Solution Structure Studies of $d(AC)_4 \cdot d(GT)_4$ via Restrained Molecular Dynamics Simulations with NMR Constraints Derived from Two-Dimensional NOE and Double-Quantum-Filtered COSY Experiments†

Miriam Gochin and Thomas L. James*

Departments of Pharmaceutical Chemistry and Radiology, University of California, San Francisco, California 94143

Received March 8, 1990; Revised Manuscript Received June 7, 1990

ABSTRACT: The structure of $d(AC)_4 \cdot d(GT)_4$ is investigated by constrained molecular dynamics simulations. The constraints include proton pair distances derived from 2D NOE intensities by using the iterative relaxation matrix analysis algorithm MARDIGRAS and sugar pucker phases and amplitudes derived from double-quantum-filtered COSY spectra. Molecular dynamics runs on simulated intensity and distance sets as well as the experimental data were carried out to determine the effects of starting structure, distance constraint derivation, energy functions, and experimental errors on the end result. It was found that structural details could not be elucidated within about 1.5-Å overall atomic deviation. This limitation is due in part to the accuracy of the experimental data but, more importantly, is attributable to the quantity of experimental constraints available and to imperfections in the force field utilized in the molecular dynamics calculations. Within the limits of the method, some structural characteristics of $d(AC)_4 \cdot d(GT)_4$ could be elucidated.

In the preceding paper (Gochin et al., 1990), a quantitative description of the double-quantum-filtered COSY (DQF-COSY) and 2D NOE spectra of $d(AC)_4 \cdot d(GT)_4$ was given. The following information about structure could be deduced from the NMR parameters alone: (1) *J*-Coupling analysis of the DQF-COSY spectra yielded an alternating sugar pucker along the purine/pyrimidine sequence, with purines (A, G) having a time-averaged sugar pucker around 180° and pyrimidines (C, T) having a sugar pucker around 144°. (2) Comparison of the NOE intensities to expected intensities for possible structures yielded the result that $d(AC)_4 \cdot d(GT)_4$ is most like a wrinkled D form of DNA (Arnott et al., 1983), a member of the B family of DNA structures. This form was originally described for poly[d(AT)] (Arnott et al., 1983).

The items to be addressed in this paper are (a) the applicability of the MARDIGRAS algorithm (Borgias & James, 1989, 1990) to obtain distances from experimental 2D NOE spectra, (b) the applicability of both distance and torsion angle constraints in restrained molecular dynamics calculations to determine molecular structure in solution, and (c) whether the NMR constraints are sufficient to enable us to define the structure explicitly. Molecular dynamics and mechanics will be used to find "the" structure that agrees with NMR data and is energetically feasible. Is there just one such structure? What is the effect of the energy potential on the structure derived? What is the effect of limited data and experimental imperfections on our ability to define the structure?

Such questions have been addressed before. Interproton distances derived from two-dimensional NOE experiments have been used in a number of studies utilizing distance geometry methods (Pardi et al., 1988; Boelens et al., 1988; Banks et al., 1989) or molecular mechanics and dynamics (Nilsson et al., 1986; Nilges et al., 1987a,b; Boelens et al., 1989; Broido et al., 1985; Pardi, 1990; Suzuki et al., 1986; Zhou et al., 1987) to calculate a DNA structure. Some of these studies (Nilsson et al., 1986; Nilges et al., 1987a,b) have produced results that converge from different starting points to essentially identical structures (RMS difference <1.0 Å), while others (Pardi et al., 1988; Boelens et al., 1989; Pardi, 1990) have shown only limited convergence (RMS difference 1.4–2.0 Å). Convergence does not necessarily imply accuracy, but obviously the accuracy of the final structure depends on the correct measurement and interpretation of the NOE intensities to give an accurate set of distances. Recently, a number of the techniques used to derive distance constraints in earlier studies have been called into question (Pardi, 1990; Reid et al., 1989; Borgias et al., 1988, 1990). The Nilsson and Nilges studies used the isolated spin-pair approximation to derive distance constraints, ignoring the effect of neighboring spins. Under this assumption, the NOE cross-relaxation rate between two spins, *i* and *j*, is given by

$$R_{ij} = R_{\mathrm{ref}}(r_{\mathrm{ref}}^6/r_{ij}^6) \tag{1}$$

i.e., is scaled according to a known fixed distance and cross-relaxation rate. This equation arises from the truncation of the true rate expression:

$$\mathbf{a}(\tau_m) = \exp(-\boldsymbol{R}\tau_m) \tag{2}$$

to the first term in $\boldsymbol{R}\tau_m$ in the Taylor series expansion, where **a** is the matrix of normalized cross-peak intensities and $\boldsymbol{R}$ is the full relaxation matrix. The accuracy of this approach depends on the use of a sufficiently short mixing time τ_m to ensure that initial cross-relaxation rates are being measured. It appears that the use of the H2′–H2″ geminal proton pair as a reference cross-peak for sugar–base and sugar–sugar NOE's in the above studies does not fit this requirement, since the extremely short H2′–H2″ distance of 1.77 Å implies that the cross-relaxation rate builds up very rapidly in the first 50 ms of NOE mixing (Borgias & James, 1988; Reid et al., 1989). In addition, even without any error in R_{ref}, ignoring the effects of spin diffusion has been shown to cause errors of 60–205% for distances in the range of 1.77–2.27 Å and 50–80% in the range of 3.60–7.15 Å (Borgias et al., 1990). It is precisely the longer distances that will allow us to get more extended information about molecular structure.

† This work was supported by National Institutes of Health Grants GM 39247, CA 27343, and RR 01695.

* Author to whom correspondence should be addressed.

0006-2960/90/0429-11172$02.50/0

A second problem to consider is the nature of the molecular motion. Studies to date on DNA oligomers have assumed effective isotropic motion. It has been shown that, for globular molecules (e.g., DNA oligomers with less than 14–18 base pairs), the error introduced by the assumption of isotropic motion is small, i.e., ≤10% (Keepers & James, 1984). Thus we may write for the relaxation matrix $\boldsymbol{R}$:

$$R_{ii} = 2(n_i - 1)(W_1^{ii} + W_2^{ii}) + \sum_{j \neq i} n_j(W_0^{ij} + 2W_1^{ij} + W_2^{ij}) + R_{1i} \quad (3a)$$

$$R_{ij} = n_i(W_2^{ij} - W_0^{ij}) \quad (3b)$$

Here n_i is the number of equivalent spins in a group such as a methyl rotor, and the zero, single, and double transition probabilities W_m^{ij} are given (for isotropic random reorientation of the molecule) by

$$W_0^{ij} = \frac{q\tau_c}{r_{ij}^6}; \quad W_1^{ij} = 1.5\frac{q\tau_c}{r_{ij}^6}\frac{1}{1 + (\omega\tau_c)^2}; \quad W_2^{ij} = 6\frac{q\tau_c}{r_{ij}^6}\frac{1}{1 + 4(\omega\tau_c)^2} \quad (4)$$

where $q = 0.1\gamma^4\hbar^2$, r_{ij} is the effective internuclear distance between protons i and j, τ_c is the correlation time, and ω is the Larmor frequency of the proton. The term R_{1i} represents external sources of relaxation such as paramagnetic impurities.

