Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a pseudo-16-mer DNA with stacked guanines and two G–A mispairs complexed with the N-terminal fragment of Moloney murine leukemia virus reverse transcriptase

Marie L. Coté and Millie M. Georgiadis*

Waksman Institute and Department of Chemistry, Rutgers University, 190 Frelinghuysen Road, Piscataway, NJ 08854, USA

Correspondence e-mail: georgiadis@waksman.rutgers.edu

Received 14 March 2001
Accepted 7 June 2001

PDB Reference: pseudo-16-mer DNA, 1ibj.

The X-ray crystal structure at 2.0 Å resolution of a DNA molecule complexed with the N-terminal fragment of Moloney murine leukemia virus reverse transcriptase (MMLV RT) has been determined. This method allows the study of nucleic acids in a unique and largely unfettered environment without the complicated lattice interactions typically observed in DNA-only crystal structures. Molecular-replacement phasing using only the protein provided readily interpretable electron density with no model bias for the DNA. The asymmetric unit of the structure consists of the protein molecule bound to the blunt end of a DNA 6/10-mer, which is composed of a six-base strand (5′-GTCGTC-3′) and a ten-base strand (3′-CAGCAGGGCA-5′), resulting in a six-base-pair duplex with a four-base single-stranded overhang. In the crystal structure, the bases of the overhang reciprocally pair to yield a doubly nicked pseudo-hexadecamer primarily B-form DNA molecule. The pairing between the single strands gives two standard (G–C) Watson–Crick pairs and two G(*anti*)–A(*anti*) mispairs. The mispairs reside in a G–C-rich environment and the three consecutive guanines on the 10-mer impart interesting structural features to the pseudo-hexadecamer, such as the preference for a guanine stack, stretching the C–G base pairs flanking the mispair to the point of loss of intra-base-pair hydrogen bonding. The DNA was designed for the purpose of comparison with a previous structure, which was determined in the same crystal lattice. In all of the authors' previous fragment–DNA complexes, the nucleotide at the blunt-ended 3′-hydroxyl was a purine. Consistent with the proposed mechanistic role of interactions with the 3′-hydroxyl in processive DNA synthesis by RT, it was found that a pyrimidine at this position in the DNA makes indentical interactions with the strictly conserved Gly191 and the main chain of Leu115 of MMLV RT.

1. Introduction

Our structural studies focused on understanding the mechanism of processive DNA synthesis by MMLV RT have led to a novel crystallographic approach to the analysis of nucleic acids. As previously reported, this approach involves complexation of the nucleic acid sequence of interest to the N-terminal fragment of MMLV RT (residues 24–278) and subsequent crystallographic analysis of the resulting complex (Coté *et al.*, 2000). The protein fragment includes the fingers and palm domains of MMLV RT and binds nucleic acid in a novel site in the fingers domain and not in the polymerase active site. The highly conserved residues Asp114, Arg116, Asn119 and Gly191 are involved in the binding of the nucleic acid duplex and may play a mechanistic role in processive

DNA synthesis by RT (Najmudin *et al.*, 2000). The DNA atoms interacting with the protein include minor-groove base atoms and sugar atoms from the $n - 2$ and $n - 3$ template strand positions as well as the 3′-OH of the primer strand, where n is the template base that would pair with an incoming nucleotide. Thus, protein–nucleic acid interactions are limited to the three terminal base pairs of the duplex. Since the fingers domain binding site binds the blunt end of the duplex (which can contain single-stranded overhangs extending beyond the duplex), the intervening nucleotides are free of the sort of contacts that are seen in crystal structures of DNA only. Additionally, the contacts from the protein to the DNA are such that the DNA and its phosphodiester backbone are minimally distorted, to an even lesser extent than that observed in phage 434 repressor–DNA complexes (Aggarwal *et al.*, 1988; Rodgers & Harrison, 1993; Shimon & Harrison, 1993) or λ repressor–DNA complex (Jordan & Pabo, 1988). Also, the interactions observed between the protein and the DNA are possible for any DNA or RNA sequence.

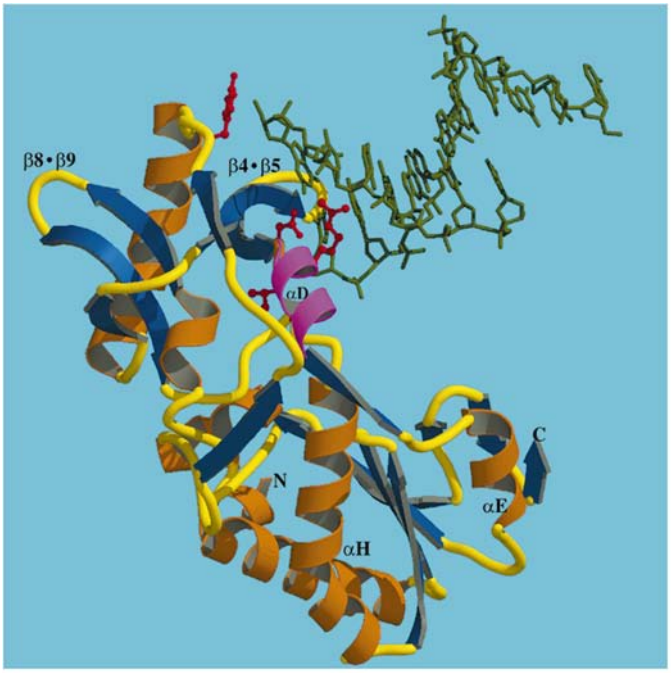
We have previously determined crystal structures for four different DNA duplexes complexed to the N-terminal fragment of MMLV RT in three distinct crystal lattices and find that the protein–DNA interactions are quite similar in each case (Coté *et al.*, 2000; Najmudin *et al.*, 2000). In our form I and II structures, two protein molecules and one DNA 8/8-mer of sequence d(5′-CATGCATG-3′)$_2$ comprise the asymmetric unit. The DNA in a second form II lattice (IIb) has the sequence d(5′-TTTCATGCATG)$_2$ and forms the same 8/8-mer duplex with the additional conformationally variable three-thymidine single strands at each of the 5′ termini (Najmudin *et al.*, 2000). Both of the form I and form II lattices are monoclinic. For historical reasons the third lattice, which is orthorhombic, is referred to as form IV, since preliminary crystallographic studies identified a form III lattice which has not been fully characterized (Sun *et al.*, 1998). The crystal form of the present structure is also that of form IV and hereafter will be referred to as form IVb. The previously determined form IVa structure has a DNA 6/10-mer with oligonucleotide sequences 5′-CTCGTG-3′ and 3′-GAGCACGGCA-5′ (Coté *et al.*, 2000). The form IVb structure shown in Fig. 1 contains another 6/10-mer having the oligonucleotide sequences 5′-GTCGTC-3′ and 3′-CAGCAGGGCA-5′. As in the case of the form IVa structure, the four-base single-stranded overhangs of the 10-mer reciprocally pair in the crystal structure to give a pseudo-hexadecamer. The form IVb pseudo-hexadecamer, which has two G–A mispairs, is shown in Figs. 1(*b*) and 2(*b*).

The DNA sequences comprising the form IVb 6/10-mer and resultant pseudo-hexadecamer were chosen for several reasons. Firstly, in all of our previous MMLV RT fragment–DNA crystal structures, the initial base pair of the DNA duplex was 5′-C–G-3′. In order to address the sequence specificity at the 3′-terminal position of the DNA duplex, the nucleotides for the first base pair of the form IVb DNA duplex were transposed, giving 5′-G–C-3′. In theory, there should be no discrimination between a purine *versus* a pyrimidine in the context of this binding site. In reverse transcriptases, the protein residue that corresponds to Gly191 of MMLV RT is strictly conserved. This Gly191 residue forms a strong hydrogen bond from its peptide oxygen to the 3′-OH of the nucleoside in the initial base pair of the DNA duplexes in our fragment–DNA complexes. Furthermore, modeling studies have shown that any DNA, DNA/RNA or RNA duplex can be accommodated in this fingers domain binding site (INSIGHT II, 1993). Secondly, in the form IVa pseudo-hexadecamer, the nucleotides flanking the mispaired adenine (A7) of the G–A mispair are (5′→3′) G and C (see Fig. 2*a*). This adenine (A7) residing at the 5′ terminus of the 10-mer strand adopts with equal facility either the standard *syn* or *anti* conformation with respect to its ribose ring (Coté *et al.*, 2000). In the present form IVb pseudo-hexadecamer, however, this similarly situated A7 is flanked by two cytosines (see Fig. 2*b*). A consequence of the flanking of A7 by two cytosines is that three consecutive guanines occur on the opposite strand, with the central guanine (G10) as the partner in the mismatch with A7 (see Fig. 2*b*). In constructing the form IVb DNA such that the adenine is flanked by two cytosine nucleosides in the pseudo-hexadecamer, the importance of alignment of bases at a mismatch site is further underscored. Also, as a direct consequence of our design of the form IVb DNA, a purine stack running from 5′→3′ in our form IVb pseudo-hexadecamer ...ACGGGAC... imparts quite different characteristics to the DNA compared with our form IVa pseudo-hexadecamer, which has the sequence ...ACGGCAC... (the excerpted portions having the same positions in each DNA structure). Thirdly, in designing the 6/10-mer the idea was to maintain a G–C-rich structure and modify the already purine-rich 10-mer strand such that a four-purine run would be flanked by two cytosines. The notion was to obtain a doubly nicked pseudo-hexadecamer in order to observe any changes that may occur, such as the A-like characteristics observed in the form IVa structure (Coté *et al.*, 2000).

