Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 3104–3108, May 1989
Biochemistry

Linker insertion mutagenesis of the human immunodeficiency virus reverse transcriptase expressed in bacteria: Definition of the minimal polymerase domain

(DNA polymerase/RNase H/domain structure/subunit)

VINAYAKA R. PRASAD AND STEPHEN P. GOFF

Departments of Biochemistry and Molecular Biophysics and of Microbiology, Columbia University, College of Physicians and Surgeons, 630 West 168th Street, New York, NY 10032

Communicated by Richard Axel, January 27, 1989

ABSTRACT **A plasmid construct expressing the p66 version of the human immunodeficiency virus reverse transcriptase as a bacterial fusion protein was subjected to *in vitro* mutagenesis, and the resulting variant proteins were assayed to define the locations of the two major enzymatic activities. The DNA polymerase activity was localized to the N-terminal portion of the protein; mutations altering or eliminating the C-terminal portion had little or no effect on that activity. The results suggest that, in contrast with previous reports, the p51 subunit found in virions should exhibit DNA polymerase activity. Mutations in many parts of the protein eliminated RNase H activity, suggesting that several areas are needed for proper folding and generation of that activity.**

The reverse transcriptase (RT) encoded by most mammalian retroviruses is a monomeric protein of approximately 80 kDa (reviewed in ref. 1). The enzyme is bifunctional: the single polypeptide exhibits both a DNA polymerase activity that is active on RNA or DNA templates and an associated RNase H activity that is able to degrade RNA only in RNA·DNA hybrid form (2–5). These two activities have been predicted (6) and indeed demonstrated (7) to reside in separate domains, with the DNA polymerase lying at the N terminus and the RNase H at the C terminus. The RT of the human immunodeficiency virus type 1 (HIV-1) (8, 9), the etiologic agent of acquired immunodeficiency syndrome (AIDS), also exhibits both enzymatic activities, and it has several distinctive biochemical properties in addition (10–12). The structure of the HIV-1 RT is unusual in that the enzyme contains at least two subunits: p66, exhibiting both DNA polymerase and RNase H activity, and p51, reported as exhibiting neither activity (13–16). Recently a third subunit, p15, has been identified and described as having RNase H activity (16). The structure of the subunits suggests that the enzyme is initially formed as a homodimer of p66 subunits, and that one subunit is cleaved to form the p51 and p15 molecules. The functional consequences of the cleavage and the localization of activities to the resulting subunits remain unclear. Ultimately a deeper understanding of the working of this enzyme is critical to the development of specific and potent antiviral drugs to block replication of the virus and to retard progression of the associated immunodeficiency (17, 18).

We have previously constructed a plasmid that expresses the region corresponding to the p66 subunit of the HIV RT encoded by the HXB2 provirus (19), as a trpE–RT fusion protein (20, 21). The fusion protein is enzymatically active and displays both the DNA polymerase and RNase H activities found in the natural enzyme. To localize the two activities on the molecule and to define the regions critical for these activities, we have now generated a library of mutations in the expression plasmid and assayed the resulting mutant proteins for their enzymatic functions.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Construction of Linker Insertion Mutants. The parent construct for all the manipulations was pHRTRX2, a plasmid encoding a trpE-RT fusion with high stability and enzyme activity (21). Enzymes (New England Biolabs) were used according to the specifications of the manufacturer in standard protocols (22). The DNA was partially cleaved with one of a selected set of restriction enzymes (*Alu* I, *Rsa* I, *Dde* I, *Hin*fI, *Nla* IV, or *Hae* III), and the linear molecules resulting from a single cut were purified by agarose gel electrophoresis. When necessary the ends were blunted by treatment with the Klenow fragment of DNA polymerase I, and oligonucleotide linkers containing the *Eco*RI cleavage site were ligated to the termini. Linker sequences (5′-3′) were GAATTC-GAATTC and CGCGAATTCGCG. The DNAs were cleaved with *Eco*RI, cyclized with T4 DNA ligase, and used to transform *Escherichia coli* strain HB101 to ampicillin resistance. DNA was isolated from various clones and screened for the presence of an *Eco*RI cleavage site, and the position of each insertion was determined by restriction mapping and nucleotide sequencing. In some clones tandem copies of the linker were found to be present.

