Purification and Characterization of Murine Retroviral Reverse Transcriptase Expressed in Escherichia coli*

Monica J. Roth‡, Naoko Tanese, and Stephen P. Goff
From the Department of Biochemistry and Institute for Cancer Research, Columbia University, College of Physicians and Surgeons, New York, New York 10032

Expression of a region of the Moloney murine leukemia virus (M-MuLV) pol gene in Escherichia coli resulted in the synthesis of reverse transcriptase activity which could be detected in crude extracts. Construction of deletions at the 3' terminus of this gene resulted in a 4-fold increase in the level of the reverse transcriptase activity in the soluble fraction of crude lysates and yielded the high level production of a stable protein species of Mᵦ = 71,000. Purification of this protein by column chromatography on DEAE-cellulose, phosphocellulose, polyriboctydylidic acid-agarose, and hydroxylapatite indicated that it was a multifunctional enzyme containing RNase H and reverse transcriptase activity. The Mᵦ = 71,000 species had a sedimentation coefficient of 4.65 S by glycerol gradient centrifugation, indicating that the enzyme was a monomer. Using poly(A)* mRNAs primed with oligo(dT), the enzyme synthesized double-stranded DNA copies between 1.3 and 9.9 kilobases in length. Synthesis of long cDNA required 8 mM Mg²⁺, 4 mM Mn²⁺, 2 mM dNTPs, and saturating levels of enzyme. Actinomycin D efficiently limited the enzyme to the first strand synthesis. Additional characteristics of the fusion protein are described.

In the early stages of the retroviral life cycle, viral RNA is copied to form a double-stranded DNA, which is integrated into host DNA to generate the provirus (for review, see Ref. 1). The viral pol gene encodes many enzymatic activities which participate in these steps. The pol gene product is initially expressed as a polyprotein Pr200pol (2), which is proteolytically processed to yield several mature proteins (3). Genetic and biochemical data support the localization of a protease function to the 5' terminus of the pol gene (4), an RNase H and reverse transcriptase activity to the large central region (5), and a function required for the establishment of the integrated provirus to the 3' end (5, 6).

The viral M-MuLV reverse transcriptase has been purified (7, 8) and shown to have a nuclease activity (RNase H) which degraded RNA contained in an RNA-DNA hybrid (7, 8). Detailed analyses of the properties of the enzyme have been limited by the small quantity of enzyme which can be isolated from retroviral virions. An alternate means to prepare the enzyme is to express the appropriate section of the pol gene in Escherichia coli, allowing overproduction of the protein and facilitating further characterization. We have previously described the construction of a plasmid which directs the synthesis of a fusion protein containing the first 37,000 daltons of the trpE gene product and a large portion of the M-MuLV pol gene product. Reverse transcriptase activity was readily detected in crude extracts of E. coli carrying the plasmid extracts (9). A modification of this construct, termed pSNHB6, was also made (9) in which only the first 18 codons of the trpE gene were joined to the pol gene. This modification resulted in a 4-8-fold increase in the level of the reverse transcriptase activity in the soluble fraction; the gene products encoded by the plasmid, however, were not stable. In the present work, this construct was further modified such that the level of expression and the stability of the reverse transcriptase fusion protein was increased. The purification of this novel reverse transcriptase and the initial characterization of this enzyme and its associated activities are reported herein.

EXPERIMENTAL PROCEDURES*

RESULTS

Construction of a Gene Fusion Expressing Stable Fusion Protein—An outline of the scheme used to stably express the MuLV reverse transcriptase activity is presented in Fig. 1. The initial construct used as the starting material for further manipulation was the plasmid pSNHB6, containing the region of M-MuLV from nucleotide position 2574-4893 (10) inserted in-frame downstream of the first 18 codons of the E. coli trpE protein (9). Our earlier work demonstrated that extracts of E. coli strains bearing this plasmid contained high levels of reverse transcriptase activity (9). Analysis of the proteins synthesized in these strains, however, indicated that the major product was broken down into smaller species; partial purification of the soluble reverse transcriptase also indicated that multiple species were active. The breakdown of the fusion protein was not prevented by the addition of...
protease inhibitors in the lysis procedure. An additional problem was that the majority of the fusion proteins partitioned into the insoluble fraction after cell lysis.

