Journal of Biomolecular NMR, 5 (1995) 193–201
ESCOM

J-Bio NMR 231

An estimate of spin diffusion in a spin subset: Application to iterative distance calculation from 3D ^{15}N NOESY-HMQC

Thérèse E. Malliavin[a], Marc A. Delsuc[a,*], Vladislav Yu. Orekhov[b] and Alexander S. Arseniev[b]

[a]*Centre de Biochimie Structurale, INSERM U414-CNRS UMR9955, Faculté de Pharmacie, Université Montpellier-I, 15 Avenue Ch. Flahault, F-34 060 Montpellier, France*
[b]*Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Ul. Miklukho-Maklaya 16/10, 117871 Moscow, Russia*

Received 6 June 1994
Accepted 14 September 1994

Keywords: ^{15}N-labelled protein; Heteronuclear spectroscopy; 3D NOESY-HMQC; Spin-lattice relaxation time; Relaxation matrix; Iterative distance calculation; Bacteriorhodopsin

Summary

A method for quantification of distances between amide hydrogens using only the 3D NOESY-HMQC experiment recorded on a ^{15}N-labelled protein is presented. This method is based on an approximate expression of the NOE intensities between amide hydrogens obtained from continuum modelling of the non-amide spins; this expression is used in a distance calculation algorithm. The algorithm has been named CROWD, standing for *C*ontinuum approximation of *R*elaxati*O*n path *W*ays between *D*ilute spins. This approximation as well as the CROWD algorithm are tested on a simulated case; the CROWD algorithm is then applied to experimental data, measured on a fragment of bacteriorhodopsin.

Introduction

The structure determination of large molecules by NMR relies on the observation of spectral features which can be related to geometrical information (Wüthrich, 1986). Vicinal J-coupling constants and nuclear Overhauser enhancement intensities can be measured on NMR spectra. From these parameters, dihedral angles and distance constraints are derived and used for structure determination.

The NOE intensities measured on NMR spectra are mostly used qualitatively, in the form of ranges of validity for the interspin distances associated with the observed NOE. The accuracy of the structural information thus obtained relies on the large redundant number of observations rather than on the accuracy of each individual signal. One drawback of this approach is the need to use short NOESY mixing times, to minimise the effect of the spin-diffusion phenomenon, thus hampering the use of experiments where long-range NOEs would be maximal.

Many authors have proposed using the measured NOE intensities in a more quantitative fashion (Massefski and Bolton, 1985; Boelens et al., 1988; Sobol and Arseniev, 1988; Borgias and James, 1989; Yip and Case, 1989; Koehl and Lefevre, 1990; Bonvin et al., 1991a; Madrid et al., 1991; Mertz et al., 1991; Nilges et al., 1991; Van de Ven et al., 1991; Baleja, 1992; Mirau, 1992; Lai et al., 1993; Suri and Levy, 1993) through modelling of spin diffusion by the use of the relaxation matrix. It is also possible to directly evaluate interspin distances by fitting the NOE intensity dependence upon mixing time (Boelens et al., 1989; Fejzo et al., 1989; Hyberts and Wagner, 1989; Malliavin et al., 1992). Likewise, a direct fit of the molecular structure to the observed NOE, by back-calculating the NOE pattern associated with a given structure, has been proposed (Nerda et al., 1989; Bonvin et al., 1994). Finally, another approach (Boelens et al., 1988; Borgias and James, 1989) consists in the construction of a 'mixed NOE matrix' from the experimental and calculated NOEs of any random structure. A regularisation of this matrix leads to somehow restrained validity ranges compared to the qualitative approach.

Isotopic labelling is commonly used in the study of large proteins by NMR, because this method facilitates the assignment process by using multidimensional heteronuclear experiments. In the case of ^{15}N labelling, the NOE information extracted from these experiments is much

*To whom correspondence should be addressed.

more partial than the NOE information extracted from homonuclear 2D or 3D NOESY spectra, because only proximity information on amide hydrogens is gathered. The methods developed to refine the relaxation matrix from NOE intensities cannot be used on heteronuclear 3D spectra, and usually one has to rely on regular homonuclear 2D NOESY or 3D NOESY-NOESY experiments to extract quantitative distance estimates used for the structural reconstruction (Bonvin et al., 1991b; Abergel and Delsuc, 1993; Berstein et al., 1993).

We present here a method which permits extraction of estimates of the distances between amide hydrogens from a 3D NOESY-HMQC experiment recorded on a ^{15}N-labelled protein (Clore and Gronenborn, 1991). This method is based on an approximate expression of the NOE intensities between amide hydrogens, obtained from a continuum modelling of the non-amide spins. This approximation is used in a distance calculation algorithm similar to MARDIGRAS (Borgias and James, 1989). It has been named CROWD, for *C*ontinuum approximation of *R*elaxati*O*n path *W*ays between *D*ilute spins.

The approximation requires measurement of the NOE autorelaxation amide peaks, the amide–amide cross peaks, and the amide–non-amide cross peaks, which are all readily available from the 3D NOESY-HMQC experiment. It also requires recording of the 2D HMQC experiment for the calibration of the transfer efficiency.

This approximation as well as the CROWD algorithm are tested on a simulated case. The proposed method is then applied to experimental data measured on the (1–71) fragment of bacteriorhodopsin, and the applicability and limitations of the present approach are discussed.

Theory

The NOE intensities of a 2D NOESY experiment with a mixing time τ_m can be expressed as a symmetric matrix $I(\tau_m)$, where the element (i,j) holds the intensity of the cross peak between the spins i and j. The relaxation rates of the spin system can also be presented as a matrix, the relaxation matrix R. The NOE intensity matrix $I(\tau_m)$ is related to R by the expression (Macura and Ernst, 1980; Keepers and James, 1984):

$$I(\tau_m) = \exp(-R\ \tau_m)\ I_O \qquad (1)$$

where I_O is a diagonal matrix, the elements of which are equal to the equilibrium magnetisations of the spins.

