Probing Contacts between the Ribonuclease H Domain of HIV-1 Reverse Transcriptase and Nucleic Acid by Site-specific Photocross-linking

Received for publication, December 6, 1999, and in revised form, March 9, 2000
Published, JBC Papers in Press, March 13, 2000, DOI 10.1074/jbc.M909808199

Jason W. Rausch‡, B. K. Sathyanarayana‡, Marion K. Bona¶, and Stuart F. J. Le Grice§¶

From the ‡HIV Drug Resistance Program, ¶Science Applications International Corporation, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21072 and the §Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-4255

Cys^{38} and Cys^{230} of p66/p51 human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT) can be converted to Ser without affecting enzyme function. We have exploited this feature to construct and purify “monocysteine” RT derivatives for site-specific modification, with the photoactivable cross-linking agent, p-azidophenacyl bromide. Acylation of a unique cysteine residue introduced at the extreme C terminus of the p66 subunit (C^{561}) with an azidophenacyl group allowed us to probe contacts between residues C-terminal to α-helix E’ of the RNase H domain and structurally divergent nucleic acid duplexes. In a binary complex of RT and template-primer, we demonstrate efficient cross-linking to primer nucleotides 21 to 24/25, and template nucleotides 18 to 21. Cross-linking specificity was confirmed by an analogous evaluation following limited primer extension, where the profile is displaced by the register of DNA synthesis. Finally, contact with a DNA primer hybridized to an isogenic RNA or DNA template indicates subtle alterations in cross-linking specificity, suggesting differences in nucleic acid geometry between duplex DNA and RNA/DNA hybrids at the RNase H domain. These data exemplify how site-specific acylation of HIV-1 RT can be used to provide high resolution structural data to complement crystallographic studies.

Bioconjugation represents site-specific tethering of two molecules to generate a novel complex displaying the combined properties of its individual components. Although in its infancy, this approach is emerging as a powerful complement to high resolution crystallographic and spectroscopic methods in providing structural information on protein-nucleic acid complexes. Since Ebright et al. (1) first reported attachment of the chemical nucleos 5-iodoacetyl)-1,10-phenanthroline to Cys^{178} of the Escherichia coli catalolite gene activator protein (CAP), photoactivable (2-4), fluorescent (5-7), nucleolytic (8-10), and proteolytic agents (11-13) have been site specifically tethered to probe the protein and nucleic acid components of nucleoprotein complexes. More recently, this strategy has been applied to study intein and reverse transcriptase (RT)¹ of human immunodeficiency virus-type 1 (HIV) via thiol tethering of the photocross-linker p-azidophenacyl to nucleic acid (14, 15). The complementarity between crystallographic and bioconjugate techniques has also been elegantly demonstrated by Huang et al. (16), who have reported a high resolution structure of HIV-1 RT containing both duplex DNA and the incoming deoxynucleoside triphosphate. This was achieved by linkage of DNA containing a single tethered thiol to p66/p51 HIV-1 RT containing a site-specific cysteine mutation at residue 258 of the p66 thumb. In addition to defining the structure of a covalently trapped catalytic complex, these authors also demonstrated that the single-stranded template ahead of the polymerase catalytic center is not colinear with the duplex, but rather bends away and extends across the face of the fingers subdomain.

In the course of our structure/function studies, several HIV-1 RT mutants have been constructed whose phenotype is not readily explained by currently available crystal structures (16–21). For example, alteration of p66 residues 227 (Phe → Ala) or 234 (Leu → Ala) confer increased fidelity of DNA synthesis (22) and loss of dimer function (23), respectively. Moreover, deleting 13 amino acids from the p51 C terminus (p66/p51Δ13 RT) selectively eliminates polymerization-independent RNase H activity, despite leaving the RNase H domain of p66 intact (24). Finally, HIV-1 RT harboring a mutation at residue 232 (Tyr → Ala) retains exclusively polymerization-independent RNase H phenotype, which theoretically stems from polymerization-dependent hydrolysis (23). In an attempt to better understand the consequences of these amino acid substitutions, we have elected to construct a series of “mono-Cys” HIV-1 RT variants for site-specific acylation or disulfide linkage of bioconjugates to investigate the manner in which subunit and subdomain function has been compromised by mutagenesis.

