Structural Influence of RNA Incorporation in DNA: Quantitative Nuclear Magnetic Resonance Refinement of d(CG)r(CG)d(CG) and d(CG)r(C)d(TAGCG)†

T. N. Jaishree,‡ Gijs A. van der Marel,§ Jacques H. van Boom,§ and Andrew H.-J. Wang*,‡

Biophysics Division and Department of Cell & Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, and Gorlaeus Laboratories, Leiden State University, Leiden 2300RA, The Netherlands

Received December 11, 1992; Revised Manuscript Received February 22, 1993

ABSTRACT: RNA and DNA adopt different types of conformations, i.e., A-type with C3′-endo sugar pucker for RNA and B-type with C2′-endo sugar pucker for DNA, respectively. The structural influence of the incorporation of RNA nucleotides into DNA is less understood. In this paper, we present the three-dimensional structures of two RNA-containing oligonucleotides, d(CG)r(CG)d(CG) and d(CG)r(C)d(TAGCG), as determined by the NMR refinement procedure, and assess the possible structural perturbation of DNA induced by RNA. With a single RNA insertion into an octamer DNA, its overall conformation remains as the canonical B-DNA, except that the sugar pucker of the rC3 residue is C3′-endo (pseudorotation angle $P = 3.6°$). In contrast, the hybrid hexamer is neither the pure B-DNA nor the pure A-DNA conformation. Instead, we propose a model in which the DNA parts adopt B conformation, whereas the RNA part adopts A conformation, with the overall conformation closer to A-DNA. To ensure an exhaustive search of the conformational space, the model was subjected to 100-ps simulated annealing with slow cooling or 100-ps molecular dynamics with subsequent quenching. Models obtained at different time points of the trajectories were further subjected to the SPEDREF NOE refinement [Robinson & Wang (1992) *Biochemistry 31*, 3524] and they appeared to arrive at a convergent model (<0.5 Å RMSD for the central four base pairs). The consensus hexamer structure contains a significant discontinuity at the (rG4)p(dC5) step with a base pair tilt angle of 6.7° and roll angle of 11.5°. This discontinuity may be related to the structural "bend" that occurs at the junction of the RNA and DNA helices.

DNA and RNA play different biological roles. Many of their biological activities are closely associated with the preferred conformations that DNA and RNA adopt. In general the biologically relevant conformation for DNA is the B genus right-handed double helix in which the deoxyribose has the C2′-endo (S-type) pucker. RNA double helix, in contrast, adopts A conformation with the ribose in the C3′-endo (N-type) pucker. This observation illustrates the profound consequence of the presence of an additional hydroxyl group at the C2′ position of the ribose having a strong effect on the ability of nucleic acids to form different conformations. An interesting question is what happens structurally when RNA nucleotides are incorporated into a DNA molecule, as in the case of Okazaki fragment (Okazaki et al., 1973) during replication or in the RNA–DNA hybrid during transcription (Kornberg & Baker, 1992).

This question was addressed earlier by the crystal structure analyses of the RNA–DNA hybrids r(GCG)d(TATACGC)[1] (Wang et al., 1980) and r(GCG)d(TATACCC) + d(GGG-TATACGC) (Egli et al., 1992). In both crystals, the hybrid molecules adopted an overall A conformation, although small distortions were noted at the junctions of the RNA–DNA hybrid duplex and the DNA–DNA duplex. Every nucleotide (RNA and DNA) in both molecules adopted the C3′-endo pucker conformation. However, an earlier NMR study of the r(GCG)d(TATACGC) hybrid decamer suggested that not all nucleotides adopt the C3′-endo conformation (Mellema et al., 1983). Specifically, the central four DNA nucleotides have predominantly C2′-endo puckers on the basis of the 1D NMR coupling constant data.

Several DNA–RNA hybrids in which a short segment of pure RNA duplex is flanked by two pure DNA duplexes have been studied by NMR. These include the hexamer d(CG)r-(CG)d(CG) (Haasnoot et al., 1983) and two dodecamers, d(CGCG)r(AAUU)d(CGCG) and d(CGCG)r(AUAU)d-(CGCG) (Chou et al., 1991). In those studies the information on the local conformations of the individual nucleotide, specifically the sugar pucker conformations, was obtained. However, the detailed helical parameters were not available due to the nature of the structural analysis in which the use of NOE data was semiquantitative. Another hybrid, r(GCG)d(CGC), has been studied by laser Raman study and it was shown to be in the A-DNA family (Benevides et al., 1986).

In this paper, we analyze two RNA–DNA hybrids, d(CG)r-(C)d(TAGCG) and d(CG)r(CG)d(CG), to probe the structural effects of RNA on DNA in a quantitative manner. The self-complementary octamer forms a DNA duplex with two RNA substitution sites, whereas the hexamer duplex consists of three distinct segments with the pure RNA duplex sandwiched between two DNA duplexes. Here we collected the complete 2D-NOESY and COSY data and refined the structures by the NOE spectral-driven refinement procedure SPEDREF (Robinson & Wang, 1992) to extract quantitative

† This work was supported by grants from NIH (GM-41612 and CA-52506) to A.H.-J.W.

* To whom correspondence should be addressed.

‡ University of Illinois at Urbana–Champaign.

§ Leiden State University.

[1] Abbreviations: A, T, G, and C, adenine, thymine, guanine, and cytosine or their respective nucleotides in DNA; NMR, nuclear magnetic resonance; 2D-NOESY, two-dimensional nuclear Overhauser effect spectroscopy; 2D-COSY, two-dimensional correlation spectroscopy; pr. tw., propeller twist; RMSD, root mean square deviation; SA, simulated annealing; MD, molecular dynamics.

0006-2960/93/0432-4903$04.00/0

information regarding the global as well as local conformation. Simulated annealing and molecular dynamics were both used to obtain a more exhaustive sample of different starting models.

