An Expanded Model of Replicating Human Immunodeficiency Virus Reverse Transcriptase†

Birgitta M. Wöhrl,‡ Chris Tantillo,§ Edward Arnold,§ and Stuart F. J. Le Grice*,‡

Division of Infectious Diseases, Case Western Reserve University School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, and Center for Biotechnology and Medicine, Chemistry Department, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854-5638

Received September 27, 1994; Revised Manuscript Received February 7, 1995⊗

ABSTRACT: Replication complexes containing wild-type and RNase H-deficient p66/p51 human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT) were analyzed by DNase I and S1 footprinting. While crystallography and chemical footprinting data demonstrate that 15−18 bases of primer and template occupy the DNA polymerase and RNase H active centers, enzymatic footprinting suggests that a larger portion of substrate is encompassed by the replicating enzyme. Independent of the position of DNA synthesis arrest, template nucleotides +7 to −23 and primer nucleotides −1 to −25 are nuclease resistant. On both DNA strands, position −20 remains accessible to DNase I cleavage, suggestive of an alteration in nucleic acid structure between exiting the RNase H catalytic center and leaving the C-terminal p66 domain. A model of HIV-1 RT containing an extended single-stranded template and duplex region was constructed on the basis of the structure of an RT/DNA complex. Mapping of footprint data onto this model shows consistency between biochemical and structural data, implicating a contribution from domains proximal to the catalytic centers.

Solution of the three-dimensional structure of human immunodeficiency virus reverse transcriptase (HIV RT;[1] Kohlstaedt et al., 1992; Arnold et al., 1992; Jacobo-Molina et al., 1993) represents an important step toward designing improved therapeutic agents to combat the spread of acquired immunodeficiency syndrome (AIDS). In the RT/DNA cocrystal (Jacobo-Molina et al., 1993), 18 bp of duplex DNA, adopting both A- and B-form geometry, are accommodated by the DNA polymerase and RNase H active centers. Furthermore, these A- and B-form regions are separated by a 40−45° bend in the DNA substrate. A parallel study of replication complexes via susceptibility to cleavage by iron−EDTA-generated hydroxyl radicals (Metzger et al., 1993) provided a picture of retroviral plus-strand DNA synthesis remarkably similar to the RT/DNA cocrystal. Within 18 template and 15 primer nucleotides contacted by the replicating enzyme, a "window" of reactivity on both strands corresponds to the position at which DNA in the cocrystal is distorted. Hydroxyl radical footprinting also suggested that the DNA substrate adopts A-form geometry. Finally, both approaches indicate that 4 bp of substrate could be accommodated by the RNase H active center.

Similarities between the cocrystal of HIV-1 RT containing template-primer (Jacobo-Molina et al., 1993) and chemical footprinting data of polymerization complexes (Metzger et al., 1993) have raised the possibility that these approaches, both of which provide high-resolution data, highlight only the portion of substrate most tightly contacted by the enzyme, i.e., between its polymerase and RNase H catalytic centers. This would be particularly true for the RT/DNA cocrystal, which used a 19-base template/18-base primer. Extending the model of Jacobo-Molina et al. (1993) to a generalized scheme of plus-strand DNA synthesis predicts that template sequences preceding the polymerase catalytic site could be shielded by the enzyme, while duplex DNA exiting the RNase H catalytic site might interact with structural elements near the C-terminus. Thus, additional domains of RT might be less intimately associated with nucleic acid, yet still play an important role in polymerization and/or translocation. An attractive candidate is the finger subdomain of p66 (residues 1−84 and 120−150), originally proposed by Kohlstaedt et al. (1992) to participate by contacting the single-stranded template. This possibility is strengthened by Nanni et al. (1993), whose revised structural models suggest that the β3−β4 hairpin of the p66 finger subdomain positions extended templates, a notion supported by Boyer et al. (1994). If the p66 finger subdomain participates in template recognition, thereby shielding the single-stranded template, it is not unreasonable that nucleic acid leaving the RNase H active center is still covered by the replicating enzyme. This would be in keeping with the recently proposed "helix clamp" motif of HIV-1 RT and related nucleic acid polymerases, where Hermann et al. (1994) have modeled a substrate extended by ~9 bp toward the RNase H domain.

To address these possibilities, we have performed enzymatic footprinting, with the goal of revealing "external" features of replication complexes containing p66/p51 HIV-1 RT. DNase I footprinting, while offering low-resolution information, has complemented chemical probing agents such

† Supported by NIH Grants GM 46623, AI 31147 (S.F.J.L.G.), and AI 27690 (E.A.). B.M.W. was supported by a grant from the German Bundesministerium für Forschung und Technik.

* Corresponding author (telephone, (216) 368-6989; Fax, (216) 368-2034).

‡ Case Western Reserve University School of Medicine.

§ Rutgers University.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1995.

[1] Abbreviations: bp, base pairs; DNase I, deoxyribonuclease I; HIV, human immunodeficiency virus; IPTG, isopropyl β-D-thiogalactopyranoside; nt, nucleotides; RNase H, ribonuclease H; RT, reverse transcriptase; SDS, sodium dodecyl sulfate.

0006-2960/95/0434-5343$09.00/0

as dimethyl sulfate (Siebenlist et al., 1980), iodine (Schatz et al., 1991) and potassium permanganate (Sasse-Dwight & Gralla 1989), in revealing overall features of several nucleic acid/protein complexes (Johnson et al., 1979; Le Grice et al., 1982; Emerson et al., 1985; Plumb et al., 1985; Sigman et al., 1991). When applied to replication complexes containing HIV-1 RT variants, enzymatic probing shows that nucleic acid adjacent to the active centers is resistant to digestion. Template nucleotides +7 to −23 and primer nucleotides −1 to −25 are nuclease-resistant, independent of the position of DNA synthesis arrest. While protection of template sequences at the leading edge of the replication complex is uninterrupted, a portion of duplex DNA near the RNase H catalytic center remains accessible to cleavage. Transposition of footprinting data onto a model of HIV RT containing an extended substrate confirms that additional structural elements lie close to the nucleic acid during DNA synthesis.

MATERIALS AND METHODS

Enzymes and Substrates. p66/p51 HIV-1 RT and the RNase H$^-$ mutant p66$^{E\rightarrow Q}$/p51 (Schatz et al., 1989) were prepared from recombinant *Escherichia coli* by metal chelate and ion-exchange chromatography as described (Le Grice & Grüninger-Leitch, 1990). Both enzymes were at least 95% pure and free of contaminating nucleases. Enzymes were stored at −20 °C in a buffer of 50 mM Tris-HCl, pH 7.9, 25 mM NaCl, 1 mM EDTA, and 50% (v/v) glycerol at a concentration of 10 mg/mL or greater. RT concentration was determined spectrophotometrically, where a 1 mg/mL solution yields an $OD_{280} = 1.8$ (Lederer et al., 1992).