Although structures exist for the isolated RNase H domain (25) and the parental p66/p51 HIV-1 heterodimer (16–21), the extreme p66 C terminus, i.e. residues C-terminal to α-helix E’, is not well defined. Nuclease footprinting of several retroviral replication complexes (26-28) suggests that up to 24 bp of duplex DNA can be accommodated by the combined DNA polymerase and RNase H domains, although these estimates must take into account the size of the probing agents. Lastly, despite the availability of high resolution structures for enzyme complexed with duplex DNA (16, 19), the manner in which the enzyme accommodates RNA/DNA hybrids and A-form duplex RNA remains to be elucidated. Presently, the only information available on accommodation of RNA/DNA hybrids has been obtained through “localized” hydroxyl radical footprinting, using enzymes where Fe^{2+} is substituted for Mg^{2+} at the RNase H catalytic center (29). Despite the novelty of this approach, a

¹ The abbreviations used are: RT, reverse transcriptase; HIV-1, human immunodeficiency virus type 1; bp, base pair(s); nt, nucleotide.

This paper is available on line at http://www.jbc.org
FIG. 1. Strategy for construction of Cys-less (p66C/His-p51C) and mono-Cys (p66C.C561/His-p51C) HIV-1 RT heterodimers. Both p66 subunits (upper panel) contain Cys38 → Ser38 and Cys280 → Ser280 substitutions, while p66RT−C.C561 was modified by introduction of Cys as the last residue (561) of the RNase H domain. p66 subunits were reconstituted in vitro with a Cys-less, poly-histidine-extended p51 subunit (middle panel) and purified by metal chelate and ion exchange chromatography, generating the purified by metal chelate and ion-exchange chromatography, generating the Cys-less heterodimer p66−/His-p51C or its mono-Cys counterpart, p66C.C561/His-p51C, containing a unique Cys on the RNase H domain (lower panel). A similar strategy was used to construct the RNase H-deficient monocysteine derivative p66E478Q.C561/His-p51C.

potential limitation is that it cannot reveal information on mutants altered in metal binding, since these by default will not bind Fe2+.

In order to address these issues, we have substituted Cys38 and Cys280 of wild type and RNase H-deficient HIV-1 RT with Ser, and thereafter introduced a unique Cys at the extreme p66 C terminus of RNase H-proficient (p66C.C561/p51C) and deficient (p66E478Q.C.C561/p51C) HIV-1 RT. We show here that these alterations are achieved without compromising enzyme function. The resulting “mono-Cys” RTs were then site specifically acylated with p-azidophenacyl to determine the spatial proximity of template and primer nucleotides in a variety of replication complexes. A unique cross-linking pattern to template and primer nucleotides, and moving in concert with the replication enzyme, is obtained with duplex DNA. Moreover, subtle alterations in the cross-linking pattern are observed when this substrate is replaced by an RNA-DNA hybrid.

MATERIALS AND METHODS

Construction of HIV-1 RT p66 and p51 Mutants—Plasmids pRT and p6HRT51 contain sequences encoding p66 and poly-histidine-tagged p51, respectively, of HIV-1 RT (30). Cys → Ser substitutions were introduced at position 38 and 280 of both polypeptides to ensure the absence of native cysteines in subsequent gene products. The polymerase chain reaction was used to amplify adjacent, overlapping regions of the RT gene using oligonucleotides encoding the Cys38 → Ser mutation. Fragments were purified and a portion of each added to a second polymerase chain reaction to merge the products into a single ~800-bp DNA. This was inserted into pRT and p6HRT51 as a BanHI-EcoRV restriction fragment to create pRT−C38S and p6HRT51C38S. A similar procedure was used to introduce the Cys280 → Ser mutation into these vectors as an EcoRV-HindIII restriction fragment, creating pRTC−C and p6HRT51C. pRTC−C.C561 was generated using a mutagenic oligonucleotide to introduce a unique cysteine at the C terminus of the p66 RT gene (C561). pRT−C38S (31) served as an alternative polymerase chain reaction template to create the RNase H-deficient variant pRTC−E478Q.C.C561. All clones were verified via restriction site analysis and DNA sequencing. p66/p51 heterodimers were reconstituted from strains expressing the modified subunits by previously described protocols (30). Alternatively, p66 expression cassettes were inserted into p6HRT51C− downstream of p51 to create co-expression vectors. A polylinker containing a Sall restriction site was introduced at a unique HindIII site 3’ to the C terminus of the p51 coding region, and p66 expression cassettes inserted as Xhol-Sall restriction fragments. The protocol for RT construction is summarized in Fig. 1.

Enzyme Preparation and Purification—Purification of recombinant RT variants was achieved using a combination of metal chelate (Ni2+-NTA-Sepharose) and ion-exchange (S-Sepharose) chromatography as described previously (32). Enzyme purity and subunit stoichiometry was evaluated by SDS-polyacrylamide gel electrophoresis followed by Coomassie Blue staining. Purified enzymes were stored at −20 °C in 50 mM Tris-HCl (pH 7.0), 1 mM EDTA, 25 mM NaCl, 5 mM dithiothreitol, and 50% glycerol.

