

## Direct Structure Refinement Using 3D NOE–NOE Spectra of Biomolecules

A. M. J. J. BONVIN, R. BOELENS, AND R. KAPTEIN

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

Received July 31, 1991

In the application of NMR to the structure determination of biomolecules of increasing size, new three-dimensional NMR experiments have been developed to overcome the overlap problems and enhance the resolution (1–5). Thus far, most 3D NMR techniques have been used in a qualitative sense, e.g., the identification of NOE cross peaks for the sequential assignment of proton resonances or the identification of long-range NOEs in proteins. Only in a few cases have cross-peak intensities been used quantitatively to derive constraints needed by structure refinement methods (6, 7). Recently, we described a direct NOE simulation approach for structure refinement (DINOSAUR) (8) based on the direct use of cross-peak intensities obtained from 2D NOE spectra as first proposed by Yip and Case (9). Here we extend this approach to 3D spectroscopy.

The intensities of cross peaks in a 3D spectrum are proportional to the transfer efficiencies of the two mixing periods of the experiment. The net transfer can be described in general as

$$T_{ijk}^{3D} = T_{ij}(1)T_{jk}(2), \quad [1]$$

where  $T_{ij}(1)$  and  $T_{jk}(2)$  describe the magnetization-transfer efficiencies for each separate mixing period and correspond to NOE transfer, exchange, or  $J$  transfer. The use of cross-peak intensities in a structure refinement procedure requires calculation of the gradients of these 3D intensities with respect to the coordinates of the protons in a molecule. The gradient of a 3D peak can be written as the sum of two terms containing the separate gradients of  $T_{ij}(1)$  and  $T_{jk}(2)$ :

$$\nabla T_{ijk}^{3D} = T_{jk}(2)\nabla T_{ij}(1) + T_{ij}(1)\nabla T_{jk}(2). \quad [2]$$

These equations can be generalized for even higher-dimensional experiments. Thus, given a valid model for the calculation of the magnetization-transfer efficiencies and their derivatives, 3D peaks can be used as constraints in a direct refinement procedure. In particular, 3D NOE–NOE spectra appear to be suitable for this purpose. This latter technique was developed for the direct observation of indirect NOE magnetization transfer (“spin diffusion”) as occurs in larger biomolecules (10) and has been applied to oligonucleotides (10) and proteins (11). In a recent paper Kessler *et al.* (7) described the quantitative use of 3D NOE–NOE cross peaks, reducing contributions from spin diffusion in NOEs. These NOEs can be translated into constraints, which are used by

distance geometry of restrained molecular dynamics computations. We will show here that it is also possible to directly refine the intensities of such 3D spectra.

In a 3D NOE–NOE spectrum the 3D cross-peak intensities  $A_{ijk}^{3D}$  can be calculated from their separate 2D NOE transfer efficiencies. The exponential matrix of normalized 2D NOE intensities  $A_{ij}$  can be written as (12)

$$\mathbf{A} = \exp(-\tau_m \mathbf{R}), \quad [3]$$

where  $\mathbf{R}$  represents the cross-relaxation matrix and  $\tau_m$  the mixing period. The exponential matrix of Eq. [3] can be expanded in a power series,

$$\exp(-\tau_m \mathbf{R}) = 1 - \tau_m \mathbf{R} + \frac{1}{2} \tau_m^2 \mathbf{R}^2 \dots, \quad [4]$$

but an exact solution can be obtained as well by diagonalization of  $\mathbf{R}$ ,

$$\mathbf{A} = \mathbf{X} \exp(-\tau_m \mathbf{\Lambda}) \mathbf{X}^{-1}, \quad [5]$$

where  $\mathbf{\Lambda} = \mathbf{X}^{-1} \mathbf{R} \mathbf{X}$  is the diagonal eigenvalue matrix. Given a structure and a motional model the cross-relaxation matrix can be calculated. The computation of the 3D NOE–NOE intensity  $A_{ijk}^{3D}$  then amounts to the multiplication of the calculated 2D contributions  $A_{ij}$  and  $A_{jk}$ .