Derivation of Additional Variant Plasmids from Linker Insertion Mutants. Selected mutations were expanded by cleavage with *Eco*RI, filling in the four-base recessed termini with the Klenow fragment of DNA polymerase I, and inserting a new 8-mer linker (GGAATTCC). Mutations denoted "TLI" were generated by cleavage of a selected linker insertion mutant with *Eco*RI plus *Xba* I to remove the C-terminal fragment, blunting with the Klenow fragment of DNA polymerase I, and closure with insertion of an *Xba* I terminator linker (CTAGTCTAGACTAG). Mutants denoted "TL" contained a terminator codon generated by the original *Eco*RI linker insertion, and those denoted "FS" contained a gratuitous loss of bases at the site of insertion, forming a frameshift mutation. Mutants ΔR1/R1 and R1/R1 30 were formed by exchanging the matching *Eco*RI/*Hpa* I fragments between the two mutants R6 and A2. Mutants denoted "PDD" were formed by cleavage of the indicated insertion mutant by *Bam*HI plus *Eco*RI, filling in with Klenow fragment, and closure with *Eco*RI linkers of the appropriate length to conserve the reading frame.

Enzyme Assays. Growth of bacteria, induction of the *trp* operon, lysis with lysozyme and detergent, and preparation of extracts were all as described previously (23). Determi-

Abbreviations: RT, reverse transcriptase; HIV, human immunodeficiency virus.

nation of the levels of RNA-dependent DNA polymerase activity, measured on the synthetic homopolymer substrate oligo(dG)·poly(rC), were as before (20, 24), using 1 μl and 0.1 μl of crude lysates for most assays. RNase H assays of proteins after separation by SDS/polyacrylamide gel electrophoresis (25) were similar to assays of the Moloney murine leukemia virus enzyme (7), but incubations of the gel were in buffer containing 5 mM $MgCl_2$ rather than 2 mM $MnCl_2$.

RESULTS

Construction of Mutants and Analysis of DNA Polymerase Activity. As parent plasmid for mutagenesis, we selected pHRTRX2 (21), encoding an extremely stable trpE–RT fusion protein, 110 kDa in size and containing the complete p66 region of the HIV RT. To introduce defined localized alterations in the HIV RT molecule, we created a battery of *Eco*RI linker insertion mutations at various positions along the RT coding region. A total of 21 distinct mutations with net insertions ranging in length from 6 to 24 base pairs (bp) were obtained and mapped; the amino acid changes generated by each insertion are summarized in Table 1. To test the variants for activity, cultures of *E. coli* containing each mutant plasmid were starved for tryptophan to induce expression of the fusion protein, and the cells were harvested and lysed. Crude lysates were assayed for the stable accumulation of the fusion protein by SDS/polyacrylamide gel electrophoresis and for the level of RNA-dependent DNA polymerase activity as measured on homopolymer substrates [poly(rC) primed with oligo(dG); ref. 24].

The results showed that most of the insertions mapping to the N-terminal part of the protein abolished DNA polymerase activity, while most mapping to the C-terminal part did not (examples shown in Fig. 1; summarized in Fig. 2). These results are similar to those obtained for the Moloney murine leukemia virus RT (7), and they suggest that the DNA polymerase function is similarly localized to the N-terminal region. There were, however, several exceptional mutations. First, two mutations near the C terminus significantly reduced polymerase activity; a similar mutation has been found in the C terminus of the murine RT (7). Second, several mutations in the N-terminal region, including an isolated one (R3) and a cluster of insertions near the center of the protein, were tolerated with little loss of activity (Fig. 2). Increasing the size of two of these insertions from 6 to 18 bp (mutants R3 and R6) did not affect the enzyme activity (Fig. 2), although detailed analysis of these mutant proteins revealed an increased thermolability relative to the parent (data not shown). Thus, we suggest that these parts of the protein are not critical for catalytic activity; perhaps they lie on the surface of the protein, where insertion of additional sequence is easily accommodated.