Evaluation of Enzyme Activities—DNA-dependent DNA polymerase activity was evaluated on a 71 nt DNA template annealed to a 5’- [32P]-end-labeled 36 nt DNA primer as described previously (26). Enzyme (10 nM) and template/primer (20 nM) were incubated at 37 °C in buffer containing 10 mM Tris-HCl (pH 8.0), 6 mM MgCl2, and 80 mM NaCl. DNA synthesis was initiated by adding dNTPs to a final concentration of 50 nM, and terminated by mixing with urea-based gel loading buffer at the times indicated in the text.

RNase H activity was evaluated on a 5’-end-labeled 90 nt RNA annealed to a 36 nt DNA primer (28). Enzyme (50 nM) and template/primer (20 nM) were incubated at 37 °C for 1 min in buffer containing 10 mM Tris-HCl (pH 8.0), 80 mM NaCl. Hydrolysis was initiated by addition of MgCl2 to a final concentration of 6 mM, and terminated as described above. In both cases, products were fractionated by high voltage electrophoresis through 10% (w/v) polyacrylamide gels containing 7 M urea in Tris borate/EDTA buffer and visualized by autoradiography of the dried gel.

Conjugation of 4-Azidophenacyl to RT Variants—Each enzyme was dialyzed at 4 °C overnight against 50 mM Tris-HCl (pH 7.0), 25 mM NaCl, 1 mM EDTA, 5% (v/v) glycerol to remove dithiothreitol and reduce
FIG. 2. A, DNA polymerase activity of mono-Cys and azidophenacyl-derivatized HIV-1 mutants. All enzymes were analyzed on a 71-nt template/36-nt primer (P) as described previously (40). For all enzymes, DNA synthesis was evaluated after 15 s (lanes 1), 1 min (lanes 2), 3 min (lanes 3), and 10 min (lanes 4). The notation around P+15 represents an intramolecular hairpin present on the DNA template which transiently pauses the replication machinery (B) RNase H activity of mono-Cys HIV-1 mutants and their azidophenacyl derivatives. Substrate was a radiolabeled 90-nt RNA template hybridized to a 36-nt DNA primer (28). Lane and panel designations, as well as incubation times, are similar to those described for evaluation of DNA polymerase activity.

glycerol content. Immediately prior to conjugation, 4-azidophenacyl bromide (Sigma) was dissolved in molecular biology grade methanol (Fisher Scientific) to a concentration of 40 mM, and diluted to 1 mM in 25 mM Tris-HCl (pH 8.0), 50 mM NaCl, 5% (v/v) glycerol. In the same buffer, enzyme and 4-azidophenacyl bromide (4 and 200 μM, respectively) were incubated for 10 min at 37 °C, then transferred immediately to microdialysis tubes (Tubodialyzer, Research Products International Corp.) for overnight dialysis (4 °C). Derivatization efficiency was evaluated by determining the concentration of free sulfhydryl groups before and after exposure to 4-azidophenacyl bromide using Molecular Probes’ Thiol and Sulfide Quantitation Kit. Derivatized enzymes were stored in aliquots at −80 °C. Throughout this procedure care was taken to minimize exposure of 4-azidophenacyl bromide to conjugated enzyme.

Photocross-linking of Conjugated Enzymes to Nucleic Acid—The protocol for photocross-linking and piperidine cleavage of cross-linked products was adapted from Dumoulin et al. (33, 34). 5-[32P]-End-labeled DNA-DNA and RNA-DNA duplexes described in the previous section were incubated with RT derivatives (−4 μM) in the dark at 37 °C for 20 min in 25 mM Tris-HCl (pH 8.0), 80 mM NaCl, 6 mM MgCl2. In some reactions, nucleotide mixtures containing a dideoxynucleoside triphosphate were included to permit limited primer extension. Reactions were then UV irradiated using a 23 W Fischer Biotech hand-held UV lamp (312 nm) at a distance of 10 cm for 2 min, and denatured by incubation in 1 M Tris-HCl (pH 8.0), 1% SDS at 60 °C for 10 min. Cross-linked complexes were isolated by extraction with Tis-saturated phenol followed by ethanol precipitation of the organic phase. Pellets were washed twice with 70% ethanol, vacuum-dried, and resuspended in 10% piperidine at 90 °C to cleave cross-linked DNA. Solutions were then lyophilized to dryness, resuspended in a urea-based gel-loading buffer, and fractionated on 8 or 15% denaturing polyacrylamide gels. Fragment sizes were determined by comparison to DNA sequencing ladders (35).

