Debus, R. J., Barry, B. A., Sithole, I., Babcock, G. T., & McIntosh, L. (1988b) *Biochemistry 27*, 9071–9074.
East, J. M., & Lee, A. G. (1982) *Biochemistry 21*, 4144–4145.
Eftink, M. R., & Ghiron, C. A. (1976) *J. Phys. Chem. 80*, 486–493.
Ferreira-Rajabi, L., & Hill, B. C. (1989) *Biochemistry 28*, 8028–8033.
Fersht, A. (1985) *Enzyme Structure and Mechanism*, 2nd ed., W. H. Freeman and Co., New York.
Frerman, F. E. (1987) *Biochim. Biophys. Acta 893*, 161–169.
Frerman, F. E. (1988) *Biochem. Soc. Trans. 16*, 416–418.
Frerman, F. E. (1990) in *Fatty Acid Oxidation: Clinical, Biochemical and Molecular Aspects* (Tanaka, K., & Coates, P. M., Eds.) pp 69–77, Alan R. Liss Inc., New York.
Gorelick, R. J., Schopfer, L. M., Ballou, D. P., Massey, V., & Thorpe, C. (1985) *Biochemistry 24*, 6830–6839.
Hearst, J. E., & Sauer, K. (1984) *Z. Naturforsch. 39C*, 421–424.
Hill, B. C., Horowitz, P. M., & Robinson, N. C. (1986) *Biochemistry 25*, 2287–2292.
Husain, M., & Steenkamp, D. J. (1983) *Biochem. J. 209*, 541–545.
Johnson, M. A., Morningstar, J. E., Oliver, M., & Frerman, F. E. (1987) *FEBS Lett. 226*, 129–133.
Lakowicz, J. R. (1983) *Principles of Fluorescence Spectroscopy*, Plenum, New York.
Leher, S. S. (1971) *Biochemistry 10*, 3254–3263.
Ljungdahl, P. O., Pennoyer, J. D., Robertson, D. E., & Trumpower, B. L. (1987) *Biochim. Biophys. Acta 891*, 227–241.
Loehr, J. P., Goodman, S. I., & Frerman, F. E. (1990) *Pediatr. Res. 27*, 311–315.
Marquardt, D. W. (1963) *J. Soc. Ind. Appl. Math. 11*, 431–441.
McKean, M. C., Frerman, F. E., & Mielke, D. M. (1979) *J. Biol. Chem. 254*, 2730–2735.
McKean, M. C., Beckmann, J. D., & Frerman, F. E. (1983) *J. Biol. Chem. 258*, 1866–1870.
Miller, G. L. (1959) *Anal. Chem. 31*, 964.
Ohinishi, T. (1987) *Curr. Top. Bioenerg. 15*, 37–65.
Parker, C. A. (1968) *Photoluminescence of Solutions*, Elsevier, Amsterdam.
Prince, R. C., & George, G. N. (1990) *Trends Biochem. Sci. 15*, 170–172.
Ragan, C. I. (1987) *Curr. Top. Bioenerg. 15*, 1–36.
Ramsay, R. R., Steenkamp, D. J., & Husain, M. (1987) *Biochem. J. 241*, 883–892.
Ruzicka, F. J., & Beinert, H. (1977) *J. Biol. Chem. 252*, 8440–8445.
Samworth, C. M., Degli Esposti, M., & Lenaz, G. (1988) *Eur. J. Biochem. 171*, 81–86.
Sivaraja, M., Goodin, D. B., Smith, M., & Hoffman, B. M. (1989) *Science 245*, 738–740.
Stryer, L. (1978) *Annu. Rev. Biochem. 47*, 819–846.
Stubbe, J. (1990) *J. Biol. Chem. 265*, 5329–5332.
Usui, S., Yu, L., & Yu, C.-A. (1990) *Biochemistry 29*, 4618–4626.
Yu, C.-A., & Yu, L. (1982) *Biochemistry 21*, 4096–4101.

# NMR Structural Refinement of a Tandem G·A Mismatched Decamer d(CCAA*GA*TTGG)$_2$ via the Hybrid Matrix Procedure†

Edward P. Nikonowicz, Robert P. Meadows, Patricia Fagan, and David G. Gorenstein*
*Department of Chemistry, Purdue University, West Lafayette, Indiana 47907*
*Received May 16, 1990; Revised Manuscript Received September 28, 1990*

ABSTRACT: A complete relaxation matrix approach employing a matrix eigenvalue/eigenvector solution to the Bloch equations is used to evaluate the NMR solution structure of a tandemly positioned G·A double mismatch decamer oligodeoxyribonucleotide duplex, d(CCAA*GA*TTGG)$_2$. An iterative refinement method using a hybrid relaxation matrix combined with restrained molecular dynamics calculations is shown to provide structures having good agreement with the experimentally derived structures. Distances incorporated into the MD simulations have been calculated from the relaxation rate matrix evaluated from a hybrid NOESY volume matrix whose elements are obtained from the merging of experimental and calculated NOESY intensities. Starting from both A- and B-DNA and mismatch syn and anti models, it is possible to calculate structures that are in good atomic RMS agreement with each other (<1.6 Å RMS) but differ from the reported crystal structure (>3.6 Å). Importantly, the hybrid matrix derived structures are in excellent agreement with the experimental solution conformation as determined by comparison of the 200-ms simulated and experimental NOESY spectra, while the crystallographic data provide spectra that are grossly different.

Mismatched bases in DNA are naturally occurring, and while they may lead to mutations during replication and transcription, they are also known to be repaired. The efficiency of repair of mismatched bases appears to be dependent upon both the nature of the mismatch and the sequences flanking the mismatch. The G·A-type mismatch has been shown to reflect these observations (Kramer et al., 1984; Fazakerley, 1986) and thus provides an excellent model to decipher the fine structural characteristics that impart its unique recognizable properties.

The G·A mismatches have been shown to adopt three grossly

†Supported by NIH (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 Biological Facilities Center on Biomolecular NMR, Structure and Design at Purdue (Grants BBS 8614177 and BBS 8714258 from the Division of Biological Instrumentation), and the National AIDS Research Center at Purdue (AI727713).

0006-2960/91/0430-1323$02.50/0 

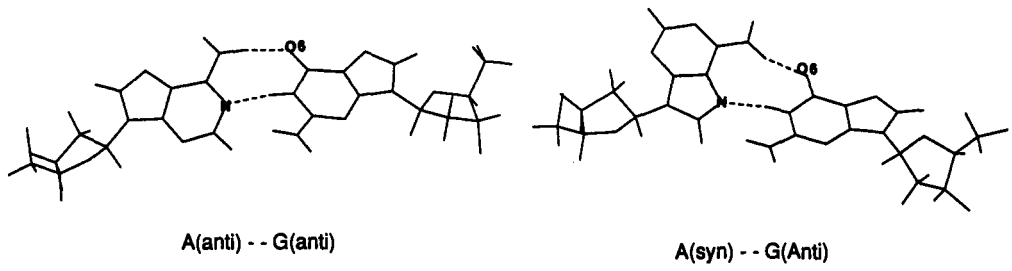


FIGURE 1: Nucleoside G·A base pair in either (left) $G_{anti}/A_{anti}$ conformation or (right) $G_{anti}/A_{syn}$ conformation for the decamer duplex, d($C^1C^2A^3A^4G^5A^6T^7T^8G^9G^{10}$).

different conformations, $G_{anti}$–$A_{anti}$, $G_{anti}$–$A_{syn}$, and $G_{syn}$–$A_{anti}$, in both the crystal and solution states. It has been demonstrated that the guanine base at the site of the G·A mismatch in the dodecamer d(CGGGAATTCACG)$_2$ is oriented in a syn conformation under acidic conditions in solution (Gao & Patel, 1988). The high-resolution three-dimensional crystal structures of G·A mismatched duplexes have been reported for a dodecamer, d(CGC*G*AATT*A*GCG)$_2$ (Brown et al., 1986), and the tandem mismatch d(CCAA*GA*TTGG)$_2$ (Prive et al., 1987). The dodecamer was found to exist in the $G_{anti}$–$A_{syn}$ conformation while the latter, henceforth referred to as the decamer, adopted the $G_{anti}$–$A_{anti}$ conformation (Figure 1).

Although the crystal structure agrees qualitatively with the NMR solution conformation reported for the duplex decamer, d(CCAA*GA*TTGG)$_2$ (Kan et al., 1983; Nikonowicz & Gorenstein, 1990), details of the solution conformation have not been addressed. Because the $G_{anti}$–$A_{anti}$ and $G_{anti}$–$A_{syn}$ structures (syn and anti) are energetically quite similar, the possibility of the existence of both conformations in solution cannot be excluded. Thus, detailed structural analysis by 2D NMR data has been undertaken without predisposition of the model structures toward the anti conformation.

