Linse, S., Brodin, P., Johansson, C., Thulin, E., Grundstrom, T., & Forsén, S. (1988) *Nature 335*, 651–652.
Long, J. E., Durham, B., Okamura, M., & Millet, F. (1989) *Biochemistry 28*, 6970–6974.
MacKinnon, R., Latorre, R., & Miller, C. (1989) *Biochemistry 28*, 8092–8099.
Martin, S. R., Andersson Teleman, A., Bayley, P. M., Drakenberg, T., & Forsén, S. (1985) *Eur. J. Biochem. 151*, 543–550.
Matsumura, M., Becktel, W. J., & Matthews, B. (1988) *Nature 334*, 406–410.
Perry, K. M., Onuffer, J. J., Gittelman, M. S., Barmat, L., & Matthews, C. R. (1989) *Biochemistry 28*, 7961–7968.
Rigler, R., Roslund, J., & Forsén, S. (1990) *Eur. J. Biochem.* (in press).
Sali, D., Bycroft, M., & Fersht, A. R. (1988) *Nature 335*, 740–743.
Shaw, W. V. (1987) *Biochem. J. 246*, 1–17.
Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987) *Nature 326*, 563–567.
Szebenyi, D. M. E., & Moffat, K. (1986) *J. Biol. Chem. 261*, 8761–8777.
Thomas, P. G., Russell, A. J., & Fersht, A. R. (1985) *Nature 318*, 375–376.
Tsai, M. D., Drakenberg, T., Thulin, E., & Forsén, S. (1987) *Biochemistry 26*, 3635–3643.
Vogel, H. J., Drakenberg, T., Forsén, S., O'Neil, J. D. J., & Hofmann, T. (1985) *Biochemistry 24*, 3870–3876.
Weber, P. C., Lukas, T. J., Craig, T. A., Wilson, E., King, M. M., Kwiatkowski, A. P., & Watterson, D. M. (1989) *Proteins: Struct., Funct., Genet. 6*, 70–85.
Wendt, B., Hofmann, T., Martin, S. R., Bayley, P. M., Brodin, P., Grundstrom, T., Thulin, E., Linse, S., & Forsén, S. (1988) *Eur. J. Biochem. 175*, 439–445.
White, H. D. (1988) *Biochemistry 27*, 3357–3365.

NMR Structural Refinement of an Extrahelical Adenosine Tridecamer d(CGC*A*GAATTCGCG)$_2$ via a Hybrid Relaxation Matrix Procedure†

Edward P. Nikonowicz, Robert P. Meadows, and David G. Gorenstein*
Department of Chemistry, Purdue University, West Lafayette, Indiana 47907
Received September 25, 1989; Revised Manuscript Received November 20, 1989

ABSTRACT: Until very recently interproton distances from NOESY experiments have been derived solely from the two-spin approximation method. Unfortunately, even at short mixing times, there is a significant error in many of these distances. A complete relaxation matrix approach employing a matrix eigenvalue/eigenvector solution to the Bloch equations avoids the approximation of the two-spin method. We have calculated the structure of an extrahelical adenosine tridecamer oligodeoxyribonucleotide duplex, d-(CGC*A*GAATTCGCG)$_2$, by an iterative refinement approach using a hybrid relaxation matrix method combined with restrained molecular dynamics calculations. Distances from the 2D NOESY spectra have been calculated from the relaxation rate matrix which has been evaluated from a hybrid NOESY volume matrix comprising elements from the experiment and those calculated from an initial structure. The hybrid matrix derived distances have then been used in a restrained molecular dynamics procedure to obtain a new structure that better approximates the NOESY spectra. The resulting partially refined structure is then used to calculate an improved theoretical NOESY volume matrix which is once again merged with the experimental matrix until refinement is complete. Although the crystal structure of the tridecamer clearly shows the extrahelical adenosine looped out way from the duplex, the NOESY distance restrained hybrid matrix/molecular dynamics structural refinement establishes that the extrahelical adenosine stacks into the duplex.

Two-dimensional nuclear magnetic resonance (2D NMR), combined with distance geometry (Braun & Go, 1983; Havel et al., 1983; Wüthrich, 1986) or restrained molecular mechanics/dynamics (Nilsson et al., 1986), is now capable of elucidating the fine structure of short DNA strands in solution. Typically, the evaluation of interproton distances from a 2D NMR nuclear Overhauser effect spectroscopy (NOESY) experiment depends on the so-called "two-spin approximation" (Clore & Gronenborn, 1985; Wüthrich, 1986). The approximation requires that the NOESY-derived distances be obtained from vanishingly short experimental mixing times where the buildup of NOE intensity is proportional to the inverse sixth power of the interproton distance and the effects of spin diffusion (NOE intensity mediated by multiple relaxation pathways) are minimal. Because many of the structurally important longer range NOEs are not observed at these short mixing times, the use of the two-spin approximation has raised concern over the validity of highly refined NMR structures derived by this methodology (Gorenstein et al., 1989; Keepers & James, 1984). To obtain a large number of more accurate distances, we have invoked the use of a complete relaxation

†Supported by NIH (GM36281, AI27744), 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), 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) and the National AIDS Research Center at Purdue (AI727713).

0006-2960/90/0429-4193$02.50/0

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 higher degree of precision and accuracy (Bothner-by & Noggle, 1979; Gorenstein et al., 1989; Keepers & James, 1984). A very promising approach to NMR structural refinement involves the direct calculation of the NOESY rate matrix (and hence distances) from the experimental NOESY volume matrix (Gorenstein et al., 1989; Olejniczak et al., 1986). However, the use of the relaxation matrix method is sensitive to the completeness of the experimental NOESY data (Post et al., 1990). One solution to this problem is provided by a "hybrid matrix approach".

Originally proposed by Kaptein and co-workers (Boelens et al., 1988, 1989), the hybrid matrix approach addresses the problem of incomplete experimental data by combining the information from the experimental NOESY volumes, v_{ij}^{exp}, and calculated volumes, v_{ij}^{c}. The matrix elements v_{ij}^{c} may be derived from a structure refined from a two-spin analysis of the NOESY data. This hybrid volume matrix, $\mathbf{V}^{hyb}$, is used to calculate the relaxation rate matrix. Distances derived from this hybrid relaxation rate matrix are then utilized as constraints in a restrained molecular dynamics simulation. This process is repeated until a satisfactory agreement between the calculated and observed crosspeak volumes is obtained.

As described in this paper, the hybrid matrix methodology provides a very powerful means to automate the refinement process for deriving solution structures from NMR data. This refinement procedure has allowed us to accurately extract interproton distances which would not have been obtainable with the two-spin approximation. We demonstrate the application of the hybrid matrix methodology to the solution structure of an extrahelical adenosine tridecamer oligodeoxyribonucleotide duplex, d(CGC*A*GAATTCGCG)$_2$.

Extrahelical base insertions occur naturally in the frameshift type of mutation (Fresco & Alberts, 1960). Their potential to form unusual conformations has prompted recent studies into their structure (Hare et al., 1986; Joshua-Tor et al., 1988, Kalnik et al., 1989a,b; Miller et al., 1988; Morden et al., 1983; Patel et al., 1982; Roy et al., 1986, 1987). This is of particular interest since it is the accommodation in the duplex DNA of these nucleotides which is likely to determine (at least in part) whether or not the mutation is recognized by the cell's repair mechanisms. X-ray crystallography and NMR spectroscopy are two methods currently capable of providing information on the fine structure of short (<25 bp) fragments of nucleic acids. The crystal structure of the duplex extrahelical containing tridecamer (13-mer) d(CGC*A*GAATTCGCG)$_2$ has recently been reported as well as that of a related pentadecamer d(CGCGAAATTTACGCG)$_2$ (Joshua-Tor et al., 1988; Miller et al., 1988). It was found in both cases that the extrahelical adenosine bases "looped" out way from the duplex, contrasting an earlier paper (Patel et al., 1982) in which it was concluded through one-dimensional (1D) NMR and calorimetric techniques that the adenosine of the 13-mer, d(CGC*A*GAATTCGCG)$_2$, was "stacked" into the duplex in aqueous solution. Neither the degree of "stacking" nor the structural perturbations of the residues immediately surrounding the adenosine insert have been elucidated.

When the solution NMR data of the extrahelical adenosine tridecamer oligodeoxyribonucleotide duplex d(CGC*A*GAATTCGCG)$_2$ are analyzed by the hybrid matrix methodology, we are able to show that the extrahelical adenosine stacks into the duplex, in opposition to the crystal structure results. Significantly, application of the two-spin approximation analysis to the data does not allow us to differentiate between a number of equally valid but quite different structures.

Experimental Procedures

Synthesis and Sample Preparation. The self-complementary tridecamer duplex (dCGC*A*GAATTCGCG)$_2$ was synthesized by a manual modification of the phosphite triester method on a solid support as previously described (Lai et al., 1984; Nikonowicz et al., 1989; Schroeder et al., 1987; Shah et al., 1984a,b).

