Valeur, B., & Weber, G. (1977) *Photochem. Photobiol. 25*, 441–444.
van Dam-Mieras, M. C. E., Slotboom, A. J., Pieterson, W. A., & de Haas, G. H. (1975) *Biochemistry 14*, 5387–5393.
Verheij, H. M., Slotboom, A. J., & de Haas, G. H. (1981) *Rev. Physiol., Biochem., Pharmacol. 91*, 91–203.
Vincent, M., Brochon, J. C., Merola, F., Jordi, W., & Gallay, J. (1988) *Biochemistry 27*, 8752–8761.
Volwerk, J. J., & de Haas, G. H. (1982) *Lipid–Protein Interactions*, Vol. I, pp 69–149, Wiley-Interscience, New-York.
Volwerk, J. J., Dedieu, A. G. R., Verheij, H. M., Dijkman, R., & de Haas, G. H. (1979) *Recl. Trav. Chim. Pays-Bas 98*, 214–220.
Wahl, Ph. (1979) *Biophys. Chem. 10*, 91–104.
Wahl, Ph. (1983) in *Time-Resolved Spectroscopy in Biochemistry & Biology* (Cundall, R. B., & Dale, R. E., Eds.) NATO Advanced Science Institutes Series A: Life Sciences, Vol. 69, pp 497–522, Plenum Press, New York and London.
Waite, M. (1987) *The Phospholipases, Handbook of Lipid Research*, Vol. 5, Plenum Press, New York.
Werner, T. C., & Forster, L. S. (1979) *Photochem. Photobiol. 29*, 905–914.
White, S. P., Scott, D. L., Otwinowski, Z., Gelb, M. H., & Sigler, P. B. (1990) *Science 250*, 1560–1563.
Willis, K. J., Szabo, A. G., Zuker, M., Ridgeway, J. M., & Alpert, B. (1990) *Biochemistry 29*, 5270–5275.

How Accurately Can Oligonucleotide Structures Be Determined from the Hybrid Relaxation Rate Matrix/NOESY Distance Restrained Molecular Dynamics Approach?†

Kumaralal Kaluarachchi, Robert P. Meadows, and David G. Gorenstein*
Department of Chemistry, Purdue University, West Lafayette, Indiana 47907
Received March 8, 1991; Revised Manuscript Received June 26, 1991

ABSTRACT: The accuracy and precision of structures derived from a combined hybrid relaxation rate matrix/NOESY distance restrained molecular dynamics methodology were examined with simulations that included typical experimental errors. NOESY data were simulated for a DNA dodecamer duplex, d-$(CGCGAATTCGCG)_2$, with added volume error of ~20% and low-level thermal noise. Distances derived from a hybrid relaxation matrix analysis of the NOE data were used as constraints in molecular dynamics driven structural refinements of several initial model geometries. The final structures were compared against results obtained from the traditional isolated two-spin approximation treatment of these NOESY volumes and also against refined structures that employed error-free data. Results show that the structures derived from the relaxation rate matrix analysis of the NOESY data are more accurate than those derived from a simple two-spin approximation analysis and it is possible to achieve refinement to the level of simulated experimental error. Results may be significantly improved with the use of either more accurately measured NOESY volumes or additional matrix-derived constraints. Many of the helical parameters and backbone torsional angles may be accurately reproduced by the hybrid matrix methodology.

Of the currently available techniques for structural elucidation, only nuclear magnetic resonance (NMR) and X-ray crystallography allow one to observe the often subtle variations in DNA topology. Many fine reviews of the NMR method provide specific details and historical perspective (Hosur et al., 1988; Kearns, 1984; Patel et al., 1987b; Reid, 1987; Van De Ven & Hilbers, 1988). Briefly, the NMR method relies upon the use of the two-dimensional nuclear Overhauser effect (NOESY) to produce a set of interproton distance constraints that are used to refine model geometries (Broido et al., 1984; Feigon et al., 1983; Frechet et al., 1983; Gorenstein et al., 1990; Hare et al., 1983; Kearns, 1984; Scheek et al., 1984; Schroeder et al., 1987). Refinement methods are varied but have included distance geometry (Havel et al., 1983) and restrained molecular dynamics (Clore et al., 1985a,b, Gorenstein et al., 1990; Kaptein et al., 1985). While NMR spectroscopy has been largely successful in defining overall conformation as well as some sequence-specific variations in the local conformation of DNA (Assa-Munt & Kearns, 1984; Clore et al., 1985a,b; Gorenstein et al., 1988; Lefevre et al., 1987; Nilges et al., 1987; Patel & Shapiro, 1987; Patel et al., 1987a; Rinkel et al., 1987), several studies show disagreement between structures derived from X-ray crystallography and NMR-derived solution conformations (Joshua-Tor et al., 1988; Nikonowicz et al., 1990; Rinkel et al., 1987; Sklenâr & Bax, 1987). For example, a ^1H NMR study of a duplex decamer failed to reproduce sequence-specific variations in the sugar ring conformation predicted by X-ray crystal analysis (Rinkel et al., 1987); similar differences have been found in the backbone torsional angles (Sklenâr & Bax, 1987). The discrepancy between solution NMR structures and crystallographic X-ray structures raises the question whether the sequence-specific structural variations observed in the X-ray crystallographic studies are the result of packing forces. Indeed, variations in the local conformation may be found in different crystalline forms of the same duplex (Jain & Sundaralingam, 1989; Shakked et

†Supported by the NIH (AI27744), the Purdue University Biochemical Magnetic Resonance Laboratory, which is supported by the NSF National Biological Facilities Center on Biomolecular NMR, Structure and Design at Purdue (Grants BBS 8614177 and DIR-9000360 from the Division of Biological Instrumentation), and the National AIDS Research Center at Purdue (AI727713).

0006-2960/91/0430-8785$02.50/0

al., 1989). Alternatively, perhaps the differences between solution and crystal structures simply reflect the intrinsic inability of NMR to accurately define these subtle structural details. In either case, a methodology by which highly accurate NMR-derived distances are obtained will, in principle, be able to provide structures of sufficient accuracy to allow the fundamental differences in crystallographic and solution conformations to be observed.

It has been suggested that the accuracy of the interproton distances obtained from the NOESY spectra has been limited primarily by failure to properly account for spin diffusion—indirect magnetization transfer from multiple spins. The errors in the interproton distances obtained by analyzing NOESY data while not accounting for indirect transfer of magnetization has been well documented (Borgias & James, 1988; Clore & Gronenborn, 1985; Dobson et al., 1982; Olejniczak et al., 1986; Post et al., 1990; Van De Ven & Hilbers, 1988) and may be as large as 1.3 Å. Distances are systematically underestimated for interproton vectors longer than the distance ruler and overestimated for vectors shorter than the ruler. The bias of the error in interproton distances obtained by the two-spin method is of particular worry for three-dimensional structure determination by NMR since such a bias could produce a systematic error in the structure.

In order to avoid spin diffusion many structural studies have relied on the so-called "isolated spin pair approximation" analysis (ISPA; Clore & Gronenborn, 1985; Wüthrich, 1986). The approximation requires that the NOESY-derived distances be obtained from relatively short experimental mixing times where the rate of buildup of the NOE cross-peak intensity is proportional to the inverse sixth power of the interproton distance and the effects of spin diffusion are minimal. Because most of the structurally important longer range NOEs are not observed at these short mixing times, the use of the two-spin (ISPA) approximation has raised concern over the validity of highly refined NMR structures derived by this methodology [Gorenstein et al., 1990; Keepers & James, 1984; Metzler et al., 1990; Nikonowicz et al., 1990; Van De Ven & Hilbers, 1988; however, see Clore and Gronenborn (1989)].

Recently, in order to obtain more accurate distances, the Bloch equations of magnetization have been solved numerically (Banks et al., 1989; Lefevre et al., 1987) or by a complete relaxation matrix approach (Baleja et al., 1990a,b; Boelens et al., 1988; Bothner-by & Noggle, 1979; Gorenstein et al., 1990; Keepers & James, 1984; Nikonowicz et al., 1989a, 1990). These methods account for the effects of spin diffusion, which allows for the measurement of interproton distance with a higher degree of precision and accuracy. 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" [Boelens et al., 1988, 1989; Borgias et al., 1990; Gorenstein et al., 1990; Nikonowicz et al., 1989a,b, 1990, 1991; see also Lancelot et al. (1989) for alternative solutions to this problem]. The hybrid matrix approach involves the substitution of NOE cross-peak volumes calculated from a model structure for overlapped or suspect experimental volumes. This hybrid volume matrix is diagonalized, producing a hybrid relaxation rate matrix. The rates are calculated with the inclusion of multispin effects and may more accurately reflect the local environment; spin diffusion is explicitly treated. The resulting cross-relaxation rates are converted into interproton distances, assuming (at this time) a simple isotropic motional model for the spectral density functions, and the distances are applied as constraints in a molecular dynamics search of Cartesian space. Coordinates are obtained from molecular dynamics (MD), energy minimized, and serve as the model geometry for the next hybrid matrix/MD cycle. The process is continued until there is minimum difference between the experimental NOESY spectra and the spectrum calculated from the refined structure. As shown by our laboratory (Gorenstein et al., 1990; Nikonowicz et al., 1989a, 1990, 1991), Kaptein and co-workers (Boelens et al., 1988, 1989), and James and co-workers (Borgias et al., 1990) a moderate number of iterations appears to be adequate to achieve convergence to a "refined" structure.

Table I: RMS Difference (Å) in Global Cartesian Coordinates

	A-model	B-model	target
A-model		5.71[a]	5.58[a]
B-model	4.35[b]		1.21[a]
target	5.58[b]	2.41[b]	

[a] RMS of model-built energy-minimized structures. [b] RMS after unconstrained molecular dynamics on B-DNA model.

