Biochemistry

 Volume 30, Number 26 July 2, 1991

*Perspectives in Biochemistry*

# HIV Reverse Transcriptase Structure–Function Relationships†

Alfredo Jacobo-Molina and Edward Arnold*

*Center for Advanced Biotechnology and Medicine (CABM) and Department of Chemistry, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854-5638*

*Received April 3, 1991*

ABSTRACT: HIV reverse transcriptase (RT) is the target of the most widely used treatments for AIDS. Biochemical and mutagenesis studies performed on HIV-1 RT are reviewed in light of the enzyme's structure and functions. Features described include domain arrangement, dimerization, proteolytic processing, and specific recognition of the priming tRNA. Possible regions of functional importance as determined by comparative amino acid sequence analysis and by site-directed mutagenesis are identified. Among the conclusions of the analysis is the unexpected realization that the substrate for proteolytic maturation of the HIV-1 RT p66/p66 homodimer to the p66/p51 heterodimer is most likely an unfolded RNase H domain. In addition, the current progress in crystallization and structure determination of HIV-1 RT is described. Finally, a functional model of the active reverse transcription complex is presented.

The reverse transcriptase (RT) enzyme of the human immunodeficiency virus (HIV) and other retroviruses is the sole viral enzyme required for the catalytic formation of proviral DNA from viral RNA. Consequently, it is an extremely favorable target for antiviral design. While reverse transcription is absolutely required for retroviral replication, only a few eukaryotic cellular examples have been identified. Furthermore, since RT is the only nonorganellar polymerase that operates in the cytoplasm of retrovirus-infected cells, it invites the specific targeting of compounds against cytoplasmic polymerization. Among the drugs that are most widely used in current treatment of AIDS are RT inhibitors [reviewed in Mitsuya et al. (1990)]. Many RT inhibitors are nucleoside analogues that function by prematurely terminating elongation of nucleic acids by virtue of their own incorporation during synthesis. However, some of the most potent and selective inhibitors, such as two different sets of diazepinone derivatives (Merluzzi et al., 1990; Pauwels et al., 1990), are not substrate analogues and thus may inhibit by different mechanisms. That these recently described antiviral compounds show specificity in inhibiting only HIV-1 RT, and not other polymerases including HIV-2 RT, suggests that they may work in a similar fashion even though their structures are substantially different. In fact, it has been proposed that these compounds share a common binding site on the enzyme (Wu et al., 1991). Representative reviews of RT have concentrated on the enzyme's functional roles in retroviruses (Varmus, 1987; Goff, 1990) and in the context of antiviral design against AIDS (Mitsuya et al., 1990; Arnold & Arnold, 1991). This article surveys our current knowledge of the relation between the structure and some of the functions of HIV-1 RT and other RTs and integrates the results of many laboratories. The topics described include the overall functional mapping of the enzyme at the level of domain structure, dimerization, proteolytic processing, specific recognition of the priming tRNA, and identification of possible regions of functional importance as determined by comparative amino acid sequence analysis and by site-directed mutagenesis studies. As each of these functional aspects of RT become better understood, additional targets for antiviral intervention against retroviral diseases will emerge. The conserved regions of the RNase H domain of retroviral RTs are analyzed in the context of the recently solved *Escherichia coli* RNase H (ribonuclease hybrid) structure and are shown either to map onto the putative substrate-binding face of the enzyme or to be part of its hydrophobic core. Surprisingly, the HIV-1 protease cleavage site on the HIV-1 RT p66/p66 homodimer, proteolyzed to form the p66/p51

† Research in the laboratory of E.A. on HIV RT has been supported by NIH Grants AI26790 and GM39558, the Center for Advanced Biotechnology and Medicine, the New Jersey Commission on Science and Technology, Hoffmann-La Roche, and Johnson & Johnson. E.A. is an Alfred P. Sloan Foundation Research Fellow. A.J.-M. thanks Johnson & Johnson for postdoctoral fellowship support.

* Author to whom correspondence should be addressed.

0006-2960/91/0430-6351$02.50/0 

heterodimer, is in the middle of a putative β-sheet, suggesting that one of the RNase H domains in the p66/p66 homodimer is partially or completely unfolded. In addition, since a three-dimensional structure of HIV-1 RT is expected to greatly enhance RT inhibitor design, some of the current progress in crystallization and structure determination of HIV-1 RT is described. Finally, a functional model of the active reverse transcription complex is proposed.

## Overall Properties and Domain Arrangement of HIV RT

RT can polymerize a DNA strand from either an RNA or a DNA template. The enzyme also encodes an RNase H activity that digests the RNA moiety of RNA–DNA duplexes into short RNA oligomers. In both virions (Veronese et al., 1986; Wondrak et al., 1986) and infected cells (Lightfoote et al., 1986), HIV-1 RT has been characterized as a heterodimer consisting of 66- and 51-kDa chains. The amino termini of the two chains are identical and the smaller subunit of the heterodimer is apparently derived by proteolytic processing of the p66 chain by the HIV-1 protease (HIV-1 PR). Expression of the HIV-1 RT gene in bacteria yields both a p66/p66 homodimer and a p66/p51 heterodimer that is processed by bacterial proteases.

It has been established that the polymerization and RNase H activities of HIV-1 RT map to the amino- and carboxy-terminal portions of the p66 chain, respectively, and that the p51 chain lacks the RNase H domain. The high similarity of RT polymerase and RNase H amino acid sequences across a wide range of RTs suggests a high degree of structural relatedness. By contrast, the composition of the so-called "tether" domain linking the two functional regions is more variable (Johnson et al., 1986). For some reverse transcriptases it has been possible to separate fully active polymerase and RNase H domains (Lai & Verma, 1978; Hizi & Hughes, 1988; Tanese & Goff, 1988). However, for HIV-1 RT, a functional interdependence of the two domains has been suggested, since linker insertion mutations at the amino terminus of HIV-1 RT can affect RNase H activity while such mutations at the carboxy terminus can disrupt polymerization (Hizi et al., 1989; Prasad & Goff, 1989; Hizi et al., 1990). This interdependence of domains has recently been demonstrated through an elegant set of experiments: whereas the RNase H domain expressed in bacteria is inactive by itself, addition of the p51 polypeptide leads to enzymatic activity (Hostomsky et al., 1991). Different groups have reported that the p51 chain was either inactive in DNA polymerization (Starnes et al., 1988), had very low levels of activity (Hizi et al., 1988; Tisdale et al., 1988; Bathurst et al., 1990), or was only active as a weakly associated dimer (Restle et al., 1990).

## Dimerization of HIV-1 RT

Reverse transcriptases have been characterized at naturally occurring monomers [murine leukemia virus (MuLV) RT] or dimers (HIV-1 RT, HIV-2 RT, avian myeloblastosis virus RT, and Rous sarcoma virus RT). Sequence alignments of RTs and other polymerases suggest that the p66 subunit of the HIV-1 RT contains all the sequence information needed for activity, implying that monomeric p66 should be a fully active RT. Similar analysis suggests that the p51 chain should be active in polymerization. However, it has been reported that enzymatic activity is almost exclusively confined to the dimeric forms of the protein (Restle et al., 1990). In addition, competition studies of HIV-1 RT p66/p51 heterodimer with different primer–template complexes indicate the presence of only one binding site (Painter et al., 1990). These results suggest that dimerization of HIV-1 RT, and probably that of other dimeric RTs, is required for optimal activity of the enzyme. Evidence that the p66 chain is providing the polymerization active site in the p66/p51 heterodimer comes from the observation that partial proteolysis of p66/p51 with trypsin produces an active p66/p30/p29 RT "complex" (Lowe et al., 1988) and that the cleavage site is lysine residue K220, which is highly conserved in RNA-dependent polymerases (Poch et al., 1989). The adjacent K219 has been suggested to be part of the triphosphate binding site (Larder & Kemp, 1989). In addition, cross-linking studies of p66/p51 with substrate analogues have shown that only the p66 subunit is labeled (Chen et al., 1991). Furthermore, a non-nucleoside-analogue inhibitor of HIV-1 RT selectively binds the p66 chain in the p66/p51 heterodimer (Wu et al., 1991).