NMR. The COSY and NOESY spectra were recorded on a Nicolet NT-470 (470-MHz ^{1}H) spectrometer at ambient temperature as previously described (Nikonowicz et al., 1989). NOESY buildup curves were recorded on a Varian VXR600S spectrometer (600-MHz ^{1}H) with 4096 complex points in t_2 and 512 (256 real) increments in t_1 and a recycle delay of 4.5 s. The T_1 experiments were collected on a Varian VXR600S spectrometer with 16K data points. A nonselective π inversion pulse and a recovery time of 36 s were used. A total of 16 different delay times were used, ranging from 0.125 to 32 s.

NOESY Distance Restrained Molecular Mechanics/Dynamics Calculations of the Duplex. The molecular modeling program MIDAS (Ferrin & Langridge, 1980) operating on a Silicon Graphics Iris 3030 workstation was used to generate both the base loop out structure and the base insert models for d(CGCAGAATTCGCG)$_2$. NOESY distance constraints were incorporated into the potential energy function in the molecular mechanics/dynamics program AMBER (Weiner & Kollman, 1981) through addition of a flatwell potential (Gorenstein et al., 1989; Powers et al., 1989; see Results). The different model-built structures with 228 NOESY distance constraints were then energy refined until a RMS gradient of 0.1 kcal/mol·Å was achieved or until the change in energy was less than 1.0×10^{-7} kcal/mol for successive steps.

All energy minimization and restrained molecular dynamics calculations were carried out in vacuo by using AMBER on MicroVax II and MicroVax III computers. The MD calculations were performed at 300 K, and velocities were scaled to maintain constant temperature. A Maxwellian distribution was used to calculate the initial velocities at 10 K. The time step for the integration was set to 1 fs, and coordinates were stored every 50 steps. The cutoff distance for nonbonded pairs was set to 8.5 Å, and a distance-dependent dielectric was used.

Hybrid Matrix/MORASS Refinement of Structures. A relaxation matrix program (MORASS: Multiple Overhauser Relaxation AnalysiS and Simulation) (Post et al., 1989; the program is available upon request) was used to calculate volume and rate matrices as well as to implement the hybrid matrix methodology. Typically, the well-resolved and measureable crosspeaks in the NOESY spectrum replace the corresponding crosspeaks in the calculated volume matrix, while overlapping or weak crosspeaks and diagonals are retained from the calculated spectrum. This hybrid volume matrix, $\mathbf{V}^{hyb}$, is then used to evaluate the rate matrix, whose off-diagonal elements include the effects of spin diffusion. Distances derived from this hybrid relaxation rate matrix (we assume a single isotropic correlation time of 4.5 ns) are then utilized as constraints in a 5-ps restrained molecular dynamics simulation. Energy minimization of the averaged, last 3-ps structures derived from molecular dynamics completes one cycle of refinement. This process is repeated until a satisfactory agreement between the calculated and observed crosspeak volumes is obtained. As shown by our laboratory (Gorenstein et al., 1989) and Kaptein and co-workers (Boelens et al., 1988, 1989), three to eight iterations appear to be

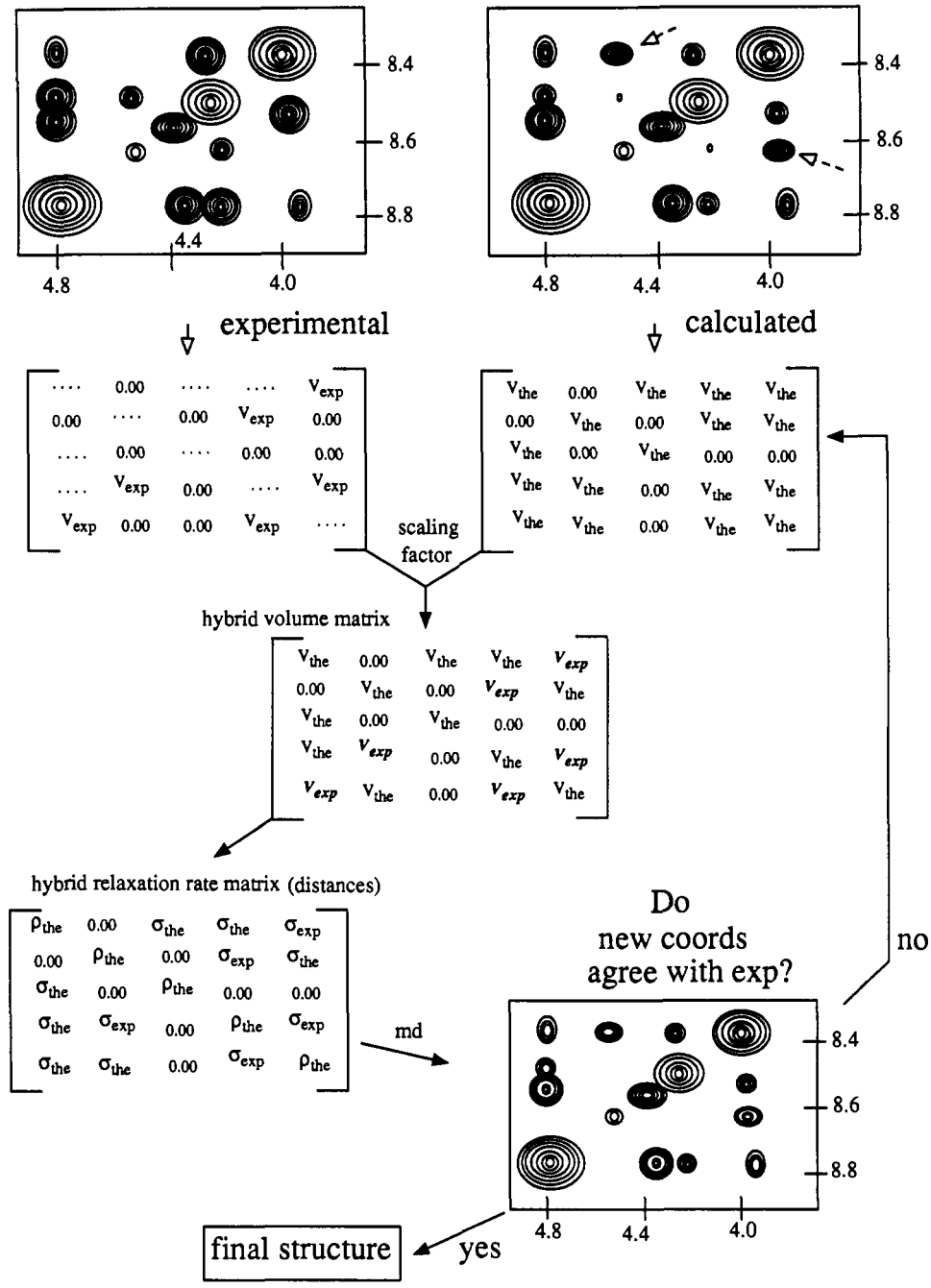


FIGURE 1: Schematic description of the hybrid relaxation matrix method. A hybrid volume matrix V^{hyb} is created by replacing the theoretical volume matrix elements V_{the} with the well-resolved experimental volume matrix elements V_{exp}. The relaxation matrix program MORASS is used to calculate the hybrid (σ/ρ) rate matrix from the hybrid volume matrix (V^{hyb}). Distances from the rate matrix are then used in a restrained molecular dynamics refinement (AMBER) to yield a new set of coordinates that are used to calculate new theoretical volume and rate matrices. Iteration continues until the experimental and theoretical NOESY volume matrices converge.

adequate to achieve convergence to a "refined" structure. This iterative scheme is represented in Figure 1.

Convergence is monitored by using eqs 1 and 2. The latter criterion is analogous to that used in X-ray crystallography.

$$\%\ \mathrm{RMS}_{vol} = \left[\frac{1}{N}\sum_{ij}\left(\frac{v_{ij}^{c} - v_{ij}^{exp}}{v_{ij}^{c}}\right)^2\right]^{1/2} \times 100 \qquad (1)$$

$$R\ \text{factor} = \frac{\sum_{ij}|v_{ij}^{exp} - v_{ij}^{c}|}{\sum_{ij}v_{ij}^{exp}} \qquad (2)$$

Convergence is achieved when the RMS_{vol} is within the reliability of the experimental volume measurement. Since most structurally important distances are those from longer range NOEs, and since these small off-diagonal volumes (<2% of the diagonal volumes) are the most sensitive to experimental noise, we feel an acceptable RMS error is 20–50% with an R factor of comparable size.

