Sugar Conformations at Hybrid Duplex Junctions in HIV-1 and Okazaki Fragments†

Miguel Salazar,‡ James J. Champoux,§ and Brian R. Reid*,‡,∥

Departments of Chemistry, Microbiology, and Biochemistry, University of Washington, Seattle, Washington 98195

Received October 20, 1992

ABSTRACT: We have carried out a solution study of the local conformation in a hybrid–chimeric duplex of the $\frac{r-D}{D-D}$ type (where r and D represent RNA and DNA). The object of this study was to investigate the sugar conformations at the internal junction in the hybrid–DNA octamer duplex (gccaCTGC)·(GCAGTGGC)—where the lower-case letters represent RNA residues. Such duplexes represent good models for Okazaki fragments in which RNA primers are covalently extended into DNA strands during DNA replication of the lagging strand. Furthermore, this particular sequence occurs during HIV-1 retrovirus reverse transcription. The chimeric RNA–DNA strand and the complementary pure DNA strand chosen for this study result from the priming of (–)-strand DNA synthesis by $tRNA^{Lys}$ and subsequent (+)-strand DNA synthesis by reverse transcriptase prior to HIV-1 retrovirus integration. Despite the unusual specificity of the RNase H activity of reverse transcriptase, which cleaves the RNA c–a phosphodiester rather than the junction a–C linkage, we found no major structural differences among the RNA sugar conformations—all RNA sugars were found in the normal C3′-endo A-form conformation. Instead, we find that the first DNA residue of the chimeric strand (5C) assumes a sugar conformation in the C4′-exo to O4′-endo range (P = 54–90°). Furthermore, the hybrid segment of this duplex is more heteronomous than previously assumed for duplexes of the $\frac{r-D}{D-D}$ type. In particular, the hybrid base pairs formed by the last and next to last RNA residues (4a·13T and 3c·14G, respectively) appear to be more heteronomous than the junction base pair 5C·12G as well as the two other hybrid base pairs (1g·16C and 2c·15G) in the hybrid half of this hybrid–chimeric duplex. These observations suggest the interesting possibility that, with respect to the RNase H cleavage of the $tRNA^{Lys}$ primer at the c–a position, the specificity of the HIV-1 reverse transcriptase RNase H activity may be related, at least in part, to the very high heteronomy of these base pairs in the hybrid segment of this duplex. The crystal structure of the similar hybrid duplex–DNA duplex (gcg-TATACCC)·(GGGTATACGC) was found to be A-form with all sugars, including the deoxynucleosides, in the N conformation [Egli, M., Usman, N., Zhang, S., & Rich, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89,* 534–538). NMR studies of this hybrid–chimera in solution lead to completely different conclusions that are similar to those in the present study, suggesting that the crystal structure reflects the effects of partial dehydration rather than the actual structure in solution.

The structure of nucleic acid acid duplexes in the A-form, the B-form, and the Z-form has been solved by X-ray diffraction methods. RNA duplexes invariably have A-form structure in both the crystalline state (Langridge & Gomatos, 1963; Tomita & Rich, 1964; Arnott et al., 1968; Dock-Bregeon et al., 1989) and the solution state (Chou et al., 1989a). When fully hydrated, DNA typically exists in the B-form in fibers (Franklin & Gosling, 1953) and in single crystals (Dickerson & Drew, 1981; Heineman & Alings, 1989), although most DNA crystal structures up to 1981 were found to be in the A-form (Shakked & Kennard, 1985; Kennard & Hunter, 1989)—an effect attributed to "hydration economy" (Saenger et al., 1986) during the partial dehydration that occurs when DNA is crystallized from aqueous organic alcohol solutions. These crystalline A-form DNAs have been shown by NMR methods to have B-form structure in solution (Reid et al., 1983a). A-form duplexes are characterized by sugar conformations in the N, or C3′-endo, range while B-form duplexes have S, or C2′-endo, sugar conformations (Saenger, 1984). The question of duplex morphology, and hence sugar conformations, in DNA–RNA hybrid duplexes is the subject of controversy. Claims have been made that the double-helix strand structure is A·A (O'Brien & MacEwan, 1970; Gray & Ratliff, 1975; Wang et al., 1982), B·A (Zimmerman & Pheiffer, 1981; Arnott et al., 1986), and even B·B (Reid et al., 1983b; Gupta et al., 1985) in hybrid duplexes. NMR studies from this laboratory (Salazar et al., submitted) suggest that none of these claims is correct and that the DNA strand in hybrid duplexes has neither B nor A structure; i.e., the sugar conformation is neither S nor N (nor is it an N–S equilibrium). In light of the confusion concerning simple hybrid duplexes, it is not surprising to learn that the situation is even more confused concerning chimeric duplexes where half of the double helix is a hybrid duplex and the other half is pure DNA (or RNA). Such duplexes have only recently become amenable to physicochemical study, yet they are important intermediates in retrovirus replication and in normal DNA replication of the lagging strand (Okazaki fragments).

