Structure of Chimeric Duplex Junctions: Solution Conformation of the Retroviral Okazaki-like Fragment r(ccca)d(AATGA)·d(TCATTTGGG) from Moloney Murine Leukemia Virus[†,‡]

Miguel Salazar,[*,§] Oleg Yu. Fedoroff,[||] and Brian R. Reid[*,||,⊥]

Drug Dynamics Institute, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, and Chemistry and Biochemistry Departments, University of Washington, Seattle, Washington 98195

Received December 8, 1995; Revised Manuscript Received March 26, 1996[⊗]

ABSTRACT: We have determined the solution structure of the synthetic chimeric duplex r(ccca)d(AATGA)·d(TCATTTGGG) by two-dimensional NMR, distance geometry, restrained molecular dynamics, and full relaxation matrix simulation of the two-dimensional nuclear Overhauser effect spectra at various mixing times. The chimeric strand of this duplex consists of the last four residues of the $tRNA^{Pro}$ primer for (−) strand DNA synthesis of Moloney murine leukemia virus and the first five residues of the (−) strand DNA produced by extending this primer; the complementary DNA strand corresponds to the (+) strand product from this template. The hybrid section of this chimeric duplex assumes a structure similar to that found for pure hybrid duplexes of mixed sequence, while the DNA section assumes a conformation closer to B-form DNA. There is significant distortion of the duplex at the hybrid−DNA junction which is manifested in marked changes in the helical parameters buckle, roll, and tip, changes in glycosidic torsion angles, and changes in the backbone torsion angles δ, ϵ, and ζ. The sugar conformations also undergo large changes, from heteromerous puckers in the hybrid section to a more B-form in the DNA section. Furthermore, the intrastrand phosphate separation in the chimeric strand is more typical of A-form duplexes in the RNA section but more like B-form duplexes in the DNA section. In the DNA section the minor groove width changes gradually from B-form at the periphery and approaches hybrid-like dimensions closer to the junction. The structural discontinuities act synergistically to produce a bend of 18 ± 3° at the junction. The global structure of this sequence is similar to that previously found in the chemically analogous Okazaki fragment r(gcg)d(TATACCC)·d(GGGTATACGC) in solution. Such structure homology suggests a possible link between structure and function with respect to the recognition and cleavage of the junction RNA residues in both retroviral chimeras and Okazaki fragments during reverse transcription and normal DNA replication.

Nucleic acid duplexes normally adopt either of two principal structural classes, known as A-form and B-form in solution. Thus, with some exceptions, DNA duplexes generally assume a B-form conformation, while RNA duplexes generally assume an A-form conformation. The structure of A-form duplexes is characterized by, among other parameters, C3′-endo sugars and a wide shallow minor groove, while the structure of B-form duplexes is characterized by C2′-endo sugars and a deep, narrow minor groove (Saenger, 1984). Until recently very little was known about the structure of nucleic acid duplexes resulting from the pairing of a chimeric DNA−RNA strand to a pure DNA strand to form a duplex consisting of a hybrid section covalently linked to a pure DNA section. These types of structures, normally referred to as Okazaki fragments, are important in molecular biology since they occur frequently during normal DNA replication and also during reverse transcription (*vide infra*). Recently, the structure of the synthetic Okazaki fragment duplex r(gcg)d(TATACCC)·(GGGTATACGC) was determined both in the crystalline (Egli et al., 1992, 1993) and in the solution state (Salazar et al., 1994). The duplex structure of this Okazaki fragment in solution was found to adopt a conformation consisting of a hybrid-form (H-form) segment connected to a B-form segment by a junction (Salazar et al., 1994). This is in contrast to the uniform A-form structure found in the crystalline state (Egli et al., 1992, 1993).

Structures that are chemically equivalent to Okazaki fragments are ubiquitous during retroviral reverse transcription. All known retroviruses replicate *via* a double-stranded DNA intermediate which is then stably integrated into the host cell genome. At the double helical level, the double-stranded replicative intermediate contains long terminal repeats at each end arranged 5′-U3-R-U5···, where U3 and U5 are derived from sequences adjacent to the viral RNA 3′ and 5′ termini, respectively, and R is a repeated sequence derived from the actual viral RNA termini. Reverse transcription is initiated by a cellular tRNA primer which binds to a complementary viral RNA sequence called the primer binding site (PBS) located near the 5′ end of the retroviral genome. This tRNA primer is extended by the viral-encoded reverse transcriptase to form the (−) DNA strand using the viral RNA as a template (Gilboa et al., 1979). The RNA

[†] This work was supported by NIH Grants GM-42896 and GM-32681 to B.R.R.

[‡] Coordinates have been deposited in the Brookhaven Protein Data Bank under the file name XXXX.

* Address correspondence to either author.

[§] The University of Texas at Austin.

[||] Chemistry Department, University of Washington.

[⊥] Biochemistry Department, University of Washington.

[⊗] Abstract published in *Advance ACS Abstracts,* June 1, 1996.

S0006-2960(95)02891-1 CCC: $12.00

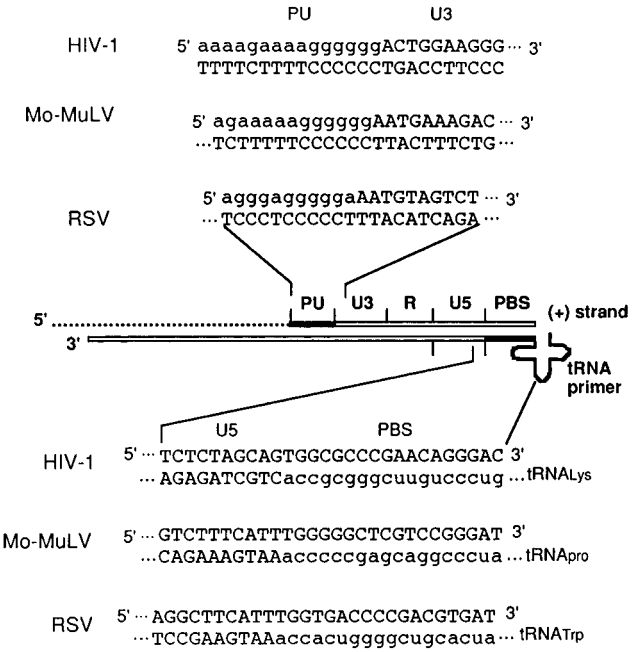

FIGURE 1: Sequences of chimeric retroviral fragments (solid black-filled rectangles represent RNA) for HIV-1 (Muesing et al., 1985; Huber & Richardson, 1990; Pullen & Champoux, 1990), Mo-MuLV (Shinnick et al., 1981; Mitra et al., 1982; Finston & Champoux, 1984), and RSV (Schwartz et al., 1983; Resnick et al., 1984; Mitra et al., 1982; Rattray & Champoux, 1987). At the top are the the sequences of the chimeric duplexes formed after extension of the (+) DNA strand from the polypurine fragment (PU) using the (−) DNA strand as a template. The bottom shows the sequences of the chimeric duplexes formed after extension of the (−) DNA strand from the tRNA primers and in the early stages of (+) DNA strand synthesis.

template is then degraded by RT[1] −RNase H, except for a purine-rich segment (PU) which is resistant to RT−RNase H. The uncleaved purine-rich short RNA segment then serves as the primer for (+) strand DNA synthesis using the (−) strand DNA as a template. Thus, during the first half of (+) strand DNA synthesis, after extension of the PU-primed (+) strand beyond the original tRNA (−) strand primer site, the complex now contains two chimeric duplex segments which are chemically equivalent to Okazaki fragments. Figure 1 shows the actual sequences at these sites and the respective tRNA primers for the three important retroviruses, human immunodeficiency virus type 1 (HIV-1), Moloney murine leukemia virus (Mo-MuLV), and Rous sarcoma virus (RSV).

