

Direct NOE refinement of biomolecular structures using 2D NMR data

A.M.J.J. Bonvin, R. Boelens and R. Kaptein*

Bijvoet Center, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received 25 July 1991

Accepted 31 July 1991

Keywords: Direct NOE refinement; Biomolecular structure; Molecular dynamics; 2D NMR

SUMMARY

A set of computer programs called DINOSAUR has been developed, which allows the refinement of biomolecular structures directly from 2D NOE intensities. The NOE restraining potential implemented emphasises the weak intensities corresponding to larger distances which are more likely to determine the three-dimensional structure. An approximation based on a two-spin approximation is proposed for the gradient of the NOE intensities instead of the exact solution which is extremely time-consuming. The DINOSAUR routines have been implemented in various refinement programs (Distance bound Driven Dynamics, Molecular Dynamics and Energy Minimisation) and tested on an eight-residue model peptide.

NMR spectroscopy has proven a valuable method for the determination of the solution structure of proteins and nucleic acids. 2D NOE spectra whose intensities are correlated with proton-proton distances are a widely used source of information to achieve this goal. Various approaches have been proposed to obtain distance constraints from NOE intensities: qualitative distance ranges (Wüthrich, 1986), initial-rates analysis (Kumar et al., 1981), relaxation matrix approaches (Olejniczak et al., 1986; Boelens et al., 1988). Recently, new methods were proposed where NOE intensities calculated from a model structure were compared to experimental ones and directly used in a refinement procedure (Yip and Case, 1989; Baleja et al., 1990). We have developed a set of routines called DINOSAUR, *Direct NOe Simulation Approach for strUcture Refinement*, where, instead of the exact derivative of the NOE intensity as proposed by Yip and Case (1989), an approximation is used together with the exact calculation of the theoretical NOEs via diagonalisation of the relaxation matrix. This approach allows for considerable reduction of the computational time. The DINOSAUR routines have been implemented in three different programs:

* To whom correspondence should be addressed.

distance bound driven dynamics (DDD) where only the covalent information is used in addition to the experimental constraints (Kaptein et al., 1988) and in the GROMOS programs for energy minimisation and molecular dynamics which make use of a complete force field (van Gunsteren and Berendsen, 1987). The DINOSAUR routines, written in FORTRAN 77, can be easily adapted to other programs and are available upon request.

The normalised intensities in a 2D NOE spectrum recorded with mixing time τ_m are given by the matrix equation (Macura and Ernst, 1980):

$$\mathbf{A} = \exp(-\tau_m \mathbf{R}) \quad (1)$$

where \mathbf{R} represents the relaxation matrix. Given a model structure, NOE intensities can be calculated exactly via diagonalisation of \mathbf{R} . Internal motions like aromatic ring flip, methyl group rotation and fast local motion can be easily incorporated in the calculations (Koning et al., 1990). These theoretical intensities can then be compared to the experimental ones and the difference used to determine forces in a refinement procedure (Yip and Case, 1989). Exact calculation of the NOE intensities at every step of a refinement procedure can become very cumbersome especially in a molecular dynamics simulation. A suitable approximation consists in calculating the NOE intensities with a two-spin approximation and correcting them for spin diffusion.

$$\begin{aligned} \mathbf{A}^{\text{approx.}} &= \mathbf{A}^{2\text{spin}} + \Delta\mathbf{A} \quad (\text{each step}) \\ \Delta\mathbf{A} &= \mathbf{A}^{\text{exact}} - \mathbf{A}^{2\text{spin}} \quad (\text{each } n^{\text{th}} \text{ step}) \end{aligned} \quad (2)$$

The spin diffusion correction is given by the difference between the exact NOE intensities calculated according to Eq. 1 and those obtained with a two-spin approximation. This contribution is held constant during a predetermined number of refinement steps. Thus, the diagonalisation of \mathbf{R} , which is the time-consuming procedure, need not be carried out at every refinement step.

The theoretical NOE intensities are then compared with the experimental data using a NOE restraining potential function V_{NOE} defined as:

$$\begin{aligned} V_{NOE} &= \left[\sum_{\tau_m} \sum_i W_{\tau_m,i} \right]^{-1} \sum_{\tau_m} \sum_i W_{\tau_m,i} (A_{\tau_m,i}^{\text{theo}} - A_{\tau_m,i}^{\text{exp}})^2 \\ \text{with } W_{\tau_m,i} &= \frac{1}{(N + \epsilon A_{\tau_m,i}^{\text{exp}})^2} \end{aligned} \quad (3)$$

The sum runs over the NOE peaks i for the different mixing times τ_m ; N represents the experimental noise level and ϵ a relative error on the experimental NOE intensities. The relative term should account for integration errors. The restraining function of Eq. 3, depending on the choice of the noise level and the relative error, covers the range between an absolute ($\epsilon = 0$) and a relative error function ($\epsilon = 1, N = 0$). This gives more emphasis to the weaker NOE intensities (corresponding to longer distances), which are more likely to determine the tertiary structure.