Modeling of the Polymerase and RNase H Domain of RT—Molecular modeling of RNA-DNA complexes was carried out using InsightII software (MSI Corp.) on a Silicon Graphics computer. All energy calculations were performed using the cff91 forcefield within the Discover module of InsightII running on a CRAY-J90 supercomputer at NCI-FCRDC. The x-ray crystal structure of HIV-RT of Huang et al. (16) was chosen as the starting structure for the modeling and the coordinates for 1rtd were downloaded from the Protein Data Bank. Seven nucleotide bases built separately in the B-form were attached to the DNA at the 5′-end and a steepest decent minimization with 1000 iterations was performed to achieve a continuous DNA helix void of any unfavorable stereochemical contacts. Only 2 bp on either side of the base pair at the linkage where the new residues were attached were included in the initial calculations while the rest of the DNA and the entire p66 RT polypeptide were constrained. Seven amino acid residues, GIRQKLC, were added to the RNase H C terminus. The first four were modeled based upon the available x-ray structure of HIV-RT structure of Jacobo-Molina et al. (19). The last three residues, ILC, were added individually, and steepest decent energy minimization involving 1000 iterations was performed for each added residue. For this part of the calculations the entire DNA and residues 1 to 540 of the p66 domain were held constrained. Once a stereochemically acceptable model of both the protein and DNA was obtained, a final energy minimization for the entire complex of DNA and p66 was performed using the steepest decent energymization, each involving 1000 iterations. When the modeled structure superimposed on the x-ray structure of Huang et al. (16), the overall root mean square deviation between the structures is −0.2 Å indicating that the modeled structure does not deviate significantly from the x-ray structure.

RESULTS

Enzymatic Activities of Monocysteine Derivatized RT—Prior to evaluating the extent to which the RNase H C terminus could be cross-linked to template and primer nucleotides, it was important to determine whether the modifications we introduced compromised enzyme function. Fig. 2A illustrates DNA-dependent DNA polymerase profiles obtained with a 71-nt template/36-nt primer used in several of our studies (26). A fortuitous feature of this substrate is the presence of an intramolecular hairpin on the template immediately ahead of the template-primer duplex, which evaluates the strand displacement capacity of RT. Although minimal synthesis is required to polymerize through this structure, it has been shown to halt progression of the purified p51 subunits of HIV-1 RT and its equine infectious anemia virus counterpart (36), as well as p66/p51 HIV-1 heterodimers harboring mutations in the p66 primer grip (23) following polymerization of 10–15 deoxynucleotides. Clearly, substituting Cys338 and Cys280 of p66 and p51 with...
Ser (Fig. 2A, left) does not seriously compromise enzyme activity. Although pausing at the template hairpin is apparent at early time points, this does not represent a major barrier to the replication machinery. The DNA synthesis pattern is likewise minimally affected when Cys-less RTs are modified via introduction of a novel Cys at the p66 C terminus (Fig. 2A, center). Note that at this stage, a mono-Cys variant of HIV-1 RT with the RNase H-inactivating mutation Glu478 → Gln (31) was also constructed and evaluated in order to investigate the interaction with RNA/DNA hybrids in the absence of hydrolysis. Similarly, data of Fig. 2A, right, indicates that wild type and RNase H-deficient RTs retain full activity following covalent tethering of azidophenacyl to Cys561 of the mono-Cys p66 subunit. Since colorimetric evaluation of the amount of free thiol indicated that derivatization was >90%, we can also eliminate the possibility that the DNA polymerase activity we observe reflects a non-derivatized population on enzymes.

Although derivatization of mono-Cys RTs was not expected to affect DNA polymerase activity, it was important to confirm that RNase H activity was likewise unimpaired, since minimal alteration of the C terminus of p66 (27) or p51 (24) can significantly impair function. RNase H activity was evaluated on a closely related RNA-DNA hybrid whose RNA template was uniquely 5'-end-labeled, the results of which are presented in Fig. 2B. As previously demonstrated (24, 27), initial hydrolysis of the 90-nt template, defined by the spatial separation of the DNA polymerase and RNase H catalytic centers (17–18 bp), yields a 71-nt product. With prolonged incubation, polymerization-independent hydrolysis from this site as far as template nucleotide −8 is characterized by accumulation of a 62-nt product. Within experimental error, the profiles presented in Fig. 2B indicate that Cys-less, mono-Cys, and azidophenacyl-derivatized enzymes retain full RNase H function. The obvious exception to this are the mono-Cys RNase H mutant and its derivatized counterpart. However, it is also important to note that the labeled RNA template undergoes no hydrolysis in the presence of these enzymes, indicating they are free of both ribonuclease A and E. coli RNase H activity. Taken together, the combined data of Fig. 2 demonstrate that derivatization of HIV-1 RT at the C terminus of the RNase H domain can be achieved without perturbing the dimer interface and catalytic centers of the enzyme.

