Journal of Biomolecular NMR, 2 (1992) 477–484
ESCOM

J-Bio NMR 077

Quantitative comparison of experimental and simulated NOE intensities: Correlation with accuracy of oligonucleotide structure determination

R. Nibedita, R. Ajay Kumar, A. Majumdar and R.V. Hosur*

Chemical Physics Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Bombay 400 005, India

Received 6 January 1992
Accepted 9 June 1992

Keywords: DNA structure; NOESY simulation

SUMMARY

The relation between the match of experimental and simulated NOE intensities with the accuracy of structure determination of oligonucleotides has been investigated. A hypothetical experimental spectrum of the oligonucleotide d(CCAACGTTGG) from its known X-ray crystallographic structure (Privé, G.G. et al. (1991) *J. Mol. Biol.*, **217**, 177–199) has been generated with simplifying assumptions of single correlation time, leakage rate etc., and this spectrum has been simulated imposing various constraints in a manner as one would do in a real case. The hypothetical spectrum represents the case of an infinitely good experimental spectrum and therefore the study of quality of fit against quality of structure represents the limiting case of real situations. It has been shown that even with a limited number of NOEs, it is possible to approach the correct structure of the molecule by demanding a highly accurate fit between experimental and simulated NOE intensities. Distance geometry calculations have been used to probe the extent of structural degeneracies in the NOE intensity matching exercise.

INTRODUCTION

NOE-based refinement has come to be accepted as the most reliable approach to structure determination of biological macromolecules (Nerdal et al., 1989; Baleja et al., 1990; Borgias et al., 1990; Gochin and James, 1990; Bonvin et al., 1991; Mertz et al., 1991; Withka et al., 1992). The standard protocol of refinement looks as follows: first, a set of intramolecular interproton distance constraints is derived by a crude interpretation of NOE intensities at short mixing times, using either the initial-rate approximation (Macura and Ernst, 1980; Anil Kumar et al., 1981) or the quadratic approximation (Majumdar and Hosur, 1989). Then, incorporating these constraints, a family of consistent structures is derived either by distance geometry (Boelens et al.,

* To whom correspondence should be addressed.

1988; Pardi et al., 1988; Banks et al., 1989) or by restrained molecular dynamics calculations (Nilsson et al., 1986; Nilges et al., 1987; Boelens et al., 1989; Baleja et al., 1990; Gochin and James, 1990). These structures are sometimes further screened on the basis of energy (in molecular dynamics calculations energy minimisation is already included) and a smaller set of possible structures is obtained. In the final step the NOE intensities are directly compared with simulated NOE intensities for these structures and the structures are altered to get the best fit of experimental and simulated intensities. Suitable functions (R-factors) are defined to reflect the deviation of simulat-

