Articles

Structural Model for an Oligonucleotide Containing a Bulged Guanosine by NMR and Energy Minimization†

Sarah A. Woodson and Donald M. Crothers*
Department of Chemistry, Yale University, New Haven, Connecticut 06511
Received August 3, 1987; Revised Manuscript Received December 1, 1987

ABSTRACT: We present three-dimensional structural models for a DNA oligomer containing a bulged guanosine based on proton NMR data and energy minimization computations. The nonexchangeable proton resonances of the duplex $^{5'}$d(GATGGGCAG)·d(CTGCGCCATC) are assigned by nuclear Overhauser effect spectroscopy (NOESY) and correlated spectroscopy connectivities, and the NMR spectrum is compared with that of a regular 8-mer of similar sequence, $^{5'}$d(GATGGCAG)·d(CTGCCATC). Experimental proton–proton distances are obtained from NOESY spectra acquired with mixing times of 100, 150, and 200 ms. A refined three-dimensional structure for the bulge-containing duplex is calculated from regular B DNA starting coordinates by using the AMBER molecular mechanics program [Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765–784]. We compare structures obtained by building the helix in three and four base pair increments with structures obtained by direct minimization of the entire nine base sequence, with and without experimental distance constraints. The general features of all the calculated structures are very similar. The helix is of the B family, with the extra guanine stacked into the helix, and the helix axis is bent by 18–23°, in agreement with gel mobility data for bulge-containing sequences [Rice, J. A. (1987) Ph.D. Thesis, Yale University].

In recent years, a large amount of information on the structure of DNA helices containing mispaired, unpaired, or modified bases has become available. The double helix appears to be remarkably able to accommodate these abnormalities, and this will no doubt have important implications for our understanding of mutation recognition and repair. In particular, we are interested in the structure of helices containing extra or bulged nucleotides, which are proposed to be intermediates in frame-shift mutagenesis (Streisinger et al., 1966). A series of bulge-containing oligomers based on a frame-shift hot-spot sequence in the λ C_I gene $^{5'}$d(GATGGGGCAG) (Skopek & Hutchinson, 1984) have been compared by one-dimensional proton NMR[1] (Woodson & Crothers, 1987). We have also attempted to model the three-dimensional structure of this sequence containing an extra adenine using molecular mechanics and NOE distance information (Woodson, 1987). Here we present the NMR data of an oligomer containing an extra guanine, $^{5'}$d(GATGGGCAG)·d(CTGC*G*CCATC), and describe energy-minimized structures computed for this molecule by various approaches.

A number of oligonucleotides have been examined by two-dimensional proton NMR, including sequences with G·T wobble pairs (Hare et al., 1986a), modified bases (Fazakarley et al., 1985; Patel et al., 1985), and extrahelical bases (Hare et al., 1986b). The methods for sequential assignment of the nonexchangeable proton resonances of DNA oligomers have been described by a number of groups (Hare et al., 1983; Scheek et al., 1984; Weiss et al., 1984; Chazin et al., 1986). While qualitative comparison of NMR spectra can yield useful information about the structure of the molecule, it is desirable to translate the NMR data into a three-dimensional picture in a more quantitative way. Two general approaches to structure determination have been applied to small DNA duplexes. Distance geometry methods (Havel et al., 1979) compute a three-dimensional structure based on correct bond lengths and angles for the molecule that best satisfies a set of distance constraints obtained from the NMR data. This approach has been applied to several DNA oligomers, including one containing an extrahelical adenosine (Hare et al., 1986a,b). Clore and Gronenborn (Clore et al., 1985a,b; Nilsson et al., 1986) have used energy minimization and molecular dynamics with experimental distance constraints included in the total force field to refine X-ray coordinates of A- and B-form helices. Mitomycin and psoralen DNA adducts as well as thymine dimers have been modeled by simple molecular mechanics (Rao et al., 1984, 1986; Pearlman et al., 1985).

Our goal is to model the perturbation of regular DNA conformation caused by a specific lesion, in this case a bulged nucleotide. Our strategy has been to refine standard B DNA coordinates for the desired sequence by using the AMBER molecular mechanics program (Weiner et al., 1984) with distance constraints from the NMR data. Starting from the B DNA geometry indicated by the NOESY connectivities, we ask that the final structures satisfy the distances reflected in the NOESY spectrum as nearly as possible, while maintaining correct molecular geometry and minimizing the interactions of the theoretical force field. This results in a family of

†Supported by Grant GM21966 from the National Institutes of Health.
*Address correspondence to this author.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; FID, free induction decay; EDTA, ethylenediaminetetraacetic acid; TSP, sodium 3-(trimethylsilyl)propionate-*2,2,3,3-d*$_4$; HPLC, high-performance liquid chromatography; NOESY, nuclear Overhauser effect spectroscopy; COSY, correlated spectroscopy; rms, root mean square.

0006-2960/88/0427-3130$01.50/0

structures whose general features agree with the NMR data and are reasonable in terms of what is known about DNA structure and which could serve as a starting point for further refinement by more powerful computational techniques such as restrained molecular dynamics.

In the present work, we extend our previous efforts for a duplex containing a bulged adenosine, 5′d-(GATGGGCAG)·d(CTGACCCATC) (Woodson, 1987), to a duplex containing a bulged guanosine. One of the difficulties with refining a structure of this size is the inefficiency with which large changes are made in the coordinates to find the global minimum. In order to overcome this problem, we first minimized a bulge-containing "core" consisting of a trinucleotide strand paired with a dinucleotide. Sections of regular B DNA were then added on to this core helix to build the complete structure, with minimization of the coordinates at each stage. This approach is similar to that of Kim and co-workers (Pearlman et al., 1985). The series of structures generated in this manner were very similar to each other, and surprisingly similar to a structure calculated in one piece, following the procedure used to refine the A-bulge 9-mer (Woodson, 1987). This gives us added confidence in the ability of the experimental contraints in conjunction with the molecular mechanics force field to determine the general features of the final structure, independent of the path of the refinement. Furthermore, this approach results in structures that appear reasonable in terms of the current understanding of DNA structure.

Materials and Methods

The oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer in 10-μmol scale and purified by reversed-phase chromatography on a Vydac C_4 peptide column. NMR samples were prepared as described previously (Woodson & Crothers, 1987). The final concentration of duplex was 3.0 mM in 10 mM sodium phosphate, 83.9 mM NaCl, and 0.1 mM EDTA, pD 7.0, in 99.98% D_2O.

Two-dimensional NMR spectra were recorded on the Yale 490 spectrometer at 288 K with quadrature detection in both dimensions. A total of 256 t_1 values with 2048 data points (Bruker) in t_2 were collected, resulting in 1024 points in both dimensions following transformation. A total of 128 transients were acquired per t_1 experiment. Phase-sensitive NOESY spectra (Macura & Ernst, 1980) were phase cycled according to States et al. (1982), with a 15% variation in the mixing time to suppress zero-quantum peaks. Magnitude COSY spectra were acquired as above. The solvent resonance was suppressed by low-power irradiation during the relaxation delay. The data were processed on a VAX 11/750 with the software package for NMR processing written by Dennis Hare. FIDs were weighted by a skewed sine-bell function in t_2 and a skewed or shifted sine-bell function in t_1. NOESY spectra were symmetrized about the diagonal.

Interproton distances were determined from volume integrals of NOESY cross-peaks as described (Woodson, 1987), using the relationship

$$r_{ij}/r_{kl} = (a_{kl}/a_{ij})^{1/6} \quad (1)$$

where r_{ij} and r_{kl} are the two distances being compared and a_{kl} and a_{ij} are cross-peak intensities. Distances were determined for all adequately resolved cross-peaks between assigned resonances from data sets acquired with mixing times of 100, 150, and 200 ms. The mean intensities of the CH5–CH6 cross-peaks (set to 2.39 Å) and the H2′–H2″ cross-peaks (set to 1.8 Å) were used as references. The intra-ribose distances determined from either reference differed by less than 0.10 Å. As discussed previously (Woodson, 1987), it was necessary to correct the sugar–base proton distances so that the average H1′–H6, H8 distance was 3.7 Å. In many cases, at least two out of the three data sets agreed closely (within 0.2 Å) on a given interproton distance, which was chosen to be the constrained distance. For the remaining constraints, an average was taken of the three experimental distances, unless one integral was clearly less reliable than the others. The few interproton distances that varied by more than 0.6 Å over the data sets were not included in the calculation.

