3.3. Constraints Available From NMR Spectroscopy

3.3.1. Distance constraints from Nuclear Overhauser Enhancements

3.3.1.1. Quantitative interpretation of NOEs

Given a molecule with known distance relations, the cross peak intensities for a special mixing time, $\mathbf{V}(\tau_m)$, in a 2D NOESY spectrum can be recalculated from known diagonal peak intensities $\mathbf{V}(0)$ at mixing time τ_m=0 by:

$$\mathbf{V}(\tau_m) = \exp(-\mathbf{R}\tau_m)\,\mathbf{V}(0) \qquad [1]$$

Here, **R** is the relaxation matrix whose elements are defined by:

$$R_{ij} = \sigma_{ij} = \frac{\gamma_i^2 \cdot \gamma_j^2 \cdot h^2}{40 \cdot \pi^2 \cdot r_{ij}^6} \cdot [6J_{2,ij}(\omega) - J_{0,ij}(\omega)] \qquad [2]$$

$$R_{ii} = \rho_i = \frac{\gamma_i^2 \cdot \gamma_j^2 \cdot h^2}{40 \cdot \pi^2 \cdot r_{ij}^6} \cdot \sum_{i,j} [J_{0,ij}(\omega) + 3J_{1,ij}(\omega) + 6J_{2,ij}(\omega) + A_{1,i}] \qquad [3]$$

$A_{1,i}$ is the contribution by non-dipolar relaxation mechanisms which can be neglected in the absence of paramagnetic nuclei.

Cross and diagonal elements of the relaxation matrix are a function of the spectral density *J* (in order to distinguish notations for a spectral density and a coupling constant, we denote the spectral density in italics) which describes the frequency dependence of a motion:

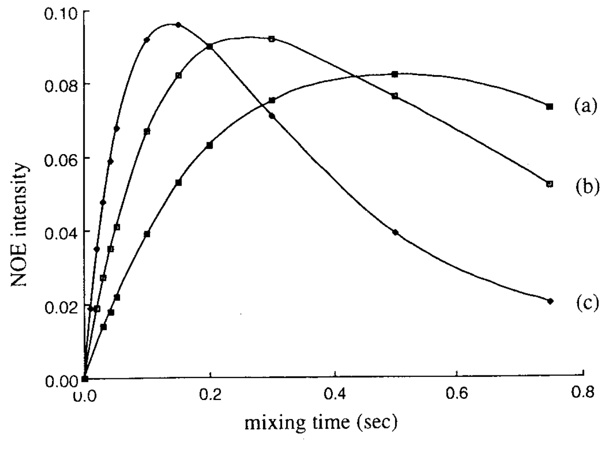
$$J_{n,ij}(\omega) = \frac{\tau_{ij}}{\left(1 + n^2\omega_0^2\tau_{ij}^2\right)} \qquad [4]$$

Since spectral densities are dependent on the correlation time τ_{ij} of the interatomic vector, the assumption of one single correlation time valid for all nuclei in the molecule is introduced.

Several approaches to quantify distances from NOE-signal intensities exist:

- Isolated spin pair approximation (ISPA);
- Direct comparison of cross and diagonal peaks (DIRECT);
- Reference structure based iterative methods (IRMA, MARDIGRAS).

The figure shows the theoretical time dependence of the cross-peak intensity (H^{δ},H^{ε} in tyrosine) calculated for different correlation times [2 ns (a), 4 ns (b), 8 ns (c)].



Isolated Spin Pair Approximation:

Although the relationship between NOE-peak intensity and distance is nonlinear, the time course can be linearized under the conditions (Gronenborn & Clore, *Progr.NMR Spectr.* (1985) 17):

- short mixing time;
- short correlation time.

Truncation of a Taylor series expansion for [1] after the second term mathematically describes this linear relationship:

$$\exp(-\boldsymbol{R}\tau_m) \approx 1 - \boldsymbol{R}\tau_m + \frac{\boldsymbol{R}^2\tau_m^2}{2} - \ldots + \left[\frac{(-1)^n}{n!}\right]\boldsymbol{R}^n\tau_m^n \approx \boldsymbol{V}(t_m) \qquad [5]$$

The cross-peak intensities depend only on the corresponding off-diagonal element in **R**.

