Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 1218–1222, February 1988
Medical Sciences

Expression of soluble, enzymatically active, human immunodeficiency virus reverse transcriptase in *Escherichia coli* and analysis of mutants

AMNON HIZI*, CAROLYN MCGILL, AND STEPHEN H. HUGHES

National Cancer Institute–Frederick Cancer Research Facility, Bionetics Research Inc.–Basic Research Program, P.O. Box B, Frederick, MD 21701

Communicated by Richard E. Dickerson, November 16, 1987

ABSTRACT **We have constructed a plasmid that, when introduced into *Escherichia coli*, induces the synthesis of large quantities of a protein with an apparent molecular mass of 66 kDa that differs from human immunodeficiency virus (HIV) RNA-dependent DNA polymerase (deoxynucleoside-triphosphate:DNA deoxynucleotidyltransferase or reverse transcriptase, EC 2.7.7.49) only in that it has two additional amino-terminal amino acids. This protein is soluble in *E. coli* extracts, is active in reverse transcriptase assays, and shows inhibition profiles with dideoxy-TTP and dideoxy-GTP that are indistinguishable from the viral enzyme. The deletion of 23 amino-terminal or carboxyl-terminal amino acids or the insertion of 5 amino acids at position 143 substantially decreases the polymerizing activity of the HIV reverse transcriptase made in *E. coli*. The properties of a 51-kDa reverse transcriptase-related protein made in *E. coli* suggests that the p51 found in the virion probably does not have substantial polymerizing activity. The full-length HIV reverse transcriptase and the various mutant proteins produced in *E. coli* should be quite useful for structural and biochemical analyses as well as for the production of antibodies.**

Retroviruses encode an enzyme, RNA-dependent DNA polymerase (deoxynucleoside-triphosphate:DNA deoxynucleotidyltransferase or reverse transcriptase, EC 2.7.7.49), that is responsible for copying the RNA genome found in the virion into DNA. This copying of viral RNA into DNA is an obligate step in the viral life cycle, and the DNA copy is subsequently inserted into the genome of the infected cell (1). No cellular homolog of viral reverse transcriptases has been discovered, although many normal cells carry endogenous retroviruses and retrotransposons that encode reverse transcriptases. The viral reverse transcriptase is, for this reason, a potential target for drug therapies designed to interfere with the life cycle of retroviruses, most notably the human immunodeficiency virus (HIV) (2–9). The drug 3′-azido-3′deoxythymidine (also called "AZT") is an effective anti viral agent presumably because the derivative triphosphate is recognized as thymidine triphosphate by HIV reverse transcriptase (4–6). Only small amounts of reverse transcriptase are found in virions, and the search for drugs that would interrupt the HIV viral life cycle by interfering with reverse transcription would be facilitated by a better source of the HIV enzyme.

In the virion, the reverse transcriptase is released from a larger polypeptide precursor by proteolytic cleavage (1). To permit the expression of the free HIV reverse transcriptase, we have taken the viral segment encoding the reverse transcriptase and modified the region at the ends of this segment so that there are initiation and termination codons at the positions where proteolytic cleavage takes place. The modified segment was introduced into an *Escherichia coli* expression plasmid. When introduced into *E. coli*, the plasmid with the HIV reverse transcriptase insert induces the synthesis of a large amount (several percent of the total *E. coli* protein) of soluble, enzymatically active, HIV reverse transcriptase. We show here that the ability of the HIV reverse transcriptase produced in *E. coli* to be inhibited by the dideoxy analogs of thymidine and guanosine is identical to the inhibition of virion reverse transcriptase. To study the catalytic domains of the HIV reverse transcriptase, we have begun a mutagenic analysis of this enzyme using the *E. coli* system. The *E. coli*–produced HIV polymerase can be used to search for inhibitors, to facilitate structural and biochemical studies, and to produce both monoclonal and polyclonal antibodies.

MATERIALS AND METHODS

Bacteria and Plasmids. *E. coli* strain DH5 (a derivative of DH-1, see ref. 10) was used as a recipient for DNA transformations (frozen competent cells were purchased from Bethesda Research Laboratories). The HIV DNA polymerase gene constructs were derived from the HIV proviral clone BH10 (11). The construction of the plasmid is given in detail in Fig. 1. *E. coli* DH5 cells transformed with the various plasmids were grown in NZYM liquid broth (25) supplemented with 100 μg of ampicillin per ml.