The aim of this study is to investigate how accurately we can define a structure for d(AC)$_4$·d(GT)$_4$ from limited NOE data. Distance constraints are derived with the novel method of MARDIGRAS (Borgias & James, 1990), which applies a complete relaxation matrix analysis (CORMA) to the NOE intensity, thus avoiding truncation of eq 2. This is an alternative approach to those described in the literature, and its reliability will be assessed. The previously described method of IRMA (Boelens et al., 1988, 1989) is somewhat similar, except that it does not rely on self-consistency of NOE intensities alone in determining structural constraints but involves the energy of the structures as well.

In addition, the effect of including coupling constant data on sugar puckers will be assessed. This has not been described as yet in the literature. We are dealing with the very real experimental problems of signal to noise and resolution, the latter made more difficult in this case by the fact that d(AC)$_4$·d(GT)$_4$ is non-self-conplementary and thus has twice the number of peaks as a self-complementary helix. With these real limitations on the number and quality of structural constaints, as described below, we are able to determine a structure that should be described as partially refined; the degree to which it represents the true solution structure is unknown. This is particularly complicated by the fact that the NMR constraints are time-averaged values which, if there are multiple conformers, do not necessarily represent one low-energy conformation.

Experimental Procedures

The sample preparation and NMR methods were described in the previous paper (Gochin et al., 1990), as well as a new method of numerical integration of NOE cross-peaks.

Distance Constraint Generation. Distance constraints were generated with the program MARDIGRAS (Borgias & James, 1989, 1990). On the basis of a starting structure, a 2D NOE relaxation matrix is set up, assuming isotropic motion and a correlation time τ_c = 2 ns (derived from T_1 and T_2 studies in the preceding paper). Diagonalization of the rate matrix is followed by calculation of an NOE intensity matrix (CORMA) (Borgias et al., 1987; Borgias & James, 1988). Those elements for which experimental data are available are then replaced in the intensity matrix. For example, 109 experimental values at a mixing time of 100 ms were used. The experimental values are scaled by the ratio of the sum of calculated intensities to the sum of observed intensities. The substituted NOE intensity matrix is back-transformed to a new relaxation matrix, from which a new set of distances can be derived. Certain elements of the new relaxation matrix are reset, namely, cross-peak rates corresponding to fixed distances and cross-peak rates corresponding to distances for which no observed NOE is available. Diagonal rate constants (R_{ii}) are replaced by appropriate sums based on the calculated and constrained cross-peak rates (cf. eq 3). A new cycle of intensity matrix calculation, substitution, and back-calculation is then repeated until the calculated and experimentally observed NOE intensities converge, i.e., the residual index

$$\mathbf{R}_1 = \frac{\sum|a_o - a_c|}{\sum|a_o|} \leq 0.003 \quad (5)$$

a value designed to reflect the noise level. a_o and a_c are the observed (scaled) and calculated intensities, respectively. The final distance constraints are then calculated and assigned upper and lower bounds according to a distance range calculated by

$$\Delta r_{ij} \approx -(k\tau_m)^{1/6}(\Delta a_{ij}/6a_{ij}^{7/6}) \quad (6)$$

where k is a constant that incorporates the correlation time and other physical constants and $\Delta a_{ij} = (a_{ij})_o - (a_{ij})_c$.

We might note that the accuracy of the distances obtained with MARDIGRAS increases with (a) fraction of experimental cross-peaks observed, (b) improved signal to noise, and (c) accuracy of initial model structure, but surprisingly accurate distances can still be obtained with a quite poor initial structure (Borgias & James, 1990; P. D. Thomas et al., unpublished results). The distances thus derived have not been influenced by any energetic considerations but solely reflect the simultaneous fitting of all experimental 2D NOE intensities. With distances and bounds available, they can be utilized in calculations entailing distance geometry, restrained molecular dynamics, or a hybrid approach. For the present studies, we chose to employ restrained molecular dynamics.

Restrained Molecular Dynamics. Molecular mechanics and dynamics calculations were carried out with the program AMBER, version 3.0A (Singh et al., 1986), and were done on a Cray Y-MP computer at the Pittsburgh Supercomputing Center. The force field parameters used in the calculations have been reported previously (Weiner et al., 1986). NMR distance constraints were included by allowing for a fixed upper bound for proton pair distances determined by MARDIGRAS:

$$E_{constr}^{ij} = 0, \quad \text{if } r_{ij} \leq r_{ij}^{upper}$$
$$= k_f(r - r_{ij})^2, \quad \text{if } r_{ij} > r_{ij}^{upper} \quad (7)$$

where k_f is the force constant for the interaction. Reliance on upper bound constraints is feasible because lower bounds are mostly avoided due to steric constraints (Chiche et al., 1990). All constraints were given the same value of k_f since the size of the distance range calculated by MARDIGRAS reflects the accuracy of the measurement and all protons moving beyond upper bound contraints are considered to have the same impact on the energy.

In most calculations with experimental data, dihedral angle constraints corresponding to experimentally determined sugar puckers (Gochin et al., 1990) were included, with a standard

Table I: Restrained Molecular Dynamics Calculations on a Simulated Data Set: A wD-DNA Model Structure for the Octamer $d(AC)_4 \cdot d(GT)_4$[a]

	force const [kcal/(mol·Å²)]	atomic RMS difference (Å)[b]					constraint violation[c]	NOE $\mathbf{R}_1$ value[d]	energy[e] (kcal/mol)
		wB	wD	wBsimA	wBsimB	wDsimA			
wB	10	0.00	1.98	1.79	1.87	2.29	0.29	0.29	−1344
	25			1.66	2.00	1.97			
wD	10		0.00	1.83	1.74	1.85	0.00	0.00	−1392
	25			1.55	1.55	1.57			
wBsimA	10			0.00	1.49	2.60	0.04	0.23	−1487
	25				0.81	1.44	0.04	0.20	−1410
wBsimB	10				0.00	2.31	0.05	0.27	−1490
	25					1.52	0.04	0.22	−1503
wDsimA	10					0.00	0.02	0.18	−1484
	25						0.02	0.16	−1469

[a] Structures are defined in the text. Starting structures for the calculation were either wB- or wD-DNA, with either exact distance constraints (suffix A) or MARDIGRAS-derived distances (suffix B). A comparison is made between the starting structures, final structures, and the simulated structure (wD) in terms of atomic position deviations, distance constraint and NOE cross-peak intensity violations, and energies. [b] Terminal residues are not included. [c] Defined as $(\sum_{i=1}^{N} e_i)/N$, where $e_i = (r - r_{upper})$ when $r > r_{upper}$, $e_i = 0$ when $r_{lower} < r < r_{upper}$, or $e_i = (r_{lower} - r)$ when $r_{lower} > r$. [d] Defined by eq 5 in the text. [e] Defined by restrained energy minimization of the structure.

harmonic potential and a force constant of 25 kcal/(mol·Å²).