Principally, solution structures have been obtained by evaluation of interproton distances from 2D NMR nuclear Overhauser effect spectroscopy (NOESY) (Clore & Gronenborn, 1989). Most often this depends upon the so-called "two-spin approximation" (Clore & Gronenborn, 1985; Wuthrich, 1986; Nilsson et al., 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 (Keepers & James, 1984; Nikonowicz et al., 1989a,b, 1990; Gorenstein et al., 1990).

In order to obtain a large number of more accurate distances, we have invoked 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 higher degree of precision and accuracy (Bothner-by & Noggle, 1979; Nikonowicz et al., 1989a, 1990; Gorenstein et al., 1990; Keepers & James, 1984). 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".

The hybrid matrix approach (Boelens et al., 1988, 1989; Gorenstein et al., 1990; Nikonowicz et al., 1989a, 1990) 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 $v_{ij}$ matrix elements may be derived from a structure refined from a two-spin analysis of the NOESY data. This hybrid volume matrix, $V^{hyb}$, 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 cross-peak volumes is obtained. We demonstrate the application of the hybrid matrix methodology to the solution structure of the tandem G·A mismatched decamer oligodeoxyribonucleotide duplex, d(CCAA*GA*TTGG)$_2$. When the solution NMR data of the decamer are analyzed by the hybrid matrix methodology, we observe structures similar in appearance to that seen in the crystalline state. However, application of the two-spin approximation analysis, using 80-ms NOESY data, produces structures that are quite different (as measured by the RMS deviation of all atom Cartesian coordinates) from the hybrid matrix solution structure as well as the crystallographically determined structure. Further, criteria established to measure the validity of refined structures indicate that some structural differences do exist between the decamer in the crystalline state and in the solution state.

## Experimental Procedures

*Synthesis and Sample Preparation.* The self-complementary decamer duplex (dCCAA*GA*TTGG)$_2$ was synthesized by a manual modification of the phosphite triester method on a solid support as previously described (Lai et al., 1984; Nikonowicz & Gorenstein, 1990; Schroeder et al., 1987; Shah et al., 1984a,b).

*NMR.* The TOCSY and NOESY spectra were recorded on a Varian VXR500 (500-MHz $^1$H) spectrometer at 22 °C as previously described (Nikonowicz & Gorenstein, 1990). Typically 48 transients were acquired for each FID. NOESY buildup curves were recorded on a Varian VXR600S (600-MHz $^1$H) spectrometer with 4096 complex points in $t_2$, 512 (256 real) increments in $t_1$, and a recycle delay of 5.5 s. The $T_1$ experiments were also collected on a Varian VXR600S spectrometer with 16K data points. A nonselective $\pi$ inversion pulse with a recovery time of 40 s was 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.* Model A- and B-DNA duplexes of the sequence d(CCAAGATTGG)$_2$ were constructed by using the molecular mechanics/dynamics program AMBER3 (Weiner & Kollman, 1981), and the duplexes were then modified to accommodate the A6 adenosine base in either

an anti or syn form by using the molecular modeling program MIDAS (Ferrin & Langridge, 1980), operating on a Silicon Graphics Iris 3030 workstation. NOESY distance constraints were incorporated into the potential energy function of the AMBER3 program through addition of a flatwell potential (Gorenstein et al., 1990). The different model-built structures were then energy refined with either 162 80-ms NOESY constraints or 260 200-ms NOESY constraints until a RMS gradient of 0.1 kcal $mol^{-1}$ $Å^{-1}$ was achieved or until the change in energy was less than $1.0 \times 10^{-7}$ kcal $mol^{-1}$ for successive steps.

All energy minimization and restrained molecular dynamics calculations were carried out in vacuo by using AMBER3 on Vax 3200 Workstation computers. The molecular dynamics (MD) calculations were performed in 6-ps blocks by stepping the temperature from 400 K during the first 2 ps down to 300 K for the final 2 ps and scaling the velocities to maintain constant temperature. During each 6-ps MD block, which was determined to be the minimum time necessary for each iterative equilibration cycle, the force constants and percent errors of the constraining distances were reset after 3 ps. A Maxwellian distribution was used to calculate the initial velocities at 10 K, and the random-number seed was incremented by 7 for each iteration. The time step for the integration was set to 1 fs, and coordinates were stored every 50 steps. The residue-based cutoff distance for nonbonded pairs was set to 8.5 Å, and a distance-dependent dielectric was used to simulate the effect of solvent.

*Hybrid Matrix/*MORASS *Refinement of Structures.* The relaxation matrix program MORASS (*m*ultiple *O*verhauser *r*elaxation *a*nalysi*s* and *s*imulation) (Post et al., 1989) was used to calculate volume and rate matrices as well as implement the hybrid matrix methodology (Figure 2).

Typically, the well-resolved and measurable cross-peaks in the NOESY spectrum replace the corresponding cross-peaks in the calculated volume matrix, while overlapping or weak cross-peaks 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 3.6 ns) are then utilized as constraints in a 6-ps restrained molecular dynamics simulation. Energy minimization of the averaged last 2 ps of structures, derived from molecular dynamics, completes one cycle of refinement. This process is repeated until a satisfactory agreement between the calculated and observed cross-peak volumes is obtained. As shown by our laboratory (Nikonowicz et al., 1989a, 1990; Gorenstein et al., 1990) and others (Boelens et al., 1988, 1989), three to five iterations appear to be adequate to achieve convergence to a "refined" structure.

Convergence is monitored by using the equation

$$\%\ \mathrm{RMS}_{vol} = \sqrt{(1/N)\sum_{ij}[(v_{ij}^{a} - v_{ij}^{b})/v_{ij}^{a}]^2} \times 100 \quad (1)$$

where $\%\ \mathrm{RMS}_{vol} = \%\ \mathrm{RMS}_{the}$ when a = the calculated volume and b = the experimental volume and $\%\ \mathrm{RMS}_{vol} = \%\ \mathrm{RMS}_{exp}$ when a = the experimental volume and b = the calculated volume.

Convergence is achieved when the $\%\ \mathrm{RMS}_{vol}$ 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%–45%.

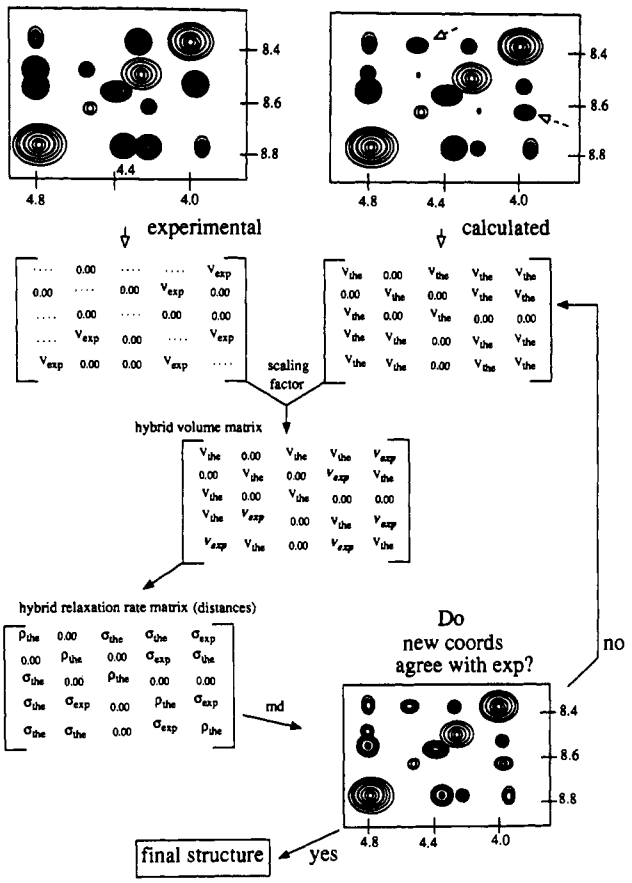


FIGURE 2: Schematic description of the hybrid relaxation matrix method. A hybrid volume matrix, $\mathbf{V}^{hyb}$, is created by replacing the theoretical volume matrix elements $V_{the}$ with the well-resolved experimental volume matrix elements $V_{exp}$. The relaxation matrix program MORASS is used to calculate the hybrid ($\sigma/\rho$) rate matrix from the hybrid volume matrix ($\mathbf{V}_{hyb}$). 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.

