A Hybrid Complete Relaxation Matrix Structural Refinement of an Extra-helical Adenosine Tridecamer d(CGC*A*GAATTCGCG)$_2$ From 2D ^{1}H NMR

Edward Nikonowicz, Robert Meadows and David G. Gorenstein*

Department of Chemistry
Purdue University
West Lafayette, Indiana 47907

Until very recently interproton distances from NOESY experiments have been derived solely from the two-spin approximation method. The approximation requires that the NOESY derived distances be obtained from vanishingly short experimental mixing times where the build-up of NOE intensity is proportional to the inverse 6th power of the interproton distance and the effects of spin diffusion (NOE intensity mediated by multiple relaxation pathways) are minimal. Unfortunately, it has been shown that even at short mixing times there is a significant error in many of these distances. In order to avoid the approximation of the two-spin method we have used a complete relaxation matrix program (*MORASS*: *M*ultiple *O*verhauser *R*elaxation *A*nalysi*S* and *S*imulation) which employs a matrix eigenvalue/eigenvector solution to the Bloch equations. We have recently characterized the extrahelical adenosine containing tridecamer d(CGCAGAATTCGCG)$_2$ in aqueous solution using 2-D ^{1}H/^{1}H NMR by an iterative hybrid matrix refinement approach using MORASS and restrained molecular dynamics calculations (1, 2).

The 13-mer sequence is of particular interest since it has the potential to accommodate the extrahelical adenosines (position A3') by three different modes: 1) a stem-loop hairpin structure, 2) a "looped" structure and 3) a "stacked" structure. Precedence exists for each of these three possibilities concerning extrahelical bases from both NMR and X-ray crystallography (3, 4, 5).

We have assigned the non-exchangeable ^{1}H resonances of the tridecamer including some of the H5' and H5" resonances from the 250 ms and 400 ms t_m (mixing time) NOESY spectra (1). The sequential connectivities of the H8/H6 - H1' region of the spectrum were consistent with those expected for a right-handed B type DNA conformation. Additionally, at shorter mixing times, the H1'-H2'/2" and H6/H8-H2'/H2" crosspeak intensities supported a B type conformation. The sequential connectivity was continuous through the C3-A3'-G4 extrahelical step in the H6/H8- H1' region, but was weak at the C9-G10 step.

Two-dimensional nuclear magnetic resonance, combined with distance geometry or restrained molecular mechanics/dynamics (6-8), is now capable of elucidating the fine structure of short DNA strands in solution. In order to obtain more accurate distances than are available by the "two-spin approximation" analysis of NOESY data, we have envoked the use of a complete relaxation matrix approach for solving the Bloch equations of magnetization. The matrix approach removes the effects of spin diffusion allowing the measurement of interproton distance with a high degree of precision and accuracy (9, 10). However, the use of the complete matrix method is sensitive to the completeness of the experimental data. An effective solution has been provided by a "hybrid matrix approach" (2, 10–12).

Originally proposed by Kaptein and coworkers (11, 12), the hybrid matrix approach addresses the problem of incomplete experimental data by combining the information from the experimental NOESY volume matrix, $\mathbf{V}_{ij}^{exp}$, and calculated volume matrix, $\mathbf{V}_{ij}^{c}$, derived from an initial assumed structure. The well-resolved and measureable crosspeaks in the NOESY spectrum replace the corresponding crosspeaks in the calculated volume ma-

