Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 1777-1781, March 1988
Biochemistry

Domain structure of the Moloney murine leukemia virus reverse transcriptase: Mutational analysis and separate expression of the DNA polymerase and RNase H activities

(gene fusion/linker insertion/temperature sensitive)

NAOKO TANESE AND STEPHEN P. GOFF

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

Communicated by Harold S. Ginsberg, November 18, 1987 (received for review September 25, 1987)

ABSTRACT **The reverse transcriptase of Moloney murine leukemia virus, like that of all retroviruses, exhibits a DNA polymerase activity capable of synthesis on RNA or DNA templates and an RNase H activity with specificity for RNA in the form of an RNA·DNA hybrid. We have generated a library of linker insertion mutants of the Moloney murine leukemia virus enzyme expressed in bacteria and assayed these mutants for both enzymatic activities. Those mutations affecting the DNA polymerase activity were clustered in the 5′-proximal two-thirds of the gene, and those affecting RNase H were in the remaining 3′ one-third. Based on these maps, plasmids were made that expressed each one of the domains separately; assays of the proteins encoded by these plasmids showed that each domain exhibited only the expected activity.**

Soon after a retrovirus enters a permissive cell, the virion core directs the synthesis of a double-stranded DNA copy of the encapsidated genomic RNA. The reaction is catalyzed by the virus-encoded enzyme reverse transcriptase, which is carried into the cell within the virion particle (1, 2). Although the overall process of reverse transcriptase is complex, the general outline of the reaction has been deduced by the study of *in vitro* systems (3). Synthesis of the first DNA strand is initiated with a tRNA primer and proceeds by copying RNA to ultimately form full-length viral DNAs. In the course of first-strand DNA synthesis, the genomic RNA is degraded (4); that degradation exposes the first DNA strand for use as template and also forms the primer RNA for initiation of the second DNA strand (5). The final product is a duplex DNA free of RNA, after the primer RNAs at the 5′ ends of each DNA strand are removed (6, 7). Reverse transcriptase enzymes purified from virion cores exhibit many of the activities thought to be needed for these reactions (8). In particular, all known reverse transcriptases exhibit two catalytic activities: a DNA polymerase activity and an associated RNase activity (9–11). The DNA polymerase activity can extend the 3′ end of a primer and copy either RNA or DNA templates, as is required to form the first and second strands of the viral DNA. The RNase activity, termed RNase H, degrades RNA only when it is in the form of an RNA·DNA hybrid duplex; single-stranded RNA and RNA·RNA duplexes are resistant to degradation (9, 10, 12). This activity can account for the degradation of the viral RNA, for second-strand primer formation, and for primer removal.

All reverse transcriptases are formed by proteolytic processing from a large polyprotein precursor (13), but the mature enzymes from different virus families show different subunit structures. The enzyme of avian viruses is an α-β heterodimer; the larger β subunit contains all of the sequences in the α subunit but includes extra sequences at the C terminus (14). Chromatographic separation of the subunits has shown that each subunit has both DNA polymerase and RNase H activities (15). The enzyme from murine and feline viruses is a simple monomer that also exhibits both activities. A number of mutants have been isolated that form altered enzymes, and in many cases both activities are coordinately affected (16–18). Thus, the two activities appeared to be intimately associated, and the separate isolation of a protein with DNA polymerase activity without the RNase H activity has not heretofore been achieved. Nevertheless, there have been hints that the two activities may be separable. Limited proteolysis has permitted the recovery of fragments exhibiting RNase H activity in the absence of DNA polymerase (19), and some chemicals exhibit differential inhibitory activity (20–23).

Recently, a computer-aided comparison of the amino acid sequences of various reverse transcriptases and those of other enzymes has led to a proposal for the functional organization of the viral proteins (24). The positions of the similarities suggested that the N-terminal part of reverse transcriptase should contain the DNA polymerase activity, whereas the C-terminal part should contain RNase H. To map the two activities and to determine whether the expression of each activity in the absence of its partner could be achieved, we have carried out a mutational analysis of a cloned copy of a viral reverse transcriptase gene (25). These studies show that the two activities are encoded by distinct regions of the gene, and that the computer-based predictions are correct. Furthermore, the two functions must use fully separable active sites, since nonoverlapping proteins can be produced that exhibit either one of the two activities without the other.