Using simulated distance data, several papers have addressed the question of the accuracy and precision of distance geometry and NOESY distance restrained molecular dynamics to define duplex structures in solution (Gronenborn & Clore, 1989; Nilsson et al., 1986; Pardi et al., 1988). In a previous publication from our laboratory (Post et al., 1990), we tested the ability of the complete relaxation method to provide accurate and precise structures of proteins and nucleic acids. The basic conclusions were that the complete relaxation method was indeed superior to the two-spin approximation methodology and that back-calculation of the NOESY volume matrix could more accurately reproduce the correct cross-relaxation rates and distances. However, if there were an insufficient number of NOESY volume matrix cross-peak elements, the matrix methodology was severely limited. As indicated above, one possible solution to this problem is the hybrid matrix method. In this paper we test the accuracy and precision of this hybrid matrix/restrained MD procedure by simulating the refinement against a known target structure and simulated NOESY spectrum. A comparison with the ISPA, two-spin approximation as well as the distance restrained molecular dynamics methods is also provided. Finally, it has often been assumed that the phosphate ester backbone is essentially disordered in crystal and NMR-derived solution structures. Analysis of a variety of helical parameters and backbone torsional angles is given, and it is shown that the hybrid matrix/restrained MD methodology can indeed provide direct information on the time-averaged conformation.

Materials and Methods

Molecular Modeling. The target molecule in all cases was the crystallographic coordinates for a DNA dodecamer d-$(CGCGAATTCGCG)_2$ (MPD7; Fratini et al., 1982) onto which hydrogens had been added. The crystal structure with added protons was subjected to a few cycles of energy minimization to remove unfavorable van der Waals contacts.

Energy-minimized, AMBER-built A- and B-form DNA were considered as starting geometries. The Cartesian coordinate rms deviations between the three geometries are given in Table I. As shown, the difference between model B-form and the target geometry is very small. Thus, the B-DNA model was subjected to 12 ps of unconstrained molecular dynamics at 298 K. Averaging and energy minimization of the last 2 ps produced a B-form geometry that was suitably removed from the target. We felt this was necessary in order to analyze the ability of the method to obtain the correct conformation from models with initial geometries that were dissimilar to each

other and the target. Visualization and plotting of all structures was performed with the molecular modeling program MIDAS (Ferrin & Langridge, 1980) operating on a Silicon Graphics 4D Personal Iris workstation.

NOESY Volume and Distance Constraint Sets. The target "experimental" NOESY spectra were simulated from the minimized crystallographic coordinates of the dodecamer with an isotropic correlation time of 3.0 ns (additional motion due to fast methyl rotation was included), at NOESY mixing times of 50 and 200 ms. For the experimental data simulations, a Gaussian distribution of errors with a σ of $\sim$20% was added proportionally to the intensity of all NOESY volumes to approximate inaccurate cross-peak integration. For the 200-ms simulation, thermal noise of $\sim$0.1% was added additionally to all peaks regardless of intensity to approximate the effects of thermal noise. Thermal noise of 0.2% was added at 50 ms along with the 20% integration error. In order to estimate the minimum amount of rms volume error one could expect from the following analysis, the NOESY matrix derived from the target with added error was subtracted from the NOESY data matrix without error. Errors of 27.2 and 26.3% for RMS_{vol}^{exp} and RMS_{vol}^{the} were observed for constraint set 1 (see below for description of RMS_{vol} and constraint sets). Thus, the effects of noise and integration error in this examination limit the minimum expected volume errors to $\sim$26%.

Five different sets of constraints were considered. Set 1 consisted of interproton NOE *volumes* typical of what may be observed at a 200-ms mixing time for a 12 base pair oligonucleotide, giving a total of 134/strand or $\sim$11/nucleotide. In constraint set 2, a total of 239 NOEs/strand were included. While set 2 is unrealistic in number if one considers only nonoverlapped cross-peaks, it may represent the results obtainable from volume integration of crowded regions by spectral deconvolution methods. Importantly, both sets 1 and 2 are NOE volumes. It is these volumes that will be used in the hybrid matrix analysis to derive the corresponding interproton distance constraints. It should also be noted that after the last iteration cycle interproton distances were obtained from the relaxation rates and applied back onto the initial model geometries in a single 40-ps dynamics run to give the final structures.

The third set of constraints, set 3, was obtained for the same interproton vectors as in set 1. However, these constraints were derived from an analysis of the NOESY volumes via the ISPA at 200 ms and thus resulted in *distance* constraints that were applied to the model geometries directly in a single restrained MD calculation. As stated earlier, the two-spin approximation assumes that the rate of buildup of NOESY cross-peak intensity, v_{ij}, which is proportional to the cross-relaxation rate, σ_{ij}, between protons i and j, is inversely proportional to the sixth power of the distance between the protons r_{ij}. An "internal ruler" is used as a reference NOE (v_{kl}) that provides a reference distance (r_{kl}) between protons k and l (Wagner & Wüthrich, 1979). For this study, C3 H5–H6 (2.45 Å) was used as the "internal ruler" reference distance.

It is now well appreciated that significant errors can arise from ISPA data acquired at 200 ms, and thus an additional data set (set 4) was created assuming a more realistic NOESY mixing time of 50 ms. This data set was also analyzed via the two-spin approximation (C3 H5–H6 as the reference distance) and resulted in a generous 87 interproton distances/strand.

Finally, set 5 consisted of the same interproton vectors as in set 1, but the distance constraints in set 5 were derived directly from the target structure. No errors were introduced in these distances, henceforth referred to as error-free data.

Table II: Analysis of Interproton Distance and Volume Constraints/Strand

constraint type	50 ms[a]	200 ms[b]	200 ms[c]
$sugar–sugar_{intra}$	47	59	119
$sugar–base_{intra}$	15	40	49
$sugar–sugar_{inter}$	0	1	23
$sugar–base_{inter}$	21	26	41
other	4	8	7
total	87	134	239

[a] Set 4: 87 interproton distance constraints derived from the two-spin approximation. [b] Set 1: 134 volume constraints/strand set for the hybrid matrix. Set 3: 134 distance constraints derived from the two-spin approximation. Set 5: 134 distance constraints derived from model geometry. [c] Set 2: 239 volume constraints/strand set for the hybrid matrix.

A summary and analysis of all constraint sets is given in Table II. In addition to those listed in Table II, a set of 12 hydrogen-bond constraints (1/base pair at 1.90 Å) were applied during all refinements.

Molecular Mechanics/Dynamics Calculations. Energy minimizations and molecular dynamics were carried out with the molecular mechanics/dynamics program AMBER 3.0 (Weiner & Kollman, 1981). Constrained dynamics employed a flat-well modification of the AMBER potential (Gorenstein et al., 1990). All simulations were performed in vacuo using a distance-dependent dielectric potential with the nonbonded cutoff distance set to 8.5 Å. Charges on the 3′ and 5′ protons were modified to prevent unwanted bond formation. Conjugate gradient minimizations were calculated with an initial step length of 5×10^{-4} and a maximum step length of 1.0. Energy minimizations were continued 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. For the molecular dynamics simulations, 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 obtained at 50-fs intervals.

Refinement for the hybrid matrix analysis utilized separate 7-ps cycles of molecular dynamics (MD). The temperature during each dynamics simulation was varied as follows: 0–1 ps at 800 K, 1–2 ps at 600 K, 2–3 ps at 400 K, 3–4 ps at 350 K, and 4–7 ps at 298 K. Cartesian coordinates for the last 2 ps of each dynamics run were averaged and energy minimized with and without the NOE constraints. Initial force constants and error bars for the flat-well modified AMBER NOE pseudoenergy term were 12 kcal/mol·Å^2 and 15%, respectively. At each cycle, the force constants were increased 2–3 kcal/mol and error bars were decreased 2–3% to reflect the increasing confidence in the accuracy of the distances. The final hybrid matrix/restrained MD cycles for all constraint sets were run at ±5% error and 30 kcal/mol force constant. As stated above, after the last iterative cycle, the set of interproton distances obtained from the relaxation rates were applied back onto the initial model geometries in a single 40-ps dynamics run to give the final structures. This iterative scheme is represented in Figure 1.

For the 40-ps dynamics simulations the temperature was varied as follows: 0–5 ps at 800 K, 5–10 ps at 600 K, 10–15 ps at 400 K 15–20 ps at 350 K, and 20–40 ps at 298 K. Cartesian coordinates from the last 10 ps of the single-run 40-ps simulations (a total of 200 sets of x,y,z coordinates) were averaged and energy minimized to give the final geometries. The reported torsional angles were calculated by averaging the last 10 ps of the single-run 40-ps simulations, without energy minimization. Constraining distances for the 40-ps simulations came from the two-spin approximation treatment

Table III: Analysis of Intermediate Structures Derived from Constraint Set 1 Starting from the A-DNA Model

structure	$\%RMS_{vol}^{exp}$ [a]	$\%RMS_{vol}^{the}$ [a]	*R* factor[b]	RMS_{xyz} [c]	RMS_{dij} [d]	energy[e]	energy[f]	cons.[g]
iteration no.[h]								
1	201.0	326.3	0.62	5.58	0.83	−861		
2	105.8	83.8	0.45	4.24	0.52	−936	−838	46.9
3	90.1	58.3	0.33	3.39	0.37	−949	−882	30.6
4	64.1	39.1	0.21	2.86	0.30	−945	−843	56.7
5	42.3	30.3	0.20	3.00	0.28	−943	−874	37.1
6	31.0	28.4	0.17	2.33	0.27	−935	−849	46.9
7	29.8	30.2	0.17	3.01	0.27	−941	−813	65.4
final MD[i]								
40 ps	23.6	24.6	0.15	3.04	0.23	−951	−766	107.4
80 ps	23.7	24.2	0.15	2.54	0.24	−954	−772	107.0

[a] %RMS from eq 1. [b] Crystallographic *R* factor [see Gorenstein et al. (1990)]. [c] RMS (Å) in Cartesian coordinates between structure and the target dodecamer. [d] RMS (Å) of distance constraints between structure and target dodecamer. [e] Total energy of structure minimized with no NOE constraints. [f] Total energy of structure minimized with NOE constraints. [g] Constraint energy of structure minimized with NOE constraints. [h] Each iteration represents a MORASS hybrid matrix calculation of a new set of distances from the rate matrix, followed by a 7-ps cycle of molecular dynamics (MD). The temperature during each dynamics simulation was decreased from 800 to 298 K (see Materials and Methods). Cartesian coordinates for the last 2 ps of each dynamics run were averaged and energy minimized with and without the NOE constraints. [i] After the final, 7th iterative cycle, the set of interproton distances were obtained from the final relaxation rate matrix and were applied back onto the initial model geometries in a single 40-ps dynamics run to give the final structures. The temperature during the 40-ps dynamics simulation decreased from 800 to 298 K. Cartesian coordinates from the last 10 ps (at 298 K) of the single-run 40-ps simulation were averaged and energy minimized to give the final structure.