The relaxation matrix elements are equal to:

$$[R]_{ij} = c/r_{ij}^6 \qquad (2)$$

$$[R]_{ii} = 1/T_{1i} + \sum_{k \neq i} c'/r_{ik}^6 \qquad (3)$$

where r_{ij} is the distance between the spins i and j and T_{1i} is the spin-lattice relaxation rate of hydrogen i.

The coefficients c and c' depend on the spectral density function of the motion of the interproton (i,j) vector. As a first approximation, we suppose that the protein motion is governed by a unique global correlation time τ_c; the spectral density function is thus the same for all the hydrogen pairs and the coefficients c and c' can be considered as being constant.

From the expression of the relaxation matrix elements, the separation of the relaxation rates between amide hydrogens and those concerning other hydrogens (non-amide) is straightforward. The matrix R can be written as the sum of the matrices R_1 and R_2:

$$R = R_1 + R_2 \qquad (4)$$

The elements of the matrices R_1 and R_2 are defined in Table 1.

The matrix R_1 is a 'relaxation matrix' containing only the amide hydrogen subset. It should be noticed that its exponential $\exp(-R_1\tau_m)$ is also limited to the amide hydrogen subset. This property will be used below.

The matrix R_2 is what remains in R when R_1 has been removed. Thus, off-diagonal elements of R_2 hold relaxation rates of non-amide hydrogens with both amide and non-amide hydrogens. The special element $[R_2]_{ii}$, where i is an amide hydrogen, can be seen as a pseudo spin-lattice relaxation rate, which takes into account the dipolar relaxation leakage from amide hydrogen i to all non-amide hydrogens. In the following, this parameter will be noted as:

TABLE 1
EXPRESSION OF THE ELEMENTS IN MATRICES R_1 AND R_2

	Matrix R_1		Matrix R_2	
	amide hydrogen i	non-amide hydrogen m	amide hydrogen i	non-amide hydrogen m
amide hydrogen j	$[R_1]_{ij} = R_{ij}$ if $i \neq j$	$[R_1]_{mj} = 0$	$[R_2]_{ij} = 0$ if $i \neq j$	$[R_2]_{mj} = R_{mj}$
	$[R_1]_{ii} = \sum_{\substack{k \neq i \\ k \in \{HN\}}} c'/r_{ik}^6 + 1/T_{1i}$		$[R_2]_{ii} = \sum_{m \notin \{HN\}} c'/r_{im}^6$	
non-amide hydrogen n	$[R_1]_{in} = 0$	$[R_1]_{mn} = 0$	$[R_2]_{in} = R_{in}$	$[R_2]_{mn} = R_{mn}$

The amide hydrogen indexes are named i, j or k. The non-amide hydrogen indexes are named n and m.

$$\theta_2^i = 1/[R_2]_{ii} \tag{5}$$

Due to the exponential relation between relaxation rate and NOE intensity matrices, the NOE intensity between two interacting amide hydrogens is influenced by all indirect dipolar relaxation through any other set of hydrogens. This well-known phenomenon is named spin diffusion. Thus, in the NOE intensity matrix $I(\tau_m)$, the separation of amide from non-amide hydrogens is not so easy.

A model of the relaxation of amide hydrogens in a bath of non-amide hydrogens is proposed here. The spin diffusion due to the amide hydrogen network can be completely taken into account by calculating the exponential of the matrix R_1. The non-amide hydrogens also influence the relaxation pathways between the amide hydrogens; they will be modelled here as a continuum. The spin diffusion through this continuum, which is the predominant part of the total spin diffusion, is taken into account using the pseudo spin-lattice relaxation time θ_2^i.

Our purpose is thus to evaluate an approximation of the amide spin dipolar relaxation from the knowledge of the geometry of the amide spins only. We shall see from measurements of the NOE between the amide and the non-amide spins that it is possible to evaluate the part of the spin diffusion due to these non-amide hydrogens.

The approximation is derived in several steps. First, the NOE intensity matrix $I(\tau_m)$ is expanded as a function of the R_1 and R_2 matrices. Only the first expansion term is kept, to calculate an approximation of NOE intensity between amide hydrogens. The second approximation step is the expression of the NOE intensities as a function of the matrix R_1 and the spin-lattice relaxation times θ_2^i. Finally, an approximate expression of θ_2^i is derived as a function of the NOE intensities between amide hydrogen i and the non-amide hydrogens.

In the following, the amide hydrogens will be designated by the indexes i, j, k and the non-amide hydrogens by the indexes m, n, p; the indexes s and s_i will specify all protein hydrogens.

An expansion of the matrix exp(A + B) as a function of the matrices A and B is derived in the Appendix. The relation between A + B, A and B, obtained from Eq. A7 by discarding the terms of order greater than three, is applied to the matrices R, R_1 and R_2. Neglecting the terms of order greater than three means that the part of the spin diffusion concerning more than one relay is neglected. An approximation of the NOE intensity $[I]_{ij}$ between two amide hydrogens i and j can thus be written as:

$$[I]_{ij} \approx \frac{1}{2}\sum_s \left([I_1]_{is}[I_2]_{sj} + [I_2]_{is}[I_1]_{sj}\right) \tag{6}$$

where $I_1 = \exp(-R_1\,\tau_m)$ and $I_2 = \exp(-R_2\,\tau_m)$. In the following, I_1 will be named the pseudo-intensity matrix, because it is the exponential of a relaxation submatrix.