Interproton distances are calculated from the relaxation rate by rearrangement of

$$\sigma_{ij} = \frac{\gamma_H^4\hbar^2}{10\langle r_{ij}^3\rangle^2}[6J(2\omega) - J(0)] \qquad (3)$$

the ^{1}H dipolar cross-relaxation rate equation, where

$$J(\omega) = \left(\frac{\tau_c}{1 + \omega^2\tau_c^2}\right)$$

Table I: T_1 (s) of Selected Protons

residue	base ^{1}H type	T_1	H1′ proton	T_1
C1	H6	2.3		
C3	H6/H5	2.4/2.7		
A3′	H8	2.3	H1′	2.5
G4	H8	2.8	H1′	2.8
A6	H8	3.3	H1′	3.3
C9	H6	2.7	H1′	2.7
G10	H8	2.6	H1′	2.0
G12	H8	1.9	H1′	2.1
A3′	H2	6.1		
A5	H2	6.4		
A6	H2	5.7		

Results and Discussion

"Stacking" of the Extrahelical Adenosine. Evidence of the stacked nature of the extrahelical adenosine was first presented on the basis of an observed downfield shift of the H8 and H2 protons of A3′ upon melting of the duplex DNA into the single-strand form [T_m = 57.5 and 55 °C for the H8 and H2 protons, respectively (Patel et al., 1982)].[1] In addition, we find that the G4 H1′ resonance is shifted ~0.21 ppm upfield (Nikonowicz et al., 1989) relative to the G4 H1′ resonance of the parent dodecamer sequence (Hare et al., 1983). This indicates a movement of the H1′ proton into the shielding portion of the ring-current cone of either G4 or A5. "Wedging" of the extrahelical adenosine into the stacked duplex could produce such an effect, while looping out of the adenosine would be expected to produce the opposite effect.

As previously reported (Nikonowicz et al., 1989), NOESY spectra were collected at mixing times from 50 to 700 ms in D_2O. Although at a mixing time of 250 ms, sequential connectivities in the base–H1′ region were unbroken and indicated a typical B-form DNA, there were no specific crosspeaks indicative of a unique stacked conformation of the A3′ base. The ^{1}H spectrum was assigned from the 250- and 400-ms spectra. NOESY spectra acquired up to a mixing time of 150 ms lacked the C9 H1′–G10 H8 crosspeak which we believe is due to a separation of these residues to accommodate the A3′ base. A 120-ms NOESY collected in H_2O at 5 °C contained a crosspeak between A3′ H2 and G4 NH5, thus confirming the relative proximity of these two protons in space. We note that the relative positions of crosspeaks in spectra at 5 °C (H_2O) and 22 °C (D_2O) were the same. Also noteworthy was the collapse of the sequential connectivity of the base-paired imino protons at the C9–G4 to G10–C3 step, indicating a "gap" caused by the A3′ nucleotide. No attempt to quantify the H_2O spectrum was made due to inherent problems with proton exchange in water and the method of water suppression chosen (Sklenar & Bax, 1987).

We have measured the T_1 relaxation times for a few selected, well-resolved protons in the base and H1′ region (Table I). As seen in Table I, the A3′ relaxation times are not unusual. While these values may not be used to calculate an accurate correlation time, τ_c, they do suggest that the A3′ residue is not exceptional in its solvation or motional properties.

Molecular Modeling of the Tridecamer Duplex. Two alternative models for the tridecamer were considered, which we define as the stacked (S) or looped out (L) conformation. Using the molecular modeling program MIDAS (Ferrin & Langridge, 1980), we model built a standard B-DNA structure for the parent dodecamer duplex, d(CGCGAATTGCGC)$_2$. The linkage between C3 and G4 of one strand was broken and an extrahelical adenosine inserted either in a looped out orientation or with the adenosine base stacked into the double helix. To ensure starting symmetry and to remove manual inconsistencies, the extrahelical containing halves of both structures were duplicated and their halves docked together to form either the symmetric doubly stacked (3/4–3/4) or doubly looped (loop–loop) duplexes. The two different conformations were energy minimized without constraints using the molecular mechanics/dynamics program AMBER (Weiner & Kollman, 1981). Initial two-spin distance constraints were calculated after integrating the crosspeak volumes from the 50-ms NOESY spectrum using the program FTNMR (D. Hare) and by using the cytidine H5–H6 crosspeak volumes as distance rulers (2.46 Å). The NOESY-derived distances were then incorporated into a distance-constrained energy minimization (using a flatwell harmonic function modified AMBER; Gorenstein et al., 1989) of the tridecamer duplex structures. The two different initial structures are shown in Figure 2.

Two-Spin Approximation Analysis. The initial model-built structures were subjected to restrained molecular dynamics (MD) equilibration using distance constraints from the two-spin data prior to MORASS refinement; 15 ps of 300 K MD was carried out on each of the minimized model-built structures in 5-ps steps, with 206 50-ms NOESY distances (103 per strand) and 12 imino hydrogen bonds reflecting the observed imino proton signals were used as input constraints for the "flatwell" modified restrained MD runs. The harmonic force constants and error limits of the NOESY-derived distance constraints are listed in Table II. The A3′ residue on both strands of the minimized 5-ps averaged structure, SI, starting from the model-built stacked structure (Figure 2B), remained stacked between residues C3 and G4. In the corresponding looped structure LI (starting from the model-built looped structure, Figure 2A), however, one of the extrahelical adenosines has moved between base pairs G2–C11 and C3–G10 (2/3 stack), while the other extrahelical adenosine has formed a "triplex" hydrogen-bonding base triplet with C3 and G10 (Figure 2C). 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 this triplex does not grossly disobey the constraints for the 50-ms NOESY distances and is energetically allowed (Tables II and III), suggesting that the time-averaged or ensemble-averaged structure for the 13-mer might include a population of such base triplets. (Alternatively, this may simply reflect errors associated with the distances derived from the two-spin approximation.) Although tRNA and triple-strand DNA are known to contain this type of H-bonding triplet (Saenger, 1984), A–C–G triplets of the type shown in our structure are not known. It is also rather remarkable that the initial restrained MD calculations would produce an extrahelical insert between base pairs G2–C11 and C3–G10. There appears to be sufficient flexibility in the sugar phosphate backbone for the A3′ to produce this highly unusual structure. Both the triplex and the 2/3 stacked structures are energetically favorable and satisfy the 32 constraints which are either constraining A3′ or are located directly opposite A3′ (Tables II and III). Importantly, there was no improvement in the constraint errors after a total of 60 ps of two-spin restrained MD (Table III). 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.

[1] Numbering is based upon the Dickerson dodecamer, d-(CGCGAATTCGCG)$_2$, with the extrahelical adenosine between residues 3 and 4 labeled as residue 3′; see Figure 2 legend.

Table II: Calculated Energies for the Tridecamer during Restrained Molecular Dynamics Structural Refinement