Table 1. Structure of insertion mutations in HIV RT

Mutant	Restriction site*	Net size of insertion, bp	Amino acid changes	
			From	To
Ha1	2634	12	G	G A N S R
Ha2	2653	12	W P	W R E F A P
Ha3	2734	24	G P	G R E F A R E F A P
R3-6	2785	6	S T	S E F T
R3-18	2785	18	S T	S E L E F Q F T
R4	2904	6	V	V N S
D2	2939	9	L D	L E F L D
D3	3083	15	L	L E F E F L
D4	3158	9	L E	L E F L E
R6-6	3303	6	V	V N S
R6-18	3303	18	V	V N W N S N S
A2	3333	12	S	R N S N S
D6-15	3429	15	L R	L T R I R V R
D6-27	3429	27	L R	L T R I G I P I R V R
A3	3481	12	E L	E E F E F L
H2	3506	15	I L	I E F E F I L
R7-6	3522	6	V	V N S
R7-18	3522	18	V	V N W N S N S
Ha4	3585	12	G	G A N S R
H4	3812	9	I P	I E F I P
R8	3860	12	Y	S R I R D
N3	3859	24	Y	S R I R A R I R D
A4	3917	12	A	G I R I P
A5	4034	6	A	G I P
N4	4187	24	V	V A N S R A N S R

*Map position based on nucleotide sequence of proviral DNA, starting with first base pair at the left edge of the left long terminal repeat defined as position 1.

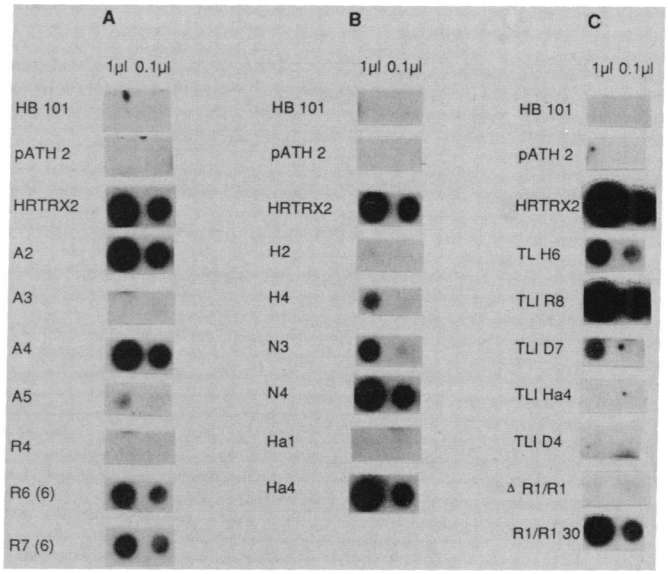

FIG. 1. RNA-dependent DNA polymerase assays of crude bacterial lysates from HIV RT mutant clones. Cultures were starved for tryptophan to induce expression of the fusion proteins, and lysates were prepared and assayed as described (21); autoradiographic signals reflect labeled DNA synthesized on a synthetic RNA template. For each extract, assays of 1 and 0.1 μl are shown. Examples of clones in three different separate experiments (A, B, and C) are shown; the autoradiographic exposures were different for the sets, but were the same for all samples within a given set. Examples with wild-type activity and with various levels of lesser activity are included. Control cultures with no plasmid (HB101) or with the pATH2 vector alone demonstrate the low level of background activity.