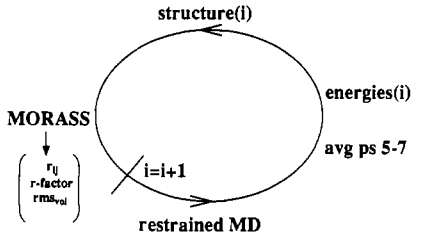


FIGURE 1: Schematic description of the MORASS hybrid relaxation matrix method, iterative cycles. Starting from an initial structure (i), a hybrid volume matrix is created by replacing the theoretical volume matrix elements, v_{ij}^{the}, with the well-resolved "experimental" volume matrix elements, v_{ij}^{exp}. These "experimental" volume matrix elements (i.e., the NOESY spectrum) have been simulated in this study by the MORASS program on the basis of the crystal structure of the dodecamer. MORASS is then used to calculate the hybrid rate matrix, from which the distances are derived. RMS deviations of the NOESY volumes and *R* factors are also computed at this cycle. Distances from the rate matrix are then used in a distance restrained molecular dynamics refinement, to yield a new set of coordinates for the *i*+1 structure (following averaging of the 5–7th ps structures and minimization) that are used to calculate new theoretical rate and volume matrices. Iteration continues until the experimental and theoretical NOESY volume matrices converge.

of the experimental data (sets 3 and 4) and were derived from the last hybrid matrix/restrained MD cycle for each of the models (sets 1 and 2) or came directly from the target geometry (set 5). Force constants and error bars for the 40-ps simulations were 40 kcal/mol and ±5%, respectively.

Relaxation Matrix Calculations. The program MORASS (MORASS: Multiple Overhauser Relaxation AnalysiS and Simulation) (Meadows et al., 1989; Post et al., 1990; available upon request) was used to calculate volume and rate matrices by the complete relaxation matrix approach (Gorenstein et al., 1990; Keepers & James, 1984; Olejniczak et al., 1986) as well as to implement the hybrid matrix methodology (Boelens et al., 1988; Nikonowicz et al., 1989a, 1990). In our laboratory, the refinement process is monitored by the following:

$$\%RMS_{vol} = \sqrt{\frac{1}{N}\sum_{ij}[(v_{ij}^{the} - v_{ij}^{exp})/v_{ij}^{c}]^2} \times 100\% \quad (1)$$

where v_{ij}^{the} and v_{ij}^{exp} represent the theoretical and experimental volume matrix elements, respectively, the RMS_{vol}^{exp} is the error between calculated and experimental volumes determined with the experimental volume in the denominator of eq 1, while RMS_{vol}^{the} contains the calculated volume in the denominator (Nikonowicz et al., 1990). An *R* factor analogous to the X-ray crystallographic *R* factor is also defined (Nikonowicz et al., 1990).

A stepwise or perturbational merging of the experimental and theoretical NOESY volumes was implemented to improve the diagonalization behavior of the hybrid volume matrix. In order to form the hybrid volume matrix, the experimental and theoretical data sets were scaled via the appropriate adjustment of the theoretically derived volumes by using the NOEs from C1, C3, C9, and C11 H5–H6 as scaling markers; an average scaling factor (Boelens et al., 1988) was used.

Structure Analysis. The structural RMS calculations were determined by using a program written by Julian Tirado-Rives. Helical analysis was performed with the program NEWHEL90 (Dickerson, personal communication). The program ANGLES (R. Meadows, unpublished) was used to calculate the time average of the dihedral angles and obtain the time dependency of the six backbone torsional angles from the dynamics simulations. Note that the torsional angles α–ζ are defined as follows: (P^{α}–$O5'^{\beta}$–$C5'^{\gamma}$–$C4'^{\delta}$–$C3'^{\epsilon}$–$O3'^{\zeta}$–P).

Variations in some of the local helical parameters (Saenger, 1984) are dependent upon the method by which the parameters are defined relative to a local or global helical axis for curved helices. For this analysis a global helix axis was defined via the positions of successive C1′, N1, and N9 atoms.

RESULTS

Several criteria were used to determine the accuracy and precision of the different methodologies used for refinement. First, for the simulations involving the hybrid relaxation rate matrix methodology, the deviation between the simulated "experimental" NOESY volumes and the calculated spectrum from the dynamics simulation was compared via eq 1. Also, the distances of the constrained protons pairs when compared to the same distances in the target geometry give the RMS_{dij} as shown in Tables III–V. The RMS_{vol} and RMS_{dij} results are given for constraint set 1 in Table III and Table 1S (see supplementary material) and are summarized for all final structures in Tables IV and V. Note that the structures derived from sets 3, 4, and 5 did not undergo the cyclic MORASS/MD refinement; the RMS_{vol} reported in the tables for these constraint sets were derived by simulating the NOESY

Table IV: Summary of Final Structures Derived from the A-DNA Model with Constraint Sets 1–5

set no.	$\%RMS_{vol}^{exp}$	$\%RMS_{vol}^{the}$ [a]	*R* factor[b]	RMS_{xyz}[c]	RMS_{dij}[d]	energy[e]	energy[f]	cons.[g]
1	23.6	24.6	0.15	3.04	0.23	−951	−766	107.4
2	22.2	17.2	0.14	1.05	0.21	−947	−878	43.4
3	47.7	66.5	0.24	4.13	0.27	−959	−899	23.2
4	30.6	41.9	0.15	3.00	0.27	−947	−899	23.2
5	17.5	16.4	0.12	1.36	0.17	−953	−939	16.3

[a–g] See Table III for footnotes.

Table V: Summary of Final Structures Derived from the B-DNA Model with Constraint Sets 1–5

set no.	$\%RMS_{vol}^{exp}$ [a]	$\%RMS_{vol}^{the}$ [a]	*R* factor[b]	RMS_{xyz}[c]	RMS_{dij}[d]	energy[e]	energy[f]	cons.[g]
1	24.6	30.0	0.17	1.55	0.26	−961	−860	49.1
2	22.8	17.7	0.15	1.25	0.20	−950	−900	31.6
3	44.3	64.0	0.23	3.58	0.24	−960	−930	13.9
4	29.7	25.5	0.16	3.20	0.25	−958	−933	16.0
5	18.9	16.2	0.14	2.58	0.17	−957	−935	15.8

[a–g] See Table III for footnotes.

data from the final structures and comparing the simulated NOESY against the target.

As shown in Table III and Table 1S, the RMS_{vol} deviations start off rather high, especially for the A-model (200–330%), and decline uniformly to values in the mid-20s. Importantly, although the RMS_{vol} for set 2 are slightly lower, both constraints sets produce structures that meet or surpass the minimum expected rms errors of 26%. The NOESY spectra for the final structures derived from constraint set 3 (Tables IV and V) show large RMS_{vol} errors (44–67%; *R* factor 0.23–0.24) regardless of the initial geometry. The spectra calculated from set 4, the 50-ms ISPA data, while having smaller errors than the set 3 derived structures, also exhibit some deviation from the target NOESY spectra (Tables IV and V; $RMS_{vol} \sim$ 25–42%). The error-free data (set 5) produced structures that show excellent results, starting from both A- or B-form models ($RMS_{vol} \sim$ 16–19%; *R* factor 0.12–0.14). Importantly, set 2, with 239 constraints/strand obtained via the MORASS matrix refinement was able to reproduce the target geometry with the highest degree of accuracy of any of the "experimental" data sets (Tables IV and V; $RMS_{vol} \sim$ 17–23%; *R* factor 0.14–0.15).

Importantly then, the final RMS_{vol} errors for the MORASS cycles show clearly that, in this simulation study, the combined hybrid matrix and molecular dynamics methodology is able to obtain interproton distances that are effectively limited by the amount of experimental error introduced into the experimental NOESY volumes, not by an inability to account for spin diffusion. Any further improvement in the RMS_{vol} or distance rms errors will require more accurately integrated volumes. An even greater improvement in the structures may be derived by including additional crosspeaks (i.e., compare set 1 vs set 2).