For the structure refinement a similar NOE-restraining potential function is used, as was suggested for the direct refinement with 2D NOE data (8),

$$V_{3DNOE} = \left[ \sum_{ijk} w_{ijk} \right]^{-1} \sum_{ijk} w_{ijk} (A_{ijk}^{3D\text{theo}} - A_{ijk}^{3D\text{exp}})^2, \quad [6]$$

with

$$w_{ijk} = \frac{1}{(N + \epsilon A_{ijk}^{3D\text{exp}})^2}, \quad [7]$$

where  $N$  represents the experimental noise level and  $\epsilon$  a relative error on the experimental NOE intensities. The relative term accounts for integration errors. Forces used in the refinement can be derived from  $V_{3DNOE}$  by calculating the gradient of this function. We need therefore to calculate the derivative of 3D NOE–NOE transfer functions with respect to the coordinates, according to Eq. [2]. An exact solution for the derivative of 2D NOE intensities has been given by Yip and Case (9). It consists of a product of five matrices and becomes extremely time consuming when the dimension of the problem increases. However, we recently proposed (8) for the direct NOE refinement with 2D data an approximation for the derivative of the NOE intensity based on the first two terms of the expansion of the exponential function (Eq. [4]):

$$\nabla A_{ij} = -\tau_m \nabla R_{ij}. \quad [8]$$

Combining this with Eq. [2] gives the gradient of a 3D NOE–NOE intensity,

$$\nabla A_{ijk}^{3D} = -[\tau_{m1} A_{jk}(\tau_{m2}) \nabla R_{ij} + \tau_{m2} A_{ij}(\tau_{m1}) \nabla R_{jk}], \quad [9]$$

and the gradient of  $V_{3DNOE}$  becomes

$$\begin{aligned} \nabla V_{3DNOE} = & \left[ \sum_{ijk} w_{ijk} \right]^{-1} \sum_{ijk} -2w_{ijk} (A_{ijk}^{3D\text{theo}} - A_{ijk}^{3D\text{exp}}) \\ & \times [\tau_{m1} A_{jk}(\tau_{m2}) \nabla R_{ij} + \tau_{m2} A_{ij}(\tau_{m1}) \nabla R_{jk}]. \quad [10] \end{aligned}$$

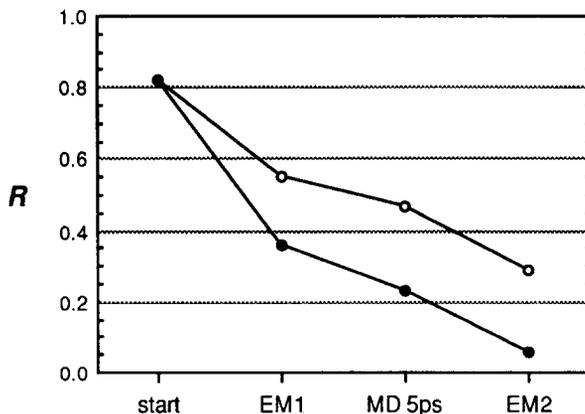


FIG. 1.  $R$  factors according to Eq. [11] calculated on interresidue peaks at different stages during the refinement. ●, Standard distance restraining; ○, DINOSAUR refinement based on 3D NOE-NOE intensities.

We implemented this 3D NOE-NOE restraining potential in a 3D version of our DINOSAUR programs (*direct NOE simulation approach for structure refinement*) (8) consisting of three different procedures: NOE-driven dynamics based on the distance-bound-driven dynamics program (13), where only the covalent information is used in addition to the experimental constraints, NOE-restrained energy minimization, and NOE-restrained molecular dynamics which make use of the GROMOS force field (14).

Due to the nature of the NOE intensities, which have an inverse sixth power dependence on the distances,  $V_{3\text{DNOE}}$  is an asymmetric function which may generate very large forces if the distances in the model structure are underestimated. As observed for the refinement with 2D NOE data as well, it is necessary to limit the NOE forces in the simulations. This corresponds to a linearization of the NOE potential above a certain value.

For a test of the quality of fit we use a 3D NMR  $R$  factor similar to what we have proposed previously for 2D data (15):

$$R^{3\text{D}} = \frac{\sum_{ijk} |A_{ijk}^{3\text{Dtheo}} - A_{ijk}^{3\text{Dexp}}|}{\sum_{ijk} A_{ijk}^{3\text{Dexp}}} \quad [11]$$

This  $R$  factor has the advantage that it is independent of the simulation parameters ( $N$ ,  $\epsilon$ ) and thus allows the comparison of various computational techniques.

For the calculation, we chose as a model target structure an eight-residue  $\alpha$  helix (56 protons) corresponding to the second helix of crambin. A synthetic 3D NOE-NOE data set was generated at 500 MHz with mixing times of 0.12 and 0.16 s for the two NOE transfers, respectively, using a correlation time of 3 ns, which resulted in a set of 3526 3D cross peaks corresponding to transfers up to 4.0 Å. A start structure was generated, which deviates by 2.5 Å for all atoms (1.4 Å for backbone atoms) from the original model. This was then used in a direct NOE refinement procedure including