The importance of chimeric structures of the type (r-D)·(D-D) in molecular biology prompted us recently to carry out a qualitative NMR study of the sugar conformations in the chimeric octamer duplex r(gcca)d(CTGC)·d(GCAGTGGC) (Salazar et al., 1993), a sequence that occurs at the PBS−U5 junction during HIV-1 replication (see Figure 1). The results of that study suggested a heteromerous A−B sugar conformation in the hybrid section of the chimeric duplex with B-like sugars in the DNA section and a significant discontinuity at the junction. A more quantitative study of the structurally related Okazaki fragment duplex r(gcg)d(TATACCC)·(GGGTATACGC) (Salazar et al., 1994) revealed a chimeric H-form−B-form structure with a significant discontinuity at the junction, in agreement with the qualitative study of the sugar conformations in r(gcca)d(CTGC)·d(GCAGTGGC). Furthermore, modeling studies of complexes with the crystal structure of HIV-1 reverse transcriptase suggested that the bent chimeric duplex structure assumed by these nucleic acids may facilitate RT binding of these substrates and subsequent cleavage of the RNA residues at the hybrid−DNA junction by RT−RNase H (Salazar et al., 1994).

The possible correlation between structure and function in chimeric structures of this type and their interaction with RT−RNase H prompted us to carry out an NMR-based structure study of a chimeric duplex containing the 3′ tail of the tRNA primer for Moloney murine leukemia virus. For this purpose we chose to study the nine-base-pair fragment r(ccca)d(AATGA)·d(TCATTTGGG), which contains the last four residues of the $tRNA^{Pro}$ primer for (−) strand DNA synthesis of Moloney murine leukemia virus and the first five residues of the (−) strand DNA product; the complementary strand is the product of (+) strand DNA synthesis (see Figure 1). The results suggest similarities in the global and local structure in retroviral chimeric duplexes and chemically equivalent Okazaki fragments which may be important for the recognition and cleavage of the RNA residues in these types of sequences by RT−RNase H during reverse transcription and by cellular RNase H during normal DNA replication.

EXPERIMENTAL PROCEDURES

Sample Preparation. Ten micromole syntheses of the pure DNA strand 5′-d(TCATTTGGG) and the chimeric RNA−DNA strand 5′-r(ccca)d(AATGA) were carried out on an automated DNA synthesizer (Applied Biosystems Model 392). The DNA strand was synthesized and purified using standard methods. The chimeric RNA−DNA strand was synthesized by introducing the ribonucleoside phosphoramidites in the appropriate cycles using base-protected 5′-DMT 3′-methoxy *N,N*-diisopropylphosphoramidites protected with *tert*-butyldimethylsilyl groups at the 2′-position; the phosphoramidites were purchased from Biogenex/ABN and Applied Biosystems. Deprotection and cleavage from the support were carried out with anhydrous methanolic ammonia overnight. The RNA 2′-TBDMS protecting groups were removed by dissolving the chimeric strand in a 1 M solution of tetrabutylammonium fluoride in tetrahydrofuran and allowing the solution to stand overnight at room temperature. Purification of the chimeric strand was carried out as previously described (Salazar et al., 1993a). The pure DNA and the chimeric strands were annealed by dissolving equal amounts of the two strands in 500 mM NaCl, 0.2 mM EDTA, and sodium phosphate buffer (pH 6.8), followed by heating to 55 °C and slow cooling to room temperature. Single-stranded material was subsequently removed by hydroxylapatite column chromatography with a phosphate gradient, and the annealed sample was then desalted via Sephadex G-10 column chromatography. The purified r(ccca)d(AATGA)·d(TCATTTGGG) duplex (∼20 mg) was dissolved in 0.4 mL of a buffer consisting of 75 mM KCl, 40 mM NaCl, 2.0 mM EDTA, and 75 mM sodium phosphate (pH 6.8). After repeated lyophilization from 99.96% ^2H_2O,

[1] Abbreviations: RT, reverse transcriptase; Mo-MuLV, Moloney murine leukemia virus; NMR, nuclear magnetic resonance; CD, circular dichroism; EDTA, ethylenediaminetetraacetic acid; NOE, nuclear Overhauser effect; 2D, two dimensional; NOESY, nuclear Overhauser effect spectroscopy; rmsd, root mean squared difference; rMD, restrained molecular dynamics.

the sample was redissolved in 0.4 mL of 99.996% ^2H_2O for the NMR experiments.

NMR and CD Spectroscopy. All NMR data were collected at 28 °C on a home-built 500 MHz NMR spectrometer (J. Gladden and G. Drobny, unpublished design). The data were processed on a Silicon Graphics IRIS 4D workstation with the FTNMR and FELIX software programs (Hare Research). Four NOESY spectra with mixing times of 60, 90, 120, and 180 ms were collected in the phase-sensitive hypercomplex mode (States et al., 1982) using 1024 complex points in t_2 and 400 pairs of real and imaginary experiments in t_1. The data were zero filled to 2048 points in both dimensions and apodized with a 90° phase-shifted sine-squared window function. The E.COSY spectrum (Griesinger et al., 1985) was collected in the phase-sensitive TPPI mode (Drobny et al., 1979; Marion & Wüthrich, 1983) using 2048 complex points in t_2 and 700 experiments in t_1. The E.COSY data were zero filled to 4096 points in t_2 and 2048 points in t_1 and processed using 6 Hz of exponential line narrowing and 6 Hz of Gaussian line broadening in t_2 and a skewed sine bell function in t_1. A relaxation delay of 5 s was used between scans in both the NOESY and the E.COSY experiments. The CD spectrum was acquired at room temperature on a JASCO 720 spectrometer at a duplex concentration of ~50 μM. The scan speed, step resolution, and equilibration time were set at 100 nm/min, 0.2 nm, and 2 s, respectively. The spectrum of the buffer was subtracted, and the final CD spectrum represents the sum of 10 scans subjected to a noise reduction routine.

Structure Refinement and Full Relaxation Matrix Simulations of the NOESY Spectra. The solution structure of the r(ccca)d(AATGA)·d(TCATTTGGG) retroviral fragment was analyzed on the basis of an isotropic tumbling model with a global correlation time. Assumption of isotropic tumbling in this case is quite valid since the calculated ratio of parallel to perpendicular diffusion ($D_{\parallel}/D_{\perp}$) is ~1.7 for a duplex of this length [Tirado & Garcia de la Torre, 1980; see also Wang et al. (1992)]. Thus, distance errors due to anisotropy are minimal and are not expected to lead to significant distance errors in the structure (Wang et al., 1992). With the exception of the 5′-terminal residues, the observed H6/H8−H3′, H6/H8−H2′, H1′−H2′/H2″ (data not shown), and H5−H6 NOESY cross peaks are consistent with a single average correlation time for the bases and sugars.

Our previous studies have shown that the sugars in the hybrid duplex d(GTCACATG)·r(caugugac) assume conformations in the O4′-endo to C1′-exo range, and depending on the sequence, sugars in the hybrid section of Okazaki fragments and other chimeric duplexes also assume conformations within this range (Fedoroff et al., 1993; Salazar et al., 1994; Zhu et al., 1995). In the present duplex the deoxyribose sugars in the hybrid section assume conformations in the C1′-exo range (*vide infra*). Recently, on the basis of their interpretation of coupling constants with respect to torsion angles, González et al. (1994, 1995) have concluded that there is significant large amplitude motion of the DNA sugars in the thiophosphate modified DNA−RNA hybrid duplex d(GCTATAA$_{ps}$TGG)·r(ccauuauagc) as a result of N/S (C3′-endo to C2′-endo) interconversion. From these results González et al. (1994, 1995) infer that the deoxyribose sugars of DNA−RNA hybrid duplexes have conformational flexibility. However, it should be mentioned that Harbison (1993) and, most recently, Zhu et al. (1994) have shown that *J*-couplings are modulated by dipolar effects in the spin diffusion limit, rendering torsion angle interpretation *via* classical Karplus relationships dubious at correlation times greater than ~2 ns. This effect has generally been ignored, and failure to take it into account could lead to erroneous conclusions in regard to the populations of deoxyribose sugar conformers in dynamic equilibrium. While the confirmation of the extent of dynamics in the deoxyribose sugars of DNA−RNA hybrid duplexes awaits solid-state NMR studies, the results of González et al. (1994, 1995) show that our conclusions regarding RNase H discrimination of minor groove widths in DNA−RNA hybrid and chimeric junctions (*vide infra*) should be unaffected by the presence or absence of any conformational flexibility of the deoxyribose sugars. In the present work we used primarily NOE information to arrive at approximate values for the sugar conformations. Typically, the sugars were constrained to *P* values with bounds of ±15°. This would allow for the presence of any small amplitude internal motion in the deoxyribose sugars. Coupling constant information from E.COSY was used only as an external check on the back-calculation results, keeping in mind the potentially deleterious effects of cross-relaxation on the *J*-couplings.