Because the coordinate geometry from which the experimental NOEs were derived is known, one measure of the reliability of the molecular dynamics/MORASS methodology is observed in the rms of Cartesian coordinates (RMS_{xyz}) between each structure and the target geometry. In general, the rms deviations between the target and the MORASS calculated structures are lower than for structures derived from either of the ISPA data sets (Tables IV and V). The greatest differences between the target and final geometries occurs for the two-spin treatment of data set 3 ($RMS_{xyz} \sim$ 3.6–4.1 Å). The results from set 4 (50-ms ISPA; $RMS_{xyz} \sim$ 3.2 Å) and set 1 (MORASS) final geometries ($RMS_{xyz} \sim$ 1.6 Å) show that the matrix analysis produces a much better fit (starting from the B-DNA model geometry; Table V). If one increases the number of constraints, while still maintaining the same level

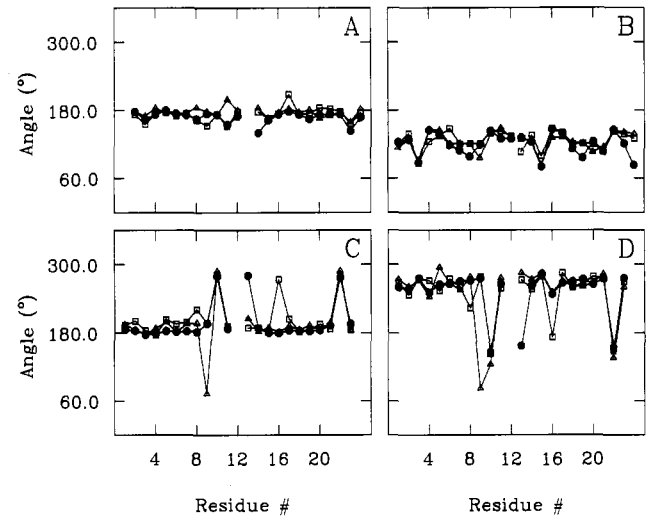


FIGURE 2: Backbone dihedral angles vs residue number for target (●) and final structures derived from A-DNA (□) and B-DNA (△) starting geometries by using MORASS volume constraint set 1. Angles shown are β (A), δ (B), ε (C), and ζ (D).

of noise and error within the NOE volumes (set 2), the accuracy of the matrix method increases remarkably, as shown in Tables IV and V. Indeed, set 2 produces structures whose overall fit to the target molecule as measured by RMS_{xyz} (1.1–1.3 Å) is better than even the error-free constraints (remembering, of course, that set 2 has 239/strand while the error-free data contained 134/strand).

The comparison of local helical parameters and backbone torsional angles utilizes results from averaging the last 10 ps of a 40-ps simulation of distance constrained MD. Figures 2–7 and supplemental Figures 1S–4S show plots of the backbone dihedral angles and local helical parameters for structures derived from sets 1–5. As can be seen, the backbone dihedrals from the perfect data (set 5; Figure 5) constrained MD runs are in excellent agreement with the torsional angles in the target molecule. Further, the major and minor groove widths, base roll, tilt and helix twist angles are also in general in reasonable agreement with the target model (filled circles in supplementary Figure 4S, the results are comparable to those of Figure 7). These perfect data simulations thus suggest that, indeed, interproton distance restraints in principle derivable from NOESY spectra function to adequately define *both* the local helical parameters *and the backbone torsional conformation.*

Many of the target helical parameters and backbone torsional angles may also be reproduced by the structures derived

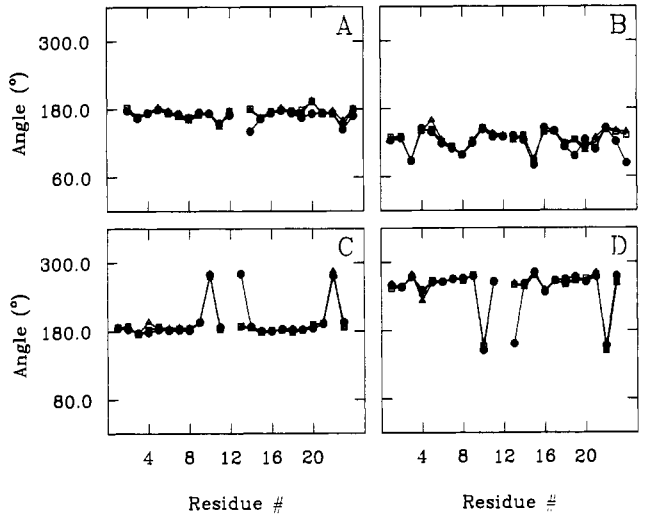


FIGURE 3: Backbone dihedral angles vs residue number with MORASS volume constraint set 2. See the legend to Figure 2.

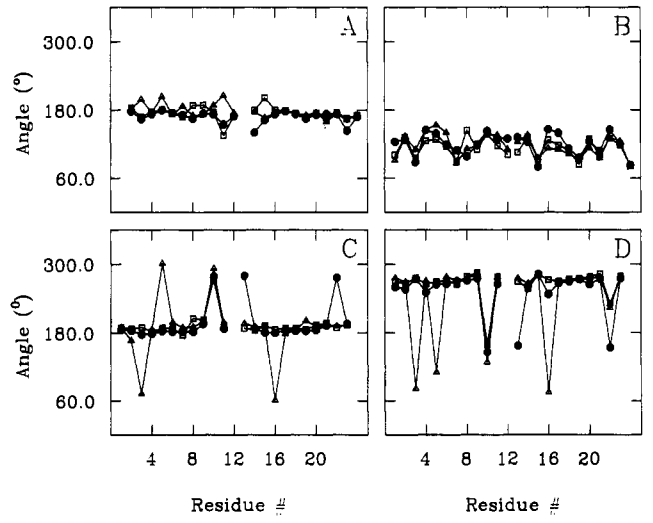


FIGURE 4: Backbone dihedral angles vs residue number with ISPA distance constraint set 3. See the legend to Figure 2.

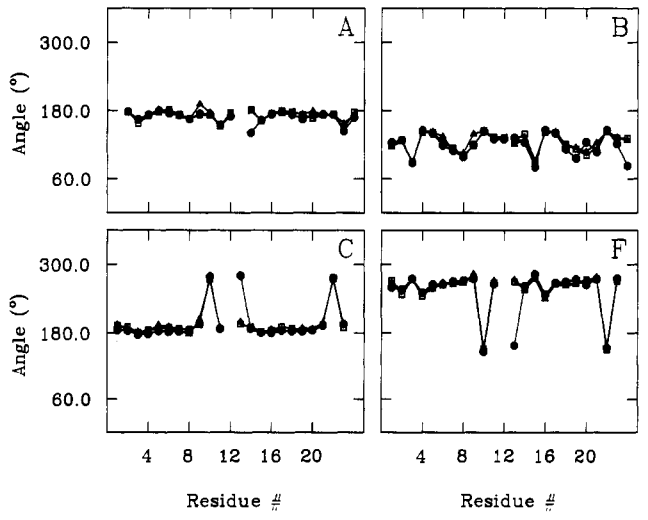


FIGURE 5: Backbone dihedral angles vs residue number with perfect distance constraint set 5. See the legend to Figure 2.

from the multi-spin, hybrid matrix MORASS analysis (as seen in Figures 2, 3, and 6 and supplementary Figure 2S). If one examines the results from set 1, the structures refined with the matrix method reproduce the sequence-specific variation

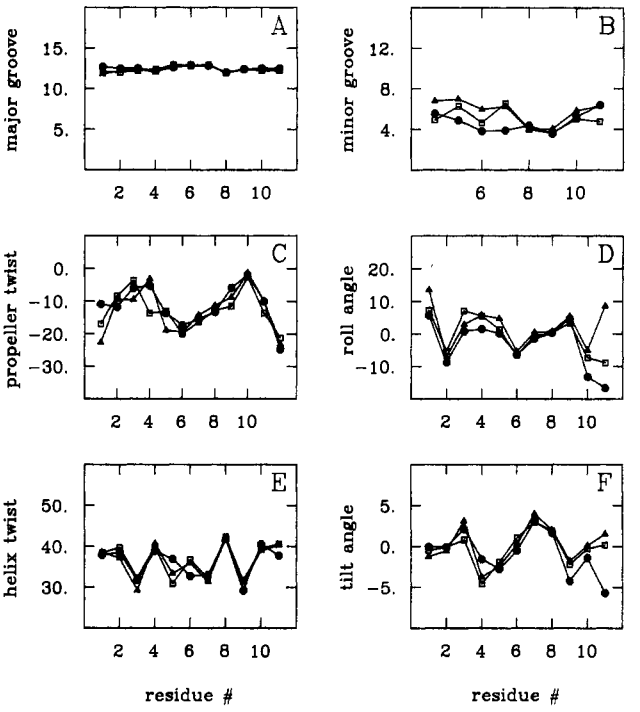


FIGURE 6: Various helical parameters for target (●) and final structures derived from A-DNA (□) and B-DNA (△) starting geometries by using constraint set 2. Local geometry parameters are major groove width (A), minor groove width (B), propeller twist (C), helix roll (D), helix twist (E), and helix tilt angles (F).

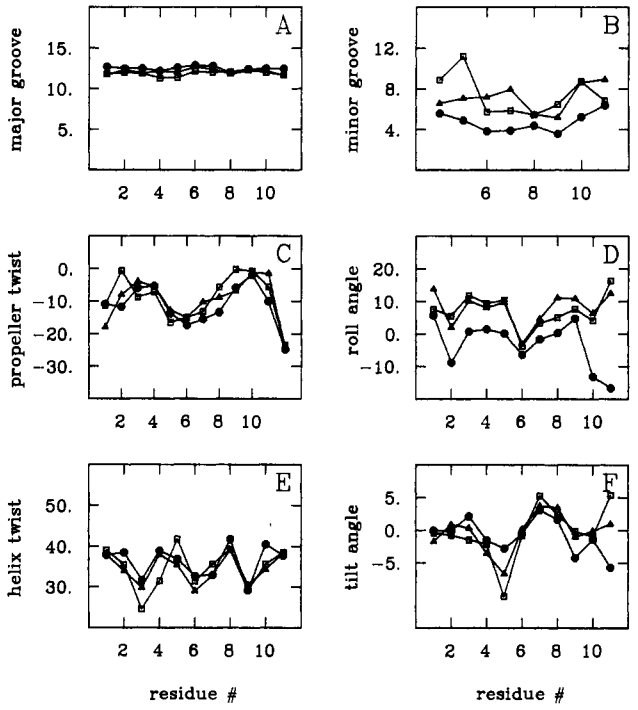


FIGURE 7: Various helical parameters obtained with constraint set 4. See the legend to Figure 6.

in the backbone dihedrals for four (β, δ, ϵ, ζ) of the six torsional angles remarkably well (Figure 2A–D). The two torsional angles that do not reproducibly fit the target structure very well, α and γ, show a strongly coupled "crankshaft" variation, where torsion about one of the bonds results in an equal and opposite rotation about the other sugar phosphate bond (Dickerson, 1983; Dickerson & Drew, 1981). The net effect of this crankshaft motion is to keep the length of the sugar–phosphate backbone approximately the same (see below).