MATERIALS AND METHODS

Construction of Mutants. *Escherichia coli* strain HB101 was the host for expression plasmids; CC114 (*lacZ*$^-$ am) was indicator strain for the SuIII$^+$ marker (26). Parent DNAs were pSH1 (25), pB6B15.23 (27), and pSC1 (28). Mutagenesis of plasmid DNAs was by the "suppressor-linker" method (26): plasmid DNA was treated with *Alu* I, *Hae* III, or *Mnl* I for limited times to generate full-length linear molecules, and either of two *Pvu* II suppressor-linker fragments was ligated to these linear molecules. One fragment, derived from plasmid pVSU-11, results in the insertion of the 12-base-pair (bp) sequence 5′ CTGGAATTCCAG 3′ into blunt-ended sites (26). A second fragment, derived from a variant plasmid termed pVSU-12, contained an extra 4 bp on one side of the *Eco*RI sequence and resulted in the insertion of the 16-bp sequence 5′ CTGGAATTCCAGCCAG 3′ or its complement into blunt-ended sites. This

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.

Abbreviation: Mo-MuLV, Moloney murine leukemia virus.

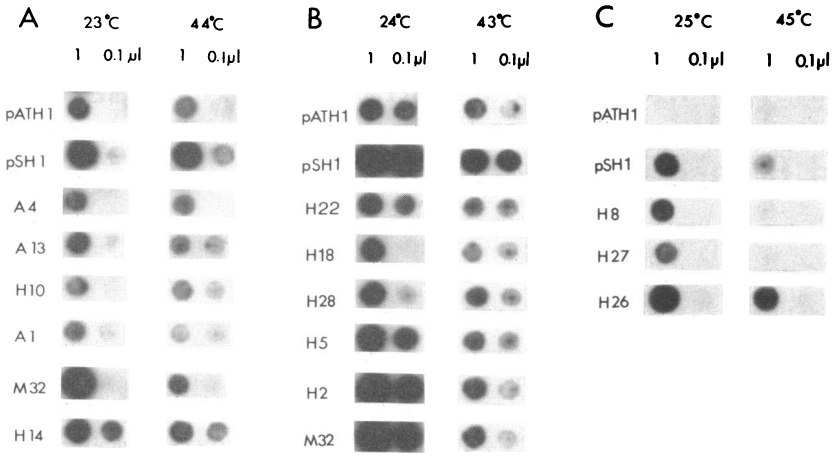

FIG. 1. RNA-dependent DNA polymerase assays of crude lysates of mutant clones. Examples of three separate sets of assays are shown; different sets were derived from very different x-ray film exposures, but samples within each set were from the same film. For each set, control or mutant bacterial cultures were induced, lysates were prepared, and DNA polymerase assays were performed with poly(rA)·oligo(dT) as substrate. For each sample the results of assays on 1 and 0.1 μl of extract at high and low temperatures are shown. Examples are included of mutants with only background (*E. coli* DNA polymerase I) activity (25), thermosensitive activity, and wild-type activity.

fragment was also used to generate net insertions of 15 bp into *Mnl* I sites that yielded 1-bp staggered cuts. Sequence analysis (29) of these mutants revealed that insertions at these sites were only recovered with loss of 1 bp of target DNA, presumably due to removal of the protruding ends by nonspecific nuclease activity in the enzyme preparation.

Construction of Plasmids Expressing Single Activities. Plasmids pRTS7 and pRTS12 were derived by transfer into the