Energy-minimized structures were computed with the AMBER program (Weiner et al., 1984). Starting coordinates were standard B DNA from fiber diffraction (Arnott & Hukins, 1972). Counterions were added to the molecule with a distance-dependent dielectric function with the dielectric constant set to 1.0, and the nonbonded interactions were cut off at 8.0 Å. Distance constraints were added to the force field with a very heavy weighting (30.0 kcal) at the start of the refinement and a much lighter one (5.0 kcal) in the final stage of the minimization. In order to prevent distortion of the bond angles by the high energy of the abnormally long P–O bonds in the starting coordinates, angle constraints were added to the backbone and phosphate bond angles adjacent to the stretched bond.

In order to build up the entire nine base pair sequence, segments of B DNA were visually docked on to a segment of minimized coordinates by using FRODO on an Evans and Sutherland Picture System 300. An effort was made to maintain base stacking and the best backbone continuity possible for both chains. A number of spliced coordinate sets were stored at each stage, and the four structures with the lowest starting energy in the first 50 cycles of minimization were minimized to completion with the appropriate distance constraints for that segment of the molecule. The starting energies were generally quite reasonable, around 10^5 kcal. To prevent undue distortion of the molecule, it was helpful to begin the refinement with a belly minimization where only the smaller segment was allowed to move, in addition to heavily weighted bond angle constraints along the backbone at the junction of the segments. Following several hundred cycles, all the coordinates were included in the calculation. After 800 cycles the phosphate bonds had reached their equilibrium value of 1.61 Å, and the bond angle constraints were removed. The interproton distance constraints were maintained throughout the calculation until the final stages, when they were included with only a light weighting.

Results

Two-Dimensional NMR of G-Bulge Duplex. The nonexchangeable proton spectrum of d(GATGGGCAG)·d(CTGCGCCATC) containing an extrahelical guanosine at 490 MHz is shown in Figure 1 and is compared to the spectrum of a regular 8-mer, d(GATGGCAG)·d(CTGCCATC). The sequences are also diagrammed in Figure 1. The NMR of the exchangeable protons of this oligomer has been described elsewhere (Woodson & Crothers, 1988). The nonexchangeable protons in short DNA oligomers can be readily assigned from sequential NOESY connectivities in addition to the COSY connectivities (Hare et al., 1983; Weiss et al., 1984). In regular B DNA, the aromatic base protons are close enough to cross-relax with the sugar protons of both the same residue and the 3′ neighboring nucleotide. This potentially provides three independent series of sequential NOE contacts along both strands of the duplex to facilitate the correct assignment of the resonances. The aromatic and the H1′, H2′, H2″, and H3′ sugar proton resonances were assigned for the G-bulge-con-

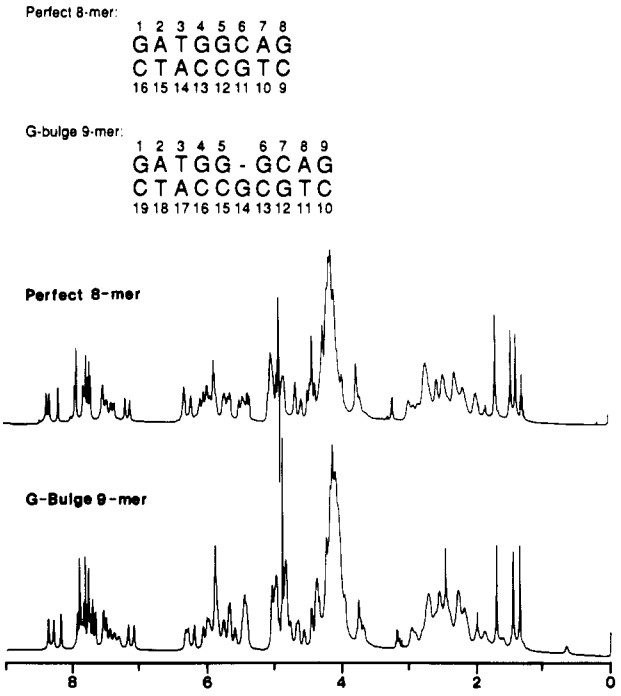


FIGURE 1: Proton magnetic resonance spectra of the G-bulge 9-mer ^{5}d(GATGGGCAG)·d(CTGCGCCATC) and the regular 8-mer ^{5}d-(GATGGCAG)·d(CTGCCATC) at 490 MHz in 10 mM phosphate and 100 mM Na^+, pH 7.0, in D_2O at 288 K. Chemical shift scale is from 0.0 to 9.0 ppm relative to TSP.

Table I: Chemical Shifts (ppm) of Nonexchangeable Protons in G-Bulge 9-mer[a]

residue	H8	H6	H5	H1′	H2′	H2″	H3′	CH_3
G1	7.91			5.68	2.61	2.80	4.87	
A2	8.31			6.29	2.73	2.96	5.05	
T3		7.10		5.70	1.88	2.26	4.83	1.37
G4	7.76			5.60	2.59	2.70	4.98	
G5	7.68			5.84	2.41	2.42	4.99	
G6	7.88			5.87	2.64	2.70	4.98	
C7		7.39	5.46	5.43	1.98	2.27	4.83	
A8	8.20			6.06	2.74	2.90	5.03	
G9	7.73			6.01	2.47	2.28	4.65	
C10		7.81	5.90	5.89	2.13	2.56	4.69	
T11		7.51		5.76	2.19	2.48	4.90	1.71
G12	7.95			5.87	2.69	2.69	5.01	
C13		7.33	5.46	5.95	1.87	2.24	4.81	
G14	7.86			5.91	2.55	2.55	4.99	
C15		7.55	5.49	5.78	2.18	2.41	4.77	
C16		7.56	5.67	5.45	2.18	2.43	4.84	
A17	8.38			6.33	2.75	2.97	5.04	
T18		7.19		5.99	2.00	2.46	4.86	1.47
C19		7.48	5.46	6.21	2.28	2.28	4.57	

[a] Measured relative to TSP at 288 K in 10 mM phosphate, 100 mM Na^+, and 0.1 mM EDTA, pH 7.0.

taining 9-mer by this method, and their chemical shifts are listed in Table I and the NOESY connectivities are outlined in Figures 2 and 3. The H4′ and H5′ sugar resonances fall in a very narrow region of the spectrum and were not assigned at this time. The two-dimensional NMR data on a regular 8-mer of similar sequence have been presented previously (Woodson, 1987). We use this molecule as a reference for the bulge-containing duplex in order to distinguish general features of the sequence from the particular effects of the bulge.

As one might expect, insertion of an extra purine into the helix primarily affects the chemical shifts of the neighboring base pairs G5·C15 and G6·C13, and the rest of the spectrum is not significantly altered. On the strand opposite the bulge,

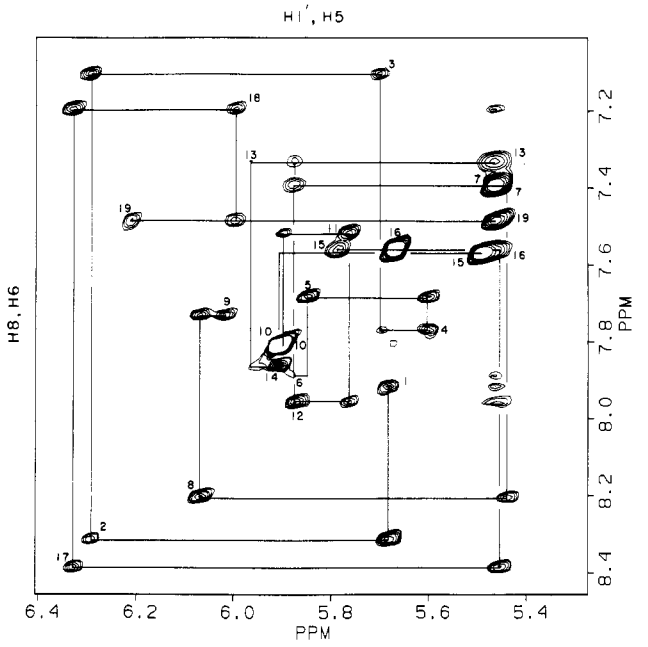


FIGURE 2: Region of the phase-sensitive NOESY spectrum of the G-bulge 9-mer showing cross-peaks between aromatic and H1′ and H5 proton resonances. The spectrum was acquired as described under Materials and Methods with a mixing time of 150 ms, at 288 K. Sequential connectivities are outlined for both strands, and intraresidue H8, H6–H1′ and H6–H5 cross-peaks are labeled by residue number.