Substrate for DNase I protection experiments was a 71-nt template/36-nt primer (Figure 1A). Both oligonucleotides were chemically synthesized (Integrated DNA Technologies, Coralville, IA) and purified by denaturing polyacrylamide gel electrophoresis. 5′-End labeling of template or primer was achieved with [γ-^{32}P]ATP and polynucleotide kinase, while 3′ labeling of template DNA required [α-^{32}P]dCTP and the Klenow fragment of DNA polymerase I. Template-primer was annealed by heating to 80 °C, followed by slow cooling to room temperature.

Limited Primer Extension. Specific primer extension was achieved with a polymerization cocktail containing deoxy/dideoxynucleoside triphosphate (dNTP/ddNTP) mixtures shown in Figure 1A. p66/p51 RT was incubated with template-primer for 5 min at 37 °C, after which polymerization was initiated by addition of the appropriate dNTP/ddNTP mixture. After 20 min at 37 °C, DNA synthesis was terminated by addition of an equal volume of a urea-based gel loading buffer (7 M urea in Tris/borate/EDTA buffer containing 0.1% xylene cyanol). Polymerization products were fractionated by high-resolution gel electrophoresis through 12% polyacrylamide gels containing 7 M urea and analyzed by autoradiography. Autoradiographic enhancement made use of the Du Pont Reflection system (Du Pont/NEN).

DNase I Protection Experiments. The DNase I protection protocol of Schmitz and Galas (1979) was followed, with minor modification. RT (200 ng, 1.7 pmol) was incubated with ~1 pmol of end-labeled template-primer in a 15-μL reaction mixture containing 10 mM Tris-HCl, pH 8.0, 6 mM $MgCl_2$, 80 mM NaCl, 5 mM DTT, and the appropriate dNTP/ddNTP mixture. The concentration of elongating dNTP was 50 mM, while that of the chain-terminating ddNTP was raised to 500 mM. After 20 min at 37 °C, 3.5 units of DNase I (Boehringer Mannheim) was added, and after 30 s at room temperature the reaction was terminated by addition of an equal volume of phenol/chloroform/isoamyl alcohol. Nucleic acid in the aqueous phase was precipitated in the presence of glycogen, and the dried sample was resuspended in 10 μL of a urea-based gel loading buffer. Samples were fractionated and processed as described above.

Since each replication complex generated increasing amounts of duplex DNA, it was necessary to prepare control DNase I digests for the +1, +4, and +10 primer extension reactions. To achieve this, the appropriate replication complex was prepared with wild-type or RNase H-deficient RT and then freed of enzyme by phenol extraction. Extended DNA substrates were recovered by ethanol precipitation and subjected to partial hydrolysis with DNase I.

S1 Protection Experiments. S1 footprinting required modification of the DNase I protection protocol. Following preparation of +4 complexes, the sample was supplemented with 40 units of S1 (Boehringer Mannheim) in a 5-fold concentrated S1 buffer to achieve a final concentration of 19 mM sodium acetate, pH 4.5, 29 mM NaCl, and 17.5 μM $ZnSO_4$. Following 30 s of S1 treatment, the reaction was stopped and nucleic acids were processed as described above. Under these conditions, the replication complex remains stable over the digestion period. Control S1 digests of extended substrates in the absence of RT were prepared as described in the previous section.

Molecular Modeling. DNA modeling was based on the crystallographic structure of the ternary complex of HIV-1 RT, a 19-mer/18-mer template-primer, and a monoclonal antibody Fab fragment (Arnold et al., 1992; Jacobo-Molina et al., 1993). Model building was performed with the programs O, version 5.9 (Jones et al., 1991), and Insight (Biosym Inc.), version 2.2, on a Silicon Graphics 4D/240GTX computer. Positioning of the extended template overhang at the polymerase domain assumes A-form helical geometry as recently outlined by Boyer et al. (1994). Modeling the template overhang at the polymerase domain and duplex extension at the RNase H domain were based only on the double-stranded DNA structure, and potential contacts with the protein were not used as a guide. In order to extend the duplex portion of the nucleic acid beyond −18, the phosphate positions of the last 9 bp (positions −10 to −18) were used to anchor the terminus of a B-form double helix. Unfavorable contacts between the modeled DNA and RT were not evident. Only the template overhang and duplex extension were modeled; the DNA structure between the catalytic centers, including the region containing the "window of accessibility" from hydroxyl radical footprinting data (Metzger et al., 1993), corresponds to the structure of the DNA in the crystalline complex.

Routine Chemicals. Radiochemicals were purchased from Du Pont/NEN. dNTP and ddNTP substrates were provided as aqueous solutions from Boehringer Mannheim. Routine chemicals were from Sigma and of the highest purity available.

RESULTS

Construction and Utilization of a Template Containing a Hairpin Structure. The template-primer combination used in our study is indicated in Figure 1A. Since DNase I cleaves