As $[I_1]_{is} = [I_1]_{sj} = 0$, for all hydrogen indexes s being non-amide hydrogens, Eq. 6 can thus be written as:

$$[I]_{ij} \approx \frac{1}{2}\sum_{k\in\{HN\}} \left([I_1]_{ik}[I_2]_{kj} + [I_2]_{ik}[I_1]_{kj}\right) \tag{7}$$

In the case of weak spin diffusion, the diagonal terms of the matrix I_2 dominate the sum of Eq. 7. From the exponential expansion, these terms can be expressed as:

$$[I_2]_{ii} = \sum_{p=0}^{+\infty} \frac{(-\tau_m)^p}{p!}[R_2^p]_{ii} \tag{8}$$

where $[R_2^p]_{ii}$ designates the (i,i) element of matrix R_2 at power p. For $p \geq 2$:

$$[R_2^p]_{ii} = \sum_{s_i} [R_2]_{is_1}[R_2]_{s_1s_2}\cdots[R_2]_{s_{p-1}i} \tag{9}$$

For increasing protein size, the slow-motion limit hypothesis ($\omega_o\tau_c \gg 1$) is usually verified, and the auto-relaxation rate $[R]_{ii}$ of hydrogen i is roughly equal to the opposite sum of the cross-relaxation rates $[R]_{is}$ between hydrogen i and other hydrogens. Thus, in the matrix R_2, the diagonal matrix elements have absolute values larger than the nondiagonal matrix elements. The huge sum of Eq. 9 is thus dominated by terms containing the largest number of diagonal elements of the matrix R_2. An approximation of $[R_2^p]_{ii}$ can now be obtained by keeping only these terms:

$$[R_2^p]_{ii} \approx ([R_2]_{ii})^p \tag{10}$$

Substituting Eq. 10 into Eqs. 7 and 8, we obtain an approximation of the NOE intensity for a cross peak or an auto peak between two amide protons:

$$[I]_{ij} \approx \frac{[I_1]_{ij}}{2}\left(e^{-[R_2]_{ii}\tau_m} + e^{-[R_2]_{jj}\tau_m}\right) \tag{11}$$

With Eq. 11 we have made a step toward obtaining an estimate of the NOE intensity for the amide spins. We now show that the remaining terms to be evaluated can be derived from measuring, on the 3D HMQC-NOESY experiment, amide–non-amide cross peaks.

An approximation for the spin-lattice relaxation rates $[R_2]_{ii}$ is first derived. From Eq. A7, the approximate expression for the intensities between amide hydrogen i and non-amide hydrogen m can be derived in the same way that Eq. 6 was derived:

$$[I]_{im} \approx \frac{1}{2}\sum_s \left([I_1]_{is}[I_2]_{sm} + [I_2]_{is}[I_1]_{sm}\right) \tag{12}$$

Using the same token as in Eq. 6, Eq. 12 can thus be written as:

$$[I]_{im} \approx \frac{1}{2}\sum_{k\in\{HN\}} [I_1]_{ik}[I_2]_{km} \tag{13}$$

As previously, only the largest term of the sum, $[I_1]_{ii}[I_2]_{im}$ is kept:

$$[I]_{im} = \frac{1}{2}[I_1]_{ii}[I_2]_{im} \quad (14)$$

By using a two-spin approximation, the elements of matrix I_2 between amide hydrogen i and non-amide hydrogen m can be approximated in the following way:

$$[I_2]_{im} \approx -[R_2]_{im}\tau_m \quad (15)$$

Using Eqs. 14 and 15, the element $[R_2]_{im}$ can thus be expressed as:

$$[R_2]_{im} \approx -\frac{2[I]_{im}}{\tau_m[I_1]_{ii}} \quad (16)$$

From the R_2 matrix element definition (Table 1), we obtain:

$$[R_2]_{ii} = \frac{c'}{c}\sum_m [R_2]_{im} \quad (17)$$

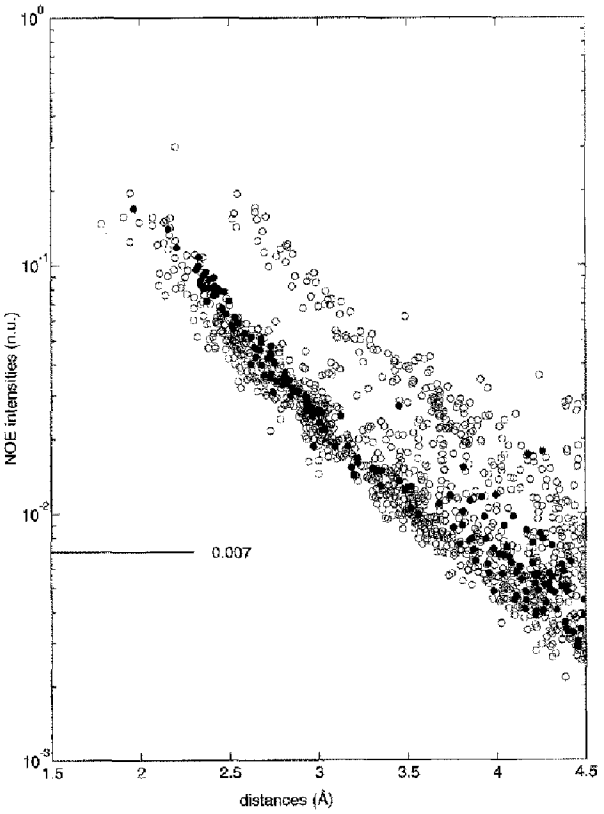

Fig. 1. Plot of the computed intensities (y-axis) as a function of the interatomic distance (x-axis) for a mixing time of 100 ms. A logarithmic scale is used for the y-axis. Only pairs of spins including at least one amide hydrogen are displayed. Pairs of amide hydrogens are indicated with closed circles (●), and pairs of hydrogens including only one amide hydrogen with open circles (○). The value of 0.007 n.u., corresponding to the mean error of the approximation for a mixing time of 100 ms, is displayed for clarity.