Virus. Twice-banded HIV-H9 was obtained from Program Resources, Frederick, MD (batch p2627).

Bacterial and Viral Lysis. The bacteria containing either plasmids that express HIV reverse transcriptase or the parental plasmid pUC12N were grown for 12–16 hr with shaking at 37°C. Bacteria were collected by centrifugation in a Microfuge for 2 min at 10,000 rpm, and the pellet was washed once with cold 100 mM NaCl/20 mM Tris chloride/1 mM EDTA, final pH 7.4 (TSE buffer). Bacteria were then disrupted in 0.2 M NaCl/20% (vol/vol) glycerol/1% Triton X-100/1 mM EDTA/2 mM dithiothreitol/25 mM Tris chloride, pH 8.0 (lysis buffer). The lysates were kept at 4°C for 15 min. The insoluble material was removed by centrifugation at 10,000 rpm in a Microfuge. The supernatant was collected and stored at −20°C. No significant decrease in DNA polymerase activity was observed after several freezings and thawings. Usually a bacterial pellet from 1 volume of stationary culture was disrupted in 1/4 volume of the lysis buffer. The protein concentration in the supernatants of the lysed cells (as measured by the Bio-Rad protein assay) was 1–1.5 mg/ml. The insoluble material must be removed after lysis by centrifugation before the RNA-dependent DNA polymerase activity can be specifically assayed in the disrupted bacteria.

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.

Abbreviations: HIV, human immunodeficiency virus; Mo-MuLV, Moloney murine leukemia virus; RSV, Rous sarcoma virus.

*Permanent address: Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

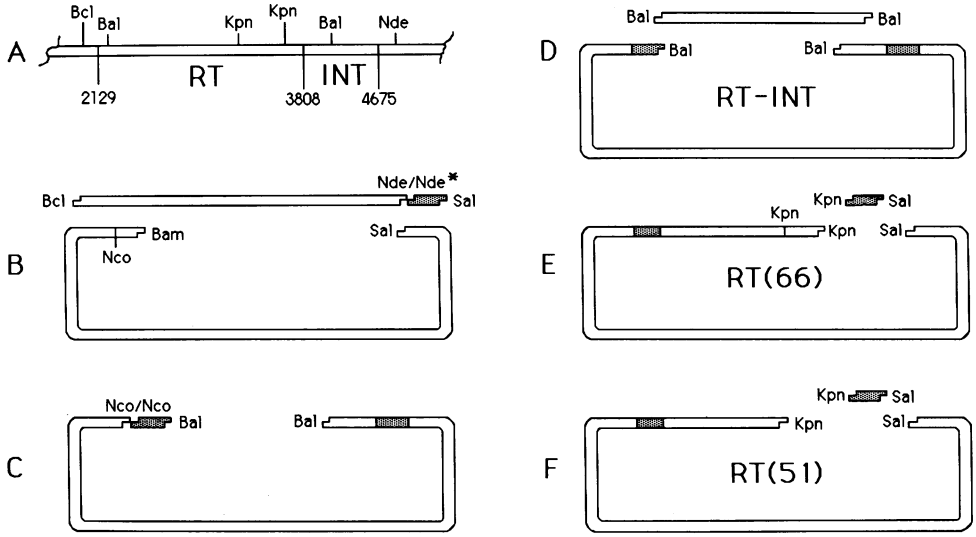

FIG. 1. Construction of the HIV reverse transcriptase expression clone. (*A*) The segment of HIV (11) encoding the reverse transcriptase (RT) and the integrase (INT) is shown. The boundaries of the reverse transcriptase and the integrase coding regions (1) are given, as are the restriction sites used in the construction. The sequential steps in the construction of the pUC12N plasmids expressing the reverse transcriptase-integrase fusion (RT-INT) (*D*), the 66-kDa reverse transcriptase [RT(66)] (*E*), and the 51-kDa reverse transcriptase [RT(51)] (*F*) are given. (*B–F*) DNA segments that derive from the viral genome are shown as open boxes, whereas synthetic DNA segments are stippled.

HIV was similarly disrupted. To one volume of virus suspension, an equal volume of 2× lysis buffer was added, mixed, and kept for 15 min on ice. The whole extract was assayed for enzymatic activity with no further treatment.

