Protein Science (1998), 7:1575–1582. Cambridge University Press. Printed in the USA.

Cloning, expression, and purification of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase: Crystallization of nucleic acid complexes*

DUNMING SUN, SVEN JESSEN, CHUNHUI LIU, XIUPING LIU, SHABIR NAJMUDIN, AND MILLIE M. GEORGIADIS

Waksman Institute and Department of Chemistry, Rutgers University, Piscataway, New Jersey 08854

(RECEIVED December 30, 1997; ACCEPTED March 30, 1998)

Abstract

Reverse transcriptase is an essential retroviral enzyme that uses RNA- and DNA-directed DNA polymerase activities as well as an RNaseH activity to synthesize a double-stranded DNA copy of the single-stranded RNA genome. In an effort to obtain high-resolution structural information regarding the polymerase active site of reverse transcriptase, we have pursued studies on a catalytic fragment from Moloney murine leukemia virus reverse transcriptase. DNA encoding the catalytic fragment, defined originally by limited proteolytic digestion, has been cloned, and the protein has been expressed and purified from *Escherichia coli*. The fragment obtained by limited proteolytic digestion and the bacterially expressed fragment retain polymerase activity. Crystallization studies involving nucleic acid complexes with a catalytic fragment from both sources are reported, including variables screened to improve crystals and cryocooling. Three crystal forms of catalytic fragment–nucleic acid complexes have been characterized, which all contain at least two protein molecules in the asymmetric unit. As isolated, the catalytic fragment is monomeric. This analysis indicates that the enzyme dimerizes in the presence of nucleic acid.

Keywords: complex; crystallization; DNA; Moloney murine leukemia virus; nucleic acid; polymerase; reverse transcriptase

Reverse transcriptase (RT) is an essential retroviral enzyme that catalyzes formation of a double-stranded DNA copy from the single-stranded retroviral RNA genome (Gilboa et al., 1979). Enzymatic properties of RT include RNA- and DNA-directed polymerase activities in addition to an RNaseH activity specific for DNA/RNA duplex that allows RT to complete the synthesis of the double-stranded DNA copy (Goff, 1990). The newly synthesized DNA is then integrated into the host genome by another retroviral enzyme, integrase. Retroviral RTs can be classified according to host specificity into one of three families: human, avian, or murine. RTs from these three families share the same functional properties but have different architectures (Goff, 1990). RTs from the human retroviral family are heterodimeric and include a 66,000 Da subunit and a 51,000 Da subunit referred to as p66 and p51 subunits. The p51 subunit results from removal of the C-terminal RNaseH domain from a p66 subunit by the retroviral protease (LeGrice, 1993). The avian RTs include, in addition to the polymerase and RNaseH activities, an integrase activity resulting from an integrase domain found on the larger 95,000 Da beta subunit and proteolytically removed from the smaller 63,000 Da alpha subunit (Leis et al., 1983). The bacterially expressed and retroviral murine RTs are 75,000 Da monomeric enzymes as isolated (Moelling, 1974; Roth et al., 1985) but have been proposed to be homodimeric in the presence of nucleic acid substrates based on gel retardation assay results (Telesnitsky & Goff, 1993; Guo et al., 1995).

Numerous crystal structures have been reported for HIV-1 RT uncomplexed and complexed with a number of non-nucleoside inhibitors (Kohlstaedt et al., 1992; Jacobo-Molina et al., 1993; Jager et al., 1994; Unge et al., 1994; Ding et al., 1995a, 1995b; Esnouf et al., 1995; Ren et al., 1995; Rodgers et al., 1995; Hsiou, 1996). At present, one RT-nucleic acid complex has been reported at 3.0 Å resolution for HIV-1 RT (Jacobo-Molina et al., 1993), which includes HIV-1 RT, Fab, and an 18/19-mer oligonucleotide. We present here our studies on nucleic acid complexes with a 30,000 Da catalytic fragment of Moloney murine leukemia virus (MMLV) RT. The catalytic fragment of MMLV RT was defined by limited proteolytic digestion, and the crystal structure of the uncomplexed enzyme has been reported at 1.8 Å resolution (Georgiadis et al., 1995). This fragment includes the fingers and palm domains of RT and retains polymerase activity as assayed in standard RT polymerization assays, albeit reduced approximately 5,000-fold from that of the full-length enzyme. Nomenclature for the

Reprint requests to: Millie M. Georgiadis, Waksman Institute and Department of Chemistry, 190 Frelinghuysen Road, Piscataway, New Jersey 08854; e-mail: georgiadis@mbcl.rutgers.edu.

*Dedicated to the memory of Dunming Sun.

domains present in RT are based on the convention reported for Klenow fragment in which the domains responsible for polymerase activity were likened to a right hand including fingers, palm, and thumb domains (Ollis et al., 1985). RT additionally contains connection and RNaseH structural domains (Kohlstaedt et al., 1992).

We are pursuing structural studies on MMLV RT–nucleic acid complexes with the goal of obtaining more detailed information on the interactions of nucleic acid with RT. In our efforts to obtain higher resolution structures, we are studying nucleic acid complexes with the catalytic fragment, which includes the polymerase active site. Here we report the crystallization, optimization, and characterization of three crystal forms of the catalytic fragment complexed with nucleic acid. Variables that were screened include N-terminal truncation of the catalytic fragment, the length of the duplex portion of the template/primer oligonucleotide substrates, the length of the single-stranded portion of the template, and molar ratios of enzyme to nucleic acid substrate in the complex. In addition, the cloning, expression, and purification of a catalytic fragment from MMLV RT are reported. Implications for the stoichiometry of the nucleic acid complexes with the catalytic fragment are discussed.

8/18-mer

5'- CATGCATG- 3'

3'- GTACGTACGGCGTACTAG-5'

8/16-mer

5'-CATGCATG-3'

3'-GTACGTACGGCGTACT-5'

8/8-mer

5'-CATGCATG-3'

3'-GTACGTAC-5'

16/16-mer

5'- TCATGCGGCATGCATG-3'

3'-GTACGTACGGCGTACT-5'

6/14-mer

5'-CTCGTG-3'

3'-GAGCACGGCATACT-5'

6/10-mer

5'-CTCGTG-3'

3'-GAGCACGGCA-3'

Fig. 1. Oligonucleotide sequences complexed with catalytic fragment are shown. The sequences are identified by primer/template lengths. The 8/14-mer sequence mentioned in Results is derived from the 8/16-mer sequence and lacks the 5′ CT found in the 16-mer sequence.