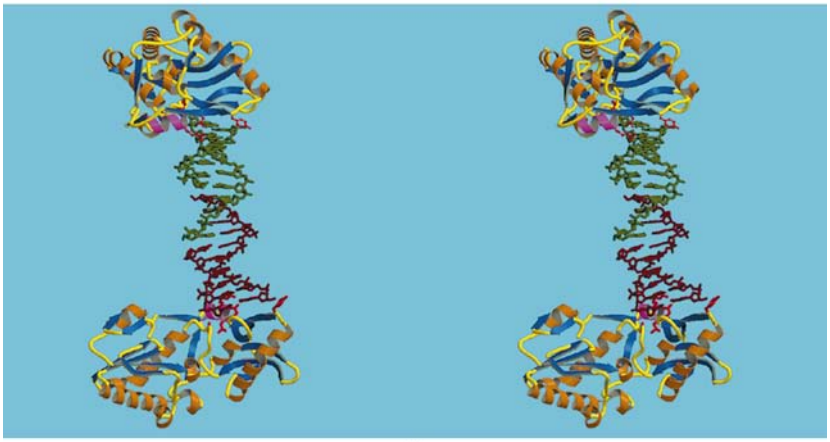
The occurrence of G–A mispairs in DNA presents opportunities for wide variations in structural conformations, both with respect to each partner and to the disposition of each purine with respect to its sugar ring. Five possible conformational motifs for G–A mispairs have been identified in rRNA (Chuprina & Poltev, 1983; Poltev & Shulyupina, 1986; Gautheret *et al.*, 1994). In crystal structures of exclusively B-DNA containing G–A mispairs, the four following conformations have been reported: the standard (non-sheared) G(*anti*)–A(*anti*) (Privé *et al.*, 1987), the sheared G(*anti*)–A(*anti*) (Shepard *et al.*, 1998; Gao, Robinson, Sanishvili *et al.*, 1999), G(*anti*)–A(*syn*) (Brown *et al.*, 1986; Webster *et al.*, 1990) and G(*syn*)–A(*anti*) (Brown *et al.*, 1989). Theoretical studies indicate that there is roughly a 4 kJ mol^{-1} energy difference between the G(*anti*)–A(*anti*) and the G(*anti*)–A(*syn*) conformations when in an intact DNA duplex (Chuprina & Poltev, 1983; Keepers *et al.*, 1984; Poltev & Shulyupina, 1986). In the cases of our form IV pseudo-hexadecamers, however, this energy barrier between the conformations is likely to be smaller since the mismatched adenine is not attached at its 5′ end. Thus, changing the nucleotides flanking the mispair could dispose the adoption of either one distinct conformation to the

exclusion of the others or impart further ambiguity to the conformations of the bases in the mispair.

Our present study further addresses the biological relevance of DNA interactions with the fingers domain binding site of RT. We report here the analysis of the form IVb structure and pseudo-hexadecamer, including a detailed analysis of its G–A mispairs, its protein–DNA interactions and the stacking of the three guanines in the region of a strand break in its DNA. A detailed comparison with a previously reported structure in the same crystalline lattice (form IVa) is presented, providing a context for evaluating specifically the effects of changes in sequence on the structure of the DNA molecule as well as the interactions of the DNA with the fingers domain binding site of RT.



(*a*)



(*b*)

Figure 1
Ribbon renderings of the (*a*) asymmetric unit and (*b*) the pseudo-hexadecamer with its bound protein molecules in the form IVb MMLV RT N-terminal fragment–DNA complex (Kraulis, 1991; Merritt & Bacon, 1997). (*a*) The asymmetric unit consists of the bound protein and the DNA 6/10-mer (depicted in dark green stick models). In the protein, the β-strands are shown in dark blue, the coils in yellow and the α-helices in orange, with the exception of the magenta αD helix. The residues Tyr64, Asp114, Leu115, Arg116 and Gly191 (barely visible) are emphasized with bright red ball-and-stick models. (*b*) The stereoview shown retains the color scheme of Fig. 1(*a*), with the addition of the symmetry-related fragment–6/10-mer (in dark red) to emphasize the pseudo-hexadecamer.

2. Materials and methods

2.1. Crystallization and data collection

The bacterially expressed N-terminal fragment from Moloney murine leukemia virus reverse transcriptase and oligonucleotides were purified as previously described (Sun *et al.*, 1998). The crystals were obtained from a 1:2 molar ratio of protein:DNA. The protein was solubilized in a solution of 0.30 *M* NaCl and 0.10 *M* HEPES pH 7.5; the DNA was solubilized in a solution of 10 m*M* $MgCl_2$ and 10 m*M* HEPES pH 7.0. The complex of protein and DNA was formed at 277 K for 1.0 h, with final concentrations of 0.4 m*M* protein and 0.8 m*M* oligonucleotide. Crystals of the fragment–DNA complex were obtained at 293 K using 1 µl each of complex and crystallization solutions, which consisted of 10% PEG 4000, 0.10 *M* NaCl and 0.05 *M* ADA pH 6.5, in vapor-diffusion hanging drops. The plate-like 210 × 120 × 90 µm form IVb crystals grew in 1 d. Microseeding was required in order to grow large single diffraction-quality crystals.

X-ray crystallographic analysis was performed on an R-AXIS IV image-plate detector with Cu $K\alpha$ radiation at 108 K using an Oxford Cryocool System. Data were collected to Bragg spacings of 2.0 Å, processed with *DENZO* and scaled with *SCALEPACK* (Otwinowski, 1993) including all data. Form IVb crystallizes in the space group $P2_12_12$, with unit-cell parameters $a = 54.26$, $b = 146.45$, $c = 46.71$ Å. The crystals were stabilized in 20% ethylene glycol, 12% PEG 4000, 0.105 *M* NaCl, 0.055 *M* ADA pH 6.5 for cryocooling. Statistics for data collected from a cryocooled crystal are given in Table 1.

2.2. Structure determination and refinement

Since the unit-cell parameters of form IVb so closely resembled those of form IVa, the form IVa protein fragment from its final refined structural model (PDB code 1d1u) was used as the starting model and subjected to rigid-body refinement using the form IVb data. A canonical B-form DNA 6/10-mer with sequences 5′-GTCGTC-3′ and 3′-CAGCAGGGCA-5′ was created using *INSIGHT* II (*INSIGHT* II, 1993) and then positioned and rebuilt into the well defined difference electron density as shown in Fig. 3. Fitting of the DNA and the addition of 204 water O atoms along with minimal rebuilding yielded R_{work} and R_{free} values of 23.8 and 26.2%, respectively, for the form IVb structure using all data in the resolution range 50–2.0 Å and a bulk-solvent correction in *REFMAC* (Murshudov *et al.*, 1997). For evaluation of R_{free},

4.9% of the reflections were excluded from the refinement. At this juncture, electron-density maps revealed virtually no density for A12 of the 10-mer and regions of the contiguous phosphodiester backbone had breaks in the density. Thus, the model was subjected to simulated annealing in *CNS* using all data in the resolution range 50–2.0 Å (Brunger *et al.*, 1998). This caused the DNA to move slightly, especially in the region of the single-stranded overhang of the 10-mer. The electron density remains minimal (though definitive) for the A12 base and regions of sparse density for the contiguous phosphodiester backbone of the single strand persist in the final model. Modest rebuilding and further individual *B*-factor refinement including a bulk-solvent correction in *CNS* gave final $R_{\rm work}$ and $R_{\rm free}$ values of 22.5 and 25.2%, respectively. Standard *CNS* parameter files were used in the refinement. The overall σ_A coordinate error based on *B* is 0.23 Å (0.15 Å) (Read, 1986).

When G–A mispairs occur in a nucleic acid structure, consideration must be given to the possibility that alternate conformations of either one or both bases may occur. In an attempt to confirm the conformations of the bases of the mispair (A7 and G10), models incorporating the standard G(*anti*)–A(*anti*), G(*anti*)–A(*syn*) and G(*syn*)–A(*anti*) conformational possibilities were all separately created and subjected to simulated annealing at the final stages of refinement when the most meaningful analysis can be ascertained. Models with the mispaired guanine (G10) in the *syn* conformation were ruled out based on refinement statistics and examination of difference electron-density maps. In the form