The HIV-1 p66/p66 homodimer isolated from bacteria is a loosely associated dimer with specific activity less than half the activity of the bacterially processed p66/p51 heterodimer. This difference could be due to the presence of inactive monomer and/or steric hindrance of the extra RNase H domain during reverse transcription. Dimerization of the p66/p66 homodimer is favored in high ionic strength solutions, while monomerization is favored in low ionic strength solutions and organic solvents (e.g., ethylene glycol and ethanol; our unpublished results). Similarly, the p66/p51 heterodimer cannot be dissociated in >1.0 M NaCl (Becerra et al., 1990) and is also stable in water. All of these data taken together suggest that the dimer interface in HIV-1 RT is hydrophobic in nature and that once the heterodimer is formed, the association is so tight that it cannot be disrupted even in low ionic strength solutions. A recent report (Restle et al., 1990) showed that 20% acetonitrile was required to dissociate the p66/p51 heterodimer and a dissociation constant of $10^{-9}$ M or lower was proposed, while the dissociation constant for the p66/p66 homodimer was $10^{-5}$ M. Finally, it has recently been shown that the dimerization domain of HIV-1 RT is localized in the last 20 kDa of the p51 subunit (Becerra et al., 1990), corresponding to the tether region separating the polymerization and the RNase H domains in the p66 chain (Johnson et al., 1986).

## Proteolytic Processing of HIV-1 RT p66 to p51

In the presence of proteases, the loosely associated HIV-1 p66/p66 homodimer can be converted into the tightly associated p66/p51 heterodimer. Once the heterodimer is formed, the remaining p66 subunit is not susceptible to cleavage. The in vivo cleavage of the p66 chain is probably carried out by the HIV-1 PR during budding of virions from infected cells or afterwards. It has not been determined whether monomeric or dimeric p66 or both are substrates for the HIV-1 PR. In addition, as stated earlier, bacterial proteases can also produce the p66/p51 form of the enzyme. However, the HIV-1 PR generates F440 as the carboxy end of the p51 chain (Mizrahi et al., 1989; Bathurst et al., 1990; Graves et al., 1990), while the bacterial p51 has a heterogeneous C-terminus centered around P433 (Lowe et al., 1988). Therefore, although this region of HIV-1 RT p66 is especially susceptible to proteases, the fact that the second subunit of the p66/p51 heterodimer is not proteolyzed might be explained by it being inaccessible to cleavage or in a conformation no longer susceptible to proteolysis.

Until the recent structure determination of *E. coli* RNase H (Katayanagi et al., 1990; Yang et al., 1990), an enzyme expected to be structurally related to the RNase H domain of RTs, it has been assumed that the cleavage site of the HIV-1 PR on HIV-1 RT p66 (the peptide bond between residues

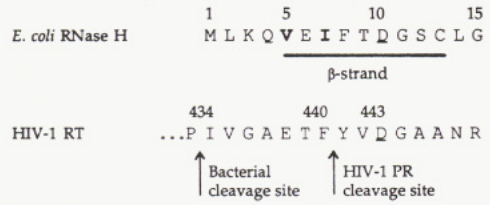


FIGURE 1: Amino acid sequences corresponding to the amino termini of *E. coli* and HIV-1 RT RNases H (Yang et al., 1990). Conserved essential aspartic acids are underlined. Amino acids in the first β-strand in *E. coli* RNase H are indicated and residues involved in the hydrophobic core are in boldface type. The primary bacterial and the HIV-1 PR sites of cleavage are indicated.

F440 and Y441) defined the beginning of the RNase H domain of RT. However, examination of *E. coli* RNase H reveals a striking finding. Sequence alignments in light of the structure suggest that Phe440 and Tyr441 of HIV-1 RT are in the middle of a putative β-strand in what is likely to be a five-stranded β-sheet (Figures 1 and 2). This strand not only includes the conserved Asp residue essential for RNase H activity of both enzymes (D10 in *E. coli* RNase H and D443 in HIV-1 RT, as described later) but also is contributing to the extensive hydrophobic core of the *E. coli* structure. The atomic structure of the HIV-1 PR with and without substrate analogues (Miller et al., 1989; Wlodawer et al., 1989) shows that *it is highly unlikely that the protease could reach the cleavage site unless the RNase H domain of HIV-1 RT p66 is partially or totally unfolded.* The active site of the HIV-1 PR is located within a cleft of the protein that is able to accommodate only a single polypeptide strand, which further

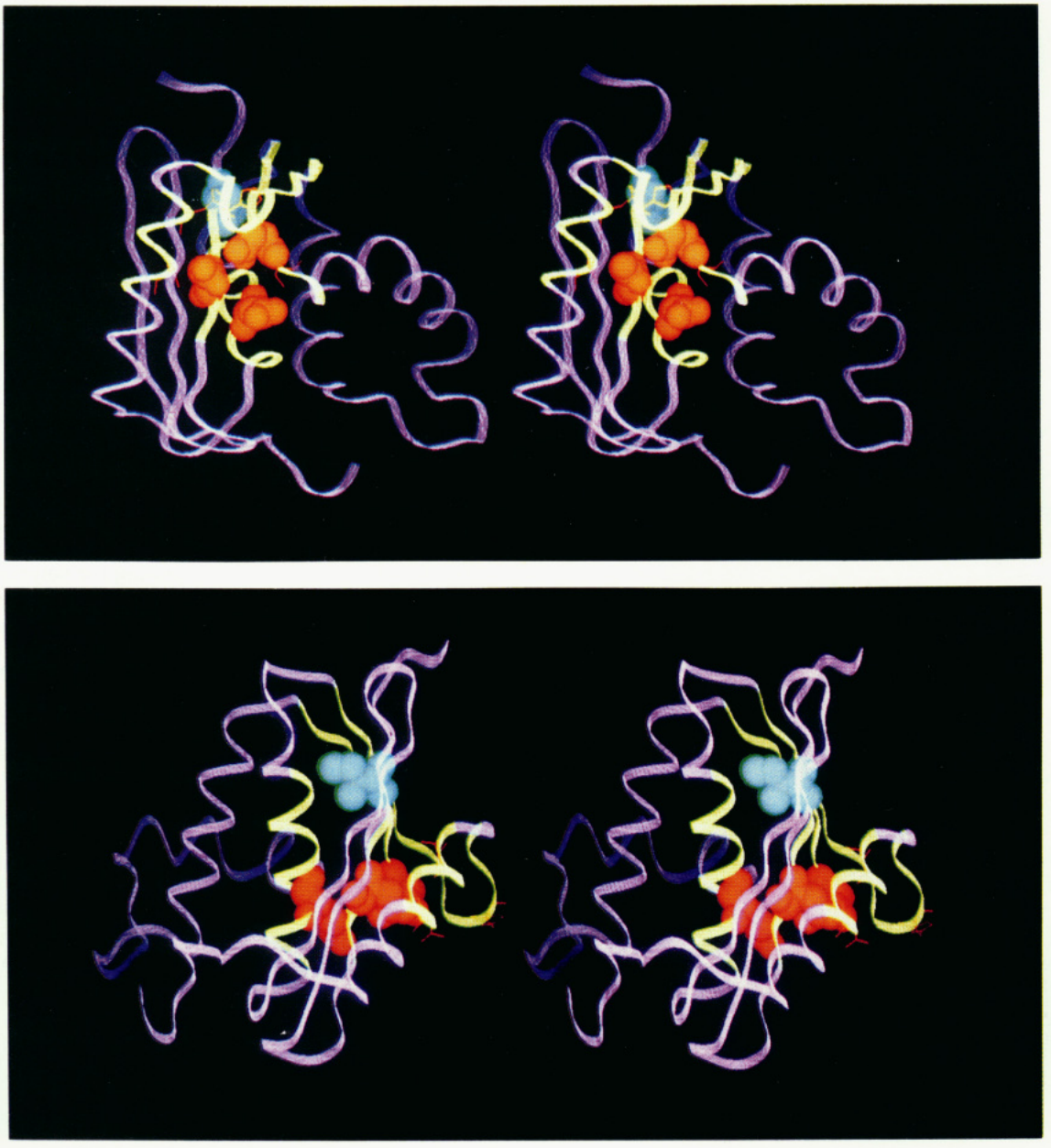

FIGURE 2: Computer graphics generated stereoviews of the *E. coli* RNase H structure [based on the coordinates of Katayanagi et al. (1990)] highlighting the corresponding conserved regions of retroviral RTs and cleavage site for the HIV-1 PR. The tracing of the polypeptide chain is illustrated as a ribbon with conserved regions yellow. Completely conserved amino acids for all RNases H are red, with the catalytic triad residues in a space-filling representation. (Top panel) *E. coli* RNase H structure oriented to highlight the location of conserved regions whose residues map onto the putative substrate binding face, which is centered at the catalytic triad, and also onto the hydrophobic core of the enzyme. (Bottom panel) Orientation of *E. coli* RNase H (rotated about a vertical axis from the view in the top panel) to emphasize that Leu7 (shown in space-filled green), which corresponds to the cleavage site in HIV-1 RT, is located in the center of the five-stranded β-sheet. This implies that the domain must be unfolded for processing of HIV-1 p66 to p51 to occur.

becomes surrounded by loops of the HIV-1 PR dimeric structure referred to as "flaps" (Miller et al., 1989). Thus, it would be improbable that the HIV-1 PR could cleave the F440–Y441 peptide bond in the central strand of a five-stranded β-sheet in a folded RNase H domain. This concept is supported by the fact that it has been possible to identify the cleaved RNase H domain even from an in vitro experiment utilizing purified HIV-1 RT p66/p66 homodimer and HIV-1 PR (Bathurst et al., 1990; Ferris et al., 1990), suggesting that the RNase H domain is further proteolyzed, as is commonly observed for unfolded proteins. This cleavage event appears to differ from other polyprotein processing by the HIV-1 PR, which produces the mature proteins in functional form by presumably cutting between structural domains. Indeed, in other examples of viral polyprotein processing, the cleavages apparently occur in regions connecting folded protein domains [e.g., picornaviral capsid proteins (Arnold et al., 1987; Ypma-Wong et al., 1988)]. Thus, the location of the cleavage site within the folded domain of retroviral RT RNase H is particularly intriguing. If the p66/p66 homodimer is the substrate for cleavage, it is possible that some of the energy required to unfold the RT RNase H domain is provided by the dimerization process. The slow dimerization from monomeric HIV-1 p66 could also be potentially explained by the unfolding of the RNase H domain as a rate-limiting step.