Obtaining the forces requires the calculation of the gradient of V_{NOE} and therefore of the derivative of the NOE intensities. The exact solution proposed by Yip and Case (1989) contains terms involving many protons in the molecule and consists of a product of five matrices and therefore

becomes extremely time-consuming when the dimension of the problem increases. To avoid this problem we use an approximation for the derivative based on the first two terms of the expansion of the exponential of Eq. 1:

$$\nabla \mathbf{A} = \nabla(1 - \tau_m \mathbf{R}) = -\tau_m \nabla \mathbf{R} \quad (4)$$

The gradient of V_{NOE} then becomes:

$$\nabla V_{NOE} = \left[\sum_{\tau_m} \sum_i W_{\tau_m,i} \right]^{-1} \sum_{\tau_m} \sum_i 2W_{\tau_m,i} (A_{\tau_m,i}^{theo} - A_{\tau_m,i}^{exp}) (-\tau_m) \nabla \mathbf{R}_i \quad (5)$$

Due to the nature of the NOE intensities (approximately inverse sixth power dependence on the distances) V_{NOE} is an asymmetric function which generates very large forces if the distances in the model structure are underestimated. It appears therefore necessary to limit the NOE forces in the simulations. This corresponds to a linearisation of the NOE potential above a defined value. In order to remain independent of the simulation parameters (N , ϵ), we describe the quality of the NMR structures in terms of an NMR R-factor (Gonzales et al., 1991).

$$R = \frac{\sum_{\tau_m} \sum_i \tau_m |A_{\tau_m,i}^{theo} - A_{\tau_m,i}^{exp}|}{\sum_{\tau_m} \sum_i \tau_m A_{\tau_m,i}^{exp}} \quad (6)$$

To test our programs, we chose as a model structure an eight-residue α -helix (56 protons) corresponding to the second helix of crambin. A synthetic data set consisting of 238 NOE peaks above a minimum intensity corresponding to a distance of 5.0 Å was generated at 500 MHz for six mixing times, using a correlation time of 3 ns. A starting structure was generated with Distance Geometry followed by 300 steps Distance bound Driven Dynamics using a set of 50 qualitative

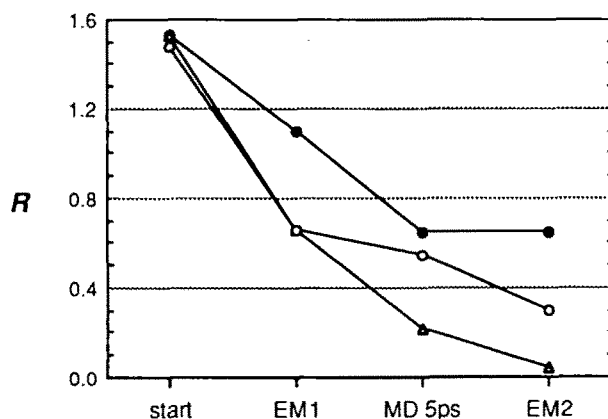


Fig. 1. R-factors following Eq. 6 calculated on interresidue peaks at different stages during the refinement. ●: standard distance restraining, ▲: DINOSAUR refinement with exact NOE calculation, ○: DINOSAUR refinement with NOE calculation using the approximation of Eq. 2 (spin diffusion correction updated every 20 steps).

constraints. This structure, which has an RMSD of 2.5 Å for all atoms (1.4 Å for the backbone atoms), was used in two parallel procedures:

- (i) a direct NOE refinement including 300 steps of energy minimisation, 5 ps molecular dynamics with an integration step of 0.001 ps followed by 300 steps energy minimisation, and
- (ii) a standard distance-restrained refinement using 203 distances of the model structure as upper bounds constraints with the same refinement scheme (EM, 5ps MD, EM).

R-factors calculated on interresidue peaks are presented in Fig. 1 for the different structures during the refinement. For the direct NOE refinement, results are given for both the exact NOE calculation and the approximation of Eq. 2 using the spin diffusion correction. In the latter, the spin diffusion correction was updated every 20 steps in the molecular dynamics simulation. In both cases, the direct NOE refinement gives better R-factors. The convergence is presented in Fig. 2 where the distances in the starting and refined structures are plotted against the distances in the target structure. The apparently larger number of lower bounds violations in the structure obtained with standard distance restraining can be due to the absence of lower bounds constraints. Figure 2 also shows the corresponding structures superimposed to the target structure. It can be seen that the DINOSAUR-refined structure is better than that from distance restrained MD and is virtually identical with the target structure (RMSD 0.5 Å for all atoms, 0.1 Å for the backbone atoms).

We tested the influence of noise on the NOE data set. The noise level can be expressed in terms of an R-factor (R_{noise}). R_{noise} values below 0.15 do not affect the convergence very much. The choice of the correlation time is more important and this parameter should be optimised before and during the refinement.