Interproton distances are calculated from the $^1H$ dipolar cross-relaxation rate

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

where

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

## Results and Discussion

*$G_{anti}$·$A_{anti}$ Nature of the Mismatch.* Evidence for the general conformation of this G·A mismatch in solution was first presented by Kan et al. (1983). The appearance of an NOE peak between A6 H2 and the G5 H1 imino proton indicated that the mismatched G and A residues were oriented in an anti conformation. Qualitative inspection of the NOESY spectra also indicates that both the G5 and A6 residues are oriented in the anti–anti conformation (Nikonowicz & Gorenstein, 1990). Also notable in these spectra were the highly upfield shifted G5 H1′ resonance and the rather broad resonances A6 H8 and A6 H1′.

As previously reported (Nikonowicz & Gorenstein, 1990), NOESY spectra were collected at mixing times of 60–800 ms in $D_2O$. Although, at a mixing time of 200 ms, sequential

connectivities in the base–H1′ region were unbroken and indicated a typical B-form DNA, the 400- and 800-ms mixing time NOESY spectra were used to assign the decamer. A 120-ms NOESY spectrum recorded in $H_2O$ at 5 °C confirmed the previously assigned exchangeable imino protons (Kan et al., 1983) and the hydrogen-bonded state of the duplex. Additionally, the relative intensities of cross-peaks involving the A6 H8 proton and the H6/H8 protons of the remaining residues in the base–H1′ and base–H2′,H2″ regions were found to be quite comparable to those measured in $D_2O$ at 20 °C. Due to problems associated with proton exchange, the $H_2O$ spectrum has been used only for qualitative purposes at this time. Additionally, the method of water suppression (Sklenar & Bax, 1987) complicates the extraction of accurate volumes from the NOESY spectrum.

The $T_1$ relaxation times for resolved H1′, H5, H6, and H8 proton resonances vary between 1.8 and 2.6 s, whereas those of the adenosine H2 protons are 5.2–5.3 s. The relaxation times of the A6 residue protons are not exceptional, suggesting that there is nothing unusual in terms of their solvation or motional properties.

*Molecular Modeling of the Decamer Duplex.* Four alternative models for the decamer were considered, which we will define as the B-$DNA_{anti}$ (BA), B-$DNA_{syn}$ (BS), A-$DNA_{anti}$ (AA), and A-$DNA_{syn}$ (AS) conformations. We have model-built Arnott A-DNA and B-DNA decamer duplexes with the NUCGEN module of AMBER. To create the syn conformers of the duplexes, the A6 base residues on each strand were rotated 180° about the N9–C1′ bond with the molecular modeling program MIDAS (Ferrin & Langridge, 1980). Unfavorable van der Waals interactions were relieved for all four model-built structures by energy minimization without constraints until RMS gradient of 0.1 kcal $mol^{-1}$ $Å^{-1}$ was achieved.

Initial two-spin distance constraints were calculated by integrating the cross-peak volumes from the 80-ms NOESY spectrum and by using the cytidine H5-H6 cross-peak 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., 1990) of the decamer duplex structures.

*Two-Spin Approximation Analysis.* A total of 162 80-ms NOESY distances (81 per strand) were used as input constraints for the flatwell-modified restrained MD runs. Six picoseconds of 400–300 K MD was carried out on each of the minimized model-built structures, followed by averaging and minimization of the coordinates generated during the final 2 ps. This procedure was repeated for a total of 36 ps. The three terminal base pairs were each constrained with an imino hydrogen bond, even though the $H_2O$ spectrum clearly indicates that all imino protons were in base-paired hydrogen bonds. This was done in order to ensure that the central G·A base pair would have adequate freedom of movement/rotation throughout the dynamics procedure and thus would not be biased toward a particular structure. The harmonic force constants and error limits of the NOESY constraints are listed in Table I. The A6 residue remained in the anti conformation after the first 6 ps of "two-spin" constrained MD in the AA and BA starting structures. However, there were sufficient constraints in the G·A mismatch region to "pull" the A6 residue of the AS and BS starting structures from a syn to the anti orientation. Interestingly, when molecular dynamics is performed on the four starting structures, with a total of six imino proton hydrogen-bond constraints (the central ApGpApTp residues are left untethered) and no interproton distance restraints, the A6 residues of structures AA and BA remain anti and the A6 residues of the AS and BS structures remain syn (Figure 1).

*Hybrid Matrix and Restrained Molecular Dynamics Calculations.* In order to obtain additional and more accurately integrated NOESY cross-peaks for further refinement of the structures, we have used the 200-ms NOESY spectrum where the cross-peaks have greater signal-to-noise and hence lower integration errors. Unfortunately, as described previously, spin diffusion effects at this point introduce considerable errors in the distances calculated from the isolated two-spin approximation methodology (Post et al., 1990; Gorenstein et al., 1990; Nikonowicz et al., 1990).

Therefore, the complete hybrid relaxation matrix procedure was employed to correct for multispin effects at this longer mixing time. At 200 ms, 260 NOESY constraints (per duplex) were measured. Only those cross-peaks that could be adequately resolved from overlapping peaks were included. An additional eight imino hydrogen-bond constraints (the central GA base pair imino protons were not included) were added, and a total of 268 distance restraints were incorporated into the next stage of the refinement. The initial structures for MORASS refinement were the four model geometries that were subjected to 6 ps of restrained molecular dynamics (MD) equilibration in which distance constraints from the two-spin data were used. The typical refinement follows the iterative merged matrix/restrained molecular dynamics methodology incorporating the NOESY distance constraints as described previously (Boelens et al., 1988, 1989; Nikonowicz et al., 1989a, 1990; Gorenstein et al., 1990). A stepwise or perturbational merging was implemented to improve the diagonalization behavior of the hybrid volume matrix (Nikonowicz et al., 1989a, 1990; Gorenstein et al., 1990). 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 that can produce an ill-conditioned mathematical problem during the transformation of the incorrect initial structure to the final structure. Full merging occurred by the third iteration.

This basic iterative scheme was followed until the % $RMS_{vol}$ (eq 1) converged to a limiting value as outlined in Table I. After each cycle, the distance-constraining pseudo force constants were gradually increased from 10 to 20 kcal $mol^{-1}$ $Å^{-2}$ (however, the eight 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 6%. Table I lists energies and % $RMS_{vol}$ for each of the structures considered.

One approach for obtaining the interproton distances, $r_{ij}$, from the cross-relaxation rates, $\sigma_{ij}$, is to use a ratio as shown in eq 3 (Boelens et al., 1988, 1989).

$$\frac{r_{ij}^6}{r_{kl}^6} = \frac{\sigma_{kl}}{\sigma_{ij}} \tag{3}$$

Our approach, however, as seen in eq 2, 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 based our correlation time on the Stokes–Einstein relation as well as estimates from the $^1H$ $T_1$ relaxation times. We have shown (Meadows et al., in preparation) that, with proper scaling of the experimental to the theoretically calculated volumes, an incorrect estimate of $\tau_c$ has negligible effects on the derived distances, within limits (up to mixing times between 250 and 300 ms for values of $\tau_c$

Table I: Comparison of MORASS/Restrained MD Refinement from Four Different Starting Structures