Initial distance restraints were calculated for each resolved NOESY cross peak by measuring the initial rate of cross-relaxation and scaling these rates to the average initial rate of cross-relaxation for the resolved cytosine H5−H6 proton pairs (d_{H5-H6} = 2.5 Å). In general, we made use of conservative bounds for the distance restraints which should adequately compensate for the uncertainties in the distances arising from the assumption of isotropic tumbling and the presence of small amplitude internal motions. Thus, typical error limits for the upper and lower bound restraints were ±0.4 Å for distances less than 3.5 Å and ±0.7 Å for distances greater that 3.5 Å. The backbone conformation was conservatively restrained as described by Kim et al. (1992). The distance restraints used are as follows: 76 interresidue (sequential) restraints, 90 intraresidue restraints, 5 interchain restraints, and 22 hydrogen bond restraints.

Initial structures were obtained by distance geometry methods in order to avoid the bias inherent in the use of standard A-form or B-form starting structures and in order to more adequately sample conformational space. Briefly, the distance restraints were entered into a bounds matrix and smoothed via the triangle inequality. The structures were then randomly embedded in 3D space from the bounds matrix and refined against the distance bounds as described by Hare and Reid (1986). The structures were further refined by restrained molecular dynamics using the DISCOVER program (BIOSYM Technologies, San Diego, CA) with AMBER potential functions (Weiner et al., 1986). The force constants in the NOE term used in the DISCOVER program were derived by assuming the left and right errors of the distance ranges were 0.1 Å (for a description of the NOE term used, see the DISCOVER Reference Guide). These force constants correspond to approximately 30 kcal/(mol·Å^2) at 300 K. Prior to the molecular dynamics simulations, the structures were first minimized using 2000 cycles of conjugate gradient minimization. The molecular dynamics simulations were then initiated at 300 K with a step size of 1.0 fs for a period of 15 ps using a distance-dependent dielectric constant of 4r. A nonbonded cutoff distance of 12 Å was used, and no counterions were included in the

simulations. Coordinate sets were recorded each 1.0 ps, and the last five coordinate sets were subjected to restrained energy minimization for 2000 steps of conjugate gradient energy refinement. The structure with the lowest energy among these last five coordinate sets was then chosen for further iterative refinement by back-calculation of the NOESY spectra.

The back-calculation refinement protocol accounts for spin diffusion effects explicitly and involves the calculation of NOE intensities from the Cartesian coordinates of interim structures *via* the full-relaxation matrix (Keepers & James, 1984). The improved NOESY simulation program BIRDER was used for back-calculation refinement as described by Zhu and Reid (1995) using typical external relaxation rates for DNA and RNA protons, as described by Zhu et al. (1995), and an empirically determined correlation time of 8.0 ns. The NOE intensities were used as input in the GNOE program (Hare Research) to simulate the 2D NOESY spectra. The distance restraints were then adjusted accordingly on the basis of the comparison of the calculated and the experimental NOE intensities in all spectral regions at 60, 90, 120, and 180 ms mixing times. After the intensities of most of the resolved cross peaks were reproduced, distance restraints for overlapped cross peaks were added in the subsequent stages of iterative refinement following the procedure described above. The rMD/relaxation matrix simulation refinement procedure was repeated until there were no significant improvements in the simulated NOESY spectra and until the NOE *R*-factor (Kim & Reid, 1992) did not improve further. The NOE *R*-factor ($R = \sum|I_e - I_c|/\sum I_c$, where I_e and I_c are the experimental and the calculated NOE intensities, respectively) was calculated for 82 resolved cross peaks in the H6/H8−H1′/H2′/H2″/H3′/H4′, the H1′−H2′/H3′/H4′, and the H2−H1′ regions of the NOESY spectra at 60, 90, 120, and 180 ms.

RESULTS

Resonance Assignments and Analysis of NMR Data. The numbering system for the r(ccca)d(AATGA)·d(TCATTTGGG) chimeric duplex is

```
       2     4     6     8
5' T C A T T T G G G
   A G T A A a c c c 5'
   18   16   14   12   10
```

where the lower-case letters designate RNA residues. Figure 2 shows the sequential assignments of the expanded H6/H8−H2′/H2″ and H6/H8−H1′ regions of the NOESY spectra at 90 and 120 ms mixing times, respectively. Although stereospecific assignments of the H5′/H5″ protons could not be made with any confidence, assignments of the rest of the nonexchangeable RNA and DNA protons were made by analysis of all regions of the NOESY spectra at various mixing times in combination with E.COSY data. These assignments are shown in Table 1. The H6/H8−H2′/H2″ region of the 90 ms NOESY spectrum shows that all the well-resolved (*n*)H6/H8−(*n*)H2′ cross peaks, including those of the T6, G7, G8, and G9, are stronger than the (*n*)-H6/H8−(*n* − 1)H2′ cross peaks. This observation indicates that none of the DNA residues in this chimeric duplex, including those in the hybrid section, assume an A-form conformation in solution. This conclusion was confirmed by measuring the vicinal coupling constants for the deoxy-

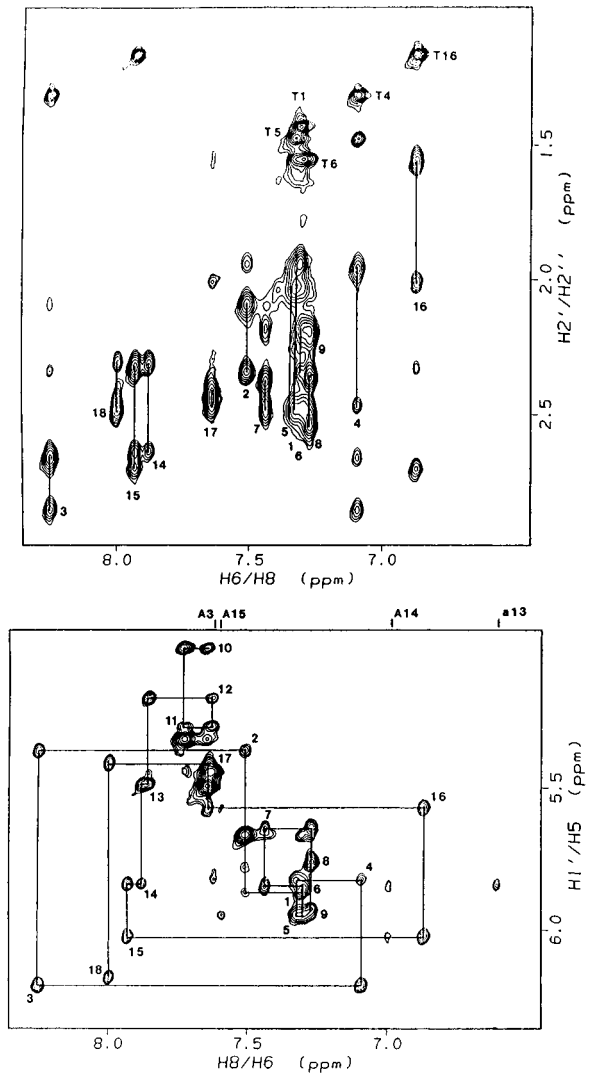

FIGURE 2: Assignment of the DNA H6/H8−H2′/H2″ (top) and RNA/DNA H6/H8−H1′ (bottom) regions of the NOESY spectra of r(ccca)d(AATGA)·d(TCATTTGGG) at 90 and 120 ms, respectively. The (*n*)H6/H8−(*n*)H2′ cross peaks in the spectrum at the top are labeled and are connected to the (*n*)H6/H8−(*n*)H2″ cross peaks by solid vertical lines. The (*n*)H6/H8−(*n*)H1′ cross peaks are labeled with the residue number, and the chemical shifts of the AH2 chemical shifts are labeled A3, A15, A14, and a13 according to the numbering scheme shown in the text.