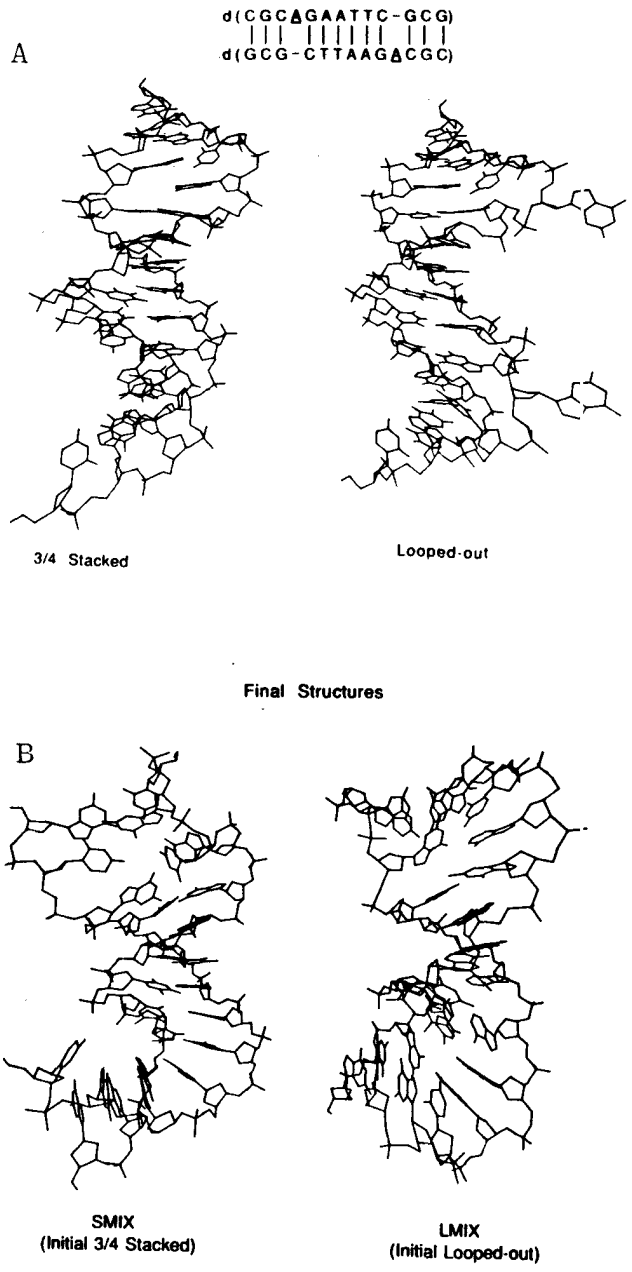

Figure 1: NOESY-distance restrained, molecular dynamics structures during the hybrid matrix/MORASS/MD refinement of the extra-helical adenosine tridecamer, starting from (A) the base loop out (LI) or 3/4 base stacked in model built structures (SI – the extra-helical adenosine is shown stacked between base pairs 3 and 4). (B) After the 9th iteration of the merged hybrid matrix/MORASS/MD refinement.

trix, while overlapping or weak crosspeaks and diagonals are from the calculated spectrum. This hybrid volume matrix, $\mathbf{V}_{ij}^{hyb}$, is then used to evaluate the rate matrix. Distances derived from this hybrid relaxation rate matrix are then utilized as input to a restrained molecular dynamics simulation for refinement of the structure. Energy minimization on the structure derived from molecular dynamics completes one cycle of refinement. This process can be repeated until a satisfactory agreement between the calculated and observed crosspeak volumes is obtained. The relaxation matrix program (*MORASS*) (13; the program is available upon request) was used to calculate volume and rate matrices as well as implement the hybrid matrix methodology. In several oligonucleotide duplexes (8-mer to 13-mer) studied by Kaptein and our group, (2, 10–12), 3-6 iterations are typically required to achieve convergence to a "unique" structure.

Two alternative models for the tridecamer were considered, which we define as the stacked (3/4-3/4) or looped out conformation (see Figure 1A). In the 3/4-3/4 stacked conformation, the extrahelical adenosine is stacked between the 3rd and 4th basepairs of the double helix. The starting input structure for the first MORASS iteration was generated using NOESY data taken with a 50 ms mixing time. 15 ps of 300K molecular dynamics (MD) was carried out on each of the minimized model built structures in 5 ps steps. The left and right force constants were gradually increased from 10 to 40 kcal/mol/Å^2 and error limits of the NOESY constraints were gradually decreased from 15% to 9%. The A3' residue on both strands of the minimized 5 ps averaged structure, SI, starting from the model built stacked structure remained stacked between residues C3 and G4, Figure 1A. However, starting from the initial looped model, one of the extrahelical adenosines has moved between base pairs G2-C11 and C3-G10 (2/3 stack) and the other extrahelical adenosine had formed a "triplex" base pairing scheme with C3 and G10. An additional 10 ps of constrained MD was carried out on structure LI to produce the averaged structure LII in which the "triplex" and 2/3 stack were maintained. It is intriguing that the "triplex" does not grossly disobey the constraints for the 50 ms NOESY distances and is energetically allowed. Indeed, at this point one is unable to confidently distinguish between the "triplex", the G2/C3 stacked

or the C3/G4 stacked structures based solely on 50 ms derived distances and restrained molecular dynamics calculations.

In order to obtain additional and more accurately integrated NOESY crosspeaks for further refinement of the structures, we have used the 150 ms NOESY spectrum where the crosspeaks have higher signal-to-noise and hence lower integration errors. At 150 ms, 258 NOESY constraints (per duplex) were measured. Only those crosspeaks that could be adequately resolved from overlapping peaks were included. An additional 12 imino hydrogen bond constraints were added and the total of 270 distance restraints were incorporated into the next stage of the refinement of the structures. Separate refinement of both loop-out and stacked models of the tridecamer was considered. The typical refinement follows the iterative merged matrix/restrained molecular dynamics methodology incorporating the NOESY distance constraints as described in the previous section.