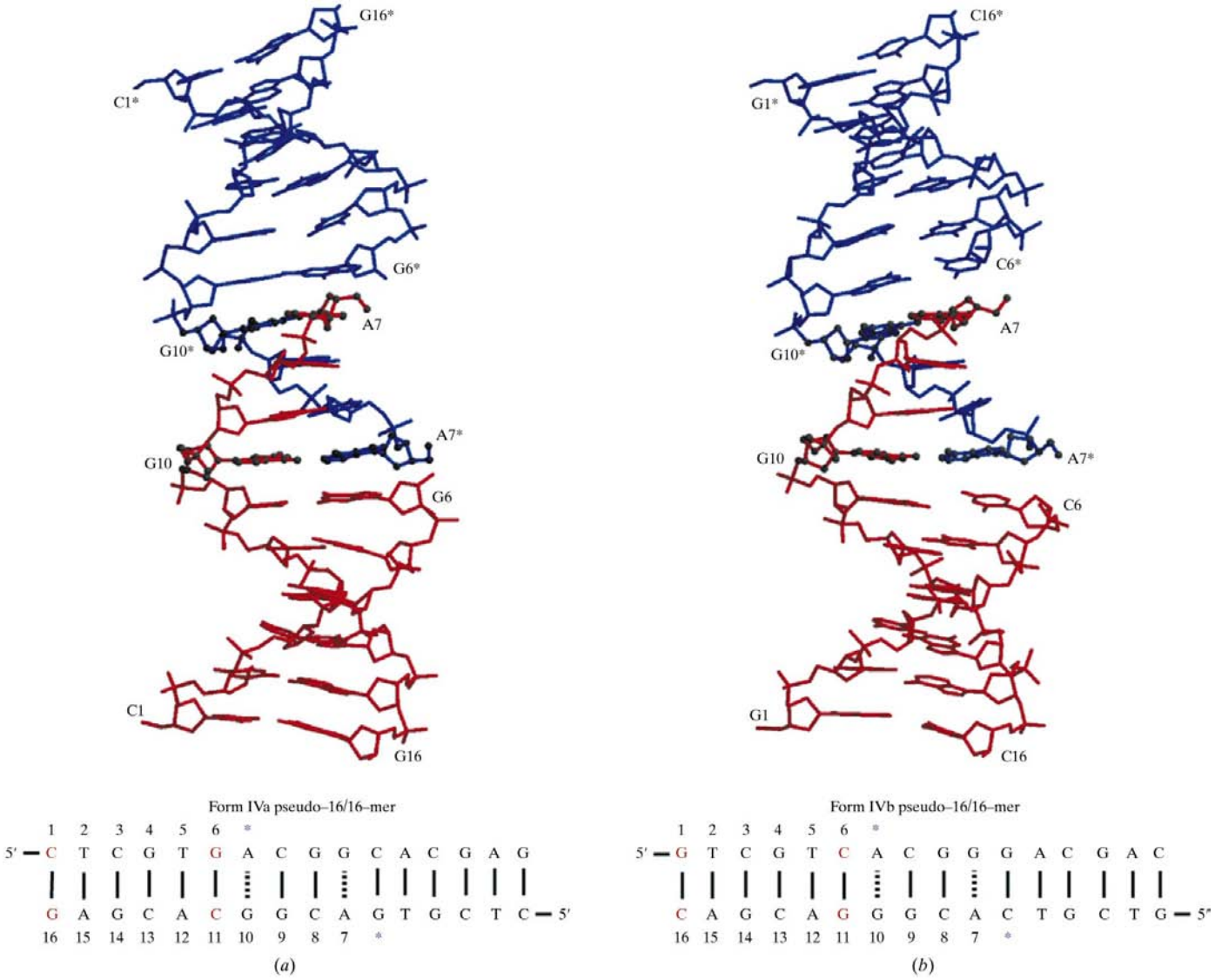


Figure 2
Comparative views (Kraulis, 1991) of the DNA pseudo-hexadecamers in the (*a*) form IVa and (*b*) IVb structures with their numbering schemes. The numbering of the template strands of each 6/10-mer is 1–6 starting from the 5′ end; primer strands are numbered 7–16, also starting from the 5′ end. Nucleotides that differ between the two structures are highlighted in red in their respective legends. Symmetry-related nucleotides are designated with asterisks. The atoms of the G–A mispairs are shown with black spheres and the bases of the mispairs are so labeled. The form IVa DNA in (*a*) retains A7 in its *anti* conformation only for clarity.

Table 1
Data-collection and refinement statistics for form IVb crystals.

Data-collection statistics. Average $I/\sigma(I)$, percentage completeness and R_{sym} are given for all data in the resolution range specified. $R_{sym} = \sum |I - \langle I \rangle| / \sum I$.

Resolution (Å)	Average $I/\sigma(I)$	Completeness (%)	R_{sym}
2.07–2.00	3.4	85.4	0.298
Total, 50.0–2.00	28.2	95.3	0.028

Crystallographic and refinement statistics for the final form IVb model.

Space group	$P2_12_12$
Unit-cell parameters (Å)	
a	54.26
b	146.45
c	46.71
R_{work} (%)	22.5
R_{free} (%)	25.2
R.m.s.d. values	
Bond lengths (Å)	0.005
Bond angles (°)	1.26
Dihedrals (°)	22.8
Improper torsions (°)	1.0
Theoretical reflections	25995
Reflections observed	24781
Unobserved	1214
Test-set reflections	1204
Average *B* factor (all) (Å^2)	44.0
Minimum *B* factor (Å^2)	16.2
Maximum *B* factor (Å^2)	102.3

IVb structure the adenine (A7) of the mispair adopts a single conformation and is best fitted in the *anti* conformation. Additionally, there is no evidence that the sheared G(*anti*)–A(*anti*) conformation occurs in the form IVb structure.

Final verification of the protein model using *PROCHECK* (Laskowski *et al.*, 1993) shows 91.7% of non-Gly/Pro residues residing in most favored regions, 7.9% in additional allowed regions and 0.0% in generously allowed regions. Val223 lies in a disallowed region in form IVb, as seen in the previously reported uncomplexed model of the N-terminal fragment (Georgiadis *et al.*, 1995) as well as our other crystal structures of fragment–DNA complexes (Coté *et al.*, 2000; Najmudin *et al.*, 2000). In the final form IVb model, the electron density in the β4–β5 loop region (residues 100–109) in the protein is ill defined and shows sparse side-chain density for Asp107. Another highly mobile loop region, β8–β9 (residues 173–180), also has weak electron density, with modest side-chain density for Arg173 and virtually none for Met177 (see Fig. 1). There is no apparent side-chain electron density for Tyr64 in the form IVb structure, indicating that it is completely disordered.

3. Results and discussion

3.1. Description of the structure

The protein fragment contains the fingers and palm domains of MMLV RT. The DNA-binding site for the fragment is located in the fingers domain, which principally involves residues comprising the αD helix (see Fig. 1). We have used the same secondary-structural assignments for the protein in form IVb as for the uncomplexed MMLV RT fragment (Georgiadis *et al.*, 1995). The oligonucleotide is bound to conserved residues in much the same manner as that seen in our other crystal structures (Coté *et al.*, 2000; Najmudin *et al.*, 2000).

Table 2
Protein–DNA hydrogen-bonding interactions.

Residue	Atom	Atom	Nucleotide	Distance (Å)
Asp114	$O^{\delta2}$	N2	G1	3.2
Leu115	N	O3′	C16	3.0
Arg116	$N^{\eta1}$	O4′	C3	2.8
		O2	T2	2.9
	$N^{\eta2}$	N2	G1	2.8
		O2	T2	2.6
		O2	C16	3.3
Gly191	O	O3′	C16	3.2

Like the form IVa structure, the form IVb structure has one protein molecule and one DNA 6/10-mer in its asymmetric unit (see Fig. 1*a*). In the form IV crystal structures the 6/10-mer forms a pseudo-hexadecamer through crystallographic symmetry (symmetry element $-x - 1, -y, z$), where the bases of the single-stranded overhang reciprocally pair to yield two Watson–Crick (G–C) base pairs and two G–A mispairs (see Figs. 1*b* and 2). As mentioned, the terminal adenine in the single-stranded overhang (A7) adopts the *anti* conformation in the form IVb structure, whereas in the form IVa structure A7 may adopt either the *syn* or the *anti* conformation (see Fig. 2) (Coté *et al.*, 2000).

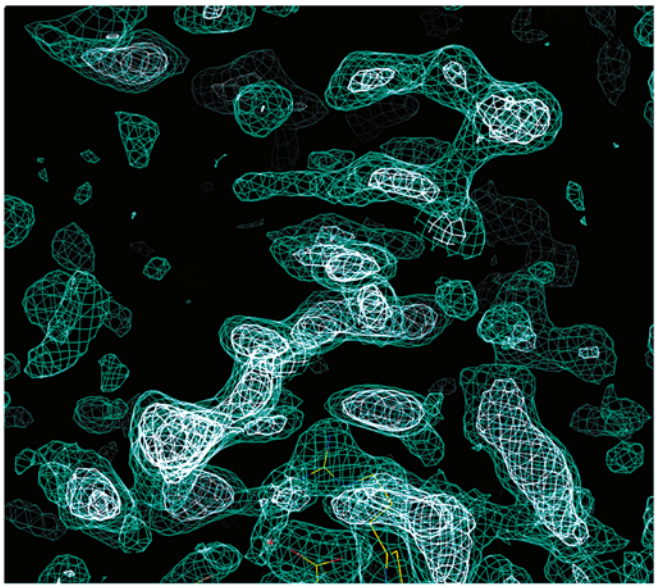
3.2. Protein–protein and protein–DNA interactions

The DNA is bound to the fingers domain of the N-terminal fragment as shown in Figs. 1, 4 and 5. In the form IVb MMLV RT fragment–DNA complex, an ion pair forms between Asp114 and Arg116, involving atoms $O^{\delta1}\cdots N^{\varepsilon}$ and $O^{\delta2}\cdots N^{\eta2}$, similar to that found in our other fragment–DNA crystal structures (Coté *et al.*, 2000; Najmudin *et al.*, 2000). This ion pair potentially serves to delocalize charge and to position the Arg116 into the minor groove of the oligonucleotide. The ion-pair hydrogen-bonding distances for $O^{\delta1}\cdots N^{\varepsilon}$ and $O^{\delta2}\cdots N^{\eta2}$ are 2.83 and 2.75 Å, respectively. Also, the ion-pair formed in form IVb is nearly planar, having $O^{\delta1}-H\cdots N^{\varepsilon}$ and $O^{\delta2}\cdots N^{\eta2}$ angles of 170.3 and 166.6°, respectively (idealized H atom from *CONTACT*; Collaborative Computational Project, Number 4, 1994), indicative of a strong interaction (see Fig. 4*b*).