An approximation of the spin-lattice relaxation rate is thus:

$$[R_2]_{ii} \approx -\frac{2c'S_i}{\tau_m c[I_1]_{ii}} \quad (18)$$

where:

$$S_i = \sum_m [I]_{im} \quad (19)$$

The proposed approximation can thus be summarised in the following way:

$$[I^*]_{ij} = \frac{[I_1]_{ij}}{2}\left(e^{-[R_2^*]_{ii}\tau_m} + e^{-[R_2^*]_{jj}\tau_m}\right) \quad (20)$$

where $[I^*]_{ij}$ is the approximate intensity between the amide hydrogens i and j, and $[R_2^*]_{ii}$ is the approximate spin-lattice relaxation rate from amide hydrogen i to all non-amide hydrogens:

$$[R_2^*]_{ii} = -\frac{2c'S_i}{\tau_m c[I_1]_{ii}} \quad (21)$$

The parameter S_i can be obtained for each amide hydrogen i (Eq. 19) from the intensities $[I]_{im}$ between hydrogen i and the non-amide hydrogens m which are measured on a 3D NOESY-HMQC spectrum. Thus, approximate NOE intensities between amide hydrogens can be obtained using only distances between amide hydrogens.

Approximation test

In the following, the NOE intensities are expressed in NOE units (n.u.); this unit is defined such that a spin of multiplicity 1 gives rise to an autorelaxation peak with 1.0 n.u. intensity in a NOESY experiment recorded with a null mixing time.

To test the quality of the approximation, we chose to run the simulation on the pike parvalbumin, a calci-protein of 109 amino acids, whose secondary structure is mainly α-helical, and which also contains a small β-sheet. Its structure was determined by NMR (Padilla et al., 1988; Padilla, A. et al., manuscript in preparation) and was deposited in the Protein Data Bank (PDB code: 1PAS). The atom coordinates of a mean structure, obtained from the PDB file conformers, are used here. During the calculation, the global protein correlation time is taken equal to 5.0 ns.

The asparagine, lysine, arginine, glutamine, histidine and tryptophan side chains contain nitrogen-bound hydrogens. These hydrogens usually give few correlations with the backbone amide hydrogens on the 3D NOESY-HMQC spectrum; they are often invisible on the spectrum, because they are annealed by the water presatura-

TABLE 2
APPROXIMATION EFFICIENCY FOR DIFFERENT MIXING TIMES

Mixing time (ms)	Δ_1	Δ_2	ε_1	ε_2	E_1 (n.u.)	E_2 (n.u.)
100	0.66	0.62	0.34	0.38	0.14	0.007
200	0.53	0.46	0.48	0.54	0.08	0.013
300	0.49	0.39	0.56	0.61	0.05	0.015
400	0.49	0.36	0.62	0.65	0.03	0.015
500	0.52	0.35	0.67	0.65	0.03	0.014

The parameters ε, Δ and E are respectively the mean relative error, the mean ratio and the mean error, as described in the text. The subscripts stand for the peak set on which the calculation was performed, i.e., 2: cross peaks; 1: cross peaks and auto peaks. ε and Δ are dimensionless parameters. The parameter E is given in n.u. The NOE unit (n.u.) is defined as follows: for intensities expressed in n.u., the autorelaxation NOE signal intensity at null mixing time is equal to the multiplicity of that signal.

tion. The amino groups present in lysine residues and in the N-terminal residue are usually not observed. Thus, amine groups and side-chain amide hydrogens are not considered in our calculations.

In all calculations, performed here on theoretical data, the spin-lattice relaxation times T_{1i}, defined in Eq. 3, were supposed infinite, as reported before (Borgias and James, 1989).

All calculations involving the relaxation matrix were performed using a newly designed program. This program consists of a package of independent commands, associated to a control language, which can be used as command files. The same program can thus be used to perform a theoretical intensities calculation (as in CORMA (Keepers and James, 1984)) or an iterative distance calculation (as in MARDIGRAS (Borgias and James, 1989); or in IRMA (Boelens et al., 1988)), simply by writing the appropriate command files. This program is available from the authors upon request.

Theoretical NOE intensities were simulated from the parvalbumin structure for different mixing times: 100, 200, 300, 400 and 500 ms. The theoretical intensities were calculated on all hydrogens, using a full relaxation matrix analysis, and the parameters S_i were obtained from these intensities. Figure 1 shows the distribution of the computed intensities relative to the interatomic distances. From the structure and the parameters S_i, the approximate intensities $[I^*]_{ij}$ were obtained and compared to the real intensities $[I]_{ij}$. The $[I]_{ii}$ NOE intensities were included in the comparison, as they can be easily measured on a 3D NOESY-HMQC experiment. The accuracy of the approximation was evaluated by computing, for each pair of spins (i,j), the relative error:

$$\varepsilon = \langle |[I^*]_{ij} - [I]_{ij}| \,/\, [I]_{ij} \rangle \qquad (22)$$

as well as the ratio:

$$\Delta = \langle [I^*]_{ij} \,/\, [I]_{ij} \rangle \qquad (23)$$

and the error induced in intensities by the approximation:

$$E = \langle |[I^*]_{ij} - [I]_{ij}| \rangle \qquad (24)$$

The means are calculated on all the (i,j) spin pairs.

These quantities were evaluated for different mixing times. Data are shown in Table 2 for two peak sets: all the experimental peaks (cross peaks and auto peaks) (Δ_1, ε_1 and E_1), and only the cross peaks (Δ_2, ε_2 and E_2). The approximate intensities are compared to the exact intensities in Fig. 2, for a mixing time of 100 ms.