Convergence is monitored using eqn. 1. This criterion is analogous to that used in x-ray crystallography.

$$\mathrm{RMS}_{vol} = \sqrt{\frac{1}{N}\sum_{ij}\left(\frac{v_{ij}^{c} - v_{ij}^{exp}}{v_{ij}^{c}}\right)^{2}} \qquad (1)$$

Convergence is achieved when the RMS_{vol} is within the reliability of the experimental volume measurement. We note that the 3/4-3/4 stacked structure [S(3/4-3/4)MIII] remains stacked throughout the iterative refinement. However in the 2nd iteration the "triplex"-2/3 structure formed a 2/3-3/4 structure [L(2/3-3/4)MII], which was shown not to occur with an additional 5 ps of 50 ms restrained MD. During the third iteration cycle the unusual 2/3 stacked extrahelical adenosine has moved into a groove-binding geometry [L(2/3-3/4)MII → L(groove-3/4)MIII] structure and by the 4th merging cycle had also converged to a 3/4-3/4 stacked structure [L(3/4-3/4)MIV].

During refinement RMS % errors in the volumes ($\%R_{NMR}$) decreased acceptably at nearly every cycle including later cycles that involved complete merging (8th, 9th, 10th, and 11th mergings) of the simulated and experimental data. The $\%R_{NMR}$ decreased from 197% to 40% for the initial 3/4 stacked model and 729% to 50% for the initial looped out model by the eighth iteration.

While the total minimized energies appear to increase as the refinement progresses, this is entirely attributed to the increased constraint energy term resulting from the increasing constraint force constants and narrower error limits imposed on the structures. In fact, unconstrained energy minimization of the 5 ps restrained MD structures shows no trend in the energy as the refinement progresses. The final minimized averaged structures are shown in Figure 1D.

As a further monitoring of the progress of our structures toward a "final" structure, we have utilized the simulation capabilities of MORASS to simulate 400 ms t_m NOESY spectra of various structures throughout the iteration process. Use of such methodology has enabled us to choose a family of possible conformations from a subset of structures which have comparable unconstrained total energies and which fulfill constraint requirements equally or nearly equally well. Spin diffusion effects, which become wide spread by 400 ms are explicitly treated in the simulated experiments.

Figure 2 shows a comparison of the experimental 400ms t_m Base–H1' region contour plot with the simulated spectrum of the final refined structure derived from the initial stacked model. Generally, the fit of crosspeak intensities between the experimental spectrum and the simulated spectra of the final structures refined from either the initial looped or stacked structures is quite good, whereas the calculated NOESY spectra of the initial models poorly agrees with the experimental spectrum.

Conclusions We have used the Hybrid Matrix/Complete Relaxation methodology in combination with restrained MD to derive a family of stuctures, for the tridecamer $d(CGCAGAATTCGCG)_2$. Use of this methodology has allowed us to accurately extract interproton distance information which would have been unobtainable using the two-spin approximation. Use of the 50 ms "2-spin" data alone does not bring the initial loop model structure to the expected 3/4-3/4 stacked structure after 50 ps of restrained MD refinement. After 60 ps of MD, no major changes in the structure occur, i.e. the molecule continues to maintain the 2/3 - 3/4 and 2/3-triplet A3' structures. Additionally, only a portion of those structures we looked at were able to reproduce the experimental NOESY data set at 400

ms t_m where we expect crosspeak intensities to be even more variable due to spin diffusion effects.

Acknowledgments Supported by the Purdue University Biochemical Magnetic Resonance Laboratory which is supported by NIH (grant RR01077 from the Biotechnology Resources Program of the Division of Research Resources) and the NSF National Biological Facilities Center on Biomolecular NMR, Structure and Design at Purdue (grants BBS 8614177 and 8714258 from the Division of Biological Instrumentation). We greatly appreciate the contributions of Dr. Claude Jones, Dr. Carol Post and Dr. Robert Santini.

REFERENCES

1. E. Nikonowicz, V. Roongta, C. R. Jones, and D. G. Gorenstein, *Biochemistry in press*, (1989).

2. E. Nikonowicz, R. Meadows, and D. G. Gorenstein, *submitted.*

3. L. Joshua-Tor, D. Rabinovich, H. Hope, F. Frolow, E. Appella, and J. L. Sussman, *Nature 334*, 82-84 (1988).

4. M. Miller, R. W. Harrison, A. Wlodawer, E. Appella, and J. L. Sussman, *Nature 334*, 85-86 (1988).

5. D. J. Patel, S. A. Kozlowski, L. A. Marky, J. A. Rice, C. Broka, K. Itakura, and K. J. Breslauer, *Biochemistry 21*, 445-451 (1982).