In an attempt to stabilize and solubilize the protein, deletions were made at the 3' terminus of the cloned pol gene. The assumption was made that random Bal31 deletions at this terminus might result in the formation of a protein that more closely resembled the authentic cleavage product, and might improve its stability in E. coli. To make deletions, plasmid pSHNB6 was linearized with HindIII, digested with Bal31 nuclease, and ligated with T4 DNA ligase; the DNA products were used to transform HB101, and the resulting colonies were screened for reverse transcriptase as described under "Experimental Procedures." The characteristics of plasmid pB6B15.23 are summarized at the bottom. The numbers in parentheses refer to the maps of M-MuLV and pBR322 as described (10 and 34, respectively). The resulting gene fusion consists of an open reading frame encoding 698 amino acids. The first 18 amino acids at the N terminus are encoded by the pol gene, and the terminal 9 amino acids by pBR322. The sequence of the 3' terminus of the gene was determined by the method of Maxam and Gilbert (35) after 5' end labeling with polynucleotide kinase at the BglII site. The DNA sequence and the deduced amino acid sequence are indicated at the bottom.

produced after Bal31 digestion for 7.5 min was similar to that of cells carrying the parent plasmid pSHNB6. Colonies produced after digestion of the DNA with Bal31 nuclease for 15 min yielded a much larger range of activities. Of 49 colonies screened, five yielded no detectable reverse transcriptase activity, 22 yielded less than 50% as much activity as cells containing the parent pSHNB6 plasmid, 17 yielded at least 50% of the activity of the parent cells, and five yielded up to 4-fold higher activity than the parent strain. The extracts from colonies which displayed enhanced reverse transcriptase activity were analyzed by polyacrylamide gel electrophoresis. One colony was selected for further study (pB6B15.23) because of the following features: 1) the specific activity of extracts from these cells was 3.5-4 times that of pSHNB6; 2) the level of induction was reproducible; 3) Coomassie Blue staining of polyacrylamide gels indicated that a single species of $M_r = 71,000$ was highly and stably overproduced; 4) comparison of the insoluble and soluble fractions by polyacrylamide gel electrophoresis indicated that at least 30% of the $M_r = 71,000$ band could be detected in the soluble fraction (data not shown).

The DNA sequence of this plasmid in the region of the

\[^3\text{M. J. Roth, N. Tanese, and S. P. Goff, unpublished observations.}\]

\[^4\text{S. Oroszlan, personal communication.}\]
deletion was determined and is shown at the bottom of Fig. 1. The deletion resulted in the removal of 204 nucleotides of the M-MuLV pol gene sequence and 64 bases of pATH1. The first stop codon is found 27 bases from the new junction between the MuLV coding sequence and the vector. The carboxyl terminus of the fusion protein would contain 9 novel amino acids encoded by the pBR322 sequence.

Analysis of the Fusion Proteins by Immunoprecipitation—Analysis of crude extracts of cells carrying the pB6B15.23 plasmid by polyacrylamide gel electrophoresis and Coomassie Blue stain indicated that a single stable fusion protein was synthesized. To determine if this was the only product made and if this product was structurally similar to the viral MuLV reverse transcriptase, immunoprecipitation was performed using various sera. Cultures of HBlOl alone, HBlOl bearing the vector plasmid pATH1, or HBlOl bearing pB6B15.23, were grown with and without induction of the trp operon and were then labeled with [35S]methionine. Extracts were prepared and were incubated with either normal goat serum (Fig. 2, lanes 1–4); normal rabbit serum (lanes 5–8); two sera produced against authentic Rauscher MuLV reverse transcriptase, termed 775-424 (lanes 9–12) and 775-454 (lanes 13–16); and a serum specific for the N terminus of the trpE protein (lanes 17–20). A protein of M, = 59,000 was precipitated nonspecifically with all the sera and extracts tested (lanes 1–20). Both anti-Rauscher reverse transcriptase sera recognized a single protein species of M, = 71,000 in the extracts of cells expressing the cloned M-MuLV reverse transcriptase (lanes 11, 12, 15, and 16). This protein was not present in HBlOl (lanes 9 and 13) or in HBlOl bearing the parent plasmid pATH1 (lanes 10 and 14). This unique protein species had the same electrophoretic mobility as the protein identified previously by Coomassie Blue staining of crude extracts on polyacrylamide gels. Analysis of cells grown under conditions repressing the trp operon (in the presence of tryptophan and without the addition of indoleacrylic acid) indicated that lower levels of the M, = 71,000 species were being synthesized, although the protein could still be detected (lanes 12 and 16). The serum specific for the trpE protein recognized a M, = 52,000 species in HBlOl, presumed to be the trpE protein encoded by the chromosome (calculated M, = 57,524) (lane 17), as well as the truncated M, = 37,000 trpE protein encoded by pATH1 (lane 18). In extracts of HBlOl containing pB6B15.23 grown under conditions of induction, only the M, = 52,000 trpE protein was detected. The M, = 71,000 fusion protein, which contains only the first 18 amino acids of the trpE protein, was not recognized by this serum.