MATERIALS AND METHODS

The oligonucleotides were synthesized by the hydroxybenzotriazole phosphotriester method described earlier (van Boeckel et al., 1981). Samples for ^{1}H NMR analysis were prepared in D_2O as described earlier (Robinson & Wang, 1992). The solution contained 2.5 mM duplex of both molecules with 150 mM NaCl and 50 mM sodium phosphate buffer (pH = 7.0) in the 0.5-mL sample. The ^{1}H NMR spectra were recorded on a GE GN500 500-MHz spectrometer. The chemical shifts (in parts per million) are referenced to the HDO peak which is calibrated to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) at different temperatures. The 2D NOESY were collected by the method of States et al. (1982) with a composite pulse NOESY sequence. The spectra were recorded with 512 t_1 complex blocks of 2048 complex points each (in the t_2 dimension) and averaged for 16 scans/block. The total recycle time for the 200-ms NOESY experiment was 2.76 s. The 2D data sets were processed with the program FELIX (Hare Research, Woodinville, WA) on Silicon Graphics workstations using the truncation apodization described earlier (Robinson & Wang, 1992) and exponential multiplication at 5 Hz. The intensities of the NOE cross peaks were measured by the program MYLOR (Robinson & Wang, 1992). There were 592 and 789 NOE cross peaks considered to be above the noise level for the hexamer and octamer hybrids, respectively, and they were used in the refinement. An iterative spectral-driven procedure SPEDREF (Robinson & Wang, 1992) using the full-matrix relaxation theory was used for the refinement.

Standard B-DNA and A-DNA and various hybrid models were prepared using QUANTA (Version 3.2.3, Polygen). Molecular dynamics (MD) and simulated annealing (SA) were carried out using the program X-PLOR (Brunger, 1992). Unless otherwise stated, we performed the simulated annealing without the NOE constraints but applied weak harmonic restraints to the end base pairs to prevent the molecule from being severely distorted, thereby losing chemical meaning. The SA was carried out by performing molecular dynamics at 1000 K for 100 ps. Models along the trajectory were selected every 10 ps and they were slowly cooled by 25 K every 25 fs to a temperature of 300 K. These models were then energy-minimized by conjugate gradient and used for further SPEDREF refinement.

RESULTS

The nonexchangeable proton 1D- and 2D-NOESY spectra of d(CG)r(CG)d(CG) and d(CG)r(C)d(TAGCG) are shown in Figures 1 and 2, respectively. The assignment of the resonances of the two hybrids was straightforward using the sequential assignment procedure (Hare et al., 1983). Figures 3 and 4 show some of the expanded regions of the 2D-NOESY spectra in which the key assignment pathways using the NOE cross peaks are indicated. The chemical shifts of all nonexchangeable protons are listed in Table I. Those of the hexamer hybrid are in agreement with the previous assignments of Haasnoot et al. (1983).

Variation of Starting Models of d(CG)r(CG)d(CG). Three initial models, B-DNA, A-DNA, and DNA–RNA–DNA (D–R–D), were tested for the SPEDREF NOE refinement. They were refined to *R*-factors of 25%, 26%, and 22.5% for the B-DNA, the A-DNA, and the D–R–D hybrid, respectively. [The *R*-factor is defined as $\sum|N_o - N_c|/\sum|N_o|$ and its implication in the assessment of models has been described more fully in Robinson and Wang (1992).] The results indicated that the hybrid model is a better representation to account for the NOE data. However, one of the major concerns in the refinement of DNA structure using NOE data is whether a global-minimum structure has been attained. We have addressed this by taking several approaches. Since we have shown that the better model (*R*-factor = 22.5%) so far for the hexamer hybrid is a D–R–D duplex (HEXDRD), we assumed that this model may be used as a starting point for further conformational scanning. The three approaches we took are described below.

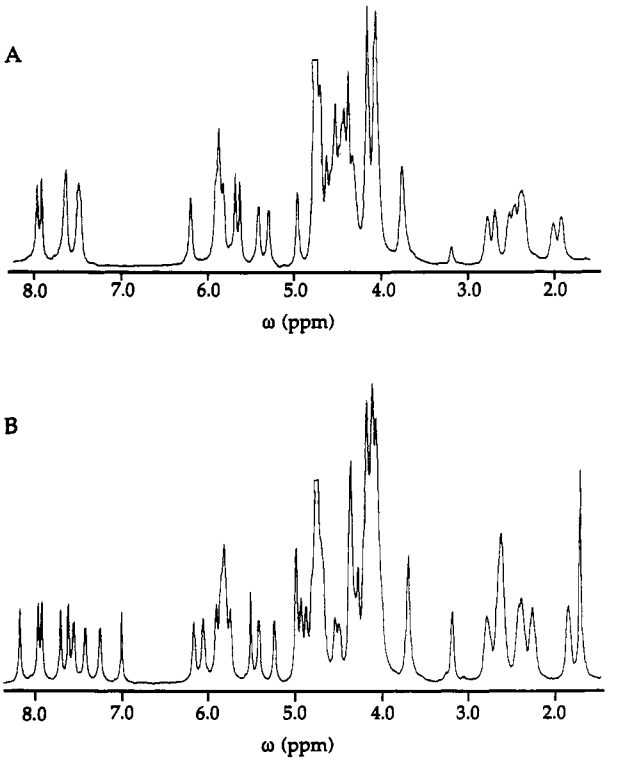

FIGURE 1: One-dimensional nonexchangeable proton 500-MHz NMR spectra of (A) d(CG)r(CG)d(CG) and (B) d(CG)r(C)d(TAGCG).

(1) SA with Slow Cooling. To probe the variability of the backbone conformation, we took the HEXDRD model and subjected it to 100-ps simulated annealing without NOE constraints but with all bases (not including the sugar–phosphate backbone) fixed in position by weak harmonic constraints. In this way, the backbone had the opportunity to move into different conformations. We sampled the molecular species at every 10 ps along the simulation trajectory and compared their structures. We found that the structure at 10 and 20 ps were significantly different (RMSD = 1.2 Å), but those at 20 ps and later were not. The structures at 0, 10, and 20 ps (Figure 5A) were used as representative families of structures and were further refined by SPEDREF to *R*-factors of 22.5%, 22.1%, and 22.7%, respectively. The refined structures are shown in Figure 5B. The 22.1% model (denoted as HEXDRD221) was chosen as an archetypal model for the hexamer hybrid. We also subjected the HEXDRD model to 100-ps SA with only the end base pairs fixed. The results were similar to the previous approach.