6. K. Wuthrich, "NMR of Proteins and Nucleic Acids", Wiley, New York, NY, 1986.

7. E. Zuiderweg, R. Scheek, R. Boelens, W. Gunsteren, and R. Kaptein, *Biochemie 67*, 707-715 (1985).

8. G. M. Clore and A. M. Gronenborn, *J. Magn. Reson. 61*, 158-164 (1985).

9. J. W. Keepers and T. L. James, *J. Magn. Res. 57*, 404-426 (1984).

10. D. G. Gorenstein, R. P. Meadows, J. T. Metz, E. Nikonowicz, and C. B. Post, "Advances in Biophysics" (C. A. Bush, Ed), JAI Press, Greenwich, Conn., in press, 1989.

11. R. Boelens, T. M. G. Koning, and R. Kaptein, *J. Mol. Struc. 173*, 299-311 (1988).

12. R. Boelens, T. M. G. Koning, G. A. van der Marel, J. H. van Boom, and R. Kaptein, *J. Magn. Reson. 82*, 290-308 (1989).

13. C. B. Post, R. Meadows, and D. G. Gorenstein, *MORASS program*, (1989).

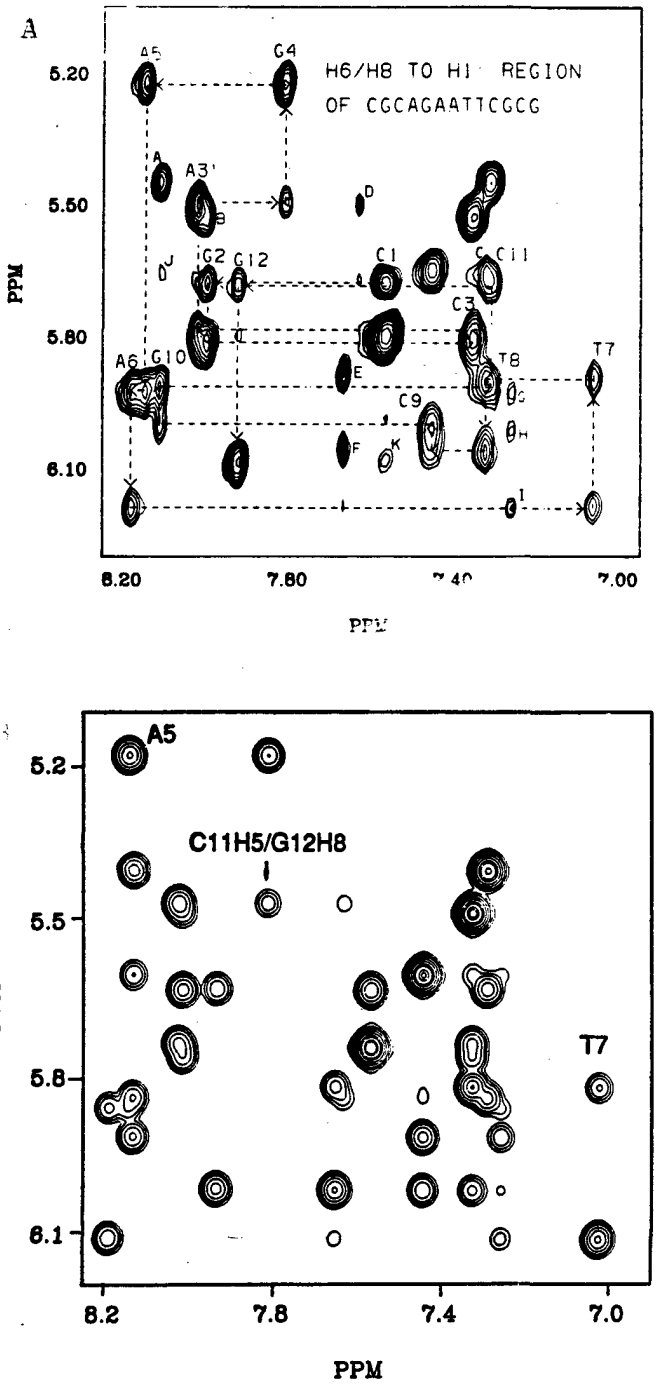

Figure 2 (A) Expansion of the pure absorption phase $^1H/^1H$ NOESY NMR spectrum contour plot of duplex extra-helical adenosine tridecamer. The sequential assignment of the base and deoxyribose H1' protons is diagrammed. (B) MORASS calculated NOESY contour plot of the base and deoxyribose H1' proton region for the final refined structure starting from the stacked model is also shown.