The fact that the quantities Δ are smaller than 1 at all mixing times shows that the proposed approximation always underestimates the NOE intensities. The Δ values decrease and the ε and E values increase with increasing mixing times, indicating that the approximation is more accurate for small mixing times than for larger ones. The error obtained on NOE cross peaks (E_2 quantity) is equal to 0.007 n.u. for a 100 ms mixing time, and is of the order of 0.015 n.u. for mixing times greater than or equal to 300 ms. From Fig. 1 it can be deduced that, for a mixing time of 100 ms, the error is negligible for spin pairs up to 3.7 Å apart.

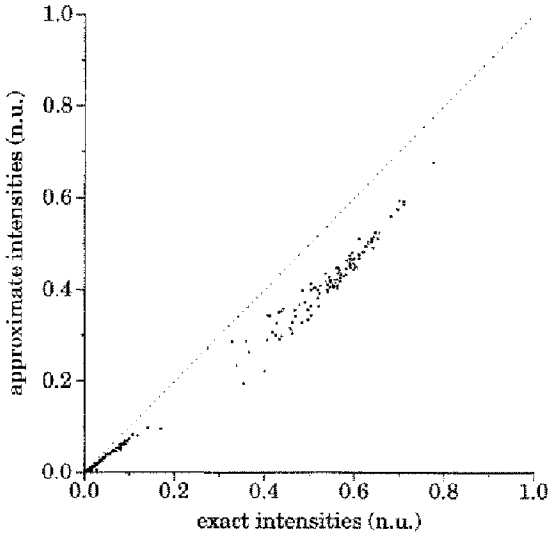

Fig. 2. Plot of the approximate intensities (y-axis) as a function of exact intensities (x-axis), for a mixing time of 100 ms. Exact and approximate intensities are calculated from the parvalbumin structure (Padilla, A. et al., manuscript in preparation).

From these data and from some other data not shown here, it is clear that the approximation works better when little spin diffusion is present, which is a direct consequence of the way it was derived. However, the proposed approximation is sufficiently robust to be applied to the calculation of distances between hydrogens.

Iterative distance calculation

A method for iterative distance calculation using the approximation is now proposed and tested in a theoretical case.

Several algorithms have been proposed to calculate iteratively distances between hydrogens from NOE intensities. Of these, the MARDIGRAS (Borgias and James, 1989) algorithm was used here, because of its simplicity and because it manipulates and produces directly relaxation rates, without using any explicit atom coordinate.

In this work, the MARDIGRAS procedure was applied in the following manner. The parameters S_i were calculated from the intensities $[I]_{im}$. The initial interproton distances were assigned in a qualitative way (Ikura et al., 1992) from the NOE intensity values: for strong NOE intensities (>0.2 n.u.), distances were randomly set between 1.8 and 2.4 Å; for medium NOE intensities (between 0.2 and 0.05 n.u.), distances were randomly set between 1.8 and 3.4 Å; for small NOE intensities (<0.05 n.u.), distances were randomly set between 1.8 and 5.0 Å. Finally, distances corresponding to hydrogen pairs for which no intensity was measured were randomly set between 4.5 Å and the protein diameter. Only the experimental intensities $[I]_{ij}$ and $[I]_{im}$ greater than a threshold value were used in the calculation. The intensity threshold used in the different simulations is shown in Table 3. For cases where a simulated noise was added to the NOE intensities, this was set equal to twice the added noise level. The initial root-mean-square (rms) deviation between initial and exact distances is larger than 1.0 Å for all calculations.

From the starting distances between hydrogens the matrices R_1 and I_1 were calculated. Using the I_1 diagonal elements and the parameters S_i, the rates $[R_2]_{ii}$ were computed from Eq. 21, and from these rates and the matrix I_1 the approximate intensities $[I^*]_{ij}$ were obtained (Eq. 20). The quality of the experimental fit was estimated by an R factor (Gonzalez et al., 1991):

$$R = \sum_{i,j} |[I]_{ij} - [I^*]_{ij}| \Big/ \sum_{i,j} |[I]_{ij}| \qquad (25)$$

If the R factor fell below a fixed threshold, the calculation was terminated. Otherwise, a mixed matrix was built by replacing the elements $[I^*]_{ij}$ by the corresponding elements $[I]_{ij}$, if these were known. From the mixed matrix and the spin-lattice relaxation rates $[R_2]_{ii}$, the matrix I_1 can be computed by inverting Eq. 20. By taking its logarithm (Massefski and Bolton, 1985), a matrix R_1' was obtained. The matrix R_1' does not have the properties of a relaxation matrix, and was regularised as proposed in the algorithm MARDIGRAS (Borgias and James, 1989). The same procedure was iterated until the R factor became smaller than the threshold. At this stage, the matrix R_1 was calculated from the current NOE intensities, as described above, and the distances between hydrogens were obtained from the R_1 nondiagonal elements.

Several simulations of the CROWD algorithm were run to check the effect of the balance between a small τ_m,

TABLE 3
ITERATIVE DISTANCE CALCULATION RESULTS FOR PARVALBUMIN

Noise level (n.u.)[a]	Mixing time (ms)	Δ[b]	Final rms (Å)[c]	Number of distances[d]	R factor[e]	Intensity threshold (n.u.)[f]
0	100	0.97	0.28	135	0.15	0.006
0	200	0.94	0.52	184	0.26	0.006
0	300	0.92	0.71	212	0.34	0.006
0.003	100	0.96	0.41	135	0.16	0.006
0.003	200	0.93	0.61	184	0.26	0.006
0.003	300	0.91	0.84	212	0.38	0.006
0.01	100	1.00	0.16	77	0.07	0.02
0.01	200	0.97	0.30	92	0.12	0.02
0.01	300	0.96	0.41	101	0.15	0.02
0.02	100	1.00	0.16	44	0.12	0.04
0.02	200	1.00	0.17	55	0.08	0.04
0.02	300	0.99	0.30	53	0.08	0.04

[a] The noise level is added to the $[I]_{ij}$ intensities and expressed in n.u.
[b] The parameter Δ is the mean ratio between final and true distances.
[c] The final rms is the rms between final and true distances.
[d] The number of distances obtained by the calculation.
[e] The final value of the R factor as obtained from Eq. 25.
[f] The intensity threshold (n.u.) is the cutoff value for the $[I]_{ij}$ intensities.

leading to an accurate approximation, and a large τ_m, permitting larger NOE intensities and thus better signal-to-noise ratios. It was also intended to check how the method would resist increasing levels of simulated noise.

