Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 167 (2004) 334–342
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Correction of spin diffusion during iterative automated NOE assignment

Jens P. Linge, Michael Habeck, Wolfgang Rieping, and Michael Nilges*

Unité de Bio-Informatique Structurale, Institut Pasteur, 25-28 rue du docteur Roux, F-75015 Paris, France

Received 1 July 2003; revised 14 January 2004

Abstract

Indirect magnetization transfer increases the observed nuclear Overhauser enhancement (NOE) between two protons in many cases, leading to an underestimation of target distances. Wider distance bounds are necessary to account for this error. However, this leads to a loss of information and may reduce the quality of the structures generated from the inter-proton distances. Although several methods for spin diffusion correction have been published, they are often not employed to derive distance restraints. This prompted us to write a user-friendly and CPU-efficient method to correct for spin diffusion that is fully integrated in our program ambiguous restraints for iterative assignment (ARIA). ARIA thus allows automated iterative NOE assignment and structure calculation with spin diffusion corrected distances. The method relies on numerical integration of the coupled differential equations which govern relaxation by matrix squaring and sparse matrix techniques. We derive a correction factor for the distance restraints from calculated NOE volumes and inter-proton distances. To evaluate the impact of our spin diffusion correction, we tested the new calibration process extensively with data from the Pleckstrin homology (PH) domain of *Mus musculus* β-spectrin. By comparing structures refined with and without spin diffusion correction, we show that spin diffusion corrected distance restraints give rise to structures of higher quality (notably fewer NOE violations and a more regular Ramachandran map). Furthermore, spin diffusion correction permits the use of tighter error bounds which improves the distinction between signal and noise in an automated NOE assignment scheme.

Keywords: Indirect magnetization transfer; Spin diffusion; Nuclear Overhauser effect; Relaxation matrix; ARIA

1. Introduction

The nuclear Overhauser enhancement (NOE) effect is the principal experimental source for solving structures by NMR. In an isolated rigid two-spin system, the NOE depends on r^{-6}, where r is the distance between two protons. In molecules, internal motion and spin diffusion complicate the conversion of NOEs into distances. Intramolecular mobility leads to non-linear distance averaging and to different effective correlation times for the inter-proton vectors. Magnetization transfer via indirect pathways (see Fig. 1) leads to enlarged NOE intensities in most cases. Thus, distances derived from long-range NOEs using the isolated spin pair approximation (ISPA) are often underestimated.

We present here a new algorithm to correct for spin diffusion in molecules and discuss the algorithm and its implementation in ARIA 1.2. To illustrate the impact of the new spin diffusion correction, we performed several ambiguous restraints for iterative assignment (ARIA) calculations using the data of the Pleckstrin homology (PH) domain of *Mus musculus* β-spectrin. The aim was to assess whether the spin diffusion correction influences the quality of the final structure ensemble, especially when employing tighter error bounds for the distance restraints.

1.1. Spin relaxation theory

The Solomon equations [1] in matrix form for N interacting spins are:

* Corresponding author. Fax: +33-1-45688719.
E-mail addresses: linge@pasteur.fr (J.P. Linge), habeck@pasteur.fr (M. Habeck), rieping@pasteur.fr (W. Rieping), nilges@pasteur.fr (M. Nilges).
URL: http://www.pasteur.fr/recherche/unites/binfs.

1090-7807/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2004.01.010

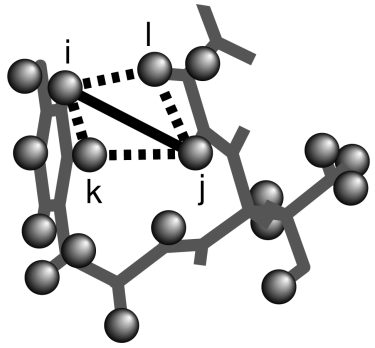

Fig. 1. Illustration of spin diffusion in a protein. Indirect magnetization transfer via spins k and l influences the NOE between i and j.

$$\frac{\mathrm{d}\Delta\mathbf{M_z}(t)}{\mathrm{d}t} = -\mathbf{R}\Delta\mathbf{M_z}(t), \tag{1}$$

in which $\Delta\mathbf{M_z} = \mathbf{M_z} - \mathbf{M_0}$ is the deviation of the longitudinal magnetization $\mathbf{M_z}$ from the equilibrium magnetization $\mathbf{M_0}$ and $\mathbf{R}$ is the relaxation matrix. The relaxation matrix is an $N \times N$ matrix with auto-relaxation rate constants $\rho_i = R_{ii}$ and cross-relaxation rate constants $\sigma_{ij} = R_{ij}$ for the exchange of magnetization.

The spectral density $J_{ij}(\omega)$ is the Fourier cosine transform of the vector auto-correlation function, $C_{ij}(t)$, between two spins i and j, respectively. If overall tumbling of the molecule is isotropic and uncorrelated to internal dynamics, the correlation function factors into contributions arising from overall tumbling, $C_{\mathrm{O}}(t) = \exp\{-t/\tau_{\mathrm{c}}\}$, and internal dynamics, $C_{\mathrm{I}}^{ij}(t)$. τ_{c} denotes the rotation correlation time of the molecule [1,2]. In the case of negligible internal dynamics, the spectral density becomes:

$$J_{ij}(\omega) = 2\int_0^\infty C_{ij}(t)\cos(\omega t)\,\mathrm{d}t = \frac{1}{4\pi r_{ij}^6}\frac{\tau_{\mathrm{c}}}{1+(\omega\tau_{\mathrm{c}})^2}. \tag{2}$$