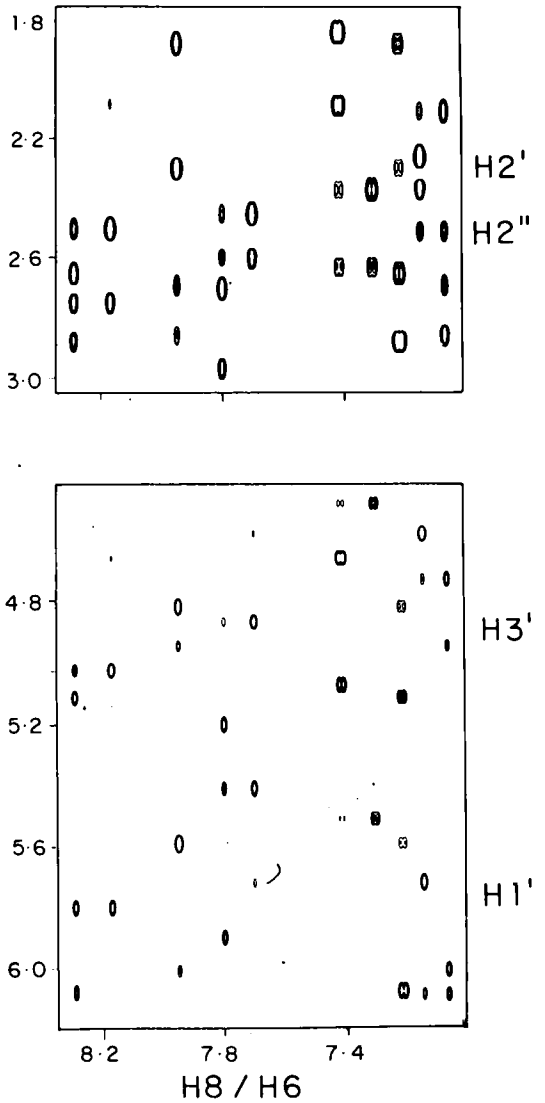

Fig. 1. Regions of a simulated NOESY spectrum at 500 MHz of the test DNA segment, d(CCAACGTTGG) whose crystal structure has been determined to high resolution (Privé et al., 1991). The spectrum has been generated by relaxation matrix calculations, assuming convenient resonance assignments for the various protons so as to avoid overlap of peaks. The mixing time is 300 ms and a single correlation time of 2 ns has been assumed.

ed intensities from experimental intensities and these are minimised to obtain optimised structures. In many cases good convergence to similar structures starting from different initial structures has been achieved (Banks et al., 1989; Nerdal et al., 1989; Baleja et al., 1990; Gochin and James, 1990).

However, good convergence alone does not ensure correctness of a structure; a good fit is required between observed and calculated NOE intensities, although it cannot be demanded to be better than the uncertainties in the observed intensities. In many of the above studies we have noticed with concern that even after extensive NOE refinements the residual individual NOE deviations are fairly high, between 10 and 30% for different peaks in the spectrum. If this represents the uncertainty in the experimental intensities, it is a rather large value and casts at least as much doubt on the validity of the structures derived. The observed deviations may be attributed in part to the inaccuracies in the structure and the rest to the methods used in quantifying experimental intensities, scaling techniques used in comparison of simulated and experimental intensities, the assumptions used in relaxation matrix calculations employed in computing NOE intensities and the number of NOEs used in the analysis. In this context, we have described in the preceding paper (Nibedita et al., 1992) a realistic approach for comparison of experimental and simulated NOE intensities. Recently Withka et al. (1992) emphasised the influence of anisotropic motions in DNA segments on the intensities of NOE cross peaks. At the end of these corrections, the residual deviations must be attributed to unknown dynamical parameters of the system under investigation.

In view of the above discussion, we focus on the relation between the goodness of fit between experimental and simulated NOE intensities and the quality of the derived structure. We search for a working relation between the quality of fit and the quality of structure within the constraint that the fit cannot be better than the uncertainties in the experimental NOE intensities. Such a question cannot be addressed with a real case because of the large number of unknowns including the true structure. We therefore consider here the generation of a 'target' spectrum of a DNA segment whose structure has been precisely determined by X-ray crystallography. The target spectrum can be generated with whatever test conditions needed and so the uncertainties about the system can be eliminated in the above exercise. The generated target spectrum would correspond to a real situation where the system parameters such as correlation times, leakage rates, relaxation times etc. are fully known and the experimental spectrum is the best possible one can ever have, with all the peaks well resolved, the noise level extremely low, the digital resolution very high compared to the line widths of the peaks etc. The deviations between experimental and simulated intensities would then be entirely due to the differences in the derived and the true structure. Recently two such exercises have been reported on proteins (Bonvin et al., 1991; Mertz et al., 1991) and the present one is the first with regard to DNA.

THE TARGET SPECTRUM

Figure 1 shows as illustrations some portions of the target spectrum at 500 MHz of a test DNA segment, d(CCAACGTTGG) whose crystal structure has been determined to very high resolution (Privé et al., 1991). Using the reported torsion angles, an interproton distance matrix has been calculated and then the expected interproton NOE intensities have been computed by relaxation matrix calculations. A fixed correlation time of 2 ns was assumed and the mixing time was 300 ms. The spectrum has been generated by assuming convenient resonance assignments and the

peaks were generated in the digital fashion as described in the preceding paper (Nibedita et al., 1992). Separate thresholds were selected for each peak and the peaks were integrated by summing all the points above the respective thresholds in each peak. The list of these peaks is used as a list of experimental peaks for the simulations. A total of 114 peaks from the regions H8/H6 – (H2′, H2″, H1′, H3′), H1′ – (H2′,H2″), and H4′ – (H5′,H5″) have been included in the analysis.