Results

Cloning, expression, and purification of a catalytic fragment

A catalytic fragment of MMLV RT (Georgiadis et al., 1995) was defined by limited proteolysis of MMLV Δ RH, which is MMLV RT lacking the C-terminal RNaseH domain (Telesnitsky et al., 1992). This fragment is a mixture and includes residues 10–274 or 10–278 of the 671 present in the full-length enzyme. MMLV Δ RH includes residues 1–497 and retains polymerase activity but lacks RNaseH activity. We have made a bacterially expressed fragment comprised of residues 24–278 from MMLV RT using the expression vector, pET15b (Novagen), which includes an N-terminal hexa-histidine tag and thrombin cleavage site allowing removal of all but four amino acids from the N-terminus encoded by the vector. The protein was overexpressed in *Escherichia coli* and purified using Ni-NTA superflow (Qiagen) chromatography and Mono-S ion-exchange (Pharmacia) chromatography. Following purification and thrombin cleavage of the affinity-tagged enzyme, the catalytic fragment was subjected to a final chromatographic step. Digestion with thrombin resulted in complete removal of the N-terminal tag as judged by its altered size on SDS gel electrophoresis and a significantly altered elution profile during ion-exchange chromatography. Final purity was assessed by analytical chromatographic separation on a Superdex 75 column and crystallization experiments. The enzyme was found to be at least 99% pure. In standard reverse transcriptase polymerization assays, the fragment was found to retain polymerase activity (M.M. Georgiadis, unpubl. obs.). Following purification, approximately 10 mg of catalytic fragment per liter of culture were obtained.

Crystallization and characterization of nucleic acid complexes with the catalytic fragment

Three crystal forms (I, II, and III) have been obtained using catalytic fragment produced by limited proteolysis of purified MMLV ΔRH or bacterially expressed catalytic fragment. Nucleic acid substrates that have been used include those shown in Figure 1. The following nomenclature will be used to describe the two catalytic fragment molecules. The catalytic fragment produced by limited proteolysis is referred to as CF10 (catalytic fragment with N-terminus starting at residue 10), and the bacterially expressed catalytic fragment is referred to as CF24 (catalytic fragment with N-terminus starting at residue 24). The oligonucleotide substrates include a template with single-stranded overhang and complementary primer forming the duplex portion of the DNA substrate (see Discussion). The complexes used in the crystallization experiments included catalytic fragment, oligonucleotide, and ddCTP, 2′,3′-dideoxycytidine 5′-triphosphate.

Form I crystals

Initial hanging drop-vapor diffusion screens were done using an incomplete factorial approach (Carter & Carter, 1979; Jancarik & Kim, 1991) with CF10 and the 8/18-mer oligonucleotide. Small crystals were obtained for several different polyethylene glycol (PEG) and salt conditions. Subsequent screens included varying the concentration of PEG, the pH, the salt concentration, the type of salt compound, the ratio of protein:DNA:ddCTP, and the length of the single-stranded template overhang used in the crystallization experiment. The best crystals for this complex were obtained using 100 mM HEPES pH 7.5, 25% PEG 4000, 100 mM NH_4Cl, and a CF10:DNA:ddCTP ratio of 1:4:8. The requirement for excess DNA in the formation of complex crystals has been previously reported (Aggarwal, 1990). Template oligonucleotides including 18, 16, and 14 bases were screened for the best crystallization condition. The template 16-mer was found to produce the largest crystals under the conditions screened. The presence of nucleic acid in all crystal forms reported here was verified by native gel electrophoresis of crystalline samples carefully washed and then dissolved. Native gels were silver stained (Blum et al., 1987), and crystalline samples were found to contain both protein and nucleic acid. The presence of ddCTP in the crystals has not been verified.

Microseeding and macroseeding techniques were used to produce crystals suitable for X-ray diffraction work. Microseeding conditions were similar to those described above with the precipitant solution containing 100 mM HEPES (pH 7.5), 15% PEG 4000, and 75 mM NH_4Cl. For macroseeding, the precipitant solution contained 100 mM HEPES (pH 7.5), 13% PEG 4000, and

75 mM NH_4Cl. The largest crystals obtained were 100 × 150 × 500 μm^3. The space group in which Form I crystals grow is $P2_1$ with cell dimensions as shown in Table 1. Using a synchrotron source, the diffraction limit of the crystals was found to be 1.9 Å. Usable data were collected to 2.3 Å resolution at the NSLS BNL, beamline X4A, for which statistics are shown in Table 2. Based on a self-rotation function analysis, there is a non-crystallographic 2-fold symmetry axis present. There are, therefore, two protein molecules in the asymmetric unit (see Materials and methods).

Crystals were also obtained for an 8/8-mer (primer/template) DNA complex with CF10 under very similar crystallization conditions. The space group and cell constants for crystals of the CF10 and 8/8-mer were identical to the Form I crystals with 8/16-mer. An experiment was done in which 8/8-mer, 8/16-mer, and 16/16-mer DNA molecules were tested for ability to form crystals in a range of microseeded crystallization conditions near the optimal condition identified for the 8/16-mer. Examples of the best crystals obtained for each substrate are shown in Figure 2. All three DNA substrates were tested as hanging drops suspended over the same reservoir. The 8/8-mer and 8/16-mer DNA substrates were found to grow substantially larger crystals than the 16/16-mer DNA complexes. Crystals obtained with the 16/16-mer required at least two weeks to grow, indicating that they were, in fact, self-nucleated.

Subsequently, two iodine sites per asymmetric unit were identified in an isomorphous difference Patterson analysis for a 5-iodouracil substituted DNA derivative of the Form I complex crystals obtained with 8/16-mer. The predicted distance between the two sites for an 8/16-mer based on ideal A-form DNA is 9.2 Å and for ideal B-form DNA is 10.8 Å (see Fig. 3 and Materials and methods). For an 8/8-mer, the predicted distances are 15.7 Å for A-form and 16.6 Å for B-form. Predicted distances for a 16/16-mer, 10.35 Å for A-form and 11.36 Å for B-form DNA, are similar to those for the 8/16-mer. The shortest observed distance from the Patterson analysis is 20.6 Å (between two sites in the asymmetric unit chosen) or 19.7 Å (between two symmetry related sites) and is consistent with the presence of one 8/8-mer per asymmetric unit.