To further explore the conformational landscape, the 22.1% model was used for another 100-ps SA, but with NOE constraints. The models along the trajectory (every 10 ps)

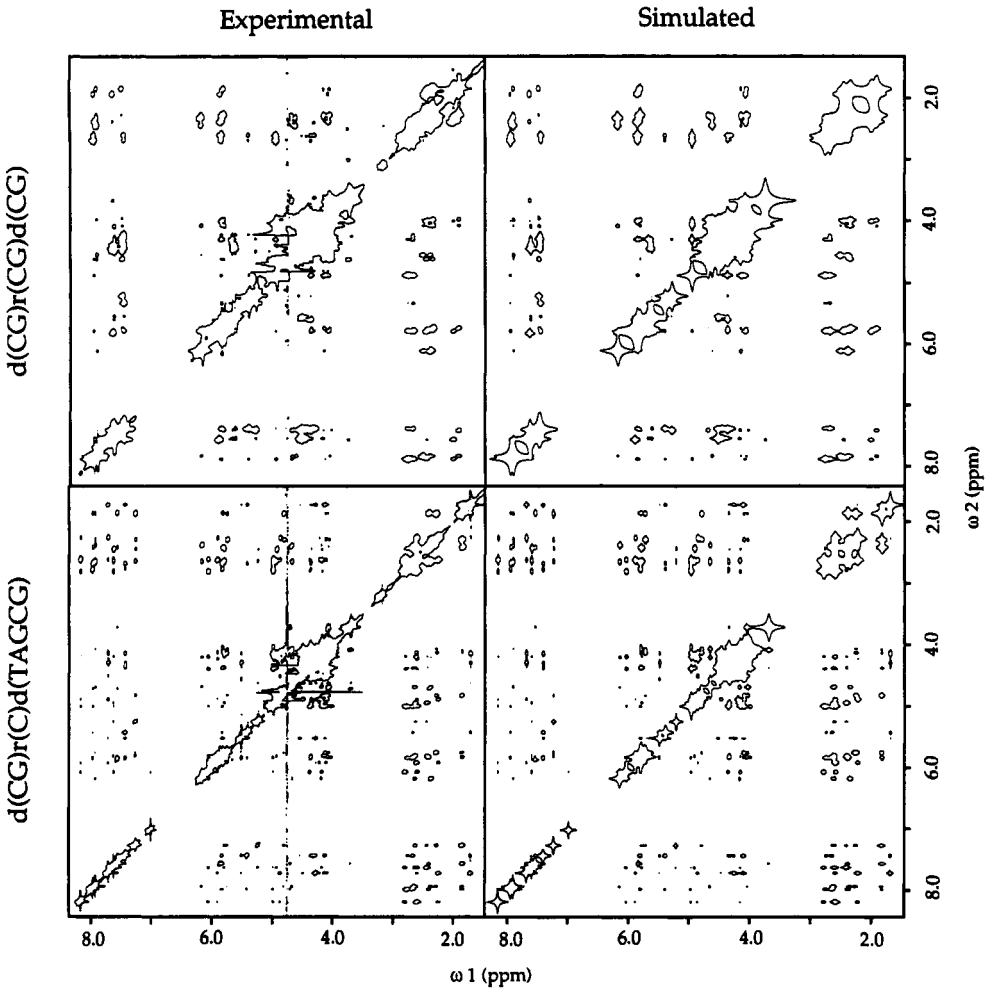

FIGURE 2: Experimental and simulated two-dimensional nonexchangeable proton NOESY spectra of d(CG)r(CG)d(CG) (top two panels), and d(CG)r(C)d(TAGCG) (bottom two panels).

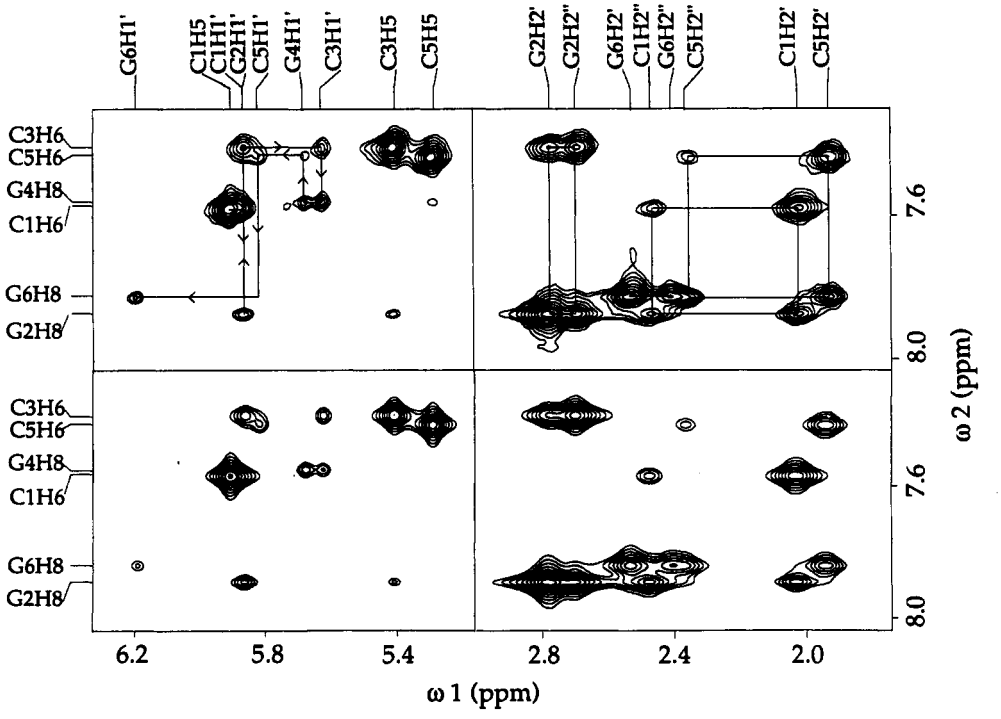

FIGURE 3: Expanded regions of the experimental (top panels) and simulated (bottom panels) 2D-NOESY spectra of d(CG)r(CG)d(CG). Key sequential assignment pathway between the aromatic and H1′ protons is indicated with arrow lines.

were slowly cooled to 300 K and energy-minimized to obtain 10 different models. Inspection of these models revealed that they are closely related, with maximum RMSD among them being 0.4 Å (all six nucleotides) (Figure 5C). The SPEDREF