The human immunodeficiency virus type 1 (HIV-1) replicates via a double-stranded DNA intermediate that contains long terminal repeats (LTRs) at each end (Gilboa et al., 1979; Gallo, 1990; Kulkosky et al., 1990). The DNA LTRs are derived from the sequences U5 and U3 located

† This work is supported by National Institutes of Health Grants GM-42896 and GM-32681 (B.R.R.) and CA-51605 (J.J.C.). M.S. acknowledges the generous support of the W. M. Keck Foundation in the form of a predoctoral fellowship.

‡ Department of Chemistry.
§ Department of Microbiology.
∥ Department of Biochemistry.

0006-2960/93/0432-739$04.00/0

Chart I[a]

PBS U5

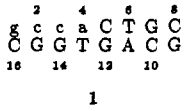
5′ $tRNA^{Lys}$···gucccuguucgggcgccaCTGCTAGAGA···(−) DNA 3′
CAGGGACAAGCCCGCGGTGACGATCTCT···(+) DNA

[a] The lower case letters represent RNA residues from the 3′ tail of the $tRNA^{Lys}$ primer.

adjacent to the viral RNA 5′ and 3′ termini, respectively, together with repeated sequences (R) derived from the actual RNA termini; at the double-helical DNA level, the LTRs are arranged 5′-U3-R-U5···. The double-stranded DNA intermediate is produced by the action of reverse transcriptase on the viral RNA and is subsequently integrated into the host cell genome. Like all known DNA polymerases, reverse transcriptase requires an RNA primer to initiate DNA synthesis. Following infection of the host cell, the first step is the synthesis of a minus-strand DNA (cDNA) that is complementary to the single-stranded viral RNA. The synthesis of this first DNA strand is initiated by a cellular tRNA primer, $tRNA^{Lys}$, which binds to a complementary viral RNA sequence called the primer binding site (PBS) located near the 5′ end of the RNA HIV-1 genome.

The sequence at the PBS–U5 junction, after completion of minus-strand DNA synthesis and in the early stages of plus-strand DNA synthesis, is shown in Chart I (Kulkosky et al., 1990). Synthesis of the (+)-strand DNA is itself primed by a polypurine-rich RNA fragment of the viral RNA generated by the action of the reverse transcriptase RNase H on the intermediate RNA–DNA hybrid (Finston & Champoux, 1984). All known mature tRNAs end in cca-OH. The deoxynucleotides covalently added to this 3′-terminal sequence ultimately define the U5 terminus of the unintegrated viral DNA (Gilboa et al., 1979; Kulkosky et al., 1990; Whitcomb et al., 1990; Pullen & Champoux, 1990). The fact that most RNase H enzymes preferentially cleave chimeric RNA–DNA sequences at the junction phosphodiester led initially to the prediction that the right LTR (U5) end of the unintegrated viral DNA should start with the sequence 5′-CTGC···3′—see Chart I. However, recent studies have shown that the major RNase H product is actually 5′-aCTGC···3′; i.e., the unusual specificity of the RNase H results in cleavage of the c–a phosphodiester bond within the $tRNA^{Lys}$ RNA segment, leaving behind a riboadenosine residue at the 5′ end of the right LTR of the proviral DNA (Kulkosky et al., 1990; Whitcomb et al., 1990; Pullen & Champoux, 1990; Furfine & Reardon, 1991; Pullen et al., 1992). Investigations of HIV-1 primer removal by avian myeloblastosis virus (AMV) and Moloney murine leukemia (Mo-MuLV) reverse transcriptase suggest that the specificity for primer removal resides with the substrate since these two enzymes also cleave at the c–a phosphodiester bond within the HIV-1 $tRNA^{Lys}$ primer (Pullen et al., 1992).

In order to gain insight into this unusual RNase H specificity at the PBS–U5 junction, we decided to investigate whether any of the sugars, particularly the ribose sugars, at or near the cleavage site might be in an unexpected or unusual conformation. We therefore carried out an analysis of the conformations of these sugars using well-established two-dimensional ^1H NMR methods. Due to the nonsymmetrical nature of this duplex junction, we decided to study the hybrid–DNA octamer duplex (**1**), in order to minimize spectral