QUALITY OF FIT VS. QUALITY OF STRUCTURE

The analysis of the spectrum has been performed in two ways to drive home the point of interest. First, the target structure is modified by a less than 5° random variation of all torsion angles and the corresponding NOESY spectrum is calculated and compared with the original target spectrum. The intensity deviations of the target and simulated structures are shown in Fig. 2b. It is observed that the intensities are a very sensitive function of the torsion angles. Such an exercise is repeated for randomisation of torsion angles of the oligonucleotide by 2°. The corresponding match of target and simulated NOE intensities is shown in Fig. 2a. The rmsd (root-mean-square deviation) of NOE intensity between target and simulated spectrum is less in this case compared to the previous case of 5° randomisation of torsion angles (see figure caption for definition of rmsd).

In the second stage, the target spectrum is simulated mimicking the real exercise, by optimising the structure starting from an initial B-DNA structure. The structure optimisation is done interactively, continuously monitoring the improvements in the fit. In this process the sugar geometries and the ε torsion angles along the backbone are constrained exactly since this information is nor-

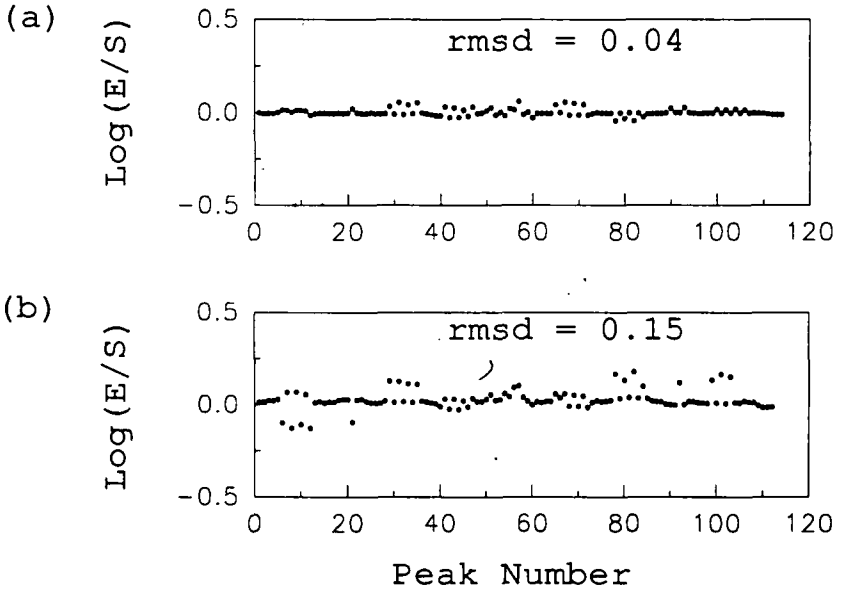

Fig. 2. Deviations in digitised simulated NOE intensities from the target intensities due to the structural changes created by random variation of all torsion angles within (a) 2° and (b) 5°. The target (E) and digitised simulated (S) NOE intensities have been obtained by the method described in the text. Root-mean-square deviation of NOE intensities between target and simulated spectra is defined as rmsd = $[\Sigma_i(1 - E(i)/S(i))^2/N]^{1/2}$ where N is the total number of peaks.