ribose residues. Table 2 shows that the DNA H1′−H2′ coupling constants are quite large (>7.5 Hz). It is only the RNA residues that have undetectable H1′−H2′ E.COSY cross peaks for which the coupling constants were too low to be measured. Thus, a qualitative analysis reveals that all the RNA residues have N-type, C3′-endo, sugars with H1′−H2′ torsion angles of ∼90°. Furthermore, the RNA (*n*)H6/H8−(*n* − 1)H2′ NOESY cross peaks at 60 ms (not shown) are much stronger than any of the DNA (*n*)H6/H8−(*n*, *n* − 1)H2′/H2″ cross peaks, indicating that the RNA segment in the chimeric strand assumes an overall A-form conformation.

Results of Structure Refinement. A total of eight final structures were obtained by the combined use of distance geometry and molecular dynamics. These structures all converged to a pairwise rmsd of *ca.* 0.6 ± 0.3 Å and are shown superimposed in Figure 3. The energy terms for these structures are shown in Table 3. Figure 4 shows good agreement between the experimental and back-calculated NOESY spectra for the H6/H8−H1′ and the H6/H8−H2′/

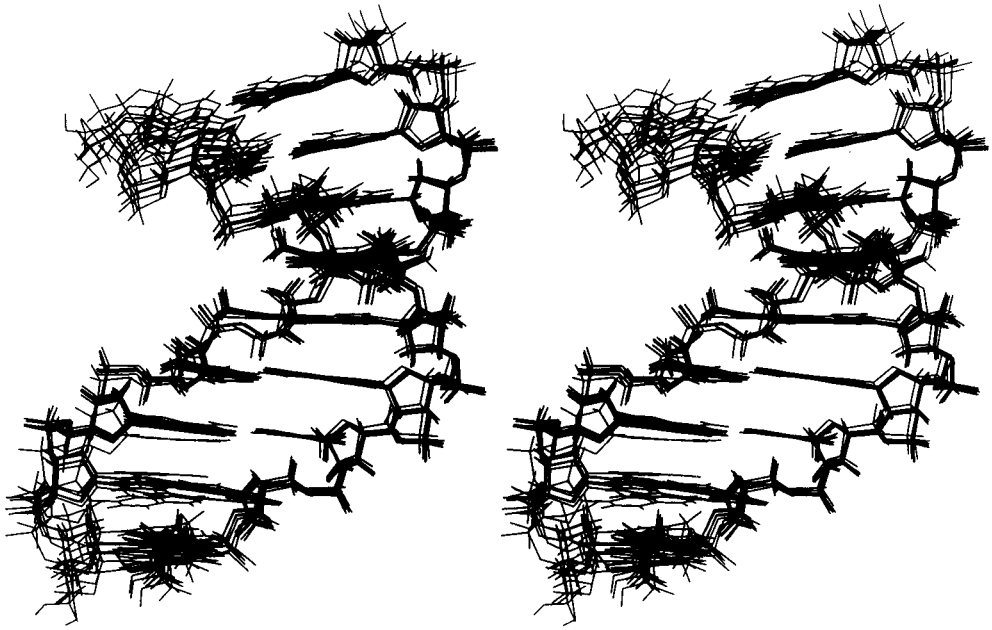

FIGURE 3: Superimposed stereoviews (wide eye) of the eight final structures of the Mo-MuLV retroviral fragment r(ccca)d(AATGA)·d(TCATTTGGG).

Table 1: Chemical Shift (ppm) of Nonlabile Protons of r(ccca)d(AATGA)·d(TCATTTGGG) at 28 °C[a]

residue	H6/H8	H5/H2/M5	H1′	H2′	H2″	H3′	H4′
T1	7.30	1.42	5.86	1.94	2.33	4.53	3.94
C2	7.49	5.65	5.36	2.08	2.32	4.72	4.02
A3	8.25	7.61	6.18	2.65	2.84	4.90	4.32
T4	7.08	1.30	5.81	1.95	2.45	4.68	4.09
T5	7.31	1.47	5.94	2.02	2.48	4.71	4.04
T6	7.28	1.54	5.85	2.18	2.45	4.76	4.04
G7	7.43		5.63	2.36	2.49	4.76	4.17
G8	7.26		5.74	2.34	2.53	4.68	4.21
G9	7.26		5.92	2.20	2.22	4.40	4.02
c10	7.63	5.48	5.00	4.26		4.15	4.15
c11	7.72	5.32	5.28	4.32		4.35	4.21
c12	7.62	5.43	5.18	4.33		4.39	4.22
a13	7.84	6.59	5.47	4.14		4.48	4.29
A14	7.87	6.98	5.83	2.31	2.80	4.75	4.25
A15	7.92	7.58	6.01	2.32	2.70	4.80	4.27
T16	6.86	1.15	5.56	1.55	2.00	4.69	3.94
G17	7.63		5.40	2.40	2.45	4.79	4.12
A18	7.99	7.73	6.15	2.47	2.29	4.53	4.09

[a] Referenced to the HDO peak at 4.6 ppm.

Table 2: H1′−H2′/H2″ Coupling Constants (Hz) in r(ccca)d(AATGA)·d(TCATTTGGG)[a]

residue	$J_{1'-2'}$	$J_{1'-2''}$	residue	$J_{1'-2'}$	$J_{1'-2''}$
T1	8.6	6.7	c10	<1.0	
C2	9.0	5.9	c11	<1.0	
A3	10.5	*b*	c12	<1.0	
T4	7.5	8.5	a13	<1.0	
T5	8.8	8.5	A14	7.5	6.4
T6	8.6	7.5	A15	8.6	7.5
G7	8.6	7.5	T16	8.0	8.0
G8	8.0	6.4	G17	*b*	*b*
G9	*b*	*b*	A18	7.5	8.6

[a] Obtained from an E.COSY spectrum with a resolution of 1.07 Hz per point. [b] Could not be measured due to spectral overlap.

H2″ regions at 90 and 120 ms mixing times, respectively, for one of the structures. The NOE *R*-factors are 0.205 at 60 ms, 0.204 at 90 ms, 0.199 at 120 ms, and 0.198 at 180 ms mixing times, indicating good agreement between the back-calculated and the experimental NOESY spectra for the resolved peaks at all mixing times.

Table 3: Energy Analysis of the Final r(ccca)d(AATGA)·d(TCATTTGGG) Structures

energy type	energy value (kcal/mol)
total energy	-103.99 ± 1.66[a]
bond energy	13.08 ± 0.07
θ energy[b]	88.5 ± 0.41
ϕ energy[c]	167.93 ± 0.71
hydrogen bond energy	-13.30 ± 0.04
nonbond energy	-175.35 ± 1.67
nonbond repulsion energy	578.35 ± 2.35
nonbond dispersion energy	-753.70 ± 3.94
Coulomb energy	-184.88 ± 0.95
forcing energy[d]	42.77 ± 0.76

[a] Energy values are expressed as mean values ± standard deviations and are based on eight final structures. [b] Bond angle deformations. [c] Dihedral angle deformations. [d] NOE constraint violation with a force constant of 30 kcal/(mol·A^2).