| structure[a] | $E_{tot}$[b] (kcal/mol) | $E_{tot}$[c] (no constr) | constr energy[d] | % $RMS_{vol}$[e] (% $R_{the}$/% $R_{exp}$) | flatwell $k$[f] | param[f] (% err) |
|---|---|---|---|---|---|---|
| **Restrained MD on 10-mer (Arnott A-DNA)** | | | | | | |
| **Residue A6 Starting in Anti Conformation** | | | | | | |
| AAMODEL | −723 | −723 | | 160/147 | | |
| AA2S | −615 | −728 | 67 | 256/66 | 10/14 | 15/12 |
| AAMI | −496 | −705 | 115 | 78/65 | 12/16 | 14/10 |
| AAMII | −538 | −717 | 90 | 71/75 | 12/16 | 13/10 |
| AAMIII | −473 | −711 | 123 | 61/71 | 14/18 | 12/8 |
| AAMIV | −529 | −729 | 102 | 55/66 | 16/20 | 10/7 |
| AAMV | −521 | −734 | 91 | 59/70 | 18/20 | 8/6 |
| AAFINAL | −490 | −736 | 48 | 42/45 | 20/40 | 6 |
| **Residue A6 Starting in Syn Conformation** | | | | | | |
| ASMODEL | −583 | −583 | | 4719/204 | | |
| AS2S | −569 | −635 | 63 | 434/65 | 10/14 | 15/12 |
| ASMI | −521 | −625 | 100 | 82/77 | 12/16 | 14/10 |
| ASMII | −547 | −660 | 101 | 71/74 | 12/16 | 13/10 |
| ASMIII | −527 | −688 | 105 | 64/78 | 14/18 | 12/8 |
| ASMIV | −546 | −676 | 73 | 58/74 | 16/20 | 10/7 |
| ASMV | −530 | −700 | 80 | 63/69 | 18/20 | 8/6 |
| ASFINAL | −487 | −723 | 45 | 47/61 | 20/40 | 6 |
| **Restrained MD on 10-mer (Arnott B-DNA)** | | | | | | |
| structure[a] | $E_{tot}$[b] (kcal/mol) | $E_{tot}$[c] (no constr) | constr energy[d] | % $RMS_{vol}$[e] (% $R_{the}$/% $R_{exp}$) | flatwell $k$[f] | param[f] (% err) |
| **Residue A6 Starting in Anti Conformation** | | | | | | |
| BAMODEL | −759 | −759 | | 173/90 | | |
| BA2S | −629 | −736 | 41 | 339/67 | 10/14 | 15/12 |
| BAMI | −602 | −731 | 57 | 190/74 | 12/16 | 14/10 |
| BAMII | −507 | −706 | 99 | 74/72 | 12/16 | 14/11 |
| BAMIII | −483 | −730 | 105 | 54/111 | 14/18 | 13/9 |
| BAMIV | −546 | −735 | 72 | 54/73 | 16/20 | 11/7 |
| BAMV | −525 | −720 | 76 | 53/79 | 18/20 | 8/6 |
| BAFINAL | −470 | −740 | 53 | 41/51 | 20/40 | 6 |
| **Residue A6 Starting in Syn Conformation** | | | | | | |
| BSMODEL | −728 | −728 | | 15100/132 | | |
| BS2S | −065 | −724 | 45 | 432/64 | 10/14 | 15/12 |
| BSMI | −583 | −692 | 57 | 365/64 | 12/16 | 14/10 |
| BSMII | −497 | −690 | 92 | 54/61 | 12/16 | 14/11 |
| BSMIII | −541 | −710 | 71 | 48/72 | 14/18 | 13/9 |
| BSMIV | −502 | −695 | 82 | 44/72 | 16/20 | 11/7 |
| BSMV | −520 | −705 | 49 | 48/72 | 18/20 | 8/6 |
| BSFINAL | −518 | −732 | 43 | 47/58 | 20/40 | 6 |
| **Two-Spin and Imino Hydrogen-Bonded Final Structures from Each of the Four Model Structures** | | | | | | |
| structure[g,h] | $E_{tot}$[b] (kcal/mol) | $E_{tot}$[c] (no constr) | constr energy[d] | % $RMS_{vol}$[e] (% $R_{the}$/% $R_{exp}$) | flatwell $k$[f] | param[f] (% err) |
| AA2SFNL | −664 | −702 | 25 | 231/133 | 10 | 15 |
| AS2SFNL | −677 | −706 | 24 | 322/98 | 10 | 15 |
| BA2SFNL | −685 | −750 | 11 | 223/67 | 10 | 15 |
| BS2SFNL | −629 | −744 | 57 | 230/110 | 10/14 | 15/12 |
| AAHBND | −770 | −770 | 0.0 | 1108/114 | 18 | 25 |
| ASHBND | −747 | −747 | 0.0 | 4455–143 | 18 | 25 |
| BAHBND | −774 | −774 | 0.0 | 200/94 | 18 | 25 |
| BSHBND | −752 | −752 | 0.0 | 12200/137 | 18 | 25 |

[a] Structure at various refinement cycles: AA, initial A-DNA model with anti A6 residues; AS, initial A-DNA model with syn A6 residues; BA, initial B-DNA model with anti A6 residues; BS, initial B-DNA model with syn A6 residues; 2S, two-spin constrained structures; MODEL, model-built structures; FNL and FINAL, final structures; HBND, only six imino proton hydrogen-bond constraints, the central four residues remaining unconstrained; M, merge matrix iterations; roman numerals represent the refinement cycle. [b] Total energy (kcal/mol) including the NOESY distance constraint term. The structures during the last 2 ps of a 6-ps restrained molecular dynamics simulation (300 K) were averaged and then subjected to distance-constrained molecular mechanics (MM) energy minimization, until the RMS gradient was less than 0.1 kcal $mol^{-1}$ $Å^{-1}$ or until the change in energy was less than $1.0 \times 10^{-7}$ kcal $mol^{-1}$ 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 6-ps MD/MM calculations. [c] Potential energy (kcal $mol^{-1}$) not including NOESY distance constraints. The structures during the last 2 ps of a 6-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 $mol^{-1}$). See footnote *b* for basis of calculation. [e] RMS error in NOESY volumes calculated from MORASS. % $RMS_{vol}$ from eq 1; both % $RMS_{the}$ and % $RMS_{exp}$ are reported. [f] Flatwell potential function parameters [$k$, harmonic force constant in kcal $mol^{-1}$ $Å^{-2}$; % err, the permitted error (±6%–15%) in the constraining distance that does not result in any constraining energy penalty]. Force constants ($k$) and permitted error (% err) are reported as those used for 1–3 ps/4–6 ps of the 6-ps MD simulation. Hydrogen-bond constraints were kept fixed at 1.93 ± 0.3 Å and a force constant of 25 kcal $mol^{-1}$ $Å^{-2}$. [g] 2S structures generated after 36 ps of "two-spin"-derived 80-ms NOESY distances. MD was carried out analogous to that of the 200-ms MORASS structures (i.e., 6-ps MD → average last 2 ps of structures → restrained minimization → 6-ps MD). [h] HBND structures generated by the same methodology as in footnote *g*; however, only six imino proton distance restraints were used: $d(C^1pC^2pA^3pApGpApTpT^4pG^5pG^6)_2$.

from ca. 3.5 to 6.0 ns for the decamer). We are therefore confident that although our method is in part dependent on a reasonably correct $\tau_c$ estimate, much of the potential error possibly introduced by the choice of a poor $\tau_c$ is corrected by scaling at each merging cycle. The Kaptein laboratory approach (eq 3) and our own (eq 2) differ only in the point at

which the scaling is performed.

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 (Baleja et al., 1990; Meadows et al., in preparation). However, it is unlikely that "real" data, especially data derived from a macromolecule of the type reported here, may be analyzed with a single isotropic rotational model. Thus, it is possible that a portion of the 45% residual volume errors could be due to such an effect; note however that the distances calculated from these volumes would have less than 8% error.

*Comparison of Structures.* Table I shows the % RMS errors in the volumes for the four different starting structures at each stage of the refinement process. The unusually large % RMS for the starting syn structures is the result of the A6 residues on each strand being in a syn orientation and clearly indicates that a syn or syn-like geometry of the mismatched A6 base is an unreasonable orientation.

All final structures result in approximately 45% and 55% RMS deviations between theoretical and experimental volumes (% $RMS_{the}$ and % $RMS_{exp}$, respectively). We note that the errors between the theoretical volumes calculated from the CRYSTAL structure coordinates and the experimental volumes are % $RMS_{the}$ = 191 and % $RMS_{exp}$ = 75 (not shown). Also presented in Table I are the final two-spin constrained structures (2SFNL) derived from applying the 80-ms NOESY constraints to each starting model for 36 ps. For comparison, four "final" structures calculated with only six imino proton hydrogen-bond constraints, the central four residues remaining unconstrained, are presented after 36 ps of MD (HBND). The unconstrained molecular dynamics simulations of these latter two sets of structures were carried out by following the identical 6-ps refinement cycle used for the hybrid matrix iteration procedures (6-ps MD, followed by averaging of the last 2 ps, followed by energy minimization). While there is some small improvement in the % RMS volume deviations between the first 6-ps cycle and the sixth (36-ps) cycle for the 80-ms NOESY-constrained structures, the errors, when compared to the 200-ms NOESY-refined structures, are quite large. Also, it should be noted that unconstrained molecular dynamics calculations alone, using the six imino protons as hydrogen-bond constraints, have produced structures whose theoretical volumes do not approximate those measured experimentally, even though the total energies of these structures are rather low.

The total potential energies with and without constraints as well as the constraint energies of the different structures at each iteration of the refinement procedure are also reported in Table I. With the exception of the syn A-DNA starting structure, there does not appear to be a significant decrease in the total energy of the unconstrained minimized structures. However, the potential and constraint energies of the constrained minimized structures increase only moderately as the force constants increase from 10 to 40 kcal $mol^{-1}$, and the permitted error of the constraint distances is decreased from 15% to 6%. Thus, although the structures become more tightly constrained as the iterations proceed, the degree to which the distance constraints are obeyed generally improves (as reflected by the energy terms).