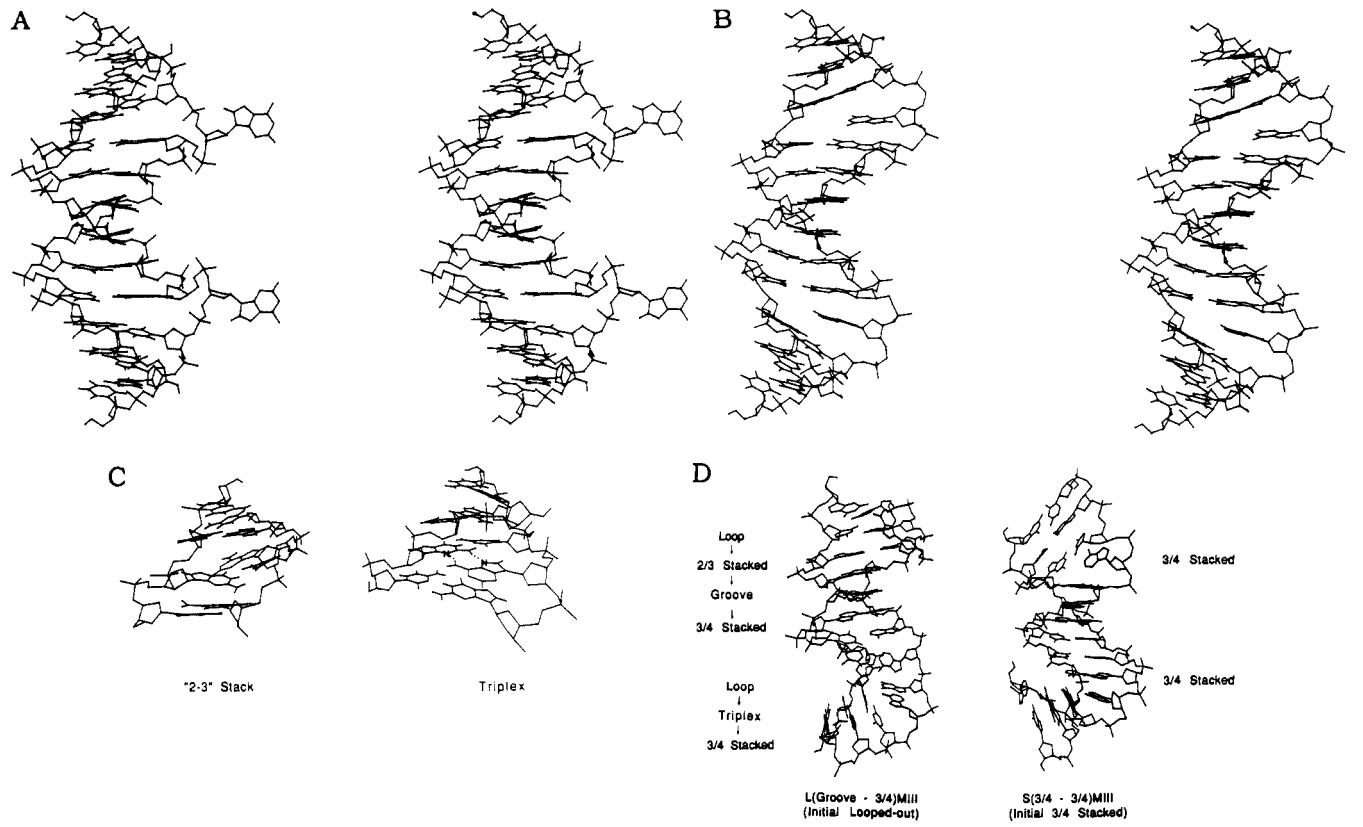
structure[a]	E_{pot}[b] (kcal/mol)	E_{pot}[c] no constr	constraint[d] (kcal/mol)	% RMS_{vol}[e]	flatwell k	parm[f] % error
(A) Starting from the Stacked (S) Model						
SI(3/4-3/4)	-650	-650		197		
SII(3/4-3/4)	-882	-950	33 (58)	167	10	15
SMI(3/4-3/4)	-856	-946	41 (105)	152	10	15
SMII(3/4-3/4)	-841	-955	43 (135)	129	12	13
SMIII(3/4-3/4)	-738	-934	100 (104)	73	12	11
SMIV(3/4-3/4)	-732	-928	106 (108)	70	13	11
SMV(3/4-3/4)	-758	-931	97 (94)	69	13	10
SMVI(3/4-3/4)	-649	-912	132 (101)	66	15	10
SMVII(3/4-3/4)	-515	-923	206 (142)	58	17	9
SMVIII(3/4-3/4)	-500	-930	205 (142)	58	19	9
SMIX(3/4-3/4)	-554	-902	115 (81)	49	30	9
SMX(3/4-3/4)	-542	-894	99 (71)	52	40	9
SMXI(3/4-3/4)	-490	-909	120 (89)	49	40	9
SMXII(3/4-3/4)	-659	-930	66 (54)		40	9
(B) Starting from Looped Out (L) Model						
LI L(loop/loop)	-702	-702		729		
LII(2/3-triplex)	-879	-952	40 (65)	358	10	15
LIII(2/3-triplex)	-901	-971	37 (61)	307	10	15
LMI(2/3-3/4)	-881	-954	30 (93)	181	10	15
LMII(2/3-3/4)	-886	-952	25 (87)	211	12	13
LMIII(groove-3/4)	-715	-938	127 (103)	70	12	11
LMIV(3/4-3/4)	-753	-943	103 (88)	65	13	11
LMV(3/4-3/4)	-806	-943	70 (66)	75	13	10
LMVI(3/4-3/4)	-684	-952	141 (100)	74	15	10
LMVII(3/4-3/4)	-590	-930	170 (104)	51	17	9
LMVIII(3/4-3/4)	-639	-922	142 (83)	52	19	9
LMIX(3/4-3/4)	-616	-941	117 (76)	50	30	9
LMX(3/4-3/4)	-660	-931	91 (54)	50	40	9
LMXI(3/4-3/4)	-628	-942	104 (59)		40	9

[a] Structure at various refinement cycles (S, initial stacked model; L, initial loop out model; M, merge matrix iterations; roman numerals represent the refinement cycle, and identifiers in parentheses represent the conformation of each half of the duplex). [b] Total energy (kcal/mol) including NOESY distance constraint term. The structures during the last 3 ps of a 5-ps restrained molecular dynamics simulation (300K) were averaged and then subjected to distance-constrained molecular mechanics (MM) energy minimization, until the RMS gradient was less than 0.1 kcal/mol·Å or until the change in energy was less than 1.0×10^{-7} kcal/mol for successive steps. The constraint energies from the latter calculation are reported in the fourth column. Sequential roman numeral labels in the structure identifier represent successive 5-ps MD/MM calculations. [c] Potential energy (kcal/mol) not including NOESY distance constraints. The structures during the last 3 ps of a 5-ps restrained molecular dynamics simulation (300 K) were averaged and then subjected to MM energy minimization without distance constraints. [d] NOESY distance constraint energies (kcal/mol). See footnote *a* for basis of calculation. The numbers in parentheses represent the distance constraint energies based upon a constant flatwell harmonic term added to the AMBER potential energy function, k = 30 kcal mol^{-1} $Å^{-2}$; % error = 9%. [e] RMS error in NOESY volumes calculated from MORASS. % RMS_{vol}; eq 1. [f] Flatwell potential function parameters (k, harmonic force constant in kcal mol^{-1} $Å^{-2}$; % error, the permitted error (±9–15%) in the constraining distance that does not result in any constraining energy penalty).

Hybrid Matrix and Restrained Molecular Dynamics Calculations. To obtain additional and more accurately integrated NOESY crosspeak volumes 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. Unfortunately, as described previously, spin diffusion effects at this point introduce considerable errors in the calculated distances from the isolated two-spin approximation methodology (Post et al., 1990; Gorenstein et al., 1989; Nikonowicz et al., 1989). Therefore, the hybrid matrix procedure was employed to correct for spin diffusion effects at this longer mixing time.

At 150 ms, 258 NOESY constraints (per duplex) were measured. Only those crosspeaks that could be adequately resolved from overlapping peaks were included. The 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 previously. A stepwise or perturbational merging was implemented to improve the diagonalization behavior of the hybrid volume matrix (Gorenstein et al., 1990; Nikonowicz et al., 1989b). The perturbational approach involves the substitution of only a fraction of the experimentally determined volumes. This "gentle nudging" of the intermediate structures avoids dramatic changes in one iteration which can produce an ill-conditioned mathematical problem during the transformation of the incorrect initial structure to the final structure. Full merging occurred by the eighth iteration.

This basic iterative scheme was followed until the % RMA_{vol} (eq 1) converged to a limiting value as outlined in Table II. After each cycle or two, the distance constraining pseudo force constants were gradually increased from 10 to 40 kcal mol^{-1} $Å^{-2}$ (however the 12 hydrogen bonds of the imino protons were kept at a constant 1.93 ± 0.30 Å with a 25.0 kcal mol^{-1} $Å^{-2}$ constraining force constant), and the estimated error brackets were gradually decreased from 15% to 9%. Table II lists energies and % RMS_{vol}, while Figure 2D shows the structures after the third, 5-ps, merge matrix iteration cycle for both the initial loop out and stacked models. We note that the 3/4–3/4 stacked structure [S(3/4–3/4)] remains stacked throughout the iterative refinement. However, in the second iteration the triplex–2/3 structure formed a 2/3–3/4 structure [LMII-(2/3–3/4)], which was shown *not* to appear with an additional 5 ps of 50-ms two-spin distance restrained MD. During the third iteration cycle the unusual 2/3 stacked extrahelical adenosine has moved into a groove-binding geometry [LMII(2/3–3/4) → LMIII(groove–3/4)] structure (shown in Figure 2D, left) and by the fourth and all subsequent merging cycles the initial loop model had also converged to a 3/4–3/4 stacked structure [the LMIX(3/4–3/4) and


A
B
C
"2-3" Stack
Triplex
D
Loop
2/3 Stacked
Groove
3/4 Stacked
Loop
Triplex
3/4 Stacked
L(Groove - 3/4)MIII
(Initial Looped-out)
3/4 Stacked
3/4 Stacked
S(3/4 - 3/4)MIII
(Initial 3/4 Stacked)
E
F
G