## Recognition of Primer tRNA by HIV-1 RT

Many retroviral RTs appear to recruit specific transfer RNA molecules that are used to prime viral reverse transcription. The recognition process between RTs and their "cognate" tRNAs may be analogous to tRNA recognition by aminoacyl-tRNA synthetases and involve an extensive surface of contact, as was found in the structure of a complex between glutaminyl-tRNA synthetase and $tRNA^{Gln}$ (Rould et al., 1989). The sequence of the HIV-1 primer-binding site (Ratner et al., 1985; Sanchez-Pescador et al., 1985; Wain-Hobson et al., 1985) suggested that $tRNA^{Lys,3}$ is used to prime DNA synthesis in HIV infection. HIV-1 RT has been shown to preferentially bind $tRNA^{Lys,3}$ even in the presence of a 100-fold excess of other tRNAs (Barat et al., 1989). In addition, the binding of the $tRNA^{Lys,3}$ induces significant structural changes in HIV-1 RT (Robert et al., 1990).

$tRNA^{Lys,3}$ inhibits polymerization by the HIV-1 RT p66/p51 heterodimer (Sallafranque-Andreola et al., 1989; Bordier et al., 1990) but does not competitively inhibit polymerization by the p66/p66 homodimer (Bordier et al., 1990). This may imply that the recognition of the tRNA by the loosely associated homodimer is more easily reversible than by the heterodimer, due to the presence of the extra RNase H domain in the homodimer, which may interfere with tRNA binding. Interestingly, the dimeric avian retrovirus RT specifically recognizes $tRNA^{Trp}$, while MuLV RT, a monomeric enzyme, binds several tRNAs with similar affinities; thus, dimeric RTs may more specifically recognize their cognate primer tRNAs than monomeric RTs. Cross-linking studies (Barat et al., 1989) indicate that the recognition of $tRNA^{Lys,3}$ by HIV-1 RT p66/p51 heterodimer occurs through interactions of the anticodon loop with the enzyme. This further strengthens the analogy with aminoacyl-tRNA synthetase–tRNA recognition: the structure of the glutaminyl-tRNA synthetase–$tRNA^{Gln}$ complex (Rould et al., 1989) also shows that the anticodon loop is in extensive contact with the protein surface.

## Conserved Regions in HIV-1 RT

In comparing retroviral RTs, it has become difficult to correlate regions of amino acid homology with the mutagenesis data since each author has defined and named the conserved regions differently. In order to best organize the current discussion of genetics and function of RT we have assigned 13 conserved, presumably functionally important regions. The aligned amino acid sequences of 15 RT polymerase domains and 14 RT RNase H domains plus *E. coli* RNase H formed the basis of the current analysis.[1] Criteria for the assignment of conserved regions were that (i) at most of the amino acid positions within a region, more than half of the amino acids in the aligned sequences share identity with the consensus residue, (ii) only rare insertions or deletions occur within a region for any of the sequences examined, and (iii) regions defined in other structure–function analyses of RT are retained for consistency. A schematic layout of the regions 1–8 in the polymerization domain of HIV-1 RT is shown in Figure 3, with

Table I: Conserved Regions of the Polymerization Domain of Retroviral Reverse Transcriptases[a]

| region | sequence | reference |
|---|---|---|
| 1 | 23QWPLTEEKIKALVEI37 | Our observations |
| 2 | 55PYNTPVFAIKKKDSTKWRKLVDFRELN81 | Our observations |
| | 57NTPVFAIKK65 | Region "A" in (*b*) |
| 3 | 102KKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPS134 | |
| | 102KKKSVTVLDVGDAYFSV118 | Region "A" in (*c*) |
| | 105SVTVLDVGDAYFSVPLDEDFRKYTAE130 | Region "A" in (*d*) |
| | 109LDVGDAYF116 | Region "B" in (*b*) |
| | 129AFTIPS134 | Our observations |
| 4 | 143RYQYNVLPQGWKGSPAIFQSSMTKILEP170 | |
| | 143RYQYNVLPQGWKGSPAIFQSSMTKILE169 | Region "B" in (*c*) |
| | 148VLPQGWKGSP157 | Region "C" in (*b*) |
| | 168LEP170 | Region "D" in (*b*) |
| 5 | 178IVIYQYMDDLYVGS191 | Region "C" in (*d*) |
| | 180IYQYMDDLYVG190 | Region "E" in (*b*) |
| | 180IYQYMDDLYVG190 | Region "C" in (*c*) |
| 6 | 211RWGLTTPDKKHQ222 | Region "D" in (*c*) |
| 7 | 227FLWMGYELHPD237 | Region "E" in (*c*) |
| 8 | 257IQKLVGKLNW266 | Region "F" in (*b*) |

[a] Individual amino acid residues conserved in 80% or more of the 15 aligned sequences are underlined. Residue numbering refers to HIV-1 RT (beginning with PISPI as residues 1–5 of RT from the HIV-1 $III_B$ isolate). Region 8 is not obviously conserved in MuLV RT. Amino acid residues that are conserved in 80% or more of the sequences but are not contained in the conserved regions are G45, G93, P95, P140, T253, P271, and L279. [b] Larder et al. (1987). [c] Poch et al. (1989). [d] Delarue et al. (1990).

Table II: Conserved Regions in the RNase H Domain of Retroviral Reverse Transcriptases[a]

| region | sequence |
|---|---|
| 9 | $_{441}$YVDGA$_{445}$ |
| 10 | $_{475}$QKTELQAIYLAL$_{486}$ |
| 11 | $_{493}$VNIVTDSQY$_{501}$ |
| 12 | $_{531}$VYLAWVPAHKGI$_{542}$ |
| 13 | $_{544}$GNEQVDKL$_{551}$ |

[a] The selection of regions was based on sequence alignments reported [Johnson et al. (1986), Ready et al. (1988), Kanaya et al. (1990)] as well as our own. Amino acid residues conserved in 80% or more of the aligned sequences are underlined. Residue numbering refers to HIV-1 RT (beginning with PISPI as amino acid residues 1–5 in the RT from the HIV-1 $III_B$ isolate).

[1] A combination of the methods of Feng and Doolittle (1987) and Smith and Waterman (1981) was utilized for constructing the amino acid sequence alignments used for this analysis. Copies of the alignments formatted by using the LINEUP and PRETTY routines of the Genetics Computer Group package (Devereux et al., 1984) are available from the authors upon request.

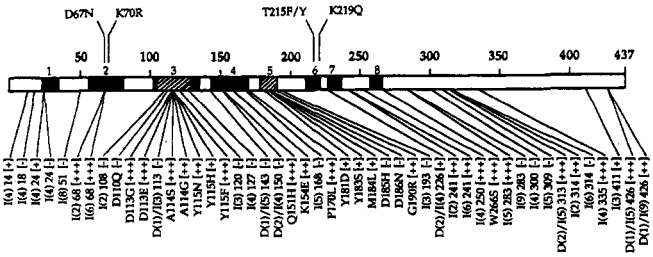

FIGURE 3: Schematic summary of mutations performed in the polymerization and tether domains of HIV-1 RT. The horizontal bar represents the first 437 amino acids of HIV-1 RT. Regions of homology among reverse transcriptases are in black, and those regions also identified in all nucleic acid polymerases (Delarue et al., 1990) are crosshatched. The different types of mutagenesis experiments performed on HIV-1 RT are indicated as follows: insertions of 2–9 amino acids [I(no.)], followed by the residue number at the insertion site, deletion/insertions (i.e., a combination of a deletion of 1 or 2 residues and the insertion of 3–9 residues) [D(no.)/I(no.)] followed by the residue number at that site, or point mutations (e.g., D110Q = substitution of aspartic acid with glutamine at position 110). The symbols in brackets indicate levels of activity: [−] = 0–1% active as compared to wild type, [+] = 2–10% active, [++] = 11–25% active, and [+++] = 26–100% active. The AZT-resistant isolates are indicated above the bar.