Purification of Reverse Transcriptase and Associated RNase H—To characterize the reverse transcriptase activity induced in cells carrying the pB6B15.23 plasmid, and to determine whether this activity did indeed reside in the novel pol-related protein, the activity was purified. The main assay for the enzyme measured the incorporation of radiolabeled dTTP into polymeric form on a poly(rA) template primed with oligo(dT). The crude extracts were prepared by detergent lysis after lysozyme treatment; the conditions of detergent and high salt utilized were important in solubilizing the activity. The presence of a nonionic detergent was required throughout the purification to prevent aggregation and loss of activity. The purification of the pB6B15.23 reverse transcriptase is detailed under “Experimental Procedures,” and is summarized in Table I. The final procedure involved column chromatography on DEAE-cellulose, phosphocellulose, polyribocytidylic acid-agarose, and hydroxylapatite.

The DEAE-cellulose step was used to remove nucleic acids from the preparation. Since the fusion protein did not bind DEAE-cellulose, only a modest purification of the enzyme was obtained by passing the activity through the column in a low salt concentration. Phosphocellulose column chromatography was the single most important step in the purification, resulting in a 6.6-fold increase in the specific activity of the preparation; the total recovery in this step was, unfortunately, rather low (24%). Polyribocytidylic acid chromatography was useful because very few proteins were capable of binding to the resin; the pB6B15.23 enzyme was essentially the only protein bound and eluted from the resin. Due to the high level of expression of the fusion protein in E. coli, the overall preparation of the reverse transcriptase activity required only a 22-fold purification.

The M-MuLV reverse transcriptase has been reported to have an associated RNase H activity (7, 8). To determine if the region of the pol gene expressed in E. coli encoded the RNase H function, fractions from each stage of the purification were assayed for RNase H as well as reverse transcriptase activity. The profile of these two activities on phosphocellulose chromatography is shown in Fig. 3. A substantial amount of RNase H activity was eluted with low salt concentrations; in addition, a predominant peak of RNase H activity was coincident with the reverse transcriptase activity and eluted at 0.22 M NaCl. The RNase H activity eluting with the low salt concentration was not identified nor characterized further. The phosphocellulose fractions containing the reverse transcriptase activity and RNase H activity were further chromatographed on polyribocytidylic acid-agarose (Figure 4). The reverse transcriptase activity and the RNase H activity co-chromatographed as a single peak on this resin. These two activities were also associated after hydroxylapatite column chromatography and after glycerol gradient centrifugation in 0.5 M NaCl (see below). These results suggest that the central portion of the pol gene of M-MuLV present in the construct encodes both reverse transcriptase and RNase H activities.