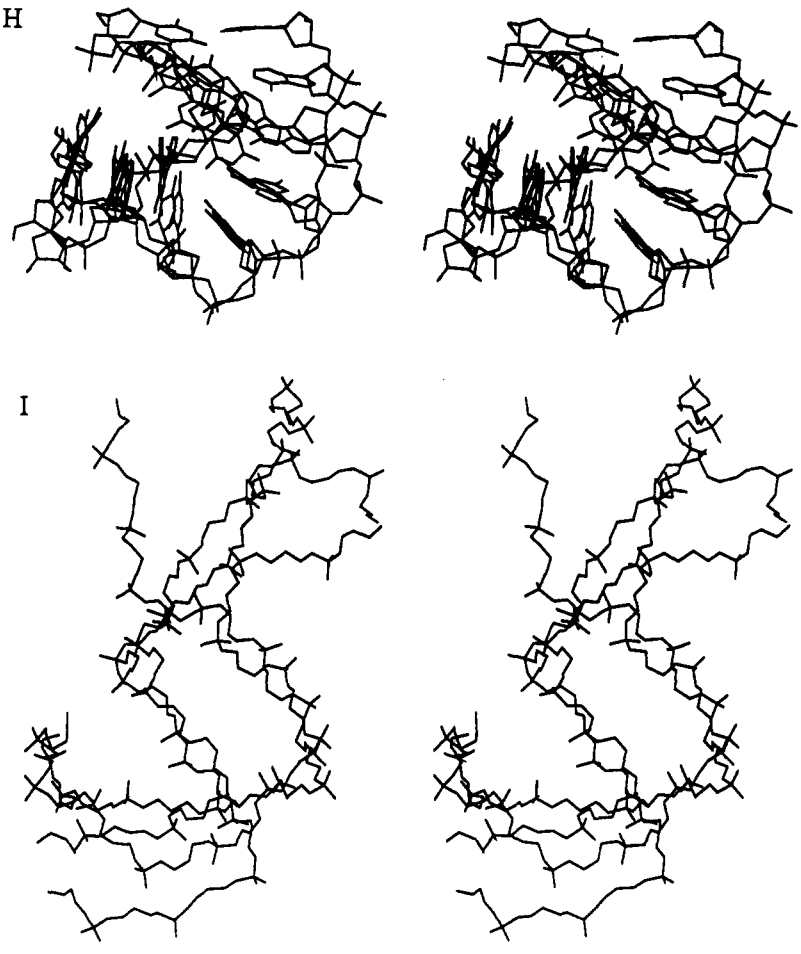


FIGURE 2: Stereoviews of NOESY-distance restrained, molecular dynamics structures during the hybrid matrix/MORASS/MD refinement of the extra-helical adenosine tridecamer d($C^1G^2C^3A^{3'}G^4A^5A^6T^7T^8C^9G^{10}C^{11}G^{12}$) (numbering of residues is shown for one strand), starting from (A) the base loop out (LI—see Table II for structure label identification) or (B) the 3/4 base stacked in, model built structures (SI—the extrahelical adenosine is shown stacked between base pairs 3 and 4). (C) Expanded nonstereoviews of both ends of the restrained MD loop out model structure showing the 2/3 stacked half and the triplet half (LII and LIII). (D) After the third iteration of the merged hybrid matrix/MORASS/MD refinement (SMIII and LMIII structures—stereoviews not shown; Table II). After the ninth iteration of the merged hybrid matrix/MORASS/MD refinement showing the LMIX refined structure (E) derived from the original looped out model and the SMIX refined structure (F) derived from the original 3/4 stacked model (Table II). (G) Stereoview overlays of full final structures LMXI and SMXII and (H) expanded stereoview of the overlain bottom halves of LMXI and SMXII. (I) Stereoview overlays of the phosphate backbones of LMXI and SMXII.

SMIX(3/4–3/4) structures are shown in Figure 2E,F].

One approach for obtaining the interproton distances from the relaxation rates is to use a ratio as shown in Boelens et al. (1988, 1989)

$$\frac{r_{ij}^6}{r_{kl}^6} = \frac{\sigma_{kl}}{\sigma_{ij}} \sim \frac{v_{kl}(t_m)}{v_{ij}(t_m)} \quad (4)$$

Our approach, however, as seen in eq 3, is to directly calculate the interproton distances from the relaxation rates. Thus, it is necessary to include a correlation time in our distance calculations. We have shown (Meadows et al., unpublished results) that with proper scaling of the experimental and theoretically calculated volumes, an incorrect estimate of τ_c has negligible effects on the derived distances. We are therefore confident that although our method is in part dependent on a reasonably correct τ_c estimate, much of the potential error possibly introduced by the choice of a poor τ_c is corrected for by scaling at each merging cycle.

Calculations using perfect volumes simulated from a single correlation time for a model structure indicate the potential to achieve as low as 15% residual error between the theoretical and "experimental" volumes by this method (Meadows et al., unpublished results). Unfortunately, it is unlikely that "real" data, especially data derived from a macromolecule of the type reported here, will be composed of a single τ_c. Thus, it is possible that a portion of the 50% residual volume errors could be due to such an effect; note, however, that the distances calculated from these volumes would have only approximately 8% error.

Comparison of Structures. During refinement, RMS percent errors in volumes (% RMS_{vol}) decreased satisfactorily at nearly every cycle including later cycles that involved complete merging (after the eighth merging) of the simulated and experimental data. As shown in Table II the % RMS_{vol} decreased from 197% to 49% for the initial 3/4 stacked model and from 729% to 50% for the initial looped out model during the final iterations. To compare the quality of fit of the theoretical volumes, we have also calculated the R factor (eq 2) directly analogous to the R value used for crystallographic refinement. In general, although there is a decrease in the R factor during the refinement cycles (decreasing from approximately 0.60–0.70 to 0.45–0.55), the changes are small. We believe our % RMS_{vol} factor better represents the quality of the NMR refinement than the R factor. The latter heavily weighs the large NOESY volumes. The absolute errors associated with the large volumes are considerably larger than those of the small volumes and dominate the summation in both the numerator and the denominator in the R factor.

Table III: RMS Deviations (Å) of the Constrained Distances for the Tridecamer during Restrained Molecular Dynamics Structural Refinement

(A) Starting from the Stacked (S) Model

structure[a]	constraints[b] full	constraints[b] intra	constraints[b] inter	full[c] $SMXII_{crd}$	full[c] $LMXI_{crd}$	CAG/GC[d] $SMXII_{crd}$	CAG/GC[d] $LMII_{crd}$
SI				0.72	0.75	0.58	0.79
SII	0.47	0.54	0.44	0.67	0.75	0.68	0.88
SIII	0.48	0.58	0.52	0.69	0.72	0.67	0.86
SIV	0.47	0.55	0.52	0.75	0.80	0.79	0.89
SMI	0.47	0.52	0.53	0.53	0.63	0.61	0.77
SMII	0.43	0.45	0.42	0.48	0.59	0.58	0.75
SMIII	0.47	0.46	0.47	0.40	0.51	0.48	0.62
SMIV	0.46	0.45	0.47	0.43	0.51	0.55	0.65
SMV	0.44	0.41	0.46	0.41	0.53	0.48	0.65
SMVI	0.42	0.40	0.44	0.40	0.51	0.50	0.66
SMVII	0.44	0.44	0.44	0.34	0.48	0.38	0.61
SMVIII	0.43	0.44	0.43	0.33	0.47	0.46	0.57
SMIX	0.35	0.34	0.35	0.27	0.48	0.35	0.60
SMX	0.33	0.33	0.34	0.29	0.50	0.42	0.60
SMXI	0.34	0.33	0.35	0.30	0.50	0.38	0.62
SMXII	0.34	0.36	0.32	0.00	0.44	0.00	0.51

(B) Starting from the Looped (L) Model

structure[e]	constraints[b] full	constraints[b] intra	constraints[b] inter	full[c] $SMXII_{crd}$	full[c] $LMXI_{crd}$	CAG/GC[d] $SMXII_{crd}$	CAG/GC[d] $LMXI_{crd}$
LI				1.02	1.00	1.53	1.56
LII	0.48	0.56	0.44	0.85	0.80	1.08	1.08
LIII	0.47	0.55	0.43	0.87	0.82	1.03	1.04
LIV	0.49	0.57	0.44	0.78	0.72	0.79	0.99
LMI	0.47	0.43	0.53	0.64	0.59	0.82	0.78
LMII	0.38	0.43	0.34	0.65	0.62	0.90	0.85
LMIII	0.49	0.48	0.50	0.50	0.40	0.62	0.51
LMIV	0.44	0.44	0.44	0.47	0.39	0.57	0.53
LMV	0.42	0.41	0.43	0.48	0.40	0.55	0.41
LMVI	0.41	0.39	0.43	0.46	0.39	0.51	0.39
LMVII	0.40	0.39	0.42	0.42	0.30	0.46	0.33
LMVIII	0.34	0.32	0.36	0.42	0.31	0.50	0.28
LMIX	0.34	0.33	0.36	0.41	0.29	0.49	0.24
LMX	0.32	0.31	0.33	0.42	0.27	0.49	0.23
LMXI	0.34	0.34	0.34	0.44	0.00	0.51	0.00

[a] See footnote *a* in Table II for definition of structures. SIV is the starting stacked structure after 60 ps of MD using two-spin constraints. Distances from the SMXII structure are used as a basis of comparison to the other structures shown in the table. [b] Constraints—the RMS deviations for the constrained proton pairs between the constraint distance input for MD/MIN and the actual proton pair distances of the minimized structure (full). These are further reduced to intraresidue and interresidue constraints. [c] Full—all 258 constraining distances are used in the RMS distance comparisons. The distances were taken from the minimized, distance-constrained molecular dynamics refined structures. Similar RMS distances are obtained if the actual constraining distances are compared. [d] CAG/GC—only those distances in the C3–A3′–G4/C9–G10 structural fragment are used in the RMS distance comparisons. [e] See footnote *a* in Table II for definition of structures. LIV is the starting stacked structure after 60 ps of MD using two-spin constraints. Distances from the LMXI structure are used as a basis of comparison to the other structures shown in the table.