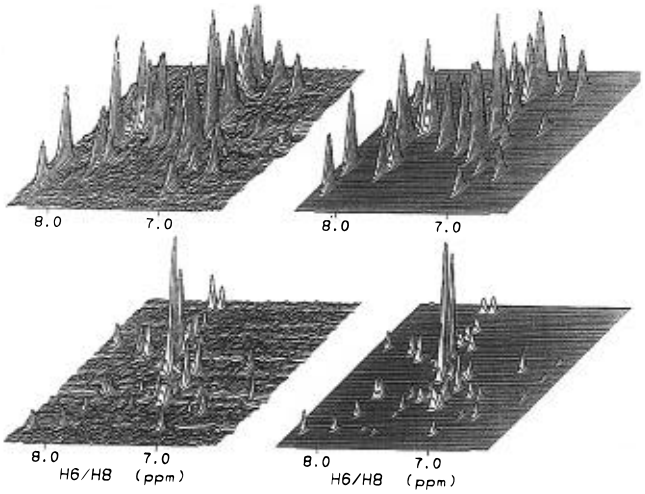

FIGURE 4: Stack plots of the experimental (left column) and back-calculated (right column) H6/H8 to H1′/H2′/H2″ NOESY spectra for one of the refined chimeric structures. The DNA H6/H8−H2′/H2″ and the RNA/DNA H6/H8−H1′ NOESY cross peaks are shown at the top and bottom rows, respectively.

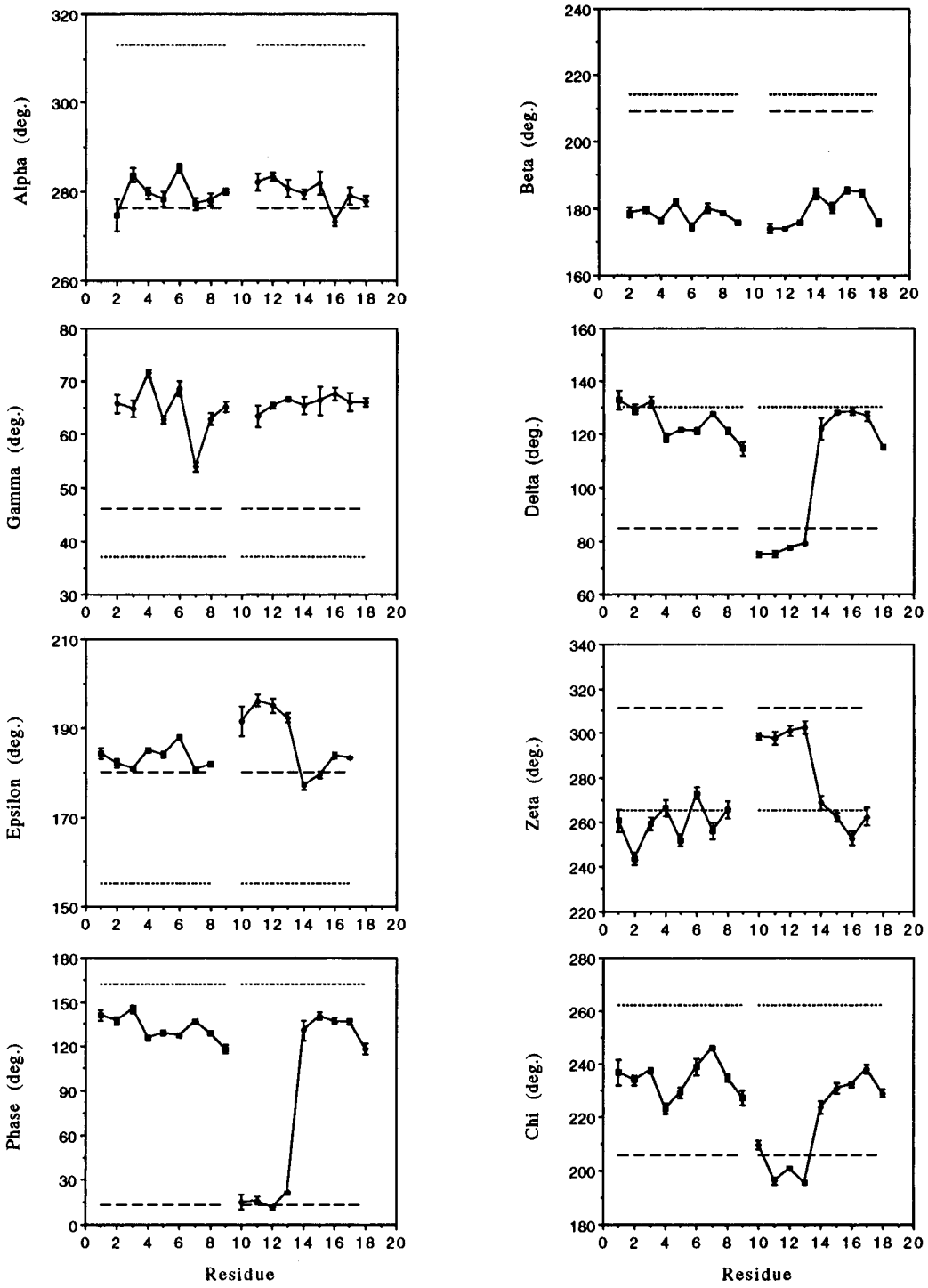

FIGURE 5: Plots (mean values ± standard deviations) of the backbone torsion angles, sugar puckers, and glycosidic torsion angles for the eight final r(ccca)d(AATGA)·d(TCATTTGGG) duplex structures compared to the corresponding values for standard A-form DNA (broken line) and standard B-form DNA (dotted line) (Arnott & Hukins, 1972, 1973).

Structural Features of r(ccca)d(AATGA)·d(TCATTTGGG). Figure 5 shows plots (mean values ± standard deviations) of the backbone torsion angles, sugar puckers, and glycosidic torsion angles for the final solution structures. Also shown for comparison are the values for standard B-form and A-form DNA. The chimeric structure of this duplex is reflected dramatically in the δ, ϵ, and ζ backbone torsion angles, in the glycosidic torsion angle χ, and in the sugar pucker values. For these torsion angles, a drastic change is observed in going from the RNA segment to the DNA segment in the chimeric strand. This observation is in agreement with previous studies of chimeric duplexes in solution (Salazar et al., 1994; Zhu et al., 1995) but is contrary to the uniform structure for the related r(cgc)d(TATACCC)·d(GGGTATACGC) chimeric duplex found in the crystalline state (Egli et al., 1992, 1993). Thus, in the chimeric strand of the r(ccca)d(AATGA)·d(TCATTTGGG) duplex in solution, the torsion angles of the RNA segment (residues 10−13) are generally closer to A-form, while the torsion angles of the DNA segment (residues 14−18) range from values that are intermediate between A-form and B-form (δ, P, and χ) to values that are closer to A-form (α, β, γ, and ϵ) or B-form (ζ) DNA, depending on the position in the chain. In the pure DNA strand (residues 1−9), the torsion angles α,