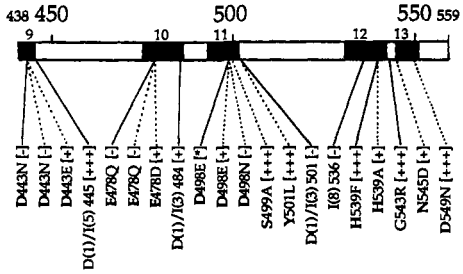

FIGURE 4: Schematic summary of mutations performed in the RNase H domain of HIV-1 RT. The horizontal bar represents the last 122 amino acids of HIV-1 RT. Regions of homology among retroviral RNases H and *E. coli* RNase H are in black. The different mutagenesis experiments and resulting activities are denoted as in Figure 3. The dotted lines indicate mutations performed on the corresponding conserved residues in *E. coli* RNase H.

the amino acid sequences for each region given in Table I. The corresponding information for regions 9–13 in the RNase H domain of HIV-1 RT is given in Figure 4 and Table II.

(*a*) *Conserved Regions in the Polymerization Domain.* Amino acid sequence alignments have identified regions of homology in the polymerization domain of RTs (Johnson et al., 1986; Larder et al., 1987), RNA-dependent polymerases (Poch et al., 1989), and among all polymerases (Delarue et al., 1990). Doolittle et al. (1989) and Xiong and Eickbush (1990) emphasize the phylogenetic relationships among RTs encoded by a diverse set of retroelements. In our suggested nomenclature for the polymerization domain of HIV-1 RT (Table I), a total of eight regions of conserved amino acids have been identified in the polymerization domain of RTs. We include one region not previously described (region 1) and extend the homology covered by regions 2 and 3. Regions 3–6 contain highly conserved residues, including the "polymerase signature sequence" (YXDD) in region 5. In the case of HIV-1 RT, the conserved regions extend only from Q23 to W266, leaving about 170 amino acids as the spacing or tether region separating it from the RNase H domain (Johnson et al., 1986). Nonetheless, it has recently been shown that the tether in HIV-1 RT includes the dimerization domain (Becerra et al., 1990), which would therefore be expected to exhibit some degree of conservation among dimeric RTs; such conservation has not been identified by this analysis. Alignment of a recently discovered 316-residue bacterial RT lacking an RNase H domain (Sun et al., 1991) with other RTs further supports the idea that the information required for polymerization is largely concentrated in the first 266 residues of HIV-1 RT.

(*b*) *Conserved Regions in the RNase H Domain.* On the basis of sequence alignments of the RNase H domain of RTs and *E. coli* RNase H (Johnson et al., 1986; Ready et al., 1988; Kanaya et al., 1990; our own observations), we propose five regions of homology in the HIV-1 RT RNase H domain (Table II) that could complement the structural information now available for the *E. coli* enzyme (Katayanagi et al., 1990; Yang et al., 1990). The regions have been numbered consecutively relative to those regions described for the polymerization domain. The conserved regions were defined prior to the determination of the three-dimensional structure of *E. coli* RNase H and, therefore, no bias was introduced on the basis of an awareness of structural motifs. It is interesting to note the correspondence between the secondary structural elements of *E. coli* RNase H [Figure 5 in Yang et al. (1990)] and the five regions of conserved amino acids. As shown in Figure 2, the corresponding conserved residues for retroviral RNases H either map to the face of the *E. coli* enzyme that is proposed to interact with the RNA–DNA hybrid substrate (Yang et al., 1990) or participate in the hydrophobic core of the enzyme.

Region 9 consists of the amino-terminal residues of the RNase H domain, including the conserved D443, which corresponds by sequence alignments to D10 in the *E. coli* enzyme. This region includes five amino acids that correspond to a portion of *E. coli* RNase H $\beta$-strand A[2] (Katayanagi et al., 1990). Region 10 is analogous to most of $\alpha$-helix 1 of the *E. coli* structure (12 out of 16 residues) and includes the conserved E478 (corresponding to E48 in *E. coli*) that along with D443 in region 9 and D498 in region 11 presumably chelates the essential $Mg^{2+}$ at the active site (Katayanagi et al., 1990). The first five residues in region 11 correspond to part of $\beta$-strand D in the *E. coli* enzyme. Among these first five residues are V493 and I495 (V65 and L67 in *E. coli* RNase H), which are conserved hydrophobic residues expected to be part of the hydrophobic core of the protein. Conserved residues D498 and S499 are also part of region 11, presumably flanked by $\beta$-strand D and $\alpha$-helix 2.

Residues 1–6 of region 12 correspond to $\beta$-strand E in the *E. coli* structure and include V531, L533, and W535 (I116, W118, and W120 in the *E. coli* enzyme), which are probably part of the hydrophobic core of the protein. The remaining amino acids in region 12 are expected to belong to the connecting loop between the $\beta$-strand E and $\alpha$-helix 5, which is thought to participate in binding to the RNA–DNA substrate (Yang et al., 1990). This putative loop in HIV-1 RT also contains the highly conserved residues V536 and H539, the latter of which has been shown to participate in substrate binding. Proline and/or glycine residues are common features of RTs in this region of the polypeptide.

Finally, the eight residues in region 13 correspond to part of the 15-residue $\alpha$-helix 5 in the *E. coli* structure. This helix includes two highly conserved residues (N545 and D549; N130 and D134 in *E. coli* RNase H) that have not been implicated in catalysis. According to the *E. coli* structure, D549 should be near the conserved H539 and in the proposed substrate-

[2] Both structures reported for *E. coli* RNase H are essentially identical (Yang et al., 1990; Katayanagi et al., 1990) and the two papers name the secondary structure elements differently. We have adopted the nomenclature used in Katayanagi et al. (1990).

Table III: Selected Mutations in the Polymerization and Tether Domains of HIV-1 RT[a]

| site | mutation | new sequence | activity | comments | ref |
|---|---|---|---|---|---|
| D67 | PM | N | nd | AZT-resistant isolate | *b* |
| S68 | I(2) | SEFT | +++ | next to AZT-resistant sites (D67 and K70) | *c* |
| S68 | I(6) | SELEFQFT | +++ | next to AZT-resistant sites (D67 and K70) | *c* |
| K70 | PM | R | nd | AZT-resistant isolate | *b* |
| D110 | PM | Q | - | conserved in all polymerases | *d* |
| D113 | PM | G | +++ | conserved in all RTs and AZT-resistant | *d* |
| D113 | PM | E | +++ | conserved in all RTs and AZT-resistant | *e* |
| D113 | D(1)/I(3) | DGIPY | - | conserved in all RTs and AZT-resistant | *f* |
| Y115 | PM | N | ++ | aromatic residue present in most polymerases | *e* |
| Y115 | PM | H | + | aromatic residue present in most polymerases | *e* |
| Y115 | PM | F | +++ | aromatic residue present in most polymerases | *e* |
| P150 | D(2)/I(4) | PPEFRW | - | deleted G152 conserved in all RDPs | *f* |
| Q151 | PM | H | +++ | present in most RDDPs and all RTs | *e* |
| Y181 | PM | D | + | hydrophobic residue present in most polymerases | *e* |
| Y183 | PM | S | + | present in most polymerases | *d* |
| M184 | PM | L | + | present in most reverse transcriptases | *e* |
| D185 | PM | H | - | conserved in all polymerases | *d* |
| D186 | PM | N | - | conserved in almost all polymerases | *e* |
| T215 | PM | F or Y | nd | AZT-resistant isolate | *b* |
| K219 | PM | Q | nd | AZT-resistant isolate | *b* |
| P226 | D(2)/I(4) | PSEFRW | + | six residues away from K220 conserved in all RDPs | *f* |
| V241 | I(2) | VNSQ | +++ | between regions 7 and 8 | *c* |
| V241 | I(6) | VNWNSNSQ | +++ | between regions 7 and 8 | *c* |
| D250 | I(4) | DRNSNS | +++ | seven residues away from region 8 | *c* |
| W266 | PM | S | +++ | present in almost all RTs and AZT-resistant | *d* |
| L283 | I(5) | LTRITVR | +++ | beginning of putative dimerization domain and tether | *c* |
| L283 | I(9) | LTRIGIPIRVR | - | beginning of putative dimerization domain and tether | *c* |
| P313 | D(2)/I(5) | PLGAPNG | +++ | within putative dimerization domain and tether | *g* |
| V314 | I(2) | VNSH | ++ | within putative dimerization domain and tether | *c* |
| V314 | I(6) | VNWNSNSH | - | within putative dimerization domain and tether | *c* |
| G335 | I(4) | GANSRQ | +++ | within putative dimerization domain and tether | *c* |
| W426 | D(1)/I(5) | WSRIRDQ | +++ | at end of the polymerization domain | *c* |
| W426 | D(1)/I(9) | WSRIRARIRDQ | ++ | at end of the polymerization domain | *c* |