**Cross-linking the RNase H Domain to Primer Nucleotides**—In Fig. 3, several RT variants were evaluated for the efficiency and specificity with which they could be cross-linked to the primer strand of duplex DNA used in the experiment reported in Fig. 2A. Complexes were irradiated at 312 nm to induce formation of azidophenacyl-mediated RT-DNA cross-links. Following phenol extraction and precipitation, cross-linked complexes were re-suspended and incubated in 10% piperidine for 30 min at 90°C to cleave the DNA specifically at the cross-linked nucleotides. Theoretically, therefore, the amount and size of cleavage products detected by autoradiography should reflect the extent and location of azidophenacyl-mediated cross-links.

As anticipated, irradiation and piperidine treatment does not result in primer hydrolysis in the absence of protein (Fig. 3, lane 1). Likewise, when binary complexes containing wild type HIV-1 RT (lane 2) or its azidophenacyl-exposed Cys-less counterpart (lane 3) are similarly treated, no cleavage products are observed. Only binary complexes containing azidophenacyl-derivatized RT (lane 4) or the RNase H-deficient variant (lane 5) yield discrete hydrolysis fragments indicative of site-specific cross-linking. Varying in length from 12 to 15 nt, these fragments position the extreme p66 C terminus in the immediate vicinity of primer nucleotides −21 to −24, which is in excellent agreement with our previous DNase I footprinting data (26–28). Finally, based upon these and other experiments in which DNA is cross-linked to RT but not cleaved (data not shown), we estimate the efficiency of azidophenacyl-mediated cross-linking to be 3–5%, which is consistent with previously published estimates.

The specificity of photocross-linking was further evaluated by supplementing binary complexes with deoxyribonucleoside/dideoxynucleoside triphosphate mixtures permitting limited primer extension (Fig. 4A). This approach was previously applied during chemical (37) and enzymatic footprinting studies of HIV-1 replication complexes (26, 28, 38) to illustrate that template and primer footprints were displaced in accordance with the register of DNA synthesis. The results of our analysis are presented in Fig. 4B. A clear cross-linking pattern was again obtained in the +4 replication complex, in this case spanning primer nucleotides −22 to −25. Since the DNA primer in this experiment is radiolabeled at its 5’ terminus, the extent of DNA synthesis could be simultaneously evaluated. Although unextended primer is evident, this virtual absence of shorter cross-linked species suggests this most likely represents excess single-stranded primer. Moreover, trace amounts of P+1, P+2, and P+3 primer extension products could also account for background products outside the major cross-linked species. Finally, a similar range of cleavage products (from positions −21 to −24) was observed upon piperidine treatment of +10 replication complexes. However, the extent of cleavage within this range varied considerably, with the greatest occurring at positions −21 and −24 (both dA) and the least at −22.
and −23 (both dC).

Cross-linking to Template Nucleotides—A similar approach was used to determine the template nucleotides in the immediate vicinity of the C terminus of the RNase H domain. In this case, the template strand of the template-primer duplex was 5'-end-labeled and the nucleoprotein complexes irradiated and treated piperidine as described above. However, such treatment of these cross-linked complexes failed to yield cleavage products detectable above background levels, suggesting that post cross-link chemistry required for breakdown of the nucleotide and cleavage of the sugar-phosphate backbone had not occurred. We speculated that the failure of this reaction might be pyrimidine specific, since (a) dC nucleotides in the primer cross-linking experiment (+4 complex) failed to yield strong cleavage products, and (b) template nucleotides −21 to −24 are all pyrimidines in the binary complex. Accordingly, we utilized an "anti-"template and primer, in which purine nucleotides were substituted for their pyrimidine complement (e.g. dA for dT), and vice versa, to generate the results of Fig. 5.

Template cross-linking is best represented in a +19 replication complex, spanning nucleotides −18 to −21. Although variations in the efficiency of cross-linking were observed in the binary, +4 and +10 complexes, the proximity of template nucleotides −18 to −21 to the C terminus of the RNase H domain was still preserved with the register of DNA synthesis. Since the DNA template was radiolabeled in these experiments, it was not possible to simultaneously follow the efficiency of primer extension. However, as was noted with primer footprints reported in Fig. 4, minor cross-linked species trailing the +4, +10, and +19 complexes most likely represent prematurely paused replication complexes. As a control, the extent of template cross-linking was also determined using a single template to which several synthetically prepared primers of varying lengths were hybridized. Under these conditions, a similar cross-linking pattern was observed (data not shown).