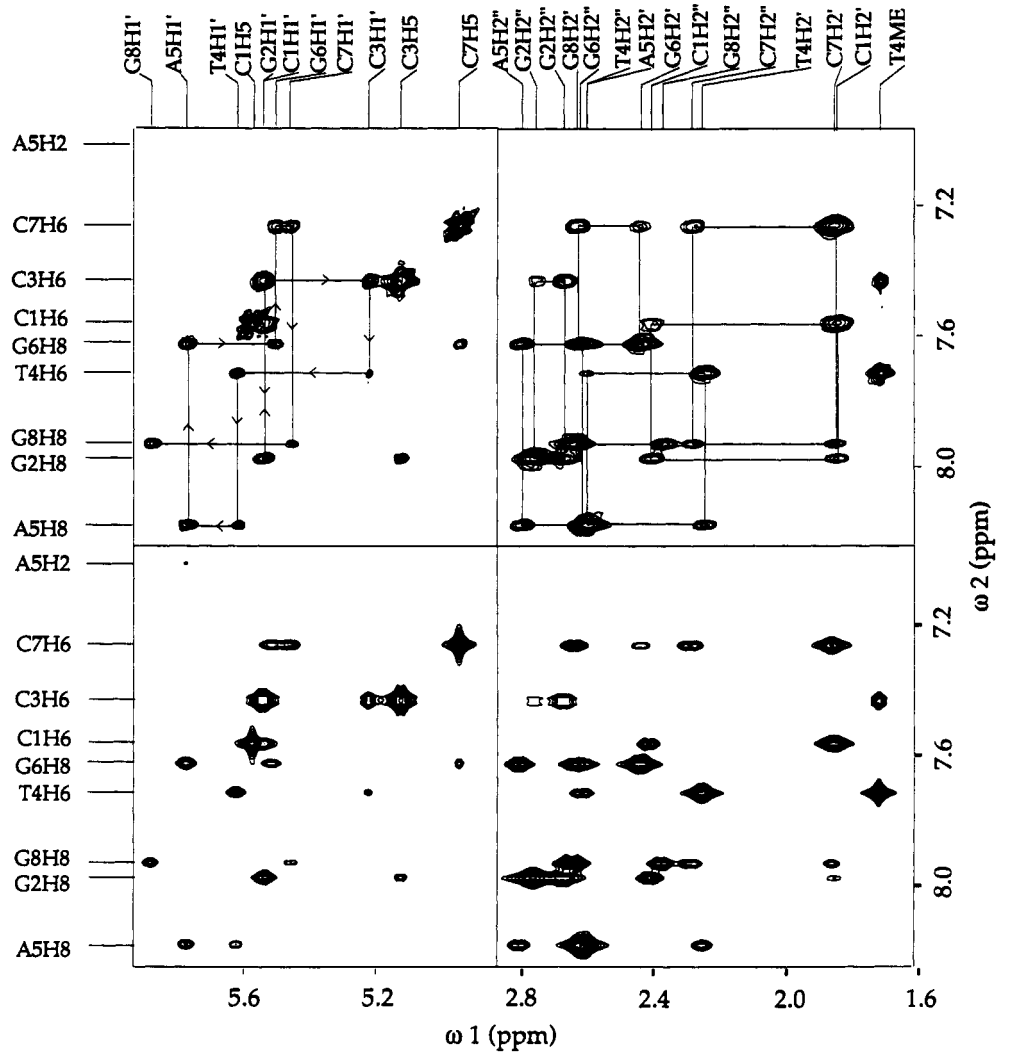

FIGURE 4: Expanded regions of the experimental (top panels) and simulated (bottom panels) 2D-NOESY spectra of d(CG)r(C)d(TAGCG).

Table I: Proton Chemical Shifts for d(CG)r(CG)d(CG) and d(CG)r(C)d(TAGCG)[a]

	H5	H6/H8	H1′	H2′	H2″	H3′	H4′	H5′	H5″
				(A) d(CG)r(CG)d(CG)					
Cyt 1	5.823	7.562	5.783	1.939	2.382	4.001	4.001	3.670	3.670
Gua 2		7.880	5.777	2.689	2.604	4.875	4.298	4.007	4.062
Cyt 3	5.325	7.378	5.539	4.449		4.378	4.378	4.260	4.260
Gua 4		7.541	5.594	4.292		4.445	4.445	4.094	4.094
Cyt 5	5.207	7.408	5.732	1.849	2.272	4.611	4.290	3.948	4.084
Gua 6		7.828	6.107	2.440	2.318	4.545	4.078	3.980	3.980
	H5/H2	H6/H8	H1′	H2′	H2″	H3′	H4′	H5′	H5″	ME
				(B) d(CG)r(C)d(TAGCG)						
Cyt 1	5.860	7.559	5.819	1.852	2.407	4.365	4.073	3.708	3.708	
Gua 2		7.970	5.829	2.669	2.762	4.988	4.374	4.014	4.117	
Cyt 3	5.412	7.428	5.509	4.281		4.129	4.546	4.353	4.353	
Thy 4		7.710	5.409	2.248	2.612	4.874	4.493	4.109	4.219	1.718
Ade 5	7.007	8.176	6.060	2.604	2.804	4.985	4.383	4.178	4.178	
Gua 6		7.621	5.802	2.435	2.636	4.927	4.357	4.179	4.179	
Cyt 7	5.236	7.257	5.748	1.860	2.283	4.810	4.118	4.056	4.056	
Gua 8		7.925	6.168	2.641	2.378	4.674	4.192	4.053	4.053	

[a] Proton chemical shifts are given in parts per million.

refinement of these models resulted in *R*-factors of approximately 22%.

(*2*) *High-Temperature MD with Quenching.* To ensure that the simulated annealing with slow cooling process did not miss some other conformational minima, we took an alternative (quenching) approach. A similar molecular dynamics with quenching approach has been used recently in the analysis of a d(CGCTGCGGC) hairpin structure (Gupta et al., 1993) and it appeared to be able to find clusters of conformers with fairly drastically different global DNA structures, in particular in the loop region of the hairpin molecule.

We again used the 22.1% model for a 100-ps MD at 1000 K with NOE constraints, but then the models along the trajectory were energy-minimized at every 10 ps to obtain 10 different models. These new models (Figure 5D) revealed that they are also closely related, but with slightly larger