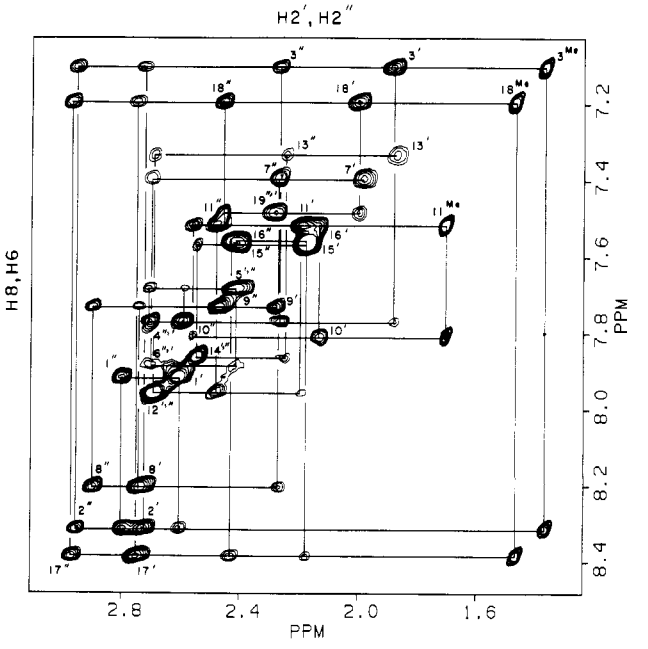


FIGURE 3: Region of NOESY spectrum containing H6, H8–H2′, H2″ and thymine methyl cross-peaks. Connectivities are outlined as in Figure 1, and intraresidue cross-peaks are labeled with the residue number; ′ and ″ denote the H2′ and H2″ resonances, respectively; M denotes methyl resonances.

G6 H8 has a chemical shift of 7.88 ppm in the G-bulge mer compared with 7.68 ppm for G5 H8 in the regular 8-mer, and the H6 and H5 resonances of C7 have also shifted downfield relative to the analogous resonances in the 8-mer. In the strand containing the bulge, the H5 and sugar resonances of C13 and H6 in C15 in the G-bulge duplex are all shifted at least 0.1 ppm from the resonances in the regular 8-mer. It is also important to note that the chemical shifts of the bulged G resonances are very similar to those of G6 and G12, which are also part of a GpC step. This implies that the chemical en-

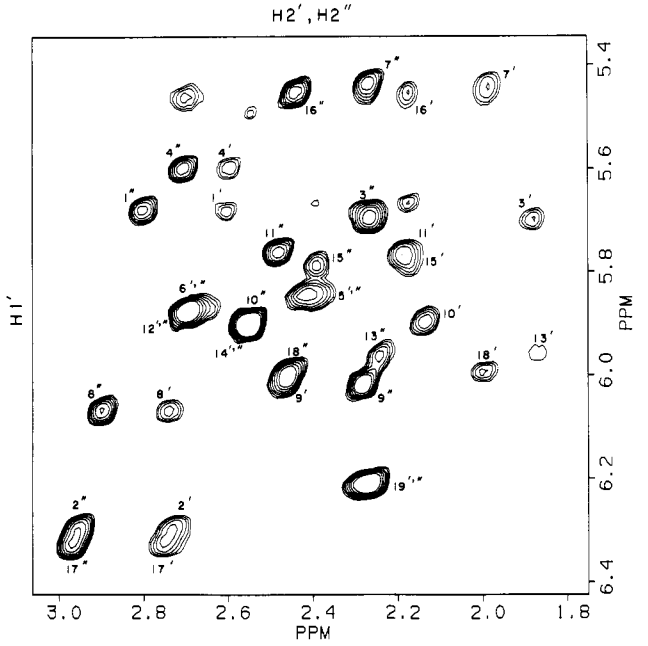


FIGURE 4: Region of NOESY spectrum containing H1′ and H2′, H2″ proton resonances.

vironments of these residues are fairly similar and strongly suggests that the bulged purine G14 is stacked into the helix. This agrees with our previous results for a bulged adenine (Woodson, 1987) and with those of Patel and co-workers (Hare et al., 1986b; Patel et al., 1982).

Comparison of NOESY Spectra. The region of the NOESY spectrum of the G-bulge 9-mer from 8.4 to 7.0 and from 6.4 to 5.3 ppm containing cross-peaks between aromatic base proton resonances and H1′ and cytosine H5 resonances is shown in Figure 2. Except for a few residues in the region of the bulge, the cross-peaks between the base protons and the anomeric sugar protons are clearly visible at a mixing time of 150 ms. The intense cross-peaks in this region of the spectrum correspond to cytosine H5–H6 resonances, and weak connectivities between H5 and the 5′ neighboring H6 or H8 resonance can also be observed, except at the bulge site, between G14 and C15.

The region from 8.4 to 7.0 and from 3.2 to 1.2 ppm containing cross-peaks between base protons and H2′, H2″, and thymine methyl resonances is shown in Figure 3. As in the previous figure, the majority of the expected contacts are observed. The intense cross-peaks between the thymine H6 methyl resonances can also be seen in this figure between 1.6 and 1.2 ppm. The relative intensities of the cross-peaks to sugar resonances are characteristic of a B-type helix. For example, the cross-peaks to T3 H6 and T18 H6 in Figure 3 show the typical "strong–medium–medium–weak" pattern of B DNA (Feigon et al., 1983). The intra-ribose NOESY and COSY connectivities provide confirmation of the sequential assignments, and the H1′–H2′, H2″ region of the NOESY spectrum is shown in Figure 4. In this case, these cross-peaks also enabled us to assign the H2 resonances of G6, C15, and C16, which were difficult to assign due to weak NOEs and resonance overlap. The relative intensities of the intra-ribose NOEs at shorter mixing times also provide information about the sugar conformation.

A qualitative comparison of the connectivities in the G-bulge duplex and the regular 8-mer highlights several significant changes in the helix around the bulged nucleotide. On the strand opposite the bulged G, the connectivities between G5 and G6, which lie directly opposite the bulge, are much weaker than usual and much weaker than those of G4 and G5 in the regular helix. The cross-peak between G5 H1′ and G6 H8 is not observed, and that between G5 H2″ and G6 H8 is very weak. On the strand containing the bulge, the cross-peaks between G12 sugar protons and C13 H6 are very weak, and the same is true for C13 and G14 and for G14 and C15. A very weak cross-peak is observed between C13 H1′ and G14 H8 and from C13 H2″ to G14 H6. The cross-peak between G14 H1′ and C15 H6 is below the noise level, but G14 H2′ and H2″ do give a cross-peak to C15 H6. Connectivities in the remainder of the molecule are fairly similar to those in the perfect duplex.

The presence of weak NOE contacts to the bulged guanosine resonances is added evidence that the base is stacked into the helix but suggests that the geometry is distorted at this site. The weak NOE contacts between the two bases bridging the bulge site on the opposite strand of the helix also suggest that the backbone is stretched more than usual between G5 and G6, as one might expect if an extra base were inserted into the duplex. Taken together with the chemical shift data, it seems unlikely that the bulged base is stacked out of the helix. The medium intensity of the G14 H8–H1′ cross-peak rules out the possibility of a syn conformation about the glycosidic bond.

An interesting feature of the NOESY spectrum of the G-bulge 9-mer is the weak intensity of all the cross-peaks assigned to resonances of G6 and C13. What is particularly important to note is that even the H5–H6 cross-peak of C13 is of much lower intensity at any given mixing time than the other cytosine H5–H6 cross-peaks, although the geometry about the aromatic ring is fixed. This suggests that G6–C13 is experiencing shorter correlation times than the rest of the helix, most likely due to greater motion at these residues. This is true only of G6–C13 and not for G5–C15, which also flanks the bulge. We did not observe such a striking variation in CH5–H6 intensities in the A-bulge 9-mer (Woodson, 1987). The terminal base pairs, which undergo substantial end fraying at this temperature and which one might expect to have greater thermal flexibility, apparently have correlation times similar to those in the rest of the molecule, judging by the relative intensity of the C19 H5–H6 cross-peak. Although there is no salient alteration of the geometry at this base pair in the calculated structures presented below, one could speculate that partial loss of base stacking at the bulge site could increase the flexibility of the helix.

Energy Minimization of G-Bulge 9-mer. Molecular mechanics is an extremely useful tool in determining energetically and sterically reasonable structures for a given molecule. We computed energy-minimized coordinates for the G-bulge 9-mer using the proton–proton distances taken from the NMR data as constraints on the minimized structure. This allows one to translate qualitative observations about the helix conformation into a three-dimensional model. The distance constraints were obtained from intensities of NOESY cross-peaks at mixing times of 0.100, 0.150, and 0.200 s, with cytosine H5–H6 and intra-ribose H2′–H2″ as reference distances. Exclusion of cross-peaks where resonance overlap prohibited reliable integration resulted in 109 usable distance constraints for the complete sequence, which are listed in Tables II–IV. The agreement between distances determined independently from each data set for a given proton pair was fairly good.