Because no direct distance constraints fix the positions of these phosphate ester atoms, it is thus not surprising that the MORASS/MD refinement has uncovered alternative but quite equivalent "solutions," (α,γ pairs may be g$^-$,g; t,t; or g,g$^-$). However, adding more volume constraints to the MORASS restrained MD simulation (set 2) results in an even better reproduction of nearly all of the backbone torsional angles throughout the sequence (Figure 3). (Interestingly, the MORASS volume data set 2 appears to do a slightly better job than even the perfect distance data set 5.) Examination of the variation in the backbone torsional angles for structures derived from either of the two-spin analyses (sets 3 and 4) yields much less satisfactory results (Figure 4 and supplementary Figure 1S).

In general, analysis of the local helical parameters for the different data set simulations (Figures 6 and 7) and supplementary Figures 2S–4S follow a similar pattern as described for the backbone torsional angles. Again, the MORASS set 2 (Figure 6) is best, followed by perfect distance data, set 5 (supplementary Figure 4S). The ISPA 200-ms data are unable to reproduce the target geometry in all but the most insensitive parameters (supplementary Figure 3S). Also, the ISPA 50-ms data, set 4 (Figure 7), while analyzed in a regime that does not exhibit large errors due to spin diffusion, apparently do not allow one to extract enough of the structurally important NOEs to give accurate structures (remembering that our analysis included a very generous 87/strand).

Progress in the MORASS/restrained MD refinement was also monitored by the change in the overall potential energy and the contribution to the overall energy from the constraining potential (Tables III–V). The energies of the intermediate structures minimized without NOE constraints decrease during the refinement (Table III). The energies of the final structures show only small differences, regardless of the initial structural model or constraint set (Tables IV and V).

Finally, the relative success of the ISPA vs MORASS methods in reproducing the overall geometry of the duplex target structure is conveyed by the stereo overlays of the target structure with the final refined structures starting from either the A or B-DNA model (Figure 8A–E).

The degree of overlap gives a visual confirmation of the quantitative differences in RMS_{xyz} found in Tables IV and V as well as qualitative measure of the accuracy of the methodologies. In addition, the difference between the A-DNA and B-DNA model refinements afford a measure of the precision of the methods. Thus the best fit of the final structure to the target dodecamer is shown by the MORASS set 2 ($RMS_{xyz} \sim 1.1$–1.3 Å). The poorest fit is found for the ISPA set 3 ($RMS_{xyz} \sim 3.6$–4.1 Å). Even though there are no long-range distance constraints, in all of the refined structures the duplex is found to be quite linear (as is the target). Note that the structures that show the greatest deviations from the target molecule (ISPA sets 3 and 4) still show relatively good RMS deviations for the distance constraints (RMS $r_{ij} \sim 0.24$–0.27; Tables IV and V). Although the MORASS structures show slightly tighter RMS r_{ij} deviations, the RMS_{xyz} deviations for the MORASS structures are considerably better, largely because the cumulative effect of small errors is magnified along the entire length of the helix. This is clearly seen in the poor fit at the bottom of the duplex in Figure 8C,D.

Discussion

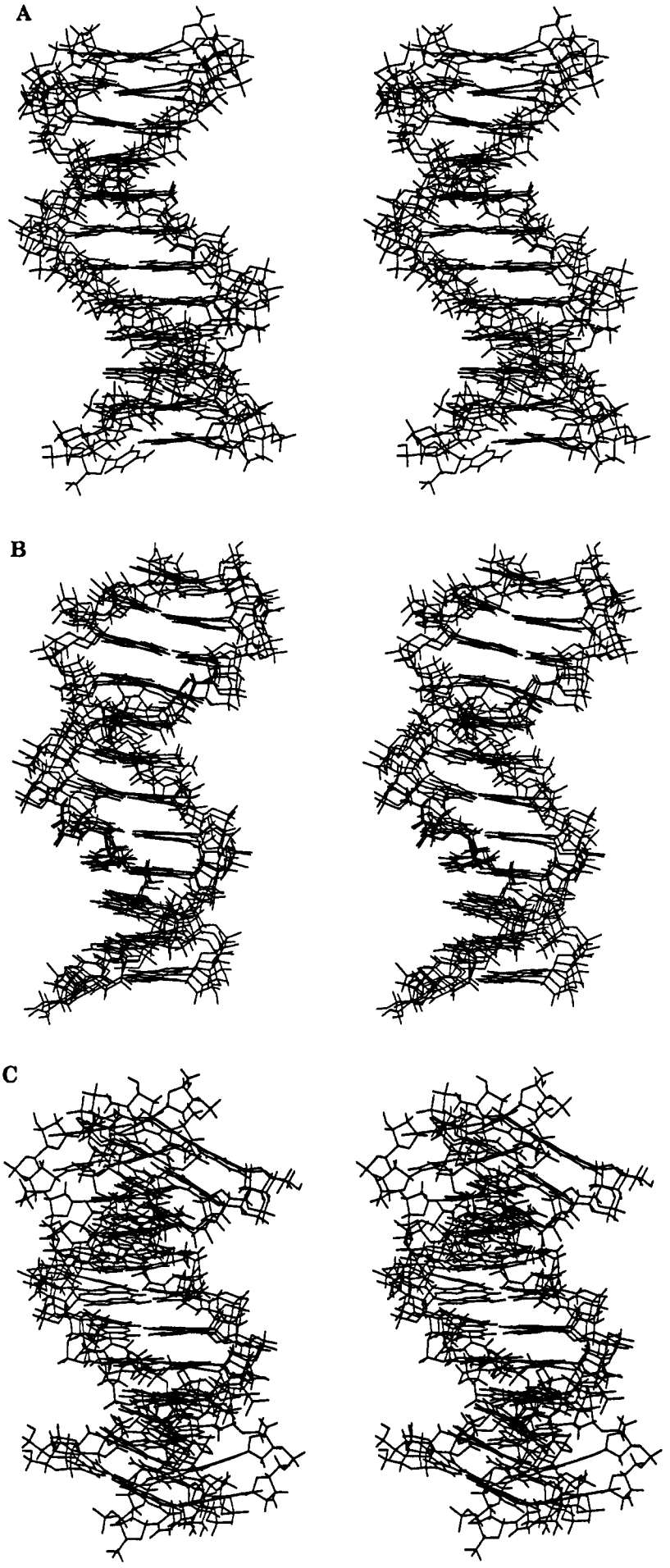
The accuracy in the hybrid matrix MORASS refinement was monitored by the $\%RMS_{vol}$ and R factor comparing the experimental and calculated NOESY volumes. The $\%RMS_{vol}$ were based on eq 1 in which the denominator contained either the theoretical or experimental volumes. Both the R factor directly analogous to the R factor used for crystallographic refinement and $\%RMS_{vol}$ decreased during the MORASS/MD iterative cycles. However, as discussed previously (Gorenstein et al., 1990; Nikonowicz et al., 1989a, 1990), we believe our $\%RMS_{vol}$ factor better represents the quality of the NMR refinement than the R factor because it equally weighs the constraint violations regardless of intensity or structural significance.

As shown above, the rms volumes and rms Cartesian x,y,z deviations for the MORASS matrix derived structures are less than those resulting from an ISPA analysis of the data. Importantly, the simulation shows that, with a moderate number of constraints (set 1), one can very closely approximate the target geometry if the starting geometry is not far removed. The same constraints, with the same degree of accuracy, are simply not enough to move the A-model geometry into the same level of refinement as the B-form geometry. However, when a larger number of constraints are obtained experimentally and the matrix method is employed, the results are very satisfying (see set 2). With this level of refinement, even the backbone torsional angles and the very subtle local geometry variations that are found within the DNA duplex may be well defined.

Typically, the A-DNA initial model refinements exhibit larger rms deviations than the B-model refinements. This is likely attributable to the greater degree of conformational space searched by the restrained MD calculations that are possible when the initial model is farther removed from the final target structure. the 1.4–2.6 Å rms deviations achieved between the error-free data restrained MD structures is comparable to the 1.5–1.8 Å rms deviations observed by Baleja et al. (1990a) in converting B- into A-DNA and the reverse using perfect data in a distance *and* torsional angle restrained MD simulation, and they are analogous to our results using only the NOE-derived distance constraints.

Although the A-DNA model did not converge quite as well as the B-DNA starting model, other criteria support the conversion of the A-DNA model into the dodecamer target structure. As shown by the stereoview overlays of the various final structures in Figure 8, all of the refined duplexes are in the B-DNA conformation. Although local differences in the structures are relatively small, their cumulative effect over the length of the helix can be quite large. 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. Each of these refined structures starting from either the A- or B-DNA model with a comparable constraint set are thus equally valid, and refinement can only be defined in terms of a family of comparable structures.