Assay for RNA-Dependent DNA Polymerase Activity. Each sample was incubated in a 0.1-ml reaction mixture containing (unless otherwise stated) 75 mM KCl, 8 mM $MgCl_2$, 2 mM dithiothreitol, 25 mM Tris chloride (pH 8.0), and either one of the following mixtures: 5 μg of poly(rA)·oligo(dT) per ml and [α-^{32}P]dTTP or 5 μg of poly(rC)·oligo(dG) per ml and [α-^{32}P]dGTP; the final concentration of the labeled deoxynucleotide triphosphates was 10 μM, and the specific activity was usually 3000–9000 cpm/pmol. In most experiments, 15-μl aliquots of enzyme were assayed for 30 min at 37°C. The amount of polymerized deoxynucleotide triphosphates was assayed by adding salmon sperm DNA carrier (at a final concentration of 100 mg/ml) in 0.2 M sodium pyrophosphate and precipitating the labeled polymer with cold 10% (wt/vol) trichloroacetic acid. The precipitates were collected on Whatman GF/C fiberglass filters by suction filtration. The incorporated triphosphate was measured by assaying for ^{32}P in a liquid scintillation counter. One unit of enzyme is defined as the amount of enzyme that incorporates 1 pmol of dGTP in 30 min at 37°C under standard assay conditions.

Polyacrylamide Gel Electrophoresis. Ten percent polyacrylamide gels containing 0.1% $NaDodSO_4$ were prepared and subjected to electrophoresis as described (12). Bacteria were grown in NZY broth and collected by centrifugation. The pellets were disrupted in gel sample buffer (1% $NaDodSO_4$/1% 2-mercaptoethanol/10% glycerol/50 mM Tris chloride, pH 6.8/10 μg of bromphenol blue per ml) for 2 min at 100°C. Gels were stained with 0.1% Coomassie brilliant blue. In most cases, the protein from 60 μl of bacterial culture was loaded on each gel slot.

Gel Filtration of Lysed Bacteria. The soluble bacterial proteins (prepared as described above for reverse transcriptase assay) were passed over Sephadex G-25 columns to remove potential inhibitors of the enzymatic activity. The supernatants from the bacterial lysates were loaded at 4°C on Sephadex G-25 columns preequilibrated with 0.2 M NaCl/2 mM dithiothreitol/0.2% Triton X-100/20% glycerol/25 mM Tris chloride, pH 7.4. After loading, the columns were washed with the same preequilibration buffer, and the fractions were assayed both for reverse transcriptase activity and for protein concentration. The main peak of enzymatic activity trails the main protein peak. Samples from the most active fractions were used for further analysis and for monitoring the effects of inhibitors. The reverse transcriptase activity recovered from the columns was 10–15 times higher than the activity that was loaded, indicating removal of inhibitors, although there was no substantial increase in the enzyme purity relative to bacterial proteins, as judged from polyacrylamide gel electrophoresis of the samples.

RESULTS

Introduction of Initiation and Termination Sites into the HIV Reverse Transcriptase Gene. Since the HIV reverse transcriptase is normally produced by proteolytic cleavage of the gag-pol polyprotein, the portion of the HIV genome encoding reverse transcriptase does not begin with an initiation codon, nor does it end with a termination codon (1, 11, 13, 14). Before the enzyme can be produced in *E. coli*, an appropriate initiation codon and termination codon must be introduced into the gene. The procedure is described in detail in Fig. 1. Briefly, restriction sites were chosen that lie just inside the HIV reverse transcriptase domain. Synthetic DNA segments were made that replaced both the segment of the gene encoding the amino- and carboxyl-terminal portion of the HIV reverse transcriptase. The segment that replaces the portion of the gene encoding the carboxyl terminus provides a termination codon at precisely the site where proteolytic cleavage takes place (see Fig. 1; see also ref. 15). The carboxyl terminus of the reverse transcriptase is inferred from the known amino terminus of the "integrase" and from the knowledge that for other retroviruses there is no gap between the carboxyl terminus of the reverse transcriptase and the amino terminus of the integrase (1). The synthetic DNA encoding the amino terminus of the HIV reverse transcriptase provides an initiation codon. Protein synthesis initiates with methionine and, to facilitate the constructions, the initiator methionine codon was incorporated into the recognition site for the restriction enzyme *Nco* I, which is CCATGG. In order to retain the ability to excise the HIV reverse transcriptase coding region for other constructions (see *Discussion*), we used a synthetic DNA segment that not only could be ligated to an *Nco* I site but also after ligation would recreate the *Nco* I site. However, the last base of the *Nco* I site (guanosine) defines the first base of the second codon. Because the first amino acid of the mature HIV reverse transcriptase is a proline, it is impossible both to recreate the *Nco* I site and to encode proline as the second amino acid. Given these facts, we have elected to add two amino acids to the amino terminus of the HIV reverse transcriptase, an amino-terminal methionine and a valine as the second amino acid and then to continue with the proline. In some cases, *E. coli* removes one or two amino-terminal amino acids from primary translation prod-

ucts. We have not yet directly examined the amino terminus of the protein that accumulates in *E. coli*.