Standard holonomic constraints for wrinkled D (wD) and wrinkled B (wB) forms of DNA (Arnott et al., 1983) were used as initial points. Standard bond lengths and angles were used to add protons. Hydrated counterions (Na^+) were added at each phosphate residue at a starting distance of 5 Å to neutralize the negative charges. All atoms up to an 18-Å radius were included in nonbonded interactions. Up to 40 ps of molecular dynamics simulations, with a time step of 0.001 ps, was carried out. Bond lengths involving hydrogen atoms were kept fixed with the SHAKE algorithm (Ryckaert et al., 1977).

The calculations proceeded by first minimizing the starting structure with low values of the constraint force constants. These force constants were then gradually increased from 1 to 10 (or 25) kcal/(mol·Å²) during the first 10 (or 12) ps of molecular dynamics calculations at 300 K. Over the next 2 ps, the molecule was heated and equilibrated at 600 K and then allowed to cool slowly to 300 K over a period of 4 ps, to avoid trapping in some local minimum. Following this annealing, the structures were again minimized by using the maximum force constant of 10 (or 25) kcal/(mol·Å²). A further 20 ps of molecular dynamics simulation at 300 K and maximum force constants was also carried out to ensure that convergence had been obtained.

Results and Discussion

First, model simulations were carried out to determine the effect of three factors on the end results, namely, (1) appropriate definition of distance constraints by MARDIGRAS from NOE intensities, (2) the effect of energy potentials in molecular dynamics on the end results, and (3) the effect of experimental error in measured intensities on the end results. Model calculations have been done before on exact and noisy distance constraint sets (Nilsson et al., 1986; Pardi et al., 1988).

How Well Does MARDIGRAS *Define Distance Ranges*? To answer this question, a simulated intensity set was generated for $d(AC)_4 \cdot d(GT)_4$ in the wD form with CORMA. This intensity set was then input into MARDIGRAS, yielding a set of upper and lower bound distance constraints. No diagonal, H5′, H5″, or methyl interactions were included. Two molecular dynamics (MD) runs starting with a wB form of $d(AC)_4 \cdot d(GT)_4$ were then carried out according to the method described under Experimental Procedures. Run A used *exact* distance constraints derived from the wD coordinate file; run B used the MARDIGRAS-derived distances. Only distances ≤4.3 Å were incorporated, to accurately mimic the NMR situation; 220 distance constraints were used. The final coordinates of the structures wBsimA and wBsimB were determined by averaging the last 5 ps of the molecular dynamics run in 0.2-ps steps. MARDIGRAS is successful if wBsimA and wBsimB are similar structures with similar distance constraint violations and NOE intensity violations. Table I shows this to indeed be the case. Note that the table gives results for two values of the maximum force constant k_f; 10 kcal/(mol·Å²) was taken from previous studies (Chiche et al., 1990), and 25 kcal/(mol·Å²) was tried to see if better convergence could be achieved. With the higher force constant, wBsimA and wBsimB converge to 0.81-Å atomic RMS difference, i.e., to essentially the same structure.

Also included in Table I is the result of a molecular dynamics run on a wD starting coordinate set, with the *exact* wD distances as constraints. Note that the effect of the energy function is to move the structure away from wD! The RMS deviation of the final structure wDsimA from wD is at least 1.6 Å. However, the distance constraint violation and NOE residual error $\mathbf{R}_1$ are only 0.02 Å and 0.16, respectively. It is noted here that an unconstrained energy-minimized wD structure, which differs from the standard wD by only 0.26 Å, has corresponding distance constraint violation and $\mathbf{R}_1$ values of 0.013 Å and 0.16, respectively. It thus appears that the overall NOE data set does not allow us to distinguish structures more closely than about 1.5-Å RMS difference. This result is similar to that found in studies by Pardi et al. (1988) and Boelens et al. (1989) but is in contrast to the results of Nilges et al. (1986). Additionally, it is noted that the RMS deviation between successive 5-ps averages in a converged molecular dynamics run is 0.3–0.5 Å. This corresponds to the level of significantly damped molecular fluctuations, the damping caused by the presence of NMR constraints. In the absence of contraints, local fluctuations are expected to be ~1 Å.

Armed with these relative values to interpret Table I, we see that wBsimA and wBsimB converge well, with approximately the same distance and NOE violations. Thus MARDIGRAS is providing us with an adequate interpretation of the NOE's. The constraints are necessarily weaker than *exact* distances, thus allowing for inherent inaccuracy and uncertainty of the NOE's; yet they allow convergence to essentially the same structure as exact distances.

What Is the Effect of Introducing Energy Potentials on the NOE Constrained Structure? This question has been addressed in the study of the effect of a molecular dynamics simulation on starting wD-DNA, using wD-DNA exact distances as constraints (vide supra). The AMBER force field moves the molecule away from the starting and constrained