Sodium Dodecyl Sulfate-Gel Electrophoresis—To assess the purity of the reverse transcriptase, the protein fractions from

![Fig. 2. Immunoprecipitation of 35S-labeled extracts with various sera. Preparation of 35S-labeled extracts, precipitation with various sera, and gel electrophoresis were as described under “Experimental Procedures.” The various sera used were as follows: lanes 1–4, normal goat serum; lanes 5–8, normal rabbit serum; lanes 9–12, serum 775-424 (raised against Rauscher reverse transcriptase); lanes 13–16, serum 775-454 (raised against Rauscher reverse transcriptase); lanes 17–20, serum raised against the N terminus of trpE protein. Grid at the bottom summarizes the cells from which the extracts were made and the level of induction. The position of migration and the M, of the marker proteins and bromphenol blue (BPB) are indicated at the right.](image-url)
Reverse transcriptase, RNase H, and total protein were assayed as described under “Experimental Procedures.” One unit of reverse transcriptase is defined as that amount of enzyme which will incorporate 1 nmol of [3H]dTMP into a form retained on DE81 paper in 15 min at 37 °C. One unit of RNase H activity is defined as that amount of enzyme which will release 1 nmol of [3H]AMP from an RNA:DNA hybrid into an acid-soluble form in 30 min at 37 °C.

Table 1

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Reverse transcriptase</th>
<th>RNase H</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>units/ml</td>
<td>units</td>
<td>% yield</td>
</tr>
<tr>
<td>DEAE load</td>
<td>2643</td>
<td>629</td>
<td>206</td>
</tr>
<tr>
<td>DEAE flow-through</td>
<td>1400</td>
<td>490</td>
<td>060</td>
</tr>
<tr>
<td>Phosphocellulose</td>
<td>2786</td>
<td>118</td>
<td>420</td>
</tr>
<tr>
<td>Hydroxylapatite</td>
<td>15120</td>
<td>529</td>
<td>200</td>
</tr>
<tr>
<td>Dialysis</td>
<td>28910</td>
<td>50,590</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Fig. 3. Phosphocellulose chromatography. The material which did not bind to DEAE-cellulose was chromatographed on phosphocellulose as described in the text. After collection of the material which did not absorb to phosphocellulose (not shown), the column was eluted with a gradient of NaCl as indicated. Aliquots of the individual fractions were diluted (1:200) as described in under “Experimental Procedures.” 25-μl aliquots of the column eluant were assayed directly for total protein (○) as described under “Experimental Procedures.”

Fig. 4. Polyribocytidylic acid-agarose chromatography. The bound fraction from the phosphocellulose column was pooled, diluted, and applied to a polyribocytidylic acid-agarose column as described in the text. The column was washed and eluted with a salt gradient as indicated. Fractions were diluted (1:200) as described in the legend to Fig. 3. Reverse transcriptase (●) and RNase H (△) activities were assayed as described under “Experimental Procedures” using 1 and 3 μl of the diluted enzyme, respectively. Total protein (○) was determined using 10 μl of the column eluant as described under “Experimental Procedures.”

Fig. 5. Analysis of purification of pB6B15.23 reverse transcriptase by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Samples from various stages of the purification were subjected to electrophoresis through a 10% polyacrylamide gel and the proteins were detected by staining with silver as described (23). Lane 1, material applied to DEAE-cellulose; 18.1 μg of protein, 10.6 units of enzyme. Lane 2, DEAE flow-through fraction; 12.6 μg of protein, 10.1 units of enzyme. Lane 3, pooled phosphocellulose fraction; 2.15 μg of protein, 11.1 units of enzyme. Lane 4, pooled polyribocytidylic acid-agarose fraction; 0.95 μg of protein, 10.0 units of enzyme. Lane 5, pooled hydroxylapatite fraction; 1.03 μg of protein, 10.8 units of enzyme. Lane 6, final preparation after dialysis; 0.89 μg of protein, 11.5 units of enzyme. The position and Mr of the marker proteins electrophoresed in the same gel are indicated at the sides. The enzymes used were the β and α subunits of E. coli RNA polymerase, myosin, β-galactosidase, albumin (bovine), albumin (egg), and carbonic anhydrase. Migration of bromphenol blue (BPB) dye is indicated.