Form II crystals

Form II crystals were obtained under nearly identical crystallization conditions to those used to obtain Form I crystals. These crystals appeared to be a different crystal form based on morphology as shown in Figure 2. For these crystals, CF24 was used instead of CF10 and the 8/8-mer DNA substrate was used in the complex. Microseeding techniques were used to obtain crystals suitable for X-ray diffraction work with dimensions of approximately 200 × 200 × 300 μm^3. Some self-nucleated crystals were obtained that were sufficiently large for X-ray diffraction work, unlike those obtained from the complexes with CF10. Crystals obtained for the complex with CF24 have the same space group, $P2_1$, but a different cell than those obtained for CF10, as shown in Table 1. A self-rotation function analysis indicates the presence of non-crystallographic 2-fold symmetry and, therefore, two protein molecules in the asymmetric unit. The limit of diffraction for Form II crystals was found to be 1.9 Å using Cu Kα X-radiation produced by a rotating anode source. Usable diffraction data were collected to 2.3 Å resolution using an *R*-axis IIC imaging plate detector mounted on a Rigaku RU200 rotating anode X-ray generator, as shown in Table 2.

Table 1. *Cell constants for crystal Forms I, II, and III*[a]

	Protein	DNA	Temp (K)	*a* (Å)	*b* (Å)	*c* (Å)	β (°)
Form I	CF10	8/16	100	61.4	38.6	129.6	100.6
Form II	CF24	8/8	108	66.0	63.5	73.4	102.9
Form III	CF24	6/10	108	107.6	39.0	156.6	90.2

[a]Cell constants are reported for the three different crystal forms. The protein molecules are designated CF10, for catalytic fragment with an N-terminus of residue 10, and CF24 for catalytic fragment with an N-terminus of residue 24. DNA molecules are designated by primer/template lengths.

Form III crystals

A third crystal form has been obtained for a complex including a 6/14-mer or 6/10-mer oligonucleotide, ddCTP, and CF24. The sequence for the duplex portion of the oligonucleotide substrate in this complex is different than that for Form I or II crystals, as shown in Figure 1. Crystals were obtained for CF24 complexed with 6/14-mer from three different PEG-containing solutions in an incomplete factorial screen. The next crystallization screens included optimization of pH, salt concentration, and PEG concentration for the initial crystallization conditions identified. Microseeding techniques were used to obtain crystals large enough for characterization using 80 mM magnesium acetate $Mg(OAc)_2$, 13% PEG 4000, and either 50 mM MES pH 6.0, 50 mM ADA (pH 6.5), or 50 mM HEPES (pH 7.5). The space group, $P2_1$, and cell constants obtained for these crystals are shown in Table 1.

Additional screens were done using the 6/10-mer and a 6/8-mer to improve the size of the crystals. In the 6/8-mer, the 8-mer is the same sequence as the 10-mer with two bases removed from the 5′ end. A significant improvement resulted from optimization of the crystallizations with 6/10-mer complexed to CF24, and the best condition for self-nucleation was identified as that containing 20% PEG 4000, 80 mM $Mg(OAc)_2$, and 50 mM Tris pH 7.9 (see Fig. 2). Microseeding methods were used under similar conditions (16% PEG 4000) to increase the size of the crystals. The space group and cell constants for the 6/10-mer were determined from cryocooled crystals, and are the same as those reported in Table 1 for the 6/14-mer complex. Similar to Form I and II crystals, Form III crystals clearly contain more than one protein molecule in the asymmetric unit. In this case, the asymmetric unit most likely includes four protein molecules and two DNA molecules. A V_m value (Matthews, 1968) of 2.65 and solvent content of 53.6% were calculated for four protein molecules and two DNA molecules in the asymmetric unit. Observed V_m values can be obtained with one to four DNA molecules per asymmetric unit but not less than four protein molecules. Typical microseeded crystals have dimensions of 40 × 40 × 500 μm^3. Macroseeding techniques have produced a crystal sufficiently large (100 × 50 × 500 μm^3) to collect a preliminary data set to 4 Å resolution (data statistics are shown in Table 2) using an *R*-axis IIC imaging plate system mounted on a Rigaku RU200 rotating anode generator. The limit of diffraction for Form III crystals on a rotating anode source is 2.5 Å.

Cryocooling of the crystals

All the crystal forms were subjected to cryocooling for characterization and subsequent data collection. Significant difficulties were

Table 2. *Data statistics for Form I, II, and III cryocooled crystals*[a]

	Resolution	Mosaicity	Observed	Unique	Ave I/σ	% Complete	R_{sym}
Form I	10.0–2.30	0.46	74,572	24,281	18.3	91.3	0.047
	2.34–2.30				4.7	84.0	0.134
Form II	20.0–2.30	0.51	91,396	26,120	21.3	98.7	0.048
	2.38–2.30				5.8	97.1	0.174
Form III	20.0–4.0	1.2	19,229	9,323	12.2	80.3	0.114
	4.14–4.0				7.6	67.9	0.157

[a]The resolution is reported in Å. Mosaicity is reported in degrees. Observed is the total number of reflections measured, and unique is the number of unique reflections. Ave I/σ, % complete, and R_{sym} are given for all data in the resolution range specified. R_{sym} is $\sum|I - \langle I\rangle|/\sum I$.