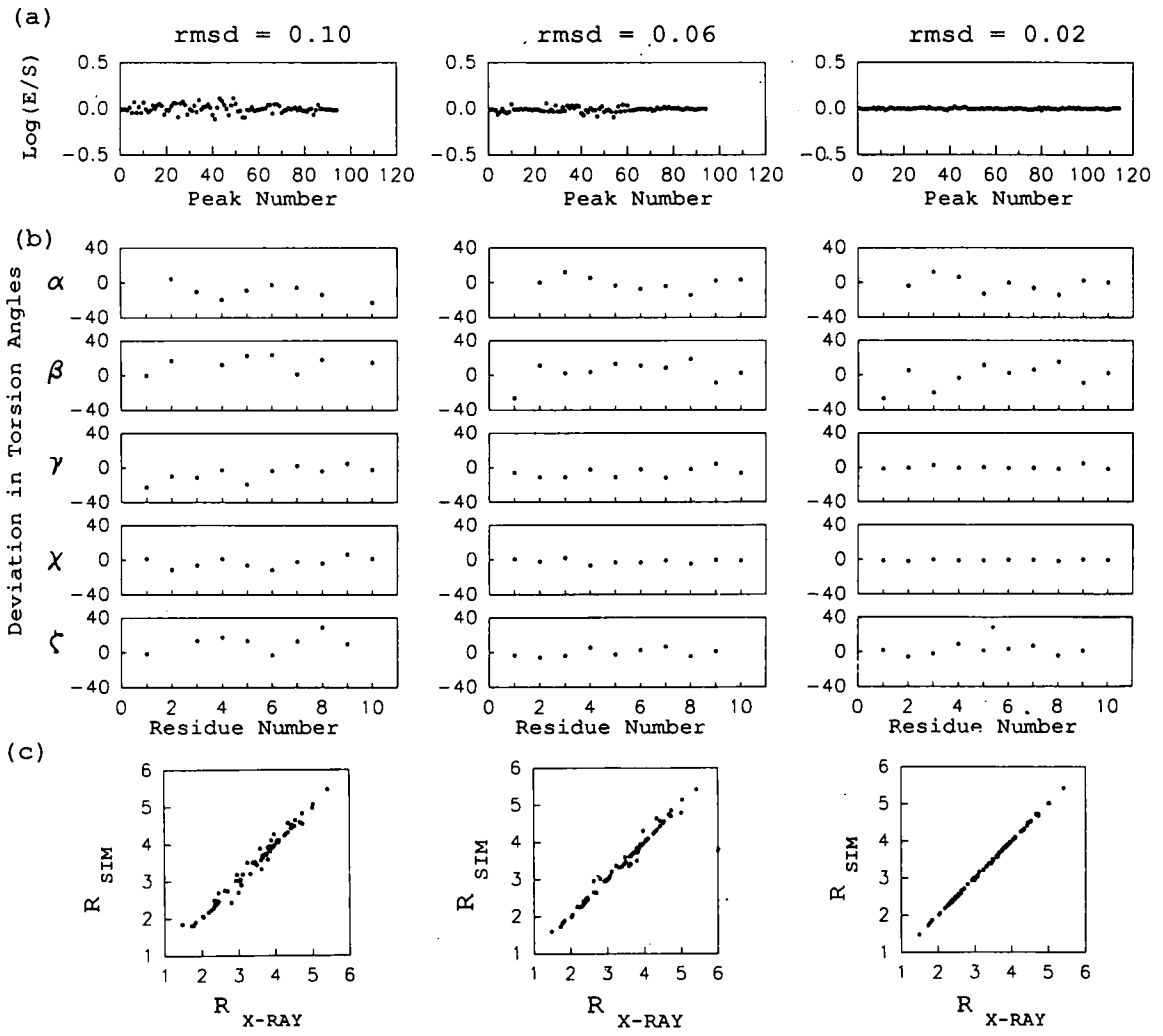

Fig. 3. Quality of fit between target NOE intensities (E) and the digitised simulated intensities (S) during interactive structure optimisation vs. the correctness of the structure at three stages of optimisation corresponding to rmsd values 0.10, 0.06, and 0.02. The three panels describe: (a) a plot of Log(E/S) vs. peak number to show the deviation of simulated intensities from target values, (b) deviations in the torsion angles and (c) the deviations in the interproton distances shown as a diagonal plot, respectively.

mally available by independent means. In addition, the duplex constraint is imposed since the DNA segment is self-complementary and is expected to exist as a duplex in solution under the normal conditions of temperature. Figure 3 shows the fits at three different stages of optimisation. In each case, the NOE intensity deviations, the torsion angle deviations from those in the target structure and the correlations between the interproton distances in the target and optimised structure, are shown. It may be noted that all torsion angles are not determined to the same level of accuracy, with the glycosidic torsion angle being most accurately determined. Nonetheless, the structure with an NOE rmsd of 0.02 can be considered to be the closest to the target structure.