Cross-linking to Duplex DNA versus RNA/DNA Hybrids—While x-ray crystallography and chemical and enzymatic footprinting have successfully probed replication complexes containing duplex DNA, there is little information available on the manner in which the retroviral polymerase accommodates either duplex RNA or RNA/DNA hybrids, both of which are encountered during replication. Exceptions to this are the modeling studies of Isel et al. (39) and a recent report of Goette et al. (29), the latter of which exploited the observation that Mg2+ in the RNase H domain can be replaced by Fe2+ to permit "site-specific" hydroxyl radical-mediated cleavage. Despite its novelty, a drawback of the hydroxyl radical approach is that altering conserved residues participating in metal binding also leads to loss of Fe2+ binding. In view of this, we reasoned that site-specific cross-linking might be useful in studying the interaction of RT with conformationally distinct nucleic acid du-
only binary complexes were evaluated. At primer nucleotides 2 cross-linking of primer nucleotides in an RNA-DNA hybrid (lane a) was cross-linked to nucleotides 21/22 through 24/25 as demonstrated in Fig. 4B. However, when the same primer was hybridized to an RNA template, subtle and reproducible differences in its cross-linking pattern were evident. Accessibility of primer nucleotide 21/22 of the RNA-DNA hybrid to azidophenacyl was substantially reduced, while at the same time accessibility of nucleotides further removed from the primer 3 terminus increased. Using site-directed hydroxyl radical footprinting, Goette et al. (29) predicted that the lower pitch of an RNA/DNA hybrid permits accommodation of an additional base pair between the DNA polymerase and RNase H catalytic centers, which is consistent with our experimental observations.

**DISCUSSION**

In this study, we have exploited site-directed photocross-linking to determine the proximity of the C terminus of the HIV-1 RNase H domain to both duplex DNA and RNA/DNA hybrids. Prior to our work, two reports had been documented employing a related approach with the HIV-1 enzyme. The first of these (15) made use of azidophenacyl-derivatized DNA to follow photocross-linking to either the p66 or p51 subunits, while Cys-substituted RT derivatives, disulfide-linked to duplex DNA containing a single tethered thiol, have been successfully used to crystallize a trapped replication intermediate (16). To complement these, we provide here the first report involving tethering of a bioconjugate directly to a solvent-exposed cysteine of the HIV-1 enzyme as a tool to probe contacts with its nucleic acid substrate. Important prerequisites to our studies were that (i) substituting Cys38 and Cys280 of p66 and p51 with Ser (ii), introduction of Cys as residue 561 of p66, and (iii) derivatization of Cys561 with the photocross-linking agent could be accomplished without seriously impairing enzyme activity. The combined enzymatic analyses presented in Fig. 2 indicates this can indeed be accomplished, which also indirectly implies that the p66-p51 dimer interface has not been altered.

Since the cross-linking sites identified in our study lie outside the limits of the template-primer duplex in the current RT/DNA co-crystals (16, 19), it was necessary to model an additional 6 bp of duplex DNA upstream of the RNase H domain onto the structure provided by Huang et al. (16). We assumed that the additional portion of DNA would retain the same B-form geometry demonstrated for nucleic acid within and proximal to the RNase H catalytic center (16, 19), and added nucleotides matching those of our model substrate accordingly. In addition, since most RT structures do not extend as far as residue 560 of the RNase H domain, the equivalent approach was used to model Ile559, Leu560, and the “extrinsic” Cys561 onto the p66 C terminus of the structure reported by Jacobo-Molina et al. (19). Energy minimization of the modeled structures positions the extreme C terminus of p66 within the DNA major groove above the 21st base pair from the 3′ primer terminus (i.e. position −21). The global structure otherwise maintains a close resemblance to those previously published. The final model is depicted in Fig. 7.

The combined length of the cysteine side chain tethered to the azidophenacyl moiety is approximately 11 Å, so a freely rotating C-terminal residue would theoretically define a sphere of the same radius within which all nucleotide bases would be subject to cross-linking. This, of course, ignores steric considerations, which we believe influences our results in two important ways. First, as illustrated in Fig. 7B, template and primer nucleotides would be differently accessible within a sphere centered in the major groove above the base pair at position −21. Specifically, while the nucleotides constituting the base pair −21 would be equally susceptible to cross-linking, primer nucleotides located closer to the primer 3′ terminus would be shielded from the probe by their base paired counterparts, while template nucleotides would be shielded at positions further removed. This most likely explains the different patterns of cross-linking to the template and primer strands.