1

overlap. Additional motivation for studying this particular duplex was the paucity of information regarding the sugar geometries in chimeric–hybrid duplexes of the $\frac{r-D}{D-D}$ type (where r and D represent RNA and DNA segments, respectively) especially at the junction between the hybrid and pure DNA segments. In addition to their role in retrovirus reverse transcription, duplexes of this type are also excellent models for RNA-primed synthesis of Okazaki fragments, which are involved in replicating the lagging strand during DNA synthesis (Ogawa & Okazaki, 1980). Theoretical models for the structure of the hybrid duplex–DNA duplex junction (including sugar conformations) in these types of duplexes have been derived from NMR and circular dichroism studies on duplexes of the type $dG_n{\cdot}rC_jdC_k$ (Selsing & Wells, 1979; Selsing et al., 1979). However, such homopolymer studies suffer from severe spectral overlap and reveal little detailed experimental data on individual sugar conformations at the junction. Previous studies on doubly chimeric h-D-h decamer sequences of the type $\frac{r_3-D_4-D_3}{D_3-D_4-r_3}$ have been carried out both in the crystalline state (Wang et al., 1982) and in solution (Mellema et al., 1983) with conflicting conclusions. However, such duplexes are not relevant to $\frac{r-D}{D-D}$ Okazaki fragments in that they contain flanking hybrid duplexes on *both* sides of the pure DNA segment, and the downstream flanking hybrid segment leads to severe distortion of the entire four base pair central DNA section (Salazar and Reid, submitted). The more relevant hybrid–chimera (gcgTATACCC)·(GGGTATACGC) containing three base pairs of hybrid duplex covalently linked to seven base pairs of pure DNA duplex has been crystallized and the structure solved by X-ray diffraction (Egli et al., 1992). The surprising observation that *all 20* nucleotides are in the A-conformation (with N sugars) prompted the present study.

MATERIALS AND METHODS

Sample Preparation. Ten micromole syntheses of the pure DNA strand 5′-d(GCAGTGGC) and the chimeric RNA–DNA strand 5′-r(gcca)d(CTGC) were carried out on an automated DNA synthesizer (Applied Biosystems Model 392). The DNA strand was prepared as described previously (Hare et al., 1986). The RNA–DNA strand was synthesized by introducing the ribonucleoside phosphoramidites in the appropriate cycles as previously described (Ogilvie et al., 1988; Chou et al., 1989a,b, 1991) using base-protected 5′-DMT 3′-methyl(*N,N*-diisopropylphosphoramidites) protected with *tert*-butyldimethylsilyl (TBDMS) groups at the 2′-position; the phosphoramidites were purchased from Biogenex/ABN and Applied Biosystems. RNA deprotection and cleavage from the support were carried out with anhydrous methanolic ammonia for 18–24 h. The 2′-TBDMS group was removed from the RNA by dissolving the chimeric strands in a 1 M solution of tetrabutylammonium fluoride (TBAF) in tetrahydrofuran and letting the solution stand overnight at room temperature. Major impurities were then removed by ethanol precipitation. Any residual TBAF was removed by sodium form ion-exchange column chromatography (AG 50W X-2, Bio-Rad). Further purification was achieved by size exclusion column chromatography (Sephadex DNA grade, superfine, Pharmacia). Column fractions were analyzed by microscale electrophoresis on a 20% polyacrylamide gel and pooled accordingly. The pure DNA and the chimeric strands were then annealed by dissolving equal amounts of the two strands in 500 mM NaCl, 0.2 mM EDTA, and sodium phosphate buffer, pH 6.8, heating to 55 °C, and slowly cooling to room temperature. The sample was subsequently desalted via

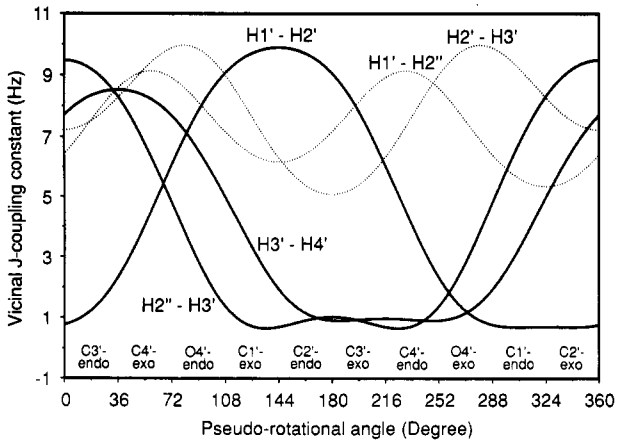


FIGURE 1: Three-bond coupling constants, $^3J_{HH}$, as a function of the sugar pseudorotational phase angle, P. A modified Karplus equation that incorporates electronegatively and orientation effects with a puckering amplitude of 38.5° was used to generate the plots [reproduced from Kim et al. (1992) and derived from Haasnoot et al. (1980)].

Sephadex G-10 column chromatography. The purified (gccaCTGC)·(GCAGTGGC) duplex (ca. 22 mg) was dissolved in 0.35 mL of a buffer consisting of 5 mM KCl, 25 mM NaCl, 0.6 mM EDTA, 25 mM NaH_2PO_4, and 25 mM Na_2HPO_4, adjusted to pH 7.0. After repeated lyophilization in 99.96% D_2O the sample was redissolved in 0.35 mL of 99.996% D_2O for all the NMR experiments.

NMR Spectroscopy. All NMR spectra were collected either on a Bruker AM-500 (E.COSY) or on a home-built 500-MHz spectrometer. The data were processed on an IRIS 4D computer with either the FTNMR or the FELIX software programs (Hare Research, Woodinville, WA).