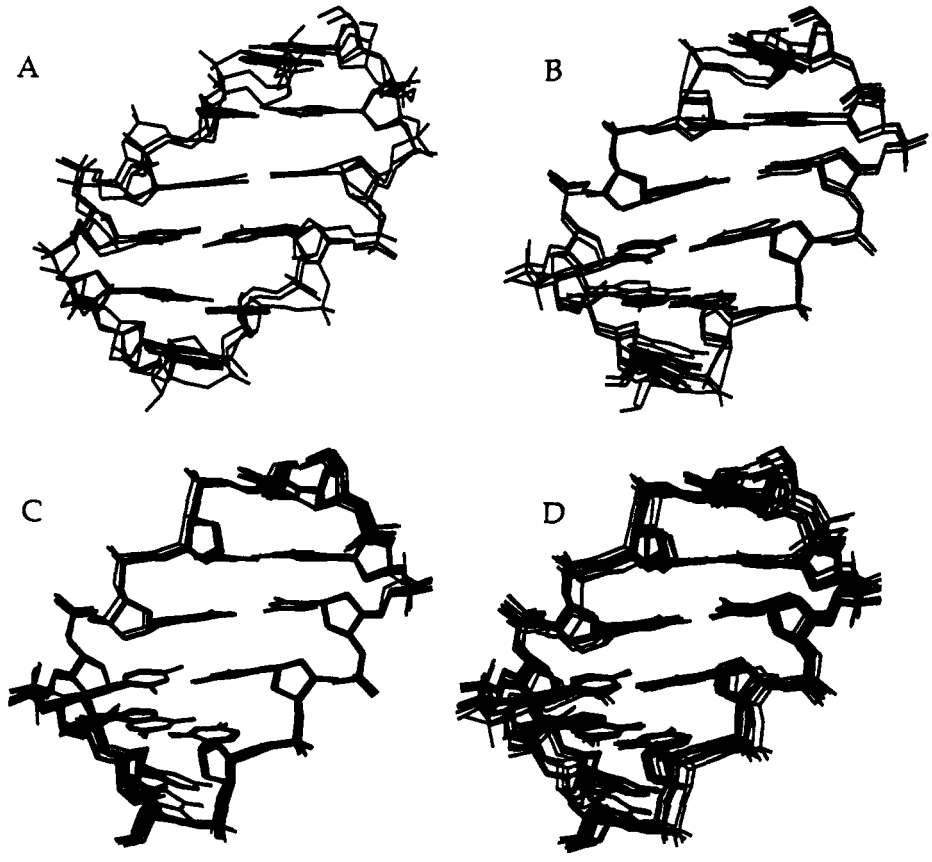

FIGURE 5: Different models of the d(CG)r(CG)d(CG) hybrid. (A) Starting models obtained from the trajectory of 100-ps simulated annealing at 0, 10, and 20 ps. (B) Refined models. (C) Ten superimposed models obtained from the trajectory of 100-ps simulated annealing with slow cooling at every 10 ps with NOE constraints and refined by SPEDREF afterward. (D) Ten superimposed models obtained from the trajectory of 100-ps molecular dynamics with NOE constraints, energy-minimized at every 10 ps and refined by SPEDREF afterward.

RMSD among them (maximum RMSD 0.7 Å for all six base pairs and 0.4 Å for the middle four base pairs). The SPEDREF refinement of these models resulted in *R*-factors of approximately 22%.

(*3*) *Underwound and Overwound Models.* The initially SPEDREF-refined HEXDRD model has a helical pitch of 12 bp/turn. We deliberately overwound and underwound the hexamer duplex by 40° and then energy-minimized the two models. They had respective helical pitches of 9.0 and 16.6 bp/turn and were subjected to 50-ps simulated annealing. Representative models were chosen from the two trajectories, resulting in the HEXDRD-OVER and HEXDRD-UNDER starting models. These three starting models have large RMSDs among them (~2 Å) (Figure 6A). When these three models were subjected to 40 cycles of SPEDREF refinement, they converged to have RMSDs of 0.8–1.0 Å for all six residues or 0.5 Å for the central four residues (Figure 6B). The *R*-factors were 22.5% for all three (HEXDRD, HEXDRD-OVER, and HEXDRD-UNDER) models.

It is interesting to note that the initially overwound and underwound models had converged to agree with the normal model. This suggested that the hybrid hexamer preferred a double helix of ~12 base pairs per turn and SPEDREF NOE refinement was sufficiently powerful to drive them to convergence. We noticed that the refined structure of the starting underwound model deviates substantially at the C5pG6 step from those of the normal and overwound models. Nevertheless, the NOE *R*-factors were identical, indicating that within the definition of NOE values the terminal base pairs appeared to have greater range of motional flexibility (fraying) or insufficient NOE data were available for the terminal base pairs. Further heteronuclear T_1 and T_2 measurements may distinguish these two possibilities.

RNA–DNA Hexamer Hybrid Structure. On the basis of the above analyses, we concluded that HEXDRD221 (Figure 7A) may be used as a representative NOE-refined structure of the hexamer hybrid. The complete simulated 2D-NOE spectrum on the basis of this model is shown in Figure 2. Some expanded regions of the 2D-NOESY spectrum are shown in Figure 3.

The hexamer hybrid structure has an overall conformation lying between the canonical A-DNA and B-DNA, as seen from the selected helical parameters (Table II). The helix *x*-displacement is between –3.6 and –4.3 Å, closer to that of A-DNA. However, the base pair tilt angle is small, comparable to that in B-DNA. It is interesting to note that there are conformational irregularities at the junction between DNA and RNA segments. The tilt angle of G2–C11 base pair is –6.7° and the roll angle of rC3–rG10 is 11.5°. This data suggests that RNA–DNA hybrids such as the Okazaki double helix may have a significant distortion at the junction of the segment of RNA–DNA hybrid and the segment of DNA–DNA duplexes.

The helix distortion in the hybrid is likely due to the preference of ribose residue in the nucleic acid backbone to adopt the N-type pucker. The analysis of the pseudorotation angles (*P*) of various NOE-refined models confirms this. As expected, the two RNA residues have very low *P* values, e.g., 0.3° and 1.5° for rC3 and rG4, respectively, in the HEXDRD221 model. Interestingly, only the first two DNA residues, dC1 (P = 158.0°) and dG2 (P = 154°), have the normal S-type sugar pucker. The two DNA residues on the