In addition, as shown in Figures 6 and 7 and supplementary Figures 2S–4S, the helical parameters such as twist, the major and minor groove widths, and the backbone torsional angles (Figures 2–5) are all generally consistent with a B-DNA geometry for all of the final refined structures. The initial values for the helix twist, roll, major groove width, and minor groove width for the A-DNA starting model were 33°, ca. 10°, 12 Å, and 11 Å, respectively. The respective values for torsional angles α, β, γ, δ, ϵ, and ζ are −50°, 175°, 41°, 79°, −149°, −78°, respectively, for an A-DNA model [Saenger (1984); from Arnott and Chandrasekaran (1982), unpublished, based upon fiber diffraction data]. The main difference is in the value of δ, which largely defines the sugar ring pucker with


A
B
C

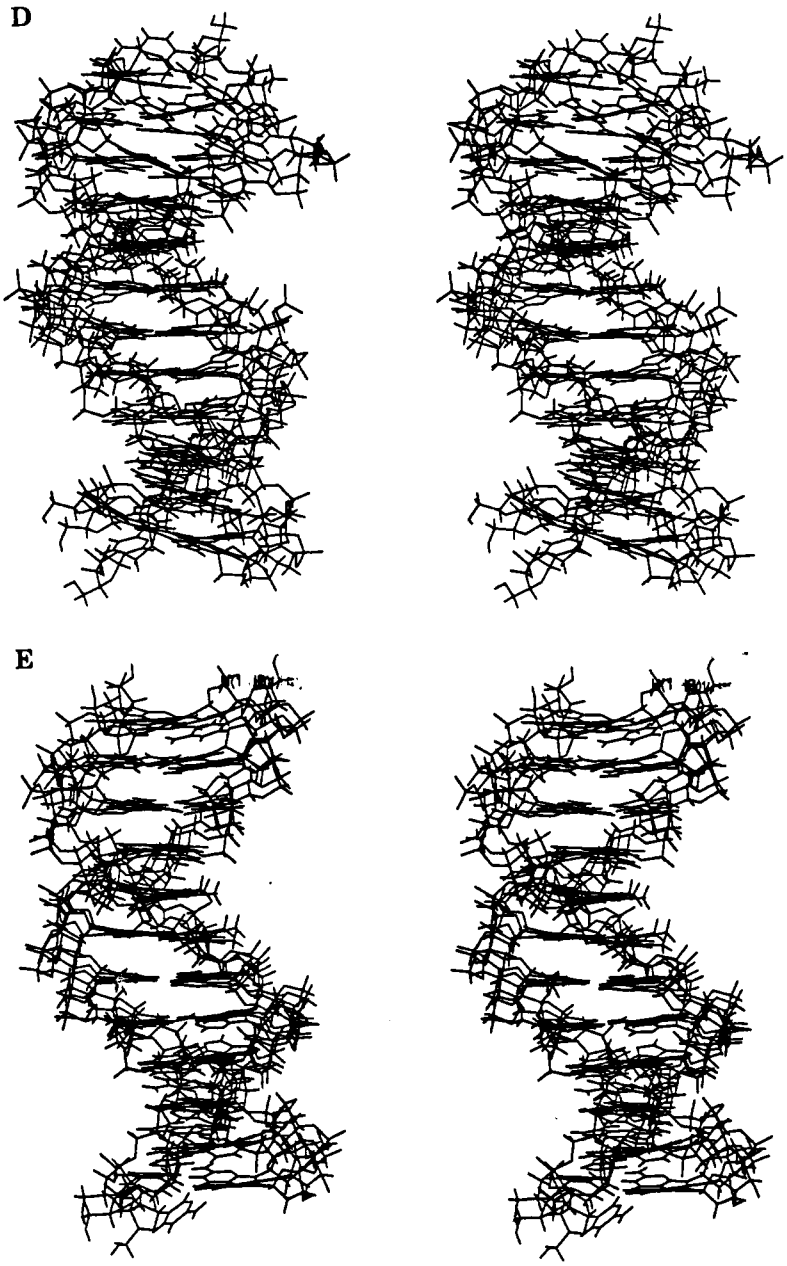


FIGURE 8: Stereoviews of the target structure as well as overlays of the final structures derived from the NOESY-distance-restrained/MORASS hybrid matrix MD refinement of the dodecamer starting from either the A- or B-DNA models. The refinement utilized either volume constraints of MORASS set 1 (A), MORASS set 2 (B), ISPA set 3 (C) and ISPA set 4 (D), or perfect distance constraint set 5 (E).

A-DNA in the C3′-endo conformation. In B-DNA the sugar puckering is generally C2′-endo or the equivalent C3′-exo. In the Dickerson dodecamer, the majority of the residues have a B-type sugar pucker (cf. Figure 2). Clearly the final refined structures have torsional angles and local helical parameters consistent with the B-DNA geometry of the Dickerson dodecamer.

Analyzing a large number of constraints obtained at long mixing times by the ISPA method yields structures that are farther from the experimental geometry. As stated earlier, this two-spin approximation underestimates the distances for longer range NOEs and overestimates the shorter ones. This systematic bias produces structures that do not satisfactorily reflect the target geometry in the observed parameters. That is to say, the distance constraints derived from the two-spin data sets are simply *wrong* and using more of them may improve the precision of the structures derived but does not increase the accuracy of the structure. Using shorter mixing times and the ISPA method does provide some improvement. However, the smaller number of only short- and medium-range distance constraints that can be derived from the poorer quality spectra are insufficient to minimize the cumulative errors that get magnified over the length of the structure. As described by Post et al. (1990), the ISPA method is quite insensitive to the integration errors in the NOESY volumes (the major source of error is the systematic bias in the lack of consideration of spin diffusion). Thus, using more accurately integrated ISPA analyzed data also will not produce more accurate structures. In contrast, the MORASS hybrid relaxation matrix method is very sensitive to the accuracy of the measured NOESY volumes (Post et al., 1990). The improvement in the precision and accuracy of the structures derived by the MORASS method would thus be even more distinctive if errors of less than 20% were achieved in analyzing the NOESY spectra.

Sequence-Specific Structural Variations. Both X-ray crystallography and NMR studies have suggested that local helical parameters vary along the DNA chain (Calladine, 1982; Dickerson & Drew, 1981; Van De Ven & Hilbers,

1988). Recent NMR studies have suggested that the duplex conformation in solution may not be identical with the static picture provided by X-ray diffraction in the crystal state (Nikonowicz et al., 1990; Sklenář et al., 1986), as one might expect. This assumes of course that the NMR data are capable of providing an accurate measure of this local structural variation.

Analysis of the local helical parameters derived from the hybrid matrix/restrained MD refined structures for the dodecamer provides support for the ability of NMR to refine these subtle structural variations. As shown by Pardi et al. (1988), distance geometry refinement of a set of simulated NMR-type distances can reproduce many of the local helical parameters with good accuracy and precision. Pardi et al. (1988) have shown that NMR-type structural refinements can define the various local helical parameters in the following order: pseudorotation angle > tilt ~ helical twist ~ δ torsion angle > roll > glycosidic angle ~ helical rise > propeller twist. As demonstrated by the data in Figures 2–7 and supplementary Figures 1S–4S, direct distance refinement by restrained molecular dynamics (set 5) as well as the MORASS/MD refinement using relaxation rate matrix derived interproton distances (set 1 and 2) produce structures whose geometries and backbone dihedral angles are quite precise and accurate. The level of accuracy for the local helical parameters follows the trend as given in Pardi (1988).

Local helical distortions originally were suggested to arise along the DNA chain due to purine–purine steric clash on opposite strands of the double helix (Calladine, 1982). As a result, 5′-PyPu-3′ sequences within the oligonucleotide represent positions where the largest helical distortions occur. Dickerson (1983) has shown that these sequence-specific variations in the conformation of duplex DNA observed in the crystal structure of a 12-mer could be quantitatively predicted through a series of simple "Calladine rule" sum function relationships (Calladine, 1982). Early crystal structure studies as well as several NMR studies initially found at least same support for these rules (Lefevre et al., 1987; Nilges et al., 1987). However, Dickerson and others have more recently noted, recent crystal and solution NMR structures have not supported the "Calladine rules" (Baleja et al., 1990b; Gorenstein et al., 1990; Heinemann & Alings, 1989; Lefevre et al., 1987; Nilges et al., 1987; Powers et al., 1989, 1990; Prive et al., 1987). There is thus considerable question over the general applicability of the Calladine-type rules or indeed whether there any unique sets of sequence-specific rules. If NMR is to be used to help define any such rules, then it is important to use a method such as the relaxation matrix/molecular dynamics methodology by which one is able to accurately reproduce the experimental geometry. Note, however, that even the two-spin approximation treatment of data (Figure 7 and supplementary Figure 3S) can still do a reasonable job of reproducing some of the sequence-specific local helical structural variations. In fact, propeller twist appears to be better defined by refinement with set 4 data than the MORASS set 1. This analysis of local helical structural variations from the NOESY distance restrained structures has thus established the ability of the NOE data to reproduce sequence-specific variations in local helical parameters even if they do not obey a simple set of the "Calladine rules".

Backbone Torsional Angle Variations from Restrained Molecular Dynamics Calculations. Unfortunately, $^1H/^1H$ 2D NOESY data give no direct information on the sugar–phosphate backbone conformation, and NOESY distance restrained structures have been suggested to be effectively disordered in this part of the structure (Pardi et al., 1988; Van De Ven & Hilbers, 1988). However, measured ^{31}P chemical shifts and $J_{H3'-P}$ coupling constants have demonstrated clearly that the backbone is not disordered but shows similar sequence specificity as the base-step geometry (Gorenstein et al., 1990; Nikonowicz & Gorenstein, 1990; Roongta et al., 1990).