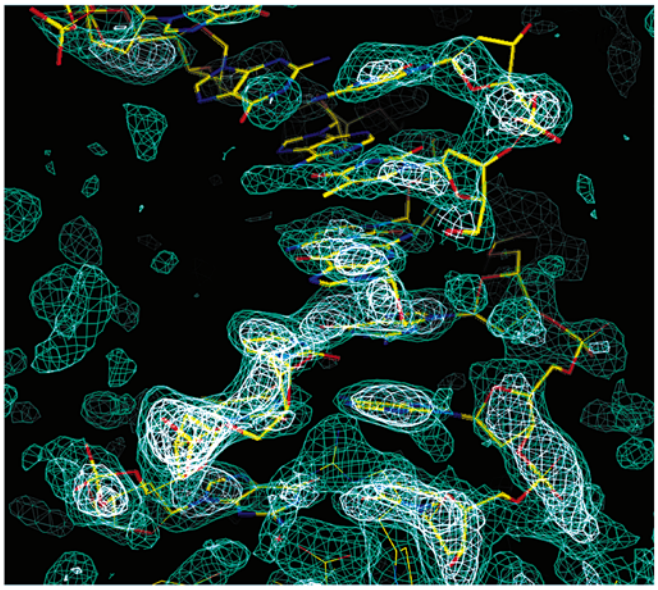
The protein–DNA interactions that occur in the fingers domain binding site of form IVb involve a significant number of contacts and hydrogen bonds with the sugar and base atoms of C16 (see Fig. 4*b*). Firstly, there are strong protein–DNA binding interactions with the 3′-OH of C16, encompassing hydrogen bonds of 3.0 Å to Leu115 N and 3.2 Å to Gly191 O and one other contact of 3.6 Å to Gln113 O. Also, the O2 atom of C16 forms a 3.3 Å contact to Arg116 $N^{\eta2}$ and a 3.6 Å contact to Arg116 N^{ε}. A unique aspect of the protein–DNA binding in the form IVb structure as opposed to all of our

other fragment–DNA structures is the hydrogen bonding of Asp114 and Arg116 to a base atom of the first 5′ nucleotide. In form IVb, the N2 atom of G1 forms hydrogen bonds to Asp114 $O^{\delta2}$ and Arg116 $N^{\eta2}$, with distances of 3.2 and 2.8 Å, respectively. The O2 atom of T2 also forms strong hydrogen bonds to $N^{\eta1}$ and $N^{\eta2}$ of Arg116, having respective distances of 2.9 and 2.6 Å. Arg116 $N^{\eta1}$ makes a 2.8 Å hydrogen bond to O4′ of C3 and a longer 3.4 Å contact to the O2 of C3. In the form IVb structure Tyr64 does not form any contacts to the nucleic acid, which is quite different from the case with form IVa. Table 2 lists the hydrogen-bonding interactions between the protein and DNA of the form IVb structure (see also Fig. 4*b*).

A C^{α} superpositioning of the protein of the form IVa model onto that of form IVb gives an r.m.s.d. of 0.82 for all residues and 0.44 for the best-matched 160 residues. Fig. 5 shows an

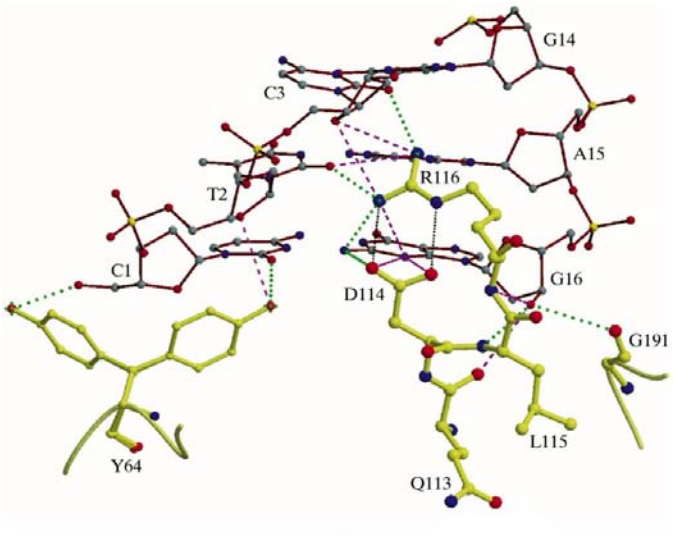


(*a*)

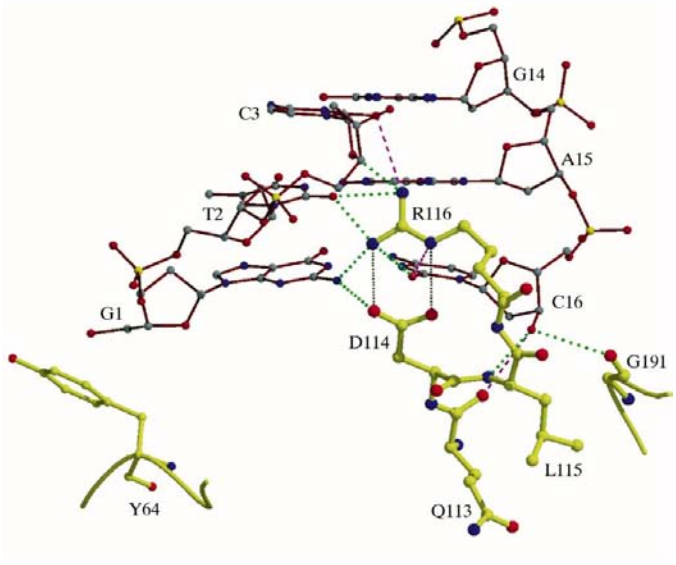


(*b*)

Figure 3
Electron-density maps (*a*) without and (*b*) with the fitted 6/10-mer of the form IVb structure. The fitted protein molecule is shown with thin stick models and the fitted DNA in (*b*) is shown with larger stick models for clarity. In both figures, the $2F_o - F_c$ map is shown in green and contoured at 1.0σ; the $F_o - F_c$ map is shown in white and contoured at 3.0σ. The figures are snapshots of the *O* session immediately following rigid-body refinement using solely the protein fragment for phasing (Jones *et al.*, 1991).



(*a*)



(*b*)

Figure 4
Comparative views (Kraulis, 1991; Merritt & Bacon, 1997) of the protein–DNA binding sites of the (*a*) form IVa and the (*b*) form IVb structures. In each view, the characteristic ion-pair between Asp114 and Arg116 is shown with black dotted lines. Green dotted lines denote hydrogen bonds whose distances range from 2.4 to 3.3 Å. Magenta dashed lines represent contacts whose distances are greater than 3.3 Å and less than 3.8 Å. Note the difference in the disposition of the Asp114–Arg116 ion pair in its interaction with the nucleic acid in the form IVa *versus* the IVb structure. Note the absence of contacts to the DNA from Tyr64 in the form IVb structure.

excerpted region of this superpositioning, emphasizing the protein–DNA interactions. Not surprisingly, there is excellent agreement between the superpositioned protein models; however, the disposition of the two form IV DNA models is remarkably different (see also Fig. 6). In addition, the protein–DNA interactions in the form IVb structure have interesting distinctions from those seen in all of our other fragment–DNA complexes (Coté *et al.*, 2000; Najmudin *et al.*, 2000).