[a] Underlined residues are the actual changes in the mutants. I(no.) = insertion (no. of residues inserted), PM = point mutation, D(no.) = deletion (no. of residues deleted). The symbols in parentheses mark levels of activity: (-) = 0–1% active as compared to wild type, (+) = 2–10% active, (++) = 11–25% active, and (+++) = 26–100% active. nd, not determined. RDP, RNA-dependent polymerase. RDDP, RNA-dependent DNA polymerase. [b] Larder and Kemp (1989). [c] Prasad and Goff (1989). [d] Larder et al. (1987). [e] Larder et al. (1989). [f] Hizi et al. (1989). [g] Hizi et al. (1988).

Table IV: Mutations Defining Functionally Important Sites in HIV-1 RNase H[a]

| site | mutation | new sequence | activity | comments | ref |
|---|---|---|---|---|---|
| D443 | PM | N | - | conserved in all RNases H | *d* |
| D443[b] | PM | N | - | conserved in all RNases H | *e* |
| D443[b] | PM | E | + | conserved in all RNases H | *e* |
| A445 | D(1)/I(5) | AGIRIPN | +++ | adjacent to conserved D443, and G444 | *f* |
| E478 | PM | Q | - | conserved in all RNases H | *g* |
| E478[b] | PM | Q | - | conserved in all RNases H | *e* |
| E478[b] | PM | D | + | conserved in all RNases H | *e* |
| L484 | D(1)/I(3) | LGIPL | + | two residues away from conserved L486 | *f* |
| D498 | PM | N | *c* | conserved in all RNases H | *d* |
| D498[b] | PM | E | + | conserved in all RNases H | *e* |
| D498[b] | PM | N | - | conserved in all RNases H | *e* |
| S499[b] | PM | A | +++ | conserved in all RNases H | *e* |
| Y501[b] | PM | L | +++ | conserved in almost all RNases H | *e* |
| Y501 | D(1)/I(3) | YGIRL | - | conserved in almost all RNases H | *h* |
| V536 | I(8) | VANSRANSRP | - | conserved in almost all RNases H | *f* |
| H539 | PM | F | +++ | conserved in all RNases H | *g* |
| H539[b] | PM | A | + | conserved in all RNases H | *e* |
| G543 | PM | R | +++ | between regions 12 and 13 | *h* |
| N545[b] | PM | D | + | conserved in all RNases H | *e* |
| D549[b] | PM | N | +++ | conserved in all RNases H | *e* |

[a] The different mutagenesis experiments and resulting activities are denoted as in Table III. [b] Mutations performed in the corresponding conserved residues in *E. coli* RNase H. [c] Mutation D498N yielded an unstable enzyme that was not possible to assay for activity (Mizrahi et al., 1990). [d] Mizrahi et al. (1990). [e] Kanaya et al. (1990). [f] Prasad and Goff (1989). [g] Schatz et al. (1989). [h] Hizi et al. (1990).

binding groove (Yang et al., 1990). A more detailed description of the mutagenesis results on which some of the above conclusions were based is discussed in the next section.

## Summary of Mutagenesis Studies on HIV-1 RT

A total of 61 site-directed mutations in HIV-1 RT have been reported to date; 52 were performed in the polymerization and the tether domains (Figure 3) and nine in the RNase H domain (Figure 4). In addition, 11 site-directed mutations were carried out on *E. coli* RNase H in residues corresponding to those conserved in the RNase H domain of RTs. Table III summarizes selected mutations in the polymerization and the tether domains, and Table IV summarizes selected mutations in the RNase H domain. For the purpose of this analysis, we

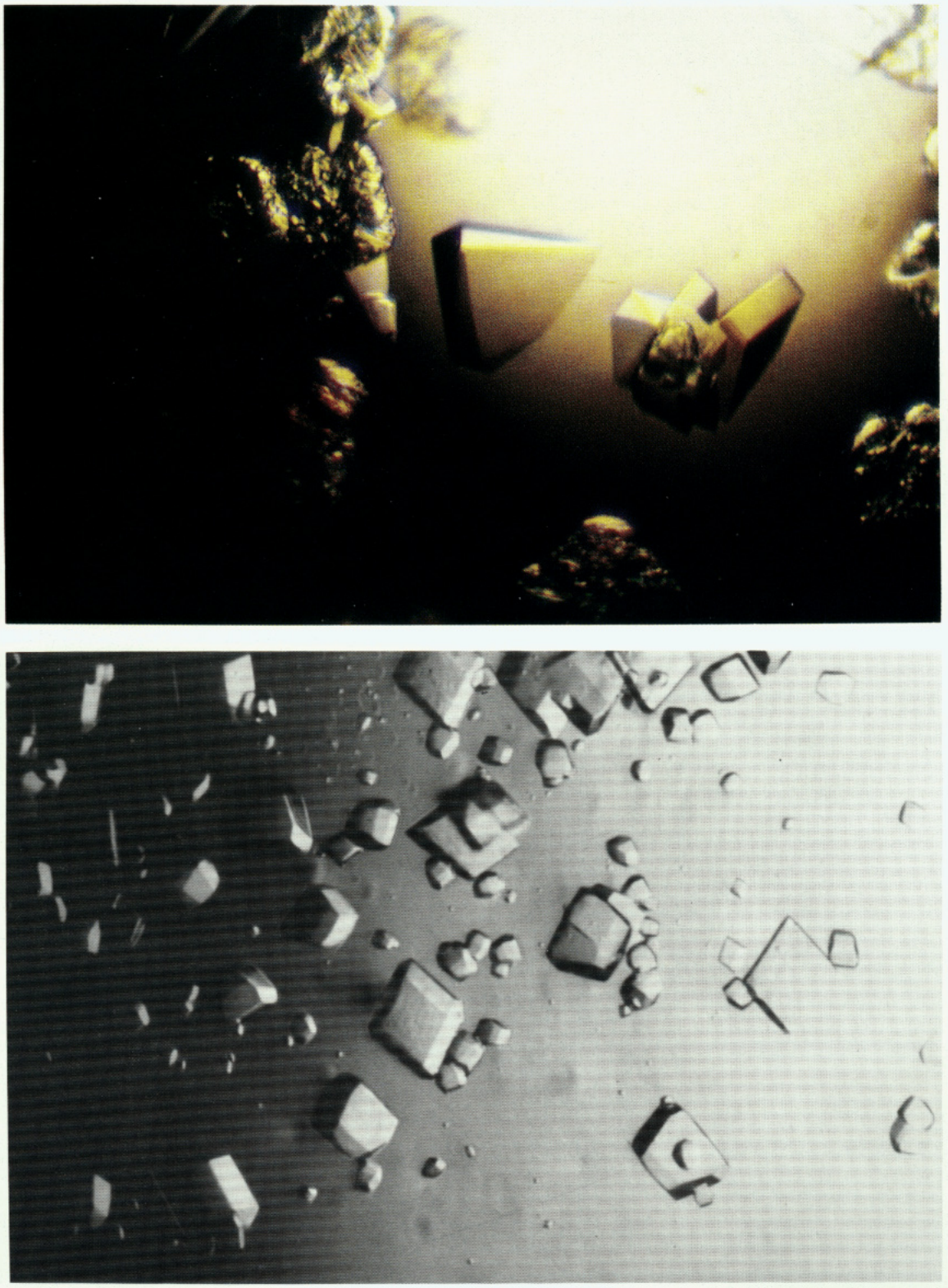

FIGURE 5: (Top panel) Crystals of HIV-1 RT p66/p51 heterodimer that diffract to 8-Å resolution. The crystals were grown in hanging drop vapor diffusion experiments in 50 mM Tris, pH 8.8, and 32.5% saturated ammonium sulfate at a protein concentration of 21 mg/mL. The largest dimension of the crystals is 0.3 mm. (Bottom panel) Cocrystals of HIV-1 RT p66/p51 heterodimer complexed with a monoclonal antibody Fab fragment that diffract to 6-Å resolution. The crystals were grown in vapor diffusion experiments in 20 mM Tris, pH 8.0, 50 mM NaCl, and 40% saturated ammonium sulfate at a total protein concentration of 8 mg/mL. The largest crystal shown measures approximately 0.2 mm in each dimension.

have arbitrarily defined W266, the last residue in conserved region 8 (Table I), as the boundary of the polymerization domain, and E438 as the first amino acid in the RNase H domain. This boundary definition is based on sequence alignment of *E. coli* RNase H with HIV-1 RT which suggests that the first residue of the RNase H domain that is participating in a regular secondary structure (i.e., β-strand A) is HIV-1 RT residue E438 (Figure 1). In describing mutations in the RNase H domain, we are assuming that the three-dimensional structures of all RNases H are closely related.