Second, the location of the probe within the DNA major groove may also explain the relative inefficiency of piperidine-mediated cleavage following cross-linking to pyrimidine residues. It has been theorized that acylation of a secondary or tertiary amine within a nucleotide base is required for scission of modified DNA under alkali conditions, and the spatial positioning of these groups are quite different in purines and pyrimidines. While the N7 secondary amines of adenine and guanine are easily accessible from the major groove, N1 and N3 of thymine and cytosine are shielded by the pyrimidine ring structure and its base pair counterpart, respectively. This notion also suggests that a probe positioned to access the minor groove would permit cleavage of both purines and pyrimidines. Studies to explore this possibility are being considered.

From our investigations, a “stabilizing” or “tracking” function for the RNase H C terminus through its location in the major groove of the nucleic acid duplex might be invoked. The results of our energy minimization of the RT-DNA structure
suggest that the presence of two positively charged residues, Arg 557 and Lys 558, could assist in positioning the RNase H C terminus in the major groove through an interaction with the phosphate backbone. The presence of positively charged residues at the C terminus of HIV-2 (-Gly 555-Ile-Arg-Gln-Val-Leu 560-), ELAV (-Lys 556-Ile-Lys-Glu-Glu-le 558-), FIV (-Lys 556-Leu-Cys-Gln-Thr-Met 558-), and SIV enzymes (-Gly 555-Phe-Arg-Gln-Val-Leu 560-) supports this hypothesis. Although our model is speculative, it is interesting to note that the RT mutant p66/p51 is speculative, it is interesting to note that the RT mutant p66/p51, which lacks residues Ser 553-Leu 560 of the RNase H domain, retains full DNA polymerase function, but supports exclusively polymerization-independent RNase H activity (27). When RNase H activity of wild type HIV-1 RT is evaluated during a single binding event, polymerization-independent hydrolysis is observed (23), suggesting this is followed by dissociation. Should a tracking function of the extreme RNase H C terminus be required to stabilize the association of RT to an RNA/DNA hybrid which has been “nicked” at position –17, the absence of residues 553–560 might induce misalignment and loss of polymerization-independent hydrolysis. The same scenario could also be invoked for the RT mutant p66/p51 Δ13 (40), which also retains exclusively polymerization-independent RNase H function although the RNase H domain remains intact. In this case, we propose that a p51 C-terminal truncation subtly re-aligns its thumb subdomain, upon which the p66 RNase H domain rests. As a consequence, increased flexibility of the RNase H domain may prevent location of its C terminus in the major groove of the RNA/DNA hybrid when re-binding to substrate which has been hydrolyzed at position –17. One means by which this postulate might be evaluated would be to investigate the cross-linking pattern with RNA/DNA hybrids which have been artificially nicked at template nucleotide –17. This could be achieved by hybridizing a single primer to two RNA template oligonucleotides such that the junction between the latter is positioned at –17. Should the enzymes retaining polymerization-dependent RNase H activity be incapable of re-binding in the appropriate manner, we would expect to lose the primer cross-linking pattern. Experiments of this nature are currently underway.

While bioconjugate approaches to studying HIV-1 integrase and RT have recently been documented (14, 15), these studies have involved tethering of azidophenacyl to the nucleic acid substrate containing a single phosphorothioate linkage. One potential disadvantage to this approach is that R and S stereoisomers of the α-S-dNTP are introduced into the oligonucleotide under synthesis. Unless these are separated the ensuing cross-linking pattern will reflect two spatially distinct populations of the chemical probe. In contrast, this ambiguity (or need for further purification) is eliminated when azidophenacyl is introduced into a monocy steine-derivatized enzyme, which we show here is biologically functional. Moreover, the availability of several high resolution structures for HIV-1 RT can be exploited to guide site-specific attachment of an increasing array of chemical probes and nucleases to other subdomains of this highly versatile enzyme to better understand both protein-protein and protein-nucleic acid interactions critical to the polymerization process.

Acknowledgments—We thank Kathryn Howard for assistance in early purification of monocy steine-derivatized enzymes. In addition, we acknowledge the National Cancer Institute for allocation of computer time and staff support at the Advanced Biomedical Computing Center of the Frederick Cancer Research and Development Center.