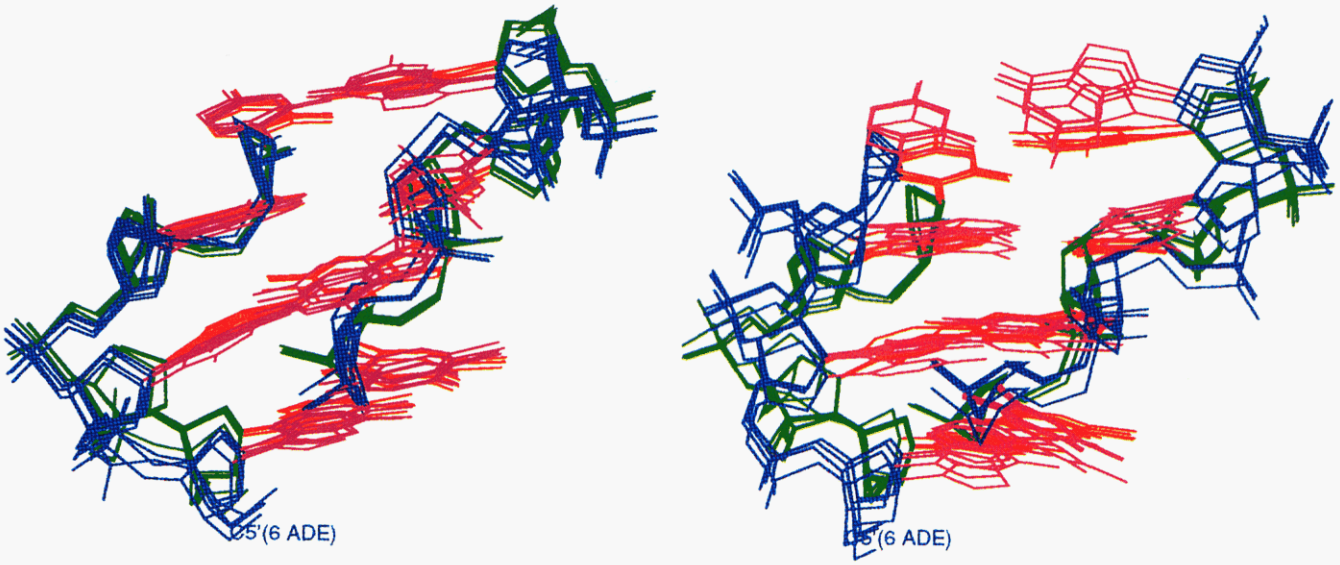

FIGURE 1: Superposition of different molecular dynamics runs on simulated data. (Left) Comparison of final structures wBsimA (green backbone, red bases) and wBsimB (blue backbone, magenta bases) (see text); (right) comparison of final structures wBsimA (green backbone, red bases) and wDsimA (blue backbone, magenta bases). The prefix wB or wD indicates the starting structure for the simulation, and the suffix indicates that either *exact* distances, A, or MARDIGRAS-derived distances, B, were used as constraints. Each structure is drawn as four samples derived from averaging over four 5-ps steps in the last 20 ps of a 40-ps MD run, in an attempt to express the flexibility of the structure within the constrained MD run. Only the four internal base pairs are shown for greater clarity. The view is looking into the minor groove.

structure. The implication clearly is that the force field utilized in the restrained molecular dynamics, especially in the absence of explicit solvent molecules, is far from perfect. In all cases, the effect of the MD simulation is encouragingly to lower the energy of the final structure compared to the starting structure. Clearly the NOE constraints and the energetic constraints both contribute to the trajectories and structures in the MD runs. Use of a higher force constant [25 kcal/(mol·Å^2) vs 10 kcal/(mol·Å^2)] for the NOE distance constraints definitely leads to improved convergence between structures. It is not considered feasible at this point to increase the force constants of NOE constraints any further, since this would apply an unnatural amount of constraint to a structure that is probably fairly flexible. As it is, NOE constraints [even at k_f = 10 kcal/(mol·Å^2)] inhibit motions such as sugar ring repuckering, which is readily observed during unconstrained dynamics. Thus, the energy potentials play an important role in determining the final structure, so a structure with simultaneously low energy, low residual distance violations, and good agreement between calculated and experimental 2D NOE spectra (low residual index) is an encouraging sign of the utility of the restrained MD approach. It is a matter of judgment to weigh the relative merits of these criteria; we would suggest that the higher force constant of 25 kcal/(mol·Å^2) be utilized only if it results in an improvement in fit to experimental data without *substantial* increase in the calculated energy.

Figure 1 shows a comparison of the final structures wBsimA, wBsimB, and wDsimA. It is obvious that all final structures clearly resemble wD-DNA (see Figure 1, previous paper). Fraying at the ends is considerable. The agreement between bases is much better than the agreement between backbone atoms. For example, the RMS difference between bases is only 0.70 Å between wBsimA and wBsimB and 1.02 Å between wBsimA and wDsimA.

What Is the Effect of Experimental Error? We now proceed to an analysis of the experimental NOE's and their errors on similar MD simulations. We are dealing not only with errors introduced by signal to noise and peak overlap in defining NOE intensities but also with a much reduced distance constraint set; only 109 useful distances could be extracted

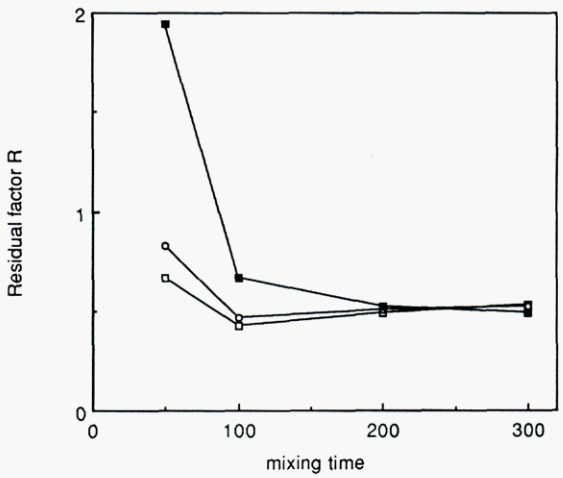

FIGURE 2: NOE residual factor $\mathbf{R}_1$ (for definition, see eq 5) vs mixing time for four 2D NOE experiments conducted at mixing times of 50, 100, 200, and 300 ms. Data are shown for three different possible structures: B- (■), wrinkled B- (○), and wrinkled D-DNA (□).

from the 2D NOE spectrum. Information about certain residues, namely C4 and C6, was missing entirely due to complete overlap of the chemical shifts of all protons of these residues, as well as all guanine H8–H2′ and H8–H2″ interactions. Information about the central part of the molecule is quite crucial because AMBER works most reliably away from the ends of the helix, where fraying is considerable.

We also included dihedral angle constraints, derived from the double-quantum-filtered COSY spectrum of d(AC)$_4$·d(GT)$_4$, from the previous paper. Time-averaged purine and pyrimidine sugar pseudorotation angles of 180° and 144°, respectively, with maximum amplitude of pucker 35° were found from analysis of the DQF-COSY spectrum. These puckers were incorporated into the constraint potential by fixing the δ and ν_i dihedral angles with a force constant of 25 kcal/(mol·Å^2). This was sufficient to constraint the sugar puckers by ±10° in the absence of other constraints.

Selection of NMR Data. Figure 2 shows the NOE residual factor $\mathbf{R}_1$ (eq 5) between experimental NOE's and NOE's