The NOESY spectra were collected in the phase-sensitive hypercomplex mode (States et al., 1982) at 25 °C with a mixing time of 100 ms and a 10-s relaxation delay. A total of 400 pairs of real and imaginary t_1 experiments were collected with 1024 complex points in t_2.

The DQF-COSY spectrum was collected in the phase-sensitive mode with time-proportional phase incrementation (Drobny et al., 1979; Marion & Wüthrich, 1983). A total of 778 t_1 experiments with 2048 complex points in t_2 were collected with 32 scans per t_1 experiment and a 5-s relaxation delay.

The E.COSY spectrum (Griesinger et al., 1986, 1987) was collected in the TPPI mode. A total of 760 t_1 experiments with 2048 complex points in t_2 were collected with 48 scans per t_1 experiment and a relaxation delay of 2.2 s. The spectral width was 4386 Hz. After zero filling to 4096 complex points the resolution is 1.07 Hz/point in t_2.

Sugar Conformations. The procedure for estimating the sugar geometry has been described previously (Kim et al., 1992). Briefly, the sugar torsion angles, ν_i, are related to the pseudorotational phase angle (P) and the puckering amplitude (T_m) as $\nu_i = T_m \cos[P + 144(j-2)]$, where $j = 0$–4 (Haasnoot et al., 1980). The individual torsion angles within a sugar can be obtained from the corresponding H–C–C–H dihedral angles determined from the three-bond coupling constant, $^3J_{HH}$, using either a simple Karplus equation or a modified Karplus relationship that incorporates electronegativity and orientation effects (Haasnoot et al., 1980; van de Ven & Hilbers, 1988; Kim et al., 1992). To facilitate discussion of the present results, the empirically calibrated $^3J_{HH}$ values as a function of the sugar conformations (P values) are reproduced from Kim et al. (1992) in Figure 1—the electronegativity/orientation parameters of Haasnoot et al. (1980) are incorporated into these plots. The coupling constants $J_{1'-2'}$, $J_{2''-3'}$, and $J_{3'-4'}$ are the most useful in that they vary widely between $P \approx 0°$ (N-type) and $P \approx 180°$ (S-type) sugar conformations. The combination of these three coupling constants is very diagnostic and was used to estimate the ranges of allowable sugar conformations in the hybrid–DNA duplex **1**. While the quantitative value of a particular ^3J for a given sugar conformation (P) may not be exactly correct, the qualitative pattern of several J-couplings taken together is remarkably restrictive in constraining the sugar conformation to a relatively narrow range of allowed conformers.

RESULTS

Our first goal was to establish the RNA sugar geometries in the top strand. In a ribose sugar, only in the C4′-exo to C1′-exo sugar conformational range (P = 54–144°) can the H4′ proton be closer or as close as the H2′ proton is to the H1′ proton; no other proton can approach the H1′–H2′ proximity regardless of sugar conformation. Thus in the RNA H1′ to H2′/H3′/H4′/H5′/H5″ region of the NOESY spectrum, the strongest cross-peaks should always be the H1′–H2′ peaks and perhaps also the H1′–H4′ cross-peaks (depending on the conformation). Figure 2A shows the RNA H1′–H2′/H3′/H4′/H5′/H5″ region of the NOESY spectrum at 100 ms. By far the strongest cross-peaks in this region are the H1′–H2′ RNA cross-peaks [the H2′ assignments were confirmed by strong H6/H8–(n–1)H2′ cross-peaks], with the H1′–H4′ cross-peaks much weaker and undetectable at this contour level. Thus the sugar P value is either <54° or >144°. Figure 2B shows this same region in a phase-sensitive double-quantum-filtered COSY (DQF-COSY) spectrum. At the top of Figure 2B one observes the DNA H3′–H4′ cross-peaks (vide infra), but no H1′–H2′ DQF-COSY cross-peaks are detectable for *any* of the RNA resides, indicating a H1′–C1′–C2′–H2′ dihedral angle close to 90° and a P value <30°. Thus all the RNA sugars in the hybrid half of this duplex assume a conformation in the general N-domain ($J_{1'-2'} < 2$ Hz, $P = 0 \pm 30°$).

To estimate the DNA sugar conformations, the diagnostic J-couplings described above were used. Figure 2C shows the H1′–H2′/H2″ region of the E.COSY spectrum. The H1′–H2′ and the H1′–H2″ coupling constants were measured from the passive couplings in the H1′–H2″ and H1′–H2′ cross-peaks, respectively. The most notable feature in this spectrum is the rather weak intensity of the cross-peaks for 5C, i.e., the chimeric strand junction DNA residue. From the H1′–H2′ passive coupling in the H1′–H2″ cross-peak, the 5C H1′–H2′ coupling constant was estimated to be 5.3 Hz. Thus, from Figure 1, dC5 at the RNA–DNA junction assumes a sugar conformation somewhere in the C4′-exo to O4′-endo range (P = 54–90°). This conclusion was further confirmed by the presence of strong H3′–H4′ (see below) *and* H2″–H3′ DQF-COSY cross-peaks (data not shown)—see Figure 1. This result differs markedly from the model study prediction of Selsing et al. (1979) that in duplexes of the $\frac{\text{r–D}}{\text{D–D}}$ type the first (junction) DNA residue of the chimeric r-D strand should have a C2′-endo (S-type) conformation. The other notable feature in Figure 2C is the readily detectable H1′–H2′ cross-peaks for the DNA residues 13T, 14G, 15G, and 16C in the hybrid half of this duplex. Clearly, from Figure 1, these residues are *not* in the general N-domain, as proposed by Selsing et al. (1979). This observation was confirmed by (n)-H6/H8 to (n)H2′ NOEs of stronger magnitude than the (n)-H6 to (n)H3′ NOEs at short mixing times (data not shown); the latter should be stronger in A-type conformations with N sugar conformations. Also, from their modeling studies Selsing