The coding region for the HIV reverse transcriptase was ligated into the pUC plasmid pUC12N, which differs from pUC12 only in that the *lacZ* initiator ATG has been converted by site-directed mutagenesis to an *Nco* I site (refs. 16 and 17; see Fig. 1). The recombinant plasmid was introduced into *E. coli* strain DH5, and colonies were chosen that were shown to contain the internal segment deriving from HIV reverse transcriptase and the two synthetic DNA segments (see *Materials and Methods*).

Expression of the HIV Reverse Transcriptase in *E. coli*. Colonies containing the appropriate plasmid were grown in NZY broth and harvested by centrifugation. The bacterial pellets were disrupted directly in gel sample buffer, and the proteins were analyzed by $NaDodSO_4$/polyacrylamide gel electrophoresis. The pUC12N plasmid carrying the complete HIV reverse transcriptase coding region induced the synthesis of a protein of ≈66 kDa (see Fig. 2). Control strains that carry the parental plasmid pUC12N did not make this protein. Treatment of the *E. coli* strains that expressed the 66-kDa protein with a buffer similar to that used to release reverse transcriptase from virions (lysis buffer, see *Materials and Methods*) resulted in the release of enzymatically active reverse transcriptase (Table 1).

Foreign proteins synthesized in *E. coli* are often insoluble. Triton X-100 released a large amount of reverse transcriptase enzymatic activity, and we estimated, based upon gel electrophoresis of the soluble and insoluble fractions, that a single extraction with lysis buffer solubilized at least half of the 66-kDa HIV reverse transcriptase from *E. coli* (data not shown).

Inhibitors of the Polymerase. To test whether HIV reverse transcriptase made in *E. coli* has the same inhibition profile as the enzyme found in virions, polymerization assays were done in the presence of increasing amounts of dideoxy-TTP and dideoxy-GTP. Sephadex G-25 was used to remove small molecules from the HIV reverse transcriptase synthesized in *E. coli*.

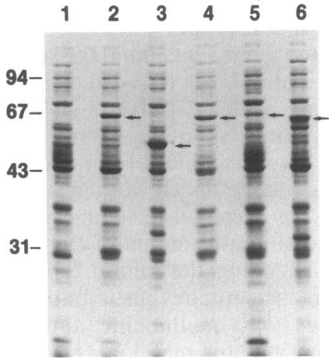

FIG. 2. HIV reverse transcriptase proteins made in *E. coli*. *E. coli* DH5 containing the parental plasmid pUC12N or derivatives that carry wild-type or mutant HIV reverse transcriptase were collected by centrifugation and lysed, and the proteins were fractionated on an $NaDodSO_4$/polyacrylamide gel. Each lane was loaded with protein that derives from 60 μl of liquid culture. Protein was visualized with Coomassie brilliant blue. The position of migration of protein molecular weight markers (in kDa) is given on the left side of the figure. Lanes: 1, proteins from DH5 carrying the parental plasmid pUC12N; 2, proteins from a strain carrying pUC12N with the wild-type HIV polymerase inserted; 3, the mutant CT-133; 4, CT-23; 5, CT-16; 6, AT-23. The mutants RV-1, RV-2, RV-3 and PM-1 all make a 66-kDa protein indistinguishable from the protein induced from the wild-type HIV reverse transcriptase (data not shown). The proteins expected to be synthesized by each mutant are described in Table 2. The HIV reverse transcriptase-related proteins are marked with arrows.

Table 1. Template-primer and divalent-cation preference of the *E. coli*-expressed HIV RNA-dependent DNA polymerase*

Template primer	Divalent cation	pUC12N control	HIV polymerase
Poly(rA)·oligo(dT)	8 mM $MgCl_2$	0.01	5.1
	0.6 mM $MnCl_2$	0.006	0.6
Poly(rC)·oligo(dG)	8 mM $MgCl_2$	<0.005	6.6
	0.6 mM $MnCl_2$	<0.005	2.4

*Activity is expressed as pmol of dTTP [for poly(rA)·oligo(dT)] or dGTP [for poly(rC)·oligo(dG)] per μg of protein of *E. coli* lysate incorporated in 30 min at 30°C.