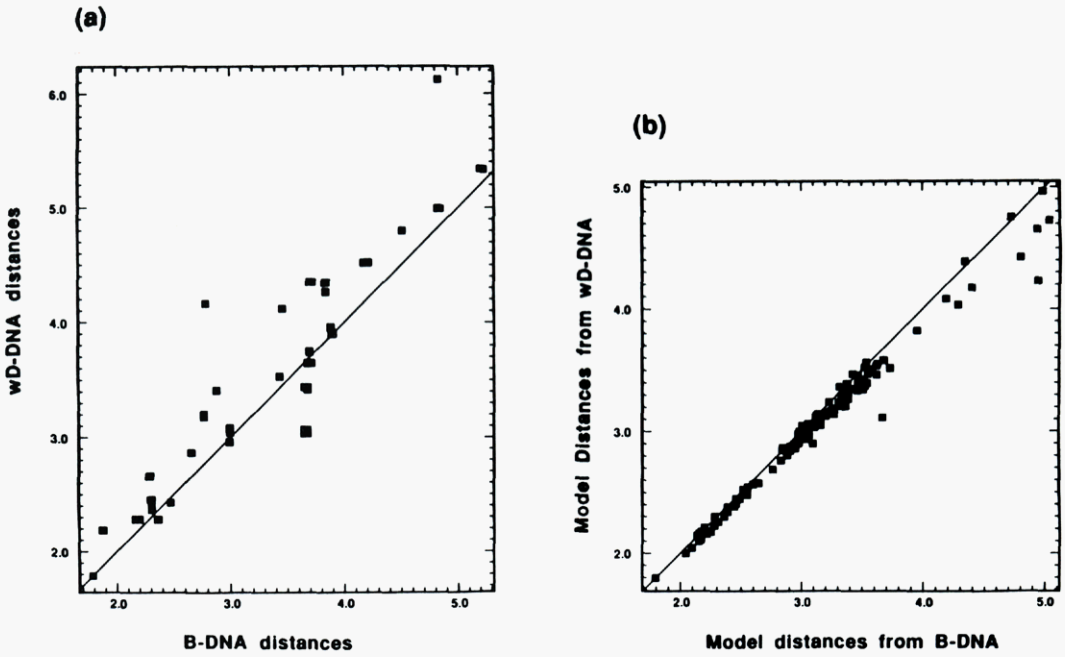

FIGURE 3: Comparison of interproton distances for $d(AC)_4 \cdot d(GT)_4$ in the B- and wD-DNA conformations (a) before and (b) after structure refinement with MARDIGRAS.

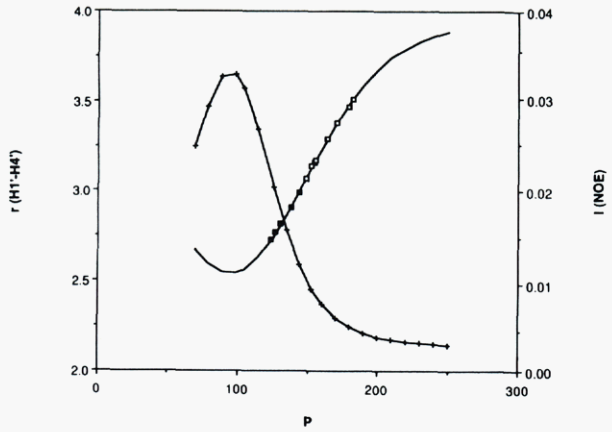

FIGURE 4: Plot of the distance between the H1′ and H4′ protons in a sugar as a function of pseudorotation angle **P**, assuming a pucker amplitude of 35° (—). Superimposed are the calculated upper bounds on the H1′–H4′ distances for the various nucleotides in $d(AC)_4 \cdot d(GT)_4$, (■) for C or T residues and (□) for G or A residues, from experimental 2D NOE intensities analyzed by using MARDIGRAS. Also shown is the corresponding intensity for the H1′–H4′ cross-peak at 100-ms mixing time, calculated from CORMA, as a function of **P** (+; arbitrary scale).

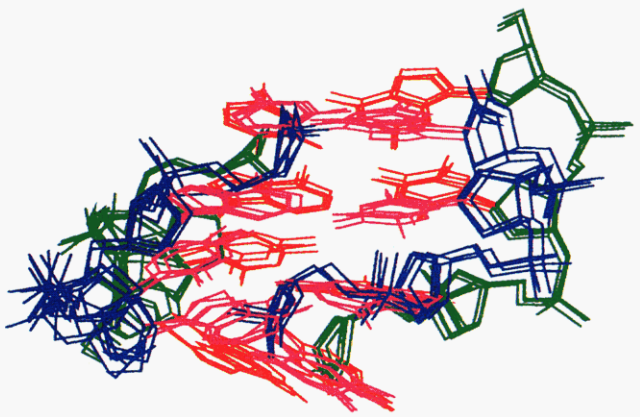

FIGURE 5: Best fit superposition of the final structures wBexp and wDexp from MD runs on experimental data. As in Figure 1, each structure is drawn as four time-averaged samples derived from averaging consecutive 5-ps MD steps. The four internal base pairs are shown with the same view as in Figure 1. For clarity, the wDexp structures have the backbone drawn in green and the bases in red, while the wBexp structures have the backbone drawn in blue and the bases in magenta.

calculated for three structures as a function of mixing time. From this figure, wD again appears to be the best initial approximation to the structure, but the other wrinkled form, wB, is not too bad either. Note also that the data at lower mixing times show greater sensitivity to structural variations, since the equalizing effect of spin diffusion is reduced. Since many cross-peaks were too small to be accurately measured at a 50-ms mixing time, the 100-ms data were used in subsequent MD simulations.

MARDIGRAS was used to obtain a set of NOE distance constraints to be used in molecular mechanics and dynamics calculations. The relaxation matrix approach has the advantage that a complete spin environment (albeit from a model) is included in distance calculations, even though sparse information is available from the experiment. Fortunately, MARDIGRAS is extremely insensitive to the starting model (Borgias & James, 1989; Borgias et al., 1990). Figure 3 shows the excellent convergence of the distance sets obtained by MARDIGRAS on the 100-ms NOE intensities, starting with B- and with wD-DNA. Five cycles of MARDIGRAS were necessary for convergence in each case.

Figure 4 shows a curve of the sugar proton H1′–H4′ distance as a function of pseudorotation angle **P**, assuming $\phi_{max} = 35°$. The corresponding "mirror-image" NOE curve was calculated by using CORMA. Superimposed on the distance curve are the H1′–H4′ distances obtained by MARDIGRAS analysis of the experimental 2D NOE intensities. It is apparent that the purine and pyrimidine H1′–H4′ distances are quite segregated and reflect a purine **P** of 150–180° and a pyrimidine **P** of 125–155°. These values are precisely in the range determined by *J*-coupling analysis of the DQF-COSY spectra, and we have by now overwhelming evidence of the alternating nature of $d(AC)_4 \cdot d(GT)_4$. We are also certain of the good match between DQF-COSY and 2D NOE data. However, we are still unable to unequivocally distinguish between single-confor-

Table II: Restrained Molecular Dynamics Calculations Using either Wrinkled B-DNA or Wrinkled D-DNA as Starting Structures with MARDIGRAS-Derived Distances from Experimental 2D NOE Data on $d(AC)_4 \cdot d(GT)_4$[a]

	force const [kcal/(mol·Å²)]	atomic RMS difference (Å)[b]					NOE R_1	
		wB	wD	wBexp	wDexp	constraint violation[c]	value	energy[d] (kcal/mol)
wB	10	0.00	1.98	1.81	2.19	0.45	0.51	−1254
	25			1.73	2.48			
wD	10		0.00	1.30	1.72	0.34	0.37	−1354
	25			2.06	1.93			
wBexp	10			0.00	1.80	0.16	0.35	−1345
	25				2.53	0.11	0.29	−1311
wDexp	10				0.00	0.15	0.35	−1459
	25					0.10	0.32	−1422

[a] Comparison is made among the starting structures, wrinkled B-DNA (wB) or wrinkled D-DNA (wD), and the structures resulting from 40 ps of restrained molecular dynamics, wBexp and wDexp, in terms of atomic position deviations, distance constraint and NOE cross-peak intensity violations, and energies. [b] Terminal residues not included. [c] Defined as in Table I. [d] Defined by restrained energy minimization of the structure.