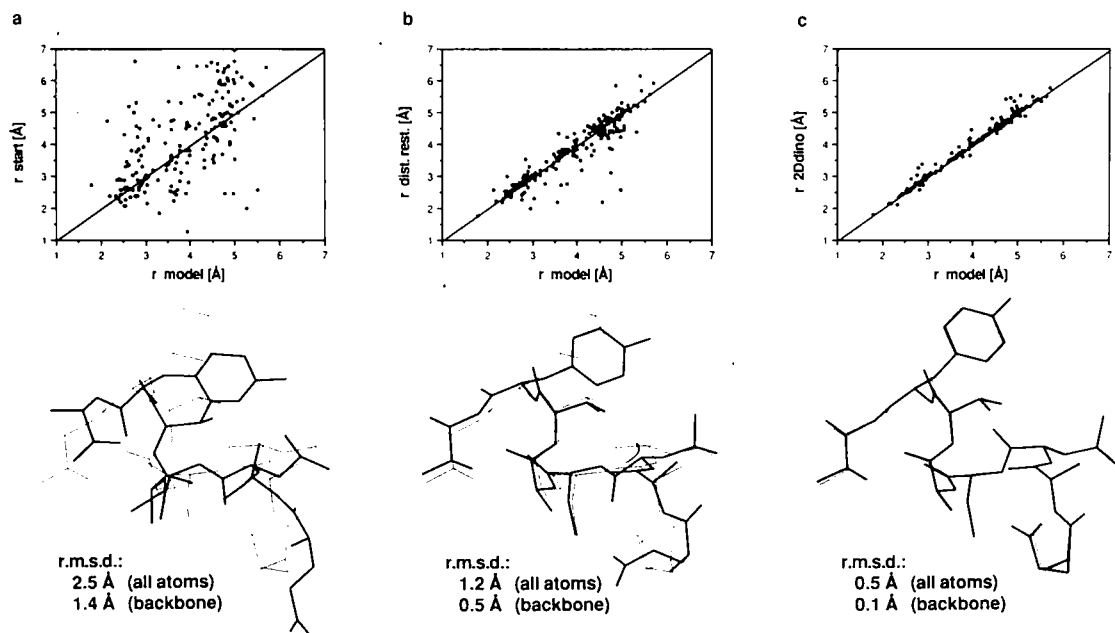


Fig. 2. Comparison of the distances in the target structure with (a) the starting structure, (b) the standard distance restraining structure and (c) the DINOSAUR structure with exact NOE calculation. The corresponding structures (broad lines) are presented superimposed to the target structure.

The entire refinement procedure with exact calculation of the NOE intensities required 77 min cpu time on an IRIS 4D/25. This corresponds to 0.85 s for one refinement step. If the spin diffusion correction is updated every 20 steps, this value decreases to 0.49 s; however, the RMSD with respect to the target structure increases (0.9 Å for all atoms and 0.5 Å for the backbone atoms). A standard distance-restrained refinement step requires 0.16 s. The diagonalisation of the relaxation matrix is the time-consuming procedure and has an N^3 dependency in time on the dimension N of the matrix. The DINOSAUR procedure is also being applied to crambin which contains 315 protons; 10 ps NOE restrained MD simulation (10000 steps) with updating of the spin diffusion correction every 10 steps can be computed in 8.5 h on a CONVEX C210.

The DINOSAUR approach appears a valuable method for structure refinement with NMR. The direct use of 2D NOE data in a refinement procedure takes into account the spin diffusion processes and avoids the transformation of NOE intensities into distances. The simplified derivative of the NOE intensities and the approximation proposed for the calculation of the NOE intensities allows for a considerable reduction of the computing time and should extend the application domain to larger molecules.

ACKNOWLEDGEMENTS

The authors thank Ton Rullmann and Joost van Opheusden for many contributions to the project. This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO).

REFERENCES

- Baleja, J.D., Moulton, J. and Sykes, B.D. (1990) *J. Magn. Reson.*, **87**, 375–384.
Boelens, R., Koning, T.M.G. and Kaptein, R. (1988) *J. Mol. Struct.*, **173**, 299–311.
Gonzalez, C., Rullmann, J.A.C., Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1991) *J. Magn. Reson.*, **91**, 659–664.
van Gunsteren, W.F. and Berendsen, H.J.C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biosmos BV, Nijenborgh 16, 9747 AG Groningen, The Netherlands.
Kaptein, R., Boelens, R., Scheek, R.M. and van Gunsteren, W.F. (1988) *Biochemistry*, **27**, 5389–5395.
Koning, T.M.G., Boelens, R. and Kaptein, R. (1990) *J. Magn. Reson.*, **90**, 111–123.
Kumar, A., Wagner, G., Ernst, R.R. and Wüthrich, K. (1981) *J. Am. Chem. Soc.*, **103**, 3654–3658.
Macura, S. and Ernst, R.R. (1980) *Mol. Phys.*, **41**, 95–117.
Olejniczak, E.T., Gaupe, R.T. Jr and Fesik, S.W. (1986) *J. Magn. Reson.*, **67**, 28–41.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Yip, P. and Case, D.A. (1989) *J. Magn. Reson.*, **83**, 643–648.