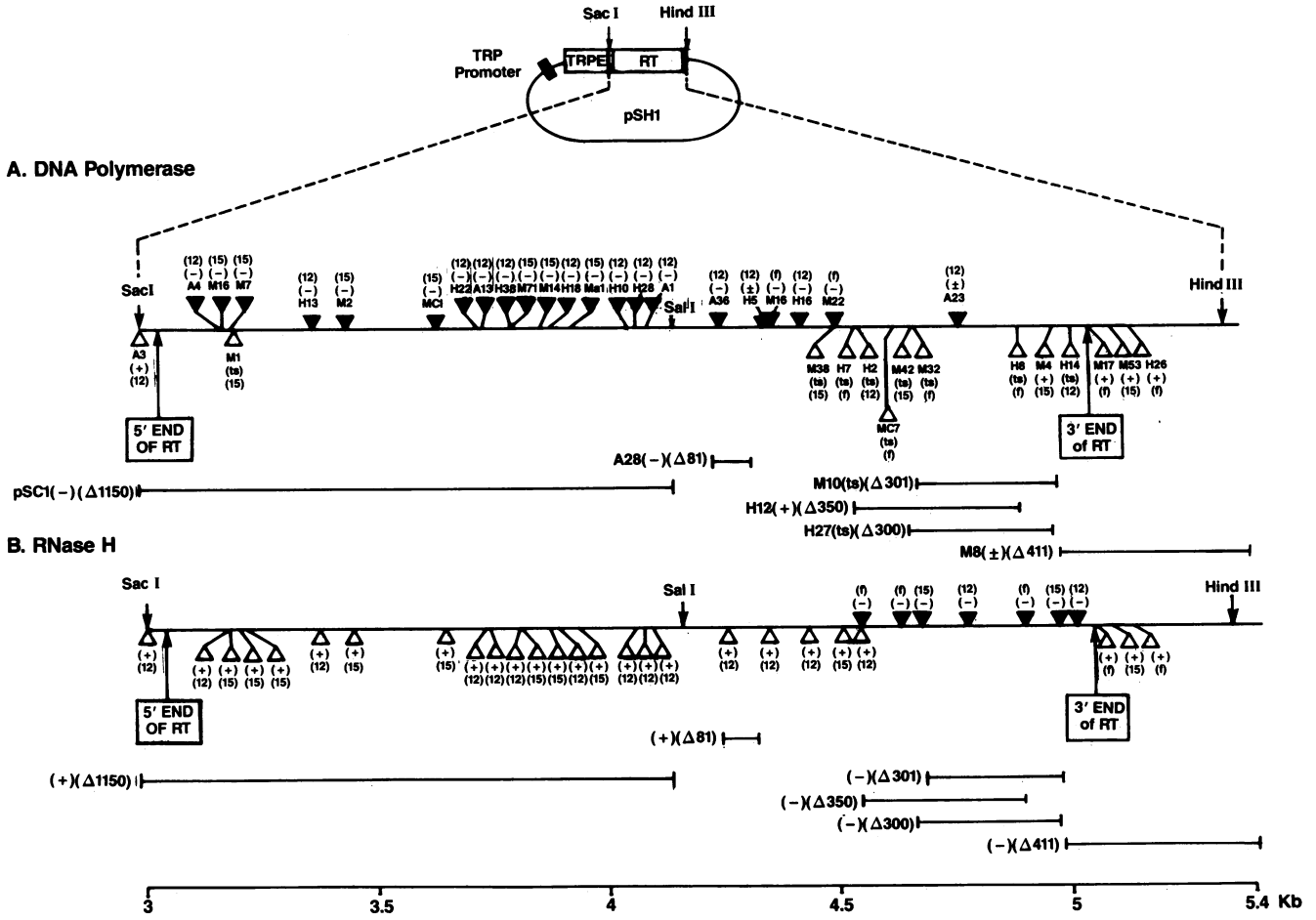

FIG. 2. Positions of linker insertion and deletion mutations and effects of the mutations on DNA polymerase (*A*) and RNase H activities (*B*). The structure of the parent plasmid, pSH1 (25), is shown at the top. An expanded view of the coding region for reverse transcriptase (RT) is shown below in both *A* and *B*. The scale gives the map position in the genome of Mo-MuLV as measured from the 5′ edge of the long terminal repeat of the proviral form. The position of each linker insertion is indicated by the position of a triangle. Solid triangles denote little (±) or no (−) activity, and open triangles denote wild-type (+) or substantial levels of temperature-sensitive (ts) activity. Numbers in parentheses indicate the number of base pairs inserted at each site. Frameshift mutations are indicated by (f). The name of each mutant is given above the triangles in *A*. The position and extent of each deletion mutation is indicated by the bars below each map. The approximate sizes of the deletions (in bp) are indicated in parentheses. Kb, kilobases.

expression construct pB6B15.23 of a *Sal* I/*Pst* I fragment from mutant plasmids pH7 and pH12, respectively, by standard procedures. pRHST1 was derived from pH2 by cleavage with *Eco*RI, blunting with mung bean nuclease, recleavage with *Sma* I, and cyclization with ligase. pRHSR1 was derived from pH2 by cleavage with *Eco*RI, blunt ending with avian myeloblastosis virus reverse transcriptase, recleavage with *Nru* I, and cyclization with ligase. pRHSK1 was made identically to pRHSR1, except that the *Eco*RI ends were blunt-ended with the Klenow fragment of DNA polymerase I. pRHSK11 was derived from pRHSK1 by replacement of the *Bgl* II/*Hind*III fragment by the corresponding fragment from mutant pM17.