For a uniformly labelled molecule, the elements of the relaxation matrix R_{ij} are:

$$\begin{aligned} R_{ii} = {} & \frac{2\pi}{5}\gamma_H^4\hbar^2\sum_{i\neq j}^{n}\left[J_{ij}(0) + 3J_{ij}(\omega_H) + 6J_{ij}(2\omega_H)\right] \\ & + \frac{2\pi}{5}\sum_S \gamma_H^2\gamma_S^2\hbar^2\sum_{k_S}\left[J_{ik_S}(\omega_H - \omega_S)\right. \\ & \left. + 3J_{ik_S}(\omega_H) + 6J_{ik_S}(\omega_H + \omega_S)\right] + R_i^{\mathrm{ext}}, \end{aligned} \tag{3}$$

$$R_{ij} = \frac{2\pi}{5}\gamma_H^4\hbar^2\left[-J_{ij}(0) + 6J_{ij}(2\omega_H)\right] \quad i \neq j, \tag{4}$$

where k_S is a heteronuclear spin of type S, ω_H is the proton frequency, and γ_H and γ_S are the gyromagnetic ratios for H and S, respectively. The external relaxation rate R_i^{ext} for spin i represents interactions with other spins in the sample. With Eq. (2), the cross-relaxation rates R_{ij} are a function of the distances r_{ij} and the correlation time τ_{c}:

$$R_{ij} = \frac{1}{10}\gamma_H^4\hbar^2\frac{1}{r_{ij}^6}\left(-\tau_{\mathrm{c}} + \frac{6\tau_{\mathrm{c}}}{1+(2\omega_H\tau_{\mathrm{c}})^2}\right). \tag{5}$$

The knowledge of the relaxation matrix would therefore permit direct calculation of the inter-proton distances.

The time evolution of the intensities $V_{ij}(t)$ in a two-dimensional NOE experiment [2,3] is:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{V}(t) = -\mathbf{R}\mathbf{V}(t). \tag{6}$$

The formal solution of this differential equation,

$$\mathbf{V}(\tau_{\mathrm{m}}) = \exp\{-\mathbf{R}\tau_{\mathrm{m}}\}\mathbf{V}(0) \tag{7}$$

allows the computation of NOE spectra at mixing time τ_{m} from a given biomolecular structure.

For heteronuclear NOESY experiments, different transfer efficiencies during INEPT and reverse INEPT durations must be taken into account:

$$V_{ij} = M_z^j(0)[\exp(-R\tau_{\mathrm{m}})]_{ij}F_i, \tag{8}$$

where F_i represents the transfer efficiencies for the INEPT and reverse INEPT processes [4]. Practically, the relative size of F_i can be estimated from the HSQC peak, and a volume V_{ij} can be corrected by dividing it by the volume of the corresponding HSQC peak [4,5].

1.2. Spin diffusion correction

Several methods exist to account for spin diffusion. In the direct approach, one minimizes the difference between the calculated NOE intensity and the experimental NOE intensity [6]. Unfortunately, the calculation of the gradient of the NOE energy term is CPU-intensive ($\mathcal{O}(N^3)$ for every cross-peak). Even with the speed-up achieved by using cut-off distances and other approximations [7–10], these methods are still considerably slower than using distance restraints.

Inversion of Eq. (7) yields spin diffusion corrected distance restraints for standard structure calculation [11,12]. Since the NOE matrix can only be inverted if all of its elements are known, we have to complement the NOE matrix elements for which there is no experimental data available or the NOE is ambiguous (cf. IRMA [11] and MARDIGRAS [12]). These methods are $\mathcal{O}(N^3)$, but need to be applied only very few times.

Numerical integration methods are the fastest means to correct for spin diffusion. Several schemes to solve Eq. (7) exist (see comparison in [13]). Numerical integration is also $\mathcal{O}(N^3)$. However, it is straightforward to use sparse matrix techniques rendering numeric integration suitable for larger systems. The relaxation matrix $\mathbf{R}$ is sparse because of the r^{-6} distance dependence of the cross-relaxation rates. Only few elements of $\mathbf{R}$ are significantly larger than zero. In order to speed up the NOE calculation process, one can thus set

cross-relaxation rates to zero if the distances are larger than a certain cut-off.

2. Materials and methods

2.1. Rapid calculation of an NOE matrix from a structure ensemble

For small time increments Δt, we can approximate the differential Eq. (6) for the NOE matrix $\mathbf{V}$ by a difference equation:

$$\frac{\mathbf{V}(t+\Delta t)-\mathbf{V}(t)}{\Delta t} = -\mathbf{R}\mathbf{V}(t). \tag{9}$$

The NOE matrix at time $t+\Delta t$ becomes:

$$\mathbf{V}(t+\Delta t) = (\mathbf{I}-\mathbf{R}\Delta t)\mathbf{V}(t) \tag{10}$$

in which $\mathbf{I}$ is the unity matrix. By iteration, the spectrum at mixing time τ_m can be built up from the NOE at $\tau_m = 0$:

$$\mathbf{V}(\tau_m) = (\mathbf{I}-\mathbf{R}\Delta t)^K \mathbf{V}(0), \tag{11}$$

where $K\Delta t = \tau_m$. For K a power of 2, we can evaluate $\mathbf{V}$ in only $\log_2 K$ operations by repeated squaring of $\mathbf{I}-\mathbf{R}\Delta t$.

In case of complete relaxation during the relaxation delay, $\mathbf{V}(0)$ is a diagonal matrix, $V_{ii} = n_i$, where n_i is the occupancy of the spin i (e.g., $n_i = 3$ for a methyl group; cf. [12]).

We calculate the relaxation matrix from a structure ensemble (e.g., from the previous iteration within an ARIA calculation). Due to the large distances dependence of its elements, most off-diagonal elements of the relaxation matrix are close to zero. Thus, a distance cut-off is a good approximation:

$$R_{ij} = \begin{cases} n_i \frac{1}{10}\gamma^4\hbar^2 \frac{1}{d_{ij}^6}\left(-\tau_c + \frac{6\tau_c}{1+4\omega^2\tau_c^2}\right) & \hat{d}_{ij} \leqslant d_{\text{cut}},\ i \neq j, \\ 0 & \hat{d}_{ij} > d_{\text{cut}},\ i \neq j, \end{cases} \tag{12}$$

where in our implementation we usually set the distance $\hat{d}_{ij}$ to the arithmetic ensemble average

$$\hat{d}_{ij} = \frac{1}{S}\sum_{s=1}^{S} d_{ij,s} \tag{13}$$

and S denotes the number of structures in the ensemble. If the structure ensemble had physical reality and represented slowly interchanging conformers, the correct average for calculating the relaxation matrix would be the $\langle r^{-6}\rangle^{-1/6}$ average. We have chosen the arithmetic average to be consistent with the standard ARIA calibration, where the arithmetic average is employed to avoid too strong weighting to the shortest distance [14].