Comparison of the data in Figs. 2 and 3 reveals some interesting points. An rmsd of 0.02 in Fig. 3, which is the lowest of all the rmsd values, corresponds to a structure that has larger deviations with respect to some torsion angles than those in either of the two cases in Fig. 2 where the maximum deviation of any torsion angle is only 2° in (a) and 5° in (b). This means that small changes in some torsion angles have compensatory effects for changes in other torsion angles with regard to overall NOE intensities. In other words, with the number of NOEs chosen, an rmsd of 0.02, although corresponding to a structure close to the target structure, cannot be considered to be unique. Considering that a good rmsd of NOE intensities implies a good set of interproton distances in the entire network, the above observation suggests that the degeneracy in the structures is not as much with respect to distances but more with respect to the torsion angles. There may be other structures with smaller and different deviations in the torsion angles from those in the target structure giving nearly the same rmsd value of 0.02. In order to probe this, we carried out distance geometry calculations using the program TANDY (Ajay Kumar et al., 1991) using all the interproton distances (including those for which the NOE is not actually seen or used) in the structure corresponding to an rmsd of 0.02 and 5 structures were generated with very tight limits of ± 0.1 Å over the distances. These are displayed superimposed over each other in Fig. 4b to show the spread in the structures. To provide a better feel for these deviations 5 structures have also been generated from the target structure by randomly varying all the torsion angles within 5° and these

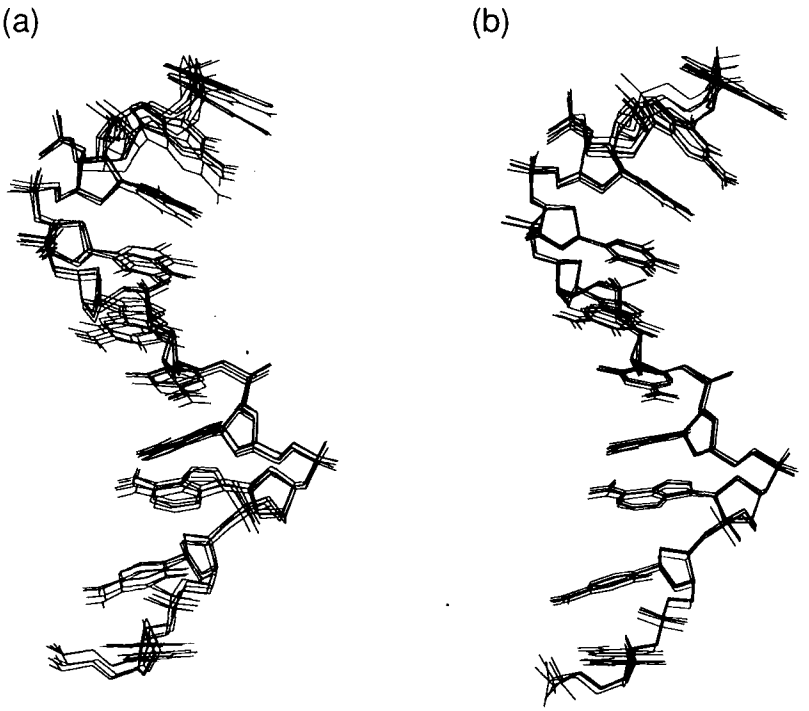

Fig. 4. From the structure corresponding to an rmsd of 0.02 in Fig. 3, a set of all interproton distance constraints has been obtained. A distance geometry program TANDY (Ajay Kumar et al., 1991) has been used with this distance constraint set to generate a family of consistent structures. The figure (b) shows a superposition of 5 such structures. For the sake of comparison, another set of 5 structures was generated by random flexing of all torsion angles in the oligonucleotide within 5°. This set of 5 structures is shown in (a). The structures generated by distance geometry seem comparable with set (a).

are shown superimposed in Fig. 4a. All these 10 structures have distance rmsds between 1.5 and 2.4 Å with respect to the target structure. There is a structure in Fig. 4b which has a lower rmsd compared to many structures in Fig. 4a indicating that it corresponds to the structure closest to the target structure and is also consistent with NOE intensities.