Enzyme Assays. RNA-dependent DNA polymerase assays of crude bacterial extracts were performed with oligo(dT)·poly(rA) substrate (30). Tests for thermolability were made by preincubation of extracts in reaction buffer for 1 hr at 25°C or 45°C, followed by assay for 15 min at the same temperature, initiated by addition of substrates. *In situ* RNase H assays were essentially as described (31). The RNA·DNA hybrid with ^{32}P-labeled RNA was prepared by incubation of M13 single-stranded DNA with [α-^{32}P]UTP, three unlabeled triphosphates, and RNA polymerase holoenzyme. The hybrid was isolated by chromatography on Sephadex G-50. NaDodSO$_4$/polyacrylamide gels (32) were copolymerized with 0.7 μg of substrate (5×10^6 cpm/μg) in a 9% polyacrylamide resolving gel (20 ml). After electrophoresis of the insoluble fraction of crude bacterial lysates, the gel was soaked in buffer (50 mM Tris·HCl, pH 8.0/50 mM NaCl/2 mM dithiothreitol/2 mM $MnCl_2$) at room temperature for 2–4 days, with several changes of buffer to permit renaturation of the proteins and digestion of the substrate, and exposed to x-ray film.

RESULTS

Construction of a Library of Linker Insertion Mutants of Moloney Murine Leukemia Virus (Mo-MuLV) Reverse Transcriptase Expressed in *E. coli*. We have previously described the construction of gene fusions between the bacterial *trpE* gene and the *pol* gene of Mo-MuLV that synthesize active fusion proteins exhibiting both DNA polymerase and RNase H activities (25, 27). To localize the regions required for each of these activities, we selected one such construct, pSH1, as the parent and generated a library of linker insertion mutants. Insertions of 12, 15, and 16 bp, all containing an *Eco*RI cleavage site, were made at the positions of cleavage by *Alu* I, *Hae* III, and *Mnl* I. The locations of the inserts were determined by restriction mapping and in some cases by nucleotide sequencing. Extracts were prepared from induced bacterial cultures containing each of the mutant plasmids (25), and the ability of the construct to form stable proteins of the expected size was confirmed by NaDodSO$_4$/polyacrylamide gel electrophoresis.

Analysis of Mutants for Reverse Transcriptase Activity. Lysates of bacterial cultures expressing each mutant protein were prepared and assayed for DNA polymerase activity on a poly(rA) template primed with oligo(dT) (Fig. 1). Each mutant was scored as retaining some DNA polymerase activity, or as exhibiting no activity above the background seen in controls. Analysis of the behavior of the mutants in comparison with the map position of their insertions gave a clear result (Fig. 2*A*). The majority of mutants with insertions in the N-terminal two-thirds of the protein showed no detectable activity over controls, whereas those with insertions in the C-terminal one-third retained full or at least partial activity. Several mutants with insertions that created frameshift mutations were also tested; those eliminating the C-terminal one-third of the protein retained activity. In addition, three mutants with large deletions affecting the C-terminal region of the protein retained full or partial activity. We conclude that the protein can be divided into two domains with distinctive responses to mutations: that the N-terminal two-thirds is critical for DNA polymerase activity, while the C-terminal one-third is largely or entirely dispensable. Only two clones were exceptional. Clone M1, with an insert very near the N terminus, retained partial polymerase activity. Another, A23, showed only minimal activity, even though its insertion mapped in a C-terminal region that could be deleted without loss of activity. We surmise that this insert was more disruptive of the structure of the protein, or interfered with its folding to a greater extent, than deletion of the entire domain.

We repeated assays of all of the mutants with full or partial activity at 25°C and at 45°C to test for thermolability. A surprising number (11/38), including insertions and large deletions in the C terminus, exhibited temperature-sensitive activity (ts in Fig. 2; examples shown in Fig. 1). These mutations presumably destabilized the whole protein to thermal denaturation and loss of activity.