Iterative calculations were performed without noise and with three different noise levels: 0, 0.003, 0.01 and 0.02 n.u. A centred Gaussian noise was simulated and added to the intensities $[I]_{ij}$ and $[I]_{im}$. For each noise level, the iterative distance calculation was performed for three mixing times: 100, 200 and 300 ms.

A histogram of the calculated NOE intensities between the amide hydrogens for parvalbumin is shown in Fig. 3, for a mixing time of 200 ms. From this figure, it is clear that a noise level of 0.003 n.u. is a mean noise level and that 0.02 n.u. is a high noise level.

The results were evaluated by calculating two parameters: the final rms between final d_{ij}^f and exact d_{ij}^r distances, and $\Delta = < d_{ij}^f / d_{ij}^r >$. The results of this comparison are summarised in Table 3.

From Table 3 it can be seen that the Δ values are always close to 1; thus, CROWD introduces little bias in the distance estimation. For the same noise level, the final rms values increase with the mixing time. Furthermore, a more detailed analysis of the results reveals that the distances between amide hydrogens closer than 3.0 Å are always precisely obtained. These distances correspond to the (i, i+1) correlations in the α-helices.

The final rms and R factor values decrease for increasing noise level, because the number of intensities used for the calculation decreases as the noise level increases, due to the fact that smaller intensities can no longer be detected. The approximation is thus applied to a hydrogen pair subset with a smaller mean distance, resulting in a better accuracy.

It should be pointed out that the first calculations using MARDIGRAS on BPTI (Borgias and James, 1989),

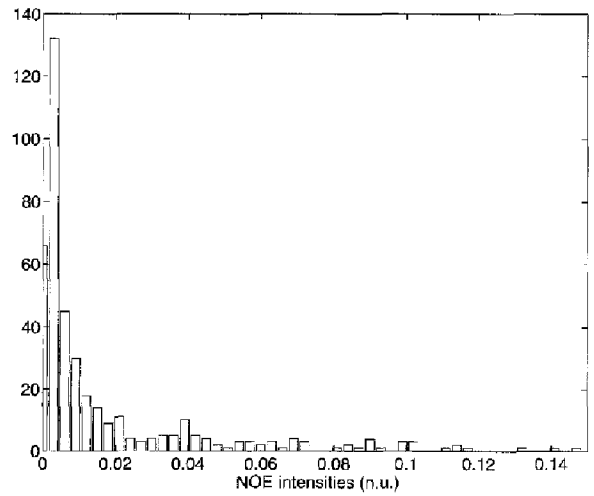

Fig. 3. Histogram of NOE intensities calculated for parvalbumin, at a mixing time of 200 ms.

with a noise level of 0.003 n.u. and a mixing time of 100 ms, provided a final rms of 0.33 Å. The calculation run here, with a mixing time of 100 ms and a noise level of 0.003 n.u., gave a final rms of 0.41 Å, which is only slightly larger.

Determination of distances in the (1–71) fragment of bacteriorhodopsin

An application of the CROWD procedure to an experimental case is now presented.

The spatial structure of a chymotryptic fragment C2 (residues 1–71) of bacteriorhodopsin, dissolved in a mixture of methanol–chloroform (1:1) and 0.1 M DCO_2/NH_4, was determined by 2D and 3D heteronuclear ^{15}N-^1H NMR techniques (Sobol et al., 1992; Pervushin et al., 1994). Distance constraints were derived from the NOE intensities observed in chloroform–methanol by using the MARDIGRAS program (Borgias and James, 1989). Twelve conformations of the (1–36) and (37–71) fragments were determined by distance geometry and simulated annealing methods using the distance constraints as described by Pervushin et al. (1994). The obtained structure has two right-handed α-helical regions from Pro^8 to Met^{32} and from Phe^{42} to Tyr^{64}. No NOE contacts between the two α-helices were found.

3D NOESY-HMQC and 2D HMQC experiments were recorded on a sample containing the C2 fragment in a mixture of methanol–chloroform. The 3D NOESY-HMQC was performed with a mixing time of 100 ms. Both experiments were acquired under equivalent conditions and were processed in the same way, using identical parameter values. The recording and processing conditions have been described elsewhere (Pervushin et al., 1994). On both spectra, peak picking and peak integration were performed using the EASY program (Eccles et al., 1991).

Several internal motions have been observed in the (1–71) fragment (Orekhov et al., 1994). The determination of spectral density functions of the amide hydrogen pairs is thus not possible in a straightforward way. Therefore, the CROWD distance algorithm was applied, assuming an isotropic rigid motion model. The protein global correlation time was taken equal to 6.6 ns.

The calculation was performed for the protein α-helical regions from Pro^8 to Met^{32} and from Phe^{42} to Tyr^{64}, because these regions are the only structured ones in the molecule (Pervushin et al., 1994).

During the CROWD calculation on the (1–71) fragment, the relaxation of the amide hydrogen with the nitrogen spins was taken into account by adding a spin-lattice relaxation rate to the amide auto-relaxation rates $[R_1]_E$.