During work on the structure of the A-bulge duplex it became apparent that one of the problems in minimizing a coordinate set of this size with a substantial defect in the center is the tendency to fall into local energy minima. Rather than

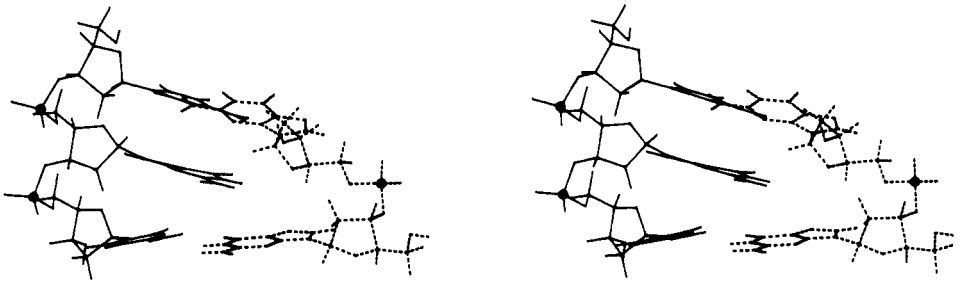


FIGURE 5: Stereo diagram of the refined coordinates for the bulge-containing dimer "core" $^{5'}$d(GpG)·$^{5'}$d(CpGpC).

Table II: Intraresidue Sugar–Base Proton Distance Constraints (Å)

H6/H8 →	H1′	H2′	H2″	H3′
G1	3.71	3.08	2.80	
A2	3.86	2.57	3.32	
T3	3.72	2.81	3.45	
G4	3.62	2.77	2.94	
G5	3.54			3.55
G6			3.50	3.21
C7		2.81	2.98	3.96
A8	3.61	2.46	2.89	
G9	3.70	2.53	3.14	
C10		3.13	4.42	
T11	3.70	2.55	3.13	
G12	3.63			3.35
C13	4.08	2.20		
G14	3.68			3.12
C15				4.10
C16				3.12
A17	3.95	2.53	3.18	
T18	3.95	2.77	3.34	
C19	3.73			3.20

Table III: Interresidue Sugar–Base Proton Distance Constraints (Å)

H6/H8 →	sugar	H1′	H2′	H2″
A2	G1	3.25	4.11	3.04
T3	A2	3.60	3.90	3.38
G4	T3	4.82	4.00	3.75
G5	G4	3.68	5.00	3.90
G6	G5		3.29	
C7	G6	3.67	4.50	4.03
A8	C7	3.63	4.76	3.93
G9	A8	3.85	4.23	3.37
T11	C10	4.50		3.19
G12	T11	4.02	3.43	3.34
C13	G12	3.59		
G14	C13	3.11	4.18	3.02
C15	G14	5.50		
C16	C15			
A17	C16	3.58	4.19	3.55
T18	A17	3.96	4.07	3.26
C19	T18	3.96	3.48	2.92

move a large number of coordinates to produce a global change in the structure, such as a kink, the coordinates in the region of the defect are deformed to lower the energy of the structure, which then becomes trapped in a local minimum. One approach is to simplify the energy surface by excluding nonbonded parameters at the start of the minimization. Another approach is to build the structure gradually, minimizing smaller helical segments, which are then joined together. We chose this latter approach in refining the G-bulge duplex.

As a first step, we wished to see whether a two base pair fragment $^{5'}$GpG·$^{5'}$CpGpC containing the bulged G would minimize in a reasonable fashion. The starting coordinates were regular B DNA with a cytidine excised from one strand opposite the bulged guanosine. Since the NMR data unambiguously indicate that the oligomer is in a B-type conformation, with the extra guanine stacked into the helix, it seemed

Table IV: Intra-Ribose Distance Constraints (Å)

residue	H1′/H3′ →	H2′	H2″
G1	H1′	2.83	2.52
T3	H1′	2.78	2.25
G4	H1′	2.83	2.50
C7	H1′	2.58	2.41
A8	H1′	2.73	2.45
G9	H1′		2.29
G9	H3′	2.74	2.59
C10	H1′	2.56	
C10	H3′	2.56	
T11	H1′		2.41
T11	H3′		2.77
C13	H1′	2.88	2.53
C15	H1′		2.68
C15	H3′	2.42	2.80
C16	H1′	2.78	2.38
T18	H1′	2.80	

reasonable to start with B DNA coordinates. In the absence of NMR constraints, the sugar rings on the strand opposite the bulge in the minimized coordinates were rotated toward each other in order to close the gap and would obviously be unsuitable for continuing the helix. With heavily weighted experimental distance constraints, however, the proper ribose conformation was maintained, with surprisingly good results. The structure of the minimized dimer is shown in Figure 5. The most outstanding feature of the structure is the wedgelike tilt of the flanking base pairs. One can imagine that a continuation of the helix from these bases would result in a kinked helix axis. The lower G·C pair, corresponding to G5–C15 in the complete sequence, is fairly flat and undistorted, but in the upper pair, G6·C13, the bases are strongly tilted relative to one another.

Following minimization of the core dimer, a four base pair segment of B DNA, $^{5'}$d(GATGp)·$^{5'}$pCATC, was visually docked onto the dimer by using stereo graphics on an Evans and Sutherland PS300. A number of "spliced" coordinate files representing slightly different attempts to fit the two helices were stored and their starting energies determined. The four coordinates with the lowest starting energies were minimized with the appropriate constraints for the hexamer. In order to prevent distortion of the phosphate backbone at the junction between the two helices, bond angle constraints were initially included in the force field. After the P–O bond distance reached its equilibrium value, the angle constraints were removed. The two hexamers, structures A and B, that best satisfied the distance constraints (see Table V) and had the lowest energies are shown in Figure 6. The general features of the dimer have been preserved, although there was a small change in the conformation of the core residues during minimization of the hexamer. The energies and rms deviation from the constraints for the intermediate structures are listed in Table V. As might be expected, the overall rms deviation from the constrained distances is lower in the dimer than the hexamers.

Table V: Relative Energies of Intermediate Structures

	dimer	A	B	C	D
total energy[a]	70.02	420.80	450.09	431.71	460.01
nonbonded VDW	−54.23	−189.85	−190.18	−189.60	−190.61
nonbonded EEL	1.58	3.45	3.41	3.71	3.45
H bond 10–12	−2.90	−7.44	−7.86	−7.68	−7.66
bond	2.57	8.19	7.75	8.08	7.75
angle	20.98	67.68	66.66	65.33	65.31
dihedral	49.21	137.38	137.74	138.96	136.15
nonbonded 1–4 VDW	23.34	65.50	65.94	65.92	65.54
nonbonded 1–4 EEL	−5.48	−13.02	−13.04	−13.01	−13.02
constraints	34.95	348.91	379.68	360.20	393.09
mean rms deviation[b]	0.197	0.323	0.337	0.328	0.343

[a] Energies (kcal/mol) relative to an arbitrary standard. [b] Mean rms deviation of calculated interproton distances (Å) for all constrained pairs.

The final nine base pair sequence was formed in the same manner by docking a trimer, $^{5'}$d(pCAG)·$^{5'}$d(CTGp), with the refined hexamer structures A and B. Again, a number of coordinate sets were tested, and the four structures with the lowest starting energies were subjected to a complete minimization, with very lightly weighted distance constraints in the final stages (structures A1, A2, B1, and B2). Two of these structures, A1 and B2, are shown in Figure 7A. The general features of all these helices are very similar, and in all cases, the wedgelike tilting of the bases flanking the bulged guanosine has been preserved from the core structure. In addition, there is a noticeable bend in the helix axis. The similarity of structures obtained from different docking attempts reassures us that the final coordinates are not dependent on the docking procedure itself.

In order to compare this approach with the method used previously to refine the structure of the A-bulge duplex (Woodson, 1987), we also minimized the G-bulge duplex from B DNA starting coordinates for the entire 9-mer with a gap in one strand opposite the bulged guanosine. The same distance constraints were used, with the same weighting. At the start of the minimization, bond angle constraints were used around the gapped phosphate residue, which were removed after 900 cycles. The calculation was continued with heavy distance constraints, followed by a final minimization with very light constraints. That structure (H) is shown in Figure 7B. It is surprisingly similar to those obtained by the segmental minimization. Most important, the tilting of the bases around the bulged G and the resultant kink in the helix axis are also apparent in this structure. This feature is clearly a result of the experimental constraints, or the force field for these bases, or both, and is independent of the model building.

Energy-Minimized Structure without Distance Constraints. In order to examine the effect of the distance constraints on the computed structures, the minimization was carried out as above, but without distance constraints. The structure (N) is shown in Figure 7B. The bases adjacent the bulged G are still tilted, and many of the overall features of the helix are similar to those obtained with experimental constraints. The helix axis is still bent slightly, but much less than in the other structures (see Table VII). The nonbonded interactions are actually slightly higher for this structure than for those computed with constraints, and the bases at the ends of the helix are less tilted than in the previous structures.