Table III: Consistent Helical Parameters Observed for the Final Structures Resulting from Restrained Molecular Dynamics (See Text) Compared to Standard B-, wB-, and wD-DNA[a]

parameter	B	wB	wD	wBsimA	wDsimA	wBexp	wDexp
P (deg)							
pyrimidine	191	161	152	*	*	140 ± 5	136 ± 7
purine	191	178	178	*	*	175 ± 3	171 ± 3
δ (deg)							
pyrimidine	156	143	142	129 ± 11	*	127 ± 3	126 ± 8
purine	156	148	154	141 ± 9	*	143 ± 5	138 ± 5
χ (deg)							
pyrimidine	262	263	257	245 ± 5	227 ± 5	243 ± 13	230 ± 24
purine	262	287	258	253 ± 10	254 ± 8	254 ± 8	261 ± 17
residues per turn	10.0	10.1	8.3	9.1	9.6	9.7	9.6
twist (deg)							
pyrimidine	34 (C); 41 (T)	28	42	37 ± 7	37 ± 2	37 ± 7	26 ± 5
purine	38 (A); 31 (G)	42	43	37 ± 8	36 ± 7	*	*
tilt (deg)							
pyrimidine	−1.3	3.5	9.0	−3 ± 5	−2 ± 3	0 ± 5	1 ± 2
purine	−1.2	5.4	9.0	*	8 ± 2	14 ± 3	7 ± 2
P–P distance[b] (Å)	13.2	12.8	10.9	9.1	11.9	9.1	10.4

[a] An asterisk indicates that no consensus value was found. [b] Corresponds to the average of the three shortest P–P distances across the minor groove in the central part of the molecule.

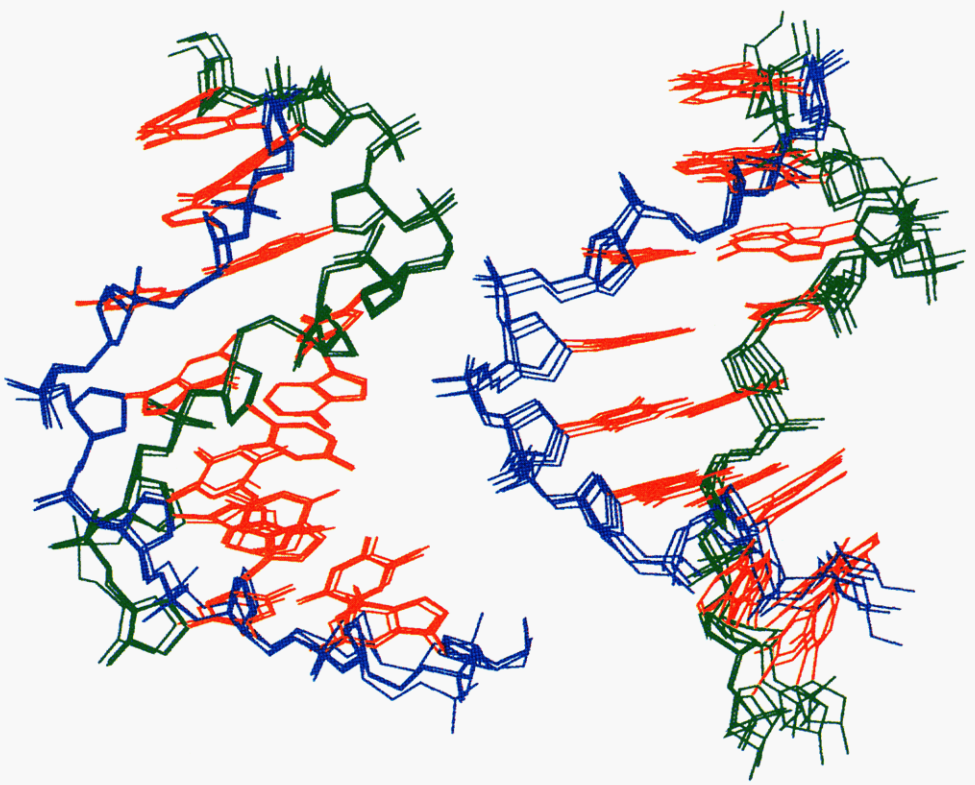

FIGURE 6: Result of MD simulations using NOE distance constraints only (excluding sugar pucker constraints): left, starting with wD; right, starting with wB. The AC backbone is drawn in green, the GT backbone is drawn in blue, and the bases are in red.

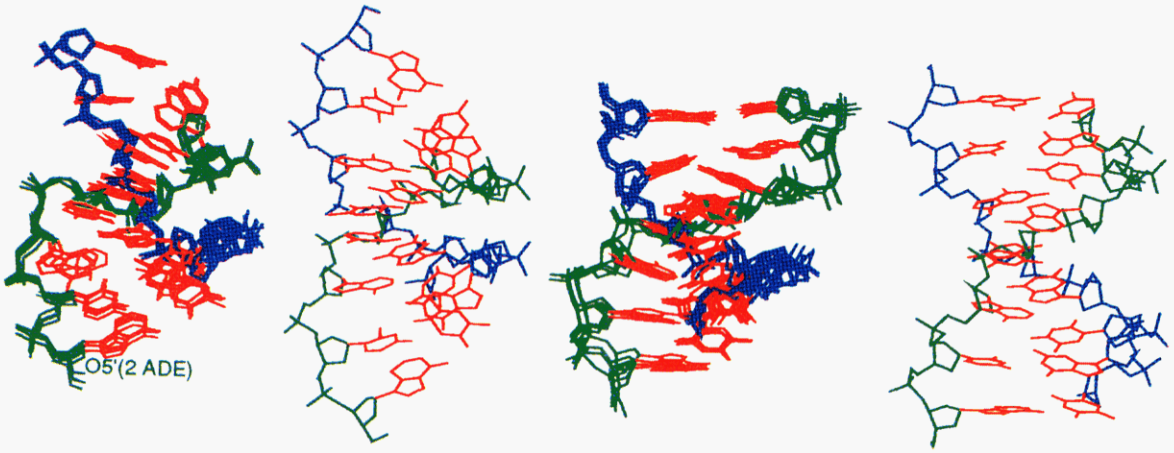

FIGURE 7: View into the narrow minor groove of the octamer structures (from left to right): wDexp, wD, wBexp, and wB. Again, each MD-derived structure is drawn as four time-averaged samples over successive 5-ps steps (see caption to Figure 1). The AC backbone is drawn in green, the GT backbone is drawn in blue, and the bases are in red.