In the amide–amide region of the 3D NOESY-HMQC spectrum, 178 peaks, corresponding to (i, i), (i, i+1) and

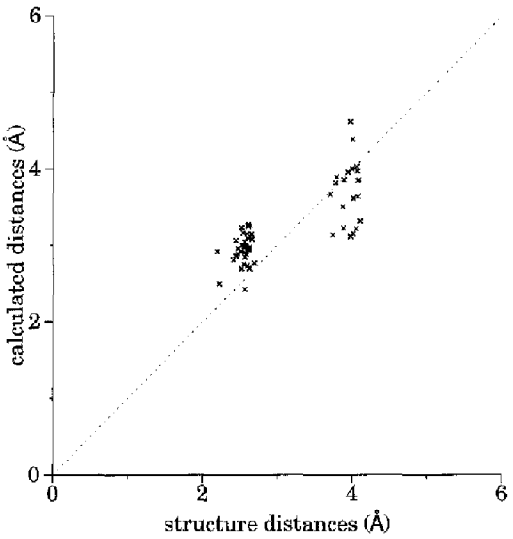

Fig. 4. Results obtained by CROWD on the (1–71) fragment of bacteriorhodopsin. Plot of the distances calculated by the CROWD method (y-axis) as a function of the distances found in the NMR structure (Pervushin et al., 1994) (x-axis).

(i, i+2) correlations in the protein α-helical regions, were detected and integrated. Fourteen peaks could not be unambiguously assigned to a single pair of amide hydrogens, due to the superposition of Gly^{16} and Gly^{23} in the HMQC spectrum, and to several superpositions of hydrogen resonances of consecutive amides. The cross-correlation peaks involved in such superpositions were not used in the CROWD calculation.

Inaccuracies due to differences in HMQC transfer efficiency, incomplete relaxation of amide signals, and differences in lineshapes were tentatively removed by dividing each cross-correlation intensity $[I]_{ij}$ between the amide hydrogen i and any other hydrogen j by the peak intensity of hydrogen i, measured on the 2D HMQC spectrum. The experimental values were then normalised to NOE units by comparing the $[I]_{ii}$ values with the simulated values calculated on the (1–71) fragment, using the same experimental parameters. The smallest 3D intensities were discarded, the dynamical range on the final normalised intensities being roughly equal to 300.

The iterative distance calculation was performed using the normalised intensities obtained from the 3D NOESY-HMQC experiment. It was stopped for an R factor equal to 0.13. Distance values were obtained for 36 (i, i+1) correlations and 21 (i, i+2) correlations.

The comparison of final distances d_{ij}^{e} with the reference distances d_{ij}^{ref} obtained from the previously determined (1–71) fragment structures is shown in Fig. 4. The distances d_{ij}^{ref} were calculated as a mean from a set of 12 conformers for the (1–36) and (37–71) fragments using the following equation:

$$d_{ij}^{ref} = (<1/(d_{ij})^6>)^{-1/6} \qquad (26)$$

The calculated distances d_{ij}^{e}, corresponding to (i, i+1) correlations, have a better correlation to d_{ij}^{ref} than the distances corresponding to (i, i+2) correlations. This discrepancy arises probably from the spectral density function used here. Indeed, for the C2 fragment it was shown (Orekhov et al., 1994) that it is not possible to fit experimental data if the spectral density function does not take into account internal motions. The mean ratio Δ between the d_{ij}^{e} and d_{ij}^{ref} distances was found to be 1.08; the iterative distance calculation slightly overestimates the distances between hydrogens.

Discussion and Conclusions

An approximate expression for NOE intensities was presented, which permits the computation of distances between amide hydrogen pairs from the evaluation of the NOE intensities measured on a 3D NOESY-HMQC spectrum. The accuracy and the efficiency of this approximation were evaluated. An iterative distance calculation algorithm, CROWD, was derived from the approximation. It was tested on theoretical data and was then applied to experimental data.

The approximation CROWD permits to calculate interproton distances by an iterative method from a dramatically incomplete NOE intensity matrix. The accuracy of the distance determination is equivalent to that obtained using other methods, as was shown above by comparing the final rms values obtained by CROWD and MARDIGRAS.

As the NOE connectivities between amide hydrogens determine α-helices more tightly than other secondary structures, a determination of distances between amide hydrogens using CROWD should be more useful for mostly helical proteins. The CROWD algorithm, used on hydrogen subsets other than the amide hydrogen subset, could also allow the processing of NOESY-HMQC spectra on a specifically ^{13}C-labelled protein.

The surprisingly good convergence of the method, despite the very crude approximation used, can be explained in the following way.

The $[I]_{ii}$ intensities cannot usually be measured on a 2D spectrum, because the diagonal is too crowded. On the other hand, on a 3D NOESY-HMQC spectrum, the $[I]_{ii}$ intensities can be determined as accurately as any other intensity. The availability of such quantities permits rescaling of the mixed intensity matrix at each iteration step. There is a weaker tendency for negative eigenvalues to occur during the inversion step of Eq. 20, whereas this appears to be a major problem in other algorithms (Pothier et al., 1993).

Finally, it is easier to accurately measure NOE intensities on a 3D HMQC-NOESY than on a 2D NOESY, because less peak superpositions are encountered on the 3D heteronuclear than on the 2D homonuclear spectrum.

Acknowledgements

We thank Dr. A.G. Sobol, Dr. P. Koehl and Dr. I. Najfeld for fruitful discussions. This work was supported by CNRS, and Université Montpellier-1. T.M. also gratefully acknowledges funding from DRET (Grant 92/1478/A000).