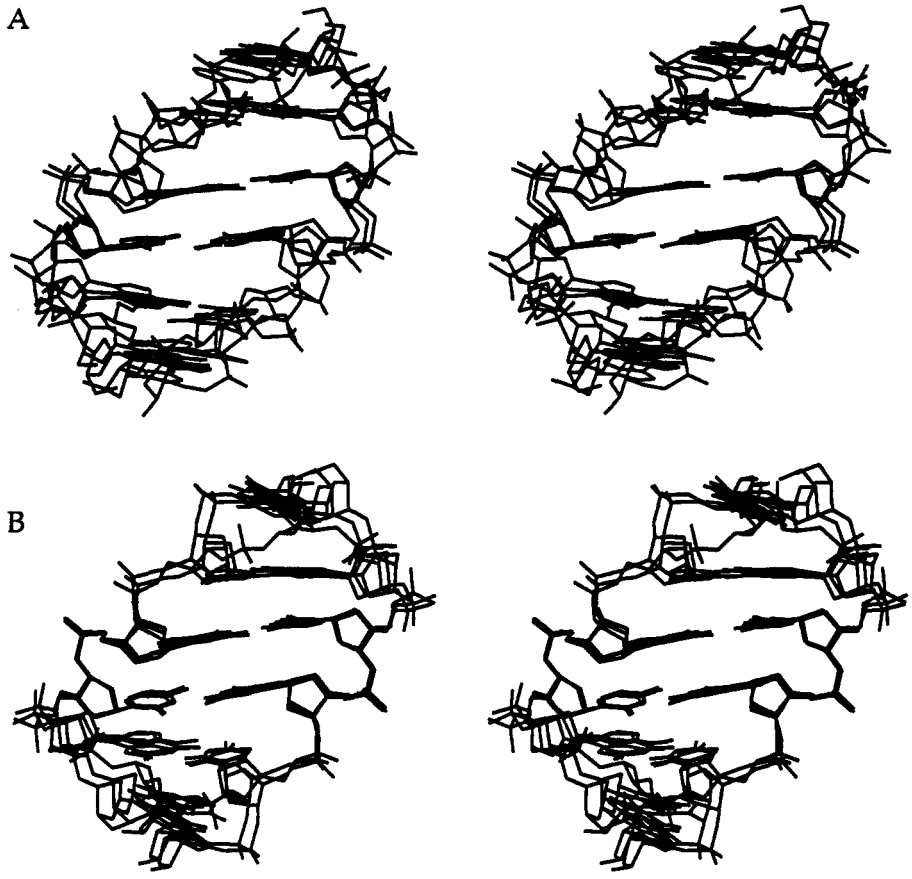

FIGURE 6: Stereoscopic skeletal drawings of the three models of the d(CG)r(CG)d(CG) hybrid, HEXDRD, HEXDRD-OVER, and HEXDRD-UNDER. (A) Starting models. (B) Refined models.

3′-end have unusual *P* values. The dC5 ribose has a low *P* of 7.2°, while the dG6 has an intermediate *P* of 81.3°. This indicates that *the deoxyribose of the DNA residue immediately following the RNA segment (on the 3′ side) is forced to adopt a near N-type pucker.*

We have calculated the coupling constants *J* between the H1′ and H2′/H2″ protons on the basis of the *P* values of the HEXDRD221 model (Table III). Those calculated *J* values agree very well with the measured *J* values using PE-COSY (purge-COSY) data (spectrum not shown). The analysis of the *J* values here also agrees with the conclusion of a similar analysis earlier using a 1D spectrum (Haasnoot et al., 1983). The latter study proposed a two-state model for each nucleotide's sugar pucker conformation. To best fit the coupling constant data, the percentages of S-type conformer for the six nucleotides were 75%, 100%, 0%, 0%, 40%, and 65%, respectively, for dC1, dG2, rC3, rG4, dC5, and dG6, with the S- and N-conformers in rapid exchange on the NMR time scale (Haasnoot et al., 1983).

Structure of d(CG)r(C)d(TAGCG). Since the octamer contains only a single RNA residue, the starting model was a simple modification of B-DNA by placing a 2′-OH group onto the sugar of the C3 residue with the energy minimization. The SPEDREF refinement of this model converged to an *R*-factor of 20.1%. As discussed above for the hexamer refinement study, the additional SA process yielded clusters of closely related conformations. We decided that no further SA study is needed in the octamer hybrid. We have also tested the A-DNA starting model from which a high *R*-factor of 28% was obtained. Therefore the model with global A conformation was rejected.

The refined structure (Figure 7B) retained the characteristics of B-DNA with the helix *x*-displacement of 0.3 Å. Most of the deoxyriboses adopt the S-type pucker (average *P* = 150°), with slight deviations for the T4 (*P* = 140°) and C7 (*P* = 100°) residues. The ribose of rC3 residue is of the N-type pucker with *P* = 3.6°. Therefore, even the insertion of a single RNA residue in DNA persists in the N-type pucker and causes some perturbation of the conformation of the neighboring residues (intrastrand T4 residue and interstrand C7 residue).

Finally, to address the mobility of each nucleotide in the molecule, we measured the T_1 relaxation time by the inversion recovery technique. We found that the rC3 ribose has a substantially longer T_1 (2.5–3.5 s) than the remaining DNA deoxyriboses' T_1 (1.0–2.0 s). The adoption of the N-type pucker seemed to rigidify the sugar ring significantly. In comparison, the hexamer hybrid had generally higher average T_1 values (2.5–4.0 s) for the two middle RNA residues (and also slightly higher values for the DNA residue on the 3′ side of the RNA residues). T_1 values of the two hybrids were measured for all nonexchangeable protons. For pure B-DNA duplex like d(CGATCG), the T_1's are in the range of 1.0–2.0 s (Robinson et al., 1992).

DISCUSSION

Conformational polymorphism of DNA is now well documented. In addition to the prevalent B-DNA, other stable alternative conformations including A-DNA and Z-DNA have been shown to exist. In contrast, RNA seems to have a substantially limited range of possible secondary structures. So far RNA is known to adopt only the A-family structure in which the large base pair tilt angle (~19°), the large helix displacement (–4 Å), the wide and shallow minor groove, and finally the narrow and deep major groove are its characteristic