(*a*) *Polymerization Domain* (*Pro1–Trp266*). Aspartic acid residues are frequently found at binding sites of enzymes,

sometimes chelating essential metal ions, sometimes directly involved in catalyzing reactions. Sequence analysis reveals two aspartic acids conserved in all polymerases, including RTs (Poch et al., 1989; Delarue et al., 1990). In HIV-1 RT, these residues correspond to D110 in region 3 and D185 in region 5 (Table I). The analogous residues of the *E. coli* Klenow fragment are D705 and D882; D882 has been shown to be involved in catalysis (Polesky et al., 1990) and it has been suggested that the spatially close D705 participates in catalysis as well (Delarue et al., 1990). Residues D110 and D185 in HIV-1 RT have been mutated to Q110 and H185 (designated D110Q and D185H, respectively), each yielding completely inactive enzymes (Larder et al., 1987). In addition, mutagenesis of D186 (D186N), part of the canonical polymerase signature sequence YXDD, also produced an inactive enzyme (Larder et al., 1989).

More experiments are needed to better understand the conserved but not essential residues of RTs that are not present in other polymerases (Table III). However, it is tempting to propose that some of these residues may be involved in the binding to the primer tRNA, a common feature of all retroviral RTs, the specific chemistry of which will vary for each enzyme–tRNA complex. It is important to note that standard assays for RNA-dependent DNA polymerization activity use an unnatural (usually homopolymeric) primer as opposed to the appropriate specific cognate tRNA and, thus, the recognition of RT by a natural primer is not required for activity in these assays.

Several mutations of HIV-1 RT have identified sites that can accommodate insertions of amino acids without disrupting the activity of the enzyme. One of these sites is S68, a residue in the middle of conserved region 2 (Table I). Insertions of two or six amino acids after S68 produce mutants with 26–100% the activity of the wild-type enzyme (Table III). Surprisingly, two of the four sites known to confer resistance to 3′-azido-3′-deoxythymidine (AZT) upon mutation [D67 and K70 (Larder & Kemp, 1989); see below] are within two residues of S68. It is fascinating to consider a site that can accommodate insertions of even six amino acids, possibly in a loop structure, that at the same time plays a critical role in the recognition of a nucleotide analogue. Although the length of this region is relatively well-conserved in retroviral RTs, analysis of a larger set of sequences including nonretroviral RTs indicates that this region is evolutionarily flexible and can tolerate insertions (unpublished observations; Xiong & Eickbush, 1990).

A variety of mutants have been analyzed for resistance to inhibition by AZT. Mutants D113G, A114G, Y115N, and Y115H in region 3, K154E in region 4, and W266S in region 8 all have high levels of resistance to inhibition by AZT. Hence, it has been suggested that regions 3, 4, and 8 contribute to the formation of the triphosphate binding site (Larder et al., 1987; Larder & Kemp, 1989). In support of this notion, most of the mutants that were AZT resistant (especially those in region 3) also showed resistance to phosphonoformic acid (PFA), a pyrophosphate analogue.

Sequencing of the HIV-1 RT gene from infected individuals who during therapy became resistant to AZT has consistently identified four sites of mutation: D67N, K70R, T215F or Y, and K219Q (Larder & Kemp, 1989). Partially resistant isolates were seen to have subsets of these mutations. It is particularly interesting to note that none of these naturally occurring mutations fall in the regions suggested by mutagenesis experiments to play a role in triphosphate binding. Two of them (D67N and K70R) are in the middle of region 2 and the other two mutations are in region 6, which includes the conserved K220, suggesting that K220 may be part of the triphosphate binding site. Since the mutations observed in naturally occurring mutants are only a subset of possible mutations that can confer AZT resistance, it is possible that these sites are mutagenic "hot spots" and that other mutations could confer AZT resistance as well. The fact that mutations at one of the AZT-resistant mutation sites requires two nucleotide changes for either of two observed single amino acid changes (T215F or Y) further supports this idea. However, until the three-dimensional structure of HIV-1 RT complexed with a variety of substrates and inhibitors is available, many of these observations will continue to challenge the imaginations of structural biochemists.

(*b*) *Tether Domain* (*Ala267–Ala437*). It is evident from the mutagenesis experiments carried out on the HIV-1 RT polypeptide that the tether domain is among the regions most tolerant to change (Figures 3 and 4, Tables III and IV). For instance, W426, at the end of the tether domain and seven residues away from the primary bacterial protease cleavage site, can permit an insertion of eight amino acids [D(1)/I(9)], with a resultant 11–25% of the activity of the wild-type enzyme (Table III). The deletion/insertion at W426 suggests that the last portion of the tether domain is rather malleable. These results together with knowledge of proteolytically accessible sites suggest that most of the region spanning W426–F440 in at least one of the chains of the p66/p66 homodimer is on the surface of the enzyme.

(*c*) *RNase H Domain* (*Glu438–Ile559*). The mutations D443N (Mizrahi et al., 1990) and E478Q (Schatz et al., 1989) in HIV-1 RT yielded enzymes with no RNase H activity but intact polymerase activity. Identical substitutions in the corresponding residues of *E. coli* RNase H (D10N and E48Q, respectively) also produced completely inactive mutants (Kanaya et al., 1990). In addition, similar results were obtained when D70 in the *E. coli* enzyme (which corresponds to D498 in HIV-1 RT) was mutated to Asn, suggesting that residues D10, E48, and D70 in *E. coli* RNase H or D443, E478, and D498 in HIV-1 RT are involved in the RNase H active sites. In fact, the three-dimensional structure of the *E. coli* enzyme shows the spatial proximity of these three residues forming the triad proposed to chelate the essential $Mg^{2+}$ ion (Figure 2) (Katayanagi et al., 1990; Yang et al., 1990). Another mutation generated at a conserved Asp residue in *E. coli* RNase H was D134N (D549 in HIV-1 RT), which, unlike the previously described mutations, yielded a fully active enzyme (Kanaya et al., 1990). The role of this residue is not clear.

The replacement of Y73 with Leu in $\alpha$-helix 2 of the *E. coli* structure, which corresponds to Y501 and HIV-1 RT, yielded an active enzyme (about 75% as active as the wild-type enzyme, Table IV), implying that, even though most RNases H have a Tyr in that position, another hydrophobic residue that favors $\alpha$-helix formation (i.e., Leu) could replace it without major consequences on the active site.

## Crystallographic Studies of HIV-1 RT

X-ray crystallography is the only technique capable of describing the detailed three-dimensional structure for a molecule the size of HIV-1 RT, whose relevant structural subunit contains at least 117 kDa (p66/p51 heterodimer) even in the absence of primer–template and substrates. Crystals of HIV-1 RT (Figure 5, top panel) have been obtained by a number of different research groups (Lowe et al., 1988, Arnold et al., 1989; Hostomska et al., 1989; Unge et al., 1990; Lloyd et al.,

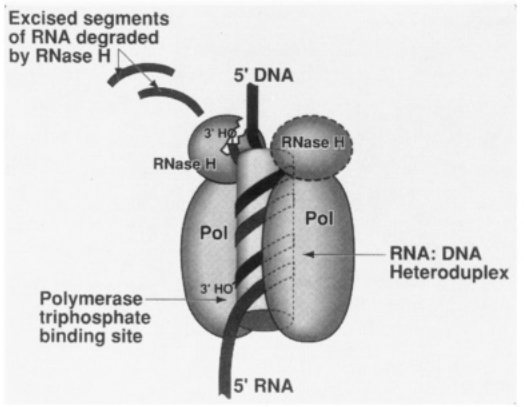


FIGURE 6: Model of HIV-1 RT in the active reverse transcription complex [modified from Arnold and Arnold (1991) with permission]. The extra RNase H domain in the p66/p66 homodimer is depicted with broken lines.