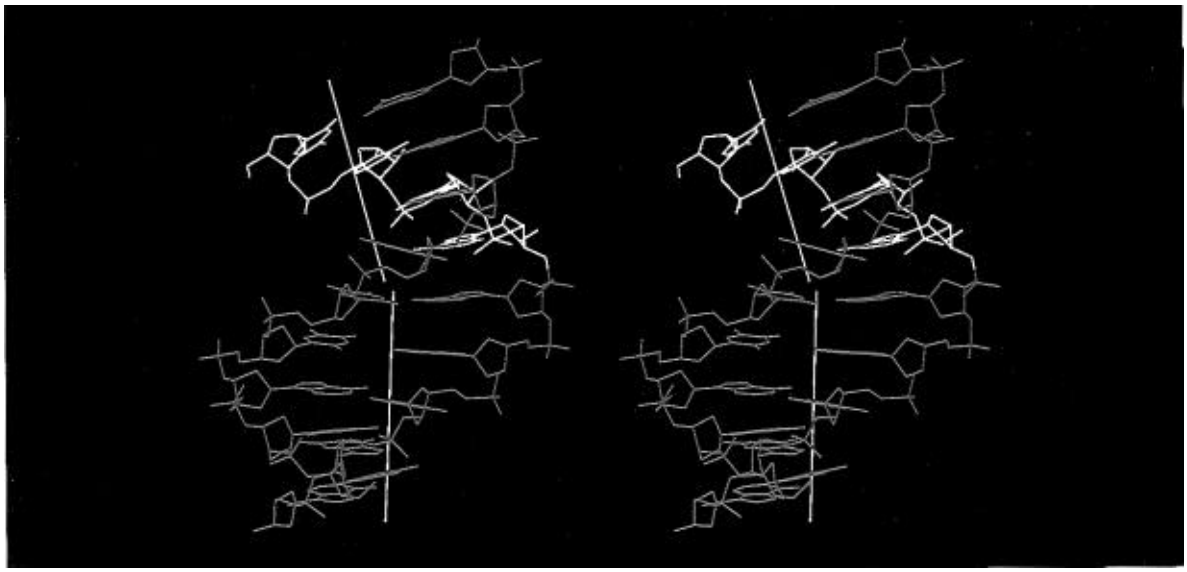

FIGURE 6: Stereoview (wide eye) of one of the final structures of the Mo-MuLV retroviral fragment r(ccca)d(AATGA)·d(TCATTTGGG) showing the distinct helical axes for the hybrid and DNA sections. The DNA residues are shown in blue, and the RNA residues are in white.

β, γ, and ϵ are closer to A-form DNA values, while δ, *P*, and χ assume intermediate conformations; only ζ assumes a conformation close to B-form.

The uniformity of the backbone torsion angles is clearly broken at the DNA junction base pair T5·A14. The sugar conformations of the DNA residues in the hybrid section of the duplex are in the O4′-endo to C1′-exo range, similar to the average *P* values found for the DNA residues in related Okazaki fragments (Salazar et al., 1993a, 1994) and in the pure DNA−RNA hybrid duplex d(GTCACATG)·r(caugugac) (Salazar et al., 1993a; Fedoroff et al., 1993). Thus, the hybrid half of this duplex has a typical hybrid double helix structure.

Figure 6 shows that there is a bend in the helix which reflects the discontinuities in the torsion angles. Thus, there are two obvious segments to the overall helix. Since the helical parameters are sensitive to the choice of local helical axis, the bend in the structure prompted us to choose two helical axes in order to measure the structural parameters *via* NEWHEL 93 (R. E. Dickerson, UCLA): one for the DNA duplex section and one for the hybrid duplex section with both axes intersecting at the junction base pair T5·A14. Figure 7 shows the helical parameters for the eight final solution structures. Also shown for comparison are the values for standard A-form and B-form DNA. The structural discontinuity at the DNA duplex−hybrid duplex junction is reflected quite dramatically in the helical parameters roll, tip, twist, and buckle. However, there does not appear to be a clear pattern in the helical parameters, and both the DNA and the RNA residues have sequence-dependent A-form, B-form, or intermediate helical structure.

The structural discontinuities in the backbone torsion angles, sugar puckers, glycosidic torsion angles, and helical parameters are also reflected in the intrastrand phosphate−phosphate separations and the width of the minor groove. Figure 8A shows the intrastrand phosphate separations for the solution structure of r(ccca)d(AATGA)·d(TCATTTGGG) compared to the values for standard A-form and B-form DNA. The intrastrand P−P separations for all the DNA residues of the solution structure are closer to those of standard B-form DNA, while those for the RNA residues, c10, c11, c12, and a13, are intermediate between A-form and B-form values and are quite similar to those found in the solution structure of the Okazaki fragment r(gcg)d(TATACCC)·d(GGGTATACGC) (Salazar et al., 1994). Figure 8B shows the minor groove width as measured by the cross-strand phosphate−phosphate separation. The minor groove width in the DNA section is similar to the value of 5.7 Å found in B-form DNA fibers (Saenger, 1984) but increases gradually as the hybrid section is approached. At the junction the minor groove width is only ∼7.5 Å and reaches a value of ∼8.2 Å at the hybrid end of the helix (which is yet to be maximal).

DISCUSSION

The structure of the r(ccca)d(AATGA)·d(TCATTTGGG) chimeric duplex in solution *is not A-form*, in either half of the chimera, in contrast to observations in the crystalline state for the related duplex r(gcg)d(TATACCC)·d(GGGTATACGC), where it is uniformly A-form (Egli et al., 1992, 1993). Instead, the structure is a chimeric mixture of hybrid (H-form) structure at one end and B-form structure at the other end, consistent with the structure of related Okazaki fragment duplexes in solution (Salazar et al., 1993a, 1994). The chimeric structure of this duplex is also reflected in its CD spectrum (Figure 9), which indicates two types of duplex helicity, with a maximum at ∼220 nm and a minimum at ∼246 nm, characteristic of B-form DNA, and a second maximum at ∼272 nm, which is intermediate between the λ_{max} for B-form DNA (∼280 nm) and the λ_{max} for A-form RNA (∼250 nm) [see Wang et al. (1992b) for typical CD spectra of A-form RNA and B-form DNA]. Thus, the CD spectrum is consistent with the chimeric characteristics of the NMR-derived solution structure.

Figure 6 shows that the coexistence of H-form and B-form structure within the same duplex results in a structural discontinuity at the junction and a change in the direction of the helix, with a bend of 18 ± 3°. The same degree of bend (*ca*. 18.1°) was found in the related chimeric duplex r(gcg)d(TATACCC)·d(GGGTATACGC) in solution (Salazar et al., 1994). This bend is less than the 26° bend predicted