encountered in attempts to successfully cryocool Form I crystals due to a 25 Å change in the *c* cell constant. The room temperature cell of Form I crystals is $a = 62.2$ Å, $b = 39.2$ Å, $c = 136.3$ Å, $\beta = 102.4°$. Cryocooling of these crystals without crosslinking resulted in the following cell: $a = 60.9$ Å, $b = 38.6$ Å, $c = 110.9$ Å, $\beta = 104.1°$. The magnitude of this change corresponds to cell shrinkage of 19% for *c*. The large change led to a large mosaic spread (>2°) for the crystals and an inability to process the data. Room temperature data were obtained to 2.8 Å resolution. A new vapor diffusion glutaraldehyde crosslinking method (C. Lusty, in prep.) was used to crosslink the crystals prior to flash cooling. Derivatization of the crystals with mercuric acetate was also found to reduce shrinkage of the *c* cell constant but without crosslinking was not sufficient to produce usable diffraction data. The cell constants for Form I crystals, which have been derivatized with mercuric acetate and crosslinked prior to cryocooling, are reported in Table 1. Crosslinked Form I crystals were successfully cryocooled using 20% ethylene glycol. Form II crystals did not require crosslinking for

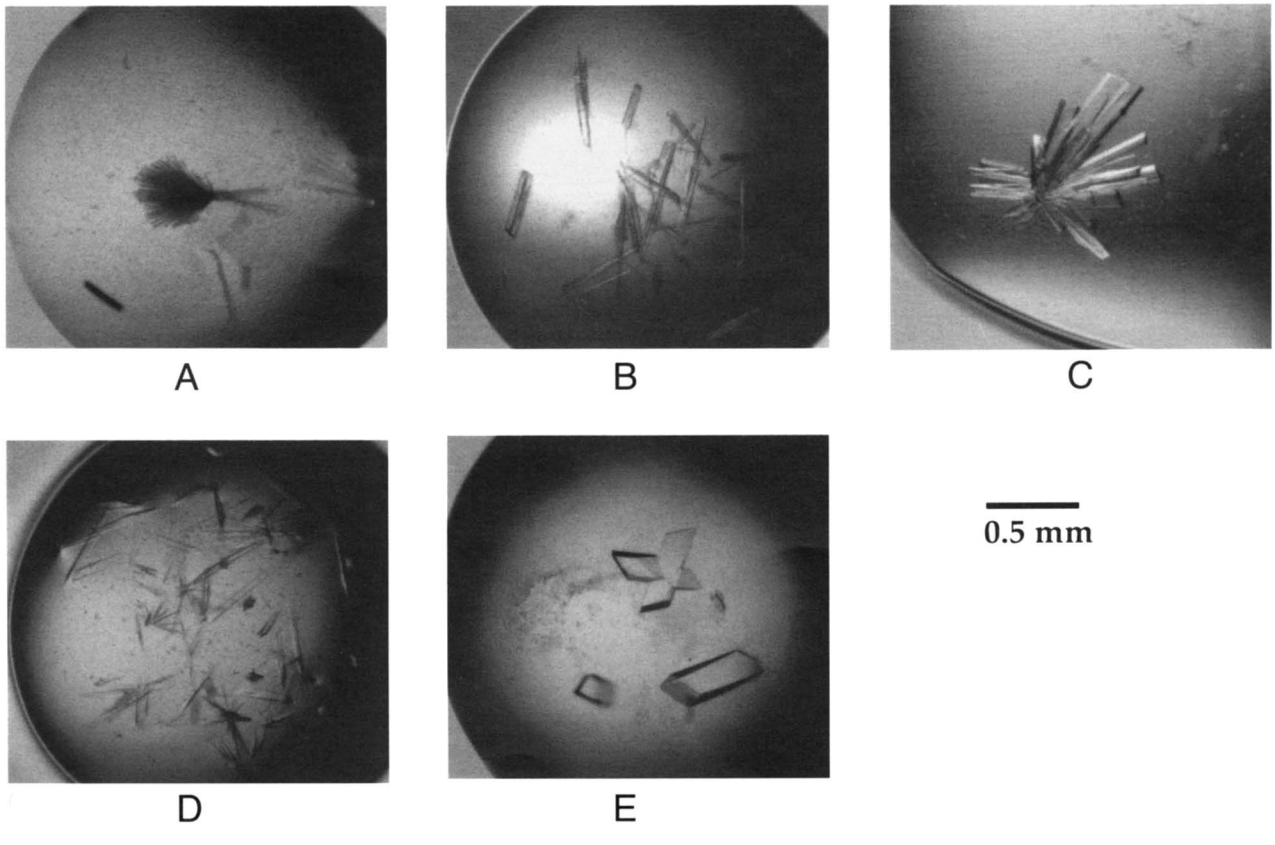

Fig. 2. Microseeded crystals grown in the presence of (**A**) 16/16-mer, (**B**) 8/16-mer, or (**C**) 8/8-mer are shown. Both 8/16-mer and 8/8-mer crystals have been characterized and shown to be Form I crystals. (**D**) Microseeded Form III crystals grown with 6/10-mer complexed to the catalytic fragment are shown. (**E**) Typical Form II crystals obtained by microseeding are shown.

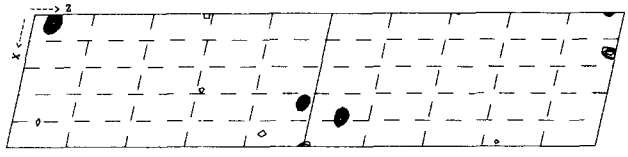

Fig. 3. The Harker section, U = 1/2, is shown for an isomorphous difference Patterson calculated for the 5-iodo-uracil–substituted derivative obtained for Form I crystals including data from 20–3 Å resolution. Contouring levels are shown starting at 3σ with 0.5σ increments. The two unique self-vectors (0.34, 0.5, 0.48) and (0.38, 0.5, 0.55) with peak heights of 6.3σ and 5.7σ result from atomic sites (0.83, 0.01, 0.76) and (0.20, 0.0, 0.28), respectively, and are present in the asymmetric unit shown in this Harker section. In addition, one cross-vector (0.04, 0.5, 0.04) with peak height of 9.1σ is present on this section. The other cross-vector (0.63, 0.0, 0.48) has a peak height of 10.7σ and falls on the U = 0 section.

successful cryocooling, and were frozen in 12–20% glycerol, depending on the size of the crystal. Form III crystals have also been successfully cryocooled in 20% glycerol without crosslinking.