To confirm and extend these results, we generated a series of deletion mutants from the linker insertion mutants (Fig. 2). Analysis of these mutants showed that the C-terminal region can be trimmed substantially without loss of polymerase activity; specifically, constructs that include only the N-terminal 50 kDa of the RT protein retained activity. These results suggest that the p51 subunit of the natural RT contains all the sequence information necessary for polymerase function and should be active. Deletions that intruded further into the p51 region of the protein abolished polymerase activity. Thus, as for the Moloney murine leukemia virus enzyme, virtually all of the N-terminal region of the protein is essential. Interestingly, a small duplication of sequence in the central region was tolerated (mutant R1/R1 30); this further suggests that this position lies on the protein surface. Further analysis showed that the observed activity did not result from loss of the insert by homologous recombination.

Analysis of Mutant Proteins for RNase H Activity. All of the mutants were also assayed for RNase H activity by using an *in situ* gel electrophoresis assay (7, 25). Radiolabeled RNA in

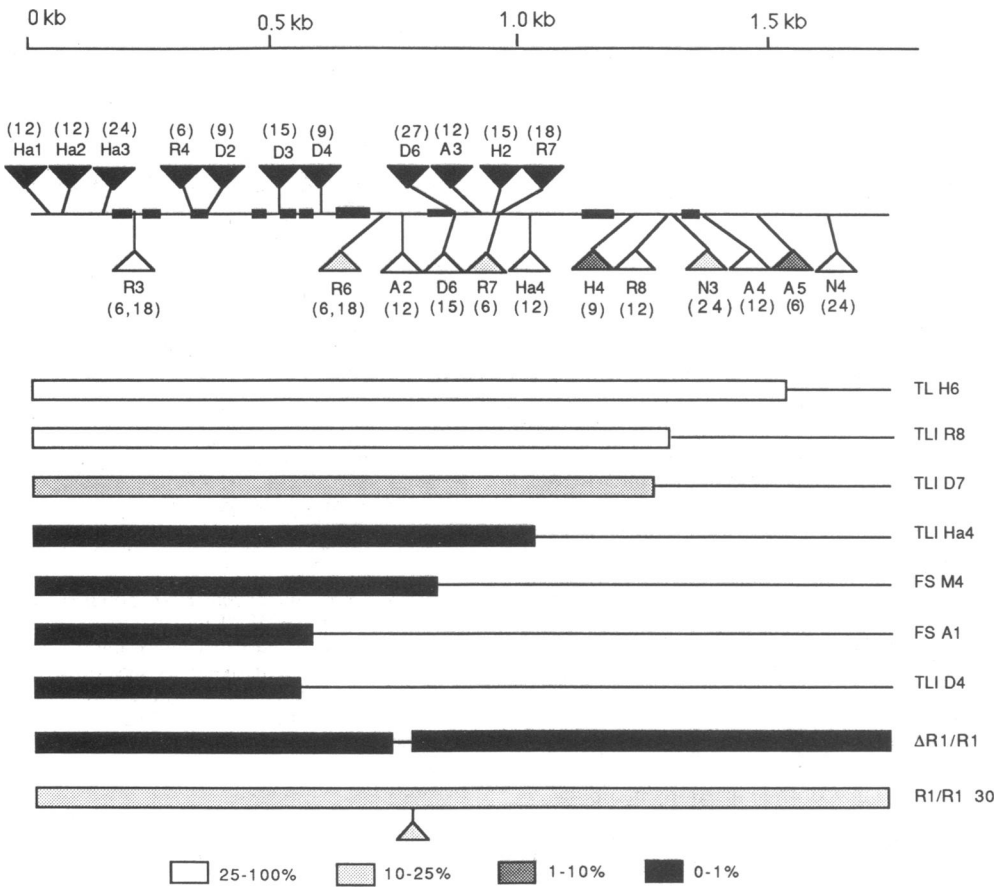

FIG. 2. Positions of linker insertion and deletion mutations and the effects of these mutations on DNA polymerase activity. The line at the top represents the sequences of the HIV *pol* gene encoding the p66 RT (nucleotides 2583–4262 of clone HXB2, numbered from the left edge of the left long terminal repeat); black boxes on the line indicate patches of close sequence similarity to the Moloney murine leukemia virus RT [ref. 26 and our alignment (unpublished data)]. The position of each triangle indicates the site of a linker insertion mutation; the size of each insertion (in bp) is indicated in parentheses. Lower lines indicate the structures of proteins encoded by various deletion mutants; boxed areas represent retained sequences and thin lines represent missing sequences. The level of DNA polymerase activity demonstrated by each mutant is indicated by the shading of the triangle or box, ranging from wild type (white) to inactive (black), as keyed at the bottom.

