Journal of Biomolecular NMR, 4 (1994) 143–149
ESCOM

J-Bio NMR 168

Time- and ensemble-averaged direct NOE restraints

Alexandre M.J.J. Bonvin, Rolf Boelens and Robert Kaptein*

Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received 18 August 1993
Accepted 22 October 1993

Keywords: Direct NOE refinement; Time averaging; Ensemble averaging; DINOSAUR

SUMMARY

NMR data are collected as time- and ensemble-averaged quantities. Yet, in commonly used methods for structure determination of biomolecules, structures are required to satisfy simultaneously a large number of constraints. Recently, however, methods have been developed that allow a better fit of the experimental data by the use of time- or ensemble-averaged restraints. Thus far, these methods have been applied to structure refinement using distance and J-coupling restraints. In this paper, time and ensemble averaging is extended to the direct refinement with experimental NOE data. The implementation of time- and ensemble-averaged NOE restraints in DINOSAUR is described and illustrated with experimental NMR data for crambin, a 46-residue protein. Structure refinement with both time- and ensemble-averaged NOE restraints results in lower R-factors, indicating a better fit of the experimental NOE data.

Structure determination by NMR is based on the collection of a large number of constraints, typically obtained from NOE and J-coupling data. These constraints are used to generate and refine a set of structures. Although NMR experimental data are collected as time- and ensemble-averaged quantities, each structure is required to satisfy simultaneously all the constraints. However, generally structures fail to satisfy all the NMR constraints at the same time and may be forced into unrealistic conformations. Solutions to this problem have been proposed by the introduction of time- (Torda et al., 1989,1990) and ensemble-averaged NMR constraints (Scheek et al., 1991; Kemmink et al., 1993). Time averaging and its implementation in restrained molecular dynamics (MD) simulations have been described and applied first for NOE-derived distances (Torda et al., 1989,1990; Pearlman and Kollman, 1991; Schmitz et al., 1992) and more recently for J-coupling data (Torda et al., 1993). Here we extend the application of time- and ensemble-averaged restraints to the direct refinement against experimental NOE data using relaxation matrix methods.

The experimental NOE intensities can be introduced directly as constraints in a refinement procedure (Yip and Case, 1989), using a penalty function of the form:

*To whom correspondence should be addressed.

$$V_{NOE} = \sum w_{ij} \left[(f\, A_{ij}^{theo})^x - (A_{ij}^{exp})^x \right]^y \quad (1)$$

where A_{ij}^{exp} and A_{ij}^{theo} represent the experimental and theoretical NOE intensities between protons i and j, respectively, f is a scaling factor and w_{ij} a weighting function. Often a quadratic potential is used (y = 2), directly with the NOE intensities (x = 1) (Yip and Case, 1989; Baleja et al., 1990; Bonvin et al., 1991; Mertz et al., 1991) or with their sixth-root (x = 1/6) (Nilges et al., 1991) or inverse sixth-root (x = −1/6) (Stawarz et al., 1992), the last two definitions being closer to a standard distance-based potential. When averaging is introduced, the theoretical intensities in Eq. 1 should be replaced by their time or ensemble averages, denoted as $\overline{A_{ij}^{theo}}$. Instead of intensities calculated from a single static structure, now the averaged intensities are required to satisfy the experimental constraints.

In the case of time averaging, for which the time course of an MD simulation can be used, the averaged NOEs, $\overline{A_{ij}^{theo}}(t)$ are given by:

$$\overline{A_{ij}^{theo}}(t) = \frac{1}{t} \int_0^t A_{ij}^{theo}(t')\, dt' \quad (2)$$

Equation 2 corresponds to the true average, which, as the time increases, might become less sensitive to instantaneous fluctuations in the MD simulation (Torda et al., 1989). To avoid this problem, Torda et al. (1989) introduced an exponentially decaying memory function with time constant τ in the summation over the time in Eq. 2, which becomes:

$$\overline{A_{ij}^{theo}}(t) = [\tau(1 - \exp(-t/\tau))]^{-1} \int_0^t \exp(-t'/\tau))\, A_{ij}^{theo}(t - t')\, dt' \quad (3)$$