Discussion

CF24 was specifically constructed to include residues 24–278 in addition to a removable vector encoded N-terminal hexa-histidine affinity tag rather than beginning with residue 10, which is the N-terminus defined by limited proteolysis. The construct was designed based on the fact that ordered electron density was observed in the crystal structure of the catalytic fragment produced by limited proteolysis starting with residue 24 at the N-terminus (Georgiadis et al., 1995). Advantages of using CF24 include the ability to overexpress the protein and purify relatively large quantities of the enzyme in addition to the ability to obtain a new crystal form of the complex with nucleic acid, discussed below.

Design of the oligonucleotides used for crystallization was based on the following considerations. Substrates for polymerization by reverse transcriptase include duplex template/primer with a single-stranded template overhang. Based on previous modeling studies (Georgiadis et al., 1995), it appeared that the binding site of the catalytic fragment could accommodate approximately 8–10 base pairs of duplex DNA, consistent with the results reported for the HIV-1 RT/Fab/DNA crystal structure (Jacobo-Molina et al., 1993). Therefore, our initial design included an 8-mer duplex with a 10 base template overhang. Template-directed incorporation of two ddCTPs was included in the design to trap the substrate in the active site (Pelletier et al., 1994).

The choice of the nucleic acid sequence to be used for crystallization is somewhat arbitrary in the sense that polymerases cannot have substrate sequence specificity. Although the self-complementary nature of the 8-mer sequence will result in the formation of 8/8-mer, 8/16-mer, and 16/16-mer (see Fig. 1) in solution when annealed in the presence of the 16-mer complementary template strand, we expected that the enzyme would utilize the 8/16-mer oligonucleotide with only one single-stranded overhanging template, incorporate ddCTP, and thus drive the equilibrium toward 8/16-mers in the crystallization experiment. The only known specificity for the substrate in the polymerase active site is the single-stranded template overhang in a template/primer duplex and the 3′-OH of the primer (Pelletier et al., 1994; Steitz et al., 1994). The considerations discussed above for the 8/16-mer are also relevant to the 8/18-mer used in the initial experiments.

Form I crystals grow under conditions quite similar to those obtained for crystals of the catalytic fragment alone. By increasing the concentration of NH_4Cl, it was possible to obtain the trigonal catalytic fragment crystals previously reported (Georgiadis et al., 1995) and easily recognized from their morphology. No crystals of uncomplexed catalytic fragment were obtained at the lower NH_4Cl concentrations used to grow the complex Form I crystals. Although the 5-iodo-uracil substituted DNA derivative data obtained for Form I crystals were not useful for phasing beyond 5 Å resolution (see Materials and methods), the Patterson analysis provides compelling evidence for the presence of one 8/8-mer DNA molecule per asymmetric unit in the complex crystals. The preference for 8/8-mer in Form I crystals, based on the isomorphous difference Patterson results, indicates that ddCTP was not incorporated and is likely not to be present in these crystals. Thus, our preliminary analysis indicates that the 8/16-mer was not favored over the 8/8-mer in the Form I crystals, as we had originally expected. The apparent discrepancy between expected and observed results of DNA binding will undoubtedly be clarified following completion of the structure determination of Form I crystals.

Optimization of crystallization conditions included using both the catalytic fragment produced by limited proteolysis (CF10) and the N-terminally truncated construct (CF24). Crystallization of the complex including CF24 and 8/8-mer oligonucleotide, which was complexed with CF10 resulting in Form I crystals, produced completely different Form II crystals. Others have reported significant differences in the quality of the crystals due to N- or C-terminal truncation of the protein (Flanagan et al., 1992; Norledge et al., 1996; Pfuetzner et al., 1997). However, in our study, removal of the N-terminal residues 10–23 allowed for a completely different packing of the molecules under nearly identical crystallization conditions. In addition, Form II crystals exhibited a much stronger diffraction pattern, which allowed us to collect data to 2.3 Å using our home source X-ray generator and *R*-axis IIC image plate detector.

Varying the length of the single-stranded portion of the template strand of the oligonucleotide substrate primarily affected the growth of the crystals but not the packing. In the rather special case of the 8/18-mer series of complexes, our initial results with 8/18-mer, 8/16-mer, and 8/14-mer had indicated that there was an optimal length for the single-stranded overhang. But subsequent experiments in which 8/8-mer, 8/16-mer, and 16/16-mer were tested indicated that the primary effect of templates with long single-stranded overhangs was to limit the growth of the crystals. This also appears to be true for Form III crystals, in which 6/10-mer complexes result in larger crystals than those with 6/14-mer. However, varying the length of the duplex portion of the substrate appears to have a larger effect than that observed for the length of the single-strand portion of the template. For a complex of CF24 and 6/10-mer, a completely different crystal form is obtained.

Crystals obtained of the uncomplexed catalytic fragment contain one molecule of protein in the asymmetric unit (Georgiadis et al., 1995). In the absence of nucleic acid, the MMLV RT molecules (full-length MMLV RT, ΔRH, and catalytic fragment) that we have purified are monomeric as isolated, based on results of gel filtration chromatography and dynamic light-scattering experiments (M.M. Georgiadis, unpubl. obs.). Our preliminary analysis of Form I data indicates that the crystals contain two protein molecules and one 8/8-mer per asymmetric unit. Form II crystals likely contain the same complex found in Form I crystals, and Form III crystals most likely contain at least four protein molecules in the asymmetric unit. The interactions that allow the formation of a complex

including two protein molecules and one 8/8-mer will be elucidated following completion of the structure determinations of Form I and Form II crystals.

Further studies currently in progress include efforts to improve the size and diffraction quality of Form III crystals and obtain phasing for crystal Forms I and II. Difficulties were encountered in preparing isomorphous derivatives for crystal Form I (see Materials and methods). As a result, the structure determination of crystal Form I will utilize molecular replacement and MAD phasing methods, if necessary, and that of Form II will utilize molecular replacement and isomorphous derivative phasing. We expect that our catalytic fragment–DNA complex structures will provide new insights for nucleic acid interactions with reverse transcriptase.