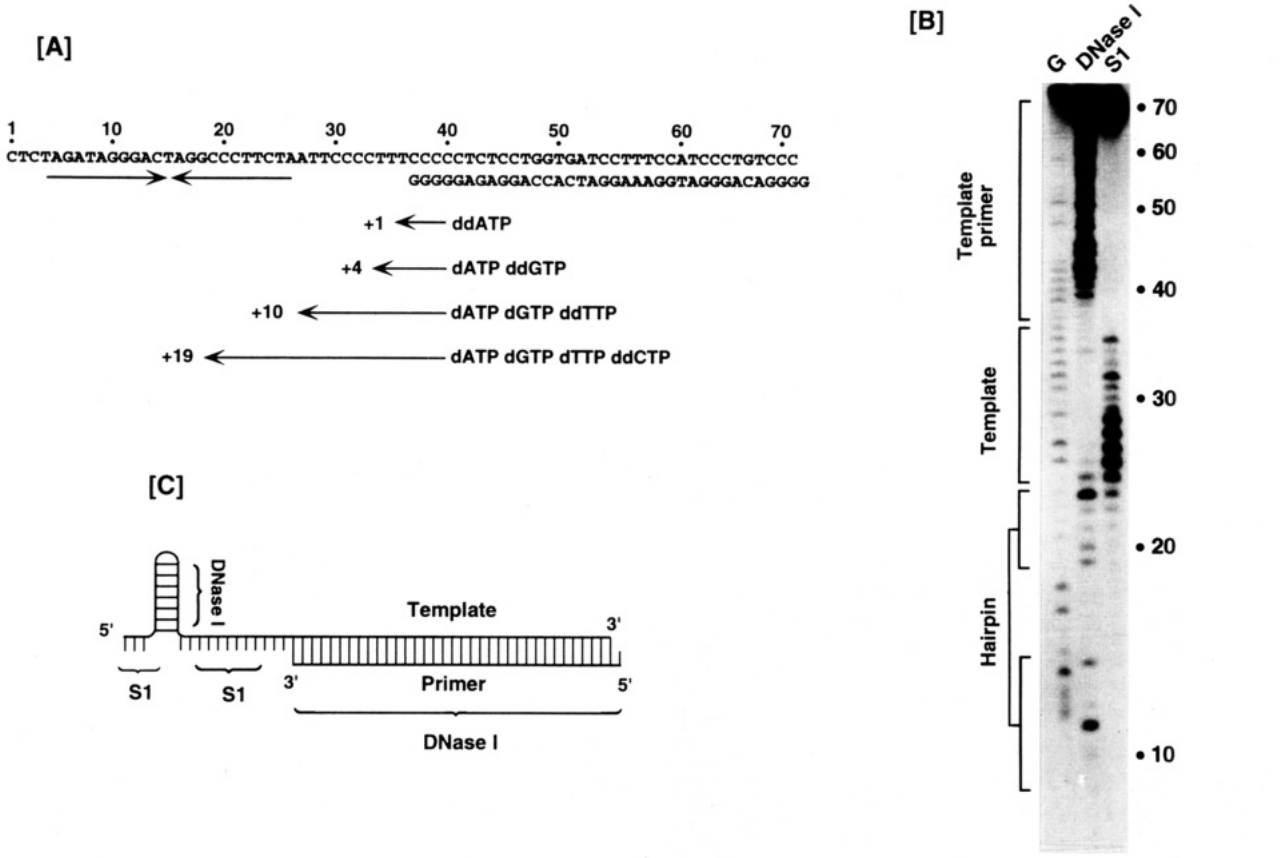

FIGURE 1: (A) Sequence of the synthetic template-primer used for enzymatic footprinting of replication complexes. Arrows below the single-stranded portion of the template represent bases which assume a hairpin structure. The dNTP/ddNTP polymerization cocktails necessary for primer extension by 1, 4, 10, or 19 nucleotides are indicated. In the absence of a ddNTP, complete primer extension is achieved. (B) Analysis of template structure by partial hydrolysis with the nucleases DNase I and S1. For these experiments, template DNA was radiolabeled at its 5′ terminus with ^{32}P and hybridized to an unlabeled primer. Two regions of secondary structure are evident from partial DNase I digestion, between which a short region of single-stranded template is accessible to nuclease S1. Base numbering is relative to the 5′ terminus of the template and was derived from a G-specific enzymatic sequencing reaction run in parallel (lane G). (C) Schematic representation of the structure assumed by the template-primer, indicating a short duplex in the vicinity of the 5′ terminus of the template. Regions displaying sensitivity to DNase I and S1 were derived from the partial DNase I and nuclease S1 hydrolysis profiles of (B).

primarily duplex DNA, partial hydrolysis of a radiolabeled template was expected to reveal only sequences paired with the primer. Conversely, treatment of template-primer with nuclease S1 should reveal ~36 bases of unpaired template. However, Figure 1B indicates that a portion of the single-stranded template was cleaved by DNase I, suggestive of intramolecular base pairing. This notion was confirmed by determining S1 susceptibility of the template, which highlighted only 13 bases adjacent to the template-primer duplex (Figure 1B). The limits of S1 sensitivity are coincident with the DNase I-sensitive region, and only a few bases at the 5′ extremity of the template retained a single-stranded configuration. The combined DNase I/S1 digestion profiles predicted the structure of Figure 1C, where a hairpin is formed near the 5′ terminus of the template.

Such a structure on the template could present an impediment to the translocating enzyme, based on observations of Klarmann et al. (1993) that specific sequences and/or structures on RNA and DNA templates induce stalling of the polymerizing HIV-1 enzyme. To address this, "programmed" primer extension reactions were performed with wild-type and RNase H-deficient p66/p51 RT. Although we demonstrated that DNA polymerase activity of mutant $p66^{E\rightarrow Q}$/p51 was unaffected quantitatively (Schatz et al., 1988), we could not rule out the possibility that altering the RNase H domain might stall the replicating enzyme on the template, making a heterogeneous population of complexes

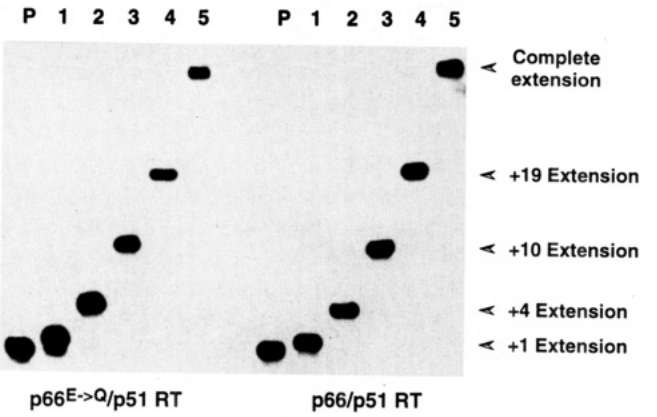

FIGURE 2: Primer extension efficiency of wild-type and RNase H-deficient p66/p51 HIV-1 RT. The substrate for these experiments contained an unlabeled template to which a 5′ ^{32}P end-labeled primer was hybridized. Lanes: P, unextended primer; 1−5, +1, +4, +10, +19, and complete primer extension reactions, respectively. Reaction products were fractionated through a 12% polyacrylamide gel containing 7 M urea.

difficult to analyze by DNase I footprinting. The results of our primer extension analysis are shown in Figure 2, where it is clear that both enzymes quantitatively extend the primer to the position at which the chain-terminating ddNTP is incorporated. From these data, we could not predict whether RT encroached upon the hairpin structure in a replication complex whose primer was extended by 10 nucleotides.