The sparse matrix multiplication technique employed in our algorithm sets those elements of the result matrix to zero for which $R_{ij} = 0$—these matrix elements are never calculated. In this way, the sparse matrix multiplication reduces the complexity from N^3 to Nq^2, where N is the total number of spins and q is the typical number of spins within the cutoff d_{cut}. A single squaring step will therefore only contain the protons i and k, say, if the corresponding relaxation matrix element R_{ik} element was non-zero, and so forth for all exponents. The consequence is that the method only calculates a non-zero NOE volume between two protons if the corresponding relaxation matrix element is non-zero. However, it is exactly the property of spin diffusion that it can produce finite NOEs through intervening "layers" of spins. One could either choose large cut-offs to alleviate this problem, or use a sparse matrix technique that would apply a cut-off criterion on the result of each squaring step. We have chosen a different solution, where we try to include the most important potential spin diffusion pathways over several atoms $k, l, \ldots$ (see Fig. 2) in the calculation. We achieve this by modifying the cut-off criterion in such a way that a relaxation element between two spins i and j is non-zero and included in the calculation if an intervening atom k is close to both of them:

$$R_{ij} \neq 0 \quad \text{if } \left(\exists k | \hat{d}_{ik} \leqslant d_{\text{cut}} \text{ and } \hat{d}_{kj} \leqslant d_{\text{cut}}\right), \tag{14}$$

similarly for a third layer involving two intermediate protons k, l:

$$R_{ij} \neq 0 \quad \text{if } \left(\exists k,l | \hat{d}_{ik} \leqslant d_{\text{cut}} \text{ and } \hat{d}_{kl} \leqslant d_{\text{cut}} \text{ and } \hat{d}_{lj} \leqslant d_{\text{cut}}\right) \tag{15}$$

and so on. This distance criterion needs only to be applied once on the average distances calculated from the structure ensemble. We set the number of layers usually to three and the distance cut-off to 5 Å.

2.2. Calibration of distance restraints with spin diffusion correction in ARIA

The integration of the relaxation matrix calculation into the ARIA procedure is straightforward: Structures

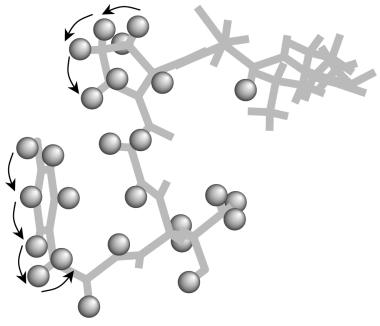

Fig. 2. Simulation of a spin diffusion pathway by applying the distance cut-off several times.

of an iteration serve as templates for the calculation of NOE intensities. These are then used as correction factors in the determination of target distances for the next iteration. The necessary steps are:

1. Calculate matrix of all average inter-proton distances $\hat{d}_{ij}$;
2. Apply the distance cut-off criterion iteratively;
3. Calculate relaxation matrix **R**;
4. Calculate NOEs.

We then derive an approximate correction factor for every distance by applying a two-step calibration (similar to [15–17]):

1. We first apply a standard calibration to the experimental NOEs, using all observed peaks, to obtain V_k^{calib}:

$$V_k^{\text{calib}} = \frac{\sum_n (V_n^{\text{th}})}{\sum_n (V_n^{\text{exp}})} V_k^{\text{exp}}, \tag{16}$$

   where n runs over all experimental NOEs, which may be ambiguous. In that case, V_k^{th} is the sum of all V_{ij}^{th} contributing to the ambiguous NOE k.
2. We then use the ratio of a calculated volume and its ISPA approximation to correct the target distance:

$$d_k^{\text{exp}} = \left(V_k^{\text{calib}} \frac{\hat{d}_k^{-6}}{V_k^{\text{th}}} \right)^{-1/6}, \tag{17}$$

where $\hat{d}_k^{-6}$ is the sum of all $\hat{d}_{ij}^{-6}$ contributing to NOE k.

By default, ARIA uses the calculated NOE intensities also to determine the contribution of each assignment possibility. We define the contribution of each possible assignment to be proportional to the NOE intensity:

$$C_a \propto V_a^{\text{th}}, \tag{18}$$

where we normalize the contributions C_a such that:

$$\sum_{a=1}^{N_\delta} C_a = 1, \tag{19}$$

where N_δ is the total number of contributions for a given frequency tolerance δ and the contributions are ordered according to size such that $C_a \geqslant C_b$ if $a < b$. We then partially assign NOEs by using Eqs. (18) and (19) and the criterion:

$$\sum_{a=1}^{N_p} C_a > p \tag{20}$$

in which p is the assignment cut-off and N_p is the number of contributions to the peak necessary to exceed a given value of p. In other words, we remove the smallest contributions depending on the choice of the cut-off parameter p which is adjusted in each iteration. By default, p decreases from 1.0 in the first iteration to 0.8 in the last iteration.

ARIA usually estimates the error of each distance from the size of the distance. Since most of the target distances will be corrected to larger values by the spin diffusion correction, this will also increase the error estimate. In order to take advantage of the improved quality of the target distances, we introduce the option to base the error estimate on the calibrated experimental volume:

$$\Delta^- = \Delta^+ = \sum_{i=0}^{3} \epsilon_i \left(V_{\text{calib}}^{-1/6} \right)^i \tag{21}$$

with $\epsilon_2 = 0.125$ and $\epsilon_0 = \epsilon_1 = \epsilon_3 = 0$ by default. The net effect of using volume-based error bounds is to shift the whole interval $[L, U]$ to larger volumes but not to increase the interval itself.