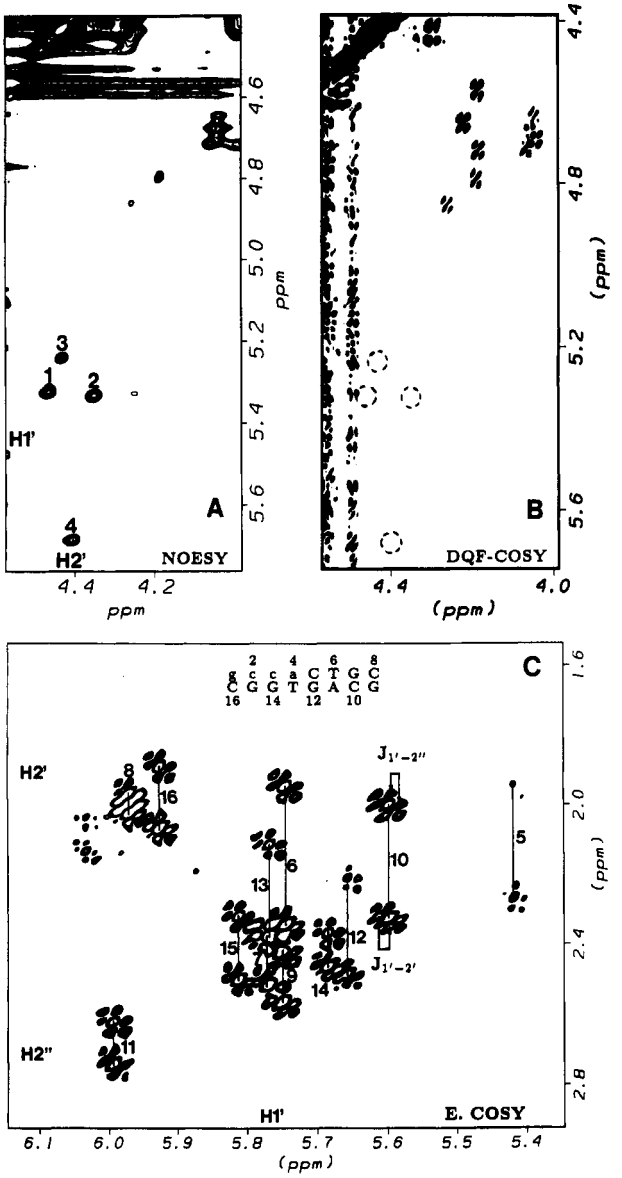


FIGURE 2: (A) RNA H1'–H2'/H3'/H4'/H5'/H5" region of the NOESY spectrum of (gccaCTGC)·(GCAGTGGC) at 100-ms mixing time. The bottom of the spectrum shows the H1'–H2' NOE cross-peaks. (B) DQF-COSY spectrum of the same region as in (A). The resolution is 2.14 Hz/point. The dashed circles represent the regions of the spectrum where the H1'–H2' cross-peaks for the RNA sugars should appear if the coupling is strong. On the basis of a 2-Hz cutoff, and undetectable $J_{1'-2'}$ corresponds to $P = 0 \pm 30°$ (see Figure 1). (C) DNA H1'–H2'/H2" above-diagonal region of the E.COSY spectrum of (gccaCTGC)·(GCAGTGGC). The resolution is 1.07 Hz/point. The lines connect the H1'–H2' and the H1'–H2" cross-peaks for each residue. $J_{1'-2'}$ and $J_{1'-2''}$ were measured from the splittings in the respective H1'–H2" and H1'–H2' cross-peaks as shown for residue 10C.