Calibration of the unknown interproton distances is then possible using the peak intensity and distance of a proton pair with a fixed distance r_{ref} as reference by:

$$r_{ij} = r_{ref} \cdot \left(\frac{V_{ref}}{V_{ij}}\right)^{\frac{1}{6}} \qquad [6]$$

Reference distances may be:

- geminal proton-proton distances (β-methylene groups):	1.78 Å;
- aromatic vicinal proton-proton distances (H^{δ},H^{ε}):	2.48 Å;
- H^{α}_i-H^{N}_i distance in a α-helix (invariant for α and 3_{10}):	2.70 Å;
- H^{α}_i-H^{N}_i distance in a β-sheet (invariant for ↑↑ and ↑↓):	2.80 Å;
- H^{5}_i-H^{6}_i distance in cytosine:	2.50 Å.

The ISPA method shows a small error for long-range distances and, therefore, can be used to reveal secondary structure elements and tertiary structures. For structural problems mainly defined by local NOE's, like nucleotide conformations (see table for B-DNA results), this methods is too error prone. In general, distances smaller than the reference are underestimated, others are overestimated by a factor of 5-6 between the percentages in distance and intensity error.

A solution for errors arising from inadequate reference distances is the utilization of two different references spanning a wider distance range and a modified formula [7] for calculation of distances.

$$r_{ij} = \frac{r_{ref1}V_{ref1}^6 - r_{ref2}V_{ref2}^6}{V_{ref1}^6 - V_{ref2}^6} + \left[\frac{r_{ref1} - r_{ref2}}{V_{ref1}^{-6} - V_{ref2}^{-6}}\right] \cdot (V_{ij})^{-\frac{1}{6}} \qquad [7]$$

Relaxation Matrix Approaches:

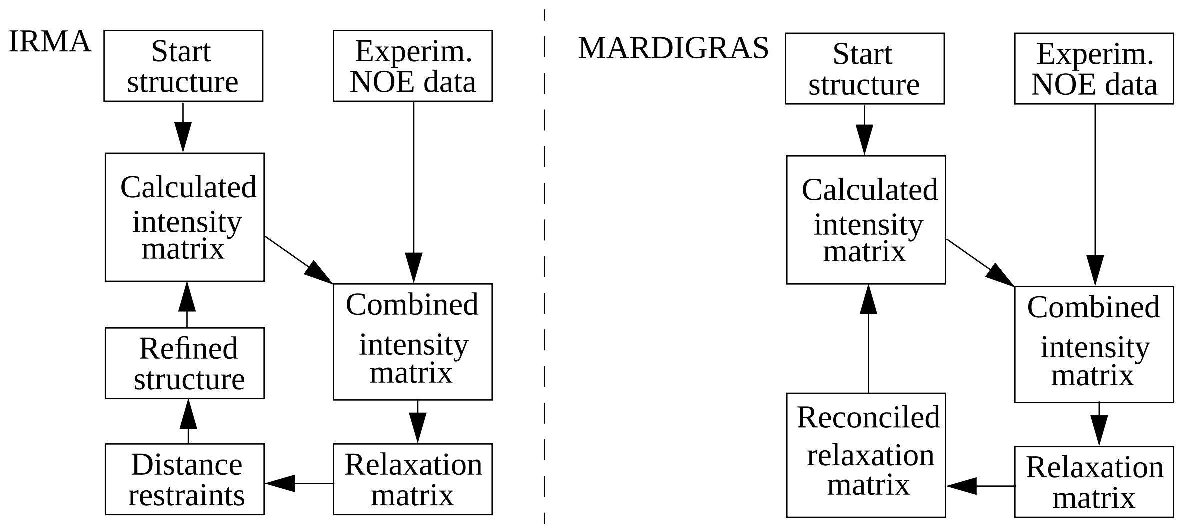
The **DIRECT** method (Olejniczak et al., *J.Magn.Reson.* (1986) 67, 28-41) implies that the (cross- and diagonal-peak) intensities for all protons can be identified and accurately quantified. Rearrangement of [1] allows to obtain the distance estimates directly from the intensity matrix:

$$\boldsymbol{R} = \frac{-(\ln \boldsymbol{V}(\tau_m) \cdot \boldsymbol{V}(0)^{-1})}{\tau_m} \qquad [8]$$

The diagonal-peak intensities at mixing time = 0 can be derived by extrapolating a series of NOE spectra run at different mixing times (**build-up rate** measurements).