2.3. Structure calculations

We used ARIA 1.2 [18] for all calculations. We employed the standard simulated annealing protocol [14] with 8000 steps during the hot stage, 10,000 and 8000 steps during first and second cooling steps, respectively. The timestep was 3 fs for Cartesian and 27 fs for torsion angle dynamics.

2.4. NMR data for the test calculations

The PH domain is a good test for the spin diffusion correction: Since the number of NOEs was comparatively small and the convergence of the simulated annealing protocol is low for the PH domain, the calibration of the distance restraints is especially important for the quality of the final ensemble.

We took the NOE data of a 106-residue construct measured by Macias et al. [19,20] from PDB entry 1MPH. Two 2D NOE spectra with mixing times of 30 and 80 ms recorded on a 600 MHz spectrometer were available. Exactly as in the original structure calculation [20], we used three different NOE lists: a manually assigned and calibrated restraint list (containing 568 restraints from both spectra), a partially assigned list for the 2D NOE spectrum with a mixing time of 30 ms (658 restraints), and a partially assigned list for the 2D NOE spectrum with a mixing time of 80 ms (1643 restraints). The manual list contained generous qualitative upper bounds derived from visual inspection of the original 2D spectra. Most of these restraints were redundant with the other peak lists; some additional restraints on this list had been derived from peak shoulders, and do not appear on the (automatically picked) peak lists. Furthermore, we employed 88 hydrogen bond restraints in all calculations.

2.5. Overview of the performed calculations

We set up a full ARIA 1.2 run with eight iterations. The final ensemble consisted of 100 structures. We used a distance cut-off of 5 Å for the setup of the relaxation matrix and a correlation time of 8.5 ns. In order to

follow spin diffusion pathways, the cut-off was applied three times (as shown in Fig. 2). The NOE matrix was calculated in 10 matrix squaring steps, corresponding to 1024 time increments Δt. Since both NOE spectra are homonuclear, we did not need to apply the heteronuclear correction in Eq. (8).

We performed 12 calculations in total, using both spectra for calculations 1–6 and the 80 ms 2D NOE for calculations 7–12 (cf. Table 1). Calculations 1–3 and 7–9 used wide error bounds with $\epsilon_2 = 12.5\%$ (cf. Eq. (21) and [14] for the calculation of the error bounds). The error bounds in calculations 4–6 and 10–12 were tighter, with $\epsilon_2 = 6.25\%$. The goal was to test the three calibration methods:

1. Calibration without spin diffusion correction.
2. Calibration with spin diffusion correction (error estimates from calculated distances).
3. Calibration with spin diffusion correction (error estimates from calculated volumes).

Since spin diffusion is more prominent in experiments with longer mixing times, the difference between the structures calculated with and without spin diffusion correction should be larger in calculations 7–12.

2.6. Structure validation

From the ensemble of 100 calculated structures of the last iteration, we chose the 20 best structures regarding total energy for structure validation. We used the program WHATCHECK [21] to calculate *Z*-scores for the Ramachandran map (RAMCHK), the packing quality (QUACHK and NQACHK), the χ_1–χ_2 correlation (C12CHK), and the backbone conformation (BBCCHK). We also determined RMS *Z*-scores for the bond lengths (BNDCHK), bond angles (ANGCHK), omega angles (OMECHK), side chain planarity (PLNCHK), improper dihedral angles (HNDCHK), and the inside/outside distribution (INOCHK). We also counted the number of inter-atomic bumps (BMPCHK), unsatisfied H-bond donors (BH2CHK), and acceptors (BA2CHK).

We also quote the content of residues with ϕ–ψ-values in the most favoured, additional allowed, generously allowed, and disallowed regions of the Ramachandran plot [22] as determined with PROCHECK [23]. We employed the program PROSA-II [24] to determine mean force potentials, averaged over all residues in a structure. We calculated the RMS differences to the X-ray structures with CNS [25] (N, C_α, and C atoms for the backbone or all heavy-atoms). For fitting the structure, we used the amino acid ranges 2–11, 24–31, 34–38, 41–46, 62–65, 75–79, 85–89, and 93–104 to exclude flexible parts in the structure and regions with significant differences between the (complexed) X-ray crystal structure and the solution NMR structure.

We validated all structures separately, and averaged their quality indices over the ensemble of 20 structures. The number quoted in Tables 2 and 3 are the average values and the standard deviations over the ensemble.

Table 1
Overview of the PH domain calculations performed with and without spin diffusion correction

No.	Spectra	Correction	ϵ_2	VOL/DIST
1	PH30, PH80	No	12.5	DIST
2	PH30, PH80	Yes	12.5	DIST
3	PH30, PH80	Yes	12.5	VOL
4	PH30, PH80	No	6.25	DIST
5	PH30, PH80	Yes	6.25	DIST
6	PH30, PH80	Yes	6.25	VOL
7	PH80	No	12.5	DIST
8	PH80	Yes	12.5	DIST
9	PH80	Yes	12.5	VOL
10	PH80	No	6.25	DIST
11	PH80	Yes	6.25	DIST
12	PH80	Yes	6.25	VOL

We used the manual list in all calculations, and did not use the peak list from the 30 ms NOESY (PH30) in calculations 7–12. The second polynomial coefficient (ERR2) was either 12.5 or 6.25%. Either the volumes or the distances (VOL/DIST) served as reference to calculate lower and upper bounds for the distance restraints.

3. Results and discussion

Fig. 3 shows the calibrated target distances against distances from the X-ray structure 1BTN [26]. Calibration with the ISPA systematically underestimates target distances. The largest distances are around 3.5 Å. The spin diffusion correction leads to a more realistic distribution of the target distances up to 5 Å.