As a control for the inhibition of the HIV reverse transcriptase, we also tested the inhibitors on Moloney murine leukemia virus (Mo-MuLV) reverse transcriptase synthesized in *E. coli* (A.H. and S.H.H., unpublished data), which also was fractionated on G-25. The HIV reverse transcriptase made in *E. coli* and the HIV reverse transcriptase from the virions responded identically to increasing amounts of dideoxy-GTP or dideoxy-TTP (Fig. 3). The Mo-MuLV polymerase responded differently, particularly to dideoxy-TTP. The HIV enzyme was much more sensitive to dideoxy-TTP than was the Mo-MuLV enzyme (see Fig. 3).

HIV Reverse Transcriptase Mutants. This system also provides a means to explore the effects of mutations on the functions of HIV reverse transcriptase. Although we tested a substantial number of amino-terminal and carboxyl-terminal deletions, *E. coli* lysates did not contain substantial quantities of some of the mutant proteins. We report here only on mutant constructions that induced the synthesis of

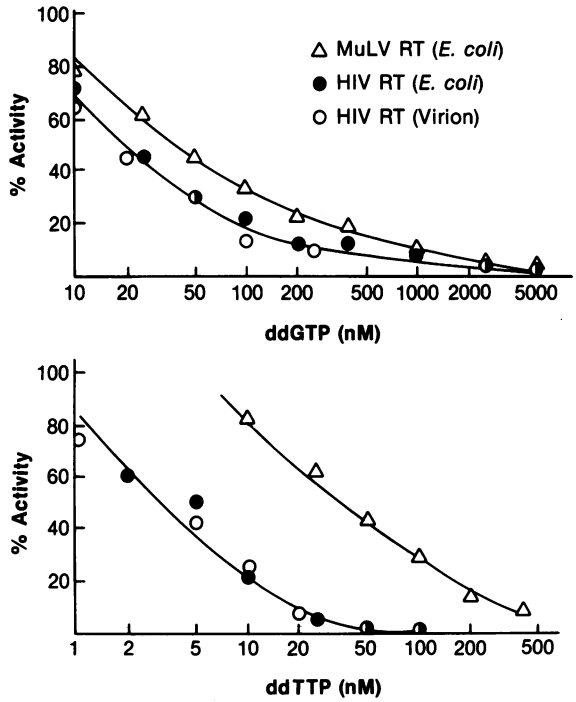

FIG. 3. Effect of inhibitors of the polymerizing function of Mo-MuLV and HIV reverse transcriptases. Extracts of *E. coli* strains that make the HIV and Mo-MuLV reverse transcriptases were passed over Sephadex G-25 and assayed for polymerizing activity in comparison with HIV reverse transcriptase from disrupted virions. Increasing amounts of either dideoxy-GTP (*Upper*) or dideoxy-TTP (*Lower*) were added to standard *in vitro* polymerization reactions. The template used to study dideoxy-TTP inhibition was poly(rA) and to study dideoxy-GTP inhibition the template was poly(rC). Activity is expressed as a percentage of the incorporation seen in the absence of added inhibitor. The concentration of the inhibitors is displayed on a logarithmic scale.

proteins that are found in large quantity in the *E. coli* lysates. By substituting a different synthetic DNA for the *Nco* I–*Bal* I segment, it was possible to synthesize an HIV reverse transcriptase protein that lacks the first 23 amino acids of the virion enzyme (although it does retain the methionine and the additional valine required to reconstruct the *Nco* I site). This protein was made in quantities comparable to the full-length enzyme; however, it had no detectable polymerizing activity (see Table 2). Three carboxyl-terminal mutants that make substantial amounts of protein were constructed by altering the carboxyl-terminal synthetic DNA segment: one lacks the last 8 amino acids, one lacks 16 amino acids, and the third lacks 23 amino acids. Although the 8- and the 16-amino acid deletions caused slightly less activity than the full-length protein did, the strains that made these proteins appeared to contain slightly less of the deleted products than the strain that made the wild-type enzyme (see Fig. 2). By contrast, the deletion of the carboxyl-terminal 23 amino acids dramatically decreased the polymerizing activity of the protein product (see Table 2).