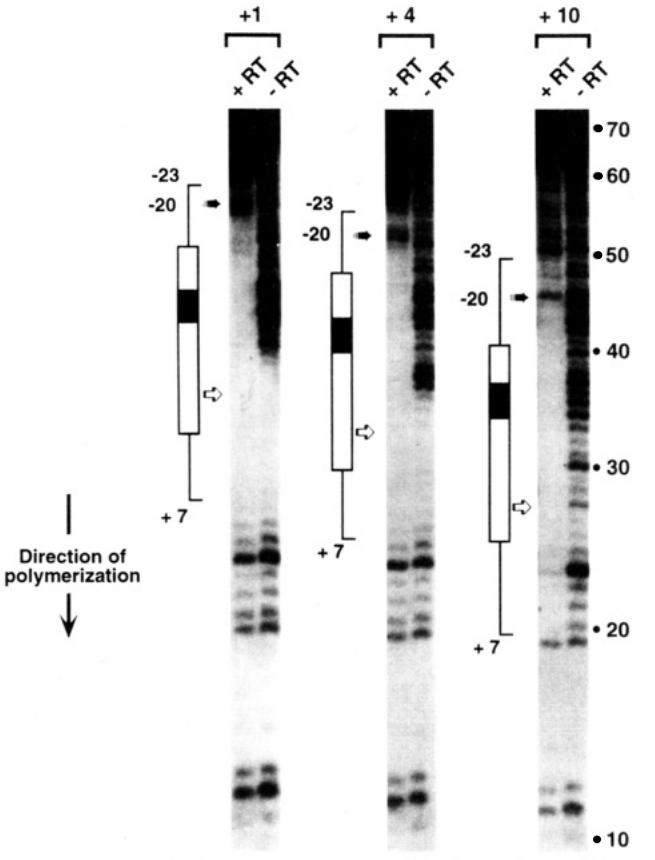

FIGURE 3: Partial DNase I digestion profile of +1, +4, and +10 replication complexes containing wild-type HIV-1 RT. Template DNA was labeled at its 5′ terminus with ^{32}P. The designation −RT represents a control digest of the appropriately extended substrate (prepared for each replication complex). Extremities of each nucleoprotein complex are enclosed by bars, within which the boxed area represents the portion resistant to cleavage by iron−EDTA-generated hydroxyl radicals (Metzger et al., 1993). Within this, filled boxes represent the "window" of nucleotides accessible to chemical cleavage. The position at which the chain-terminating ddNTP was incorporated is indicated (by an open arrow), and is defined as −1. Base numbering of the protected region is relative to this position. Bases in the vicinity of the RNase H active center displaying increased accessibility to DNase I digestion are indicated by a closed arrow.

However, the efficiency with which +19 and full primer extension reactions were catalyzed indicates that both enzymes experienced little difficulty in traversing the hairpin in the absence of accessory viral proteins (the energy for which may be provided by dNTP hydrolysis).

Replication Complexes Containing Wild-Type p66/p51 RT. The +1, +4, and +10 replication complexes containing wild-type HIV-1 RT were analyzed by DNase I footprinting (Figure 3). This immediately highlighted several differences by comparison with chemical probing. Whereas hydroxyl radical footprinting placed the boundary of the replicating enzyme at position −15 on the template (Metzger et al., 1993), DNase I footprinting suggests that a portion of RT in the vicinity of the RNase H active center is sufficiently close to the substrate to render it DNase I-resistant. Protection of template DNA to position −23 was evident in each replication complex, differing only in that it was displaced by the register of DNA synthesis. Within this footprint, increased reactivity to DNase I around position −20 was evident. Although not apparent in the autoradiogram of Figure 3, longer exposure also indicated mild reactivity at template base −14. Previous data with equivalent replication complexes (Metzger et al., 1993) indicated that template position −15 was hyperaccessible to chemical cleavage. Enhanced accessibility at different positions on the template implies that the chemical and enzymatic probing reagents address different conformational features or local surface topography of the RT-bound template. The notion that altered DNase I sensitivity arose through translocation was eliminated by analyzing substrate whose primer was preextended by a single base (i.e., a 37-mer primer), where a hydrolysis profile identical to the +1 replication complex was achieved (data not shown).

The template extremity of the polymerizing enzyme is most evident from a +10 complex. Since primer in this case is extended to the immediate vicinity of the hairpin (Figure 1A), an expanded DNase I hydrolysis profile is evident in the absence of RT. However, the majority of this is DNase I-resistant in a +10 complex, positioning the boundary of replicating RT six to seven bases beyond the chain-terminating ddNTP (i.e., template base +7). This profile of template DNA also contrasts with chemical footprinting data (Metzger et al., 1993), which indicated protection only three bases ahead of the point of chain termination. If RT of a +10 replication complex proceeds into the hairpin, enzyme arrested at position +4 should be located approximately six bases behind this, i.e., immediately preceding the duplex structure. This notion was observed experimentally, where virtually the entire hairpin remains DNase I-sensitive. Although the difference is minimal, enzyme of a +1 complex is located slightly behind that arrested at +4. The collective data of Figure 3 indicate that replicating HIV-1 RT covers approximately 30 bases of template (positions −23 to +7), independent of the point at which DNA synthesis is halted. The window of accessible bases between positions −8 and −11 evident from chemical footprinting (Metzger et al., 1993) is absent in our DNase I footprints. This may reflect accessibility of substrate between the catalytic centers to the small chemical probing reagent (i.e., the hydroxyl radical), while a substantially larger enzymatic probe fails to penetrate the nucleic acid binding cleft. The proximity of the p66 thumb subdomain to the substrate and its proposed role in template-primer binding (Nanni et al., 1993; see also the accompanying paper of Patel et al.) may be sufficient to exclude the enzymatic probe.

S1 Footprinting of a +4 Complex. To verify the downstream extremity of the replication complexes, reactivity of template sequences in a +4 complex to nuclease S1 was determined (Figure 4). A control digest of substrate containing a 3′ end-labeled template DNA, within which primer was extended by a single base (lane +1, −RT), indicates S1 susceptibility 37−47 bases from the radiolabel. This pattern represents single-stranded DNA between the template-primer and the hairpin structure of Figure 1A. The hydrolysis profile is shortened by three bases (i.e., revealing nucleotides 40−47) when the primer is extended to position +4 (lane +4, −RT). This pattern reflects a reduction in single-stranded substrate as a consequence of limited polymerization. However, when S1 susceptibility of the +4 replication complex was determined in the presence of p66/p51 HIV-1 RT, six additional template bases were shielded, restricting hydrolysis to positions 46 and 47 (lane +4, +RT). Locating the extremity of a +4 complex six bases ahead of the point of chain termination (i.e., template base +5) places it in the immediate vicinity of the hairpin. This is in good agreement