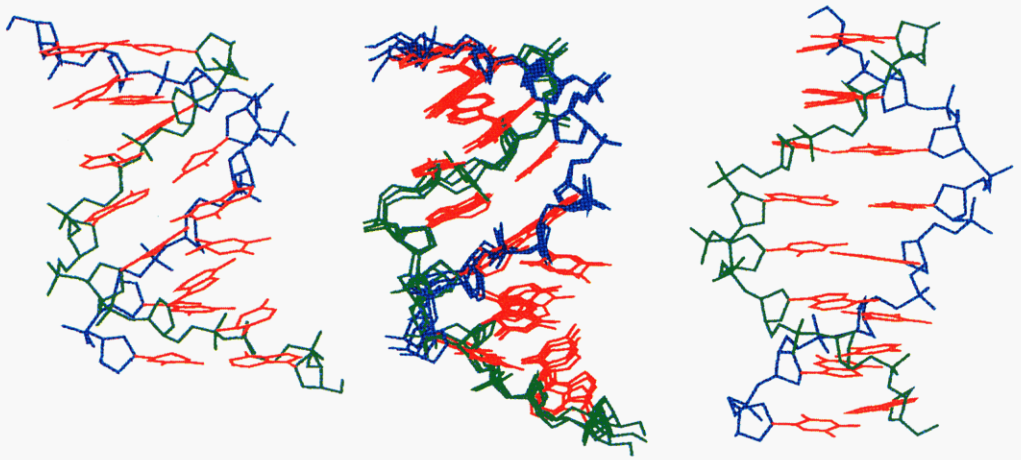

FIGURE 8: The structure of wDexp (center) compared to the structures of wD- (left) and wB-DNA (right). The central four residues of the wDexp structure are drawn as four different structures averaged over successive 5-ps steps in a 40-ps converged MD run. The AC backbone is drawn in green, the GT backbone is drawn in blue, and the bases are in red.

mational and multiconformational average states. The distance between the H1′ and H4′ protons in the N conformer (**P** = 9°) is about the same as for the S conformer (**P** = 180°), implying that a MARDIGRAS distance of 3.5 Å obtained for an H1′–H4′ distance could equally well represent a single S conformer (**P** = 180) or a mixture of N (**P** = 9°) and S (**P** = 180°) conformers in any proportions. At this point, we are also unable to specify absolutely whether our structure more closely approximates a wD- or wB-like model.

Restrained MD runs, with both distance and torsion angle constraints extracted from the experimental 2D NMR data, were carried out as for the simulated data, with the final structures taken from averaging the last 5 ps of a 40-ps run. The starting structures used were wB and wD, and the final structures are wBexp and wDexp. Table II shows the RMS deviation between structures, the energy of minimized structures, the distance constraint violation, and the NOE residual factor $\mathbf{R}_1$. The two final structures agree well with the experimental 2D NOE data, both in constraint violation and in $\mathbf{R}_1$ value; however, the RMS deviation between them is 1.80 Å. It thus appears that the limited number of NMR constraints used in the present study is insufficient to define the structure more accurately. Interestingly, the RMS deviation between wDexp and wBexp increases when the NOE force constant is increased from 10 to 25 kcal/(mol·Å^2). This is in contrast to the results presented in Table I for simulated data. Of course, use of a different force constant results in a different trajectory. A contributing factor leading to this discrepancy may be due to the simulated data representing a single conformer. However, the real experimental data may represent a time-averaged constraint that is not reflective of any potential energy minimum. Use of a different value for k_f can alter the initial trajectory and can lead to a different local minimum being found. The molecular dynamics trajectory plays a key role in the end result, since it is fairly easy for a molecule to fall into a local minimum that does not represent the true structure. At the level of NMR constraining potential used here, the resulting structures do depend on starting conformation to some extent, in that the NMR constraints fail to force the molecule into a global minimum within the time period of the 40-ps MD run. Figure 5 shows a comparison of the two final structures.

Figure 6 shows the effect of not including the sugar dihedral angle constraints in the potential function. On a starting wD structure, the effect is minimal, giving rise to a structure that has an atomic RMS difference of 1.08 Å compared with wDexp. However, the effect on a starting wB structure is dramatic, giving rise to a final structure that differs by more than 3 Å from any of the structures in Table II. This is attributed to the fact that wD more closely resembles the actual structure, and the absence of the defining central residues C4 and C6 precludes convergence from wB. Indeed, the

GT chain, obtained both with and without sugar pucker constraints, conforms fairly well in all structures, but the AC chain fluctuates significantly.

Some Details of the Structure of $d(AC)_4 \cdot d(GT)_4$. It is apparent that the structure of $d(AC)_4 \cdot d(GT)_4$ that can be derived from the above studies is not extremely specific. Missing NOE's and even their inherent level of accuracy and information content fail to allow us to determine an exact structure, down to values of dihedral angles and helical parameters. Table III lists those parameters for which consistency was seen in the final structures, i.e., for which particular values could be seen for all bases or base pairs throughout the sequence. Helix parameters were calculated by using a program (BASE PAIR) written by von Kitzing (1986). Clearly, the effect of constraining sugar puckers according to DQF-COSY results is to define well the alternating sugar pucker and δ angles for the alternating purine–pyrimidine sequence. Also, the glycosidic angles χ are fairly well-defined, probably a result of the fact that the base–sugar NOE's available give accurate information about χ. No other backbone dihedral angles showed consistency for this alternating sequence. Base twist and tilt are given, but propeller twist could not be well-defined.

The overriding feature of $d(AC)_4 \cdot d(GT)_4$ is the extremely narrow and deep minor groove. It appears in all final structures regardless of the starting structure. Table III also gives the cross-strand P–P distances across the minor groove. Such a structure allows the "spine of hydration" to form a hydration tunnel in the minor groove. Water molecules can be easily accommodated in the minor groove betwen opposing strands (Suzuki et al., 1986; Zhou et al., 1987), but any particular water molecule may reside in the minor groove for a very short time. Figure 7 illustrates the minor groove for different structures.

The best structure is considered to be that which, in addition to having a low 2D NOE $\mathbf{R}_1$ value and low distance violations, is lowest in energy after annealing, i.e., wDexp. Figure 8 compares this final structure to wB and wD. Note that although this structure seems to agree quite well with wD in overall features it exhibits >9 base pairs per turn, unlike idealized wD that characteristically has 8.3 base pairs per turn.

Conclusions

It appears that we cannot define this octamer structure very precisely, i.e., with exact helical parameters and dihedral angles. There are a number of important factors that contribute to this situation. Perhaps the most important part of the problem in this case is missing data. This can be very significant for defining interresidue contacts, which are most sensitive to structural changes (see Figure 10 in the previous paper). Excluding H5′ and H5″ interactions, the purine-to-pyrimidine step A–C or G–T can be represented by 13 NOE contacts below 4.7 Å. Many of these are intersugar interactions, which are not seen in the experimental data set due to the broadness of these resonances. Others, in the case of the G–T step, are interactions with the methyl group, which were excluded because of uncertainty in the representation of methyl group rotation. Thus, only two to four NOE contacts of this type were used. The pyrimidine-to-purine step C–A or T–G would have only five NOE contacts below 4.7 Å for a wrinkled D-like structure. Again, only two to three NOE's were experimentally available. The pyrimidine-to-purine step is an extremely important indicator of the specifics of a wrinkled structure. Absence of these defining constraints may easily account for the failure of convergence of MD simulations from the "wrong" structure.