We also made three different insertions at the unique *Eco*RV site (position 2556). These led to the synthesis of a protein that had lost a tyrosine at amino acid position 143 (measured from the amino terminus of the viral enzyme) and had an insertion of 5 amino acids at this position. Although the three insertions had substantially different sequences, the insertions all led to the production of proteins that lacked polymerizing activity. An insertion at the unique *Pfl*MI site (position 3065) led to the loss of two amino acids at amino acid position 313 and the insertion at this position of five new amino acids. In contrast to the *Eco*RV mutant products, the product of the *Pfl*MI mutation retained approximately half the activity of the wild-type protein (see Table 2).

Table 2. Relative reverse transcriptase (RT) activities of the different mutants

HIV RT construct	Activity, %
Wild type (p66)	100
CT-133 (p51)	0.4
CT-23	4.1
CT-16	45
CT-8	60
AT-23	0
RV-1	0
RV-2	0
RV-3	0
PM-1	43

HIV RT constructs in *E. coli* were assayed for RT activity in bacterial extracts as described. The various clones are named for the mutations they express: CT, deletions from the carboxyl terminus of the molecule followed by the number of amino acids deleted; NT, deletions from the amino acid terminus of the reverse transcriptase followed by the number of amino acids deleted; RV, three different insertions, each of 5 amino acids, at amino acid position 143 (DNA insertion in the *Eco*RV site); and PM, a substitution of 5 amino acids for two at amino acid position 313 (DNA insertion at the *Pfl*MI site). The specific sequences for each of the mutant proteins are as follows: wild-type has, compared with the virion enzyme, two additional amino-terminal amino acids, Met-Val; CT-133 has 133 amino acids deleted from the carboxyl terminus, and the last 11 amino acids derive from the +1 reading frame, giving a sequence Leu-Trp-Tyr-Gln-His-Thr-Lys-Glu-Leu-Glu-Glu-Met-Asn-Lys at the carboxyl terminus; CT-23 ends Tyr-Leu-Ala-Trp-Val-Leu; CT-16 ends His-Lys-Gly-Ile-Gly-Gly; CT-8 ends Asp-Lys-Leu-Val; AT-23 has a deletion of 23 amino-terminal amino acids but still has the two extra amino acids (Met-Val) present in the wild-type enzyme, hence the sequence begins with Met-Val-Trp-Pro-Leu-Thr; RV-1,2,3 have DNA insertions equivalent to 5 amino acids (italic) in the *Eco*RV site (Gly-Ile-Arg-Tyr-Gln-Tyr-Asn), giving Gly-Ile-Arg-*Leu-Gly-Ala-Pro-Asn*-Gln-Tyr-Asn (RV-1), Gly-Ile-Arg-*Trp-Arg-Ala-Arg-His*-Gln-Tyr-Asn (RV-2), and Gly-Ile-Arg-*Ser-Asp-Ala-Ser-Asp*-Gln-Tyr-Asn (RV-3); and PM-1 has Glu-Pro-*Leu-Gly-Ala-Pro-Asn*-Gly-Val instead of Glu-Pro-Val-His-Gly-Val.

We also asked whether the 51-kDa reverse transcriptase-related molecule found in virions has enzymatic activity. Since the 51-kDa protein found in virions has the same amino terminus as the 66-kDa enzyme, it is likely that the 51-kDa protein lacks the carboxyl terminus of the 66-kDa protein (18). Based on this assumption, we made a construction that would induce the synthesis of the amino-terminal 70% of the HIV reverse transcriptase molecule. This strain did not produce substantial amounts of enzyme; however, we did find that, by deleting the *Kpn* I–*Kpn* I segment from the wild-type construction, *E. coli* would synthesize large amounts of a 51-kDa protein. This protein lacks the last 133 amino acids of the 66-kDa enzyme; however, since the *Kpn* I–*Kpn* I deletion induced a frameshift, 11 unrelated amino acids were added to the carboxyl terminus. This protein had little, if any, polymerizing activity (see Table 2).