While the total and constraint energies may be useful for observing gross trends and for serving as a general monitor of the refinement process, they may not necessarily reflect the "correctness" of the structures. Although the final structures produced with only imino proton constraints throughout the MD procedure have the lowest potential energies of any of the other structures, the theoretical NOESY spectra calculated from these structures differ the most from the experimentally observed NOESY spectrum. In fact, the syn model-built decamers maintain their syn A6 orientation when only the six imino hydrogen-bond constraints are used over the entire 36 ps of dynamics. This reflects the local energy minimum problem in unconstrained molecular mechanics and dynamics. Only by incorporating the NOESY distance constraints in the dynamics calculations can we overcome the relatively high energy barrier separating the syn and anti models.

By incorporation of NOESY distance constraints in the syn model, the MD calculations are able to readily locate the more favorable anti conformation. As shown in Figure 3, the transition from the syn to the anti conformation is accomplished within the first several picoseconds of restrained MD refinement. At about 0.25 ps (Figure 3A) the base-stacking separation between the G5 residue of the G·A base mismatch and its neighboring stacked mismatched G·A base pair has begun to widen from 3.4 Å to ca. 4.4 Å. This provides enough room for the adenosine base to begin to rotate from the syn conformation, and it is thus not required to swing out from the interior of the duplex in order to "flip". By 1.6 ps (Figure 3B) the A6 C2–N7 vector is nearly parallel with the dyad axis, as full rotation begins; at 4.0 ps (Figure 3C) the base has completed its rotation and is in the anti conformation. The phosphodiester backbone considerably extends during this rotation. Rotation of the other tandem syn A base on the other strand occurs during the same time period, although it is not concerted.

We have chosen to evaluate the extent to which the constraints are obeyed by the structures at each stage of the refinement by comparing the RMS deviation between the input constraint distances used for the MD runs and the distances found in the resulting structures. The greatest possible RMS deviations at each stage would be expected if all of the constrained distances were maximally perturbed to the point at which the distance violation would contribute an energy-constraint penalty ($r_{NOE}$ − % err > $r_{ij}$ > $r_{NOE}$ + % err). The constraints are well satisfied at the beginning and during the intermediate stages of the refinement; the actual deviation of the distances from those allowed by the contraining function range from about 0.42 to 0.22 Å. At the final stage of the refinement, with a narrowed % err of ±6%, the constrained distances are just accommodated within the flatwell limits of the narrowed penalty function, thus further increasing the total and constraining energies.

Because the iterative merged matrix approach maintains an independent refinement pathway for each different starting model, the constraints determined via the MORASS program at each step will be unique to the input structure. Consequently, we can define the "final" refined structure in terms of a series of equally valid sets of families of structures and distance constraints. Table II compares the RMS difference between the final sets of flatwell distance constraints derived from the refinements of the four different starting structures as well as the distances taken from the crystal structure. The degree to which the constraints agree with the distances found in the various structures is presented as an RMS deviation. In general, it can be seen that the four solution-state-derived distance constraints are more closely related to the four MORASS-refined structures than to the crystal structure. In fact, the deviations among the final constraints and the four final structures vary around 0.34 Å, while the final set of constraints differs from the crystal structure distances by about 0.74 Å.

Table II: RMS Deviations (Å) between the Atomic Cartesian Coordinate and the Final Complete Set of MORASS-Derived Constraints

| | structure[a] | | | | |
|---|---|---|---|---|---|
| flatwell[b] | AAFINAL | ASFINAL | BAFINAL | BSFINAL | CRYSTAL |
| AAFINAL | 0.26 | 0.37 | 0.33 | 0.43 | 0.69 |
| ASFINAL | 0.33 | 0.27 | 0.36 | 0.38 | 0.72 |
| BAFINAL | 0.39 | 0.45 | 0.27 | 0.45 | 0.77 |
| BSFINAL | 0.40 | 0.39 | 0.38 | 0.24 | 0.76 |

[a] Cartesian coordinates derived from the FINAL 40-ps restrained MD structures. [b] Full set of constraining distances generated from the fifth iteration cycle of MORASS for each of the four structures.

Table III: Breakdown of Intra- and Interresidue Distance (Å) and Volume (% $RMS_{vol}$) Deviations for All Constrained Proton Pairs

| | | constraint[a] | | | | % $RMS_{vol}$[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $^1H-^1H$ type | total[c] | AA | BA | AS | BS | AA | BA | AS | BS | CRYSTAL[g] |
| $intra_{d(ribose)}$[d] | 90 | 0.17 | 0.19 | 0.20 | 0.17 | 39/45 | 47/52 | 43/65 | 41/46 | 56/50 |
| $intra_{base-d(ribose)}$[e] | 83 | 0.25 | 0.26 | 0.25 | 0.26 | 39/42 | 34/43 | 51/63 | 56/42 | 54/60 |
| $inter_{base-d(ribose)}$[f] | 87 | 0.34 | 0.35 | 0.34 | 0.28 | 52/43 | 44/49 | 62/53 | 51/66 | 413/108 |

[a] RMS deviation between the restrained distances of the final structures (coordinate sets: AA = AAFINAL; BA = BAFINAL; AS = ASFINAL; BS = BSFINAL) and the final set of flatwell MD constraints derived from the fifth iteration of the hybrid matrix procedure by using 200-ms NOESY data. [b] RMS deviation between the volumes theoretically calculated from the final structures and the experimentally measured volumes; columns are shown as % $RMS_{the}$/% $RMS_{exp}$. [c] Total number of interproton distance constraints used in each category (footnotes *d–f*). [d] Full set of intraresidue deoxyribose ring interproton constraints. [e] Set of intraresidue base–deoxyribose ring interproton constraints including the fixed cytidine H5/H6 pair. [f] All interresidue proton constraints including base–base, base–deoxyribose ring, and four (two per strand) interstrand adenosine H2 restraints. [g] Crystal coordinates (Prive et al., 1987) obtained from the Brookhaven data bank.

Table IV: Atomic Cartesian Coordinate RMS Deviations (Å) between Selected Full Structures (All Atoms)

| structure[a] | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | | 3.5 | 2.9 | 2.2 | 3.0 | 3.2 | 2.4 | 3.6 | 2.8 | 2.7 | 3.5 | 3.0 | 5.2 |
| B | | | 2.2 | 3.7 | 1.6 | 2.4 | 4.0 | 1.6 | 2.0 | 3.4 | 1.4 | 2.0 | 2.6 |
| C | | | | 3.0 | 2.1 | **1.6** | 4.0 | 2.6 | **1.7** | 3.6 | 2.4 | **1.6** | 3.5 |
| D | | | | | 3.0 | 3.3 | 2.7 | 3.9 | 3.0 | 2.9 | 3.8 | 3.2 | 5.3 |
| E | | | | | | 2.2 | 3.5 | 1.7 | 1.5 | 3.0 | 52.0 | 2.3 | 3.4 |
| F | | | | | | | 4.3 | 2.8 | **2.0** | 3.6 | 2.6 | **2.1** | 3.7 |
| G | | | | | | | | 3.8 | 3.4 | 2.1 | 4.0 | 3.8 | 5.4 |
| H | | | | | | | | | 1.8 | 3.2 | 1.7 | 2.6 | 2.9 |
| I | | | | | | | | | | 3.0 | 2.1 | **2.0** | 3.5 |
| J | | | | | | | | | | | 3.3 | 3.5 | 4.9 |
| K | | | | | | | | | | | | 1.9 | 2.6 |
| L | | | | | | | | | | | | | 3.2 |

[a] A = AA2S; B = AAMIII; C = AAFINAL; D = AS2S; E = ASMIII; F = ASFINAL; G = BA2S; H = BAMIII; I = BAFINAL; J = BS2S; K = BSMIII; L = BSFINAL; M = CRYSTAL.

The ASFINAL set of constraints shows the least deviation among the four solution structures, while the BAFINAL set of constraints appears to have the greatest deviation.

The 260 interproton restraints derived from the 200-ms NOESY experiment have been divided into three groups: those between protons of the same sugar ring [intra-d(ribose)], interactions among base protons and sugar ring protons of the same residue (intraresidue), and the interresidue type that includes *all* interresidual interactions.

In each of the four cases (AA, BA, AS, and BS) the degree to which the types of input constraints are obeyed by the structures after 40 ps of restrained MD decreases as the average distance for the constrained proton pairs increases. Generally, it can be seen that the shorter, intraresidue distances are more precisely defined (% $RMS_{vol}$) and more closely obeyed (constraint). It is interesting to note that the most significant deviations of the theoretical NOESY volumes generated from the crystal structure (Prive et al., 1987) lie in the longer interresidue $^1H-^1H$ interactions and that the deviations for the shorter interactions are on an order with those of the solution-calculated structures.