When the error estimates are determined from the volumes, the error bounds get tighter leading to a to a larger number of NOE violations. Many target distances are increased by the spin diffusion correction (see Fig. 3). For distance-based error estimates, also the error bounds become wider and the net effect is that the upper bound is shifted towards larger values but not the lower bound. On the other hand, when the estimation of errors is based on the peak volume, the allowed ranges are similar to not using spin diffusion correction, and the effect is that both upper and lower bound are shifted towards larger values, and that the upper bound is shifted less than with the distance-based error estimate.

The most important quality criterion in our study is the number of NOE violations in the final structure ensemble. In calculations 1–12, calibration with spin diffusion correction systematically gave the best results. Calculations 10–12 demonstrate the influence of the spin diffusion correction best, with $\epsilon_2 = 6.25\%$ and the 80 ms NOE spectrum. Whereas we have 284.1 NOE violations larger than 0.1 Å with the ISPA, spin diffusion corrected target distances give rise to a significantly smaller

Table 2
Quality indices for calculations 1–6 (all abbreviations are explained in Section 2.3)

	1	2	3	4	5	6
WHATCHECK Z-scores						
First generation packing quality (QUACHK)	-1.3 ± 0.1	-2.0 ± 0.2	-1.6 ± 0.2	-0.9 ± 0.1	-1.5 ± 0.2	-1.2 ± 0.2
Second generation packing quality (NQACHK)	-3.0 ± 0.3	-3.2 ± 0.3	-2.9 ± 0.3	-2.7 ± 0.3	-2.7 ± 0.3	-2.6 ± 0.3
Ramachandran plot appearance (RAMCHK)	-4.5 ± 0.4	-4.4 ± 0.2	-4.6 ± 0.4	-5.3 ± 0.3	-4.6 ± 0.3	-4.5 ± 0.2
χ_1–χ_2 rotamer normality (C12CHK)	-1.1 ± 0.4	0.3 ± 0.5	-0.3 ± 0.3	-2.7 ± 0.3	-1.3 ± 0.3	-2.2 ± 0.5
Backbone conformation (BBCCHK)	-2.5 ± 0.6	-2.0 ± 0.7	-2.5 ± 0.6	-4.5 ± 0.5	-2.4 ± 0.7	-2.4 ± 0.4
WHATCHECK RMS Z-scores						
Bond lengths (BNDCHK)	0.194 ± 0.006	0.177 ± 0.003	0.2 ± 0.003	0.468 ± 0.013	0.234 ± 0.004	0.291 ± 0.009
Bond angles (ANGCHK)	0.41 ± 0.004	0.388 ± 0.003	0.412 ± 0.003	0.628 ± 0.012	0.451 ± 0.007	0.494 ± 0.011
Omega angle restraints (OMECHK)	0.11 ± 0.01	0.06 ± 0.01	0.11 ± 0.01	0.26 ± 0.01	0.17 ± 0.01	0.2 ± 0.01
Side chain planarity (PLNCHK)	0.07 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.18 ± 0.02	0.09 ± 0.01	0.11 ± 0.01
Improper dihedral (HNDCHK)	0.18 ± 0.01	0.15 ± 0.0	0.18 ± 0.01	0.4 ± 0.02	0.23 ± 0.01	0.29 ± 0.01
Inside/outside (INOCHK)	0.98 ± 0.02	0.98 ± 0.02	0.99 ± 0.02	0.96 ± 0.01	1.0 ± 0.02	0.98 ± 0.02
Inter-atomic bumps (BMPCHK)	28.5 ± 5.0	15.5 ± 4.2	28.2 ± 3.8	62.8 ± 4.1	36.7 ± 5.5	41.1 ± 5.0
Unsatisfied hydrogen donors (BH2CHK)	15.4 ± 2.9	16.4 ± 3.5	17.9 ± 4.1	17.9 ± 3.2	16.2 ± 3.4	15.1 ± 3.0
Unsatisfied hydrogen acceptors (BA2CHK)	0.2 ± 0.4	0.5 ± 0.6	0.4 ± 0.6	1.6 ± 1.1	0.4 ± 0.8	0.6 ± 0.8
PROCHECK results						
Most favoured regions	70.7 ± 3.1	67.1 ± 3.4	69.0 ± 4.0	65.0 ± 2.3	70.3 ± 2.9	71.1 ± 3.9
Allowed regions	25.5 ± 2.5	28.0 ± 3.1	25.7 ± 3.3	27.6 ± 2.3	25.5 ± 1.8	24.4 ± 4.1
Generously allowed regions	2.0 ± 1.3	3.3 ± 1.3	3.8 ± 1.9	5.5 ± 2.0	3.0 ± 1.9	2.6 ± 1.1
Disallowed regions	1.8 ± 1.0	1.6 ± 1.0	1.5 ± 0.8	2.0 ± 1.4	1.2 ± 1.2	1.9 ± 1.1
PROSA-II mean force energy	-1.07 ± 0.08	-1.11 ± 0.09	-1.12 ± 0.1	-1.19 ± 0.06	-1.06 ± 0.09	-1.1 ± 0.08
RMSD to 1BTN (bb)	0.82 ± 0.04	0.79 ± 0.08	0.83 ± 0.07	0.86 ± 0.02	0.92 ± 0.05	0.83 ± 0.05
RMSD to 1BTN (all)	1.83 ± 0.06	1.77 ± 0.12	1.93 ± 0.11	1.71 ± 0.02	1.85 ± 0.11	1.87 ± 0.13
RMSD from mean structure (bb, 2nd)	0.27 ± 0.03	0.34 ± 0.03	0.33 ± 0.05	0.20 ± 0.03	0.23 ± 0.03	0.20 ± 0.04
RMSD from mean structure (heavy, all)	1.2 ± 0.1	1.3 ± 0.3	1.3 ± 0.1	1.0 ± 0.2	1.2 ± 0.07	1.3 ± 0.1
NOE violations > 0.3 Å	1.9 ± 0.6	0 ± 0	0.7 ± 0.6	55.0 ± 3.6	6.4 ± 1.5	18.9 ± 2.6
NOE violations > 0.1 Å	17.0 ± 2.8	2.9 ± 0.6	20.1 ± 2.7	270.1 ± 7.3	68.7 ± 3.5	140.3 ± 7.7

number of NOE violations (45.1 when using error estimates derived from distances, 138.3 from volumes).