If the simulation time t is much longer than the time constant of the exponential τ, a practical way to calculate the average of Eq. 3, suitable for implementation in MD algorithms, is given by (Torda et al., 1989):

$$\overline{A_{ij}^{theo}}(t) = [1 - \exp(-\Delta t/\tau)]\, A_{ij}^{theo}(t) + \exp(-\Delta t/\tau)\, A_{ij}^{theo}(t - \Delta t) \quad (4)$$

where Δt is the time step of the integrator in the MD simulation. The true average defined in Eq. 2 can, however, be used for the analysis of the MD trajectories.

For ensemble-averaged NOE restraints, the theoretical intensities are given by

$$\overline{A_{ij}^{theo}} = \sum_{k=1}^{nconf} p_k\, A_{ij}^{theo}(k) \qquad \text{with} \sum_{k=1}^{nconf} p_k = 1 \quad (5)$$

where p_k gives the probability of the conformer k. Ideally, a Boltzmann weighting should be chosen for the probabilities. It is, however, not possible to obtain the free energy to calculate the Boltzmann factor in the course of a simulation; thus, some other weighting function has to be used. Equation 5 requires the computation of the theoretical NOE intensities for all conformers in the ensemble, which is one of the time-consuming steps of direct NOE refinement. However, if

we assume that the various conformers are in slow exchange, the formalism of Landy and Rao (1989) can be applied and the averaged NOEs are then given by

$$\overline{A_{ij}^{theo}} = \sum_{k=1}^{nconf} p_k A_{ij}^{theo}(k) = \left[\exp\left(-\tau_m \sum_{k=1}^{nconf} p_k \mathbf{R}(k) \right) \right]_{ij} \quad (6)$$

where $\mathbf{R}(k)$ represents the relaxation matrix for the *k*th conformer and τ_m is the mixing period. With this formalism, the computation of the theoretical NOE intensities is performed only once for the ensemble, allowing a reduction in computational time.

Both time- and ensemble-averaging options have been implemented in the DINOSAUR routines for direct NOE refinement (Bonvin et al., 1991,1993a) and can be used in conjunction with the GROMOS programs for energy minimisation and molecular dynamics (Van Gunsteren and Berendsen, 1987). The routines, written in FORTRAN 77, are available from the authors.

To demonstrate the feasibility of time and ensemble averaging for direct NOE refinement we used experimental NOE data for crambin, a small protein of 46 amino acid residues. The solution structure of crambin has been determined using relaxation matrix calculations, combining the Iterative Relaxation Matrix Approach (IRMA) in the first stage to generate accurate distance constraints (Bonvin et al., 1993b) and direct refinement against experimental NOE intensities with DINOSAUR in the final stage (Bonvin et al., 1993c). The set of constraints is the same as used previously (for a complete description see Bonvin et al., 1993b). 380 NOE peaks at four mixing times (40, 80, 160 and 250 ms) have been introduced as time- or ensemble-averaged restraints in MD simulations. In addition, 159 qualitative distance restraints and 13 χ^1 dihedral angle restraints were used in a standard way in the calculations. The theoretical NOEs were calculated using a spherical cutoff of 4.5 Å around each proton pair defining a peak. A simple quadratic restraining potential was used (x = 1, y = 2 in Eq. 1) with as weighting factor an experimental error defined as $w_{ij} = (N + \varepsilon A_{ij}^{exp})^{-2}$ (Bonvin et al., 1991). N corresponds to a noise level and ε to a relative error in the experimental intensities. The force constant for NOE restraining was set to 400 kJ mol^{-1}. The scaling factor f in Eq. 1 between theoretical and experimental NOE intensities was calculated at every step during the refinement from three alanine H^{α}-methyl build-up curves. One structure obtained after DINOSAUR refinement with standard NOE restraints (Bonvin et al., 1993c) was chosen as starting conformation.