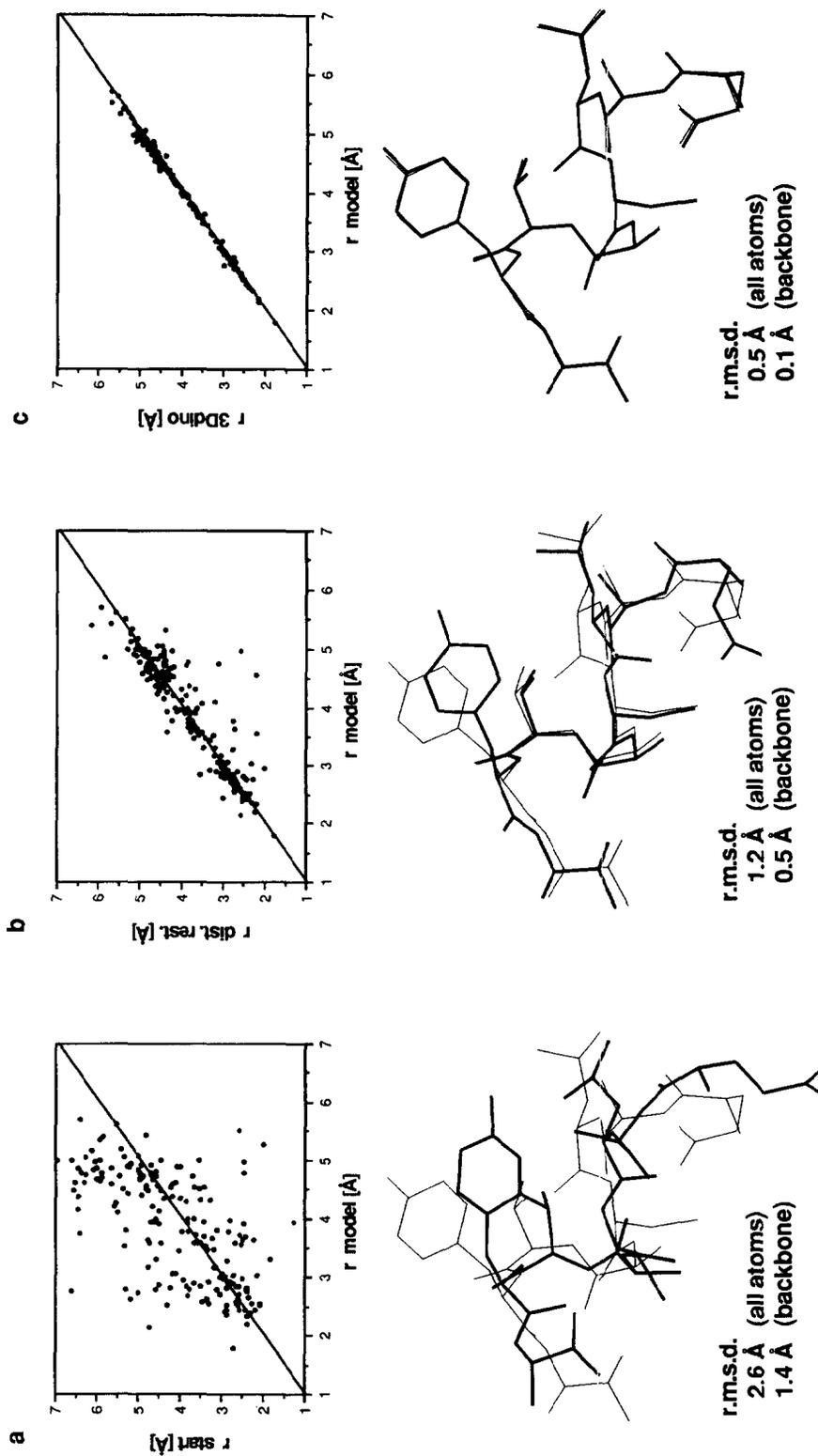


FIG. 2. Comparison of the distances in the model structure with (a) the start structure, (b) the standard distance-restraining structure, and (c) the 3D DINOSAUR structure. The corresponding structures (thick lines) are superimposed on the target structure.

300 steps of restrained energy minimization, 5 ps of restrained molecular dynamics with an integration step of 0.001 ps, followed finally by 300 steps of restrained energy minimization. For comparison, a standard distance-restrained refinement using 203 distances of the model structure as upper bound constraints was performed using the same procedure. Results in terms of  $R$  factors are presented in Fig. 1 for the different structures during the refinement. Clearly, the 3D DINOSAUR procedure gives a very low  $R$  factor. The convergence is presented in Fig. 2, where the distances in the start and refined structures are plotted against the distances in our target structure. For the structure obtained with 3D DINOSAUR the fit is almost perfect. The apparent large number of lower-bound violations in the structure obtained with standard distance restraining is probably due to the absence of lower-bound constraints. Figure 2 also presents a view of the corresponding structures superimposed on the target structure. The DINOSAUR-refined structure is virtually identical to the target structure (rmsd 0.5 Å for all atoms, 0.1 Å for backbone atoms). Thus, 3D DINOSAUR appears to be a potentially very useful method for the direct refinement of 3D NOE–NOE data.