It is clear that the shorter intra-d(ribose) distance restraints are well satisfied in the dynamics simulation, as would be anticipated since the average flatwell deviation (ca. ±0.14 Å) is smaller than that found in the other two groups. However, if a constant error limit is substituted into the flatwell for the percent deviation, the degree to which the longer interresidue distances satisfy the input constraints increases only slightly (from 0.34 to 0.31 Å RMS), but the constraint energy increases considerably (ca. 250 kcal $mol^{-1}$). Inspection of the % $RMS_{vol}$ deviations of Table III indicates that although the shorter distances appear to be more accurately defined, the difference is less pronounced than the degree to which the distances are actually satisfied. As with the constraint term, application of a fixed error flatwell into the later stage of MD refinement does not enhance the degree to which the experimental volumes are satisfied. Thus, the ability to precisely define these longer range distances may, to a large extent, at this point be limited by the larger errors inherent in the integration of small volumes by conventional methods (perhaps computer deconvolution and line-shape fitting will be able to provide more accurately integrated volumes).

In addition, we have evaluated the overall fit of the complete set of Cartesian coordinates for selected structures (Table IV). Structures that show the greatest RMS deviations are those derived from the 80-ms NOESY data and that generally have >3.0 Å RMS deviations from most of the other structures. As expected, deviations for structures within a single family (e.g., the anti A-DNA series) are generally slightly smaller than cross-family comparisons (e.g., anti A-DNA with anti B-DNA), although they do converge toward the final structures. The final structures are quite close, with an RMS deviation of ca. 1.6 Å. The BSFINAL coordinates are the poorest fitting of the four final sets, with deviations of ca. 2.0

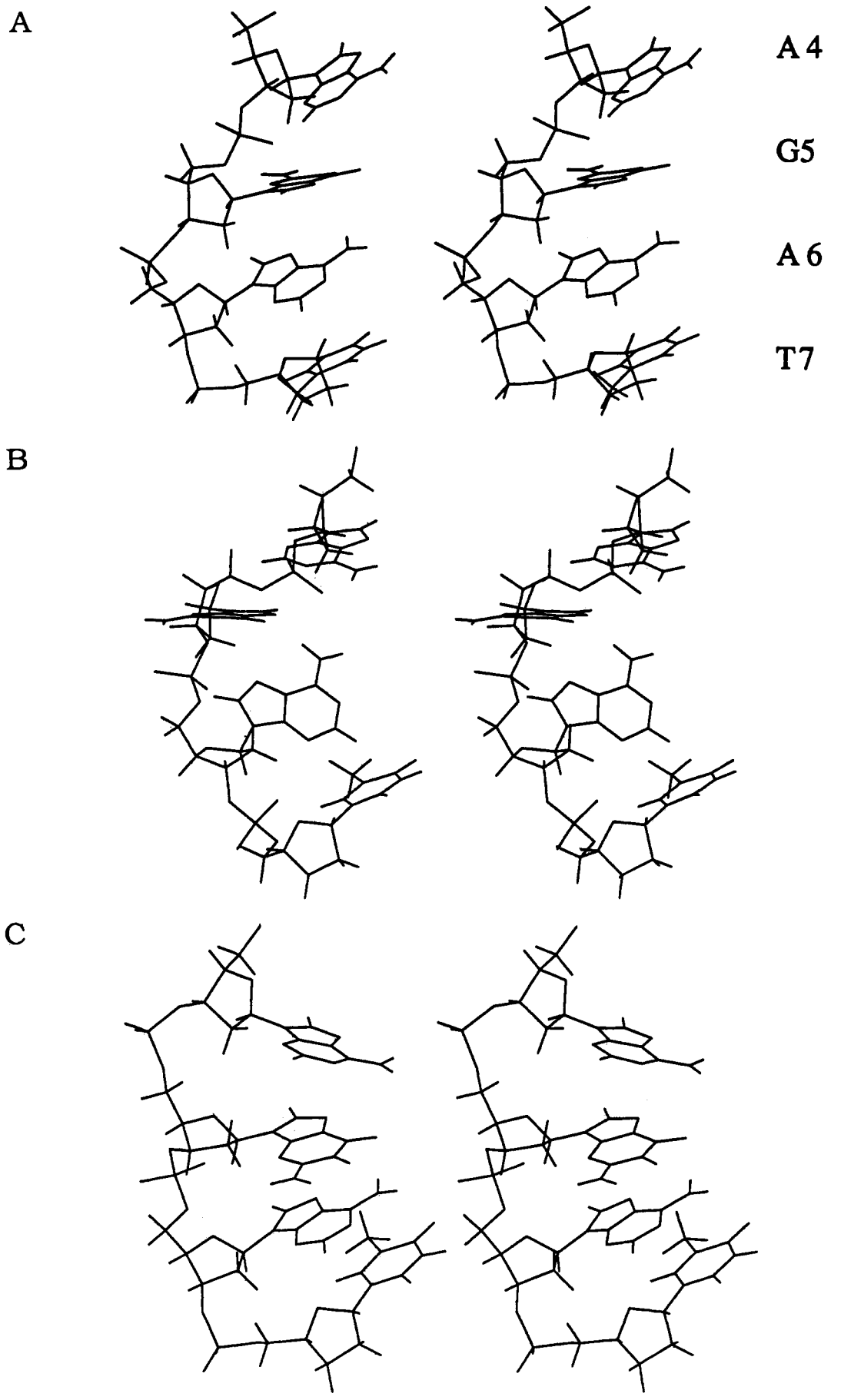


FIGURE 3: Time series for the transition from the syn to the anti conformation for A6 in the AS model. Stereoview of one strand of the central four bases in the decamer at 0.25 ps (A), 1.0 ps (B), and 4.0 ps (C) after the start of NOESY distance restrained MD calculations. Note that the four base pairs are oriented 3′ → 5′ (top to bottom) for clearer visualization.