Time-averaged restraints require simulation times that are approximately one order of magnitude longer than the decay constant of the memory function, in order to allow the system to converge. Since direct NOE refinement is a computationally expensive procedure, we chose a rather short decay constant τ of 5 ps and generated a 50-ps MD trajectory. A time step of 0.001 ps was used in the integrator and the system was weakly coupled (0.01 ps) to a heat bath at 300 K (Berendsen et al., 1984). The initial averaged NOE intensities were set equal to the experimental values. This implies that no NOE forces are acting on the structure at the beginning of the simulation. As a comparison, a reference MD trajectory was generated with conventional instantaneous NOE restraints. The computation of a 50-ps MD trajectory with direct NOE constraints required 135 h of CPU time on a Convex C220. Both trajectories were analysed in terms of R-factors to monitor the quality of the fit between experimental and theoretical NOE intensities. For this purpose, the $Q^{1/6}$ definition was used, i.e. a sixth-root residual, normalised by the sum of experimental and theoretical NOEs (Bonvin et al., 1993d).

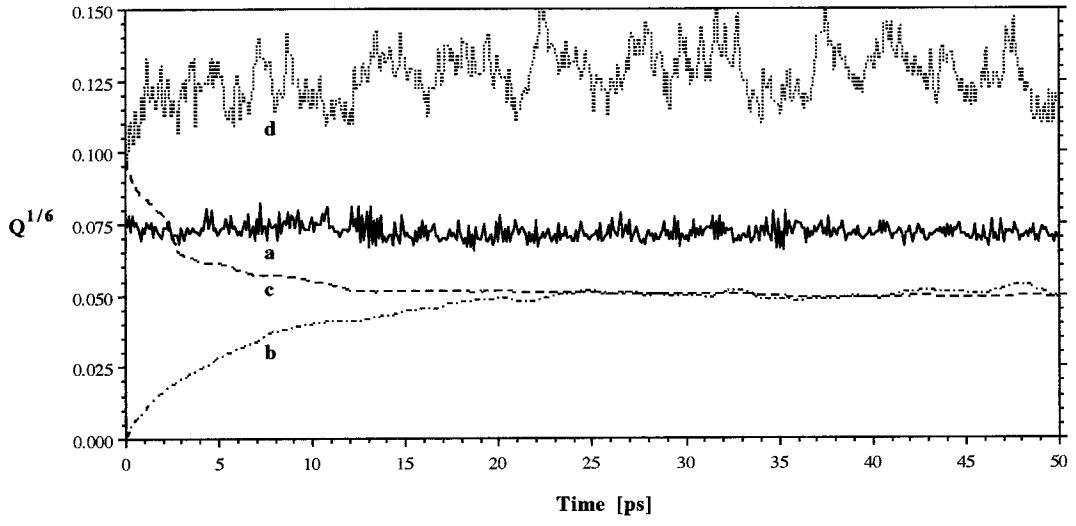

Fig. 1. R-factors ($Q^{1/6}$ definition (Eq. 7)) of crambin as a function of time during the restrained MD refinement with direct NOE constraints. The R-factors are plotted for the MD simulations with instantaneous NOE restraints (a) and with time-averaged NOE restraints ((b) time-averaged NOEs with exponential memory function (τ = 5 ps) according to Eq. 3 (the initial NOEs were set equal to the experimental intensities); (c) time-averaged NOEs (true average) according to Eq. 2; and (d) instantaneous NOEs).

$$Q^{1/6} = \frac{\sum_{ij} \tau_m \mid (A_{ij}^{theo})^{1/6} - (A_{ij}^{exp})^{1/6} \mid}{\sum_{ij} 0.5\tau_m \mid (A_{ij}^{theo})^{1/6} + (A_{ij}^{exp})^{1/6} \mid} \tag{7}$$