Comparison of Minimized Structures. The final energies and rms deviations from the constraints of the calculated structures for the G-bulge 9-mer are listed in Table VI. The energies are remarkably similar for all the minimized structures. A1 and A2 have slightly lower energies than B1 and B2 and also lower rms deviations from the distance constraints. Both H and N have slightly lower bond and angle energies than A1–B2, suggesting that the joining of the helical segments does cause some small distortions in the molecular geometry. Furthermore, it should be noted that there is no contribution from constraints to the total energy of N, which is why it has a much lower total energy than the other coordinates. As might be expected, the deviation from the constrained distances in N where the constraints were not actually included in the minimization is much greater than in the other structures. The overall mean deviation is 0.650 Å for N, compared with 0.0456–0.474 Å in the other structures. The relative rms deviations from the experimental constraints for each residue are plotted in Figure 8. Again, structure N generally deviates from the constraints more than the other structures, which are nearly the same as each other. The biggest disparity in the rms values are at the residues flanking the bulge site, G5 and G6. The constraints are satisfied most poorly at residues G6, C16, and A17, most likely due to unreasonable values for some of the interproton distances.

The base tilt and twist of structures A1, B2, H, and N are plotted in Figure 9. Here, base tilt is the angle between the normal to the base plane and the helix axis, which includes both roll and longitudinal tilt of the base plane. In general, the base tilt values are similar in all the structures. There is a noticeable difference at G6, which has a low tilt angle (9°) in A1 and B2 and a high angle in H and N (18°) that were minimized directly from the complete coordinates. Structures

Table VI: Relative Energies of Refined Nine Base Pair Helices

	A1	A2	B1	B2	H	N
total energy[a]	46.20	47.98	45.59	46.92	41.69	12.93
nonbonded VDW	−309.35	−310.11	−306.65	−306.30	−306.31	−303.79
nonbonded EEL	3.31	3.45	3.37	3.47	3.12	3.84
H bond 10–12	−12.70	−12.55	−12.76	−12.78	−12.34	−12.70
bond	7.45	7.41	7.44	7.53	7.20	7.21
angle	72.45	72.50	72.19	73.21	70.01	65.48
dihedral	182.98	184.12	183.58	184.61	178.15	178.97
nonbonded 1–4 VDW	93.34	93.35	93.00	93.11	93.91	93.10
nonbonded 1–4 EEL	−19.17	−19.17	−19.18	−19.17	−19.19	−19.18
constraints	27.89	28.97	24.49	23.24	23.14	0.00
mean rms deviation						
heavy constraints[b]	0.422	0.420	0.397	0.396	0.395	
final (light)[c]	0.456	0.457	0.474	0.462	0.461	0.650

[a] As in Table V. [b] Mean rms deviation of calculated interproton distances from constraints with a heavy weighting. [c] Mean rms deviation from distance constraints in final structures, calculated with lightly weighted constraints. For structure N, the distance constraints were not actually included in the minimization.

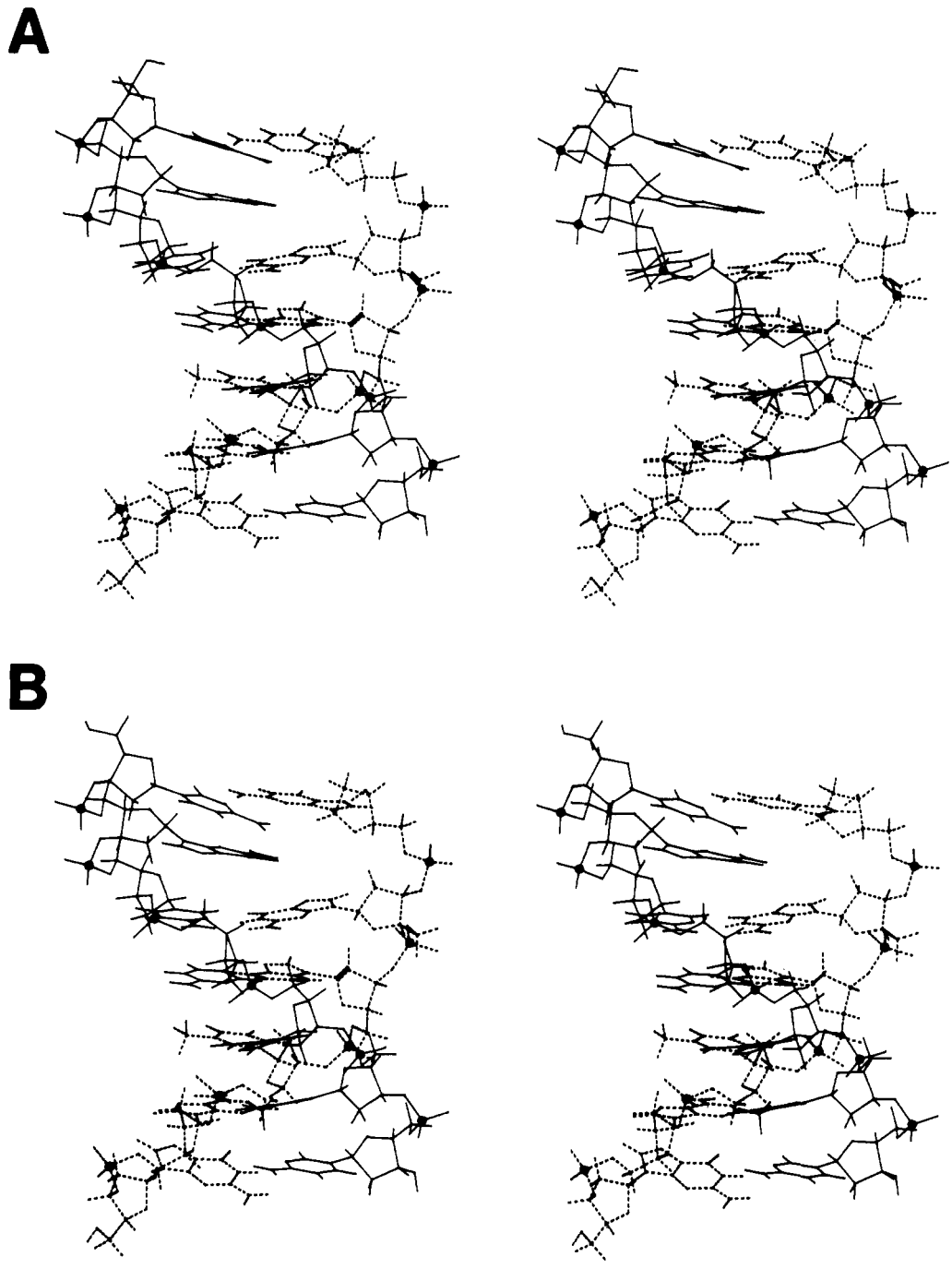


FIGURE 6: Stereo diagrams of two refined hexamers, structures A and B, of sequence $^{5'}$d(GATGGG)·$^{5'}$d(CGCCATC).

computed with the constraints tend to have more propellor twist than structure N (Figure 7B). It is conceivable that the experimental constraints are biasing the energy-minimized structure in favor of the extremes of thermal motion rather than the average, since the r^{-6} dependence of the NOE intensities will weight the shortest distances much more heavily than the longer ones. Molecular dynamics simulations would be one way of answering this question.

The sugar pucker phase is plotted versus sequence in Figure 10. As one would expect, the majority of the ribose rings are in the C2′ endo–C3′ exo conformation. The bulged guanosine (G14), however, as well as G5, C16, and C19 are closer to a C3′ endo puckering. The sugar conformation seems to be very sensitive to the intra-ribose constraints. All of the structures minimized with distance constraints are quite similar to each other, but the sugar puckers in structure N differ at several positions. The puckers of G5, C16, and C19 are C2′ endo in N, instead of C3′ endo as in the other structures.

The helical twist and helix axis kink angles are listed in Table VII. Surprisingly, the helical twist is 10.5–10.6 base pairs per turn for the entire duplex if one considers it to be a 10-mer instead of 9-mer. The twist angles at the bulged residue are slightly smaller than for a normal base pair, around 27°, and the remainder of the helix is overwound (10 base pairs per turn), expecially at G6–C7·G12–C13, which makes a step of 38–40°. This agrees with the results for the A-bulge duplex, where the helical twist was between 10.6 and 10.8 base pairs per turn with a twist angle of 27° at the bulge (Woodson, 1987). It is also encouraging that the helical twist is identical in all the segmentally minimized structures and in H, which was minimized in one step.