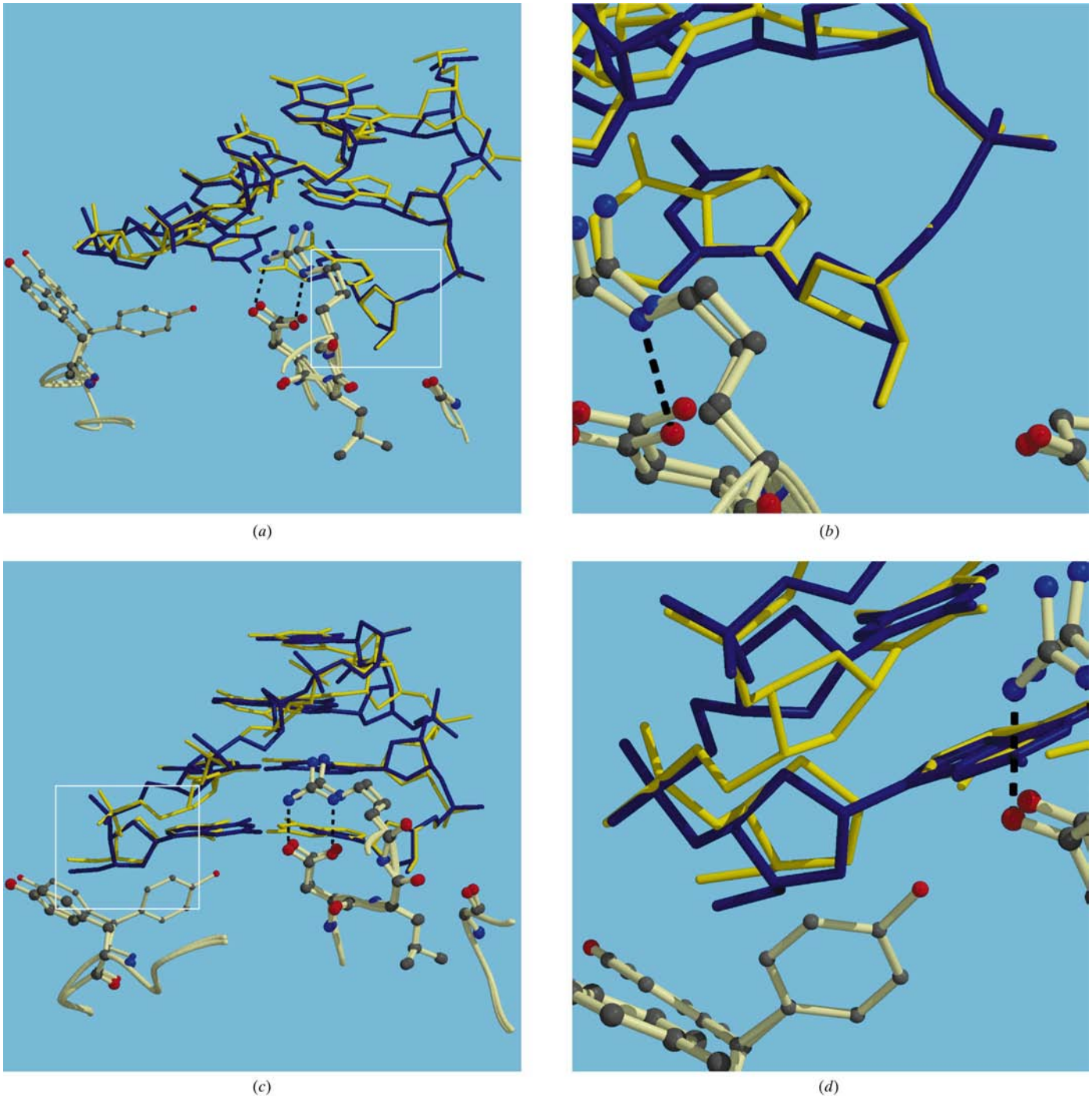


Figure 5
Views of the superpositioning (Jones *et al.*, 1991) of the DNA-binding sites of the form IVa structure onto the form IVb structure, based upon the C^{α} loci of the best matching 160 residues of each protein molecule. In all views, the ion pair of the form IVb structure is shown for increased perspective. In all views, gold stick models represent the DNA model of form IVa and the navy blue stick models represent the DNA of form IVb. (*a*) View of the binding site and the first three steps of the DNA of each structure, optimized to highlight the 3′-OH region. (*b*) Magnified view (rotated for increased clarity) of the 3′-OH regions of the blunt-ended duplexes, showing the near-exact mapping of the 3′-OH ribose rings of the two structures. (*c*) View emphasizing the 5′-OH region of the binding site. (*d*) Closeup (also rotated slightly) highlighting the large discrepancy between the two DNA structures, which is representative of the lack of agreement for the loci of the atoms of the two DNA molecules in the remainder of the structures. Note the shift of the ion-pair of form IVb toward its 5′ guanine (Kraulis, 1991; Merritt & Bacon, 1997).

3.3. DNA analysis

3.3.1. The pseudo-hexadecamer helix. The overall structure of the form IVb pseudo-hexadecamer is essentially that of B-form DNA. Fig. 2 shows both the form IV pseudo-hexadecamers with their numbering schemes. The symmetry in the pseudo-hexadecamers is such that each is bisected between the C8–G9* and G9–C8* base pairs (where * denotes the $-x-1, -y, z$ relationship). Each A–G mispair in the form IVa pseudo-hexadecamer is bordered $5'\rightarrow3'$ by a G–C and C–G pair. In the form IVb structure, the A–G mispair is flanked on both sides by a C–G pair. Both pseudo-hexadecamers have similarly situated C–G and G–C pairs sandwiched between their mispairs.

The form IVb pseudo-hexadecamer was analyzed (Lu & Olson, 1999) both as an intact DNA duplex and as a doubly nicked molecule (the observed condition) using the programs *CURVES* (Lavery & Sklenar, 1997) and 3*DNA* (Lu *et al.*, 2000). In constructing an intact pseudo-hexadecamer to assess the overall helical properties and groove characteristics, the constituent nucleosides were constrained to the positions and conformations they retain in the observed doubly nicked pseudo-hexadecamer. Thus, for the overall helical properties, idealized bridging phosphate groups were placed between nucleotides C6 and A7*, and A7 and C6* (symmetry relationship as above) using *O* (Jones *et al.*, 1991). No regularization of the idealized groups was invoked.

Analysis of the global features of the intact form IVb DNA using *CURVES* (Lavery & Sklenar, 1997) gives average major-groove and minor-groove widths of 11.1 and 6.7 Å, respectively, with respective ranges of 9.4–13.5 and 6.1–7.4 Å. The average minor-groove width value is far wider than the 5.8 Å value observed in canonical B-form DNA; all calculated minor-groove width values for the form IVb pseudo-hexadecamer exceed this value. The form IVa pseudo-hexadecamer, however, which has far more A-like character, has an even wider average minor-groove width of 7.0 Å. The average major-groove width value for the form IVb pseudo-hexadecamer is slightly smaller than the canonical B-form value of 11.6 Å; however, several values fall within the canonical range. The regions with a wider minor groove do correspond to those regions having a smaller major groove; however, the overall trend for the form IVb pseudo-hexadecamer is to B-form with merely a wider minor groove. A similar conclusion is reached when groove-analysis calculations are performed using the program 3*DNA* (Lu *et al.*, 2000) invoking the convention of El Hassan & Calladine (1998). Using this convention, the form IVb DNA is found to have average major- and minor-groove widths of 17.1 and 12.7 Å, respectively, with respective ranges of 15.2–19.8 and 12.5–13.1 Å. For B-form DNA, the convention of El Hassan and Calladine lists 17.4 (±1.3) Å as its standard value for the major-groove width and 10.8 (±1.4) Å as that for the minor-groove width. Using *CURVES*, the calculated average minor-groove depth is 5.1 Å, with values ranging from 3.8 to 7.8 Å. Those regions exhibiting the shallowest values also correspond to those with a wider minor groove. The average major-groove depth is 4.7 Å, with ranges from a very shallow 1.0 Å to a quite deep 7.0 Å. The average diameter is 19.2 Å, with the widest points of 20.3 Å occurring at C6–G11 and C6*–G11*, the steps preceding the strand breaks.

DNA structures containing G–C only or G–C-rich regions are prone to having wider minor grooves and other A-DNA-type characteristics. In a G–C-only octamer $d(GGGGCCCC)_2$, McCall and coworkers describe a model for poly(dG)–poly(dC) whose groove widths, inclination angles and other characteristics reveal it to be essentially A-form (McCall *et al.*, 1985). In the self-complementary decamer $d(CCGGCGCCGG)_2$, Heinemann and coworkers describe a significantly widened and deepened minor groove in an otherwise B-form DNA duplex. They attribute the changes in the groove characteristics to the sliding of the base pairs along their long axes (Heinemann *et al.*, 1992). The form IVb DNA shares another trait with the Heinemann decamer in that it maintains its geometrical integrity across a strand break. In yet another G–C-rich DNA octamer, $d(GGGATCCC)_2$, Lauble and co-workers describe a 75% G/C structure as an A-like DNA duplex with intermediate tilt, rise, and groove widths (Lauble *et al.*, 1988). In a recent DNA-only crystal structure determined at 1.3 Å of the decamer having 80% G/C bases and the sequence $d(AGGGGCCCCT)_2$, Gao and coworkers report two different crystal forms of predominantly A-form DNA. The $P2_12_12_1$ form is completely A-form, having an average step rise of 2.45 Å, an average inclination of 17.7°, an average pseudorotation angle of 15° and all of its deoxyribose puckers as C3′-*endo* (Gao, Robinson & Wang, 1999). In their $P6_122$ crystal form, although the DNA is A-form overall, several intermediate A/B characteristics occur, one being a larger average rise of 2.81 Å. Despite our form IVb pseudo-hexadecamer

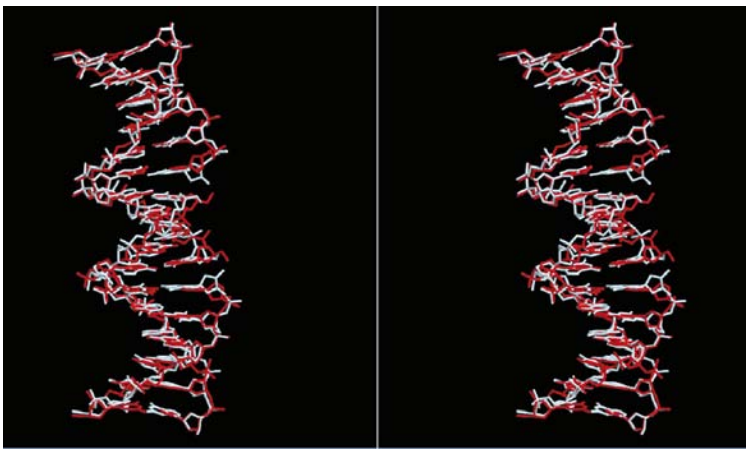


Figure 6
Stereoview (Kraulis, 1991; Merritt & Bacon, 1997) of the form IVa and form IVb pseudo-hexadecamers resulting from the superpositioning of the C^{α} atoms of the protein molecules of their structures. Note the near-exact match of the 3′-OH ribose rings and the lack of matches elsewhere. The form IVb DNA is shown in red and the form IVa DNA is shown in white, retaining its A7 base in the *anti* conformation.

having 68% G/C bases, it has less A-like character than we would have expected given the three consecutive guanines on a strand and the previous result of the form IVa structure.