et al. (1979) reported that the pure DNA (bottom) strand residue complementary to the chimeric strand junction DNA residue (12G in the present study) should adopt a C3'-endo (N-type) conformation in duplexes of the $\frac{r-D}{D-D}$ type. In Figure 2C the H1'–H2' cross-peak for residue 12G is readily detectable (i.e., not C3'-endo) but is weaker than that of 10C, 11A, or 13T. Therefore, from Figure 1, it is clear that, at least for the present sequence, the sugar of the pure DNA strand residue complementary to the junction DNA residue of the chimeric strand *does not* assume a C3'-endo N-type conformation as proposed by Selsing et al. (1979). From the H1'–H2' and H3'–H4' couplings, this residue (12G) appears to have an E-type (ca. O4'-endo) conformation. Table I lists the H1'–H2' and the H1'–H2" coupling constants for the 12 DNA residues of the hybrid chimera. With the possible exception of 5C, 7G, and 13T, all the deoxyribose H1'–H2' coupling constants are either of equal magnitude to or of greater magnitude than the H1'–H2" coupling constants. From Figure 1 this indicates that the sugar conformations in these residues lie in the O4'-endo to C3'-exo range (P = 72–216°). This is in sharp contrast to the observation that all DNA sugars are in the N conformation ($P = 20 \pm 15°$) in hybrid duplex–DNA duplex chimeras in the crystalline state (Egli et al., 1992). The sugar geometries were further constrained from the intensities of the H3'–H4' DQF-COSY cross-peaks. From Figure 1 the observation of readily detectable H3'–H4' DQF-COSY cross-peaks indicates a P value of <144° ($J_{3'-4'}$ > 2 Hz). As shown in Figure 3, all the DNA H3'–H4' cross-peaks were readily observed. Thus all the DNA sugars must be in the O4'-endo to C1'-exo range (P = 72–144°).

The H2"–H3' COSY or DQF-COSY cross-peaks are also very useful in narrowing the range of allowed sugar geometries. Because of the presence of passive couplings to phosphorus and other protons, the H2"–H3' coupling constant could not be accurately measured and only the relative H2"–H3' cross-peak intensities are listed in Table I. Thus the absence of detectable H2"–H3' DQF-COSY cross-peaks for residues 11A, 10C, and 9G indicates P values >100° ($J_{2''-3'}$ < 2 Hz) for these residues. The sugar conformations for the other DNA residues, which showed medium to strong H2"–H3' cross-peaks, were estimated from the relative intensities of these cross-peaks combined with the relative H3'–H4' cross-peak intensities in Figure 3. The H2"–H3' cross-peak intensities for residues 6T and 13T could not be determined due to spectral overlap. However, these sugar geometries were estimated by comparison of their H3'–H4' DQF-COSY cross-peak intensities with other more resolved residues. Thus, although the H3'-H4' cross-peak of residue 6T is slightly stronger, it is of comparable intensity to that of 14G (P = 72–100°). Likewise the H3'–H4' cross-peaks of 13T and 10C are of similar intensity (P = 100–144°).

The sugar geometries for all residues, determined as described above, are listed in Table I. The lower P values for residues 5C and 12G at the junction were confirmed by (n)-H6/H8 to (n)H2' NOEs at short mixing times (50 ms) which are much weaker than the corresponding NOEs for the other DNA residues (except for 15G and the terminal residues).

DISCUSSION

The above analysis assumes a relatively small conformational range for each nucleoside and ignores the possibility of C2'-endo to C3'-endo interconversions for the DNA residues. Although N–S equilibria ($\Delta P \approx 150°$) have been proposed to rationalize the inability to fit observed couplings to empirically calibrated J-couplings (Haasnoot et al., 1980; van de Ven et al., 1988; Rinkel & Altona, 1987; Schmitz et al., 1990), we note that recent ^{2}H solid-state NMR studies argue against the existence of large amplitude motions for DNA sugars and thus against the existence of N–S sugar ring interconversions (Huang et al., 1990). Thus, we feel that it is not necessary to invoke such interconversion to explain the observed experimental couplings; a similar conclusion has been reached by Gochin et al. (1990).

From the data in Figures 2 and 3, the base pairs between residues 3c and 14G and between residues 4a and 13T appear to be more heteronomous than the other hybrid base pairs in that residues 3c and 4a have N-type sugars ($P \approx 15°$) while their complements 14G and 13T have S or SE ($P \approx 122°$) sugar conformations (ΔP = ca. 107°)—see Table I. This

Table I: Coupling Constants (Hz) and Estimated Sugar Geometries in (gccaCTGC)·(GCAGTGGC)[a]

	1g	2c	3c	4a	5C	6T	7G	8C
$J_{1'-2'}$	≤2	≤2	≤2	≤2	5.3	6.8	o	6.6
$J_{1'-2''}$					x	6.2	o	6.6
$J_{2''-3'}$					s	o	u	s
$J_{3'-4'}$					s	m	m (–)	s
geometry	C3′-endo	C3′-endo	C3′-endo	C3′-endo	C4′-exo/O4′-endo	O4′-endo/C1′-exo	C1′-exo	C4′-exo/O4′-endo
P value (deg)	0 ± 30	0 ± 30	0 ± 30	0 ± 30	54–90	90–126	108–144	54–90