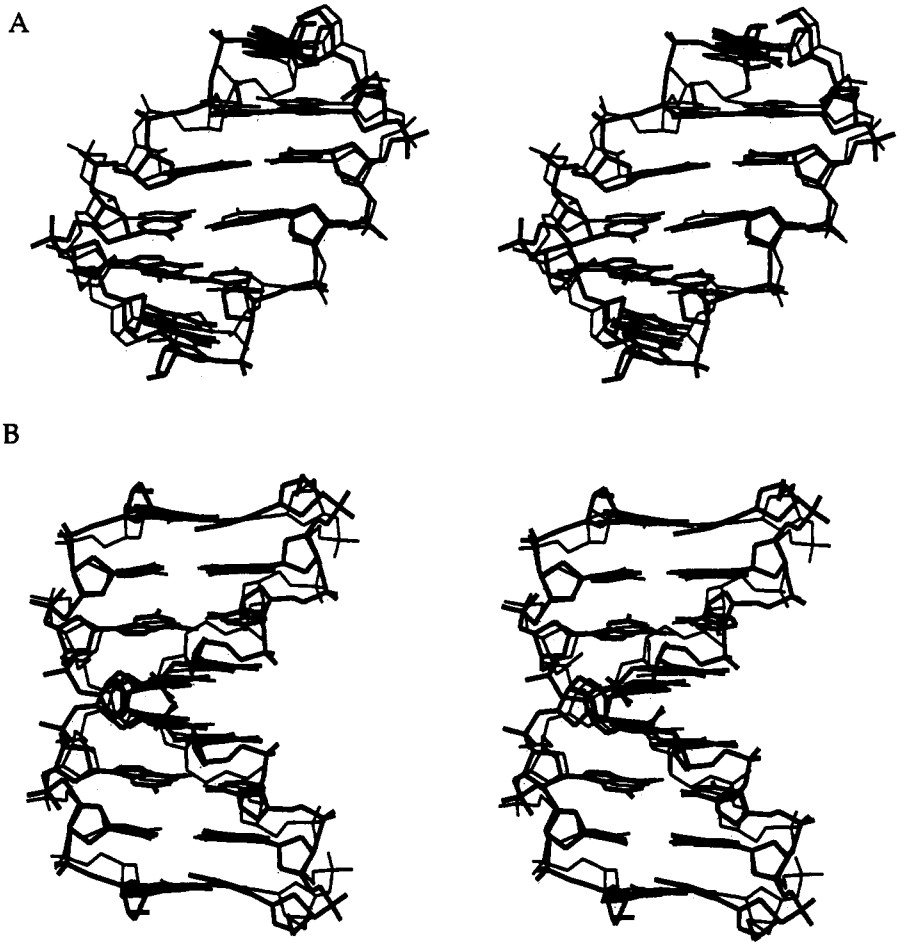

FIGURE 7: Superposition of the starting model (thin lines) and the refined model (thick lines). (A) d(CG)r(CG)d(CG) hybrid. The refined model is HEXDRD221. (B) d(CG)r(C)d(TAGCG).

Table II: Selected Helical Parameters of the HEXDRD221 Model

base pair	roll (deg)	tilt (deg)	D_x[a] (Å)
C1–G12	−32.2	−2.2	−4.3
G2–C11	−5.4	−6.7	−4.0
C3–G10	11.5	−0.1	−3.5
G4–C9	−5.0	6.7	−3.5
C5–G8	−32.7	2.2	−4.0
G6–C7			−4.3

[a] D_x is the x-displacement.

Table III: ^3J Coupling Constants for d(CG)r(CG)d(CG)

		^3J(H1′–H2′) (Hz)		^3J(H1′–H2″) (Hz)	
residue	P	exp[a]	theory[a]	exp[a]	theory[a]
dC1	158.0	8.3	10.1	6.2	6.1
dG2	154.0	10.0	10.2	5.4	5.7
rC3	0.3	<1	<1		
rG4	1.5	<1	<1		
dC5	7.2	5.0	2.4	6.5	6.8
dG6	81.3	7.1	7.4	6.6	7.4

[a] Experimental values were measured from the 2D-PE-COSY spectrum, and theoretical values were calculated on the basis of a modified Karplus equation (Haasnoot et al., 1980).

global helical features. This unique property of RNA is likely associated with the propensity of ribose for adopting the C3′-endo pucker conformation.

The influence of an RNA strand on the stability of nucleic acid structures, including duplexes and triplexes, have been studied recently (Roberts & Crothers, 1992). That work suggested that an RNA–RNA hybrid is the most stable duplex among the different combinations of duplexes. In contrast, RNA–DNA–RNA (third strand) is the most stable triplex structure. This data pointed out the intricacy of the role RNA plays in influencing the stability of nucleic acid structures. It is important to elucidate the basis of this at the molecular level.

Interestingly, very few RNA crystal structures are available. The crystal structures of a number of transfer RNA structures and their complexes with tRNA synthetases have been determined and they displayed the remarkable ability of RNA in folding into complex tertiary structures (Saenger, 1984; Steitz, 1990). Nonetheless, the stem regions of tRNA are in general in the A conformation. The only pure RNA oligonucleotide duplex structure is that of r[U(UA)$_6$A] which is a "typical" A structure (Dock-Bregeon et al., 1989).

It is less clear what happens if RNA is inserted into a DNA structure. The crystal structures of two related RNA–DNA hybrids, r(GCG)d(TATAGCG) and r(GCG)d(TATAGGG)+d(CCCTATACGC), have been solved (Wang et al., 1980; Egli et al., 1992). They adopted the canonical A conformation with most of the sugars in the N-type pucker conformation. However, the results from the NMR study of the r(GCG)d(TATACGC) decamer hybrid clearly indicated significant components (40–65%) of S-type pucker in the DNA segment of the hybrid (Mellema et al., 1983). It is likely that the solution structure of the decamer is more heterogeneous and only a small fraction of the conformation population is the same as the crystal structure. This observation seems to be quite common. A number of DNA oligomers have been crystallized in the A family (Wang et al., 1982; McCall et al., 1986; Frederick et al., 1989), but the solution study had

indicated otherwise (Clark et al., 1990). This should not be surprising, as DNA in solution exists as an ensemble of many structures, including principally the B-DNA family and a small fraction of other conformers (e.g., A or Z, dependent on the sequences). The crystallization process may select out the minor species of conformer for the formation of crystal lattice.

Not unlike the above situation, the crystal structure of the hexamer hybrid d(CG)r(CG)d(CG) has been determined at 1.5-Å resolution and shown to be in the left-handed Z conformation (Teng et al., 1989). The structural analyses revealed that the O2′ hydroxyl group of the ribosylcytosine nucleoside residue forms a hydrogen bond to the amino group of the guanine in the 5′-(dG)p(rC) step. However, it is clear from the present NMR study that there is no evidence of Z-DNA/Z-RNA conformation in solution for the hexamer hybrid. Instead, it adopts a slightly distorted structure intermediate between A-DNA and B-DNA. There is a slight discontinuity between the RNA and DNA steps, probably induced by the persistence of RNA in the A conformation (N-type pucker) and the adoption by the 3′-DNA residue at the junction of the A conformation. We noticed that the DNA residue that precedes an RNA residue was not affected and adopted an S-type sugar conformation. Our data are consistent with earlier NMR studies (Haasnoot et al., 1983; Mellema et al., 1983; Chou et al., 1991).

The discontinuity at the DNA–RNA junction of the hexamer hybrid, i.e., tilt angle of 6.5° and roll angle of 11°, is interesting, as it may serve as a model for understanding the RNA–DNA junction of polymer. Selsing et al. (1978, 1979) have suggested by model-building studies that there was a large kink of 29° between the RNA and DNA segments. Our study indicated that the extent to which the helix axis is kinked is not as large. However, the hexamer hybrid may be too short to have the full extent of structural effect exerted by the RNA. A quantitative analysis of other hybrids, consisting of longer RNA and DNA segments, not unlike the two dodecamers mentioned earlier (Chou et al., 1991), should provide a more definitive answer.