Table IV: Various Energy Terms for the Final Structures

structure	energy (kcal/mol) E_{pot}	bond	constr	angle	dihed	H bond	electro	VdW
$SMXII_{constr}$	−690	22	66	154	319	−7	−1045	−201
$SMXII_{unconstr}$	−929	11		101	288	−10	−1078	−245
$LMXI_{constr}$	−659	22	104	170	323	−6	−1052	−218
$LMXI_{unconstr}$	−941	11		113	286	−8	−1091	−256

[a] Note that $E_{pot} = E_{bond} + E_{constr} + E_{angle} + E_{dihed} + E_{H\text{-}bond} + E_{VdW} + E_{electro}$. constr, constrained structure values. unconstr, unconstrained structure values.

These large volumes of course correspond to short distances, and many of these short distances in DNA represent sugar ^1H–^1H intra-furanose distances which are not very sensitive to the overall geometry of the duplex. The longer distances define the interresidue constraints and are therefore much more important in refining features of the structure such as the orientation and position of the extrahelical base. Our % RMS_{vol} factor *equally* weighs the small and large volumes by summing the *percentage* errors in these volumes (equally weighing a 20% error in a large and small volume while that of eq 2 would place a much heavier weight on the error in the large volume). Thus, we believe the % RMS_{vol} factor defined in eq 1 is a more useful criterion for convergence and quality of fit for NMR data.

While the minimized potential energies shown in Table II 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 the 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 (Table II). Note as shown in Table II that the LIII(2/3-triplex) structure, rather surprisingly, has the lowest unconstrained energy (−971 kcal/mol) of any of the structures [because separate energy minimization was performed with and without distance constraints, the E_{pot} values shown in Table II are not equal to the sum of the E_{pot} (no constraint) plus constraint energy]. Table IV summarizes the covalent and

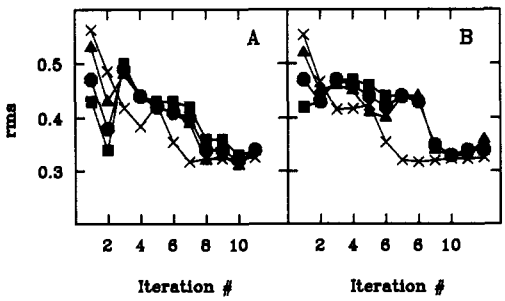


FIGURE 3: Diagramatic representation of the RMS deviation at each iteration between the input flatwell distance constraint for molecular dynamics and the actual distance for the position-averaged minimized output structure. The (×) curves represent the largest theoretical rms deviation allowed by the error bars for the flatwell used at each iteration (see also Table II). (A) Starting stacked structure series; (B) starting looped structure series; (○) RMS for all (intra- and interresidue) constraints; (□) RMS for intraresidue constraints; (△) RMS for interresidue constraints.

van der Waal energy terms of the final stacked and looped structures with and without constraints. Note that for comparison purposes the E_{constr} term has been removed from the E_{bond} such that E_{bond} now represents the summation of only covalent terms. The final refined structures, $SMXII_{unconstr}$ and $LMXI_{unconstr}$, have comparable energy term contributions, although the latter appears to be an energetically more favorable structure.

The major purpose of employing the MD algorithm was to "drive" the refinement into obeying the new set of distance constraints generated by the merge–MORASS procedure during each iteration, but to do so in a minimal amount of computer time. We found a cycle time of 5 ps was adequate to equilibrate the structure to a constant constraint energy. Using a constant flat well constraint penalty function (% error = 12%; harmonic force constant, k = 15 kcal mol^{-1} Å^{-2}), the constraint energy decreases from ca. 100–135 to 50–80 kcal/mol during the merge matrix refinement steps (Table II). However, because the equilibrium constraint distances are also changing with each iteration, comparison of constraint energies and distance violations may be misleading [note especially that the two-spin distances *appear* to provide low constraint energies, and yet these distances are quite different from the final hybrid matrix distances (Table III)].

Perhaps a better measure of the progress of the refinement methodology is achieved by monitoring the behavior in the NOESY distance constraints at each step and by comparison of the final distances with all other structures along the refinement pathway (Table III).

The degree to which the constraints are satisfied at each stage of the refinement procedure is followed in two ways. First, the 3-ps averaged coordinate set from each MD refinement cycle was energy minimized with a standard flat well (k = 30 kcal mol^{-1} Å^{-2}; % error = 9%). The constraint energies of these minimized structures are shown in Table II. Second, the deviations in distances between the structures derived from the MD calculations and the input set of constraint distances for that MD run may be calculated (Table III). Shown in Figure 3 are the RMS deviations for the total set of distance constraints as well as the subset of inter- and intraresidue distance constraints. Also shown (×) is the greatest possible RMS deviation of the maximal errors allowed from the input distances calculated by using the flat well % error for each iteration.

We observe no significant differences in the behavior of the inter and intraresidue constraints as the refinement proceeds. We also note that after 60 ps of two-spin restrained MD (SIV and LIV), the RMS deviation between the distances in the SIV and LIV structures and the constraint distances are the same as after 10 ps.

Table V: Atomic Cartesian Coordinate RMS Deviations (Å)

(A) Between Selected Full Structures (All Atoms)

structure[a]	SI	LI	SII	LIII	SMVII	LMVII	SMXII	LMXI
SI		2.8	3.2	3.8	6.3	5.3	5.4	5.4
LI			3.8	4.3	6.3	5.9	5.5	5.9
SII				3.7	5.0	5.4	4.9	4.8
LIII					5.9	4.2	4.5	5.0
SMVII						5.7	4.1	4.5
LMVII							3.8	2.9
SMXII								4.3

(B) About the A3′ Region (CAG/GC) for Selected Structures

structure[a]	SI	LI	SII	LIII	SMVII	LMVII	SMXII	LMXI
SI		3.9	3.1	3.5	3.4	3.5	3.0	3.4
LI			4.9	4.9	4.7	5.3	4.6	5.1
SII				3.3	3.2	2.7	2.9	2.8
LIII					3.6	3.0	3.5	3.4
SMVII						4.0	1.7	3.3
LMVII							3.3	1.8
SMXII								2.7

[a] See Table II, footnote *a* for a definition of the structure nomenclature.

The RMS differences between the distances derived from the final structures [LMXI(3/4–3/4) and SMXII(3/4–3/4)] are compared for the full tridecamer duplex (full) and those constraining distances involving only the residues immediately adjacent to the extrahelical adenosine C3–A3′–G4/C9–G10 (CAG/GC) for both halves of the duplex (Table III). The final 0.44-Å RMS difference between the full set of constrained distances and the 0.51-Å RMS difference between the extrahelical insert region set of constrained distances in the LMXI and SMXI structures is comparable to the RMS deviation of distances between any two sets of refined structures (for example, the RMS deviation of distances between SMXII and SMIX, SMX, or SMXI structures is 0.3–0.4 Å and the RMS deviation of distances between LMXI and LMVIII, LMIX, or LMX is 0.2–0.4 Å). Each of these latter refinement cycle structures is thus equally valid, and refinement can only be defined in terms of a family of comparable structures. RMS Cartesian coordinate comparisons between the structures show similar trends (Table V). However, the RMS differences are considerably larger (>5 times the values shown in Table III) because throughout these MD refinement cycles considerable gross changes in the duplex are observed. Although the atomic Cartesian RMS deviation between successive structures is low (<1.0 Å), the deviation between final structures and final and initial structures is relatively high (Table V). As can be seen the extrahelical insert causes the duplex to bend at the site of the insert (Figure 2), and the degree of bending is quite variable in each iteration cycle. Thus, while the short-range distances are well-defined, it is very difficult, if not impossible, by NMR to define long-range structural differences when only short-range distances are used as constraints.

Parts G–J of Figure 2 depict overlays of the final structures SMXII and LMXI. As can clearly be seen, the fit of the bottom halves is rather good but that of the upper half is quite poor. The extrahelical adenosines cause the DNA to "kink", and the direction of this "kinking" is random other than generally bending away from the direction of insertion of the extrahelical adenosine. Although local differences in the structures may be small, their cumulative effect over the length of the helix can be quite large, as is indicated by the overlay of the phosphate backbone for the final structures (Figure 2J).