The evolution of the R-factors as a function of time is presented in Fig. 1. For the trajectory with time-averaged NOE restraints, the R-factors were calculated from the averaged NOEs with memory function according to Eq. 3, from the true average using Eq. 2 and from the instantaneous values. Note that, due to the choice of setting the initial NOE intensities equal to the experimental values in Eq. 4, the R-factors calculated from the averaged NOEs with exponential memory function (curve b) start from zero. It is clear from Fig. 1 that the R-factors calculated from the exponential memory and true averages converge and reach a plateau (0.049) after ~ 25 ps, which corresponds to the lowest $Q^{1/6}$ factor value for crambin. The use of time-averaged direct NOE restraints results in a better fit of the experimental data, the R-factors for the reference run having higher values, fluctuating around 0.074. We also notice from the R-factors, calculated from the instantaneous NOEs in the time-averaged trajectory, that time-averaged restraints result in larger instantaneous deviations from the experimental values and increased variability in the structures. This can also be seen from a plot of the rmsd on C^{α} atoms, calculated from the last 25 ps of the MD trajectory in Fig. 2. The rmsd for the run with instantaneous restraints are very low (0.1 Å on average), because the structures are required to satisfy the experimental constraint at any time during the simulation. For time-averaged NOE restraints, however, much higher rmsd (1.0 Å on average) are obtained, with maxima in the N- and C-terminal parts and in the loop regions around Gly^{20}, Gly^{37} and Gly^{42}. Similar effects were already noted for time-averaged distance and J-coupling restraints (Torda et al., 1990,1993; Pearlman and Kollman, 1991).

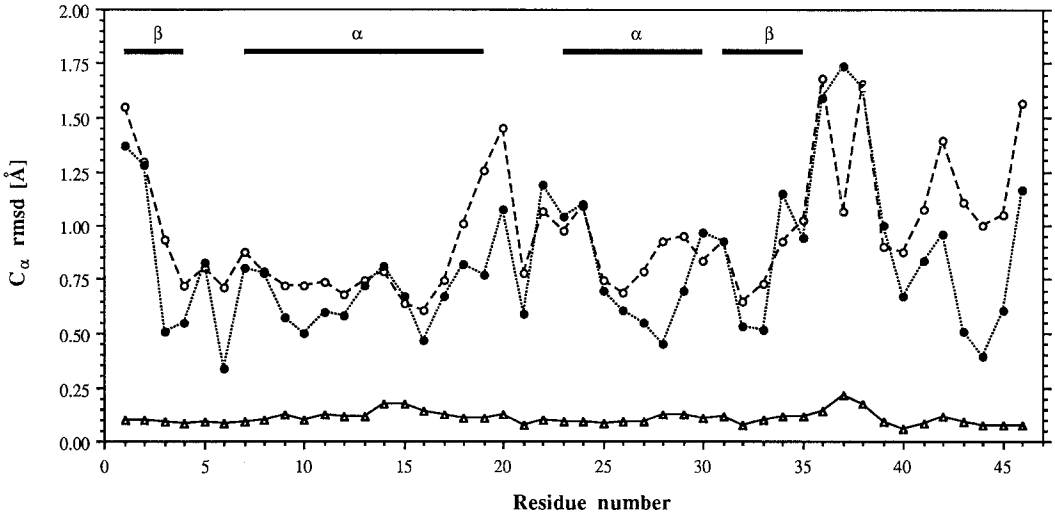

Fig. 2. Rmsd of C^α atoms as a function of residue sequence for crambin, calculated from: (△) the last 25 ps of the MD trajectory with instantaneous NOE restraints; (○) the last 25 ps of the MD trajectory with time-averaged NOE restraints; and (●) the ensemble of eight structures, obtained after the slow-cooling simulated annealing with ensemble-averaged NOE restraints.