In order to control for any sequence effect outside of the bulge on the helical twist of the refined structures, a regular 9-mer of the same sequence $^{5'}$d(GATGGGCAG)·d-

Table VII: Helical Parameters for Refined Structures

	A1	A2	B1	B2	H	N
helical twist (bp/turn)[a]						
10-mer	10.58	10.50	10.62	10.54	10.61	10.48
without bulge	10.03	10.00	9.97	9.92	10.22	10.27
helical twist angles (deg)						
at bulged G[b]	28.43	29.08	26.73	27.20	30.19	29.19
	26.75	27.74	25.66	26.13	28.54	27.60
kink angles (deg)[c]						
end to end	11.80	15.40	14.45	14.51	15.94	12.26
at bulge	18.03	21.13	21.94	22.00	19.55	13.97

[a] Average helical twist over all dinucleotide units; "without bulge" excludes the angle between G5 and C7 and C13 and C15; "10-mer" is the twist where the helix is considered to be 10 base pairs (bp) instead of 9. [b] Half the helical twist angles between G5–C7 and C13–C15. [c] Apparent kink angles in the helix axis calculated from phosphate positions. "End to end" is the overall kink angle between the ends of the helix; "at bulge" is the kink angle at the bulge site.

(CTGCCCATC) and a regular 10-mer $^{5'}$d-(GATGGCGCAG)·d(CTGCGCCATC) were minimized from B DNA coordinates without experimental distance constraints. The minimized 9-mer and 10-mer both had a helical screw of 10.4 base pairs per turn. This demonstrates that the overwinding of the flanking base pairs in the structures calculated for the bulge-containing duplex is a result of the extra nucleotide and not a property of this particular sequence as determined by the AMBER force fields. There is experimental evidence, however, that the unwinding due to the insertion of an extrahelical nucleotide is not fully compensated by the flanking base pairs (Rice, 1987).

The kink angles of the helix axis at the bulged guanosine are 18–24°, slightly larger than those observed in the A-bulge structures. This is in excellent agreement with gel mobility experiments, which predict a bend of 20° for each bulged nucleotide (Rice, 1987). The phasing of the bend in the gel mobility experiments agrees with the wedge in the base pairs flanking the bulge seen in these structures. The end-to-end bend angles of the calculated duplexes are somewhat less, 12–16°, due to smaller kink angles of 2–6° in the adjacent base pairs that compensate for the angle at the bulge site. These smaller kink angles are also present in the minimized regular 9-mer described above, which has an end-to-end bend of 12°. "Regular" B DNA is known to contain small changes in the direction of the helix axis (Dickerson, 1983). However, there is no independent evidence at this time that these small kink angles are correct for this sequence and not an artifact of the molecular mechanics.

Discussion

Three-Dimensional Conformation from NMR and Energy Minimization. Several important features about the helical structure of the bulge duplex are apparent upon consideration of the two-dimensional NMR spectra. The chemical shift of G14 H8 and the NOE connectivities suggest strongly that the extra guanine is stacked into the helix. The changes in chemical shift and NOE connectivities are mainly confined to the residues near the bulged G and indicate that the insertion of an isolated extrahelical nucleotide causes only a local perturbation of helix structure. The reduction of NOE intensities between G4 and G5, which are opposite the bulge, and the relatively weak contacts between G14 and its neighbors support the idea that the helix is spread slightly at the bulge position, but not disrupted entirely.

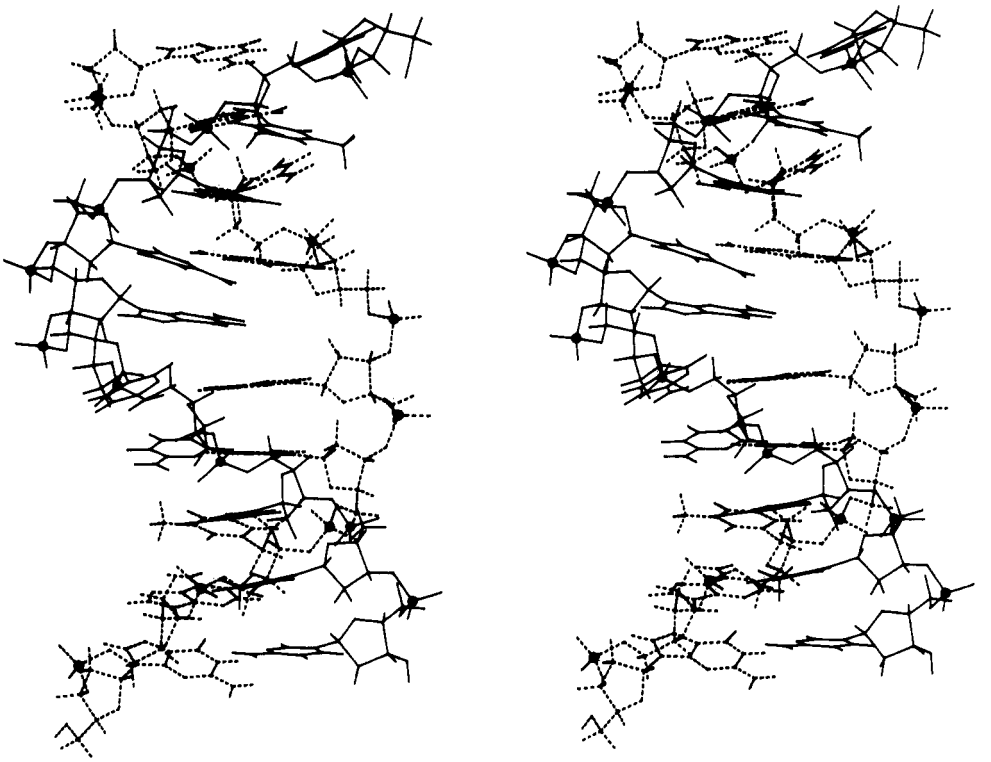
As one would hope, the general structural features predicted by the NOESY spectrum are in fact apparent in structures calculated with distance constraints taken from the NMR experiments. We chose to start from B DNA coordinates with the extra base inserted into the helix, since the NOESY connectivities clearly represent a B-type helix. The bases flanking the bulge adopt a wedgelike conformation around the bulged base in all of the minimized helices. This would account for a slight spreading of G4 and G5, but not to the extent that all connectivities are lost. The helical twist angles at the bulge position (27°) also imply a stretching of the residues on the opposite strand, reducing the internucleotide NOEs. This spreading of the strand opposite a bulge rather than compaction of the bulge-containing strand is also a consistent feature of the energy-minimized structures calculated for the A-bulge 9-mer (Woodson, 1987).

Comparison of A- and G-Bulge Duplexes. The tilting of the bases is also largely a function of the base-stacking energies, which are a result of the nonbonded van der Waals and electrostatic terms of the molecular mechanics force field. This may explain in part the differences between the A-bulge 9-mer and the G-bulge 9-mer structures. In the A-bulge duplex, of sequence $^{5'}$d(GATGGGCAG)·d(CTG*A*CCCATC), the extra adenine is inserted at a GpC, whereas the extra guanine in the G-bulge 9-mer is between two cytosines. In the A-bulge structure, the cytosine on the opposite strand (C7) was strongly tilted into the gap created by the unpaired adenine, while the guanine on that strand (G6) was not strongly tilted and remained stacked with the base pair below. In the G-bulge sequence, there are now two purines opposite the bulge site. The observed "wedge angle" of the flanking base pairs may be the best way of optimizing purine–purine stacking. In both the A-bulge and G-bulge structures, the orientation of the bases did not change significantly even when the distance constraints were left out of the minimization. It appears that while the constraints may be contributing to the final conformation of the helix, especially of the backbone, the base-stacking energies as modeled in the computational force field are an important determinant of the final structure. The overall fit to the experimental distances is somewhat better in the G-bulge- than the A-bulge-calculated structures. The backbone conformation varies slightly in each duplex. Interestingly, the extrahelical riboses in both bulge-containing helices adopt an N (C3′ endo) conformation, while the flanking sugars in both molecules are in an S (C3′ exo) conformation.