References

Abergel, D. and Delsuc, M.A. (1993) *J. Mol. Struct. (THEOCHEM)*, **286**, 65–70.
Baleja, J.D. (1992) *J. Magn. Reson.*, **96**, 619–623.
Berstein, R., Ross, A., Cieslar, C. and Holak, T.A. (1993) *J. Magn. Reson. Ser. B*, **101**, 185–188.
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.
Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1991a) *J. Biomol. NMR*, **1**, 305–309.
Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1991b) *J. Magn. Reson.*, **95**, 626–631.
Bonvin, A.M.J.J., Vis, H., Breg, J.N., Burgering, M.J.M., Boelens, R. and Kaptein, R. (1994) *J. Mol. Biol.*, **236**, 328–341.
Borgias, B.A. and James, T.L. (1989) *J. Magn. Reson.*, **79**, 493–512.
Clore, G.M. and Gronenborn, A.M. (1991) *Prog. NMR Spectrosc.*, **23**, 43–92.
Eccles, C., Güntert, P., Billeter, M. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 111–130.
Fejzo, J., Zolnai, Zs., Macura, S. and Markley, J.L. (1989) *J. Magn. Reson.*, **82**, 518–528.
Gonzalez, C., Rullmann, J.A.C., Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1991) *J. Magn. Reson.*, **91**, 659–664.
Hyberts, S.G. and Wagner, G. (1989) *J. Magn. Reson.*, **81**, 418–422.
Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B. and Bax, A. (1992) *Science*, **256**, 632–638.
Keepers, J.W. and James, T.L. (1984) *J. Magn. Reson.*, **57**, 404–426.
Koehl, P. and Lefevre, J.F. (1990) *J. Magn. Reson.*, **86**, 565–583.
Lai, X., Chen, C. and Andersen, N.H. (1993) *J. Magn. Reson. Ser. B*, **101**, 271–288.
Macura, S. and Ernst, R.R. (1980) *Mol. Phys.*, **41**, 95–117.
Madrid, M., Llinàs, E. and Llinàs, M. (1991) *J. Magn. Reson.*, **93**, 329–346.
Malliavin, T.E., Delsuc, M.A. and Lallemand, J.Y. (1992) *J. Biomol. NMR*, **2**, 349–360.
Massefski, W. and Bolton, P.H. (1985) *J. Magn. Reson.*, **65**, 526–530.
Mertz, J.E., Güntert, P., Wüthrich, K. and Braun, W. (1991) *J. Biomol. NMR*, **1**, 257–269.
Mirau, P.A. (1992) *J. Magn. Reson.*, **96**, 480–490.
Nerda, W., Hare, D.R. and Reid, B.R. (1989) *Biochemistry*, **28**, 10008–10021.
Nilges, M., Habazettl, J., Brünger, A.T. and Holak, T.A. (1991) *J. Mol. Biol.*, **219**, 499–510.
Orekhov, V.Yu., Pervushin, K.V. and Arseniev, A.S. (1994) *Eur. J. Biochem.*, **219**, 887–896.
Padilla, A., Cavé, A. and Parello, J. (1988) *J. Mol. Biol.*, **204**, 995–1017.
Pervushin, K.V., Orekhov, V.Yu., Popov, A.I., Musina, L.Yu. and Arseniev, A.S. (1994) *Eur. J. Biochem.*, **219**, 579–583.
Pothier, J., Gabarro-Arpa, J. and Le Bret, M. (1993) *J. Comput. Chem.*, **14**, 226–236.
Sobol, A.G. and Arseniev, A.S. (1988) *Bioorg. Khim.*, **14**, 997–1014.
Sobol, A.G., Arseniev, A.S., Abdulaeva, G.V., Musina, L.Yu. and Bystrov, V.F. (1992) *J. Biomol. NMR*, **2**, 161–171.
Suri, A.K. and Levy, R.M. (1993) *J. Magn. Reson. Ser. B*, **101**, 320–324.
Van de Ven, F.J.M., Blommers, M.J.J., Schouten, R.F. and Hilbers, C.W. (1991) *J. Magn. Reson.*, **94**, 140–151.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.
Yip, P. and Case, D.A. (1989) *J. Magn. Reson.*, **83**, 643–648.

Appendix

The exponential of a matrix A can be expressed as a sum of a series:

$$\exp(A) = \sum_{n=0}^{+\infty} \frac{A^n}{n!} \tag{A1}$$

$$\exp(A) = Id + A + \frac{A^2}{2} + \zeta(A) \tag{A2}$$

where $\zeta(A)$ contains the expansion terms of order greater than 3. The exponential of a matrix sum (A+B) can thus be expressed in the following way:

$$\exp(A+B) = Id + A + B + \frac{1}{2}(A^2 + AB + BA + B^2) + \zeta(A+B) \tag{A3}$$

By using the expressions for exp(A) and exp(B) analogous to Eq. A3, the two products exp(A) exp(B) and exp(B) exp(A) can be expressed as follows:

$$\exp(A)\exp(B) = Id + (A + B) + \frac{1}{2}(A^2 + 2AB + B^2) + \xi(A,B) \tag{A4}$$

and:

$$\exp(B)\exp(A) = Id + (A + B) + \frac{1}{2}(A^2 + 2BA + B^2) + \xi(B,A) \tag{A5}$$

where $\xi(A,B)$ and $\xi(B,A)$ contain the terms of order greater than 3. Thus, by adding Eqs. A4 and A5 we obtain:

$$\frac{1}{2}(\exp(A)\exp(B) + \exp(B)\exp(A)) = Id + (A + B) + \frac{1}{2}(A^2 + AB + BA + B^2) + \psi(A,B) \tag{A6}$$

where $\psi(A,B)$ contains the terms of order greater than 3. The comparison of Eqs. A3 and A6 gives the following expression for the NOE intensity matrix:

$$\exp(A+B) = \frac{1}{2}(\exp(A)\exp(B) + \exp(B)\exp(A)) + \omega(A,B) \tag{A7}$$

where $\omega(A,B)$ contains the terms of order greater than 3.

The matrices A and B, used for dipolar relaxation studies in NMR, are the matrices $-R_1\tau_m$ and $-R_2\tau_m$, where τ_m is the mixing time. In this case, the term $\omega(A,B)$ decreases to zero as the mixing time decreases to zero.