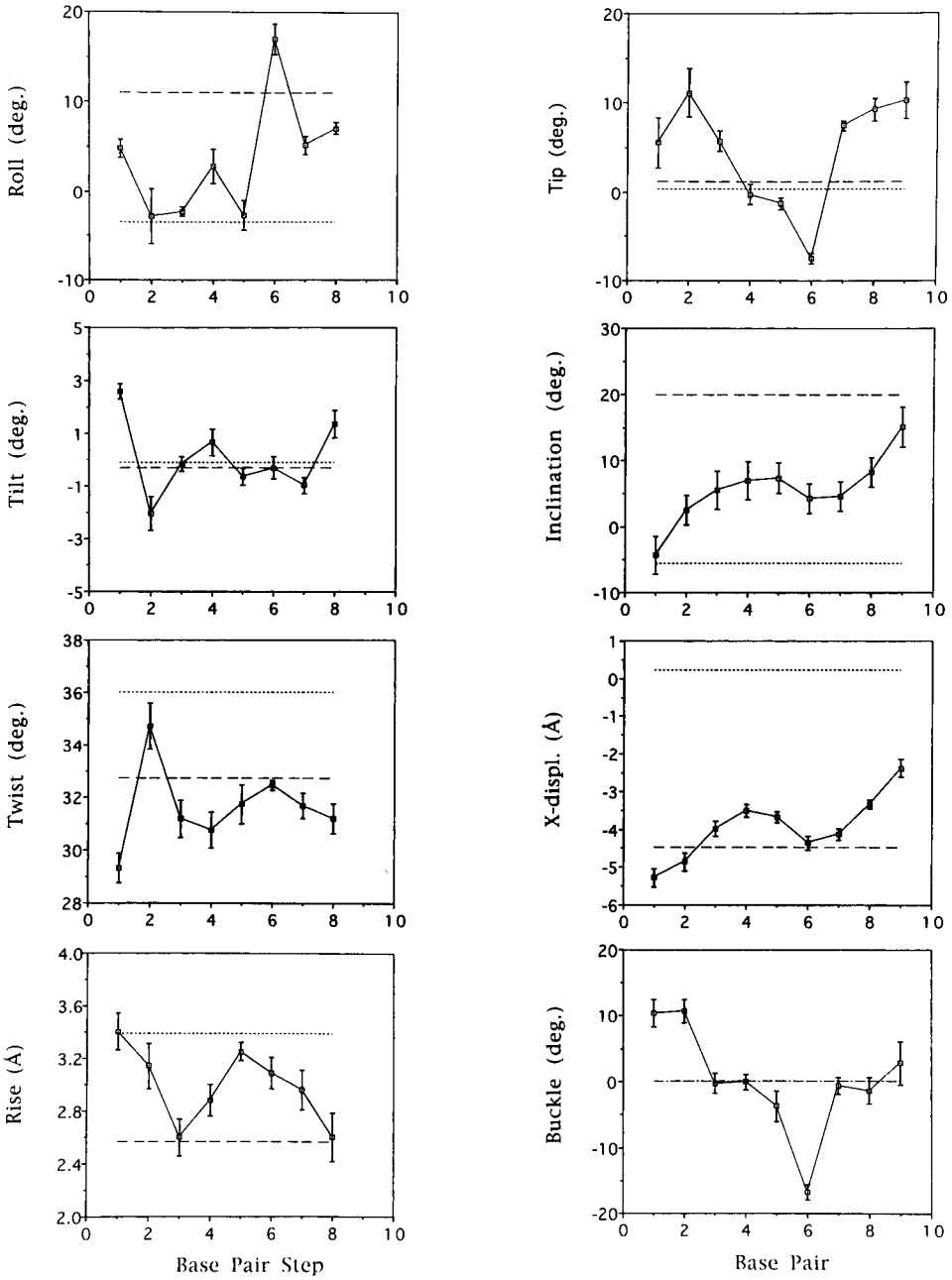

FIGURE 7: Plots (mean values ± standard deviations) of selected helical parameters for the eight final r(ccca)d(AATGA)·d(TCATTTGGG) duplex structures compared to the corresponding values in standard A-form DNA (broken line) and standard B-form DNA (dotted line). The helical parameters for the solution structure were calculated assuming two helical axes: one for the hybrid section and one for the DNA section, as described in the text.

by Selsing et al. (1978, 1979) for duplexes of the type dG_n·rC_idC_k on the basis of modeling studies. However, it should be noted that the 26° bend in the modeling studies of Selsing et al. (1978, 1979) may be overestimated since they assumed a uniform A-form structure for the hybrid section. On the other hand, the hybrid section of the present duplex and that of the r(gcg)d(TATACCC)·d(GGGTATACGC) duplex studied previously (Salazar et al., 1994) is rather short, and this precludes an accurate determination of the helix axis in the hybrid section due to end-fraying effects. Thus, care must be taken in comparing bend angles between related nucleic acid duplexes since the measured bend angle will depend on how the separate helix axes are defined and in the accuracy in determining the helix axes. Nonetheless, this bend appears to be a conserved feature of Okazaki fragments in solution.

In addition to the helix bend, there are other important structural features that are conserved between the present duplex and the chimeric duplex r(gcg)d(TATACCC)·d(GGGTATACGC) in solution. Thus, in both duplexes the same general trend is observed in the backbone torsion angles. In the present duplex, a large negative buckle is observed at the junction base pair step T5·A14−T6·a13 (see Figure 6). The same high negative buckle has been observed at the junction base pair step A7·T14−C8·g13 in the r(gcg)d(TATACCC)·d(GGGTATACGC) structure in solution (Salazar et al., 1994) and in the crystalline state (Egli et al., 1992, 1993); negative buckle was also reported for the junction base pair step A8·T17−C9·g16 in the solution structure of $[r(cgcg)d(TATACGCG)]_2$ (Zhu et al., 1995). Thus, although most helical parameters vary from structure to structure in a sequence-dependent way, it is particularly significant that a

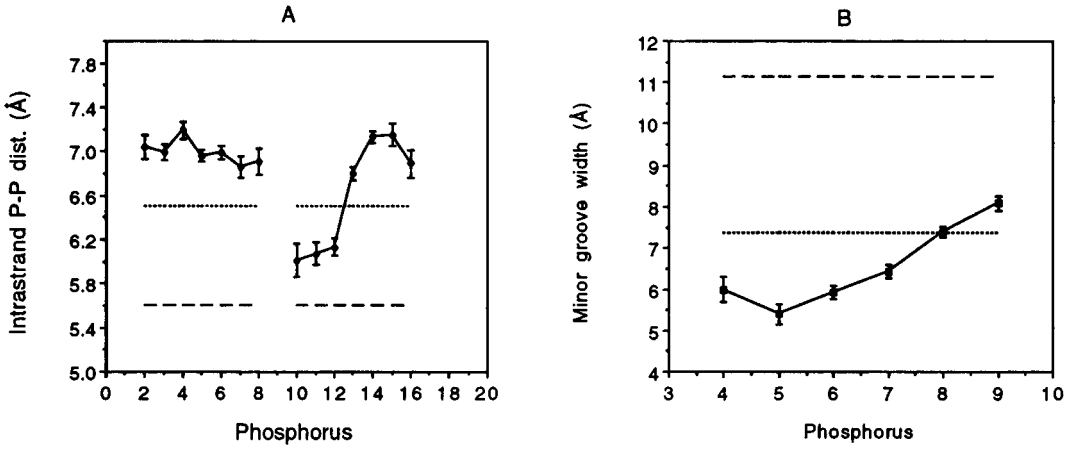

FIGURE 8: Plots (mean values ± standard deviations) of the intrastrand phosphorus−phosphorus separations (A) and minor groove widths (B) for the eight final r(ccca)d(AATGA)·d(TCATTTGGG) duplex structures compared to the corresponding values in standard A-form DNA (broken line) and standard B-form DNA (dotted line).

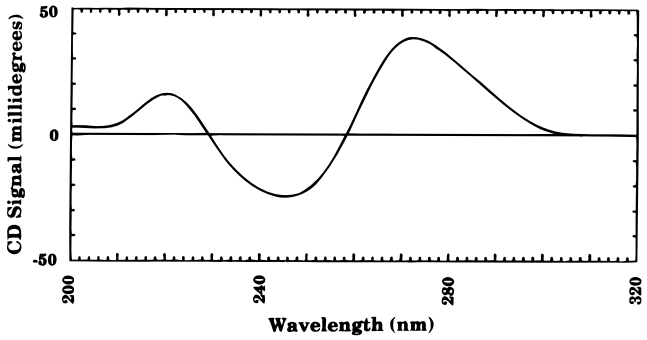

FIGURE 9: Circular dichroism spectra of the chimeric retroviral fragment duplex r(ccca)d(AATGA)·d(TCATTTGGG).

high negative value of buckle is conserved in all chimeric duplex junctions studied thus far.

Two structural features that are similar between chimeric duplexes stand out: the intrastrand phosphate−phosphate separation and the interstrand phosphate−phosphate separation, *i.e.*, the width of the minor groove. The intrastrand phosphate−phosphate separations of ∼7.0 Å for the DNA residues and ∼6.0 Å (see Figure 8A) for the RNA residues are similar to those found in the solution structure of r(gcg)d(TATACCC)·d(GGGTATACGC) (Salazar et al., 1994). This supports the observation made earlier (Fedoroff et al., 1993) that this structural feature is highly conserved for the DNA and RNA segments of chimeras and hybrids. The gradient in the width of the minor groove across the junction is also very similar to that observed previously in duplexes of this type (Salazar et al., 1994). In this duplex the minor groove width increases gradually from the DNA section toward the hybrid section, reaching a value of ∼8.2 Å at the junction (see Figure 8B).