Calculations 10–12 show the most prominent differences since we are using only the spectrum with $\tau_m = 80$ ms while applying tighter error bounds. Calculation 11 shows the best values for Ramachandran plot appearance, χ_1–χ_2 rotamer normality, backbone conformation, bond lengths, bond angles, omega side chain planarity, improper dihedral distribution, and van der Waals clashes.

The packing quality decreases while all the other quality parameters get better when correcting for spin diffusion (see QUACHK results in Tables 2 and 3). However, we cannot expect to improve packing quality by means of spin diffusion correction. Application of the spin diffusion correction increases the target distances. We have often observed that slightly inconsistent (and too short) distance restraints give a better packing quality than completely consistent restraints. The decreasing packing quality is therefore an artefact of the distance bounds and the van der Waals representation used in the calculation. It would be easy to correct this by running a short trajectory of the protein in a water shell [27].

Table 3
Quality indices for calculations 7–12 (all abbreviations are explained in Section 2.3)

	7	8	9	10	11	12
WHATCHECK Z-scores						
First generation packing quality (QUACHK)	-1.3 ± 0.2	-2.2 ± 0.3	-1.8 ± 0.2	-1.0 ± 0.2	-1.6 ± 0.2	-1.5 ± 0.2
Second generation packing quality (NQACHK)	-3.0 ± 0.2	-3.3 ± 0.4	-3.0 ± 0.4	-2.9 ± 0.2	-2.9 ± 0.2	-2.9 ± 0.2
Ramachandran plot appearance (RAMCHK)	-4.8 ± 0.5	-4.6 ± 0.4	-4.2 ± 0.6	-5.5 ± 0.3	-4.4 ± 0.4	-4.9 ± 0.4
χ_1–χ_2 rotamer normality (C12CHK)	-1.3 ± 0.5	0.0 ± 0.4	-0.2 ± 0.5	-2.8 ± 0.3	-1.1 ± 0.3	-2.4 ± 0.4
Backbone conformation (BBCCHK)	-2.8 ± 0.9	-2.4 ± 0.8	-2.1 ± 0.9	-4.9 ± 0.7	-2.1 ± 0.7	-2.8 ± 0.5
WHATCHECK RMS Z-scores						
Bond lengths (BNDCHK)	0.195 ± 0.009	0.177 ± 0.004	0.184 ± 0.002	0.477 ± 0.012	0.218 ± 0.005	0.288 ± 0.007
Bond angles (ANGCHK)	0.41 ± 0.006	0.39 ± 0.003	0.397 ± 0.003	0.622 ± 0.011	0.425 ± 0.006	0.5 ± 0.01
Omega angle restraints (OMECHK)	0.11 ± 0.01	0.06 ± 0.01	0.08 ± 0.01	0.26 ± 0.02	0.13 ± 0.01	0.21 ± 0.01
Side chain planarity (PLNCHK)	0.07 ± 0.02	0.06 ± 0.01	0.06 ± 0.01	0.16 ± 0.02	0.09 ± 0.01	0.13 ± 0.01
Improper dihedral (HNDCHK)	0.18 ± 0.01	0.15 ± 0.01	0.16 ± 0.01	0.4 ± 0.01	0.21 ± 0.01	0.3 ± 0.01
Inside/outside (INOCHK)	0.97 ± 0.02	0.99 ± 0.02	1.0 ± 0.02	0.96 ± 0.01	1.0 ± 0.02	0.99 ± 0.02
Inter-atomic bumps (BMPCHK)	27.9 ± 4.8	17.3 ± 4.6	19.0 ± 5.6	63.9 ± 4.3	27.8 ± 5.1	44.9 ± 4.0
Unsatisfied hydrogen donors (BH2CHK)	15.8 ± 2.7	17.4 ± 3.0	17.2 ± 2.6	20.6 ± 2.8	17.4 ± 3.4	15.8 ± 3.3
Unsatisfied hydrogen acceptors (BA2CHK)	0.4 ± 0.6	0.5 ± 0.7	0.5 ± 0.5	1.0 ± 0.8	0.5 ± 0.8	1.0 ± 0.8
PROCHECK results						
Most favoured regions	67.6 ± 3.6	67.2 ± 2.8	70.0 ± 3.5	63.3 ± 3.0	68.4 ± 3.8	70.6 ± 2.5
Allowed regions	27.1 ± 3.2	27.0 ± 2.7	25.5 ± 3.1	29.7 ± 2.8	27.5 ± 3.8	24.2 ± 2.8
Generously allowed regions	3.6 ± 1.9	3.9 ± 1.7	3.1 ± 1.8	5.6 ± 1.5	3.2 ± 1.5	3.6 ± 1.9
Disallowed regions	1.7 ± 0.9	1.9 ± 0.8	1.4 ± 1.3	1.3 ± 0.7	1.0 ± 0.7	1.6 ± 1.2
PROSA-II mean force energy	-1.09 ± 0.1	-1.03 ± 0.08	-0.96 ± 0.09	-1.16 ± 0.06	-1.07 ± 0.08	-1.06 ± 0.07
RMSD to 1BTN (bb)	0.85 ± 0.04	0.90 ± 0.04	0.77 ± 0.03	0.80 ± 0.06	0.84 ± 0.10	0.86 ± 0.08
RMSD to 1BTN (all)	1.79 ± 0.04	2.01 ± 0.09	1.87 ± 0.07	1.81 ± 0.03	1.84 ± 0.12	1.81 ± 0.10
RMSD from mean structure (bb, 2nd)	0.32 ± 0.03	0.41 ± 0.04	0.33 ± 0.04	0.21 ± 0.02	0.21 ± 0.03	0.23 ± 0.03
RMSD from mean structure (heavy, all)	1.1 ± 0.1	1.5 ± 0.1	1.3 ± 0.2	1.1 ± 0.2	1.1 ± 0.1	1.3 ± 0.2
NOE violations > 0.3 Å	1.6 ± 0.5	0 ± 0	0.3 ± 0.4	58.0 ± 1.5	3.3 ± 1.2	20.9 ± 1.6
NOE violations > 0.1 Å	12.0 ± 1.9	1.4 ± 0.5	7.7 ± 1.4	284.1 ± 12.1	45.1 ± 1.2	138.3 ± 13.1