DISCUSSION

Amino- and carboxyl-terminal deletions of the HIV reverse transcriptase were used in an attempt to define the limits of the polymerizing function of reverse transcriptase. Removing 23 amino acids from either the amino or the carboxyl terminus markedly reduces the polymerizing activity of the recombinant enzyme. The extreme ends of retroviral reverse transcriptases are not well conserved; however, when the sequence of the HIV reverse transcriptase is compared with the relatively unrelated reverse transcriptase from Mo-MuLV, there are conserved amino acids within the regions covered by the deletions. These conserved amino acids fall at a boundary of the 23-amino acid deletions, which may explain why the deletion of 16 amino acids from the carboxyl terminus is tolerated relatively well. A comparison of the sequences of several retroviral reverse transcriptases, including that of HIV, with *E. coli* DNA-dependent RNA polymerase and *E. coli* RNase H, has suggested that the amino-terminal portion of retroviral reverse transcriptases contains the polymerizing function and the carboxyl-terminal portion contains the RNase H function (19). However, our data implies that both the amino- and carboxyl-terminal portions of the reverse transcriptase are required for the polymerizing functions. This does not necessarily conflict with the sequence comparisons; however, it does suggest, at least in the case of HIV reverse transcriptase, that the polymerizing function and the RNase H function do not comprise readily separable domains. Any one of three separate blocks of 5 amino acids inserted in place of a tyrosine at a position 143 amino acids from the amino terminus destroys the polymerizing function. This position is immediately adjacent to a region that is conserved among viral reverse transcriptases (19). However, a substitution of 5 amino acids for 2 at a position 313 amino acids from the amino terminus decreases the polymerizing activity about 50%. This is a region that is not particularly well conserved between HIV and Mo-MuLV reverse transcriptases.

Several other groups have produced proteins in *E. coli* or in yeast that are related to either the HIV or the Mo-MuLV reverse transcriptases (20–23); however, the protein product we report here is substantially closer to the viral enzyme in structure and made larger amounts. Since it is possible that there are posttranslational modifications made on the HIV reverse transcriptase made in eukaryotes that would not be made in *E. coli*, the HIV reverse transcriptase coding region in the construction we have made can be precisely excised and introduced into yeast and higher eukaryotic expression systems (24). The corresponding yeast plasmid has already been made (C.M., A.H., D. Garfinkel, S.H.H., and J. Strathern, unpublished data) and also produces large

amounts of enzymatically active protein. The corresponding higher eukaryotic constructions will soon be tested.

We wished to ask whether the 51-kDa protein found in HIV virions has reverse transcriptase activity. We approached this problem by examining the properties of carboxyl-terminal deletions of the HIV enzyme. Based on the properties of the 23-amino acid carboxyl-terminal deletion and on a protein of ≈51 kDa produced by the *Kpn* I–*Kpn* I deletion strain, we predict that the 51-kDa protein found in the virion would have, at most, only a small residual amount of reverse transcriptase activity. This interpretation is tempered by the fact that the 51-kDa protein made in *E. coli* has 11 unrelated amino acids at the carboxyl terminus and by the fact that the carboxyl terminus of the virion p51 is unknown and may not precisely correspond to the end point of the p51 made in *E. coli*. In addition, there are potential roles for p51 other than reverse transcription, and the function of the p51 in the viral life cycle, if any, remains obscure. It should be remembered that the relationship of the 51-kDa and 66-kDa HIV reverse transcriptase proteins is quite different from the situation with the α and β subunits of the Rous sarcoma virus (RSV) reverse transcriptase. The RSV enzyme is composed of two unlike subunits, α and β. The larger subunit, β, is composed of the reverse transcriptase and the endonuclease, while the smaller, α, differs from β in that it lacks the endonuclease domain. However, in the case of RSV, not only is the $\alpha\beta$ enzyme active, but so is the α monomer and the $\beta\beta$ dimer (12).

There are several purposes behind synthesizing the HIV reverse transcriptase in *E. coli*. The *E. coli*-synthesized reverse transcriptase is soluble, enzymatically active, produced in large amounts and, when analyzed with two inhibitors, dideoxy-TTP and dideoxy-GTP, has properties indistinguishable from the viral enzyme. The *E. coli* enzyme should be useful for the preliminary screening of drugs *in vitro*, although it will not replace replication assays in cultured cells. However, there is one caveat: most of the substrate analogs should be tested in the *in vitro* reverse transcriptase assay in the form of triphosphates; to be used as drugs, the analogs would be given as free nucleotides or nucleosides.