Under typical experimental conditions for bigger biomolecules not all diagonal peaks are resolved and cross peaks may be disturbed by overlaps and close proximity to the diagonal. Hence, the ideal and complete relaxation matrix representing the spatial arrangement of all protons will not be evaluable.

Iterative algorithms are designed to cope with incomplete intensity informations and are based on **formula [8]**. All approaches aim to minimize the deviation between measured an calculated intensities while they refer to a proposed **structural model**. The programs differ in the modified subject to reach agreement between calculated and measured intensities, as shown in the schematic.



The **MARDIGRAS** approach (Borgias & James, 1990) is one of this iterative methods in which the self-consistency of the relaxation matrix is used to reconcile experiment and theory. The following recipe is performed until a convergence criterion is satisfied.

1. From a model structure a intensity matrix is calculated by the module CORMA (Keepers & James, 1984). Here, the rate matrix **R** [9] is represented by a product of unitary matrices of orthogonal eigenvectors χ and χ^T and the matrix of eigenvalues λ. Since λ is diagonal, the series expansion for its exponential, i.e. the mixing coefficient matrix, collapses [10] and for a given mixing time, the cross-peak intensities can be calculated.

$$\boldsymbol{R} = \chi\lambda\chi^{\boldsymbol{T}} \qquad [9]$$

$$\boldsymbol{V}(t_m) = 1 - \chi\lambda\chi^{\boldsymbol{T}}\tau_m + \frac{1}{2}\chi\lambda\chi^{\boldsymbol{T}}\chi\lambda\chi^{\boldsymbol{T}}\tau_m^2 - \ldots \approx \chi e^{-\lambda\tau_m}\chi^{\boldsymbol{T}} \qquad [10]$$

2. The measured NOE intensities are used to substitute the corresponding calculated ones in the intensity matrix building a semi-experimental combined intensity matrix.

3. From the combined matrix a relaxation matrix is evaluated by [8].

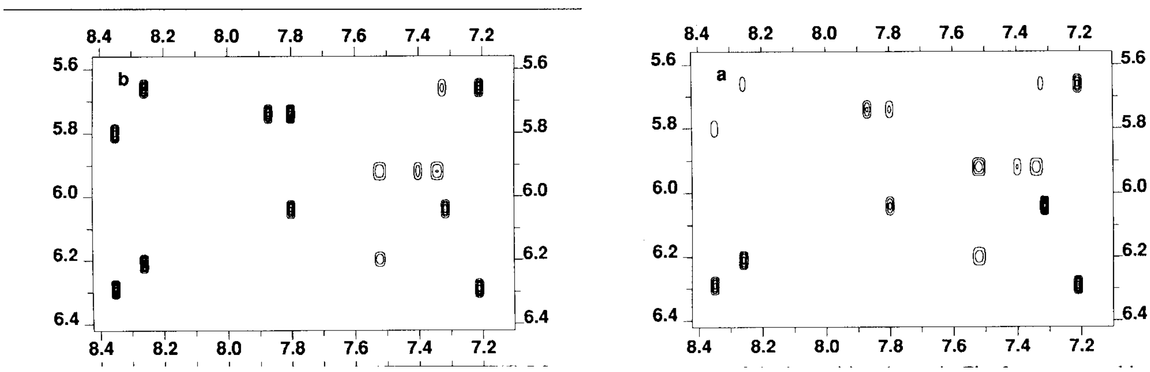
4. The relaxation matrix is checked and modified for correspondence between diagonal and off-diagonal elements, known values for off-diagonal elements (fixed distances) and upper limits for the off-diagonal elements coming from the shortest possible interproton distance.

5. A new intensity matrix is calculated.

6. Repeat 2.-5. until convergence is reached.

7. Extraction of cross-peak distances from the final relaxation matrix for DG/MD-calculations.

CORMA results (a) vs. experimental intensities (b) (Borgias & James, *Meth.Enzym.* (1989) 176,169-183).



IRMA (Boelens et al., *J.Mol.Struct.* (1988) 173, 299-311) is a program devised for an accurate determination of H-H distances from 2D NOE intensities. The use of full cross-relaxation matrices in the procedure allow a proper treatment of spin diffusion in the NOE interpretation. In contrast to MARDIGRAS it refines the starting structure rather than relying only on the relaxation matrices. Thus, in each iteration a constraints set is extracted from the relaxation matrix whose combination with the intensity matrix is performed as described under MARDI-GRAS. As shown in the figure, the constraint set is employed in a DG-/MD-refinement program and the resulting structure used as starting model for a new cycle.