The average intra-base C1′—C1′ distance in canonical B-form DNA is 10.5 (±0.2) Å. In G–A mismatches where the bases adopt the standard G(*anti*)–A(*anti*) conformation, the intra-base C1′–C1′ distances are typically much longer. These greater distances may lead to increased bulging, buckling and/or twisting in the structure (Chuprina & Poltev, 1983; Keepers *et al.*, 1984). In three solution NMR studies at near-neutral pH, DNA molecules containing the standard G(*anti*)–A(*anti*) conformation all report rather long C1′—C1′ intra-base distances in the G–A mispairs (Kan *et al.*, 1983; Patel *et al.*, 1984; Gao & Patel, 1988). The structure of the Kan decamer, $d(CCAAGATTGG)_2$, was later confirmed in a 1.3 Å resolution X-ray structure by Privé, who reports the C1′—C1′ distance in the G(*anti*)–A(*anti*) mispair as 12.5 Å (Privé *et al.*, 1987). In the Privé structure there is no significant bulging or excessive twisting, which is likely to be a consequence of the centrally located mispairs, which exhibit a significant propeller twist of 24.8° and are flanked by traditional Watson–Crick base pairs with gradually increasing/decreasing P—P distances. In the form IVb pseudo-hexadecamer the C1′—C1′ intra-base-pair distance in the G–A mispair is exceptionally long at 13.1 Å. This distance is far longer than the Privé distance and similarly longer than the 12.4 Å C1′—C1′ distance in the G(*anti*)–A(*anti*) model of our form IVa pseudo-hexadecamer (Coté *et al.*, 2000). In the form IVb pseudo-hexadecamer there is slight bulging in its center; however, gross bulging is alleviated since the mispairs reside at the strand breaks. Also, as in the case with the Privé structure, there is a gradual increase/decrease in the P—P distances of the approaching/departing steps in conjunction with the longer P—P distances in the C–G and G–C steps sandwiched between the mispairs. The C1′—C1′ intra-base-pair distances of C6–G11 and C8–G9* are quite long at 12.0 and 11.7 Å, respectively, reflecting the phenomenon seen in the P—P distances. The average intra-base-pair C1′—C1′ distance for the Watson–Crick pairs in the form IVb pseudo-hexadecamer is 11.1 Å, indicating a widened B-form structure. Several of the angles reflecting the pseudosymmetry about the glycosyl bonds in the form IVb pseudo-hexadecamer are more acute than the values in the accepted range of 52–62° (Rosenberg *et al.*, 1976; Seeman *et al.*, 1976). This occurs principally in the region of the stacked guanine nucleotides on each strand, which includes the G–A mispair. Table 3 shows the C1′—C1′ bond-distance data and the glycosidic bond parameters for the form IVb pseudo-hexadecamer.

A least-squares fitting of the bases of the form IVb pseudo-hexadecamer to standard bases shows virtually no deviation from standard forms. This is in contrast to the case with the form IVa pseudo-hexadecamer, whose bases on the 6-mer strand have significant A-like character. Based upon a global analysis using *CURVES* (Lavery & Sklenar, 1997), the form IVb pseudo-hexadecamer shows some minor bending, as evidenced by the global axis curvature value of 26.2°, which is slightly more than that observed in the form IVa DNA. The path length for the form IVb pseudo-hexadecamer is 50.0 Å and its end-to-end length is 48.2 Å; thus, there is an overall 3.1% shortening of the molecule.

Table 3
C1′—C1′ distances and glycosidic parameters for the form IVb pseudo-hexadecamer.

Y = pyrimidine, R = purine.

BP	λ(1)† (°)	λ(2)† (°)	C1′—C1′‡ (Å)	RN9—YN1§ (Å)	RC8—YC6¶ (Å)
G–C	50.3	53.6	10.7	8.9	9.7
T–A	53.7	57.2	10.4	8.8	9.7
C–G	58.3	52.0	10.8	9.1	10.1
G–C	57.0	48.4	11.0	9.2	10.1
T–A	56.2	48.2	10.9	9.1	9.9
C–G	47.2	58.5	12.0	10.2	11.1
A–G	39.4	37.4	13.1	10.8††	10.9††
C–G	50.1	52.7	11.7	9.9	10.7
G–C	52.7	50.1	11.7	9.9	10.7
G–A	37.4	39.4	13.1	10.8††	10.9††
G–C	58.5	47.2	12.0	10.2	11.1
A–T	48.2	56.2	10.9	9.1	9.9
C–G	48.4	57.0	11.0	9.2	10.1
G–C	52.0	58.3	10.8	9.1	10.1
A–T	57.2	53.7	10.4	8.8	9.7
C–G	53.6	50.3	10.7	8.9	9.7

† The angles between C1′—YN1 or C1′—RN9 glycosidic bonds and the bp C1′—C1′ line. ‡ Distance between C1′ atoms for each base pair. § Distance between RN9 and YN1 atoms for each base pair. ¶ Distance between RC8 and YC6 atoms for each base pair. †† Column headings do not strictly apply for these values owing to purine–purine mismatch.

With regard to the local base-pair step and helical parameters, the mean rise in the form IVb pseudo-hexadecamer is 3.3 Å; however, the greatest rise values occur on either side of the mispair. It is interesting to note that the rise values in the form IVb DNA closely parallel those of the form IVa G(*anti*)–A(*anti*) model, but the steps in the form IVa G(*anti*)–A(*syn*) model are far more evenly spaced, which is likely to be a consequence of the better base stacking and the closer C1′—C1′ distance resulting from the *syn* adenine (Coté *et al.*, 2000). The large shift at A7*C6*/G10G9 moves C6*–G10 into the major groove, with an intimate stacking of G10 over G9. An A-like inclination angle of 26° occurs in the center of the form IVb structure. This corresponds to the 15° roll in the same CG/CG step. This is likely to serve to further reduce any helical strain engendered by the G(*anti*)–A(*anti*) conformation of the mispairs. The mean helical twist is 35°, close to the standard B-form DNA value; however, there are wide variations among values in the mispair region. Table 4 shows the local base-pair step and helical parameters for the form IVb pseudo-hexadecamer.

Table 5 lists the intra-base-pair parameters for the form IVb pseudo-hexadecamer. Most notably, there is a distinct stretching evident in the A–G mispair and its flanking steps. Modest buckling is evident in the first two base pairs. This is observed in all of our fragment–DNA structures and is likely to be a consequence of interactions with the protein. The C–G step just preceding the mispair has an 11° buckle, which is likely to further accommodate the observed widening in the mispair region as well as alleviating helical tension. There is

Table 4
Local base-pair step and helical parameters.

Step	Shift (Å)	Slide (Å)	Rise (Å)	Inclin. (°)	Tip (°)	Twist (°)
GT/AC	0.2	−0.4	3.1	10	6	35
TC/GA	0.2	−0.1	2.9	−3	−13	38
CG/CG	−1.0	0.6	3.4	8	16	36
GT/AC	0.3	−1.2	3.2	1	−1	34
TC/GA	0.8	0.4	3.4	−9	−7	35
CA/GG	−0.6	−0.6	3.1	3	−13	29
AC/GG	2.1	0.7	3.6	−6	−18	39
CG/CG	0.0	0.5	3.5	26	0	36
GG/AC	2.1	0.7	3.6	−6	18	39
GG/CA	0.6	−0.6	3.1	3	13	29
GA/TC	−0.8	0.4	3.4	−9	7	35
AC/GT	0.3	−1.2	3.2	1	1	34
CG/CG	1.0	0.6	3.4	8	−16	36
GA/TC	−0.2	−0.1	2.9	−3	13	38
AC/GT	−0.2	−0.4	3.1	10	−6	35
Avg.	0.0	−0.1	3.3	2	0	35
Std. dev.	1.0	0.6	0.2	9	12	3

Table 5
Intra-base-pair parameters.

BP	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propeller (°)	Opening (°)
G–C	−0.4	−0.2	−0.1	−18	−7	−5
T–A	−0.4	−0.1	0.5	−16	−9	1
C–G	0.7	0.0	−0.4	−3	−4	1
G–C	0.7	0.0	0.6	2	−1	−3
T–A	1.0	0.0	0.5	11	−3	8
C–G	−0.9	1.0	0.6	5	−14	−3
A–G	0.3	1.4	0.1	9	−7	−32
C–G	−0.1	0.7	−0.2	5	−16	−5
G–C	0.1	0.7	−0.2	−5	−16	−5
G–A	−0.3	1.4	0.1	−9	−7	−32
G–C	0.9	1.0	0.6	−5	−14	−3
A–T	−1.0	0.0	0.5	−11	−3	−8
C–G	−0.7	0.0	0.6	−2	−1	−3
G–C	−0.7	0.0	−0.4	3	−4	1
A–T	0.4	−0.1	0.5	16	−9	1
C–G	0.4	−0.2	−0.1	18	−7	−5
Avg.	0.0	0.4	0.2	0	−8	−7
Std. dev.	0.6	0.6	0.4	10	5	10

increased opening in the mispair itself and significant propellering in the neighboring steps, particularly the steps sandwiched between the mispairs. The greatest shear occurs at C6–G11, which displaces C6 from the minor groove and is a result of the preferential stacking of G11, G10 and G9. The greatest stretching (1.4 Å) occurs in the mispair and the 5′-flanking C–G pair (1.0 Å). There is little staggering in the form IVb DNA, which comports well with the strong intra-base-pair hydrogen bonding observed in the first five base pairs. However, the C6–G11 base pair lacks hydrogen bonding and C6 is detached at its 3′ end, so the lack of commensurate stagger must be attributed to the interbase spacing dictated by the stacked guanines on the 10-mer strand. Also, the opening of the G–A mispair into the major groove is a consequence of both its *anti–anti* conformation and the intra-base hydrogen bonding that this entails as well as its being centrally located in the 3-guanine stack.

In the form IVb DNA, further evidence for overall B-form lies in the values for the sugar-conformation parameters. Unlike the form IVa DNA, whose 6-mer strand has most of its bases as C3′-*endo*, the form IVb structure has predominantly C2′-*endo* sugar conformations. The different protein–DNA interactions for the form IVa *versus* the form IVb structure account for the differences seen in the DNA structures. Table 6 shows the sugar-conformation parameters for the form IVb structure.