The method could be sensitive to various factors related to the quality of the 3D experiment and to parameters chosen for the simulations. We tested the influence on the NOE data set of noise of the choice of the correlation time and of the size of the data set. The noise level can be expressed in terms of an  $R$  factor ( $R_{\text{noise}}$ ). Values for  $R_{\text{noise}}$  below 0.15 do not affect the convergence significantly. The correlation time is more important and this parameter should certainly be optimized before and during the refinement. The same refinement procedure as that described above using transfers up to 4.0 Å was repeated with two smaller 3D NOE–NOE data sets corresponding to transfers of 3.5 Å (2719 peaks) and 3.0 Å (882 peaks), respectively. A very good convergence is observed with the set of 2719 peaks. With the data set corresponding to two transfers up to 3.0 Å, the procedure fails to find the right conformation. The information in the latter case is not sufficient to determine the three-dimensional structure precisely.

The entire refinement procedure with exact calculation of the NOE intensities required two and one-half hours cpu on a Silicon Graphics Personal IRIS 4D/25 for an eight-residue system with 56 protons. This corresponds to 1.65 s for one refinement step. The procedure scales as  $n^3$  problem for large  $n$ .

The use of 3D NOE–NOE intensities in a direct refinement procedure as implemented in DINOSAUR appears to be a promising method for structure refinement. The main problem in the application of the method lies in the assignment of cross peaks of a 3D NOE–NOE spectrum. The huge amount of data contained in a 3D experiment requires an efficient program for automatic assignment and peak picking before a real refinement based on 3D NOE–NOE intensities can be undertaken. Such programs are presently in development and will be used to test the current method on an experimental data set.

#### ACKNOWLEDGMENTS

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

#### REFERENCES

1. H. OSCHKINAT, C. GRIESINGER, P. J. KRAULIS, O. W. SØRENSEN, R. R. ERNST, A. M. GRONENBORN, AND G. M. CLORE, *Nature* **332**, 374 (1988).

2. C. GRIESINGER, O. W. SØRENSEN, AND R. R. ERNST, *J. Magn. Reson.* **84**, 14 (1989).
3. L. E. KAY, D. MARION, AND A. BAX, *J. Magn. Reson.* **84**, 72 (1989).
4. S. W. FESIK AND E. R. P. ZUIDERWEG, *Q. Rev. Biophys.* **23**, 97 (1990).
5. R. BOELENS, G. W. VUISTER, A. PADILLA, G. J. KLEYWEGT, P. DE WAARD, T. M. G. KONING, AND R. KAPTEIN, "Structure and Methods. Vol. 2. DNA Protein Complexes and Proteins," p. 63, Adenine Press, Guilderland, New York, 1990.
6. P. DE WAARD, "Structural Studies on Carbohydrates by Means of 1D, 2D and 3D NMR Spectroscopy," Ph.D. Thesis, University of Utrecht, 1990.
7. H. KESSLER, S. SEIP, AND J. SAULITIS, *J. Biomol. NMR* **1**, 83 (1991).
8. A. M. J. J. BONVIN, R. BOELENS, AND R. KAPTEIN, *J. Biomol. NMR*, submitted.
9. P. YIP AND D. CASE, *J. Magn. Reson.* **83**, 643 (1989).
10. R. BOELENS, G. W. VUISTER, T. M. G. KONING, AND R. KAPTEIN, *J. Am. Chem. Soc.* **111**, 8525 (1989).
11. J. BERG, R. BOELENS, G. W. VUISTER, AND R. KAPTEIN, *J. Magn. Reson.* **87**, 646 (1990).
12. S. MACURA AND R. R. ERNST, *Mol. Phys.* **41**, 95 (1982).
13. R. KAPTEIN, R. BOELENS, R. M. SCHEEK, AND W. F. VAN GUNSTEREN, *Biochemistry* **27**, 5389 (1988).
14. W. F. VAN GUNSTEREN AND H. J. C. BERENDSEN, "Groningen Molecular Simulation (GROMOS) Library Manual," Biomos BV, Groningen, Nijenborgh 16, 9747 AG, The Netherlands, 1987.
15. C. GONZALES, J. A. C. RULLMANN, A. M. J. J. BONVIN, R. BOELENS, AND R. KAPTEIN, *J. Magn. Reson.* **91**, 659 (1991).