Analysis of Mutants for RNase H Activity. The Mo-MuLV reverse transcriptase expressed in bacteria, like authentic reverse transcriptases, exhibits RNase H activity specific for RNA·DNA hybrid duplexes (27). To localize the activity on

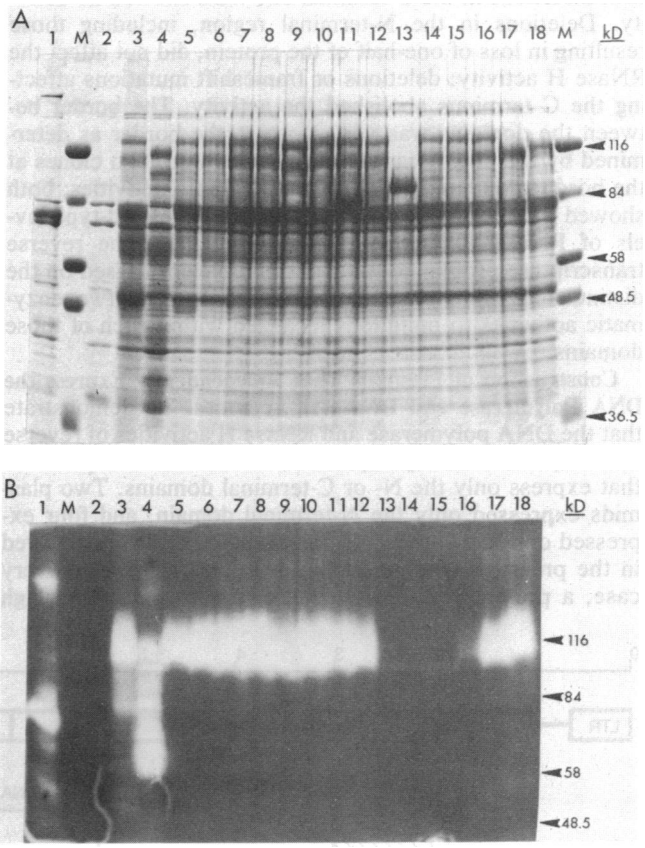

FIG. 3. *In situ* gel assay for RNase H activity. (*A*) Coomassie blue stain of polyacrylamide gel of various mutant extracts. Insertion mutants with a complete coding region show the intact fusion protein of 124 kDa and some smaller species at 110 and 75 kDa; frameshift or deletion mutants show appropriately shorter proteins. (*B*) Autoradiogram of gel run in parallel to that of *A* but preformed with ^{32}P-labeled RNA·DNA hybrid. After incubation for 4 days, the gel was dried and exposed for 2.5 days. Cultures containing the following plasmids are shown, arranged from left to right according to the map position of the mutation. Lanes: 1, pB6B15.23; 2, pATH1; 3, pSH1; 4, pSC1; 5, H13; 6, H22; 7, A13; 8, A1; 9, A28 (deletion of 81 bp); 10, A36; 11, H16; 12, M38; 13, MC7 (frameshift); 14, M42; 15, A23; 16, H14; 17, M17 (frameshift); 18, H26 (frameshift). Molecular mass standards of indicated sizes (in kDa) were run in lanes M.

the protein, we devised an *in situ* polyacrylamide gel assay capable of screening mutant proteins in crude lysates. Standard $NaDodSO_4$/polyacrylamide gels were prepared normally, but radiolabeled RNA in RNA·DNA hybrid form was cast into the resolving gel. Cell lysates were prepared and applied to the gel; after electrophoresis, the gel was incubated in buffer at room temperature to allow renaturation of the protein, digestion of the substrate, and release of the radioactivity. Autoradiography of the dried gel then revealed bands of clearing against the labeled background (Fig. 3). Extracts from control cells containing only the pATH vector plasmid showed no RNase H activity (lane 2), while those containing the pSH1 parent plasmid showed strong activity comigrating with the 124-kDa fusion protein (lane 3). Smaller proteins on these gels, possibly breakdown products, showed little or no activity. Assays of extracts of the mutant clones (examples are shown in lanes 5–18) allowed assignment of each one as active or inactive. A summary of the RNase H activity of each mutant in comparison with the position of its insertion mutation yielded a simple picture, neatly opposite the results for DNA polymerase activity (Fig. 2*B*). All the mutants with insertions in the N-terminal two-thirds of the protein retained RNase H activity; all those with insertions in the C-terminal one-third showed no activity. Deletions in the N-terminal region, including those resulting in loss of one-half of the protein, did not affect the RNase H activity; deletions or frameshift mutations affecting the C terminus abolished the activity. The border between the domains was very close to the border as determined by the polymerase assays. There were two clones at the border, M38 and H2, that retained both activities; both showed thermolabile polymerase activity and wild-type levels of RNase H activity. We conclude that the reverse transcriptase protein consists of two domains, based on the distinct and complementary sensitivities of its two enzymatic activities to mutations localized within each of those domains.