3.3.2. Analysis of the G–A mispairs. Several X-ray crystal structures of DNA only with otherwise unmodified bases containing G–A mispairs are known (Brown *et al.*, 1986, 1989; Privé *et al.*, 1987; Webster *et al.*, 1990; Shepard *et al.*, 1998; Gao, Sanishvili *et al.*, 1999). The form IVb DNA is identical to the form IVa DNA insofar as its G/C base percentage, but upon interchanging the 5′-flanking step to the A–G mispair from G–C to C–G (see Fig. 2), the G(*anti*)–A(*anti*) mispair conformation results. Although the mispair resides at a strand break, which would seemingly afford greater opportunities for *anti–anti* to *anti–syn* conformation transitions (as seen in the form IVa DNA), that does not occur here and the conclusion must be drawn that flanking sequence is the driving force for the single conformation. Thus, the form IVb DNA structure determined at pH 6.5 is better compared with the solution NMR structure of Gao and Patel determined at pH 6.8, which has the sequence d(CGGGAATTCACG)$_2$, where the adenine of the G–A mispair is also flanked by two cytosines (Gao & Patel, 1988). In their structure, the mispair retains the standard G(*anti*)–A(*anti*) conformation, as is the case with the form IVb pseudo-hexadecamer. An isolated G–A mispair having two cytosines flanking the adenosine of the mispair necessarily results in the guanosine of the mispair being flanked by two other guanosines. Such is the case for the form IVb DNA, with the added commonality of the ...CGGGA... run of nucleosides.

When unmodified G–A mispairs adopt the standard G(*anti*)–A(*anti*) conformation, the intra-base-pair hydrogen bonds occur between Gua N1···N1 Ade and Gua O6···N6 Ade. In our G(*anti*)–A(*anti*) model, these distances are 2.9 and 2.4 Å, respectively. This is similar to the G(*anti*)–A(*anti*) form IVa model, whose same hydrogen-bond distances are 2.5 and 2.7 Å, respectively. In our form IVb DNA model, true base-pair hydrogen bonding of the flanking C–G pairs does not occur. In the flanking C6–G11 pair, the closest contact between potentially hydrogen-bonded base atoms is 4.0 Å and occurs between Cyt N3 and Gua N2, which is not a conventional C–G hydrogen-bonding atom pair. A pulling away of these bases has occurred, which is likely to be a consequence of the residence of C6 at a strand break in conjunction with G11 being the first nucleotide in the three-guanine stretch and bordering the mismatch. In the form IVa pseudo-hexadecamers [either the G(*anti*)–A(*anti*) or the G(*anti*)–A(*syn*) model], this base pair is G6–C11 (see Fig. 2*a*). In that base pair, virtually normal hydrogen bonding occurs with a slight departure involving a longer distance of 3.3 Å from Gua O6···O4 Cyt, reflecting a shift into the major groove similar to that seen in the form IVb structure. There are far more favorable stacking interactions for G6 with A7* across

Table 6
Sugar-conformation parameters.

Owing to the symmetry of the hexadecamer, only one 16-member strand is shown.

Base	$\nu0$† (°)	$\nu1$‡ (°)	$\nu2$§ (°)	$\nu3$¶ (°)	$\nu4$†† (°)	Tm‡‡ (°)	P§§ (°)	Pucker
G	−15	28	−31	23	−5	31	171	C2′-*endo*
T	−6	26	−35	32	−16	35	189	C3′-*exo*
C	−28	36	−32	17	6	36	152	C2′-*endo*
G	−18	26	−26	17	1	27	160	C2′-*endo*
T	−22	36	−36	26	−3	38	166	C2′-*endo*
C	−20	31	−35	23	−2	36	166	C2′-*endo*
A	−20	29	−28	18	1	30	159	C2′-*endo*
C	−24	35	−33	20	2	35	157	C2′-*endo*
G	−25	37	−34	20	3	37	157	C2′-*endo*
G	−21	34	−34	23	−1	36	163	C2′-*endo*
G	−32	43	−37	20	7	42	151	C2′-*endo*
A	−24	40	−40	27	−2	41	165	C2′-*endo*
C	−4	23	−32	31	−17	33	192	C3′-*exo*
G	−20	34	−35	24	−2	36	165	C2′-*endo*
A	−26	37	−34	20	4	38	156	C2′-*endo*
C	−27	36	−31	16	7	35	150	C2′-*endo*

† $\nu0$, C4′—O4′—C1′—C2. ‡ $\nu1$, O4′—C1′—C2′—C3′. § $\nu2$, C1′—C2′—C3′—C4. ¶ $\nu3$, C2′—C3′—C4′—O4. †† $\nu4$, C3′—C4′—O4′—C1′. ‡‡ Tm, the amplitude of pseudorotation of the sugar ring. §§ P, the phase angle of pseudorotation of the sugar ring.

the strand break in the form IVa DNA than there are for C6 and A7* of form IVb. Thus, it is clear that the structural driving force in this region of the form IVb DNA is the stacking of the guanines. Although not egregiously stretched like the C6–G11 pair, the C8*–G9 pair flanking 3′ to the A–G mispair is also stretched. In the C8*–G9 pair, the intra-base-pair hydrogen-bonding values between O2···N2, N3···N1 and N4···O6 are 3.8, 3.7 and 3.5 Å, respectively. These values far exceed not only canonical values, but also those seen in other B-DNA structures with G–A mispairs (Brown *et al.*, 1986, 1989; Webster *et al.*, 1990; Coté *et al.*, 2000). The AC/GG step (containing the mispair) has a 12° tilt, which appears to be coupled to the 2.1 Å shift, in accordance with the predictions made by Packer *et al.* (2000). This is also reflected in the increased rise seen in the steps bordering the mispair, where the two C nucleotides have little opportunity to stack with the mispaired A7.

3.3.3. Protein–DNA interactions. Even though the interactions in the binding site of our fragment–DNA structures are similar to each other in a general sense, the differences between the form IVb structure and all of our other fragment–DNA structures are quite interesting. The results remain consistent with the proposed mechanistic role of interaction with the fingers domain binding site, which the enzyme uses during processive synthesis. The proposed mechanism involves binding of the nucleic acid substrate alternatively to the polymerase active site and fingers domain binding site in a ratcheting type of mechanism. Potentially, the enzyme keeps track of the 3′-OH of the primer strand in this way. As the 3′-OH is the chemically reactive site for continued polymerization of DNA, the specificity for interactions in the fingers domain binding site provides a direct mechanism for understanding translocation of the substrate during processive synthesis.

In interchanging the C and G nucleotides at the initial DNA duplex step, we have observed specific differences in the protein–DNA interactions with the MMLV RT fragment molecule. Prior to this transposition, a cytosine resided at the 5′ position in all of our fragment–DNA complexes. This C1 nucleoside forms no contacts from its base atoms to the protein in our form I and form II structures. The form IVa protein–DNA interactions are essentially similar to those observed in all of our other fragment–DNA complexes (Coté *et al.*, 2000; Najmudin *et al.*, 2000) and will be used as a general comparison model with the form IVb structure. Of all of our fragment–DNA structures, only one of the half-occupancy Tyr64 molecules in the form IVa structure makes a contact to a base atom of the 5′ nucleoside, which is a 3.0 Å hydrogen bond Tyr64 OH···O2 C1 (see Fig. 4*a*). The other half-occupancy Tyr64 OH atom forms a 2.7 Å hydrogen bond to O5′ of C1. In the form I and form II structures, Tyr64 forms no contacts to C1 and only interacts with the O4′ atom of the second nucleoside of the template strand (A2 in the cases of forms I and II; T2 in the case of form IVa). The highly conserved Asp114 and Arg116 do not interact at all with C1 in any of our other structures. What is unique about the protein–DNA interactions of the form IVb structure is that both Asp114 and Arg116 form hydrogen bonds to the N2 atom of G1 (see Fig. 4*b*).

In viewing Fig. 4, note the 'triangle' of hydrogen bonds formed among Asp114 $O^{\delta2}$, Arg116 $N^{\eta2}$ and Gua N2. In the form IVa structure, this triangle has its minor-groove vertex with N2 of G16 on the 3′→5′ strand (see Fig. 4*a*). In the form IVb structure, however, the transposition of G and C in the first base pair of the DNA duplex has caused a shift of this triangle such that its minor-groove vertex with N2 of G1 is on the 5′→3′ strand (see Fig. 4*b*). The greater buckling of the G1–C16 pair of form IVb is evident and occurs owing to the fixed distance dictated by the strong ion pair and its forming contacts to the base atoms of the 5′ purine as well as maintaining the contacts with the 3′ nucleotide.

One reason that less A-like character is observed in the form IVb structure as opposed to the form IVa structure is because of the transposition of G and C in the first base pair. In the form IVa structure, the ion pair forms greater hydrogen-bonding interactions with the 3′ G16 and the pyrimidine–pyrimidine stack on the opposite 5′ strand has greater helical tension exerted upon it. Also, one of the Tyr64 rotamers of form IVa forms a contact to the O4′ of T2, locking it into an A-like C3′-*endo* conformation. In the form IVb structure, the fixed ion pair forms the interaction with the N2 of the 5′ G1, and Tyr64 forms no contacts to the DNA at all. Thus, less helical tension is imparted to the form IVb duplex and its predisposed B-form is maintained in the first two steps. Also, the CT/AG step of the form IVa DNA has less stacking advantage, predisposing it to greater flexibility in its interactions with protein.

One consequence of the comparison of the form IVa and form IVb structures was the discovery that upon super-

imposing the best defined C^{α} atoms of each protein model (r.m.s.d. = 0.4), the DNA models of the two structures had virtually no atomic sites in common, except for those of the ribose ring on the primer strand (see Fig. 5). The peptide O from the strictly conserved Gly191 residue forms a strong hydrogen bond to the 3′-OH of the nucleoside (as does the main-chain N of Leu115). The transposing of G and C has not altered these critical interactions at all and the DNA of each structure has adjusted to this common set point. Fig. 6 shows the overall comparison for the two DNA models of form IVb and IVa when only their protein molecules are superimposed. The very specific binding of the 3′-OH is consistent with a mechanistic role in polymerization by MMLV RT.