In addition to being a tool for testing drugs directly, the *E. coli* enzyme should be a convenient source for preparing large quantities of the enzyme for biochemical and physical analyses, including x-ray structure determination. If the structure of the enzyme were known in detail, it might be possible to use this detailed structure to design drugs that would specifically inhibit the enzyme. The *E. coli* enzyme also is being used for the preparation of both monoclonal and polyclonal antibodies. Finally, it should be possible to use the *E. coli* system as a tool to explore the limits imposed on the function of the HIV reverse transcriptase by its primary sequence. This can be explored in several ways. Naturally occurring variants of the HIV reverse transcriptase could be produced in *E. coli* in the same fashion, and their physical and biochemical properties could be compared. In addition, other less-related reverse transcriptases can be expressed by using the same system. We already have expressed the Mo-MuLV reverse transcriptase and demonstrated that, with respect to inhibitors, it has properties distinct from the HIV enzyme. Finally, we have begun to make mutations in the HIV reverse transcriptase. The analyses of such mutations should, coupled with biochemical and structural studies, not only provide information on the relationship between the primary sequence and the catalytic activity of the enzyme but also should help to define the limits of variability that will be encountered in the reverse transcriptase of different HIV isolates.

We are grateful to Drs. Jeffrey Strathern, David Garfinkel, and Stephen Oroszlan for helpful discussions and to Hilda Marusiodis for preparing the manuscript. This research was sponsored by the National Cancer Institute under Contract N01-CO-74101 with Bionetics Research.

1. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (1985) *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
2. Chandra, P., Vogel, A. & Gerber, T. (1985) *Cancer Res.* **45,** 4677s–4684s.
3. de Clercq, E. & Balyarini, J. (1985) *Antiviral Res. Suppl.* **1,** 89–94.
4. Mitsuya, H., Weinhold, K. J., Furman, P. A., St. Clair, M. H., Lehrman, S. N., Gallo, R. C., Bolognesi, D., Barry, D. W. & Broder, S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7096–7100.
5. Furman, P. A., Fyfe, J. A., St. Clair, M. H., Weinhold, K., Rideout, J. L., Freeman, G. A., Lehrman, S. N., Bolognesi, D. P., Broder, S., Mitsuya, H. & Barry, D. W. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8333–8337.
6. Mitsuya, H. & Broder, S. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1911–1915.
7. Cheng, Y.-C., Dutschman, G., Bastow, K., Sarngadharan, M. G. & Ting, R. (1987) *J. Biol. Chem.* **262,** 2187–2189.
8. Mitsuya H., Jarrett, R. F., Matsukura, M., Di Marzo Veronese, F., DeVico, A. L., Sarngadharan, M. G., Johns, D. G., Reitz, M. S. & Broder, S. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2033–2037.
9. Mitsuya, H. & Broder, S. (1987) *Nature (London)* **325,** 773–778.
10. Hanahan, D. (1983) *J. Mol. Biol.* **166,** 557–580.
11. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starchich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313,** 277–283.
12. Hizi, A. & Joklik, W. K. (1977) *J. Biol. Chem.* **252,** 2281–2289.
13. 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.
14. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985) *Cell* **40,** 9–17.
15. Lightfoote, M., Coligan, J., Folks, T., Fauci, A. S., Martin, M. A. & Venkatesan, S. (1986) *J. Virol.* **60,** 771–775.
16. Norrander, J., Vieira, J., Rubenstein, I. & Messing, J. (1985) *J. Biotechnol.* **2,** 157–175.
17. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
18. Di Marzo Veronese, F., Copeland, T. D., DeVico, A. L., Rahman, R., Oroszlan, S., Gallo, R. C. & Sarngadharan, M. G. (1986) *Science* **231,** 1289–1291.
19. Johnson, M. S., McClure, M. A., Feng, D. F., Gray, J. & Doolittle, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7648–7652.
20. Kotewicz, M., D'Alessco, M., Driftmier, M., Blodgett, K. & Gerard, G. (1985) *Gene* **35,** 249–258.
21. Roth, M.J., Tanese, N. & Goff, S. P. (1985) *J. Biol. Chem.* **260,** 9326–9335.
22. Tanese, N., Sodroski, J., Haseltine, W. & Goff, S. P. (1986) *J. Virol.* **59,** 743–745.
23. Barr, P. J., Power, M. D., Lee-Ng, C. T., Gibson, H. L. & Luciw, P. A. (1987) *Bio/Technology* **5,** 486–489.
24. Hughes, S. H., Greenhouse, J., Petropoulos, C. J. & Sutrave, P. (1987) *J. Virol.* **61,** 3004–3012.
25. Maniatis, T., Fritsch, E. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).