The program allows to introduce a set of spectra with different mixing time in order to improve the accuracy. Also a set of structures (e.g. a MD trajectory) may be used as starting structure for a new refinement cycle since this allows to reflect influences of local mobility.
The following effects of motion can be incorporated:

- aromatic ring flips are described by averaging the cross-relaxation rate elements of protons with the exchange position (r^{-6} averaging);
- methyl rotation is faster than the molecules overall tumbling and described by a r^{-3} averaging;
- fast local motions can be included via an order parameter S^2 according to the model-free approach by Lipari and Szabo.

For a **build-up rate** or initial rate analysis a series of NOE spectra with different mixing times is processed and integrated. The intensities (x-) are plotted against the mixing time (y-axis) and under the assumption of a single-exponential behaviour the initial y_i and final values y_f are estimated. The NOE build-up is then fitted to the function

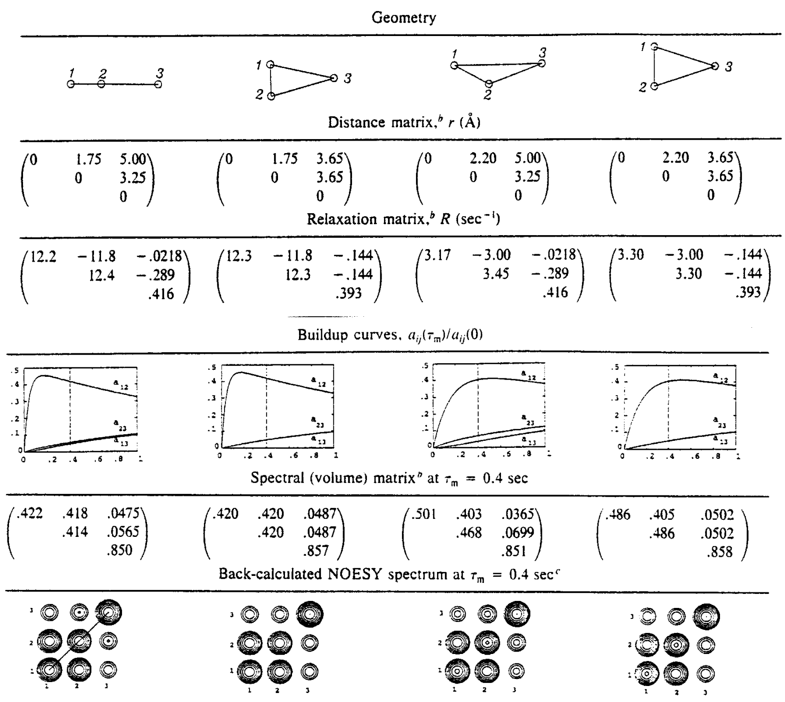
$$y(t) = y_f + (y_i - y_f) \cdot \exp[-kt] \qquad [11]$$

The graphical analysis of each cross peak allows to identify and exclude informations biased by spin diffusion or relaxation leakage at an early stage of the process, since these phenomena result in an obvious non single-exponential behaviour. For such cross peaks the higher mixing time data is then excluded from analysis.

The initial build-up rate is

$$B_{ref} = k \cdot (y_f - y_i) \qquad [12]$$

Substituting V_{ref} against B_{ref} and V_{ij} against the cross-relaxation rates B_{ij} in equation [6] allows to extract the distances.
The time course of a magnetization transfer for a three-spin system is given in the next figure. The cross peaks build up as the diagonal peaks decay. For long mixing times cross and diagonal peaks reach equal intensities and by T_1 relaxation furtheron decay to 0 (Macura et al., *Meth.Enzym.* (1994) 239,106-144).



[a] Rigid body isotropic motion with τ_c = 6.0 nsec and R_{ex} = 0.1 sec^{-1} at $\omega/2\pi$ = 500 MHz.
[b] Only upper triangles are shown since matrices are symmetric, $x_{ij} = x_{ji}$.
[c] Gaussian lines with equal widths. Contours are drawn with exponential scaling, $c(n) = I_{33}^{MAX} \times 2^{n-8}$.