Construction of Plasmids That Independently Express the DNA Polymerase and RNase H Activities. To demonstrate that the DNA polymerase and RNase H activities of reverse transcriptase are fully separable, we constructed plasmids that express only the N- or C-terminal domains. Two plasmids expressed only the N-terminal domain, and four expressed only the C-terminal domain; the constructs differed in the precise endpoints of the domains (Fig. 4). In every case, a protein of the predicted size accumulated to high levels after induction of the *trp* operon, as judged from Coomassie blue staining of gels (Fig. 5*A*).

Lysates of cells carrying pRTS7 and pRTS12, which retained only the N-terminal domain, exhibited high levels of RNA-dependent DNA polymerase activity as assayed on the standard poly(rA)·oligo(dT) template, comparable to that of the parent pB6B15.23 (data not shown). Analysis of these constructs for RNase H activity by the *in situ* gel assay showed no detectable activity (Fig. 5*B*). Lysates of cells carrying pRHST1, pRHSR1, pRHSK1, and pRHSK11, which retained only the C-terminal domain, showed no detectable RNA-dependent DNA polymerase activity over control cultures. Analysis for RNase H activity showed that all four of these proteins exhibited high levels of activity (Fig. 5*B*); the smaller of those proteins showed significantly higher activity than the larger parent constructs. We conclude that the two activities are indeed separable; that the N-terminal two-thirds is necessary and sufficient for DNA polymerase activity; and that the C-terminal one-third is necessary and sufficient for RNase H activity.

DISCUSSION

These mutational studies strongly suggest that the murine reverse transcriptase enzyme can be divided into two domains, each responsible for a distinct enzymatic activity; the fact that constructs containing only one of these domains exhibit only one activity firmly proves this point. The conclusion is that the polymerase activity is N-terminal, and the RNase H activity is C-terminal. This localization of the polymerase and RNase H activities is in agreement with the recent predictions of their locations based on amino acid sequence homologies to related enzymes (24) and is in disagreement with some earlier suggestions (14, 33). The boundary between the domains as defined by the behavior of the mutants falls nicely into the "tether" region proposed in the homology studies to connect the two functions. Our results suggest that for Mo-MuLV there may be little nonfunctional tether sequence, but that most residues in the protein are important for one or the other activity.

The results presented show that the two activities must utilize distinct and nonoverlapping active sites for the binding of nucleic acids and for catalysis. The failure of most of the insertion mutations in one domain of the bifunctional protein to affect the other suggests that interactions between the domains are very limited. Since some of the C-terminal insertions rendered the polymerase activity thermolabile,