3.3.4. Dinucleotide steps and stacked guanines. The stark differences between the DNA pseudo-hexadecamers of form IVa and form IVb are all a consequence of the transposition of the C and G bases at the beginning and end of the 6-mer duplex. Fig. 7 profiles three of the dinucleotide steps encompassing the base pairs of the transposed nucleotides. In the first base pair, the changing of the nucleotides has altered the protein–DNA interactions and this augments the structure that the duplex would likely adopt if it were in a DNA-only structure. For example, the first dinucleotide step of the form IVa DNA is CT/AG, which is known to have great flexibility (Packer *et al.*, 2000), and in conjunction with the helical strain from the interactions with the protein, A-like characteristics ensue. In the comparable form IVb step, the more favorable GT/AC alignments are assisted by the protein forming contacts to maintain its B-form. In step 5, there is rather normal hydrogen bonding between G6 and C11 in the form IVa DNA; however, there is none between C6 and G11 in form IVb. The reason is that there is far greater flexibility in a T–C stack on a strand, especially since C6 is not attached at its 3′ end. Also, the G10 is the first of a three-guanine stack. In the mixed pyrimidine–purine TG/CA step of form IVa, G6 resides at the 3′ strand break and is far less likely to slide when it is flanked by T and A nucleotides. In the dinucleotide step containing the G–A mispair [the G(*anti*)–A(*anti*) form IVa model is shown], the stacking of the guanines on the 10-mer strand of the form IVb DNA is quite evident. The six-membered ring of G10 is aligned over the center of G11, maximizing the interaction between the two guanines of that step. Although not evident in these two-dimensional renditions, the uneven shearing and stretching of the base pairs of the purine-rich region including the mispairs is reflected by the lack of hydrogen bonding in the base pairs flanking the G–A mismatch. Also, since the view is down the helical axis, the difference in tilt (12° in the form IVb structure) between the CA/GG step of form IVb and the reasonably flat GA/GC step of form IVa is evident.

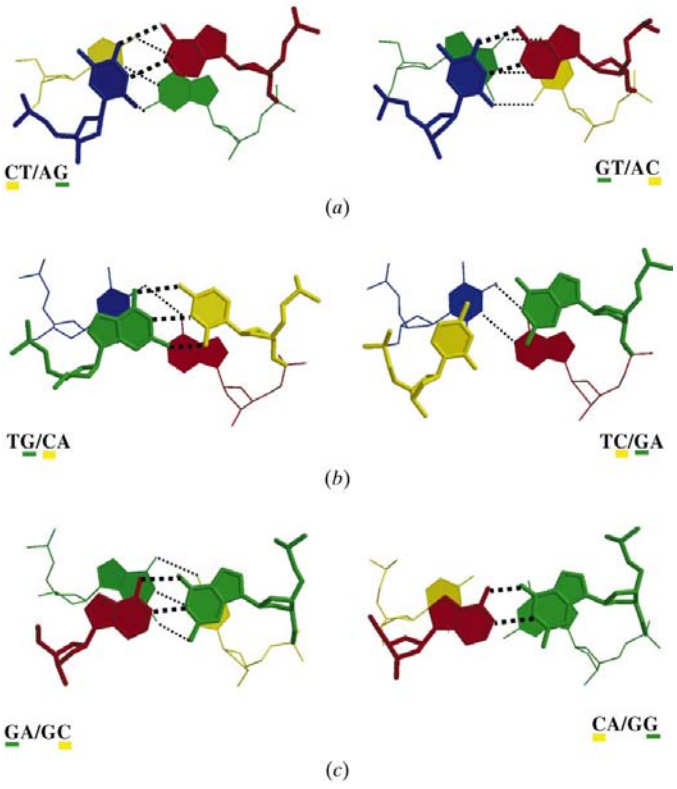


Figure 7
Comparison of selected dinucleotide steps in the DNA structures of forms IVa (left) and IVb (right). In all views (down the DNA-helical axis of each), the nucleotides which differ between the two pseudo-hexadecamers are underlined. Yellow, C; blue, T; red, A; green, G in all views. (*a*) Step 1, showing the better stacking of the form IVb structure, with its mixed-sequence purine–pyrimidine dinucleotide step. (*b*) View of step 5, just preceding the G–A mispair, *i.e.* just prior to the strand break. Note the normal hydrogen bonding of the G–C pair in the form IVa DNA, as opposed to that of the C–G pair in the form IVb DNA, where the G has completely pulled away from its partner. (*c*) View of step 6, emphasising the difference in the stacking of the guanines across the strand break seen in the form IVb structure. The mispaired adenine of the form IVa structure is shown in its *anti* conformation for best comparison. This figure was generated using the mstack2img utility of the program 3*DNA* (Lu *et al.*, 2000).

In conclusion, the DNA structure of the form IVb structure reported here differs in important ways from the form IVa structure. Through interchanging the bases of the first base pair in the DNA molecule, important questions were addressed regarding the binding of the nucleic acid duplex to the MMLV RT N-terminal fragment. Also, the modification interchanging the bases of the step preceding the G–A mispair resulted in a single G(*anti*)–A(*anti*) conformation for the mispair, emphasizing the differences flanking sequence imparts to the G–A mispair conformation. The modification further underscored the structural intricacies of three consecutive guanine residues and their behavior when in an environment that does not have them physically constrained in the DNA molecule.

We thank Xiang-Jun Lu for assistance with the analysis of the DNA, Wilma Olson for helpful discussions, Dick Leidich and Tom Weaver for technical support of the X-ray equipment, and Attilio Defalco and Greg Listner for general computing assistance. This work was supported by a grant GM55026 (MMG) from the National Institutes of Health.

References

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988). *Science*, **242**, 899–907.

Brown, T., Hunter, W. N., Kneale, G. & Kennard, O. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 2402–2406.

Brown, T., Leonard, G. A., Booth, E. D. & Chambers, J. (1989). *J. Mol. Biol.* **207**, 455–457.

Brunger, A. T., Adams, P. A., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Chuprina, V. P. & Poltev, V. I. (1983). *Nucleic Acids Res.* **11**, 5205–5222.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Coté, M. L., Yohannan, S. J. & Georgiadis, M. M. (2000). *Acta Cryst.* D**56**, 1120–1131.

El Hassan, M. A. & Calladine, C. R. (1998). *J. Mol. Biol.* **282**, 331–343.

Gao, X. & Patel, D. (1988). *J. Am. Chem. Soc.* **110**, 5178–5182.

Gao, Y.-G., Robinson, H., Sanishvili, R., Joachimiak, A. & Wang, A. H.-J. (1999). *Biochemistry*, **38**, 16452–16460.

Gao, Y.-G., Robinson, H. & Wang, A. H.-J. (1999). *Eur. J. Biochem.* **261**, 413–420.

Gautheret, D., Konings, D. & Guttel, R. R. (1994). *J. Mol. Biol.* **242**, 1–8.

Georgiadis, M. M., Jessen, S. M., Ogata, C. M., Telesnitsky, A., Goff, S. P. & Hendrickson, W. A. (1995). *Structure*, **3**, 879–892.

Heinemann, U., Alings, C. & Bansal, M. (1992). *EMBO J.* **11**, 1931–1939.

INSIGHT II (1993). *INSIGHT II User Guide.* San Diego, CA, USA: Biosym Technologies.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Jordan, S. R. & Pabo, C. O. (1988). *Science*, **242**, 893–899.

Kan, L.-S., Chandrasegaran, S., Pulford, S. M. & Miller, P. S. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 4263–4265.

Keepers, J. W., Schmidt, P., James, T. L. & Kolman, P. A. (1984). *Biopolymers*, **23**, 2901–2929.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–290.

Lauble, H., Frank, R., Bloker, H. & Heinemann, U. (1988). *Nucleic Acids Res.* **16**, 7799–7815.

Lavery, R. & Sklenar, H. (1997). *Curves* 5.2. *Helical Analysis of Irregular Nucleic Acids*.

Lu, X.-J. & Olson, W. K. (1999). *J. Mol. Biol.* **285**, 1563–1575.

Lu, X.-J., Shakked, Z. & Olson, W. (2000). *J. Mol. Biol.* **300**, 819–840.

McCall, M., Brown, T. & Kennard, O. (1985). *J. Mol. Biol.* **183**, 385–396.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Najmudin, S., Coté, M. L., Sun, D., Yohannan, S., Montano, S. P., Gu, J. & Georgiadis, M. M. (2000). *J. Mol. Biol.* **296**, 613–632.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Packer, M. J., Dauncey, M. P. & Hunter, C. (2000). *J. Mol. Biol.* **295**, 71–83.

Patel, D. J., Kozlowski, S. A., Ikuta, S. & Itakura, K. (1984). *Biochemistry*, **23**, 3207–3217.

Poltev, V. I. & Shulyupina, N. V. (1986). *J. Biolmol. Struct. Dyn.* **3**, 739–765.

Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L. & Dickerson, R. E. (1987). *Science*, **238**, 498–504.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Rodgers, D. W. & Harrison, S. C. (1993). *Structure*, **1**, 227–240.

Rosenberg, J. M., Seeman, N. C., Day, R. O. & Rich, A. (1976). *J. Mol. Biol.* **104**, 145–167.

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). *Proc. Natl Acad. Sci. USA*, **73**, 804–808.

Shepard, W., Cruse, W. B., Fourme, R., de La Fortelle, E. & Prange, T. (1998). *Structure*, **6**, 849–861.

Shimon, L. J. W. & Harrison, S. C. (1993). *J. Mol. Biol.* **242**, 826–838.

Sun, D., Jessen, S., Liu, C., Liu, X., Najmudin, S. & Georgiadis, M. M. (1998). *Protein Sci.* **7**, 1575–1582.

Webster, G. D., Sanderson, M. R., Skelly, J. V., Neidle, S., Swann, P. F., Li, B. F. & Tickle, I. J. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 6693–6697.