1991). Although extensive efforts have been invested in this endeavor to date, success in obtaining crystals of HIV-1 RT that diffract X-rays to high resolution has not been reported. A low-resolution (7-Å) structure determination is being pursued (Stammers et al., 1990) that might reveal the overall domain arrangement of the enzyme. HIV-1 RT has been cocrystallized with nucleic acid oligomers that represent model primer–template complexes (Arnold et al., 1989; Stammers et al., 1990) and with monoclonal antibody Fab fragments (A. Jacobo-Molina, A. D. Clark, R. L. Williams, S. H. Hughes, and E. A. Arnold, manuscript in preparation) (Figure 5, bottom panel). These Fab cocrystals with HIV-1 RT diffract to at least 6-Å resolution, permitting another approach to determining the structure at moderate resolution. Although the crystals of HIV-1 RT obtained to date will not allow unambiguous tracing of the entire polypeptide chain, it should be possible to map functional sites such as substrate and inhibitor binding sites at this resolution. It is possible that various modes of conformational flexibility that are required for HIV-1 RT activity are interfering with the ability to obtain high-quality crystals. Variations of the structure that might lead to different crystallization properties are being attempted through protein engineering in a number of laboratories (including our own in collaboration with S. Hughes).

## A Model for the Active Reverse Transcription Complex

In an attempt to integrate the findings reviewed in this paper, we propose a model for the spatial arrangement of the different active sites of HIV-1 RT during reverse transcription [Figure 6; modified and extended from Arnold and Arnold, (1991)]. In this model, we assume that the polypeptides of the active dimeric form of HIV RT are parallel to each other, as they might be associated in this orientation during assembly and budding in newly formed virions prior to proteolytic processing (Arnold & Arnold, 1991). This model further assumes that the helical axis of the primer–template substrate is coincident with the axis of symmetry of the dimeric enzyme. As described in the section on dimerization, the hydrophobic dimer interface probably forms part of the binding site of the nucleic acid substrate (Becerra et al., 1990; Painter et al., 1990). In order to optimize the overall efficiency and coordination of the polyfunctional RT molecule, the RNase H moiety may be in a favorable position to digest the RNA template immediately following polymerization of short segments (Oyama et al., 1989). Furfine and Reardon (1991) analyzed the linkage between the polymerization and RNase H activities of HIV-1 RT and found that the two active sites were consistently separated by 15–16 nucleotides in extending primer–template substrates. Cross-linking experiments suggest that the polymerization activity resides on the p66 subunit in the p66/p51 heterodimer (Chen et al., 1991; Wu et al., 1991). Taken together, these results indicate that the polymerization and RNase H active sites from the p66 chain bind to the same face of the primer–template heteroduplex. The observation that cleaved segments are 7–13 nucleotides in length (Oyama et al., 1989; Schatz, 1990; Furfine & Reardon, 1991) may reflect a minimum size preference for efficient endoribonucleolytic substrate cleavage.

This model assumes that the position of the active RNase H domain in the p66/p51 heterodimer and the p66/p66 homodimer is roughly the same. As hypothesized in a previous section, the extra RNase H domain in the homodimer may be partially or totally unfolded and its cleavage may allow a more extensive interaction between the polymerization domains of both subunits and the remaining RNase H domain. This suggestion is consistent with the observation that HIV-1 PR does not further process the p66/p51 heterodimer. An unusual feature of this model is that the asymmetry observed at the heterodimer stage may originate from that of a homodimer in which the two polypeptide chains are covalently, but not conformationally, equivalent. Finally, the lower specific activity of homodimer as compared to heterodimer could be due to steric interference of the extra RNase H domain with nucleic acid substrates in the active reverse transcription complex. Alternatively, the reduced polymerization activity of the homodimer may be due to differences in the contacts at the dimer interface relative to heterodimer or to the presence of significant amounts of inactive monomer in equilibrium with dimer. It has been recently proposed that the strongly conserved residues 20–190 of HIV-1 RT are likely to be in contact with the nucleic acid substrate during polymerization (Barber et al., 1990). In addition, it was suggested on the basis of monoclonal antibody epitope-mapping studies (Ferris et al., 1990; Barber et al., 1990) that the region that includes residues 220–320 is located on the outer surface.

## Conclusions and Current Perspectives

Since the only available three-dimensional structure of a polymerase, that of the Klenow fragment of *E. coli* DNA polymerase I (Ollis et al., 1985), may be related to that of RT, it is worthwhile to consider antiviral design possibilities directed against RT by using the Klenow structure as a reference. The identification of motifs that are conserved among all known polymerases (Delarue et al., 1990) provides modes for suggesting more detailed models of the RT structure in the absence of an experimental determination; however, the amino acid sequence homology of HIV RT with Klenow is barely detectable. Although generating a high-quality model of HIV RT via homology modeling from the Klenow fragment structure is very challenging, thoughtful work of this type is likely to generate additional hypotheses that can be tested, for example, by further mutagenesis experiments. On the other hand, the recently solved structure of *E. coli* RNase H, closely related to the RNase H domain of RTs, should greatly facilitate the modeling possibilities of what corresponds to about 20% of the total length of the HIV-1 RT polypeptide.

When the high-resolution structure of HIV-1 RT succumbs to the powerful methods of X-ray crystallography, it will be extremely exciting to place its known functions in a three-dimensional context. Given the intense activity directed toward this goal, it should be only a matter of time before the structure of retroviral RT is known at high resolution. The development

of antiviral inhibitors of this enzyme will also be greatly accelerated by the extension of structure–activity relationships to three dimensions.

## Added in Proof

The recently determined three-dimensional structure of the RNase H domain of HIV-1 RT (Davies et al., 1991) supports the ideas presented here about retroviral RNases H. Davies et al. independently concluded that an unfolded RNase H domain is the substrate for proteolysis of HIV-1 RT p66 to p51 by HIV-1 PR. A recently obtained form of the HIV-1 RT–antibody cocrystals shown in Figure 5, bottom panel, accommodates the binding of synthetic double-stranded DNA oligomers and diffracts X-rays to at least 3.5-Å resolution (laboratories of E.A. and of S. H. Hughes).

## Acknowledgments

We are particularly grateful to Steve Hughes and his colleagues at the National Cancer Institute–Frederick Cancer Research Facility for their tremendous help of our work on RT. We are especially thankful to Arthur Clark for his dedication and enthusiastic contributions to this work and to Gail Ferstandig Arnold and Roger Williams for critically reviewing the manuscript and for their help on the project. We thank Raymond G. Nanni and Janardan Yadav for their effort on the project and for helpful discussions. We are grateful to K. Morikawa and Y. Katayanagi for providing us with the coordinates of *E. coli* RNase H and to the staff at the Cornell High Energy Synchrotron Source for their help in the diffraction studies. We also thank Aaron Shatkin, Marvin Cassman, James Gaylor, and Manfred Weigele for their support and interest in this work.

## References

Arnold, E., & Arnold, G. F. (1991) *Adv. Virus Res. 39*, 1–87.

Arnold, E., Luo, M., Vriend, G., Rossmann, M. G., Palmenberg, A. C., Parks, G. D., Nicklin, M. J. H., & Wimmer, E. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 21–25.

Arnold, E., Arnold, G. F., Clark, A., Jacobo-Molina, A., Nanni, R., Paidhungat, M. M., & Zemaitis, G. (1989) Third Meeting of Groups Studying the Structure of AIDS-Related Systems and Their Applications to Targeted Drug Design, NIGMS, NIH, Bethesda, MD.

Barat, C., Lullien, V., Schatz, O., Keith, G., Nugeyre, M. T., Gruninger-Leitch, F., Barre-Sinoussi, F., LeGrice, S. F. J., & Darlix, J. L. (1989) *EMBO J. 8*, 3279–3285.

Barber, A. M., Hizi, A., Maizel, J. V., Jr., & Hughes, S. H. (1990) *AIDS Res. Hum. Retroviruses 6*, 1061–1072.

Bathurst, I. C., Moen, L. K., Lujan, M. A., Gibson, H. L., Feucht, P. H., Pichuantes, S., Craik, C. S., Santi, D. V., & Barr, P. J. (1990) *Biochem. Biophys. Res. Commun. 171*, 589–595.

Becerra, S. P., Kumar, A., Widen, S. G., Karawya, E. M., Abbotts, J., Hughes, S. H., Shiloach, J., & Wilson, S. H. (1990) Fourth Meeting of Groups Studying the Structure of AIDS-Related Systems and Their Application to Targeted Drug Design, NIGMS, NIH, Bethesda, MD.

Bordier, B., Tarrago-Litvak, L., Sallafranque-Andreola, M., Robert, D., Tharaud, D., Fournier, M., Barr, P. J., Litvak, S., & Sarih-Cottin, L. (1990) *Nucleic Acids Res. 18*, 429–436.

Chen, N., Painter, G. R., & Furman, P. A. (1991) *Biochem. Biophys. Res. Commun. 174*, 785–789.