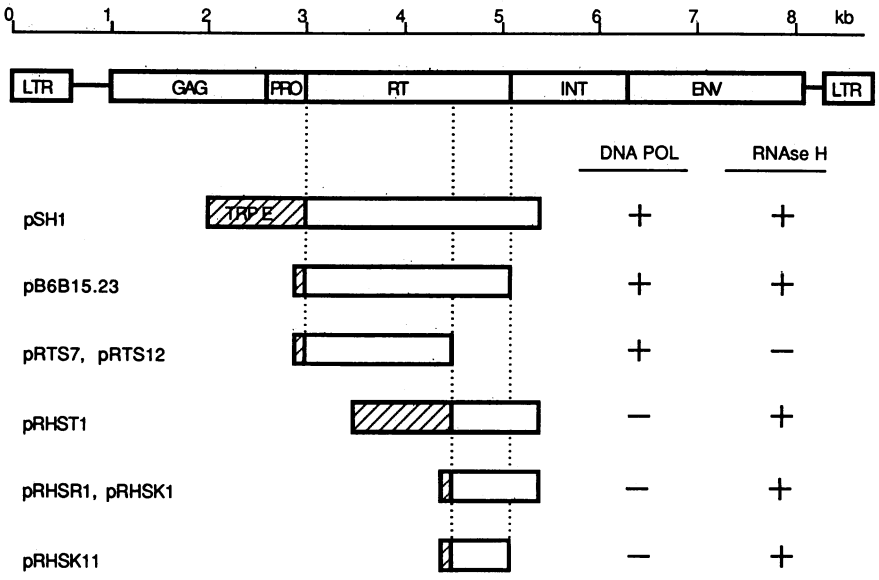

FIG. 4. Separate expression of DNA polymerase and RNase H activities. The structure of the fusion proteins encoded by various plasmids is shown, including *trpE* segments (hatched boxes) and *pol* segments (open boxes). Each protein has been aligned with the corresponding region of the Mo-MuLV genome (*Upper*). The RNA-dependent DNA polymerase (DNA POL) and RNase H activities of the various proteins are indicated on the right. LTR, long terminal repeat; kb, kilobases.

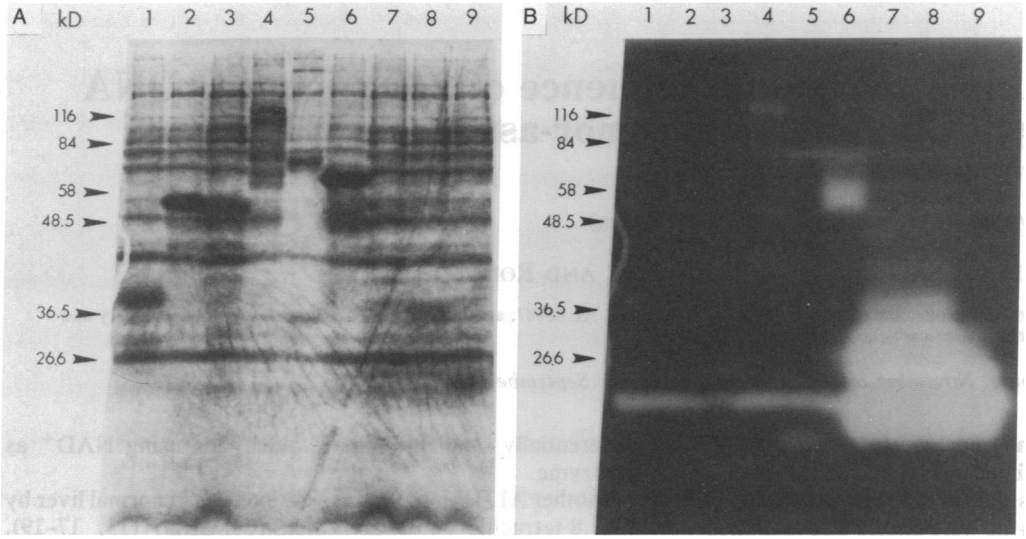

FIG. 5. *In situ* gel assay of various fusion proteins for RNase H activity. (*A*) Coomassie stain; the fusion proteins are readily detected as the major protein in each lane. (*B*) Autoradiogram of gel after digestion of substrate. The band running across the bottom of the gel is probably *E. coli* RNase H. Lanes: 1, pATH1; 2, pRTS7; 3, pRTS12; 4, pSH1; 5, pB6B15.23; 6, pRHST1; 7, pRHSR1; 8, pRHSK1 (similar structure to pRHSR1); 9, pRHSK11. The positions of migration of protein standards of the indicated sizes (in kDa) are marked by arrowheads.

and one such insertion nearly abolished that activity, there must, however, be some interactions between the regions. The situation may well be similar to that for DNA polymerase I of *E. coli* (pol I); here the polymerase and 3′ exonuclease activities also lie in separable domains. A distinction is that the activities of pol I can be separated by proteolysis, which breaks the backbone in an exposed site. Attempts to separate and recover both activities of reverse transcriptase by proteolysis were unsuccessful (19), suggesting that the linkage between domains may be hidden or buried. Recent results suggest that the 66-kDa human immunodeficiency virus reverse transcriptase may be cleaved in this region normally, presumably by the viral protease, to form a 51-kDa protein with no detectable activity and a separate 15-kDa protein.

The utility of linker insertion mutations in determining protein structure is demonstrated by their ease of preparation and mapping and the simplicity of the phenotypes induced. A critical feature of the method is that all of the mutant proteins were resistant to proteolysis in the bacteria and accumulated to high levels after induction. The results suggest that insertions of four or five amino acid residues most often abolish the activity of the affected domain, with only minimal effect on the adjacent domain. This fact seemed to be relatively independent of the amino acids inserted; the insertions of the same oligonucleotide could yield insertions of three different sets of amino acids depending on the translational reading frame at the site of insertion. Judging from our experience with cellular and other retroviral proteins (26), these two enzymes were unusually sensitive to insertions; essentially no silent insertions were recovered.