REFERENCES

1. Ebright, R. H., Ebright, Y. W., Prendergast, P. S., and Gunasekera A. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 2882–2886
2. Yang, S. W., and Nash, H. A. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 12183–12187
3. Kohlstaedt, L. A., Wang, J., Friedman, M., Rice, P. A., and Steitz, T. A. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 4466–4470
4. Muyldermans, S. (1998) Nature 395, 492–495
5. Andrich, M. M., Gehrig, P., and Gehring, H. (1999) J. Biol. Chem. 274, 5326–5332
6. Mandal, A. K., Bhattacharyya, A., Bhattacharyya, S., Bhattacharyya, T., and Roy, S. (1998) Protein Sci. 7, 1546–1551
7. Bartegi, A., Roustan, C., Kassab, R., and Fattoum, A. (1999) Biochemistry 38, 335–341
8. Pan, C. Q., Landgraf, R., and Sigman, D. S. (1995) Protein Sci. 4, 2279–2288
9. Pan, C. Q., Johnson, R. C., and Sigman, D. S. (1996) Biochemistry 35, 4326–4383
10. Chen, C. H., Landgraf, R., Walts, A. D., Chan, L., Schlonk, P. M., Terwilliger, T. C., and Sigman, D. S. (1998) Biol. Chem. 479, 283–292
11. Hall, K. B., and Fox, R. O. (1999) Methods (Orlando) 18, 78–84
12. Murakami, K., Kimura, M., Owens, J. T., Meares, C. F., and Ishihama, A. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 1709–1714
13. Heilek, G. M., and Nolle, H. F. (1999) J. Mol. Biol. 286, 355–364
14. Heuser, T. S., and Brown, P. O. (1998) Biochemistry 37, 6667–6678
15. Canard, B., Sarafi, R., and Richardson, C. C. (1998) Proc. Natl. Acad. Sci. U. S. A. 94, 11279–11284
16. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) Science 282, 1669–1675
17. Jaeger, J., Rostel, T., and Steitz, T. A. (1998) EMBO J. 17, 4535–4542
18. Kohlstaedt, L. A., Wang, J., Friedman, M., Rice, P. A., and Steitz, T. A. (1992) Science 256, 1783–1790
19. Jacobs-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Lu, X., Tantillo, C.,
20. Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C., and Harrison, S. C. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1222–1226
21. Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., and Stuart, D. (1995) *Nature Struct. Biol.* **2**, 303–308
22. Wisniewski, M., Palaniappan, C., Fu, Z., Le Grice, S. F. J., Fay, P., and Bambara, R. A. (1995) *J. Biol. Chem.* **270**, 7068–7076
23. Ghosh, M., Jacques, P. S., Rodgers, D., Ottmann, M., Darlix, J.-L., and Le Grice, S. F. J. (1996) *Biochemistry* **35**, 4279–4286
24. Cameron, C. E., Ghosh, M., Rodgers, D. W., Le Grice, S. F. J., and Benkovic, S. J. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6700–6705
25. Davies, J. P., Hostomska, Z., Hostomsky, Z., Jordan, S. R., and Matthews, D. A. (1991) *Science* **252**, 88–95
26. Wohrl, B. M., Tantillo, C., Arnold, E., and Le Grice, S. F. J. (1995) *Biochemistry* **34**, 5343–5350
27. Ghosh, M., Howard, K. J., Cameron, C. E., Benkovic, S. J., Hughes, S. H., and Le Grice, S. F. J. (1995) *J. Biol. Chem.* **270**, 7068–7076
28. Rausch, J. R., Wohrl, B. M., and Le Grice, S. F. J. (1996) *J. Mol. Biol.* **257**, 500–511
29. Gotte, M., Maier, G., Gross, H. J., and Heumann, H. (1998) *J. Biol. Chem.* **273**, 10139–10146
30. Le Grice, S. F. J., Naas, T., Wohlenstüger, B., and Schatz, O. (1991) *EMBO J.* **10**, 3905–3911
31. Schatz, O., Cromme, P., Gruninger-Leitch, F., and Le Grice, S. F. J. (1989) *FEBS Lett.* **257**, 311–314
32. Le Grice, S. F. J., Cameron, C. E., and Benkovic, S. J. (1995) *Methods Enzymol.* **262**, 139–147
33. Dumoulin, P., Ebright, R. H., Knetel, R., Kaptein, R., Granger-Schnarr, M., and Schnarr, M. (1996) *Biochemistry* **35**, 4279–4286
34. Dumoulin, P., Oertel-Buchheit, P., Granger-Schnarr, M., and Schnarr, M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 2030–2034
35. Maxam, A., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560
36. Wohrl, B. M., Howard, K. J., Jacques, P. S., and Le Grice, S. F. J. (1994) *J. Biol. Chem.* **269**, 8541–8548
37. Metzger, W., Hermann, T., Schatz, O., Le Grice, S. F. J., and Heumann, H. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5909–5913
38. Wohrl, B. M., Georgiadis, M., Telesnitsky, A., Hendrickson, W., and Le Grice, S. F. J. (1995) *Science* **267**, 96–99
39. Isel, C., Westhof, E., Massire, C., Le Grice, S. F. J., Ehresmann, B., Ehresmann, C., and Marquet, R. (1999) *EMBO J.* **18**, 1036–1046
40. Jacques, P. S., Wohrl, B. M., Howard, K. J., and Le Grice, S. F. J. (1994) *J. Biol. Chem.* **269**, 1388–1393