Davies, J. F., II, Hostomska, Z., Hostomsky, Z., Jordan, S. R., & Matthews, D. A. (1991) *Science 252*, 88–95.

Delarue, M., Poch, O., Tordo, N., Moras, D., & Argos, P. (1990) *Protein Eng. 6*, 461–467.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Doolittle, R. F., Feng, D.-F., Johnson, M. S., & McClure, M. A. (1989) *Q. Rev. Biol. 64*, 1–30.

Feng, D.-F., & Doolittle, R. F. (1987) *J. Mol. Evol. 25*, 351–360.

Ferris, A. L., Hizi, A., Showalter, S. D., Pichuantes, S., Babe, L., Craik, C. S., & Hughes, S. H. (1990) *Virology 175*, 456–464.

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

Goff, S. P. (1990) *J. Acquired Immune Defic. Syndr. 3*, 817–831.

Graves, M. C., Meidel, M. C., Pan, Y.-C. E., Manneberg, M., Lahm, H.-W., & Gruninger-Leitch, F. (1990) *Biochem. Biophys. Res. Commun. 168*, 30–36.

Hizi, A., & Hughes, S. H. (1988) *Gene 66*, 319–323.

Hizi, A., McGill, C., & Hughes, S. H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1218–1222.

Hizi, A., Barber, A., & Hughes, S. H. (1989) *Virology 170*, 326–329.

Hizi, A., Hughes, S. H., & Shaharabany, M. (1990) *Virology 175*, 575–580.

Hostomska, Z., Hostomsky, Z., Moomaw, E., Hudson, G., Kaufman, I., & Jordan, S. (1989) Third Meeting of Groups Studying the Structure of AIDS-Related Systems and Their Application to Targeted Drug Design, NIGMS, NIH, Bethesda, MD.

Hostomsky, Z., Hostomska, Z., Hudson, G. O., Moomaw, E. W., & Nodes, B. R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1148–1152.

Johnson, M. S., McClure, M. A., Feng, D., Gray, J., & Doolittle, R. F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7648–7652.

Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A., Iwai, S., Inoue, H., Ohtsuka, E., & Ikehara, M. (1990) *J. Biol. Chem. 265*, 4615–4621.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., & Morikawa, K. (1990) *Nature 347*, 306–309.

Lai, M., & Verma, I. M. (1978) *J. Virol. 25*, 652–663.

Larder, B. A., & Kemp, S. D. (1989) *Science 246*, 1155–1158.

Larder, B. A., Purifoy, D. J. M., Powell, K. L., & Darby, G. (1987) *Nature 327*, 716–717.

Larder, B. A., Kemp, S. D., & Purifoy, D. J. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4803–4807.

Lightfoote, M. M., Coligan, J. E., Folks, T. M., Fauci, A. S., Martin, M. A., & Venkatesan, S. (1986) *J. Virol. 60*, 771–775.

Lloyd, L. F., Brick, P., Mei-Zhen, L., Chayen, N. E., & Blow, D. M. (1991) *J. Mol. Biol. 217*, 19–22.

Lowe, D. M., Aitken, A., Bradley, C., Darby, G. K., Larder, B. A., Powell, K. L., Purifoy, D. J. M., Tisdale, M., & Stammers, D. K. (1988) *Biochemistry 27*, 8884–8889.

Merluzzi, V. J., Hargrave, K. D., Labadia, M., Grozinger, K., Skoog, M., Wu, J. C., Shih, C.-K., Eckner, K., Hattox, S., Adams, J., Rosenthal, A. S., Faanes, R., Eckner, R. J., Koup, R. A., & Sullivan, J. L. (1990) *Science 250*, 1411–1413.

Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., & Wlodawer, A. (1989) *Science 246* 1149–1152.

Mitsuya, H., Yarchoan, R., & Broder, S. (1990) *Science 249*, 1533–1544.
Mizrahi, V., Lazarus, G. M., Miles, L. M., Meyers, C. A., & Debouck, C. (1989) *Arch. Biochem. Biophys. 273*, 347–358.
Mizrahi, V., Usdin, M. T., Harington, A., & Dudding, L. R. (1990) *Nucleic Acids Res. 18*, 5359–5363.
Oyama, F., Kikuchi, R., Crouch, R. J., & Uchida, T. (1989) *J. Biol. Chem. 264*, 18808–18817.
Painter, G. R., Wright, L. L., Hopkins, S., & Furman, P. A. (1990) *J. Cell. Biochem.* S14D; UCLA Symposium *HIV and AIDS: Pathogenesis, Therapy, and Vaccine, Abstract L140.*
Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M. J., Breslin, H. J., Raeymaeckers, A., Van Gelder, J., Woestenborghs, R., Heykants, J., Schellekens, K., Janssen, M. A. C., De Clercq, E., & Janssen, P. A. J. (1990) *Nature 343*, 470–474.
Poch, O., Sauvaget, I., Delarue, M., & Tordo, N. (1989) *EMBO J. 8*, 3867–3874.
Polesky, A. H., Steitz, T. A., Grindley, N. D. F., & Joyce, C. M. (1990) *J. Biol. Chem. 265*, 14579–14591.
Prasad, V. R., & Goff, S. P. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3104–3108.
Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, J. S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., & Wong-Staal, F. (1985) *Nature 313*, 277–284.
Ready, M. P., Katzin, B. J., & Robertus, J. D. (1988) *Proteins: Struct., Funct., Genet. 3*, 53–59.
Restle, T., Muller, B., & Goody R. S. (1990) *J. Biol. Chem. 265*, 8986–8988.
Robert, D., Sallafranque-Andreola, M.-L., Bordier, B., Sarih-Cottin, L., Tarrago-Litvak, L., Graves, P. V., Barr, P. J., Fournier, M., & Litvak, S. (1990) *FEBS Lett. 277*, 239–242.
Rould, M. A., Perona, J. J., Soll, D., & Steitz, T. A. (1989) *Science 246*, 1135–1142.
Sallafranque-Andreola, M. L., Robert, D., Barr, P. J., Fournier, M., Litvak, S., Sarih-Cottin, L., & Tarrago-Litvak, L. (1989) *Eur. J. Biochem. 184*, 367–374.
Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D., & Luciw, P. A. (1985) *Science 227*, 484–492.
Schatz, O., Cromme, F. V., Gruninger-Leitch, F., & Le Grice, S. F. J. (1989) *FEBS Lett. 257*, 311–314.
Smith, T. F., & Waterman, M. S. (1981) *Adv. Appl. Math. 2*, 482–489.
Stammers, D. K., Tisdale, M., Powell, K. L., Lowe, D. M., Darby, G., Stuart, D. I., Jones, I. Y., Garman, E. F., & Finch, J. (1990) Fourth Meeting of Groups Studying the Structure of AIDS-Related Systems and Their Application to Targeted Drug Design, NIGMS, NIH, Bethesda, MD.
Starnes, M. C., Gao, W., Ting, R. Y. C., & Cheng, Y.-C. (1988) *J. Biol. Chem. 263*, 5132–5134.
Sun., J., Inouye, M., & Inouye, S. (1991) *J. Bacteriol.* (in press).
Tanese, N., & Goff, S. P. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1777–1781.
Tisdale, M., Ertl, P., Larder, B. A., Purifoy, D. J. M., Darby, G., & Powell, K. L. (1988) *J. Virol. 62*, 3662–3667.
Unge, T., Ahola, H., Bhikhabhai, R., Backbro, K., Lovgren, S., Fenyo, E. M., Honigman, A., Panet, A., Gronowitz, J. S., & Strandberg, B. (1990) *AIDS Res. Hum. Retroviruses 6*, 1297–1303.
Varmus, H. (1987) *Sci. Am. 257*, 56–64.
Veronese, F. D., Copeland, T. D., DeVico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C., & Sarngadharan, M. G. (1986) *Science 231*, 1289–1291.
Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., & Alizon, M. (1985) *Cell 40*, 9–17.
Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Bladwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616–621.
Wondrak, E. M., Lower, J., & Kurth, R. (1986) *J. Gen. Virol. 67*, 2791–2797.
Wu, J. C., Warren, T. C., Adams, J., Proudfoot, J., Skiles, J., Raghaven, P., Perry, C., Potocki, I., Farina, P. R., & Grob, P. M. (1991) *Biochemistry 30*, 2022–2026.
Xiong, Y., & Eickbush, T. H. (1990) *EMBO J. 9*, 3353–3362.
Yang, W., Hendrickson, W. A., Crouch, R. J., & Satow, Y. (1990) *Science 249*, 1398–1405.
Ypma-Wong, M. F., Filman, D. J., Hogle, J. M., & Semler, B. L. (1988) *J. Biol. Chem. 263*, 17846–17856.