Analysis of the X-ray crystal structures (Dickerson, 1983; Dickerson & Drew, 1981; Saenger, 1984) of B-form oligodeoxyribonucleotide duplexes has shown that torsional angles α, β, and γ on the 5′ side of the sugar are largely constrained to values g^- (–60°), t (180°), and g (+60°), whereas significant variations are observed in the 3′ side of the deoxyribose phosphate backbone. The greatest variation in backbone torsional angles is observed for ζ (P–O3′), followed by ϵ (C3′–O3′), and then δ (C4′–C3′). It is important to note that many of these torsional angle variations are correlated (Dickerson, 1983; Dickerson & Drew, 1981; Saenger, 1984).

Torsional angle δ correlates with the sugar-puckering conformation (as defined by the pseudorotation angle). The sugar ring constrains δ to values no smaller than 70°–85° (C3′-endo) and no larger than 140–160° (C2′-endo). Intermediate values of δ yield the other two common DNA sugar-puckering conformations ($\delta \sim 96°$, O1′ endo; and $\delta \sim 120°$, C1′-exo). At low values of ζ, the phosphate P–O3′ conformation is in the low energy g^- conformation. When the P–O3′ conformation is g^-, invariably the C–O3′ conformation (ϵ) is found to be t. This ϵ (t) ζ (g^-) conformation is the most common backbone conformation. In this B_I (t, g) conformation, δ can vary considerably (Dickerson, 1983; Dickerson & Drew, 1981). The other most common conformation for the (ϵ, ζ) pair is the (g^-, t) or B_{II} state. Rotation about the ϵ/ζ torsional angles interconverts B_I and B_{II} conformations with remarkably little overall movement of the C3′–O3′–P phosphate "elbow." It is largely this variation in δ, ϵ, and ζ that allows the sugar–phosphate backbone to "stretch" or "contract" to allow for variations in the local helical twist and base-pair displacement of B-DNA (Gorenstein et al., 1988; 1990).

As shown in Figures 2 and 3, the MORASS restrained molecular dynamics calculations are able to generally reproduce the variation in the backbone torsional angles for the target dodecamer. The fit to the target value for the δ, ϵ, and ζ torsional angles is rather accurately reproduced by the MORASS/MD refined data sets 1 and 2 and, of course, of the perfect distance constraints (set 5; Figure 5). The phosphates of residues 10, 13, and 22 are in the B_{II} conformation, while all of the others are B_I. None of the data set refinements reproduce the B_{II} conformation for phosphate 13, which might be expected since this is a terminal residue, and a similar lack of consistency in terminal phosphate geometry has been observed in the various crystal structures of the dodecamer. The other two phosphates are calculated to be in the B_{II} conformation with the MORASS refinement, while the ISPA refinement does a much poorer job. There is poorer agreement between the target α and γ torsional angles and the torsional angles derived from the refinement sets.

It is important to point out that comparisons of the backbone torsional angles at a single "snapshot" during the MD run or after averaging and minimization, fail to reproduce these sequence-specific variations. Presumably the origin of the rather good reproduction of the sequence-specific variation in the torsional angles derives from the ability of these distance-constrained MD runs to correctly reflect the time-averaged (and ensemble-averaged) population of different conformational states. A single minimized structure will reflect only one of many of these conformations with each of the torsional

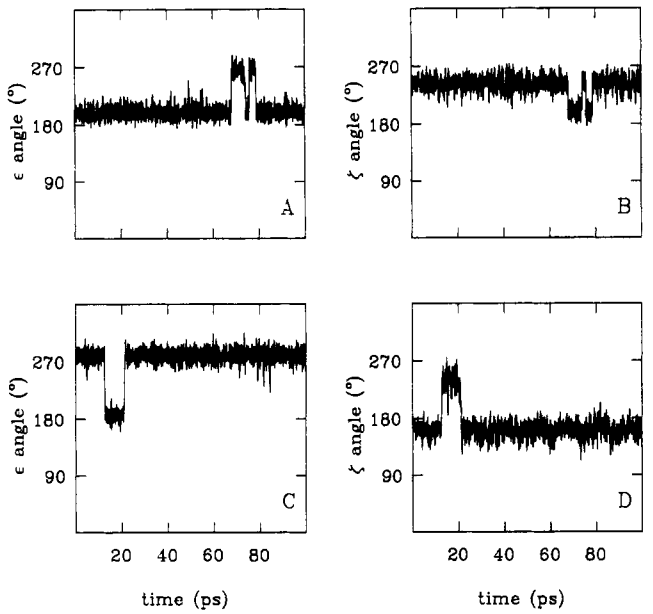


FIGURE 9: Time course for the fluctuations in the ε (A, C) and ζ (B, D) backbone torsional angles for two of the residues, 16 (top panels) and 2 (bottom panels), during 100 ps of distance-restrained molecular dynamics from constraint set 1.

angles falling into one of the local energy minima. The time dependence of the ζ and ε backbone torsional angles may be seen in the results plotted from an additional 100-ps dynamics run of several representative residues on the 3′ strand by using constraint set 1 on the B-form model (Figure 9).

With the exception of the ζ and ε angles, the other four backbone torsional angles showed relatively small amplitude fluctuations about an average B-DNA-like value during this 100-ps restrained molecular dynamics calculation (data not shown). As noted from X-ray studies (Dickerson, 1983; Dickerson & Drew, 1981) of B-DNA, the ε and ζ torsional angles show the largest variability; and indeed our calculations demonstrate that large amplitude fluctuations occur for these two torsional angles. These torsional angle changes reflect a transition from the low-energy B_I conformation to the higher energy B_{II} conformation. As shown, these transitions are short lived and typically relax back to the low-energy conformation. These calculations also provide support for the rapid conformational transition between the B_I and B_{II} conformations and the strong correlation between the ε and ζ torsion angles observed in the X-ray crystal structures (Dickerson, 1983; Dickerson & Drew, 1981) and NMR solution structures of B-DNA (Gorenstein et al., 1990; Powers & Gorenstein, 1990; Powers et al., 1990).

Origin of Sequence-Specific Variation in the Backbone Torsional Angles. Again, it must be emphasized that there are no distance or torsional angle constraints defining the phosphodiester backbone in these restrained MD calculations. Thus the conformation must be constrained by the overall local geometry of the stacked base pairs. This is in confirmation of our observation that ^{31}P chemical shifts and coupling constants follow sequence-specific variations, attributed to a strong coupling of the backbone conformation to local helical geometry (Gorenstein et al., 1988, 1990; Nikonowicz & Gorenstein, 1990; Schroeder et al., 1989).

The possible basis for the sequence-specific variation in the backbone conformation can be analyzed in terms of deoxyribose phosphate backbone distortions arising from local variations of the helical parameters and changes in the length of the deoxyribose phosphate backbone (Gorenstein et al., 1988, 1990; Schroeder et al., 1989). These local helical and backbone length changes require changes in the deoxyribose phosphate backbone angles α through ζ. Our laboratory has shown (Gorenstein et al., 1988) that when the helix winds or unwinds, the distance between the adjacent C4′ atoms of deoxyribose rings along an individual strand must change to reflect the stretching and contracting of the deoxyribose phosphate backbone between the two stacked base pairs. To a significant extent, these changes in the overall length of the deoxyribose phosphate backbone "tether" are reflected in changes in the backbone torsional angles.

Yanagi et al. (1991) point out the importance of the variations in the sugar–phosphate backbone and the limitations it imposes on local helical parameters. Indeed it is possible that the failure of the Calladine rules to reliably explain the sequence-specific local helical structural variations is at least partially attributable to the failure to properly take into account this structural feature. The maximal length of the sugar–phosphate backbone (in an all-trans conformation about each of the backbone bonds) is 7.4 Å (Yanagi et al., 1991), leading to a maximal C1′ to C1′ distance between adjacent sugars of 5.6 Å (Dickerson defines this as D_{xyz}). The mean D_{xyz} found in crystals is 5.0 Å, and thus the backbone is 90% extended. The observed range of D_{xyz} in crystals varies from 4.3 (77%) to 5.5 Å (98%). As noted by Gorenstein et al. (1988) and further by Yanagi et al. (1991) the backbone linkage (as measured by our C4′–C4′ distance or the Dickerson C1′–C1′ distance) provides a connection between various local helical parameters such as rise, twist, and roll. It again supports the importance of the sugar–phosphate conformation in helping to define or at least constrain the duplex base-pairing geometry.

Conclusion

As shown in this paper, accurate and precise distances are obtainable from the hybrid relaxation matrix methodology. We also show that starting from two different initial structures it is possible to iteratively refine the initial models to a common family of structures consistent with the NOESY-derived distances. Importantly, the use of a large number of constraints, possibly obtained via spectral deconvolution or simulations, combined with the hybrid matrix analysis remarkably increases the precision and accuracy of the structural refinement.

While there is still some disagreement over the ability of even relaxation matrix refinement procedures to reliably demonstrate sequence-specific structural variations in duplexes, there is little doubt that it is possible to reproduce some of the sequence-specific variations *both* in the local helical parameters *as well as the sugar–phosphate backbone.* It is important to recognize that structural variations in the crystal structures may arise from both "rules" as well as distortions due to crystal packing forces. Because more accurate distances (<0.2 Å) can be shown to accurately reproduce such sequence-specific structural variations, it will be important to use the NMR-determined *solution* conformation to develop more generally applicable sequence-specific structural variation rules.

Acknowledgments

We greatly appreciate the contributions of Dr. Edward Nikonowicz.

Supplementary Material Available

Four figures showing backbone dihedral angles and various helical parameters as well as one table showing analysis of intermediate structures derived from constraint set 1 starting

from the B-model (5 pages). Ordering information is given on any current masthead page.

References

Assa-Munt, N., & Kearns, D. R. (1984) *Biochemistry 23*, 791.

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

Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990b) *Biochemistry 29*, 4828–4839.

Banks, K., Hare, D., & Reid, B. (1989) *Biochemistry 28*, 6996–7010.

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.

Borgias, B. A., & James, T. L. (1988) *J. Magn. Reson. 79*, 493–512.