	16C[h]	15G[h]	14G[h]	13T[h]	12G	11A	10C	9G
$J_{1'-2'}$	6.5	7.3	7.0	o	7.8	7.2	8.0	8.2
$J_{1'-2''}$	6.6	6.4	6.0	6.4	6.2	6.4	6.1	o
$J_{2''-3'}$	s	m	u	o	m	u	u	u
$J_{3'-4'}$	s	m (+)	m (–)	m (–)	m (+)	w	m (–)	w
geometry	C4′-exo/O4′-endo	O4′-endo	C1′-exo	C1′-exo	O4′-endo	C1′-exo	C1′-exo	C1′-exo
P value (deg)	54–90	72–100	90–144	100–144	72–100	100–144	100–144	100–144

[a] J-couplings were obtained from the t_2 dimension (resolution = 1.07 Hz/point) of the E.COSY spectrum shown in Figure 2C. The symbol h designates DNA residues in the hybrid segment of the duplex; u denotes undetectable crosspeak; x denotes an E.COSY cross-peak too weak to measure the coupling; o denotes J-couplings that could not be measured due to spectral overlap. The H2″–H3′ and H3′–H4′ couplings could not be accurately measured due to passive couplings to phosphorus and other protons, and only the relative intensities are given (s = strong, m = medium, w = weak) with (+) or (–) to fine tune the qualitative observation.

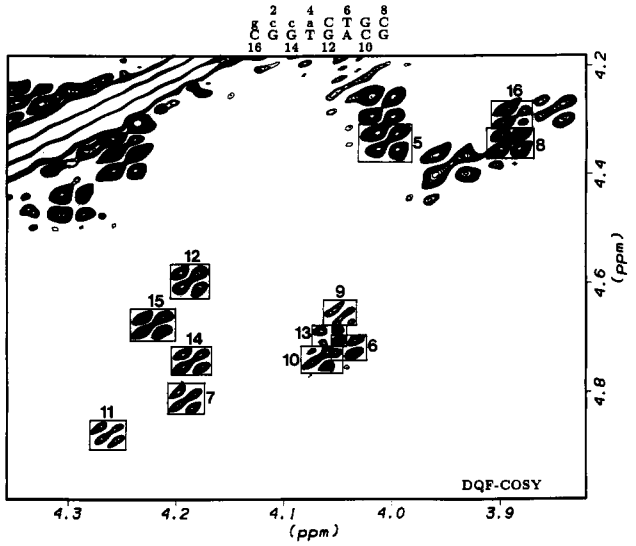


FIGURE 3: H3′–H4′ region of the double-quantum-filtered COSY (DQF-COSY) spectrum of (gccaCTGC)·(GCAGTGGC). After zero filling the resolution is 1.07 Hz/point in t_2. The resonances were independently assigned from a NOESY spectrum. The active coupling is the major contributor to cross-peak intensity. On the basis of a 2-Hz cutoff, an H3′–H4′ cross-peak of undetectable intensity would indicate a sugar P value of >144°.

observation suggests the interesting possibility that, with respect to the observed cleavage of the $tRNA^{Lys}$ primer at the c–a position, the specificity of the RNase H activity of reverse transcriptase may, at least in part, be related to the extent of heteronomy in the 3c·14G and the 4a·13T base pairs compared to the 5C·12G junction base pair (ΔP = ca. 14°) and the 2c·15G and the 1g·16C base pairs (ΔP = ca. 68°) in the hybrid section of the duplex. However, we note that it is entirely possible that other structural factors may be acting in a synergistic fashion to direct cleavage at the c–a position. We are currently in the process of determining the three-dimensional solution structure of this duplex in order to further constrain the spatial relationships of the nucleotides using NOESY buildup rates and to determine if other structural factors at or near the cleavage site may be involved in directing RNase H cleavage at the 3c–4a step. Unfortunately, due to the small size of this octamer, it is difficult to conveniently assay its cleavage by RNase H using standard gel electrophoresis assays. Nevertheless, at 25 °C end effects should not significantly perturb the structure of the central hexamer within this octamer, particularly at the junction. We are currently investigating RNase H activity in vitro on the longer decamer (gcgccaCTGC)·(GCAGTGGCGC) in conjunction with structural studies. We are also beginning studies on other Okazaki fragment models related to retroviruses to determine what, if any, are the sequence effects on the conformation of the RNA and DNA sugars, particularly at the hybrid duplex–DNA duplex junction. Finally, we note that Egli et al. (1992) very recently reported the crystal structure of the Okazaki fragment model duplex (gcg-TATACCC)·(GGGTATACGC). They found that, in the crystalline state, this hybrid chimeric duplex assumes a uniform *A-type* conformation throughout the entire duplex with no structural discontinuities at the hybrid duplex–DNA duplex junction. While these authors argue that the RNA trinucleotide in the three base pair hybrid segment converts the seven base pair adjoining DNA segment into A-type (N sugar) geometry, it remains entirely possible that this transition was in fact caused by the partially dehydrating conditions of crystal growth (Saenger et al., 1986; Kennard & Hunter, 1989). We have carried out J-coupling analysis and NOE studies similar to those described above on the same (gcgTATACCC)·(GGGTATACGC) decamer sequence used in the crystal studies (Salazar et al., in preparation). The results indicate a conformation for the RNA and DNA sugars similar to the sugar conformations found in (gccaCTGC)·(GCAGTGGC) reported in the present study. This observation strongly argues against an A-form conformation for Okazaki fragments in solution.