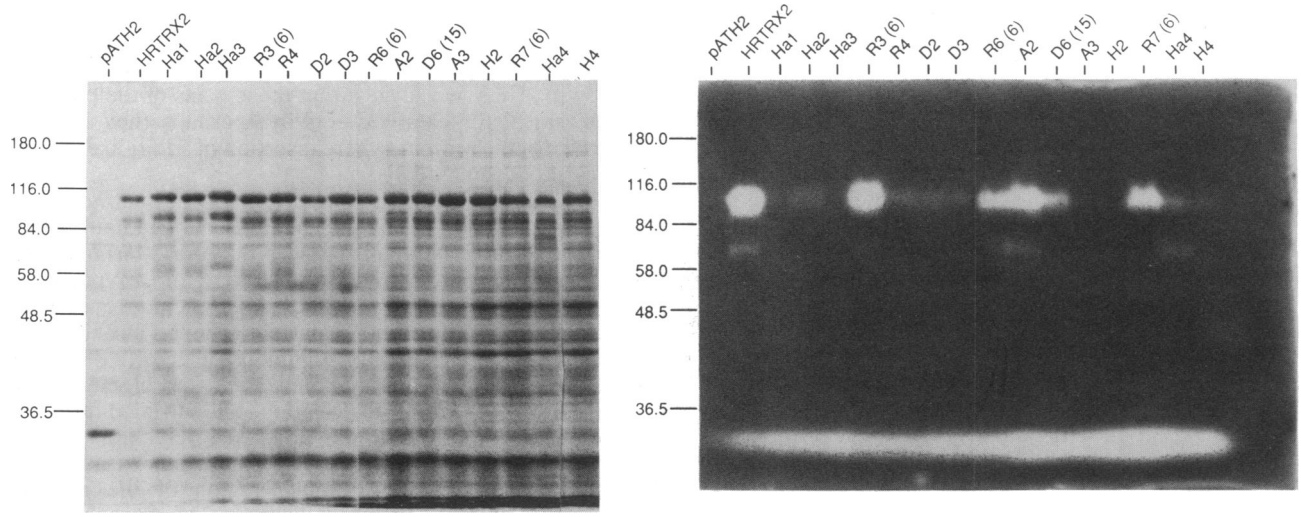

FIG. 3. *In situ* gel assay for RNase H activity of HIV RT mutant clones. (*Left*) Coomassie blue-stained SDS/polyacrylamide gel of crude extracts from mutant clones. Positions of molecular mass markers are given on the left in kDa. Each mutant expressed equal amounts of a stable fusion protein of the same size as the wild-type protein, 110 kDa. (*Right*) Autoradiogram of an SDS/polyacrylamide gel demonstrating the detection of RNase H activity. A gel was cast with ^{32}P-labeled RNA in RNA·DNA hybrid form, and extracts were applied and run as in *Left*. After incubation to permit renaturation of the protein and digestion of the substrate, the gel was fixed, dried, and exposed to x-ray film overnight. The positions of migration of molecular mass markers are indicated in kDa. The large zone of activity at the bottom of the gel is due to the bacterial RNase H enzyme.