Precision (RMSD to the average structure) and accuracy (RMSD to the X-ray structure 1BTN) do not differ greatly between the calculations. Their values are all within a standard deviation. The inside/outside distribution (INOCHK) and the PROSA-II mean energy force cannot distinguish between small conformational differences.

4. Conclusions

The new algorithm for the calculation of the NOE matrix has proven to be numerically stable and CPU-efficient. For the PH domain, we can simulate an NOE spectrum in 12 s on a single Athlon 1.5 GHz processor. We use the calculated NOE intensities in an automated NOE assignment scheme in a straightforward manner. The spin diffusion correction also works for ambiguous NOEs. In ARIA, the size of the contribution of each assignment possibility is proportional to the intensity of the calculated NOE. Thus, one does not have to merge experimental and calculated NOE matrices (as in the methods IRMA [11] or MARDIGRAS [12]) in order to make use of the calculated NOE intensities.

The described spin diffusion correction is suitable for iterative automated NOE assignment. In each round of

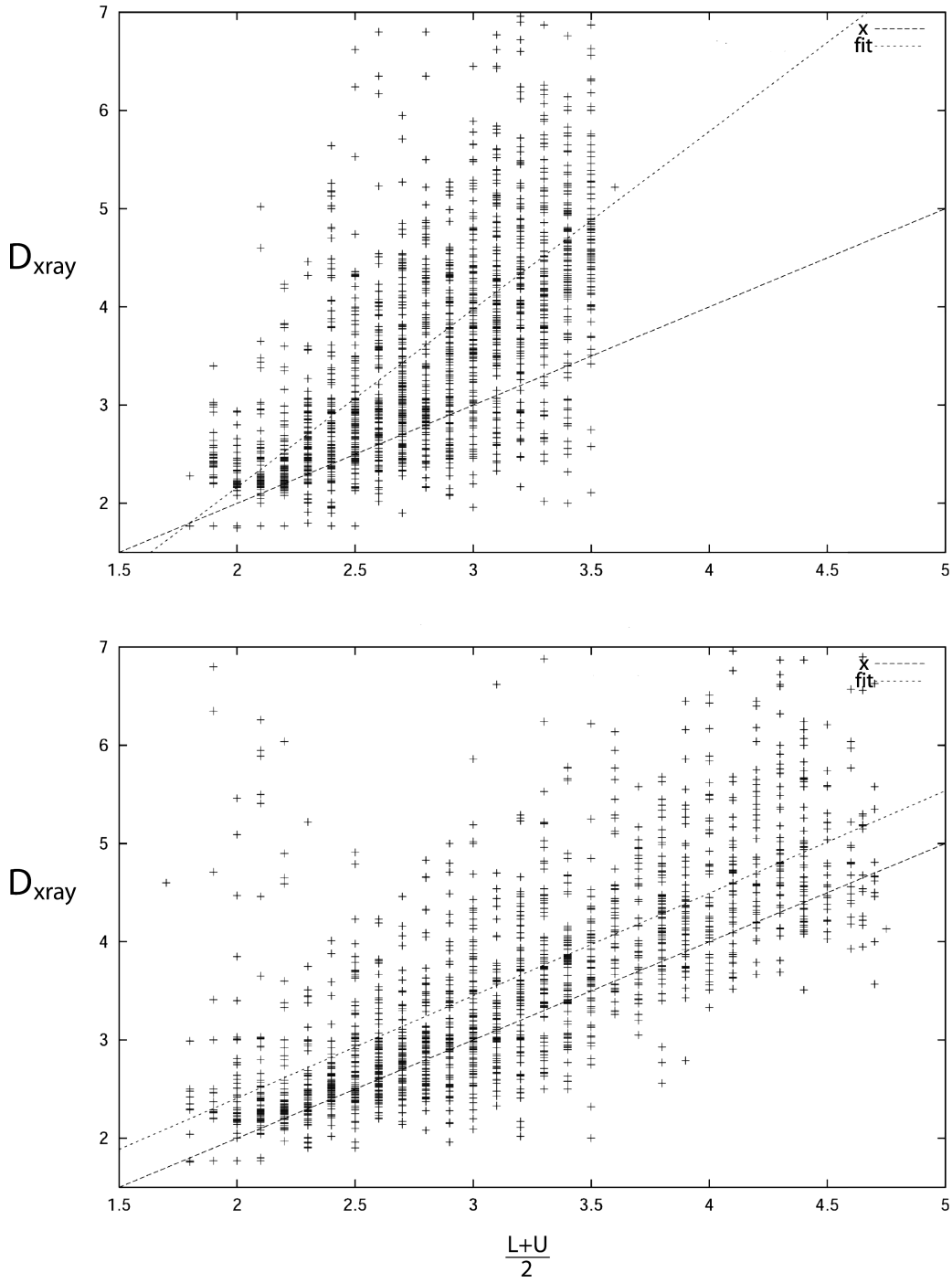

Fig. 3. Distances D in the model structures vs calibrated distances (the sum of lower bound L and upper bound U divided by two). The distances of the merged peak list of calculation 10 (without spin diffusion correction) are shown above, the merged peak list of calculation 11 (with spin diffusion correction) below. The dashed lines are a least-squares fit to all data points and the diagonal.

assignment, calibration, and structure calculation, ARIA applies the method to derive spin diffusion corrected target distances automatically. Test calculations with the PH domain data showed strong improvements such as less NOE violations and a more realistic relationship between the calibrated distances and the distances in the X-ray structure through the use of the spin diffusion correction. Common quality indices such as the Ramachandran plot show clear improvements when applying the spin diffusion correction, except for the packing quality. Calibration with the ISPA underestimates the target distances and thus leads to a tighter packing of the structures. We made the spin diffusion correction available in ARIA 1.2 [18]. The method is

thus easily accessible even for spectroscopists unexperienced in structure calculation. The user only has to provide the correlation time τ_c, the mixing time τ_m, and the spectrometer frequency.