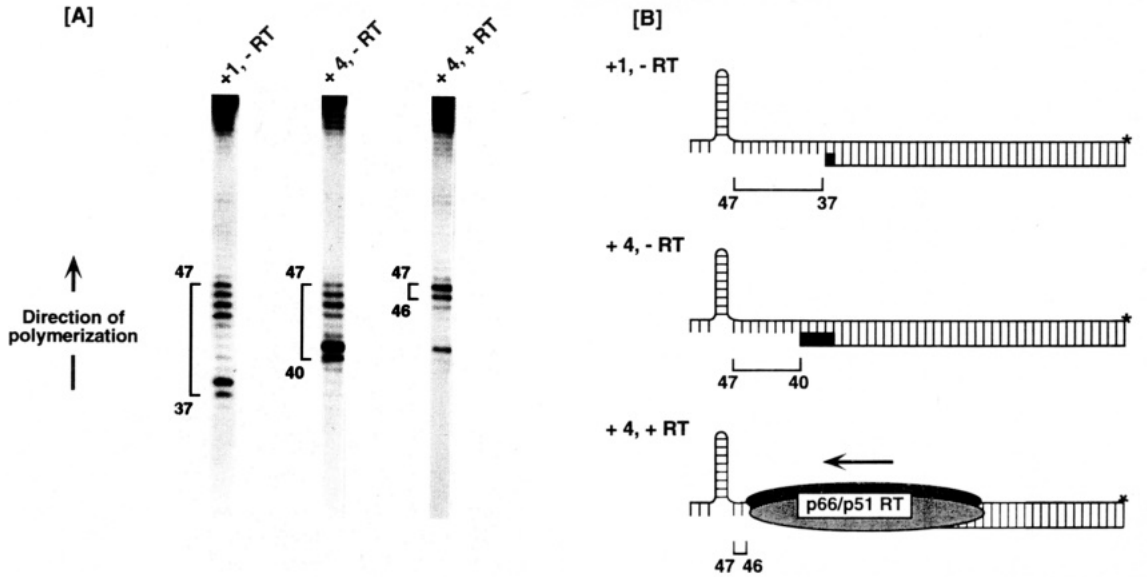

FIGURE 4: (A) Susceptibility of a +4 replication complex containing wild-type HIV-1 RT to digestion with the single-strand-specific nuclease S1. In this experiment, the DNA template was labeled at its 3′ terminus with ^{32}P. For comparative purposes, the hydrolysis profiles of template DNA containing a primer extended by 1 or 4 bases are included. Lane designation +4, +RT represents the hydrolysis profile of the +4 replication complex. (B) Schematic representation of the S1 digestion profile of a +4 replication complex. Filled boxes represent bases added to the primer for control +1 and +4 reactions. The base notation for each complex represents sequences accessible to S1 and is relative to the 3′ label on the template.

with the DNase I footprinting data of Figure 3, which places the extremity of a +4 complex possibly as far as position +7.

Replication Complexes Catalyzed by RNase H-Deficient RT. DNase I sensitivity of replication complexes containing the RNase H$^-$ mutant p66$^{E\rightarrow Q}$/p51 (Schatz et al., 1989) was also evaluated. Since enzymatic footprinting with wild-type p66/p51 RT indicated that a substantially larger portion of template was covered by the replicating enzyme, we asked if a mutation eliminating RNase H activity altered interactions at either extremity of the complex. Our rationale was based on previous *in vitro* mutagenesis studies, indicating that alterations in the RNase H domain are in many cases detrimental to polymerase function (Prasad, 1993).

The results of these experiments are presented in Figure 5, indicating that the hydrolysis profiles of +1, +4, and +10 complexes containing RNase H-deficient RT are indistinguishable from those of the wild-type enzyme. Accessibility of template and primer at position −20 to DNase I was preserved in all complexes, in contrast to chemical probing, which highlighted hyperaccessibility with only wild-type RT (Metzger et al., 1993). While the reason for altered accessibility to DNase I digestion at position −20 is not immediately clear, this event appears unrelated to reactivity of template bases to chemical cleavage.

Interaction of HIV-1 RT with Primer DNA. The interaction of RT with primer sequences is more easily assessed by DNase I footprinting, since duplex DNA is present under all conditions. Figure 6 illustrates the hydrolysis profiles of +1 and +4 complexes containing the wild-type (Figure 6A) or RNase H-deficient enzyme (Figure 6B). As might have been predicted from the combined data of Figures 3 and 5, the two enzymes bind DNA in a similar manner. In all replication complexes, primer DNA up to position −25 is resistant to DNase I digestion, indicating asymmetric distribution of RT on the template and primer. All complexes exhibited DNase I accessibility around positions −20, the position of which again moved in concert with the register of DNA synthesis. This feature of the DNA primer also

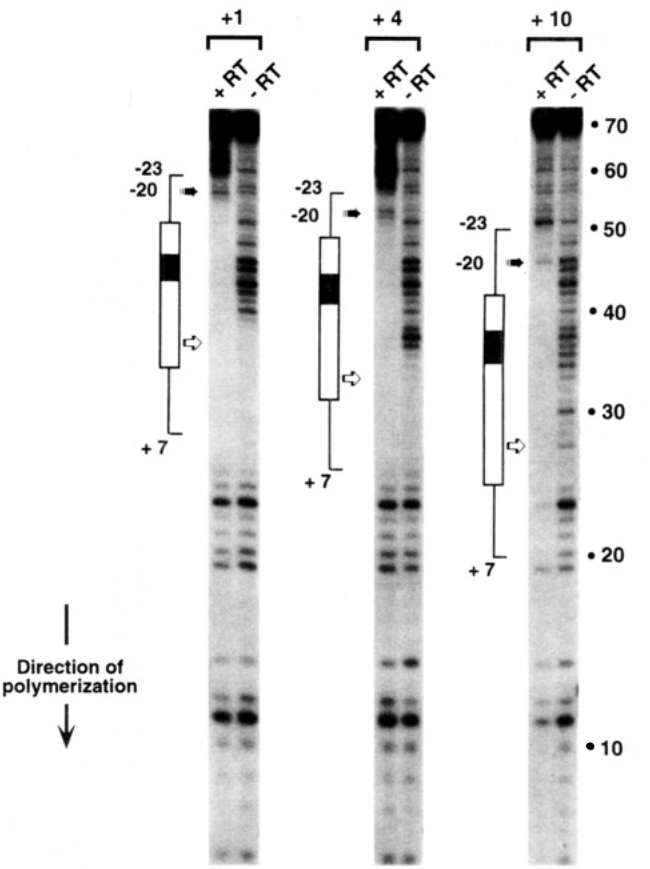

FIGURE 5: Partial DNase I digestion profiles of +1, +4, and +10 replication complexes containing RNase H-deficient HIV-1 RT p66$^{E\rightarrow Q}$/p51. Lane notations, template labeling, etc. are identical to those in the legend to Figure 3.

contrasts with chemical probing experiments (Metzger et al., 1993), where enhanced hydroxyl radical cleavage at position −15 was restricted to the *template*. Thus, as proposed in the previous section, enzymatic and chemical probing reagents address different features of substrate in and around the RNase H active center.