For the refinement with ensemble-averaged NOE restraints, eight copies of the starting structure were placed on the edges of a cubic box of 100 Å length. The dimensions of the box ensure that the structures do not interact with each other during the simulation. The same weight was attributed to all structures and the theoretical NOE intensities were computed from the averaged relaxation matrix according to Eq. 6. A slow-cooling simulated annealing protocol (Brünger and Krukowski, 1990) was followed, as previously applied for the DINOSAUR refinement of crambin (Bonvin et al., 1993c). This protocol consisted of 5 ps MD from 1000 to 1 K, with a time step of 0.001 ps and a cooling rate of 10 K/0.05 ps. The system was weakly coupled (0.01 ps) to a heating bath, the temperature of which was progressively decreased during the run. Initial random velocities were taken from a Maxwellian distribution at 1000 K. As a comparison, a single-structure reference run was performed. The 5 ps slow-cooling annealing required 13.5 and 29 h of CPU time on a CONVEX C220 for the single-structure and ensemble runs, respectively. Figure 3 shows a plot of the $Q^{1/6}$ factors as a function of time during the slow-cooling annealing for the single-structure and the ensemble-averaged runs. For both, a slow decrease in R-factors is observed during the simulation. The run with ensemble-averaged NOE restraints results, however, in a lower R-factor (0.052 against 0.066 for the single-structure run), close to the one obtained with time-averaged NOE restraints (0.049). No improvement in R-factor is found for the single-structure run, which is not surprising, since the starting structure was already refined directly against experimental NOE data. In Fig. 3, note that at the end of the simulation the temperature is 1 K and the R-factors do not decrease further. Taken individually, the various members of the ensemble have higher R-factor values (0.106 ± 0.007) than the ensemble as a whole, similar to what happens with time-averaged restraints. The ensemble of structures has rmsd of 0.9 Å on backbone atoms and 1.2 Å on all heavy atoms. The rmsd on C^α atoms per

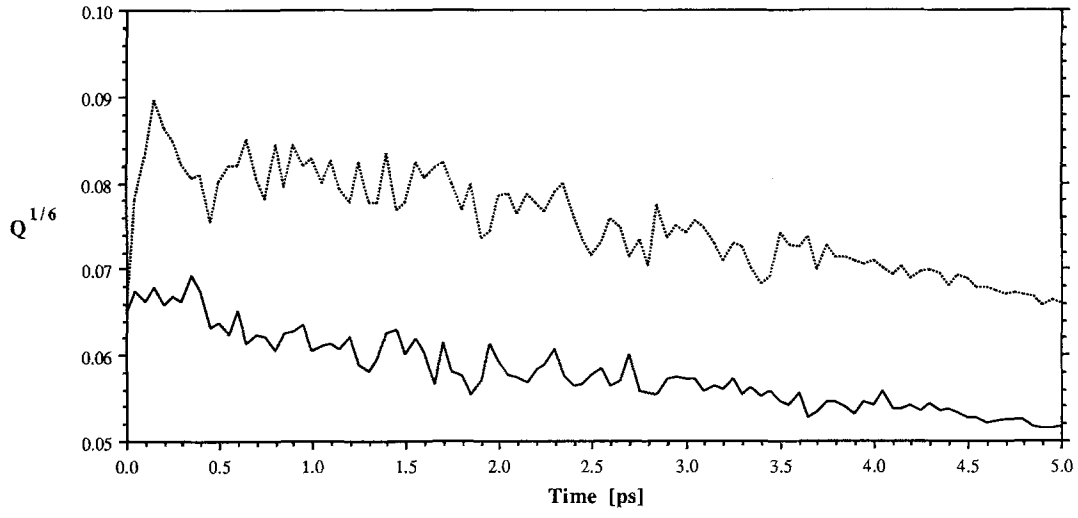

Fig. 3. R-factors ($Q^{1/6}$ definition (Eq. 7)) of crambin as a function of time during the slow-cooling simulated annealing (from 1000 to 1 K in 5 ps with a time step of 0.001 ps, cooling rate 10 K/0.05 ps). (·····) single-structure refinement; (—) ensemble-averaged refinement with averaged NOEs calculated from the ensemble of eight structures according to Eq. 6.

residue are shown in Fig. 2. Interestingly, these correspond reasonably well with the rmsd obtained from the restrained MD simulation with time-averaged NOE constraints.