Materials and methods

Cloning and expression of the catalytic fragment

DNA encoding the catalytic fragment was amplified by PCR using pRT (Telesnitsky et al., 1992) as the template and two oligonucleotide primers A and B. Upstream primer A, 5′-ATC TAG CTA CAT ATG ACA TGG CTG TCT GAT TTT-3′, includes an Nde I restriction site (underlined) and a translation start site ATG within the Nde I site. The downstream primer B, 3′-GAT TTT CTC CCA GTC TCT ATT GAG CTC ATA GCC-5′, contains an Xho I restriction site (underlined). The PCR products were purified using the Qiaex II gel extraction kit (Qiagen), digested with Nde I and Xho I, and ligated into pET15b (Novagen). DH5α were transformed with the product plasmids. Plasmid clones including the expected DNA insert were purified and used to transform BL21 (DE3). Induction experiments were done to verify that a protein of the appropriate size could be induced by addition of IPTG to cultures of transformants in BL21 (DE3). Cell pellets from the induction experiments were lysed directly in SDS sample buffer and subjected to SDS-PAGE analysis. Several clones that expressed an approximately 30,000 Da protein were then sequenced to verify that no mutations had been introduced by PCR.

Purification of the bacterially expressed catalytic fragment

The initial steps of the purification were modified slightly from those described by Qiagen for hexa-histidine tagged proteins. An 8-mL Ni-NTA Superflow (Qiagen) column was used on an FPLC for the chromatographic separation. Fractions eluted from the Ni-NTA column including catalytic fragment greater than 70% pure were combined and diluted 10-fold in Buffer A (50 mM MES pH 6.0, 1 mM DTT, and 50 mM sodium chloride) and loaded on a Mono-S 16/10 column on the FPLC. Gradient fractionation of the sample using a linear gradient from 50 mM to 1 M NaCl was used to elute nearly homogeneous catalytic fragment at approximately 0.8 M NaCl. Approximately 100 mg of purified catalytic fragment was then digested with thrombin as recommended by Novagen. Removal of the N-terminal affinity tag from the protein by thrombin was verified by SDS-PAGE, and the digested catalytic fragment was then subjected to a second Mono-S 16/10 purification step, as previously described, and eluted at approximately 0.35 M NaCl. Purified CF24 includes N-terminal residues Gly Ser His Met from the vector in addition to residues 24–278 from MMLV RT. CF24 was then concentrated with Amicon centricons to 10–15 mg/mL as isolated in 50 mM MES (pH 6.0), 1 mM DTT, and 0.35 M NaCl, and stored at −80 °C.

Purification of the MMLV RT, ΔRH, and proteolytic fragment

MMLV RT, ΔRH, and the catalytic fragment produced by limited proteolysis were purified as described elsewhere (M.M. Georgiadis, in prep.) The catalytic fragment was concentrated to 30–40 mg/mL using Amicon centricons as isolated in 50 mM MES (pH 6.0), 1 mM DTT, and 0.20 M NaCl, and stored at −80 °C. The sample was then diluted to 10–15 mg/mL for crystallization experiments in the same buffer.

Purification of oligonucleotides

An ABI 392 synthesizer was used for 1 μmol scale syntheses of the oligonucleotides with the dimethyl trityl groups left on for purification purposes. Standard deprotection, C18 reverse phase chromatography, and detritylation procedures were used to purify the oligonucleotides. Precautions taken to purify 5-iodo-uracil containing oligonucleotides included minimizing exposure to light. Following purification, samples were dried and redissolved in 10 mM HEPES (pH 7.0) and 10 mM $MgCl_2$ with final concentrations of 5 mM. Duplex oligonucleotide samples were annealed with final concentrations of 2.5 mM. Sequence length and purity were determined by visualization using 19% urea polyacylamide gel electrophoresis.

Crystallization of nucleic acid complexes with the catalytic fragment

Hanging drop vapor diffusion methods were used for crystallization of nucleic acid complexes with the catalytic fragment. Initial screens for crystallization conditions for Form I crystals were done using an incomplete factorial screen (J. Jancarik, pers. comm.). Crystallization conditions for Form III crystals were obtained from a different incomplete factorial screen using Natrix solutions (Hampton Research). Drops contained 1 μL of pre-formed complex and 1 μL of precipitant solution equilibrated over 0.5 mL of precipitant solution. The complex was preformed by mixing catalytic fragment, oligonucleotide, and ddCTP and incubating on ice for 1 h prior to crystallization. The concentrations of components used to form the complexes were as follows: catalytic fragment 13 mg/mL, oligonucleotide 2.5 mM, and ddCTP 20 mM. All of the crystal forms were grown at 20 °C.

For all crystal forms reported here, microseeding techniques were used to produce crystals suitable for X-ray diffraction work. Microseeding was done by selecting a drop containing a sufficient crystalline mass and then crushing the crystal or crystals with a needle in the drop. Several microliters (5–10 μL) of stabilizing solution were then added to the drop. Tenfold serial dilutions from 10 to 10^6 were then done by adding 2 μL of microseed-containing solution to 18 μL of the precipitant solution. Hanging drops consisting of 1.5 μL of diluted microseed–containing solution and 1.5 μL of preformed complex were then suspended over 0.5 mL of appropriate precipitant solution. Dilutions that produced the best crystals for Forms I and II were 10^3–10^5 dilutions. Form III crystals were grown from dilutions of 10^2 to 10^4.

For the Form I crystallization experiments in which oligonucleotides 8/8-mer, 8/16-mer, and 16/16-mer were tested by the hanging drop vapor diffusion method, the precipitant solutions screened included 100 mM HEPES (pH 7.5), 75 mM NH_4Cl, and 14, 16, 18, or 20% PEG 4000. Complexes were formed for each of the oligonucleotide primer/template combinations listed above as previ-

ously described, and all three complex drops were suspended over the same reservoir for each of the different precipitant solutions. Hanging drops were microseeded using 10-fold, 100-fold, or 1,000-fold seed dilutions derived from previously grown 8/16-mer complex crystals for the four different PEG concentrations and three different complexes.