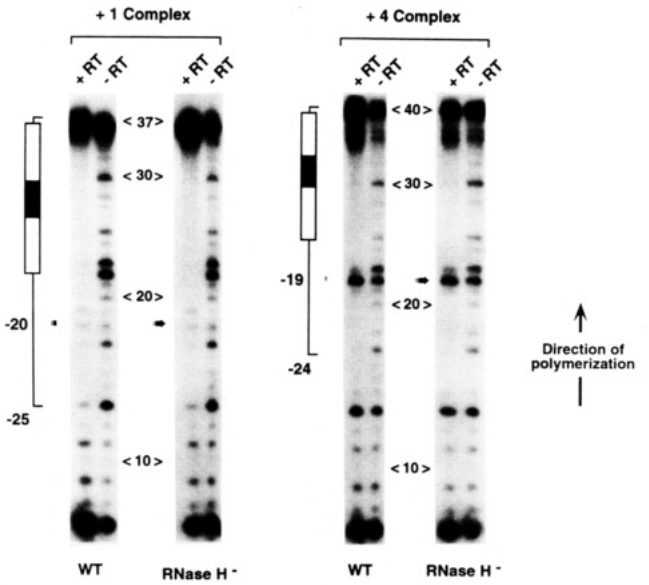

FIGURE 6: Partial DNase I digestion profiles of primer DNA containing wild-type and RNase H-deficient HIV-1 RT. The +1 and +4 extension reactions were performed with substrate containing a 5′ ^{32}P-labeled primer. Material migrating at position 36 in +4 extension reactions represents trace amounts of unextended primer, which does not make a significant contribution to the overall footprint. Definition of the HIV RT binding site is in accordance with the legend to Figures 3 and 5.

DISCUSSION

We have mapped the results of enzymatic and chemical footprinting experiments to DNA in the crystal structure of an HIV-1 RT template-primer cocrystal (Arnold et al., 1992; Jacobo-Molina et al., 1993). Since the original double-stranded substrate in the crystal structure does not extend beyond the DNA polymerase and RNase H catalytic centers, we modeled a single-stranded substrate extension at the polymerase domain and lengthened the duplex region at the RNase H domain to visualize potential protein−nucleic acid interactions adjacent to the catalytic centers. From this the model of replicating HIV-1 RT depicted in Figure 7 was generated.

The window of accessible bases determined by hydroxyl radical footprinting (Metzger et al., 1993) maps to an exposed region of DNA in the crystal structure just beyond the p66 thumb subdomain (dark blue region of Figure 7, upper). Metzger et al. (1993) suggested that the DNA substrate is more likely A- than B-form, which places this window on one side of the DNA rather than completely around it (which would imply that some accessible regions may occur at an interface between the enzyme and substrate). It was particularly interesting to map these results to DNA in the cocrystal, since this adopts A- and B-form characteristics (Jacobo-Molina et al., 1993). The A-form region corresponds to the first five to seven base pairs at the DNA polymerase domain (positions −1 to −5/−7), followed by a bend, and the remainder of the nucleic acid (positiosn −10 to −18) is primarily B-form. The portion of substrate accessible to iron−EDTA-generated hydroxyl radicals maps to the beginning of the B-form region (between positions −6 to −9 on the primer and −8 to −11 on the template), just beyond a close interaction with the p66 thumb subdomain. However, although the window resides in a B-form region, it maps to only one side of the DNA and is exposed to solvent, demonstrating that the chemical footprinting and structural data are in agreement.

The combined enzymatic footprinting data of this paper provide an expanded picture for replicating HIV-1 RT. When compared with X-ray crystallography (Kohlstaedt et al., 1992; Arnold et al., 1992; Jacobo-Molina et al., 1993) and studies on coupled polymerase/RNase H reactions (Wöhrl & Moelling, 1992; Furfine & Reardon, 1991; Peliska & Benkovic, 1993), the small hydroxyl radical appears to provide information exclusively on nucleic acid filling the cleft provided by the DNA polymerase and RNase H catalytic centers of RT. The agreement between enzymatic footprinting and the structure of the enzyme/substrate complex suggests that the "umbrella" of replicating HIV-1 RT covers template sequences +7 to −23, while its catalytic centers, or "heart", embrace 18 bases between +3 and −15. It is unnecessary to invoke steric effects for the enlarged footprint, since modeling of enzyme/nucleic acid interactions [Figure 7 and Hermann et al. (1994)] is consistent with protection extending to the borders observed in the footprinting experiments described here.

Given this scenario, what features of HIV-1 RT might contribute to inaccessibility of template DNA as far forward as position +7? Figure 7 argues that the $\beta3-\beta4$ hairpin of the p66 finger subdomain confers protection of the single-stranded template approximately seven bases upstream of the polymerase catalytic center. In support of our model, Boyer et al. (1994) have demonstrated that the p66 finger subdomain interacts with a segment of the template overhang three to six bases beyond the DNA polymerase catalytic center. The experiments of this paper, together with the data of Boyer et al. (1994), suggest that the $\beta3-\beta4$ hairpin shields the single-stranded template between positions +3 to +6/+7. A potential role for the p66 finger subdomain may be to "thread" the template in an orderly fashion into the polymerase active site. The efficiency with which +19 and complete primer extension reactions are achieved (Figure 2) indicates that the enzyme has little difficulty in resolving short hairpin structures. An analogous RNA-dependent DNA polymerase system with extensive intramolecular base pairing through the template is also efficiently transcribed by HIV-1 RT (Ghosh et al., 1995). Thus, the p66 finger subdomain could function (either alone or in concert with another subdomain) to disrupt duplex areas of the template prior to its delivery at the polymerase catalytic center. Although the hairpin structure in our substrate was fortuitous, it reflects situations encountered during minus- and plus-strand retroviral DNA synthesis. Several investigators have indicated that the 5′ end of the viral RNA genome adopts a highly ordered structure (Cobrinik et al., 1991; Berkhout & Schonveld, 1993; Baudin et al., 1993) which must be resolved for efficient synthesis of strong-stop minus-strand DNA from the primer binding site. Furthermore, displacement synthesis is a feature of plus-strand synthesis (Boone & Skalka, 1993), again mandating that duplex structures are resolved by the translocating enzyme. The observation of Whiting and Champoux (1994) that MLV RT supports plus-strand displacement synthesis over at least 1000 nucleotides in the absence of accessory viral proteins makes a strong case for the involvement of an RT subdomain in resolving intra- and intermolecular duplexes.