With the results for crambin, we have demonstrated the feasibility of time- and ensemble-averaged NOE restraints for structure refinement. Both approaches have similar characteristics: by the modelling of the NOE intensities as time or ensemble averages a better fit is obtained between theoretical and experimental NMR data, together with an increased variability in the structure. The use of time- and/or ensemble-averaged NOE restraints should result in a more realistic picture of a biomolecule in solution with a Boltzmann distribution of structures. Presently, direct NOE refinement still is, however, a computationally expensive procedure, especially in the case of time averaging which requires very long MD simulations.

ACKNOWLEDGEMENTS

This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO). A.B. acknowledges NWO for financial support.

REFERENCES

Baleja, J.D., Moult, J. and Sykes, B.D. (1990) *J. Magn. Reson.*, **87**, 375–384.

Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) *J. Chem. Phys.*, **81**, 3684–3690.

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1991) *J. Biomol. NMR*, **1**, 305–309.

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1993a) In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Vol. 2 (Eds, Van Gunsteren, W.F., Weiner, P.K. and Wilkinson, A.J.) ESCOM, Leiden, pp. 407–440.

Bonvin, A.M.J.J., Rullmann, J.A.C., Lamerichs, R.M.J.N., Boelens, R. and Kaptein, R. (1993b) *Biopolymers*, **15**, 385–400.

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1993c) *Biopolymers*, in press.

Bonvin, A.M.J.J., Vis, H., Breg, J.N., Burgering, M.J.M., Boelens, R. and Kaptein, R. (1993d) *J. Mol. Biol.*, in press.

Brünger, A.T. and Krukowski, A. (1990) *Acta Crystallogr.*, **A46**, 585–593.

Kemmink, J., Van Mierlo, C.P.M., Scheek, R.M. and Creighton, T.E. (1993) *J. Mol. Biol.*, **230**, 312–322.

Landy, S.B. and Rao, B.D.N. (1989) *J. Magn. Reson.*, **81**, 371–377.

Mertz, J.E., Güntert, P., Wüthrich, K. and Braun, W. (1991) *J. Biomol. NMR*, **1**, 257–269.

Nilges, M., Habazettl, J., Brünger, A.T. and Holak, T.A. (1991) *J. Mol. Biol.*, **219**, 499–510.

Pearlman, D.A. and Kollman, P.A. (1991) *J. Mol. Biol.*, **220**, 457–479.

Scheek, R.M., Torda, A.E., Kemmink, J. and Van Gunsteren, W.F. (1991) In *Computational Aspects of the Study of Biological Macromolecules by NMR* (Eds, Hoch, J.C., Poulsen, F.M. and Redfield, C.) Plenum Press, New York, NY, pp. 209–217.

Schmitz, U., Kumar, A. and James, T.L. (1992) *J. Am. Chem. Soc.*, **114**, 10654–10656.

Stawarz, B., Genest, M. and Genest, D. (1992) *Biopolymers*, **32**, 633–642.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1989) *Chem. Phys. Lett.*, **157**, 289–294.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Torda, A.E., Brunne, R.M., Huber, T., Kessler, H. and Van Gunsteren, W.F. (1993) *J. Biomol. NMR*, **3**, 55–66.

Van Gunsteren, W.F. and Berendsen, H.J.C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, University of Groningen, Groningen.

Yip, P. and Case, D.A. (1989) *J. Magn. Reson.*, **83**, 643–648.