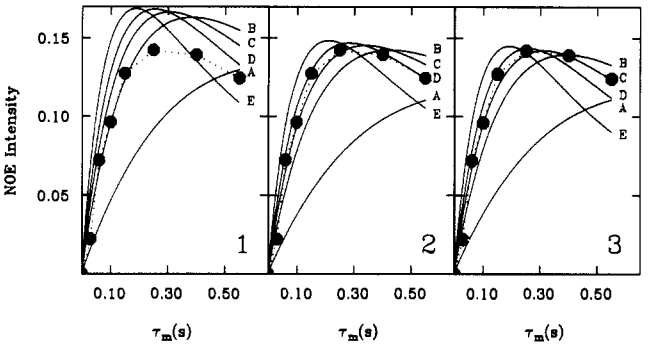


FIGURE 4: Experimental (●) and theoretical NOE buildup curves of the C11 H5/H6 proton pair for the initial stacked model (SI) (panel 1), the final stacked structure (SMXII) (panel 2), and the final looped structure (LMXI) (panel 3) as a function of mixing time. Initial volumes for each curve were scaled to the diagonal volume at mixing time zero for a single proton. The average of three well-resolved single diagonal peak volumes was used for this purpose. τ_m = mixing time; (●) experimental NOE intensity; (A) τ_c = 1.5 ns; (B) τ_c = 3.5 ns; (C) τ_c = 4.5 ns; (D) τ_c = 5.5 ns; (E) τ_c = 7.5 ns.

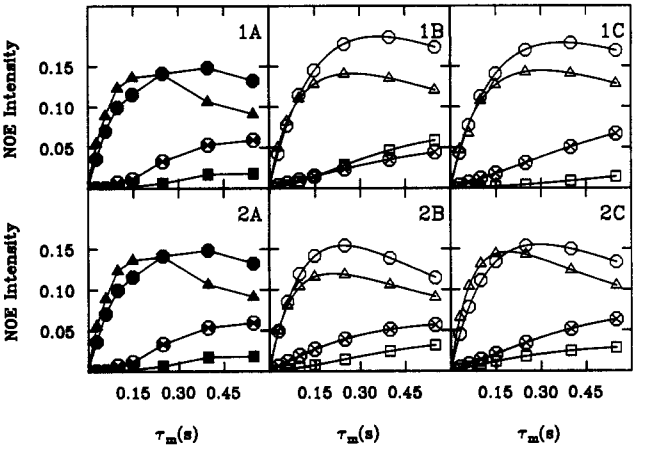


FIGURE 5: Experimental (shaded symbols; 600-MHz ^{1}H) and theoretical (open symbols) NOE buildup curves for four selected spin pairs. Panels 1A and 2A present the experimentally determined NOE intensities as a function of mixing time τ_m. Panels 1B and 2B contain the starting stacked structure SI and final refined structure SMXII, respectively. Panels 1C and 2C contain the starting looped structure LI and final refined structure LXI, respectively. (○) C3 H5–C3 H6; (△) C9 H1′–C9 H2″; (⊗) A3′ H1′–A3′ H8; (□) A3′ H1′–G4 H8.

In addition, while interresidue proton distances have been locally well-defined along the entire helix, the phosphate backbone conformation of the 13-mer has not been directly constrained by any NOESY-derived distances or coupling constant derived torsional angles [see, for discussion, Gorenstein et al. (1990)].

NOE time development curves have long been a standard for structural high-resolution NMR work. However, their main purpose has been to define the linear regime of the NOE buildup to extract "two-spin" distances. Although it was not necessary to generate a buildup curve to carry out the MORASS refinement, we have found it a useful tool for measuring the validity of our calculated structures. It should be restated that the accuracy of the theoretical curves is based upon two independent unknown parameters, i.e., τ_c and the set of distances representing the correct structure. Figure 4 depicts the NOE intensity of a fixed distance proton pair (C11 H5/H6) as a function of τ_c for three different structures: (1) unconstrained, minimized stacked structure, (2) final stacked structure, and (3) final looped structure. It can be seen that the experimental data would fit a τ_c of approximately 2 ns in the initial stacked structure. However, the NOE buildup of the H5/H6 (experimental) matches that calculated at 5 and 4.5 ns τ_c in SMXII and LMXI. We emphasize that although the NOE buildup is structure dependent as well as τ_c dependent, the matching of the τ_c used for the MORASS calculations may reflect the accuracy of our original choice of a 4.5-ns τ_c. If, however, we were to have used a correlation time suggested by the initial structure, a value 2 ns shorter would have been chosen. This points, perhaps, to a better method to optimize τ_c; that is to say, choose an initial τ_c from models and after initial MORASS refinement, reevaluate the choice of τ_c.

Figure 5 shows the NOE intensities as a function of mixing time for four proton pairs, two of which involve the extrahelical adenosine residue. The fit of theoretical and experimental buildup curves is generally quite good. However, it appears that the longer range intra- and interresidue ^{1}H–^{1}H distances better satisfy the experimentally observed intensities as well as the overall shapes of the curves. Again, these intensities are dependent not only on τ_c but also on structure. We believe that the proton pairs leading to the shorter distance crosspeaks (mostly the intra-furanose ring distances) may have internal ring motions that produce a local τ_c that is slightly different from that of the whole molecule (Van De Ven & Hilbers, 1988). However, as previously stated, this error has been greatly reduced by properly scaling the data.

As a final measure of the progress of our refinement scheme, we have simulated (using MORASS) the 400-ms mixing time NOESY spectra of various structures at various iteration cycles. Spin diffusion effects, which become quite appreciable by 400 ms, are explicitly treated in the simulations. Figure 6 shows a comparison of the experimental 400-ms t_m base–H1′ region contour plot with those simulated from initial and final structures. Generally, the fit of crosspeak intensities between the experimental spectrum and the simulated spectra of final structures refined from either initial looped or stacked models is quite good. In contrast, the simulated spectra of the initial stacked and looped models show large crosspeak intensity errors (compare Figure 6A–C). Selected critical crosspeaks have been identified in the various spectra. However, in all spectra the A6 and A7 H2 crosspeaks are not well reproduced in the simulated spectra. These differences can be explained, at least in part, by incomplete recovery of Z magnetization in the experimental NOESY. Because the adenosine H2 protons generally have a significantly longer T_1 relaxation time than the remainder of the molecule (Table I), experimental crosspeak volumes to these protons will be disproportionately small, but the theoretically calculated volumes shall have undergone essentially "complete" T_1 recovery, thus "appearing" to be too large.

CONCLUSIONS

We have used a hybrid relaxation matrix methodology in combination with restrained molecular dynamics to derive a family of structures for the extrahelical adenosine tridecamer d(CGC*A*GAATTCGCG)$_2$. The hybrid matrix/restrained molecular dynamics refinement procedure has allowed us to accurately extract interproton distance information that would not have been obtainable by using the two-spin approximation. Thus, it is not possible to differentiate between the structures in which the extrahelical adenosine stacks between the second and third base pairs if the distances derived from the two-spin approximation are used in a restrained MD refinement. Much more accurate and precise distances are obtainable from the hybrid relaxation matrix methodology, and starting from two quite different initial structures, it is possible to iteratively refine both to a common family of structures consistent with