RNA·DNA hybrid form was uniformly distributed in the gel matrix before the acrylamide was polymerized, extracted proteins were applied and fractionated by electrophoresis, and the enzyme was allowed to renature and digest the substrate in the gel. Active species were identified by their ability to form zones of clearing in the uniform label, visualized by autoradiography (Fig. 3 *Right*). The results of analysis of the linker insertion and deletion mutants permit only tentative localization of the RNase H activity (Fig. 4). Many insertions in the N-terminal RT region were tolerated, and most insertions in the C terminus abolished RNase H activity; the results are consistent with the idea that much of the C terminus is important for RNase H function. But there were many exceptional insertion mutations in the N terminus which did affect RNase H, and others in the C terminus which did not. Surprisingly, deletions affecting either the N or the C terminus of the protein abolished activity, and only those proteins with gross mutations that were restricted to the center of the protein retained activity. In this assay the protein is initially denatured and must renature to recover an active site; thus, we can conclude that alterations in many parts of the RT protein, at both N and C termini, affect the efficient recovery of this activity. We should note that the presence of the trpE sequences at the N terminus, although profoundly stabilizing the DNA polymerase activity of the fusion protein, might actually hinder the renaturation of many potentially active mutant proteins.

DISCUSSION

The analysis presented here suggests that the general organization of the HIV-1 RT is similar to that of other retroviral RTs. In particular, the RNA-dependent DNA polymerase function is localized to the N-terminal region as predicted (6) and demonstrated (7) for the murine leukemia virus RT. Most importantly, this analysis suggests that both the p51 and p66 subunits should have that activity. Previous failures to detect activity in the p51 subunit (13–16) or in recombinant proteins with deletions in the C terminus (27, 28) could be due to the proteins' lability or inability to fold properly; the expression of the protein as a fusion protein here might be responsible for its stabilization. Recent results comparing the HIV RT generated in bacteria without an N-terminal leader and pHRTRX2 show that the presence of the trpE leader does indeed increase profoundly the thermostability and ability to renature of the DNA polymerase activity of the molecule (D. Helland and W. A. Haseltine, personal communication). We consider the alternative, that the trpE domain itself contributes DNA polymerase activity, to be unlikely. The results here are supported by a very recent report that the p51 protein does indeed exhibit at least some polymerase activity (29), although our results show that appropriate C-terminal mutations can exhibit wild-type levels of the polymerase. The HIV RT in this analysis exhibited one noteworthy difference from the Moloney murine leukemia virus enzyme: several