We have transferred many of the mutations into a complete copy of the viral genome and found that all the mutations were unconditionally lethal to replication (data not shown), suggesting that both activities are important in the life cycle. Preliminary results show that mutant virions, whether defective in polymerase or RNase H activities, do not synthesize full-length proviral DNA upon infection of fresh NIH 3T3 cells. Further analysis of mutants with insertions in the RNase H domain should enable us to determine more precisely the roles of RNase H in reverse transcription.

This work was supported by National Cancer Institute Grant CA 30488.

1. Baltimore, D. (1970) *Nature (London)* **226,** 1209–1211.
2. Temin, H. M. & Mizutani, S. (1970) *Nature (London)* **226,** 1211–1213.
3. Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. (1979) *Cell* **18,** 93–100.
4. Collett, M. S., Dierks, P., Parsons, J. T. & Faras, A. J. (1978) *Nature (London)* **272,** 181–184.
5. Resnick, R., Omer, C. A. & Faras, A. J. (1984) *J. Virol.* **51,** 813–821.
6. Omer, C. A. & Faras, A. J. (1982) *Cell* **30,** 797–805.
7. Champoux, J. J., Gilboa, E. & Baltimore, D. (1984) *J. Virol.* **49,** 686–691.
8. Verma, I. M. (1977) *Biochim. Biophys. Acta* **473,** 1–38.
9. Moelling, K., Bolognesi, D. P., Bauer, W., Busen, W., Passmann, H. W. & Hausen, P. (1971) *Nature (London) New Biol.* **234,** 240–243.
10. Verma, I. M. (1975) *J. Virol.* **15,** 843–854.
11. Gerard, G. F. & Grandgenett, D. P. (1975) *J. Virol.* **15,** 785–797.
12. Moelling, K. (1976) *J. Virol.* **18,** 418–425.
13. Kopchick, J. J., Jamjoon, G. A., Watson, K. F. & Arlinghaus, R. B. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2016–2020.
14. Grandgenett, D., Quinn, T., Hippenmeyer, P. J. & Oroszlan, S. (1985) *J. Biol. Chem.* **260,** 8243–8249.
15. Grandgenett, D. P., Gerard, G. F. & Green, M. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 230–234.
16. Verma, I. M., Mason, W. S., Drost, S. D. & Baltimore, D. (1974) *Nature (London)* **251,** 27–31.
17. Verma, I. M. (1975) *J. Virol.* **15,** 121–126.
18. Lai, M. H., Verma, I. M., Tronick, S. R. & Aaronson, S. A. (1978) *J. Virol.* **27,** 576–581.
19. Lai, M.-H. T. & Verma, I. M. (1978) *J. Virol.* **25,** 652–663.
20. Gorecki, M. & Panet, A. (1978) *Biochemistry* **17,** 2438–2442.
21. Srivastava, A. & Modak, M. J. (1979) *Biochem. Biophys. Res. Commun.* **91,** 892–899.
22. Margalith, M., Falk, H. & Panet, A. (1982) *Mol. Cell. Biochem.* **43,** 97–103.
23. Freeman-Wittig, M.-J., Vinocour, M. & Lewis, R. A. (1986) *Biochemistry* **25,** 3050–3055.
24. Johnson, M. S., McClure, M. A., Feng, D.-F., Gray, J. & Doolittle, R. F. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7648–7652.
25. Tanese, N., Roth, M. R. & Goff, S. P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4944–4948.
26. Lobel, L. I. & Goff, S. P. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4149–4153.
27. Roth, M. J., Tanese, N. & Goff, S. P. (1985) *J. Biol. Chem.* **260,** 9326–9335.
28. Tanese, N., Roth, M. R. & Goff, S. P. (1986) *J. Virol.* **59,** 328–340.
29. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
30. Goff, S., Traktman, P. & Baltimore, D. (1981) *J. Virol.* **38,** 239–248.
31. Rucheton, M., Lelay, M. N. & Jeanteur, P. (1979) *Virology* **97,** 221–223.
32. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
33. Levin, J. G., Hu, S. C., Rein, A., Messner, L. I. & Gerwin, B. I. (1984) *J. Virol.* **51,** 470–478.