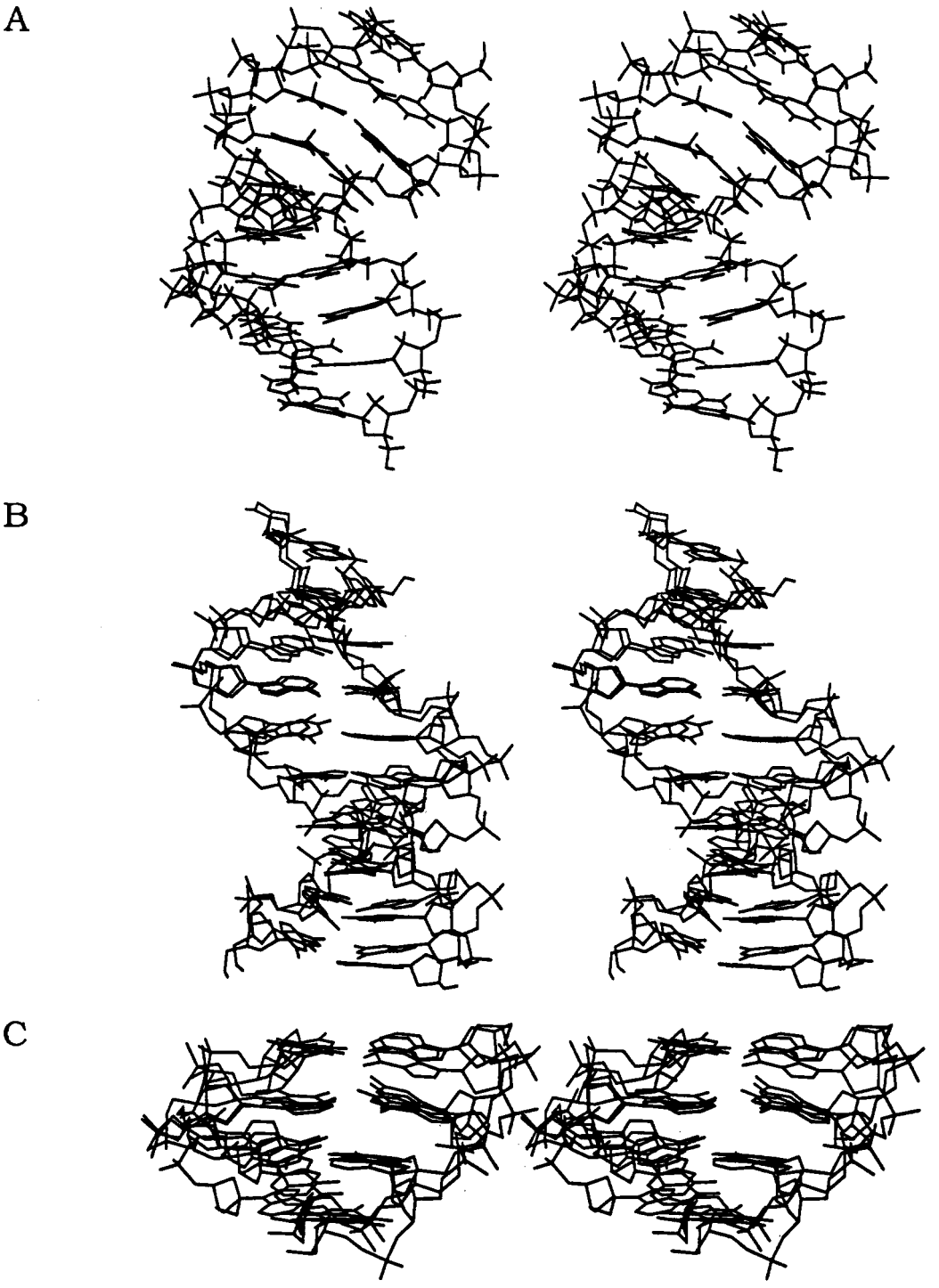


FIGURE 4: Stereoviews of the BAFINAL structure (A) as well as overlays of the final structures derived from the NOESY distance restrained/MORASS hybrid matrix MD refinement of the decamer. Structures are defined by their initial starting models; see Table II for structure label identification. (B) Overlay of AAFINAL (A-DNA with A6 anti) with BSFINAL (B-DNA with A6 syn). (C) Expanded stereoview overlay of the central four mismatch and base pairs of the final structures BAFINAL, AAFINAL, and BSFINAL.

Å. However, one set of comparisons is quite clear, that of the four final structures with the crystal structure; all deviations are greater than 3.2 Å. In fact, the lowest RMS deviation between the crystal and any other structure is that of the minimized (no constraints) model-built Arnott-type B-DNA (1.6 Å, not shown). It is likely, however, that a considerable amount of the deviation between the crystal- and MORASS-derived final structures is due to the bending or "kinking" of the helix axis in the MORASS-refined structures.

Figure 4 depicts various structures and overlays of the final BAFINAL, BSFINAL, and AAFINAL refined duplexes. As can clearly be seen, the overall fit is quite good. The bending caused by the tandem mismatched G and A at the center of the duplex follows the same direction in all four structures. This is in contrast to other kinked DNA structures derived from MD simulations in which the direction of the bend is generally random (Nikonowicz et al., 1990; Rao & Kollman, 1990). We note as before (Nikonowicz et al., 1990) that while NMR can be used to define short distances rather accurately (<5 Å), long-range distances such as that between opposite ends of the helix cannot be directly determined. Thus the degree of bending that may be observed in the full structure can only be defined as the additive effect of parameters defined through short distances, helix twist and helix roll angle, as well

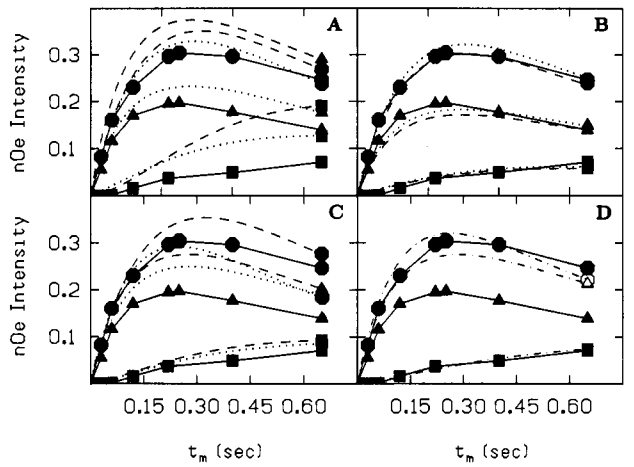


FIGURE 5: Experimental and theoretical NOE buildup curves for C4 H5–C4 H6 (●), A6 H1′–A8 (■), and A12 H1′–A12 H2″ (▲). (A) Time dependence calculated for initial structures AAMODEL (dashed lines) and BAMODEL (dotted lines) vs experimental (solid lines). (B) Time dependence calculated for final structures AAFINAL (dashed lines) and BAFINAL (dotted lines) vs experimental (solid lines). (C) Time dependence calculated for final structures AA2SFINAL (dashed lines) and BA2SFINAL (dotted lines) vs experimental (solid lines). (D) Time dependence calculated for the crystal structure CRYSTAL (dot-dashed lines) vs experimental (solid lines). Initial volumes for the theoretical curves were scaled to the diagonal volume at mixing time 0 for a single proton.

as through backbone torsional angles (Nikonowicz & Gorenstein, 1990).

NOE time development curves have generally been used to define the linear regime of the NOE buildup in order to extract "two-spin" distances. Although it was not necessary to generate a buildup curve in order 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., $\tau_c$ and the set of distances representing the correct structure.

Figure 5 depicts the NOE intensity of three representative pairs of spins (involving short, medium, and long distances) as a function of mixing time. Zero mixing time NOESY diagonal volumes were measured and used to approximate the initial intensities for the theoretical volume matrix simulation with MORASS. The correlation time used for the NOE simulated buildup was 3.6 ns, the value initially estimated. This value was chosen as a result of fitting the theoretical buildup to the experimentally observed buildup of the NOE for a fixed-distance proton pair. The fixed-distance pair, C2 H5/H6, was also one of the marker volumes used to scale the set of experimental volumes at each iteration of the merging procedure. Although the 3.6 ns value actually provides the best fit to the data, there is perhaps a better method to optimize $\tau_c$; that is, choose an initial $\tau_c$ and after a suitable period of refinement (either every merge matrix/MD iteration cycle or after final convergence) reevaluate the choice of $\tau_c$. Although $^{13}C$ relaxation parameters provide good estimates for $\tau_c$ (Davies et al., 1978), we believe that this iterative distance and correlation time MORASS hybrid matrix refinement procedure could provide quite accurate measurements of $\tau_c$ as well.

Figure 5 depicts a comparison of four different sets of theoretical NOE buildup curves with the experimentally determined NOE buildups. It can be clearly seen that the poorest fit to the experimental data is that of the minimized model-built structures (Figure 5A). Figure 5C demonstrates that after 36 ps of two-spin approximation constraints the fit of the NOE buildup for the long-distance proton pair has much improved; however, the theoretical and experimental curves for the shorter pairs are still in poor agreement. Figure 5B shows the improvement of the fit of the three spin pairs after 40 ps of MD with the constraints generated from the fifth MORASS-merging of the 200-ms NOESY data. Finally, in the last panel, the experimental buildup curves are compared to those simulated by using the crystal structure. Although two of the three NOE buildup curves are quite good, the one that does not match involves the d(ribose) ring of the mismatched adenosine residue, (A6). Importantly, the experimental NOE between the H1′ and H2″ protons, a distance that varies between 2.1 and 2.4 Å, is about 60% that of the H5–H6 NOE, a distance of 2.5 Å. Thus, it is clear that the environment in solution surrounding the A6 H1′ and/or H2″ protons is not accurately reproduced in the crystal structure but does appear to be quite good in the final calculated structures AAFINAL and BAFINAL. We also note at this point that the final structures derived from the starting syn models show a slightly poorer fit to the experimental data for these three spin pairs; however, the shapes and relative positioning of the theoretical NOE buildup curves do fit the experimentally observed data.

As a final measure of the progress of our refinement scheme, we have simulated (using MORASS) the 200-ms mixing time NOESY spectra of intermediate and final decamer structures. While there exist more cross-peaks at longer mixing times (400 and 800 ms) than at 200 ms by which to compare theoretical and experimental NOESY maps, the intensities tend to be of similar magnitude. Since spin diffusion effects are moderate at the mixing time of 200 ms, there is a wide range of cross-peak intensities, thus providing a rigorous test of the structural features leading to primary relaxation pathways. Figure 6 shows a comparison of the experimental 200-ms $\tau_m$ base–H1′/H3′ region contour plots with those simulated from the structures AAFINAL, BAFINAL, ASMODEL, and CRYSTAL. We note that the experimental cross-peaks at 4.8 ppm in the $\nu_2$ dimension of the experimental spectrum are of diminished intensity due to the saturation of the HDO peak in the spectrum. [Note: Fitting of the final theoretical NOESY spectra to the experimental data was accomplished through the inclusion of eight more base–H2′/H2″ constraints into the flatwell file for the final 40 ps of dynamics and the final constrained minimization. These were not included during the five iterations of the hybrid matrix refinement procedure because of excessive overlap. Constraint values were chosen by determining the distances between proton pairs to be constrained and adjusting (increasing or decreasing) the distances accordingly.]