The width of the minor groove is an important parameter to consider in retroviral fragments of this type since they are natural substrates for RT−RNase H. Recent results have shown that RT−RNase H from Mo-MuLV cleaves preferentially between the last and penultimate ribonucleotides (*i.e.*, c12−a13 in the present duplex) within the $tRNA^{Pro}$ primer in an extended duplex (Schultz et al., 1995). Thus, Mo-MuLV RT−RNase H leaves behind an RNA residue in the cleaved substrate. This is an unusual finding in that analysis of the sequence of a Mo-MuLV dsDNA circle junction (Shoemaker et al., 1980) indicates that Mo-MuLV RT−RNase H should cleave exactly at the RNA−DNA junction. However, as Schultz et al. (1995) pointed out, further analysis of Mo-MuLV circle junctions will be necessary to clarify this apparent discrepancy. Regardless of the cleavage specificity at the junction, RT−RNase H from Mo-MuLV does appear to have a strong structure-specific preference for the RNA−DNA junction, since it makes only minor cleavages in the remaining RNA primer (Schultz et al., 1995). An explanation for this specificity can be invoked by considering the width of the minor groove at the junction. Recently, we have shown that DNA−RNA hybrid duplexes have a minor groove width that is intermediate between A-form and B-form duplexes and that the uniqueness of the minor groove width in hybrids is important for RNase H recognition and cleavage of the RNA strand (Fedoroff et al., 1993). The optimum minor groove width for RNase H recognition and cleavage is not yet known and may even vary for different RNases H. Our model studies have shown that it appears to be around 9 Å for *Escherichia coli* RNase HI. In the present duplex, the minor groove width at the hybrid−DNA junction is only ∼7.5 Å, but one RNA base from the junction the minor groove width increases to ∼8.2 Å. It is interesting to note that, in contrast to the retroviral and eukaryotic RNases H, *E. coli* RNase HI does not have any specificity for the RNA−DNA junction per se (Furfine & Reardon, 1991; Smith & Roth, 1992; Huang et al., 1994). Thus, we may speculate that, compared to prokaryotic RNases H, retroviral and eukaryotic RNases H are tuned to a slightly narrower minor groove.

Recently, on the basis of studies docking HIV-1 RT to a chimeric duplex, we have proposed that the overall structure assumed by Okazaki fragments may be responsible for the recognition and cleavage of such substrates by RT−RNase H (Salazar et al., 1994). Thus, the mixed chimeric nature of the r(gcg)d(TATACCC)·d(GGGTATACGC) duplex and the inherent bend in the duplex lead to a structure that fits snugly within the binding cleft of the RT crystal structure of Jacobo-Molina et al. (1993). The structural similarities between r(gcg)d(TATACCC)·d(GGGTATACGC) and the present duplex suggest that, together with the width of the minor groove at the cleavage site, the overall structure assumed by these types of duplexes may in fact be an important factor responsible for RT−RNase H recognition and cleavage of chimeric junctions in retroviral segments.

ACKNOWLEDGMENT

We thank Ms. Julie M. Miller for her assistance in the purification of the chimeric duplex, Dr. Leiming Zhu for the use of his improved back-calculation program BIRDER, and Dr. James J. Champoux and members of his laboratory for illuminating discussions.

REFERENCES

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504−1510.
Arnott, S., & Hukins, D. W. L. (1973) *J. Mol. Biol. 81*, 93−105.
Drobny, G., Pines, A., Sinton, S., Weitekamp, D. P., & Wemmer, D. E. (1979) *Faraday Symp. Chem. Soc. 13*, 49−55.
Egli, M., Usman, N., Zhang, S., & Rich, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 534−538.
Egli, M., Usman, N., & Rich, A. (1993) *Biochemistry 32*, 3221−3237.
Fedoroff, O. Y., Salazar, M., & Reid, B. R. (1993) *J. Mol. Biol. 233*, 509−523.
Finston, W. I., & Champoux, J. J. (1984) *J. Virol. 51*, 26−33.
Furfine, E. S., & Reardon, J. E. (1991) *Biochemistry 30*, 7041−7046.
Gilboa, E., Mitra, S. W., Goff, S., & Baltimore, D. (1979) *Cell 18*, 93−100.
González, C., Stec, W., Kobylanska, A., Hogrefe, R. I., Reynolds, M., & James, T. L. (1994) *Biochemistry 33*, 11062−11072.
González, C., Stec, W., Kobylanska, A., Hogrefe, R. I., Reynolds, M., & James, T. L. (1995) *Biochemistry 34*, 4969−4982.
Griesinger, C., Sørensen, O. W., & Ernst, R. R. (1995) *J. Am. Chem. Soc. 107*, 6394−6396.
Harbison, G. S. (1993) *J. Am. Chem. Soc. 115*, 3026−3027.
Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341−5350.
Huang, L., Kim, Y., Turchi, J. J., & Bambara, R. A. (1994) *J. Biol. Chem. 269*, 25922−25927.
Huber, H. E., & Richardson, C. C. (1990) *J. Biol. Chem. 265*, 10565−10573.
Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H., & Arnold, E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6320−6324.
Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404−426.
Kim, S.-G., & Reid, B. R. (1992) *Biochemistry 31*, 12103−12116.
Kim, S.-G., Lin, L.-J., & Reid, B. R. (1992) *Biochemistry 31*, 3564−3574.
Lane, A. W., Ebel, S., & Brown, T. (1993) *Eur. J. Biochem. 215*, 297−306.
Marion, D., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Mitra, S. W., Chow, M., Champoux, J. J., & Baltimore, D. (1982) *J. Biol. Chem. 257*, 5983−5986.
Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C., Lasky, L., & Capon, D. (1985) *Nature 313*, 450−458.
Ogawa, T., & Okazaki, T. (1980) *Annu. Rev. Biochem. 49*, 421−457.
Pullen, K. A., & Champoux, J. J. (1990) *J. Virol. 64*, 6274−6277.
Pullen, K. A., Ishimoto, L. K., & Champoux, J. J. (1992) *J. Virol. 66*, 367−373.
Rattray, A. J., & Champoux, J. J. (1987) *J. Virol. 61*, 2843−2851.
Resnick, R., Omer, C. A., & Faras, A. J. (1984) *J. Virol. 51*, 813−821.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York.
Salazar, M., Champoux, J. J., & Reid, B. R. (1993a) *Biochemistry 32*, 739−744.
Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S., & Reid, B. R. (1993b) *Biochemistry 32*, 4207−4215.
Salazar, M., Fedoroff, O. Y., & Reid, B. R. (1994) *J. Mol. Biol. 241*, 440−455.
Schultz, S. J., Whiting, S. H., & Champoux, J. J. (1995) *J. Biol. Chem. 270*, 24135−24145.
Schwartz, D. E., Tizard, R., & Gilbert W. (1983) *Cell 32*, 853−869.
Selsing, E., & Wells, R. D. (1979) *J. Biol. Chem. 254*, 5410−5416.
Selsing, E., Wells, R. D., Early, T. A., & Kearns, D. R. (1978) *Nature 254*, 249−250.
Selsing, E., Wells, R. D., Alden, C. J., & Arnott, S. (1979) *J. Biol. Chem. 254*, 5417−5422.
Shinnick, T. M., Lerner, R. A., & Sutcliffe, J. G. (1981) *Nature 293*, 543−548.
Shoemaker, C., Golf, S., Gilboa, E., Paskind, M., Mitra, S. W., & Baltimore, D. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 3932−3936.
Smith, J. S., & Roth, M. J. (1992) *J. Biol. Chem. 267*, 15071−15079.
States, D. J., Haberkorn, R. A., & Reuben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
Tirado, M. M., & Garcia de la Torre, J. (1980) *J. Chem. Phys. 73*, 1986−1993.
Wang, A. C., Kim, S. G., Flynn, P. F., Sletten, E., & Reid, B. R. (1992a) *J. Magn. Reson. 100*, 358−366.
Wang, A. C., Kim, S. G., Flynn, P. F., Chou, S.-H., Orban, J., & Reid, B. R. (1992b) *Biochemistry 31*, 3940−3946.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230−252.
Wöhrl, B. M., & Moelling, K. (1990) *Biochemistry 29*, 10141−10147.
Zhu, L., & Reid, B. R. (1995) *J. Magn. Reson., Ser. B 106*, 227−235.
Zhu, L., Reid, B. R., Kennedy, M., & Drobny, G. P. (1994) *J. Magn. Reson., Ser. A 111*, 195−202.
Zhu, L., Salazar, M., & Reid, B. R. (1995) *Biochemistry 34*, 2372−2380.

BI9528917