In keeping with the enzyme/DNA cocrystal, Figure 7 indicates that substrate is protected by p66/p51 RT across the RNase H domain. Previous chemical probing indicated position −15 as the extremity of the RNase H catalytic center. However, Nanni et al. (1993) have proposed that

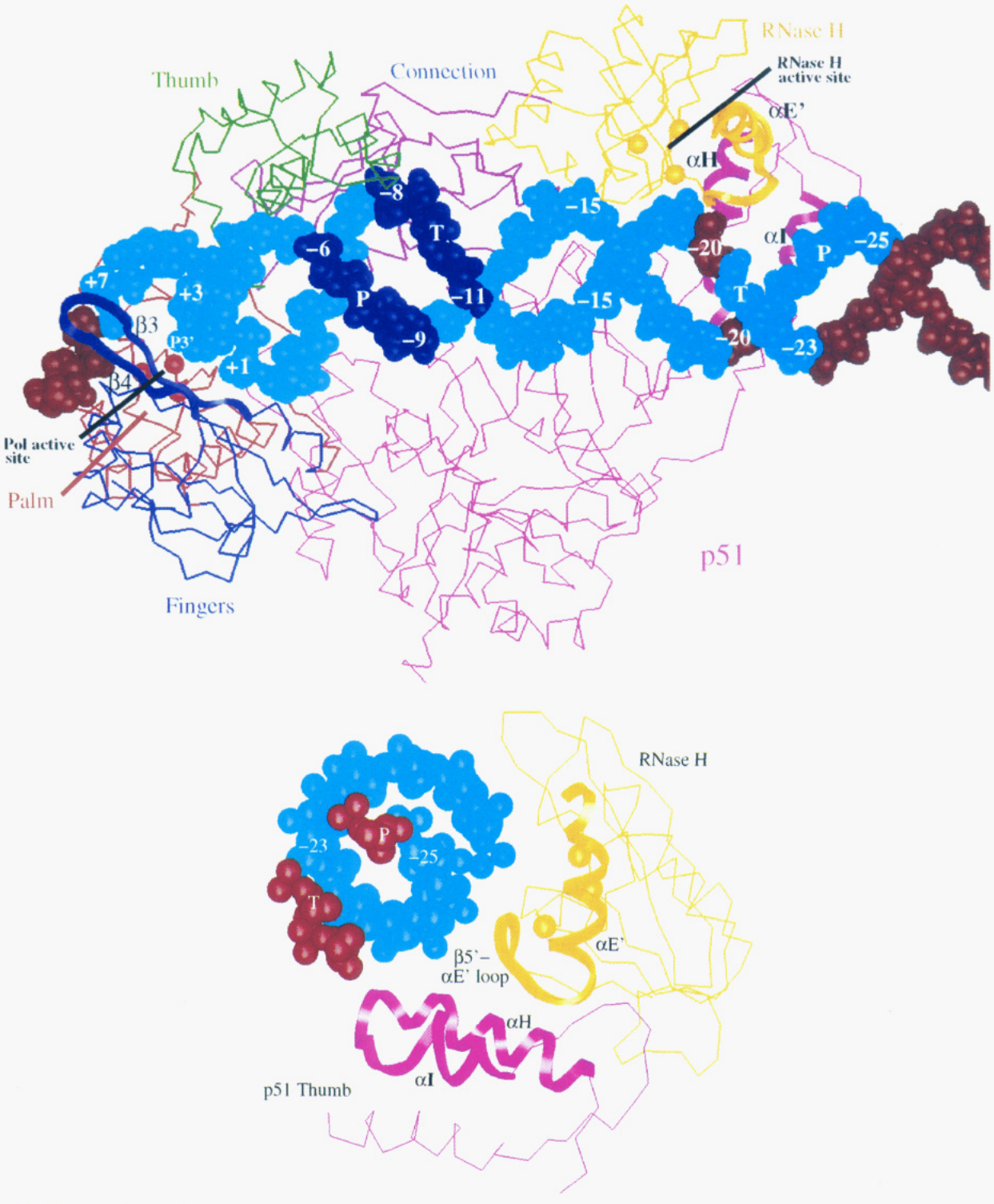

FIGURE 7: Transposition of footprinting data for replicating p66/p51 HIV-1 RT onto substrate bound by the enzyme in the crystal structure (Arnold et al., 1992; Jacobo-Molina et al., 1993). (Upper) The single-stranded template overhang at the polymerase domain and the duplex portion beyond −18 at the RNase H domain were modeled as described in Materials and Methods. p66 RT is color-coded by subdomain, i.e., fingers in blue, palm in red, thumb in green, connection in purple, and RNase H in yellow. The p51 subunit is colored violet red. Essential active site residues are represented as space-filled Cα atoms to aid their location. Sugar phosphate backbone atoms of the template-primer are depicted as space-filling spheres and colored on the basis of their accessibility to enzymatic and chemical probing agents. Red and cyan regions correspond to accessible and protected regions, respectively, determined by the combined DNase I and S1 footprinting data of this paper. Protection at the leading edge of RT extends to position +7 on the template overhang and may be conferred by the $\beta3-\beta4$ hairpin of the p66 finger subdomain (highlighted as blue solid ribbons). The upstream extremity of the replication complex extends to position −23 on the template and −25 on the primer. Structural elements which might interact with DNA beyond the RNase H catalytic center are highlighted as solid ribbons, including αE′ and the $\beta5'-\alpha E'$ loop of the RNase H domain, in addition to αH and αI of the p51 thumb. Possible explanations for enhanced DNase accessibility at position −20 are discussed in the text. The dark blue regions of the primer (−6 to −9) and template (−8 to −11) comprise the "window of accessibility" observed by chemical probing (Metzger et al., 1993). (Lower) View along the DNA axis from the C-terminal RNase H domain, highlighting regions of HIV-1 RT beyond the active site residues (yellow spheres) which may bind the template-primer. Structural elements of the enzyme which potentially confer protection from DNase I include α helices H and I of the p51 thumb subdomain (depicted in violet red), in addition to αE′ and the $\beta5'-\alpha E'$ loop of the RNase H domain (depicted in yellow). Together, these elements constitute a "floor and wall" which may position RNA/DNA hybrids for template degradation by RNase H.

duplex DNA extended by five or six bp could interact with helices αH and αI of the p51 thumb, along the floor of the substrate binding cleft and just beyond the RNase H active center. Our modeling of a duplex extension supports this proposal. Protection beyond position −15 would be afforded by both the RNase H domain and the thumb subdomain of the p51 subunit. Asymmetric protection of primer and template DNA (to positions −25 and −23, respectively) may indicate that the primer lies closer to the replicating enzyme. Specifically, the primer beyond the RNase H catalytic center may be shielded by αA′ and the β5′−αE′ loop of the RNase H domain, as well as αH and αI of the p51 thumb (Figure 7, lower).