Borgias, B. A., Gochin, M., Kerwood, D. J., & James, T. L. (1990) *Prog. Nucl. Magn. Reson. Spectrosc. 22*, 83–100.

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

Broido, M. A., Zon, G., & James, T. L. (1984) *Biochem. Biophys. Res. Commun. 119*, 663–670.

Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343–352.

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

Clore, G. M., & Gronenborn, A. M. (1989) *J. Magn. Reson. 84*, 398–409.

Clore, G. M., Gronenborn, A. M., Brunger, A. T., & Karplus, M. (1985a) *J. Mol. Biol. 186*, 435–455.

Clore, G. M., Gronenborn, A. M., Moss, D. S., & Tickle, I. J. (1985b) *J. Mol. Biol. 185*, 219–226.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419–441.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.

Dobson, C. M., Olejniczak, E. T., Poulsen, J. F. M., & Ratcliffe, R. G. (1982) *J. Magn. Reson. 48*, 97–110.

Feigon, J., Leupin, W., Denny, W. A., & Kearns, D. R. (1983) *Biochemistry 22*, 5930–5942; 5943–5951.

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

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686–14707.

Frechet, D., Cheng, D. M., Kan, L.-S., & Ts'o, P. O. P. (1983) *Biochemistry 22*, 5194–5200.

Gronenborn, A., & Clore, G. (1989) *Biochemistry 28*, 5978–5984.

Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V. A., & Jones, C. R. (1988) *Biochemistry 27*, 7223–7237.

Gorenstein, D. G., Meadows, R. P., Metz, J. T., Nikonowicz, E., & Post, C. P. (1990) *Advances in Biophysical Chemistry* (Bush, C. A., Ed.) pp 47–124, JAI Press, Greenwich.

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.

Heinemann, U., & Alings, C. (1989) *J. Mol. Biol. 210*, 369–381.

Hosur, R. V., Govil, G., & Miles, H. T. (1988) *Magn. Reson. Chem. 26*, 927–944.

Jain, S., & Sundaralingam, M. (1989) *J. Biol. Chem. 264*, 12780–12784.

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

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 172–182.

Kearns, D. R. (1984) *Crit. Rev. Biochem. 15*, 237–290.

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

Lancelot, G., Guesnet, J. L., & Vovelle, F. (1989) *Biochemistry 28*, 7871–7878.

Lefevre, J.-F., Lane, A. N., & Jardetzky, O. (1987) *Biochemistry 26*, 5076–5090.

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

Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol. 214*, 711–736.

Nikonowicz, E. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845–8858.

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

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

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

Nikonowicz, E. P., Meadows, R. P., Fagan, P., & Gorenstein, D. G. (1991) *Biochemistry 30*, 1323–1334.

Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987) *Biochemistry 26*, 3734–3744.

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.

Pardi, A., Hare, D. R., & Wang, C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8785–8789.

Patel, D. J., & Shapiro, L. (1987) *Annu. Rev. Biophys. Biophys. Chem. 16*, 423–454.

Patel, D. J., Shapiro, L., & Hare, D. (1987a) *Biophys. Chem. 16*, 423–454.

Patel, D. J., Shapiro, L., & Hare, D. (1987b) *Q. Rev. Biophys. 20*, 35–112.

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

Powers, R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 9994–10008.

Powers, R., Olsen, R. K., & Gorenstein, D. G. (1989) *J. Biomol. Struct. Dyn. 7*, 515–556.

Powers, R., Jones, C. R., & Gorenstein, D. G. (1990) *J. Biomol. Struct. Dyn. 8*, 253–294.

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

Reid, B. R. (1987) *Q. Rev. Biophys. 20*, 1–34.

Rinkel, L. J., van der Marel, G. A., van Boom, J. H., & Altona, C. (1987) *Eur. J. Biochem. 163*, 275–286.

Roongta, V. A., Powers, R., Nikonowicz, E. P., Jones, C. R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 5245–5258.

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

Scheek, R. M., Boelens, R., Russo, N., vanBoom, J. H., & Kaptein, R. (1984) *Biochemistry 23*, 1371–1376.

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

Schroeder, S. A., Roongta, V., Fu, J. M., Jones, C. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8292–8303.

Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F., & Rabinovich, D. (1989) *Nature 342*, 456–460.

Sklenář, V., & Bax, A. (1987) *J. Am. Chem. Soc. 109*, 7525–7526.

Sklenár, V., Miyashiro, H., Zon, G., Miles, H. T., & Bax, A. (1986) *FEBS Lett. 208*, 94–98.
Van De Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Wagner, G., & Wüthrich, K. (1979) *J. Magn. Reson. 33*, 675–680.
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.
Yanagi, K., Prive, G. G., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 201–214.

Effect of Metal Ion Binding on the Secondary Structure of Bovine α-Lactalbumin As Examined by Infrared Spectroscopy

Steven J. Prestrelski,*,‡ D. Michael Byler,‡,§ and Marvin P. Thompson‡,∥

Eastern Regional Research Center, U.S. Department of Agriculture, 600 East Mermaid Lane, Philadelphia, Pennsylvania 19118, and Department of Chemistry and Physical Science, Philadelphia College of Textiles and Science, Schoolhouse Lane and Henry Avenue, Philadelphia, Pennsylvania 19144

Received February 22, 1991; Revised Manuscript Received June 6, 1991

ABSTRACT: We have examined the influence of monovalent and divalent cations on the secondary structure of bovine α-lactalbumin at neutral pH using Fourier-transform infrared spectroscopy. Our present studies are based on previously reported amide I′ component band assignments for this protein [Prestrelski, S. J., Byler, D. M., & Thompson, M. P. (1991) *Int. J. Pept. Protein Res. 37*, 508–512]. The results indicate that upon dissolution, α-lactalbumin undergoes a small, but significant, time-dependent conformational change, regardless of the ions present. Additionally, these studies provide the first quantitative measure of the well-known secondary structural change which accompanies calcium binding. Results indicate that removal of Ca^{2+} from holo α-lactalbumin results in local unfolding of the Ca^{2+}-binding loop; the spectra indicate that approximately 16% of the backbone chain changes from a rigid coordination complex to an unordered loop. We have also examined the effects of binding of several other metal ions. Our studies have revealed that binding of Mn^{2+} to apo α-lactalbumin (Ca^{2+}-free), while inducing a small, but significant, conformational change, does not cause the α-lactalbumin backbone conformation to change to that of the holo (Ca^{2+}-bound) form as characterized by infrared spectroscopy. Similar changes to those induced by Mn^{2+} are observed upon binding of Na^+ to apo α-lactalbumin, and furthermore, even at very high concentrations (0.2 M), Na^+ does not stabilize a structure similar to the holo form. Binding of Zn^{2+} to the apo form of α-lactalbumin does not result in significant backbone conformational changes, suggesting a rigid Zn^{2+}-binding site. Further, as characterized by infrared spectroscopy, binding of Zn^{2+} to holo α-lactalbumin does not induce a reversion to an apolike conformer. These results suggest that current models of the metal-binding properties of α-lactalbumin need to be carefully reexamined in light of these findings.

α-Lactalbumin is a protein present in the whey portion of the milk from most species of mammals. It functions as a modifier protein in lactose biosynthesis where it directs the specificity of the enzyme galactosyltransferase (Brodbeck et al., 1967). In addition to its physiological roles, α-lactalbumin has been demonstrated to exhibit potent antitumor activity in human mammary carcinoma cell lines (Bano et al., 1985). The structural integrity of α-lactalbumin may have important implications in these and other biological processes. Previous investigations have demonstrated a strong influence of monovalent and divalent cations on the conformation, stability, and activity of α-lactalbumin. Several studies have demonstrated that this protein has a strong Ca^{2+}-binding site whose binding constant is on the order of 10^{-8}–10^{-9} M (Berliner & Johnson, 1988). Intrinsic fluorescence and aromatic circular dichroism (CD)[1] studies have suggested that binding of Ca^{2+} to this site results in a marked conformational change (Permyakov et al., 1981a; Segawa & Sugai, 1983). Further, Ca^{2+} has been demonstrated to enhance the thermal stability of α-lactalbumin (Kuwajima et al., 1986). Recently, the atomic structure of baboon α-lactalbumin has been determined at a resolution of 1.7 Å by X-ray crystallography (Acharya et al., 1989). This study revealed an extremely well-formed, high-affinity Ca^{2+}-binding site, different from the previously observed EF-hand motif (Stuart et al., 1986).

Other investigations of α-lactalbumin have demonstrated binding of other cations to the Ca^{2+} site, including Mg^{2+}, Mn^{2+}, Na^+, and K^+ (Permyakov et al., 1981b, 1985; Murakami & Berliner, 1983; Segawa & Sugai, 1983). A second, distinct cation-binding site has been implicated for α-lactalbumin which has a specificity for Zn^{2+} (Murakami & Berliner, 1983). This study has also suggested that binding to the Ca^{2+} and Zn^{2+} sites is mutually exclusive. That is, binding of Zn^{2+} to holo α-lactalbumin causes displacement of bound Ca^{2+} and

* To whom correspondence should be addressed at Amgen Inc., Amgen Center, 1840 Dehavilland Dr., Thousand Oaks, CA 91320.
‡ U.S. Department of Agriculture.
§ Philadelphia College of Textiles and Science.
∥ Present address: 1432 Manor Lane, Bluebell, PA 19422.

[1] Abbreviations: CD, circular dichroism; EGTA, ethylene glycol bis(β-aminoethyl ether)-*N,N,N′,N′*-tetraacetic acid; IR, infrared; FTIR, Fourier-transform infrared spectroscopy; PAGE, polyacrylamide gel electrophoresis; SDS, sodium dodecyl sulfate.

0006-2960/91/0430-8797$02.50/0