Generally, the fit of cross-peak intensities between the experimental spectrum and the simulated spectra of the final structures refined from either initial A-DNA or B-DNA models is quite good. In contrast, the simulated spectra of the initial A-DNA model show rather large cross-peak intensity errors, especially in the base–H2′/H2″ region (spectra not shown) where most of the inter- and intraresidue cross-peaks are of reversed intensity. Noteworthy also is the intensity of the intraresidue A6 H8–A6 H1′ cross-peak in the starting $A_{syn}$ structure and the lack of an interresidue A6 H8–G5 H1′ cross-peak. The NOESY spectrum derived from the crystal structure coordinates also contains cross-peaks such as A4 H8 and T8 H6 protons in the H1′ and H2′/H2″ regions that do not agree with those found in the experimental NOESY map. We also note that the G5 H8–A4 H1′ interaction is considerably more intense in the crystal structure than that observed in the experimental. Other critical cross-peaks are identified in the various spectra. Although the information contained

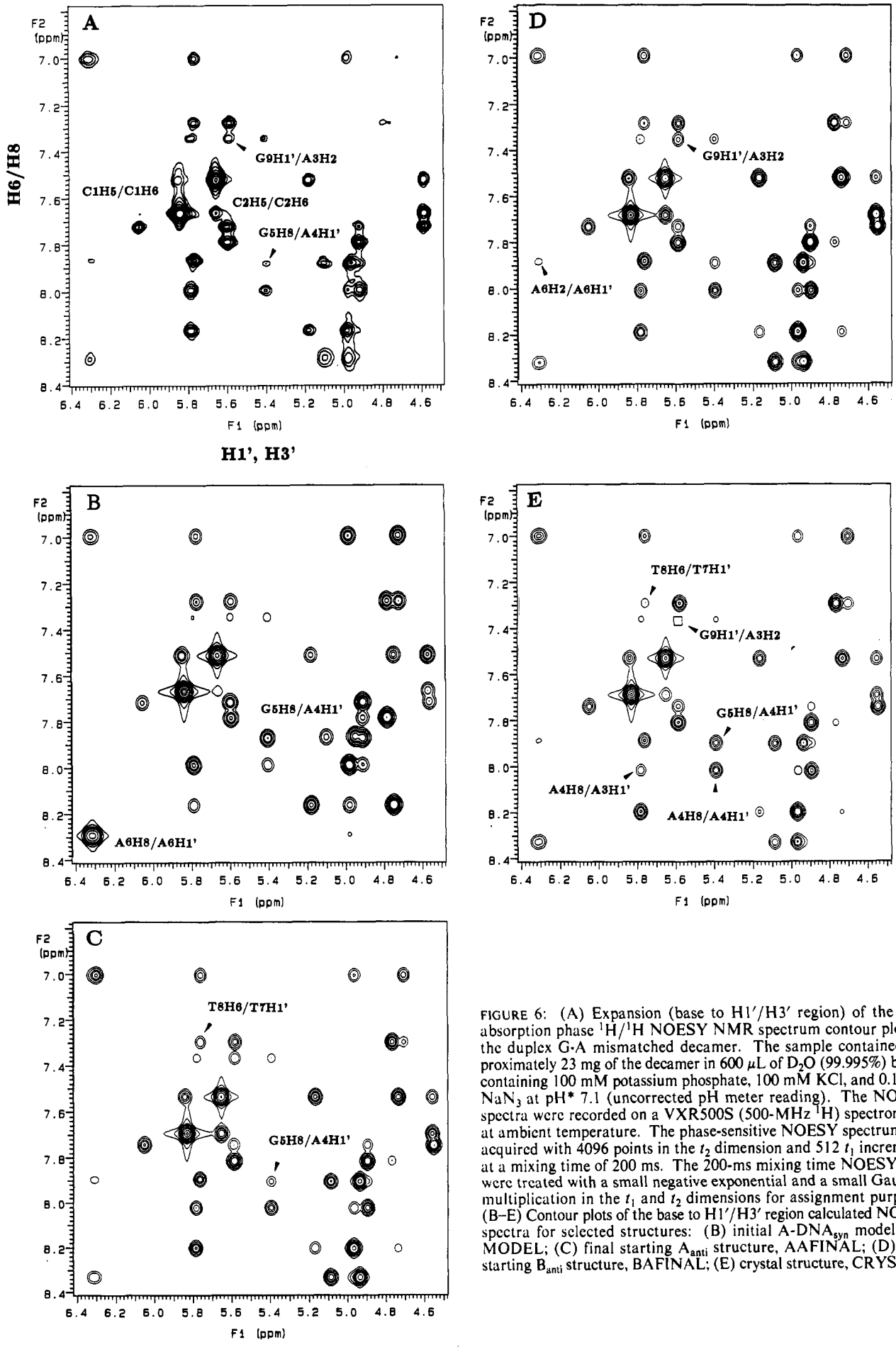


FIGURE 6: (A) Expansion (base to H1′/H3′ region) of the pure absorption phase $^1H/^1H$ NOESY NMR spectrum contour plots of the duplex G·A mismatched decamer. The sample contained approximately 23 mg of the decamer in 600 μL of $D_2O$ (99.995%) buffer containing 100 mM potassium phosphate, 100 mM KCl, and 0.1 mM $NaN_3$ at pH* 7.1 (uncorrected pH meter reading). The NOESY spectra were recorded on a VXR500S (500-MHz $^1H$) spectrometer at ambient temperature. The phase-sensitive NOESY spectrum was acquired with 4096 points in the $t_2$ dimension and 512 $t_1$ increments at a mixing time of 200 ms. The 200-ms mixing time NOESY data were treated with a small negative exponential and a small Gaussian multiplication in the $t_1$ and $t_2$ dimensions for assignment purposes. (B–E) Contour plots of the base to H1′/H3′ region calculated NOESY spectra for selected structures: (B) initial A-$DNA_{syn}$ model, ASMODEL; (C) final starting $A_{anti}$ structure, AAFINAL; (D) final starting $B_{anti}$ structure, BAFINAL; (E) crystal structure, CRYSTAL.

in these plots is dependent in part upon both structure *and* $\tau_c$, we are confident that our choice of correlation time, 3.6 ns, is quite reasonable, as shown in the buildup curves of Figure 5, and that the gross differences observed here reflect differences between the solution conformation and the crystalline state.

## CONCLUSIONS

We have used a hybrid relaxation matrix methodology in combination with restrained molecular dynamics to derive a family of solution structures for the tandem G·A decamer d(CCAA*G*ATTGG)$_2$. The hybrid matrix/restrained molecular dynamics refinement procedure has allowed us to accurately extract interproton distance information that would have been unobtainable had the two-spin approximation been used. Much more accurate and precise distances are obtainable from the hybrid relaxation matrix methodology, and starting from four different initial structures, it is possible to iteratively refine the four initial models to a common family of structures consistent with the NOESY-derived distances. This methodology has also provided a tool to *automate the refinement process* for deriving solution structures from NMR data.

Unconstrained molecular dynamics calculations fail to locate the correct solution or crystal conformation if the calculations start from model-built structures. Finally, while the solution NMR structure is quite comparable to that in the solid state, the solution and solid-state structures are not identical. 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.

**Registry No.** d(CCAACATTGG), 86880-63-9.

## REFERENCES

Baleja, J. D., Moult, J., & Sykes, B. D. (1990) *J. Magn. Reson. 87*, 375–384.

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., & Noggle, J. H. (1979) *J. Am. Chem. Soc. 101*, 5152–5155.

Brown, T., Hunter, W. N., Kneale, G., & Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2402–2406.

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

Clore, G. M., & Gronenborn, A. M. (1989) *Crit. Rev. Biochem. Mol. Biol. 24*, 479–564.

Davies, D. B. (1978) *Prog. Nucl. Magn. Reson. Spectrosc. 12*, 135.

Fazakerley, G. V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G. A., van Boom, J. H., Jones, M., & Radman, M. (1986) *EMBO J. 5*, 3697–3703.

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

Gao, X., & Patel, D. J. (1988) *J. Am. Chem. Soc. 110*, 5178–5182.

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

Kan, L., Chandrasegaran, S., Pulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.

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

Kramer, B., Kramer, W., & Fritz, H. J. (1984) *Cell 38*, 879–887.

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

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

Nikonowicz, E., Meadows, R., Post, C., Jones, C., & Gorenstein, D. G. (1989a) *Bull. Magn. Reson. 11*, 226–229.

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

Nikonowicz, E., Meadows, R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 4193–4204.

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.

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

Post, C. B., Meadows, R., & Gorenstein, D. G. (1990) *J. Am. Chem. Soc. 112*, 6796–6803.

Prive, G. G., Heinemann, U., Chandrasegaran, S., Kan, L., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.

Rao, S. N., & Kollman, P. (1990) *Biopolymers 29*, 517–532.

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.

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

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