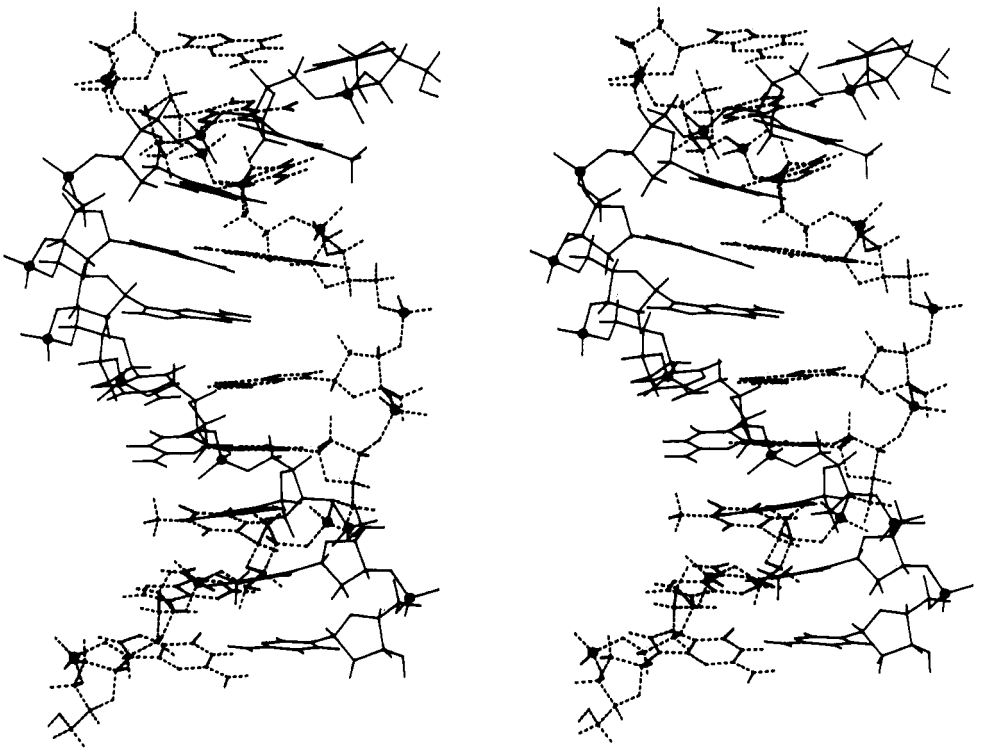
Error in Distance Constraints. The relationship in eq 1 is only valid for short mixing times, where the rate of cross-relaxation is roughly linear and where the effective correlation times of the reference and unknown distance are similar (Clore et al., 1985a,b). This linear relationship does not hold for oligonucleotides at mixing times of 100 ms or greater, but the decreasing signal to noise that results from shorter mixing times introduces large uncertainty into the value of the volume integral itself. There is also evidence that secondary NOEs via H2′ may be the most significant source of error in H1′–H6, H8 distances, which are usually grossly underestimated relative to cytosine H5–H6 (Chazin et al., 1986). This results in real

A

A1



B2



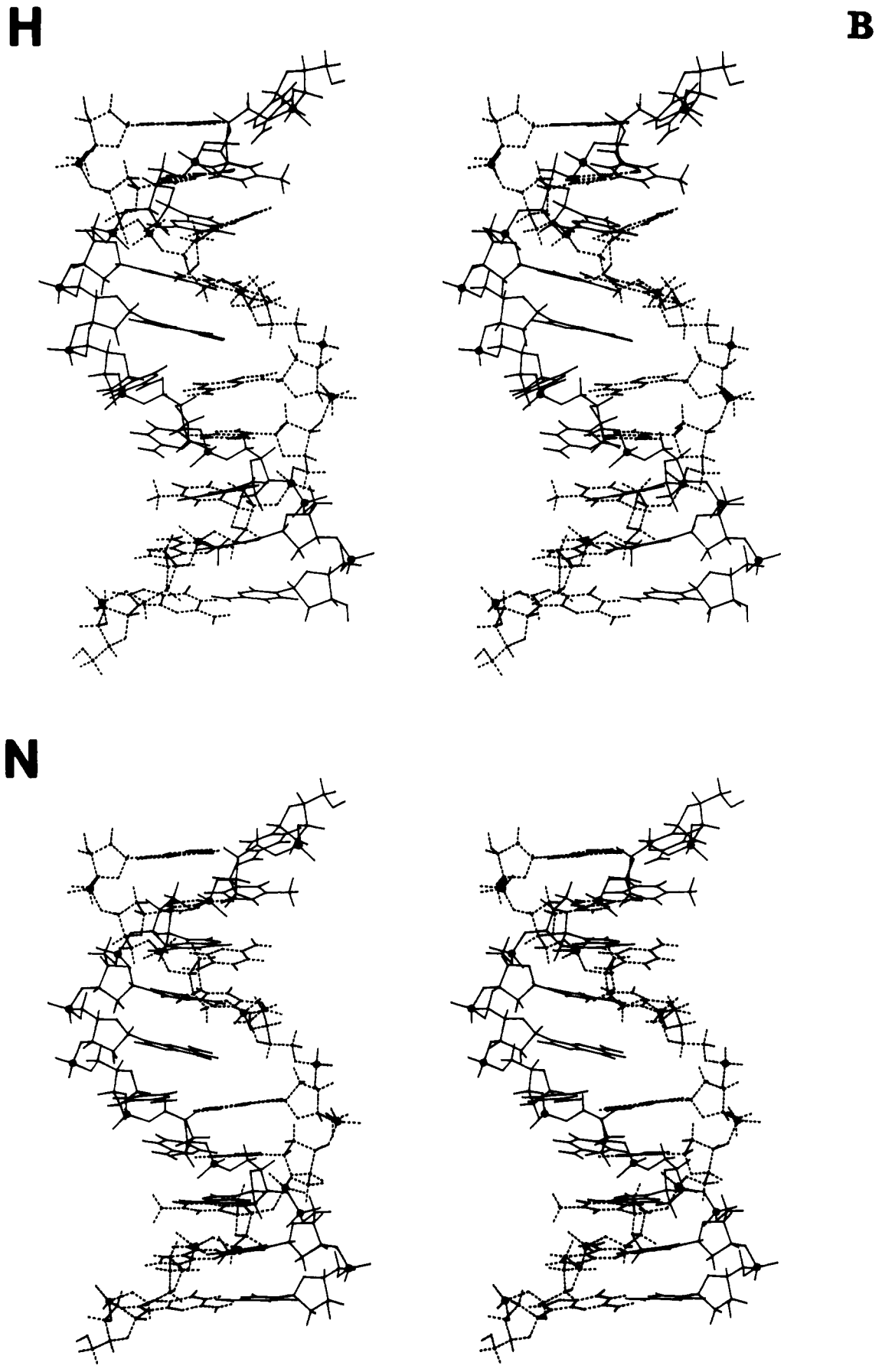


FIGURE 7: (A) Stereo diagrams of two minimized coordinates for the entire G-bulge 9-mer, structures A1 and B2. These structures were obtained by building up the final sequence in three segments (see Figures 5 and 6), with energy minimization at each stage including the appropriate experimental distance constraints. (B) Structures H and N for the G-bulge 9-mer. Starting coordinates were B DNA for the entire 10 base pair sequence with the cytosine opposite the bulged guanosine excised. H was minimized with experimental distance constraints; N was calculated without distance constraints.

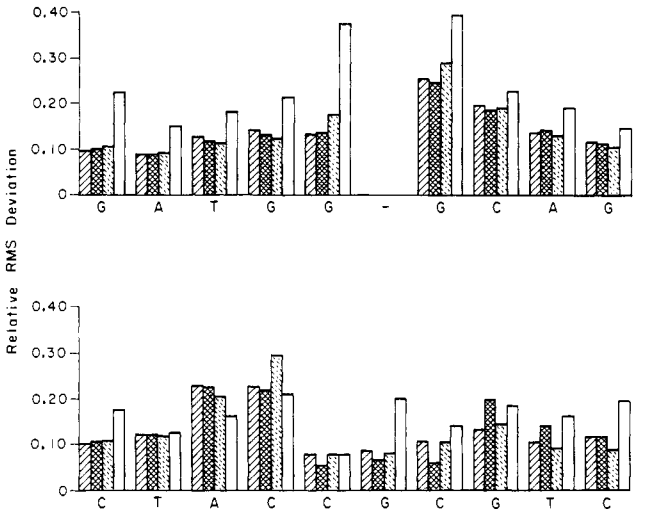


FIGURE 8: Relative rms deviations from the distance constraints at each residue plotted versus sequence for structures A1 (▨), B2 (■), H (⊡), and N (□). The relative deviation was obtained by dividing the difference between a calculated distance and the constraint by the value of the constraint.