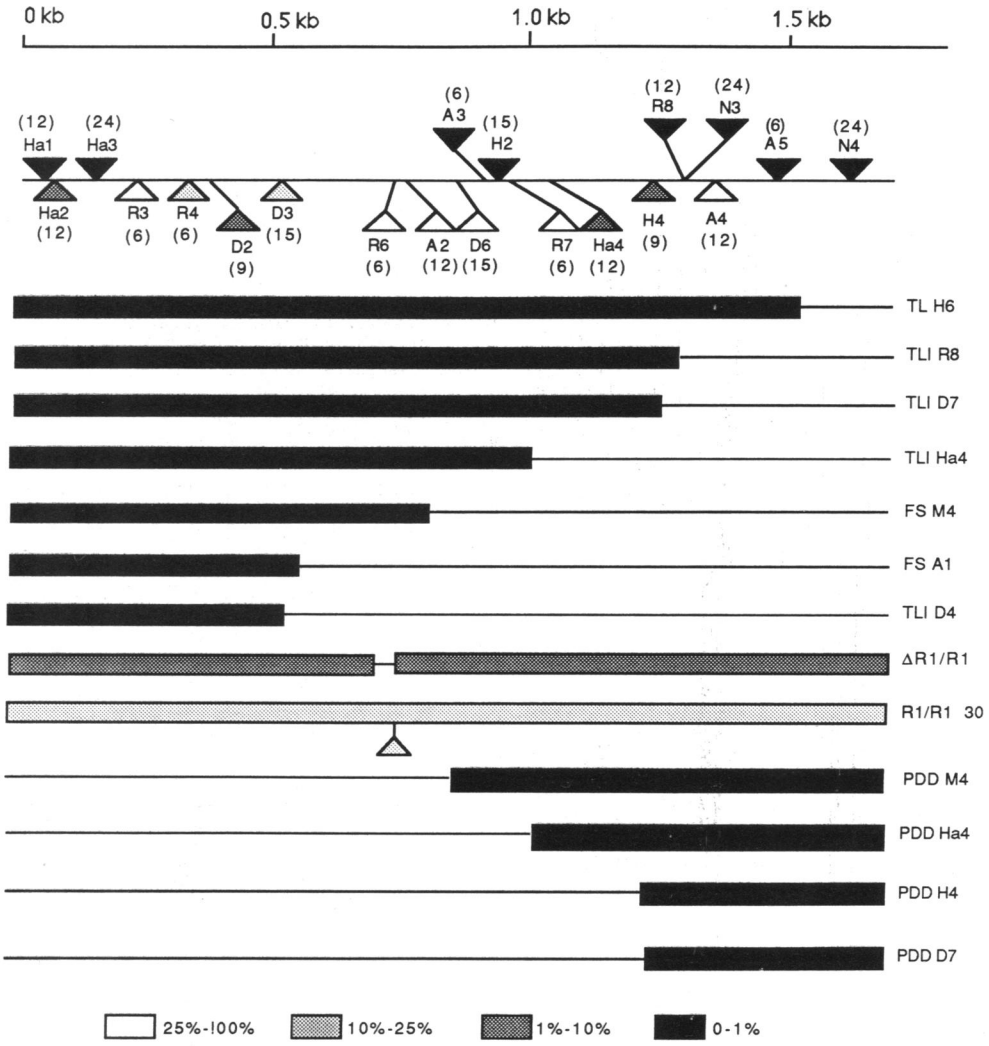

FIG. 4. Effects of linker insertion and deletion mutations on RNase H activity of HIV RT. The position and sizes of the various mutations are indicated as in Fig. 2. The levels of RNase H activity of each mutant, as judged from the *in situ* gel assay, are indicated by the shading.

insertions in the center of the HIV protein were well tolerated, suggesting that the enzyme is more stable to such alterations or contains a relatively nonessential region.

The RNase H domain could not be so simply localized, but other reports suggest that it may be restricted to the C-terminal part of the p66 and indeed localized in the p15 subunit (16). This work is fully consistent with that view. The results, taken together, suggest that the HIV RT may be formed as a p66 dimer, and that one of the two molecules is normally cleaved to yield p51, a DNA polymerase, and p15, an RNase H, thereby separating the two enzymatic activities. The selective advantage to the virus for such cleavage, not occurring in any other group of retroviruses, remains unknown. Furthermore, the protease responsible remains unclear. Cleavage could be due to a variety of host proteases, or breakage at this site could even be autocatalytic.

The constructs generated by our laboratory (20, 21) and many others (27, 30–33) that permit the expression of active HIV RT in microorganisms provide rapid access to large quantities of protein. Batteries of characterized mutants, including these and other types prepared by others (28, 34), will be useful for a variety of studies. Coupled to more detailed structural information, their analysis will facilitate the identification of active sites and domains. In this regard the localization of a highly flexible region, in the middle of the primary sequence of the DNA polymerase domain, is intriguing. The mutants may also be helpful as parents for the isolation of second-site revertants, which could ultimately identify intramolecular contacts in the folded protein.

This work was supported by U.S. Public Health Service Grant AI 24845 and by American Foundation for AIDS Research Grant 421.