ARIA 1.2 comprising all the protocols used in this study is available from our web server: www.pasteur.fr/recherche/unites/Binfs.

Acknowledgments

J.P.L. thanks the Pasteur Institute and the SPINE network (EU Fifth Framework Program, Contract No. QLG2-CT-2002-00988) for financial support. M.H. and W.R. received funding from the NMRQUAL project (EU Fifth Framework Program, Contract No. QLG2-CT-2000-01313).

References

[1] I. Solomon, Relaxation processes in a system of two spins, Phys. Rev. 99 (1955) 559–565.
[2] D. Neuhaus, M.P. Williamson, The Nuclear Overhauser Effect in Structural and Conformational Analysis, VCH Publishers, New York, 1989.
[3] S. Macura, R.R. Ernst, Elucidation of cross-relaxation in liquids by two-dimensional NMR spectroscopy, Mol. Phys. 41 (1980) 95–117.
[4] L. Zhu, H.J. Dyson, P.E. Wright, A NOESY-HSQC simulation program, SPIRIT, J. Biomol. NMR 11 (1998) 17–29.
[5] R.H.A. Folmer, M. Nilges, R.N.H. Konings, C.W. Hilbers, Solution structure of the single-stranded DNA binding protein of bacteriophage Pf3, EMBO J. 14 (1995) 4132–4142.
[6] P.F. Yip, D.A. Case, A new method for refinement of macromolecular structures based on nuclear Overhauser effect spectra, J. Magn. Reson. 83 (1989) 643–648.
[7] A.M.J.J. Bonvin, R. Boelens, R. Kaptein, Determination of biomolecular structures by NMR: use of relaxation matrix calculations, in: W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinson (Eds.), Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, vol. 2, Escom, Leiden, 1993, pp. 407–440 (Chapter 17).
[8] D.A. Case, New directions in NMR spectral simulation and structure refinement, in: W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinsonm (Eds.), Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, vol. 2, Escom, Leiden, 1993, pp. 382–406 (Chapter 6).
[9] M.J. Dellwo, J. Wand, Computationally efficient gradients for relaxation matrix-based structure refinement including the accommodation of internal motions, J. Biomol. NMR 3 (1993) 205–214.
[10] P.F. Yip, A computationally efficient method for evaluating the gradient of 2D NOESY intensities, J. Biomol. NMR 3 (1993) 361–365.
[11] R. Boelens, T.M.G. Koning, R. Kaptein, Determination of biomolecular structures from proton–proton NOEs using a relaxation matrix approach, J. Mol. Struct. 173 (1989) 299–311.
[12] B.A. Borgias, T.L. James, MARDIGRAS: a procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure, J. Magn. Reson. 87 (1990) 475–487.
[13] M.J. Forster, Comparison of computational methods for simulating nuclear Overhauser effects in NMR spectroscopy, J. Comput. Chem. 12 (1990) 292–300.
[14] J.P. Linge, S.I. O'Donoghue, M. Nilges, Automated assignment of ambiguous nuclear Overhauser effects with ARIA, Methods Enzymol. 339 (2001) 71–90.
[15] G. Lancelot, J.-L. Guesnet, F. Vovelle, Solution structure of the parallel-stranded duplex oligonucleotide α-d(TCTAAAC)-β-d(AGATTTG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations, Biochemistry 28 (1989) 7871–7878.
[16] J.-L. Guesnet, F. Vovelle, N.T. Thuong, G. Lancelot, 2D NMR studies and 3D structure of the parallel-stranded duplex oligonucleotide Acrm5-α-d(TCTAAACTC)-β-d(AGATTTGAG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations, Biochemistry 29 (1990) 4982–4991.
[17] R.C. Hoffman, R.X. Xu, R.E. Klevit, J.R. Herriott, A simple method for the refinement of models derived from NMR data demonstrated on a zinc-finger domain from yeast ADR1, J. Magn. Reson. B 102 (1993) 61–72.
[18] J.P. Linge, M. Habeck, W. Rieping, M. Nilges, ARIA: automated NOE assignment and NMR structure calculation, Bioinformatics 19 (2003) 315–316.
[19] M.J. Macias, A. Musacchio, H. Ponstingl, M. Nilges, M. Saraste, H. Oschkinat, Structure of the pleckstrin homology domain from β-spectrin, Nature 369 (1994) 675–677.
[20] M. Nilges, M.J. Macias, S.I. O'Donoghue, H. Oschkinat, Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin, J. Mol. Biol. 269 (1997) 408–422.
[21] R.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature 381 (1996) 272.
[22] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, J. Mol. Biol. 7 (1963) 95–99.
[23] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[24] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins Struct. Funct. Genet. 17 (1993) 355–362.
[25] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination, Acta Crystallogr. Sect. D 54 (1998) 905–921.
[26] M. Hyvönen, M.J. Macias, M. Nilges, H. Oschkinat, M. Saraste, M. Wilmanns, Structure of the binding site for inositol phosphates in a PH domain, EMBO J. 14 (1995) 4676–4685.
[27] J.P. Linge, M.A. Williams, C.A. Spronk, A.M. Bonvin, M. Nilges, Refinement of protein structures in explicit solvent, Proteins Struct. Funct. Genet. 20 (2003) 496–506.