In the case of the single-substituted octamer d(CG)r(C)d(TAGCG), the perturbation is more subtle. The rC3 residue persisted in the N-type pucker and its structural influence seemed to propagate only to the immediate neighboring residues. It is somewhat surprising that the DNA backbone did not force the single rC3 ribose to accommodate the S-type pucker. Also, from the T_1 data we believe that it is rigidly held in the N-type. The only time that RNA adopted the S-type pucker was either when the RNA is in a tight space (e.g., making a sharp turn in tRNA) (Rich et al., 1979) or in Z-RNA (Teng et al., 1989).

Egli et al. (1992) noted that a single RNA substitution in the d(GCGTATACGC) decamer made it crystallize in the A conformation. But again, the crystal packing forces may have a significant role in this. A case in point is that an unrelated sequence of the pure DNA decamer d(CCCGGCCGGG) (Ramakrishnan & Sundaralingam, 1992) also crystallized as A-DNA in the same $P2_12_12_1$ orthorhombic lattice as the r(GCG)d(TATACGC) hybrid. This observation suggested that this crystal lattice is particularly conducive for decamer oligonucleotides in the A conformation. A relevant analogy is that DNA dodecamers with sequences like d(CGCNNNNNNGCG), where N is A/T nucleotides, tend to crystallize in the B conformation with the central six residues adopting a narrow minor groove. Our NMR data of the hybrids indicated that a single substitution causes only a very localized distortion of the B-DNA backbone due to the strong preference of ribose for adopting the N-type pucker. Nonetheless, this small perturbation is sufficient to shift the equilibrium of various conformers to favoring the A family, thereby facilitating its crystallization in the case of the single-substituted d(GCGTATACGC) (Egli et al., 1992).

In conclusion, we show that RNA residues can be readily inserted into a DNA double helix. The structural influence depends on the length of the RNA segment. A single insertion causes only local perturbations, but a longer stretch of RNA in DNA will generate significant discontinuity at the junction of DNA and RNA helixes and possibly a kink of the helix direction. This type of distortion may be recognized by proteins.

NOTE ADDED IN PROOF

A new NMR study on a hybrid duplex, r(GCCA)d(CTGC)+d(GCAGTGGC), has just appeared (Salazar et al., 1993).

ACKNOWLEDGMENT

We thank Dr. H. Robinson for assistance in the use of SPEDREF.

REFERENCES

Benevides, J. M., Wang, A. H.-J., Rich, A., Kyogoku, Y., van der Marel, G. A., van Boom, J. H., & Thomas, G. J., Jr. (1986) *Biochemistry 25*, 41–50.

Brunger, A. T. (1992) X-PLOR, Version 3.0, The Howard Hughes Medical Institute and Yale University, New Haven, CT.

Chou, S. H., Flynn, P., Wang, A., & Reid, B. (1991) *Biochemistry 30*, 5248–5257.

Clark, G. R., Brown, D. G., Sanderson, M. R., Chwalinsky, T., Neidle, S., Veal, J. M., Jones, R. L., Wilson, W. D., Zon, G., Garman, E., & Stuart, D. I. (1990) *Nucleic Acids Res. 18*, 5521–5528.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekman, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., & Zhurkin, V. B. (1989) *Nucleic Acids Res. 17*, 1797–1803.

Dock-Bregeon, A. C., Chevier, B., Podjarny, A., Johnson, J., De Bear, J. S., Gough, G. R., Gilham, P. T., & Moras, D. (1989) *J. Mol. Biol. 209*, 459–474.

Egli, M., Usman, N., Zhang, S., & Rich, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 534–538.

Frederick, C. A., Quigley, G. J., Teng, M. K., Coll, M., van der Marel, G. A., van Boom, J. H., Rich, A., & Wang, A. H.-J. (1989) *Eur. J. Biochem. 181*, 295–307.

Gupta, G., Garcia, A., & Hiriyanna, K. T. (1993) *Biochemistry 32*, 948–960.

Haasnoot, C. A. G., de Leeuw, F. A. A. M., & Altona, C. (1980) *Tetrahedron 36*, 2783–2792.

Haasnoot, C. A. G., Westerink, G. A., van der Marel, G. A., & van Boom, J. H. (1983) *J. Biomol. Struct. Dyn. 1*, 131–149.

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

Kornberg, A., & Baker, T. A. (1992) *DNA Replication*, W. H. Freeman, New York.

Mellema, J. R., Haasnoot, C. A. G., van der Marel, G. A., Wille, G., van Boeckel, C. A. A., van Boom, J. H., & Altona, C. (1983) *Nucleic Acids Res. 11*, 5717–5738.

Okazaki, R., Sugino, A., Hirose, S., Okazaki, T., Imae, Y., Kainuma-Kuroda, R., Ogawa, T., Arisawa, M., & Kurosawa, Y. (1973) in *DNA Synthesis in vitro* (Wells, R. D., & Inman, R. B., Eds.) pp 83–106, University Park Press, Baltimore, MD.

Ramakrishnan, B., & Sundaralingam, M. (1992) Abstracts of American Crystallography Association Annual Meeting, Poster PA89.

Rich, A., Quigley, G. J., & Wang, A. H.-J. (1979) in *Stereodynamics of Molecular Systems* (Sarma, R. H., Ed.) pp 315–330, Pergamon Press, New York.

Roberts, R. W., & Crothers, D. M. (1992) *Science 258*, 1463–1466.

Robinson, H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 3524–3533.

Robinson, H., van der Marel, G. A., van Boom, J. H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 10510–10517.

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

Salazar, M., Champoux, J. J., & Reid, B. R. (1993) *Biochemistry 32*, 739–744.

Selsing, E., Wells, R. D., Early, T. A., & Kearns, D. R. (1978) *Nature 275*, 249–250.

Selsing, E., Wells, R. D., Alden, C. J., & Arnott, S. (1979) *J. Biol. Chem. 254*, 5417–5422.

Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W. B. T., Salisbury, S. A., & Viswamitra, M. A. (1983) *J. Mol. Biol. 166*, 183–201.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.

Steitz, T. (1990) *Q. Rev. Biophys. 23*, 205–280.

Teng, M.-k., Liaw, Y.-C., van der Marel, G. A., van Boom, J. H., & Wang, A. H.-J. (1989) *Biochemistry 28*, 4923–4928.

Van Boeckel, C. A. A., van der Marel, G. A., Wille, G., Westerink, H., Wang, A. H.-J., Mellema, J. R., Altona, C., & van Boom, J. H. (1981) *Recl. Trav. Chim. Pays-Bas. 100*, 389–390.

Wang, A. H.-J., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A. A., & Rich, A. (1980) *Nature 299*, 601–604.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1981) *Science 211*, 171–176.