In the above exercise, the system parameters were considered known; and they were included in the calculations. Hence these parameters were of no significance for determining the quality of fit. In real situations however, this is rarely the case and they will have to be optimised in addition to the structure of the molecule. Some information on these parameters can be obtained by relaxation time measurements (Woessner, 1962), coupling constant measurements (Rinkel and Altona, 1987; Hosur et al., 1988; van de Ven and Hilbers, 1988; Majumdar and Hosur, 1992) and also by other experimental techniques such as nuclear quadrupole resonance (NQR) and Raman spectroscopy. The uncertainties in these parameters warrant some tolerance for the quality of fit between experimental and calculated spectra.

CONCLUSIONS

We have shown that in DNA the NOE intensities are a sensitive function of the torsion angles in the molecule and therefore a reliable characterisation of the structure from NOE data requires accurate quantification of the NOE intensities and a very high level of fitting between experimental and simulated intensities. It is also demanding on the quality of the spectra with regard to signal-to-noise ratios, digital resolutions, and appropriate realistic means of comparing simulated and experimental spectra. At the same time it is also encouraging to note that, even with a limited size of the data set, a fairly good reproduction of the target structure has been obtained.

ACKNOWLEDGEMENTS

The facilities provided by the 500 MHz FTNMR National Facility supported by the Department of Science and Technology, Government of India are gratefully acknowledged.

REFERENCES

Ajay Kumar, R., Hosur, R.V. and Govil, G. (1991) *J. Biomol. NMR*, **1**, 363–378.
Anil Kumar, Wagner, G., Ernst, R.R. and Wüthrich, K. (1981) *J. Am. Chem. Soc.*, **103**, 3654–3656.
Baleja, J.D., Germann, M.W., van de Sande, J.H. and Sykes, B.D. (1990) *J. Mol. Biol.*, **215**, 411–428.
Banks, K.M., Hare, D.R. and Reid, B.R. (1989) *Biochemistry*, **26**, 6996–7002.
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. (1991) *J. Biomol. NMR*, **1**, 305–309.
Borgias, B.A., Gochin, M., Kerwood, D.J. and James, T.L. (1990) *Prog. NMR Spectrosc.*, **22**, 83–100.
Gochin, M. and James, T.L. (1990) *Biochemistry*, **29**, 11172–11180.
Hosur, R.V., Govil, G. and Miles, H.T. (1988) *Magn. Reson. Chem.*, **26**, 927–944.
Macura, S. and Ernst, R.R. (1980) *Mol. Phys.*, **41**, 95–117.
Majumdar, A. and Hosur, R.V. (1989) *Biochem. Biophys. Res. Commun.*, **159**, 886–892.
Majumdar, A. and Hosur, R.V. (1992) *Prog. NMR. Spectrosc.*, **24**, 109–158.
Mertz, J.D., Guntert, P., Wüthrich, K. and Braun, W. (1991) *J. Biomol. NMR*, **1**, 257–269.
Nerdal, W., Hare, D.R. and Reid, B.R. (1989) *Biochemistry*, **28**, 10008–10021.

Nibedita, R., Ajay Kumar, R., Majumdar, A. and Hosur, R.V. (1992) *J. Biomol. NMR*, **2**, 467–476.
Nilges, M., Clore, G.M., Gronenborn, A.M., Brunger, A.T., Karplus, M. and Nilsson, L. (1987) *Biochemistry*, **26**, 3718 3733.
Nilsson, L., Clore, G.M., Gronenborn, A.M., Brunger, A.T. and Karplus, M. (1986) *J. Mol. Biol.*, **188**, 455–475.
Pardi, A., Hare, D.R. and Wang, C. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 8785–8789.
Privé, G.G., Yanagi, K. and Dickerson, R.E. (1991) *J. Mol. Biol.*, **217**, 177–199.
Rinkel, L.J. and Altona, C. (1987) *J. Biomol. Struct. Dyn.*, **4**, 621–649.
Van de Ven, F.J.M. and Hilbers, C.W. (1988) *Eur. J. Biochem.*, **178**, 1–38.
Withka, J.M., Swaminathan, S., Srinivasan, J., Beveridge, D.L. and Bolton, P.H. (1992) *Science*, **255**, 597–599.
Woessner, D.E. (1962) *J. Chem. Phys.*, **36**, 1–4.