While DNase I footprinting reveals an expanded picture of replication complexes, it differs from chemical probing (Metzger et al., 1993) in that template and primer sequences around position −20 are accessible. This is common to wild-type *and* RNase H-deficient RT, whereas only replication complexes with the wild-type enzyme displayed hyperreactivity to hydroxyl radicals at template position −15 (Metzger et al., 1993). The current observations may indicate alterations in substrate conformation as it traverses the polymerase and RNase H active centers and finally exits from the enzyme. In the RT/DNA cocrystal of Jacobo-Molina et al. (1993), duplex substrate adopts A-form geometry in the DNA polymerase active center, while a more B-like configuration is evident at the RNase H active center. As constraints on duplex DNA departing the RNase H active center are gradually released, it would be expected to recover B configuration, raising the possibility that this transition is detected as increased susceptibility to DNase I digestion at position −20.

Finally, our data contradict work of Pelletier et al. (1994), who propose from the crystal structure of rat DNA polymerase β that HIV-1 RT is oriented on the template-primer with the bulk of the enzyme occupying the single-stranded template. In fact, we find the opposite; i.e., enzyme is positioned primarily over the template-primer duplex. In support of current models (Kohlstaedt et al., 1992; Jacobo-Molina et al., 1993), we have found that (a) wild-type MLV RT occupies the template-primer in a manner similar to the HIV enzyme and (b) elimination of the MLV RNase H domain reduces the amount of *template-primer duplex* protected from DNase I digestion by 12 bp (Wöhrl et al., 1995). Furthermore, stepwise incursions into α-helix E′ (the last structural element of the RNase H domain) of HIV-1 RT result in increased susceptibility to DNase I digestion around position −24 on the template and primer (Ghosh et al., 1995). Finally, p66 EIAV yields a DNase I footprint closely resembling those of its human and murine counterparts (J. S. Rausch and S. F. J. Le Grice, unpublished observations). Our observations with three different enzymes (and mutants thereof) provide compelling evidence that the disposition of *retroviral* RT on its template-primer is that suggested by Kohlstaedt et al. (1992) and demonstrated by Jacobo-Moline et al. (1993).

ACKNOWLEDGMENT

We thank Jianping Ding, Alfredo Jacobo-Molina, and Stephen H. Hughes for helpful discussion, in addition to their contribution in HIV-1 RT structure determination. The technical assistance of Kathryn J. Howard in RT purification and manuscript reading is gratefully acknowledged.

REFERENCES

Arnold, E., Jacobo-Molina, A., Nanni, R. G., Williams, R. L., Lu, X., Ding, J., Clark, A. D., Zhang, A., Ferris, A. L., Clark, P., Hizi, A., & Hughes, S. H. (1992) *Nature 357*, 85−89.

Baudin, F., Marquet, R., Isel, C., Darlix, J. L., Ehresmann, B., & Ehresmann, C. (1993) *J. Mol. Biol. 229*, 383−397.

Berkhout, B., & Schonveld, I. (1993) *Nucleic Acids Res. 21*, 1171−1178.

Boone, L. R., & Skalka, A. M. (1993) in *Reverse Transcriptase* (Skalka, A. M., & Goff, S. P., Eds.) pp 119−133, Cold Spring Harbor Monograph Series, Cold Spring Harbor, NY.

Boyer, P. A., Tantillo, C., Jacobo-Molina,. A., Nanni, R. G., Ding, J., Arnold, E., & Hughes, S. H. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 4882−4886.

Cobrinik, D., Aiyar, A., Ge, Z., Katzman, M., Huang, H., & Leis, J. (1991) *J. Virol. 65*, 3864−3872.

Emerson, B. M., Lewis, C. D., & Felsenfeld, G. (1985) *Cell 41*, 21−30.

Furfine, E. S., & Reardon, J. E. (1991) *J. Biol. Chem. 266*, 406−412.

Ghosh, M., Howard, K. J., Cameron, C. E., Benkovic, S. J., Hughes, S. H., & Le Grice, S. F. J. (1995) *J. Biol. Chem.* (in press).

Hermann, T., Meier, T., Götte, M., & Heumann, H. (1994) *Nucleic Acids Res. 22*, 4625-4633.

Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H., & Arnold, E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6320−6324.

Johnson, A. D., Meyer, B. J., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5061−5065.

Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjelgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.

Kohlstaedt, L. A., Wang, J., Friedman, M., Rice, P. A., & Steitz, T. A. (1992) *Science 256*, 1783−1790.

Klarmann, G. J., Schauber, C. A., & Preston, B. D. (1993) *J. Biol. Chem. 268*, 9793−9802.

Lederer, H., Schatz, O., May, R., Crespi, H., Darlix, J. L., Le Grice, S. F. J., & Heumann, H. (1992) *EMBO J. 11*, 1131−1139.

Le Grice, S. F. J., & Grüninger-Leitch, F. (1990) *Eur. J. Biochem. 187*, 307−314.

Le Grice, S. F. J., Matzura, H., Marcoli, R., Iida, S., & Bickle, T. A. (1982) *J. Bacteriol. 150*, 312−318.

Metzger, W., Hermann, T., Schatz, O., Le Grice, S. F. J., & Heumann, H. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5909−5913.

Nanni, R. G., Ding, J., Jacobo-Molina, A., Hughes, S. H., & Arnold, E. (1993) *Perspect. Drug Discovery Des. 1*, 129−150.

Peliska, J. A., & Benkovic, S. J. (1993) *Science 258*, 1112−1118.

Plumb, M. A., Nicholas, R. H., Wright, C. A., & Goodwin, G. (1985) *Nucleic Acids Res. 113*, 4047−4065.

Prasad, V. R. (1993) in *Reverse Transcriptase* (Skalka, A. M., & Goff, S. P., Eds.) pp 135−162, Cold Spring Harbor Monograph Series, Cold Spring Harbor, NY.

Sasse-Dwight, S., & Gralla, J. D. (1989) *J. Biol. Chem. 264*, 8074−8081.

Schatz, D., Leberman, R., & Eckstein, F. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6132−6136.

Schatz, O., Cromme, F., Grüninger-Leitch, F., & Le Grice, S. F. J. (1989) *FEBS Lett. 257*, 311−314.

Schmitz, A., & Galas, D. (1979) *Nucleic Acids Res. 6*, 111−137.

Siebenlist, U., Simpson, R. B., & Gilbert, W. (1980) *Cell 20*, 269−281.

Whiting, S. H., & Champoux, J. J. (1994) *J. Virol. 68*, 4747−4758.

Wöhrl, B. M., & Moelling, K. (1990) *Biochemistry 29*, 10141−10147.

Wöhrl, B. M., Georgiadis, M. M., Telesnitskt, A., Hendrickson, W. A., & Le Grice, S.F. J. (1995) *Science 267*, 96−99.

BI942280Z