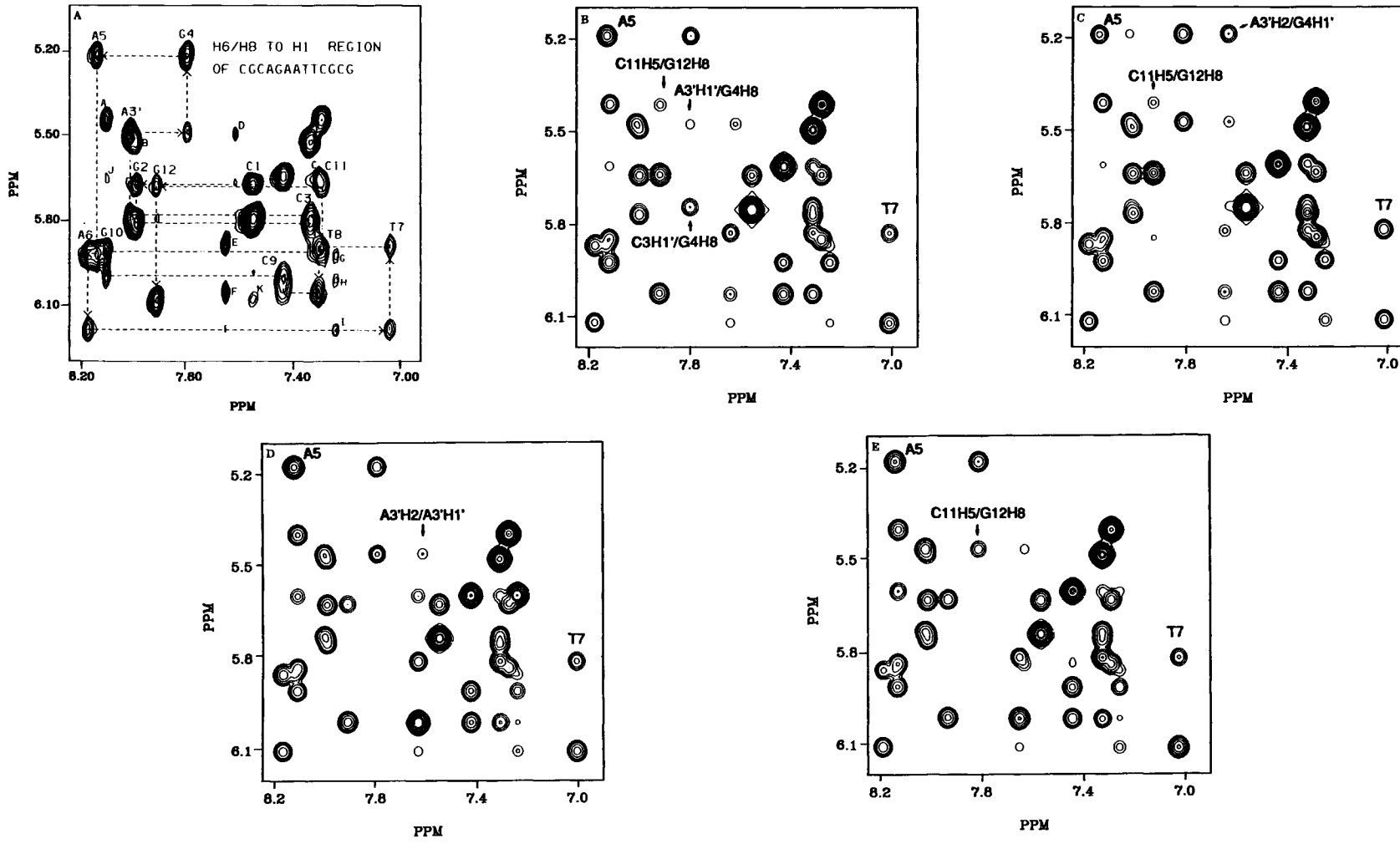


FIGURE 6: (A) Expansion of the pure absorption phase $^1H/^1H$ NOESY NMR spectrum contour plot of duplex extrahelical adenosine tridecamer. The sample contained approximately 460 O.D. units of the tridecamer in 400 μL of D_2O (99.995%) buffer containing 50 mM potassium phosphate, 100 mM KCl, and 0.1 mM NaN_3, and 100 mM phosphate buffer at pH* 7.1 (uncorrected pH meter reading). The NOESY spectra were recorded on a Nicolet NT-470 (470-MHz ^1H) spectrometer at ambient temperature. The phase-sensitive TPPI (Marion & Wüthrich, 1983) NOESY spectrum was acquired with 2048 points in the t_2 dimension and 256 t_1 increments. Mixing times of 50, 100, 150, 200, 250, 300, 400, 500, and 700 ms were used. The 400-ms mixing time NOESY data (shown) were treated with a small negative exponential and a small Gaussian multiplication in the t_1 and t_2 dimensions. The sequential assignment of the base and deoxyribose H1′ protons is diagrammed. Additional labeled crosspeaks correspond to C11 H5–G10 H8 (A), G2 H8–C3 H5 (B), T8 H6–C9 H5 (C) A3′ H2–A3′ H1′ (D), A6 H2′–T7 H1′ (E), interstrand A6 H2′–T8 H1′ (F), A5 H2–A5 H1′ (G), interstrand A5 H2–C9 H1′ (H), A5 H2–A6 H1′ (I), G10 H8–C9 H5 (J), and C1 H6–G12 H1′ (K). The large unlabeled crosspeaks correspond to the four CH5–CH6. (B–E) Contour plots of the base to H1′ region calculated NOESY spectra at various refinement cycles of the hybrid matrix/MORASS/MD procedure starting from the initial base loop out (B) and stacked models (C) of the extrahelical adenosine tridecamer (see text). Contour plots of the final structures LMXI (D) and SMXII (E) are also shown. Selective crosspeaks are labeled. Note in particular that in the initial looped model (B) the intensity of the A3′ H1′–G4 H8 crosspeak is weak compared to the experimental value (shown in plot A). In the initial stacked model (plot C) several additional peaks are calculated to have significant intensity (G12 H8–C11 H5 and A3′ H2–G4 H1′), whereas they are absent from the experimental spectrum. In contrast, both contour plots (D and E) of the final structures derived from the initial loop or stack models, respectively, agree quite well with each other and the experimental spectrum.

the NOESY-derived distances. The new methodology has also provided a tool to *automate the refinement process* for deriving solution structures from NMR data. Finally, the solution NMR data show that the extrahelical adenosine stacks into the duplex, while in the solid state the crystal structure of the tridecamer clearly shows the extrahelical adenosine looped out way from the duplex. This difference highlights the importance of utilizing both X-ray crystal and solution NMR methods to determine structure.

ACKNOWLEDGMENTS

We greatly appreciate the contributions of Dr. Claude Jones and Dr. Robert Santini.

REFERENCES

Boelens, R., Koning, T. M. G., & Kaptein, R. (1988) *J. Mol. Struct. 173*, 299–311.

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

Bothner-by, A. A., & Noggel, J. H. (1979) *J. Am. Chem. Soc. 101*, 5152–5155.

Braun, W., & Go, N. (1983) *J. Mol. Biol. 186*, 613–621.

Clore, G. M., & Gronenborn, A. M. (1985) *J. Magn. Reson. 61*, 158–164.

Ferrin, T. E., & Langridge (1980) *Comput. Graphics 13*, 320.

Fink, T., & Crothers, D. M. (1972) *J. Mol. Biol. 66*, 1–12.

Fresco, J. R., & Alberts, B. M. (1960) *Proc. Natl. Acad. Sci. U.S.A. 46*, 311–321.

Gorenstein, D. G., Meadows, R. P., Metz, J. T., Nikonowicz, E., & Post, C. B. (1990) *Advances in Biophysical Chemistry* (Bush, C. A., Ed.) JAI Press, Greenwich, CT (in press).

Hare, D., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7456–7464.

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. (1983) *J. Mol. Biol. 171*, 319.

Havel, T. F., Kuntz, I. D., & Crippen, G. M. (1983) *Bull. Math. Biol. 45*, 665–720.

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

Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., & Patel, D. J. (1989a) *Biochemistry 28*, 294–303.

Kalnik, M. W., Norman, D. G., Swann, P. F., & Patel, D. J. (1989b) *J. Biol. Chem. 264*, 3702–3712.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404–426.

Lai, K., Shah, D. O., Derose, E., & Gorenstein, D. G. (1984) *Biochem. Biophys. Res. Commun. 121*, 1021.

Lee, C. H., & Tinoco, I., Jr. (1978) *Nature 274*, 609–610.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biochem. Biophys. Res. Commun. 113*, 967–974.

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

Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I., Jr. (1983) *Biochemistry 22*, 5557–5563.

Nikonowicz, E., Roongta, V., Jones, C. R., & Gorenstein, D. G. (1989a) *Biochemistry 28*, 8714–8725.

Nikonowitz, E., Meadows, R., & Gorenstein, D. G. (1989b) *Bull. Magn. Reson., 11*, 226–229.

Nilsson, L., Clore, G. M., Gronenborn, A. M., Brunger, A. T., & Karplus M. (1986) *J. Mol. Biol. 188*, 455–475.

Olejniczak, E. T., Gampe, R. T., & Fesik, S. W. (1986) *J. Magn. Reson. 67*, 28.

Patel, D. J., Kozlowdki, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 445–451.

Post, C. B., Meadows, R. P., & Gorenstein, D. G. (1989) MORASS program.

Post, C. B., Meadows, R. P., & Gorenstein, D. G. (1990) *J. Am. Chem. Soc.* (submitted for publication).

Powers, R., Olsen, R., & Gorenstein, D. G. (1989) *J. Biomol. Struct. Dyn.* (in press).

Roy, S., Weinstein, S., Borah, B., Nickol, J. Apella, E., Sussman, J. L., Miller, M., & Shindo, H. (1986) *Biochemistry 25*, 7417–7423.

Roy, S., Sklenar, V., Appella, E., & Cohen, J. S. (1987) *Biopolymers 26*, 2041–2052.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Schroeder, S., Fu, J., Jones, C., & Gorenstein, D. G. (1987) *Biochemistry 26*, 3812–3821.

Shah, D. O., Lai, K., & Gorenstein, D. G. (1984a) *Biochemistry 23*, 6717–6723.

Shah, D. O., Lai, K., & Gorenstein, D. G. (1984b) *J. Am. Chem. Soc. 106*, 4302.

Sklenar, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469–479.

Van De Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem 178*, 1–38.

Weiner, P. K., & Kollman, P. A. (1981) *J. Comput. Chem. 2*, 287–303.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.