ACKNOWLEDGMENT

We thank Dr. Seong-Gi Kim for technical discussions and N. Susan Ribeiro and Julie M. Miller for DNA and RNA synthesis.

REFERENCES

Arnott, S., Fuller, W., Hodgson, A., & Prutton, I. (1968) *Nature 220*, 561–564.

Arnott, S., Chandrasekaran, R., Millane, R. P., & Park, H.-S. (1986) *J. Mol. Biol. 188*, 631–640.

Chou, S. H., Flynn, P., & Reid, B. R. (1989a) *Biochemistry 28*, 2422–2435.

Chou, S. H., Flynn, P., & Reid, B. R. (1989b) *Biochemistry 28*, 2435–2443.

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

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.
Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J. S., Gough, G. R., Gilham, P. T., & Moras, D. (1989) *J. Mol. Biol. 209*, 459–474.
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.
Finston, W. I., & Champoux, J. J. (1984) *J. Virol. 51*, 26–33.
Furfine, E. S., & Reardon, J. E. (1991) *Biochemistry 30*, 7041–7046.
Gallo, R. C. (1990) *J. Acquired Immune Defic. Synd. 1*, 380–389.
Gilboa, E., Mitra, S. W., Goff, S., & Baltimore, D. (1979) *Cell 18*, 93–100.
Gochin, M., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 11161–11171.
Gray, D. M., & Ratliff, R. L. (1975) *Biopolymers 14*, 487–498.
Griesinger, C., Sørensen, O. W., & Ernst, R. R. (1986) *J. Chem. Soc. 85*, 6837–6852.
Griesinger, C., Sørensen, O. W., & Ernst, R. R. (1987) *J. Magn. Reson. 75*, 474–492.
Gupta, G., Sarma, M. H., & Sarma, R. H. (1985) *J. Mol. Biol. 186*, 463–469.
Haasnoot, C. A. G., de Leeuw, F. A. A. M., & Altona, C. (1980) *Tetrahedron 36*, 2783–2792.
Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G. P., & Reid, B. R. (1986) *J. Mol. Biol. 171*, 319–336.
Heinemann, U., & Alings, C. (1989) *J. Mol. Biol. 210*, 369–381.
Huang, W.-C., Orban, J., Kintanar, A., Reid, R. R., & Drobny, G. P. (1990) *J. Am. Chem. Soc. 112*, 9059–9068.
Kennard, O., & Hunter, W. N. (1989) *Q. Rev. Biophys. 22*, 327–379.
Kim, S.-G., Lin, L.-J., & Reid, B. R. (1992) *Bichemistry 31*, 3564–3574.
Kulkosky, J., Katz, R. A., & Skalka, A. M. (1990) *J. Acquired Immune Defic. Synd. 3*, 852–858.
Langridge, R., & Gromatos, P. (1963) *Science 141*, 694–698.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
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.
O'Brien, E. J., & MacEwan, A. W. (1970) *J. Mol. Biol. 48*, 243–261.
Ogawa, T., & Okazaki, T. (1980) *Annu. Rev. Biochem. 49*, 421–457.
Ogilvie, K. K., Usman, N., Nicoghosian, K., & Cedergren, R. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5764–5768.
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.
Reid, D. G., Salisbury, S. A., Bellard, S., Shakked, Z., & Williams, D. H. (1983a) *Biochemistry 22*, 2019–2025.
Reid, D. G., Salisbury, S. A., Brown, T., Williams, D. H., Vasseur, J.-J., Rayner, B., & Imabach, J.-L. (1983b) *Eur. J. Biochem. 135*, 307–314.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621–631.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Saenger, W., Hunter, W. N., & Kennard, O. (1986) *Nature 324*, 385–388.
Schmitz, U., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 2357–2368.
Selsing, E., & Wells, R. D. (1979) *J. Biol. Chem. 254*, 5410–5416.
Selsing, E., Wells, R. D., Alden, C. J., & Arnott, S. (1979) *J. Biol. Chem. 254*, 5417–5422.
Shakked, Z., & Kennard, O. (1985) *Biological Macromolecules and Assemblies* (Jurnak, F. A., & McPherson, A., Eds.) pp 1–36, Wiley Interscience, New York.
States, D. J., Haberkorn, R. A., & Reuben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Tomita, K.-I., & Rich, A. (1964) *Nature 201*, 1160–1163.
van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Wang, A. J.-R., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A. A., & Rich, A. (1982) *Nature 299*, 601–604.
Whitcomb, J. M., Kumar, R., & Hughes, S. H. (1990) *J. Virol. 64*, 4903–4906.
Zimmerman, S. B., & Pheiffer, B. H. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 78–82.