Crystal Forms I and III both required macroseeding techniques to produce crystals suitable for X-ray diffraction work. Successful macroseeding required that freshly grown microseeded crystals be used as seeds. Form I crystals required pre-equilibration of drops placed on 25-mm cover glasses containing 2 μL of complex and 2 μL of precipitant solution in 35 × 10 mm Petri dishes with 1 mL of precipitant solution sealed with parafilm for 4–6 h prior to addition of seeds. Selected macroseeds were perfect in appearance and approximately 10–15 μm in the smallest dimension. The seeds were washed five times in 0.25 mL of washing solution (100 mM Hepes (pH 7.5), 8% PEG 4000, 50 mM NaCl, and 75 mM NH_4Cl) and subsequently four times in 0.25 mL of the precipitant solution. The seed was introduced into the pre-equilibrated drops in 0.5 μL volume using a pipetman. Macroseeded crystals required two to three days to reach maximum size.

Macroseeding for Form III crystals involved selecting microseeded crystals as seeds, which appeared to be perfect and at least 10 μm in the smallest dimension. The seeds were then washed four times in 0.25 mL of washing solution (50 mM Tris (pH 7.9), 80 mM magnesium acetate, and 6% PEG 4000) and four times in precipitant solution (50 mM Tris pH 7.9, 80 mM magnesium acetate, and 12% PEG 4000). Drops were made by adding 1.5 μL of precipitant solution including the seed to 1.5 μL of preformed complex and suspending over 0.5 mL of precipitant solution. Macroseeded crystals grew to maximal size in two to three days.

Derivatization of Form I crystals using 5-iodo-uracil–substituted oligonucleotides

5-Iodo-uracil–substituted oligonucleotides were synthesized and purified as described above. Several different substituted oligonucleotides were prepared, but only one duplex (5′-CATGCAUG-3′ annealed to 3′-GTACGUACGGCGTACT-5′) crystallized with the same cell as the native crystals. U is used to denote 5-iodo-uracil used in place of thymine as the base. Data for this derivative and a native crystal were collected to 2.8 Å at room temperature, integrated using DENZO (Otwinowsky, 1993), and scaled using CCP4 programs (CCP4, 1994). Difference Patterson and phasing calculations were performed using CCP4 (CCP4, 1994). Phasing from the two sites (0.20, 0, 0.28) and (0.83, 0.01, 0.76) proved to be usable to only 5 Å resolution based on phasing analysis done with MLPHARE, and despite the very interpretable appearance of the Patterson map. A contributing factor includes fractional y-coordinates for the two sites differing by only 0.01, which necessarily limits the phasing information in this space group $P2_1$. In addition, the phasing analysis indicated nonisomorphism problems for the higher resolution data. Due to the cryocooling problems encountered with Form I crystals, further phasing experiments involving isomorphous derivative phasing were abandoned in favor of molecular replacement phasing and MAD phasing approaches.

Crosslinking of Form I crystals

Macroseeded Form I crystals were crosslinked by introducing glutaraldehyde by vapor diffusion (C. Lusty, in prep.). This is a new method for crosslinking crystals and has been used successfully on three problems. After the crystal had grown to maximal size, a second drop of 2 μL of 25% glutaraldehyde aqueous solution (Sigma, EM grade) was added to the cover glass and the Petri dish was resealed. Crosslinking of the crystal was monitored with time by placing crystals in water and observing whether the crystal dissolved. Crosslinking of Form I crystals required 2 h. Crystals were then transferred to a stabilizer solution.

Cryo-cooling of complex crystals

Form I crystals were serially transferred through solutions including 10 mM HEPES (pH 7.5), 75 mM NH_4Cl, 50 mM NaCl, 3 mM $MgCl_2$, 14% PEG 4000, and 1–20% ethylene glycol at 1% intervals of ethylene glycol. Form II crystals were transferred through similar cryosolutions replacing ethylene glycol with 12–20% glycerol as the final cryoprotectant concentration. Form III crystals were transferred through solutions containing 50 mM Tris (pH 7.9), 80 mM magnesium acetate, 50 mM NaCl, 3 mM $MgCl_2$, 14% PEG 4000, and 1–20% glycerol in 1% increments of glycerol. Crystals were equilibrated in 0.2 mL of each cryosolution for 2–5 min and introduced into the cryostream (108 K) with a cryoloop. An Oxford cryosystem was used to produce the 108 K cryostream used for all freezing experiments.

Data collection and self-rotation functions

Data for Form II and Form III crystals reported here were collected on either an R-axis IIC or R-axis IV image-plate detector mounted on a Rigaku RU 200 rotating anode system with MSC mirror systems operating at 50 mA × 90 kV. Exposure times were typically 30 min for 1–1.5° oscillation exposures. Data for Form I crystals were collected both on our home source detectors and at NSLS, BNL beamline X4A. Details of the synchrotron data collection will be reported elsewhere. Data were integrated and processed with DENZO and SCALEPACK (Otwinowsky, 1993). Self-rotation functions were calculated for Form I and Form II crystals using data from 10–4 Å resolution with AMORE (CCP4). Non-crystallographic 2-fold axes (α = 180.0, β = 147.0, γ = 0.0) with a correlation of 68.0 and (α = 45.7, β = 169.2, γ = 134.2) with a correlation of 33.1 were obtained for Form I and Form II data, respectively.

Supplemental material

We observed in our complexation experiments that the addition of DNA to the protein at high protein concentrations resulted in the formation of precipitate. This is likely due to the requirement of NaCl to maintain solubility of the protein at high concentrations. However, in order to form the complex, the overall salt concentration must be lowered beyond that required for the solubility of the protein at near millimolar concentrations. The end result is that the complex is only soluble at protein concentrations of approximately 0.2 mM compared to the protein alone, which is soluble to concentrations of greater than 1 mM. Therefore, the observed effect of ratios of DNA to protein required to optimize crystallization conditions may simply be a dilution effect. Our stock oligonucleotide substrate concentrations were always 2–2.5 mM duplex. Although we tried using protein concentrations from 10–15 mg/mL (330–500 μM), we did not systematically change the concentrations of the oligonucleotide and ddCTP solutions used in order to

eliminate dilution effects in forming the complex. There was no discernible difference in the solubility of CF10 versus CF24 alone or in the complex solutions.

Acknowledgments

This work was supported by a grant (GM55026) from the National Institutes of Health. We thank Alice Telesnitsky for providing pRT with the MMLV RT cDNA. We thank Wayne Hendrickson and Helen Berman for helpful discussions pertaining to this project and Tax Georgiadis for proofreading this manuscript. We also thank Craig Ogata for assistance with data collection at NSLS, HHMI beamline X4A.