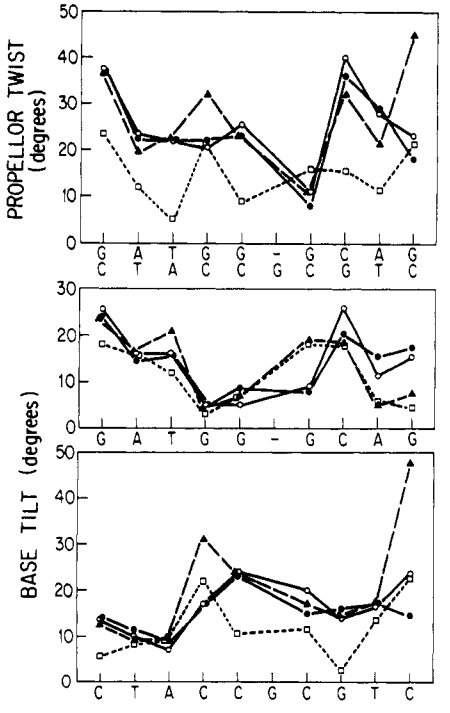


FIGURE 9: Plot of propellor twist (top) and base tilt (middle and bottom) versus sequence position: (○) A1; (●) B2; (▲) H; (□) N. Propellor twist is defined here as the angle between the base planes along the short axis, and tilt is defined as the angle between the normal to the base plane and the helix axis.

practical limitations on the use of NMR data for structure determination of oligonucleotides. Progress is being made in iterative calculations where a theoretical NOESY spectrum calculated from the refined structure is compared with the actual data (Keepers & James, 1984; Lefevre et al., 1987). This approach will undoubtedly improve the quality of solution structures available.

We estimate the overall uncertainty in the distance constraints to be ±0.2 Å for distances less than 2.50 Å, ±0.3 Å for distances 2.5–3.3 Å, and +0.5 or −0.2 Å for distances greater than 3.2 Å. We have attempted to use the greatest number of constraints possible to reduce distortion of the structure by error in any one constraint. Furthermore, the relatively good agreement between the adjusted interproton

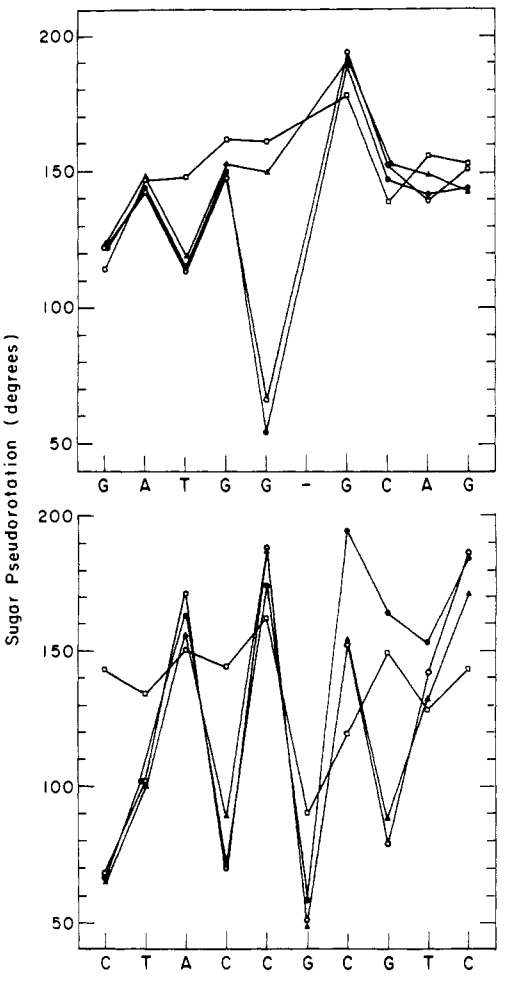


FIGURE 10: Plot of sugar pseudorotation angles versus sequence position: (○) A1; (●) B2; (▲) H; (□) N.

distances calculated at different mixing times increases our overall confidence in the structures obtained with these constraints.

Our results indicate overall similarity between structures that are energy minimized with and without use of the NMR distance constraints. However, the structures do differ in detail, as can be seen from the 0.65-Å rms deviation (Table VI) between the calculated structure and the NMR distances when no constraints are used in the refinement. This deviation is about 0.25 Å larger than the values found when the constraints are used. Introduction of the distance constraints increases the calculated energy, by an amount that is greater than the constraint energy. Hence, there is a conflict between the detailed structural adjustment required by the NMR distance constraints and the adjustment demanded by the energy-minimization algorithm. Extensive simulation of the molecule, including calculation of the NOESY spectrum for representative dynamic states, will be required before the sources of this disagreement can be identified.

CONCLUSIONS

In the work presented here, we have modeled the three-dimensional structure for a duplex containing an extrahelical guanosine, extending our previous work on a duplex of similar sequence containing an extra adenosine. Not only are the general features of the energy-minimized structures in agreement with the two-dimensional NMR data for this molecule, but the similarity of the coordinates obtained by very different paths during the energy minimization gives us added confidence in the ability of molecular mechanics in conjunction

with information obtained from the NMR data to correctly predict the overall features of the duplex. Comparison of previously calculated structures for the A-bulge 9-mer with the current structures for the G-bulge 9-mer indicates that base-stacking energies play an important role in the conformation of the helix at the bulge site. In addition to determining the conformation of the flanking base pairs, evidence from several laboratories suggests that extrahelical purines are generally stacked into the helix, while extrahelical pyrimidines are in equilibrium between stacked and unstacked states (Woodson & Crothers, unpublished results; Patel et al., 1982; Morden et al., 1983; Lomant & Fresco, 1973).

All of the bulge-containing oligomers based on the sequence $^{5'}$d(GATGGGGCAG) examined to date are in the B family, as is the G-bulge-containing duplex presented here. Insertion of the extra guanine into the double helix causes an apparent kink in the helix axis of 18–24° in the calculated structures, which is in excellent agreement with gel mobility experiments on bulge-containing sequences (Rice, 1987). The experiments presented here take one step toward an understanding of the structure of helices containing bulged nucleotides and the relation of that structure to biological function. Determination of solution structures for DNA oligomers is currently limited by the uncertainty in distances obtained from NOESY experiments and to some extent by the computational methods themselves. NOE intensities in nucleic acids are particularly affected by secondary energy transfer, which can cause significant distortion of experimentally derived distance constraints (Lefèvre et al., 1987). Methods for determining interproton distances that take the full spin system into account have been presented (Lefèvre et al., 1987; Keepers & James, 1984) and will hopefully offer a significant improvement in the reliability of structures obtained from NMR data.

Acknowledgments

We thank Abraham Brown and Janet Rice for many helpful discussions and the WERMS computing facility for use of the graphics systems.

Registry No. d(GATGGGCAG)·d(CTGCGCCATC), 111662-61-4.

References

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504–1509.

Chazin, W. J., Wüthrich, K., Hyberts, S., Rance, M., Denny, W. A., & Leupin, W. (1986) *J. Mol. Biol. 190*, 439–453.

Clore, G. M., Gronenborn, A. M., & McLaughlin, L. W. (1985a) *Eur. J. Biochem. 151*, 153–165.

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.

Fazakerley, G. V., Teoule, R., Guy, A., Fritzche, H., & Guschlbauer, W. (1985) *Biochemistry 24*, 4540–4548.

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

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.

Hare, D., Shapiro, L., & Patel, D. J. (1986a) *Biochemistry 25*, 7445–7456.

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

Havel, T. F., Crippen, G. M., & Kuntz, B. D. (1979) *Biopolymers 18*, 73–81.

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

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

Lomant, A. J., & Fresco, J. R. (1973) *J. Mol. Biol. 77*, 345–354.

Macura, F., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95–117.

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

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

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

Patel, D. J., Kozlowski, S. A., Hare, D. R., Reid, B. R., Ikuta, S., Lander, N., & Itakura, K. (1985) *Biochemistry 24*, 926–935.

Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B., & Jones, R. A. (1986) *J. Mol. Biol. 188*, 677–692.

Pearlman, D. A., Holbrook, S. R., Pirkle, D. H., & Kim, S.-H. (1985) *Science (Washington, D.C.) 227*, 1304–1308.

Rao, S. N., Keepers, J. W., & Kollman, P. A. (1984) *Nucleic Acids Res. 12*, 4789.

Rao, S. N., Singh, U. C., & Kollman, P. A. (1986) *J. Am. Chem. Soc. 108*, 2058–2068.

Rice, J. A. (1987) Ph.D. Thesis, Yale University.

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

Skopek, T. R., & Hutchinson, F. (1984) *MGG, Mol. Gen. Genet. 195*, 418–423.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286.

Streisinger, G., Okada, Y., Emrich, J., Neton, J., Tsugita, A., Terzaghi, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77–84.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765–784.

Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 130–134.

Woodson, S. A. (1987) Ph.D. Thesis, Yale University.

Woodson, S. A., & Crothers, D. M. (1987) *Biochemistry 26*, 904–912.

Woodson, S. A., & Crothers, D. M. (1988) *Biochemistry 27*, 436–445.