Another part of the problem is the complex series of steps by which an intensity set is translated into a structure. Distances are somehow generated for feeding into a molecular dynamics or distance geometry program. The isolated spin-pair approximation has been used in most studies described in the literature to date, although it is now considered to be rather inaccurate. Iterative relaxation matrix calculation using MARDIGRAS is a big improvement, since it takes into account all relaxation pathways while still retaining accuracy without being highly dependent on model structure (Borgias & James, 1989; Thomas et al., 1990). But there are still unproved assumptions about molecular and internal motions that must currently be utilized. Using these distances in an MD calculation adds a whole new dimension of complexity, viz., the role that energy terms in the force field play in determining the final structure. We have seen that this is not at all insignificant, since it moves a constrained model structure away from its $E_{constr} = 0$ position.

It is insufficient to claim a solved structure by comparison of final vs starting distances. It is rather necessary to compare intensity sets (Banks et al., 1989; Pardi, 1990). Structures with a 1.4-Å RMS difference have been found to still agree equivalently with the original intensity data. Similarly in this work, a pair of structures differing by only 0.2 Å shows a similar overall deviation from the intensities as a pair of structures differing by 1.5 Å.

Another question that arises is whether one structure can suitably fit all of the data in a real experiment if the data manifest conformational averaging. We see that we can generate several structures that fit the data equally well but that none of these fits are perfect. If there was better convergence of the structures resulting from the various rMD calculations and better agreement of the resulting structures with experimental data, one might tentatively conclude that the DNA fragment has only a single structure in solution. However, the likelihood of multiple conformations must be recognized; this would affect the values of time-averaged constraints derived from 2D NMR experiments, which in turn would affect rMD calculations in that the time-averaged distances could correspond to no physically feasible structure. An important future direction is to address the issue of multiple conformers.

Currently, we observe significant limitations of the NOE method for determining very specific DNA parameters, although it can reliably be used to gauge overall structure, and can easily distinguish between different DNA families, i.e., A- or B-DNA. Within a family, NOE can also be used to identify deviations from the canonical structure, such as the wrinkling of alternating sequence DNA and (as in this work) even the type of wrinkling that is occurring. We anticipate that additional insights into the structural parameters will result from additional experimental data; experiments such as homonuclear DQF-COSY (this work) and possibly ^{31}P and ^{13}C heteronuclear scalar coupling, NOE, and relaxation measurements do provide further information.

Acknowledgments

We gratefully acknowledge the use of the University of California, San Francisco, Computer Graphics Laboratory (supported by NIH Grant RR 01081). We also gratefully acknowledge the use of the Cray Y-MP supercomputer that was supported by a grant from the Pittsburgh Supercomputing Center through the NIH Division of Research Resources Cooperative Agreement U41RR04154 and a grant from the

National Science Foundation Cooperative Agreement ASC-8500650.

Registry No. $d(AC)_4 \cdot d(GT)_4$, 81609-64-5.

REFERENCES

Arnott, S., Chandrasekaran, R., Puigjaner, L. C., Walker, J. K., Hall, I. H., & Birdsall, D. L. (1983) *Nucleic Acids Res. 11*, 1457.

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

Behling, R. W., Rao, S. N., Kollman, P., & Kearns, D. R. (1987) *Biochemistry 26*, 4674.

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

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

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

Borgias, B. A., & James, T. L. (1989) *Methods Enzymol. 176*, 169.

Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson. 87*, 475.

Borgias, B. A., Thomas, P. D., & James, T. L. (1987, 1989) COmplete Relaxation Matrix Analysis (CORMA), University of California, San Francisco.

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

Broido, M. S., James, T. L., Zon, G., & Keepers, J. W. (1985) *Eur. J. Biochem. 150*, 117.

Chiche, L., Gaboriaud, C., Heitz, A., Momon, J.-P., Castro, B., & Kollman, P. (1990) *Proteins (3rd Ed.)* (in press).

Gochin, M., Zon, G., & James, T. L. *Biochemistry* (preceding paper in this issue).

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

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1987a) *Biochemistry 26*, 3718.

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

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

Pardi, A. (1990) Symposium on Structural Biology, University of California, Berkeley.

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

Reid, B. R., Banks, K., Flynn, P., & Nerdal, W. (1989) *Biochemistry 28*, 10001.

Ryckaert, J. P., Cicotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327.

Singh, U. C., Weiner, P. K., Caldwell, J., & Kollman, P. A. (1986) AMBER 3.0: A Molecular Mechanics and Dynamics Program, University of California, San Francisco; Seibel, G. L. (1989) A Revision of AMBER, Revision A.

Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854.

Thomas, P. D., Basus, V. J., & James, T. L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

von Kitzing, E. (1986) BASE PAIR, Max Planck Institute for Biophysical Chemistry, Am Fassberg, 3400 Goettingen, FRG.

Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230.

Zhou, N., Bianucci, A. M., Pattabiraman, N., & James, T. L. (1987) *Biochemistry 26*, 7905.

Binding of *p*-Nitrophenyl α-D-Galactopyranoside to *lac* Permease of *Escherichia coli*

Julius S. Lolkema‡ and Dieter Walz*,§

Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110

Received June 4, 1990; Revised Manuscript Received August 23, 1990

ABSTRACT: Binding of the substrate analogue *p*-nitrophenyl α-D-galactopyranoside (NPG) to *lac* permease of *Escherichia coli* in different membrane preparations was investigated. Binding was assayed with an improved version of the centrifugation technique introduced by Kennedy et al. [Kennedy, E. P., Rumley, M. V., Armstrong, J. B. (1974) *J. Biol. Chem. 249*, 33–37]. Two binding sites for NPG were found with dissociation constants of about 16 μM and 1.6 mM at pH 7.5 and room temperature. With purified *lac* permease reconstituted into proteoliposomes, it could be shown that one permease molecule binds two substrate molecules. Oxidation of *lac* permease with the lipophilic quinone plumbagin or alkylation with the sulfhydryl reagent *N*-ethylmaleimide caused a 12-fold increase in the first dissociation constant. The second dissociation constant seemed to be increased as well, but its value could not reliably be estimated. Ethoxyformylation of *lac* permease with diethyl pyrocarbonate totally abolished NPG binding. The implications of these results for the catalytic performance of the enzyme are discussed.

lac permease is responsible for the transport of lactose across the cytoplasmic membrane of *Escherichia coli*. By coupling the translocation of lactose and proton, accumulation of lactose in the cell against a concentration difference is achieved at the expense of free energy stored in the electrochemical potential difference for protons. Mechanistic studies on the carrier have been carried out on *E. coli* membrane vesicles containing *lac* permease and other components in the active state and on the enzyme purified to homogeneity and reconstituted into proteoliposomes [for a review, see Kaback (1986)]. These systems enabled detailed investigations on the kinetic

* Correspondence should be addressed to this author.

‡ Present address: Department of Physical Chemistry, State University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands.

§ Permanent address: Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

0006-2960/90/0429-11180$02.50/0