References

Aggarwal AK. 1990. Crystallization of DNA binding proteins with oligodeoxynucleotides. *Methods (San Diego) 1*:83–90.

Blum H, Beier H, Gross HJ. 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis 8*:93–99.

Carter CW Jr, Carter CW. 1979. Protein crystallization using incomplete factorial experiments. *J Biol Chem 254*:12219–12223.

CCP4. 1994. Collaborative computational project, number 4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D50*:760–763.

Ding J, Das K, Moereels H, Koymans L, Nadries K, Janssen PAJ, Hughes SH, Arnold E. 1995a. Structure of HIV-1 RT/TIBO R86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat Struct Biol 2*:407–415.

Ding J, Das K, Tantillo C, Zhang W, Clark AD Jr, Jessen S, Lu X, Hsiou Y, Jacobo-Molina A, Andries K, Pauwels R, Moereels H, Koymans L, Janssen PAJ, Smith, RH Jr, Koepke MK, Michejda CJ, Hughes SH, Arnold E. 1995b. Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor α-APA R95845 at 2.8 Å resolution. *Structure 3*:365–379.

Esnouf R, Ren J, Ross C, Jones Y, Stammers D, Stuart D. 1995. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat Struct Biol 2*:303–308.

Flanagan JM, Kataoka M, Shortle D, Engelman DM. 1992. Truncated staphylococcal nuclease is compact but disordered. *Proc Natl Acad Sci USA 89*:748–752.

Georgiadis MM, Jessen SM, Ogata CM, Telesnitsky A, Goff SP, Hendrickson WA. 1995. Mechanistic implications from the structure of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase. *Structure 3*:879–892.

Gilboa E, Mitra SW, Goff S, Baltimore D. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell 18*:93–100.

Goff SP. 1990. Review: Retroviral reverse transcriptase: Synthesis, structure, and function. *J Acquir Immun Defic Syndrome 3*:817–831.

Guo J, Wu W, Yuan ZY, Post K, Crouch RJ, Levin JG. 1995. Defects in primer-template binding, processive DNA synthesis, and RNaseH activity associated with chimeric reverse transcriptases having the murine leukemia virus polymerase domain joined to *Escherichia coli* RNase H. *Biochemistry 34*:5018–5029.

Hsiou Y, Ding J, Das K, Clark A Jr, Hughes SH, Arnold E. 1996. Structure of unliganded HIV-1 reverse transcriptase at 2.7Å resolution: Implications of conformational changes for polymerization and inhibition mechanisms. *Structure 4*:853–860.

Jacobo-Molina A, Ding J, Nanni RG, Clark AD Jr, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P, Hizi A, Hughes SH, Arnold E. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double stranded DNA at 3.0 Å resolution show bent DNA. *Proc Natl Acad Sci USA 90*:6320–6324.

Jager J, Smerdon SJ, Wang J, Biosvert DC, Steitz TA. 1994. Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. *Structure 2*:869–876.

Jancarik J, Kim S-H. 1991. Sparse matrix sampling: A screening method for crystallization of proteins. *J Appl Crystallogr 24*:409–411.

Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science 256*:1783–1790.

LeGrice SFJ. 1993. Human immunodeficiency virus reverse transcriptase. In: Goff AMS, Goff SP, eds. *Reverse transcriptase*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. pp 156–191.

Leis H, Duyk G, Johnson S, Longiaru M, Skalka AM. 1983. Mechanism of action of the endonuclease associated with the ab and bb forms of avian RNA tumor virus reverse transcriptase. *J Virol 45*:727–739.

Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol 33*:491–497.

Moelling K. 1974. Characterization of reverse transcriptase and RNaseH from friend-murine leukemia virus. *Virology 146*:46–59.

Norledge BV, Mayr E-M, Glockshuber R, Bateman OA, Slingsby C, Jaenicke R, Driessen HPC. 1996. The X-ray structure of two mutant crystallin domains shed light on the evolution of multi-domain proteins. *Nat Struc Biol 3*:267–274.

Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA. 1985. Structure of the large fragment of *E. coli* DNA polymerase I complexed with dTMP. *Nature 313*:762–766.

Otwinowsky Z. 1993. Oscillation data reduction program. In: Sawyer L, Isaacs N, Bailey S, compilers. *Proceedings of the CCP4 study weekend: Data collection and processing*. Daresbury, UK: SERC Daresbury Laboratory. pp 56–62.

Pelletier H, Sawaya MR, Kumar A, Wilson SH, Kraut J. 1994. Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP. *Science 264*:1891–1903.

Pfuetzner RA, Bochkarev A, Frappier L, Edwards AM. 1997. Replication protein A. Characterization and crystallization of the DNA binding domain. *J Biol Chem 272*:430–434.

Ren J, Esnouf R, Garman E, Somers D, Ross C, Kirby I, Keeling J, Darby G, Jones Y, Stuart D, Stammers D. 1995. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat Struc Biol 2*:293–302.

Rodgers DW, Gamblin SJ, Harris BA, Ray S, Culp JS, Hellmig B, Woolf DJ, DeBouck C, Harrison SC. 1995. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc Natl Acad Sci USA 92*:1222–1226.

Roth MJ, Tanese N, Goff SP. 1985. Purification and characterization of murine retroviral reverse transcriptase expressed in *Escherichia coli*. *J Biol Chem 260*:9326–9335.

Steitz TA, Smerdon SJ, Jager J, Joyce CM. 1994. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science 266*:2022–2025.

Telesnitsky A, Blain SW, Goff SP. 1992. Defects in Moloney murine leukemia virus replication caused by a reverse transcriptase mutation modeled on the structure of *Escherichia coli* RNase H. *J Virol 66*:615–622.

Telesnitsky A, Goff SP. 1993. RNase H domain mutations affect the interaction between Moloney murine leukemia virus reverse transcriptase and its primer-template. *Proc Natl Acad Sci USA 90*:1276–1280.

Unge T, Knight S, Bhikhabhai R, Lögren S, Dauter Z, Wilson K, Strandberg B. 1994. 2.2 Å resolution structure of the amino-terminal half of HIV-1 reverse transcriptase (fingers and palm subdomains). *Structure 2*:953–961.