1. Verma, I. M. (1977) *Biochim. Biophys. Acta* **473,** 1–38.
2. Moelling, K., Bolognesi, D. P., Bauer, W., Busen, W., Passman, H. W. & Hausen, P. (1971) *Nature (London) New Biol.* **234,** 240–243.
3. Verma, I. M. (1975) *J. Virol.* **15,** 843–854.
4. Gerard, G. F. & Grandgenett, D. P. (1975) *J. Virol.* **15,** 785–797.
5. Moelling, K. (1976) *J. Virol.* **18,** 418–425.
6. Johnson, M. S., McClure, M. A., Feng, D.-F., Gray, J. & Doolittle, R. F. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7648–7652.
7. Tanese, N. & Goff, S. P. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1777–1781.
8. Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Danguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. & Montagnier, L. (1983) *Science* **220,** 868–870.
9. Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, R., Kalyanaraman, V. S., Mann, D., Sidhu, G. D., Stahl, R. E., Zolla-Pazner, S., Leibowitch, J. & Popovic, M. (1983) *Science* **220,** 865–868.
10. Hoffman, A. D., Banapour, B. & Levy, J. A. (1985) *Virology* **147,** 326–335.
11. Rey, M. A., Spire, B., Dormont, D., Barre-Sinoussi, F., Montagnier, L. & Chermann, J. C. (1984) *Biochem. Biophys. Res. Commun.* **121,** 126–133.
12. Cheng, Y.-C., Sutschman, G. E., Bastow, K. F., Sarngadharan, M. G. & Ting, R. Y. C. (1987) *J. Biol. Chem.* **262,** 2187–2189.
13. DiMarzo Veronese, F., Copeland, T. D., DeVico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C. & Sarngadharan, M. G. (1986) *Science* **231,** 1289–1291.
14. Lightfoote, M. M., Coligan, J. E., Folks, T. M., Fauci, A. S., Martin, M. A. & Venkatesan, S. (1986) *J. Virol.* **60,** 771–775.
15. Starnes, M. C., Gao, W., Ting, R. Y. C. & Cheng, Y.-C. (1988) *J. Biol. Chem.* **263,** 5132–5134.
16. Hansen, J., Schulze, T., Mellert, W. & Moelling, K. (1988) *EMBO J.* **7,** 239–243.
17. Chandra, P. (1974) *Top. Curr. Chem.* **52,** 99–139.
18. Chandra, P., Vogel, A. & Gerber, T. (1985) *Cancer Res. (Suppl.)* **45,** 4677s–4684s.
19. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Joseph, S. F., Doran, E. R., Rafalski, J. A., Whitwhoen, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. S., Papas, T. K., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313,** 277–284.
20. Tanese, N., Sodroski, J., Haseltine, W. & Goff, S. P. (1986) *J. Virol.* **59,** 743–745.
21. Tanese, N., Prasad, V. & Goff, S. P. (1988) *DNA* **7,** 407–416.
22. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
23. Tanese, N., Roth, M. & Goff, S. P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4944–4948.
24. Goff, S. P., Traktman, P. & Baltimore, D. (1981) *J. Virol.* **38,** 239–248.
25. Rucheton, M., Lelay, M. N. & Jeanteur, P. (1979) *Virology* **97,** 221–223.
26. Xiong, Y. & Eickbush, T. H. (1988) *Mol. Cell. Biol.* **8,** 114–123.
27. Larder, B., Purifoy, D., Powell, K. & Darby, G. (1987) *EMBO J.* **6,** 3133–3137.
28. Hizi, A., McGill, C. & Hughes, S. H. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1218–1222.
29. Tisdale, M., Ertl, P., Larder, B. A., Purifoy, D. J. M., Darby, G. & Powell, K. L. (1988) *J. Virol.* **62,** 3662–3667.
30. Hansen, J., Schulze, T. & Moelling, K. (1987) *J. Biol. Chem.* **262,** 12393–12396.
31. Farmerie, W. G., Loeb, D. D., Casavant, N. C., Hutchinson, C. A., III, Edgell, M. H. & Swanstrom, R. (1987) *Science* **236,** 305–308.
32. Le Grice, S. F. J., Beuck, V. & Mous, J. (1987) *Gene* **55,** 95–103.
33. Le Grice, S. F. J., Zehnle, R. & Mous, J. (1988) *J. Virol.* **62,** 2525–2529.
34. Larder, B., Purifoy, D., Powell, K. & Darby, G. (1987) *Nature (London)* **327,** 716–717.