The various stages of the purification were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and the proteins were visualized by the silver staining procedure (23). The data presented in Fig. 5 show the polypeptide composition of the material applied to the DEAE-cellulose column (lane 1); of the material that flowed through the DEAE-cellulose
column (lane 2) of the pooled fractions after chromatography on phosphocellulose (lane 3), polyribocytidylic acid-agarose (lane 4), and hydroxylapatite (lane 5); and of the final fraction (lane 6). The analysis showed the presence of a single major band of $M_r = 71,000$ in the late stages of the purification. Although the largest loss of reverse transcriptase activity occurred during phosphocellulose chromatography, extensive purification of the $M_r = 71,000$ protein was achieved by this step. The $M_r = 71,000$ protein is greater than 95% pure after polyribocytidylic acid-agarose chromatography. The hydroxylapatite chromatography and dialysis steps were mainly useful as a means of concentrating the protein.

Glycerol Gradient Centrifugation—To determine the subunit structure of the purified pB6B15.23 reverse transcriptase, the enzyme was further characterized by glycerol gradient sedimentation (Fig. 6). The reverse transcriptase and RNAse H activities co-sedimented as a single peak with a sedimentation coefficient of 4.65 S. Based on this sedimentation coefficient, the molecular mass of the species was estimated to be about 65,000 daltons. The proteins from the gradient fractions surrounding this peak of activity were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and the proteins detected by silver stain. The results (Fig. 6) showed that the presence and intensity of the $M_r = 71,000$ species paralleled the reverse transcriptase and associated RNAse H profile. The purified fusion protein appears to behave as a monomer, as has been reported for the authentic M-MuLV reverse transcriptase (7, 8).

Requirements for Reverse Transcriptase Activity of Fusion Protein.—The assay used for the purification of the pB6B15.23 reverse transcriptase activity measured the incorporation of dTMP on a poly(rA) template primed with oligo(dT). The requirements and the optimal conditions for DNA synthesis by the purified enzyme on this synthetic substrate were determined (Table II A). Under optimal conditions, incorporation of dTMP occurred linearly with time from 2 min to up to 2 h. Maximal DNA synthesis saturated at 500 pmol of dTMP incorporated, equivalent to one-third of the available single-stranded template. Incorporation was almost totally dependent on the presence of the template poly(rA) and primer oligo(dT). In addition, DNA synthesis required a divalent cation; either Mn$^{2+}$ or Mg$^{2+}$ were capable of supporting synthesis. The maximal incorporation occurred between 0.5-1.0 mM MnCl$_2$ and was greatly inhibited at levels higher than 2 mM (data not shown). Mg$^{2+}$ supported DNA synthesis to a much lower degree than Mn$^{2+}$, with optimal activity occurring at 0.5 mM; higher concentration of Mg$^{2+}$ also inhibited the reaction (data not shown). The addition of Mg$^{2+}$ to a reaction mixture that already contained Mn$^{2+}$, however, did not inhibit the Mn$^{2+}$-dependent synthesis.

Maximal DNA synthesis on poly(rA):oligo(dT) occurred in the presence of 60-80 mM NaCl; standard reaction mixtures contained 60 mM NaCl. DNA synthesis was decreased to 64% of the maximal synthesis at 20 mM NaCl, to 56% of the maximal at 120 mM NaCl, and to only 12% of the maximal at 240 mM NaCl. The enzyme activity was inhibited by inorganic phosphate. In a standard assay (see "Experimental Procedures") activity was reduced to 73% of maximal by the addition of 5 mM inorganic phosphate, to 46% of maximal by 20 mM phosphate, and to 14% by 40 mM phosphate. The enzyme was extremely sensitive to pyrophosphate; 68% and 50% of maximal TMP incorporation was seen in the presence of 0.1 and 0.5 mM Na$_2$P$_2$O$_7$, respectively, and only 6% of maximal activity was found at 1.0 mM Na$_2$P$_2$O$_7$. Similar results...
TABLE II

Requirements for pB6B15.23 MuLV reverse transcriptase

Experiment A, the complete reaction mixture contained 50 mM Tris-HCl, pH 8.3, 20 mM dithiothreitol, 0.5 mM MnCl\(_2\), 60 mM NaCl, 10 μg/ml oligo(dT), 20 μg/ml poly(rA), 20 μM dTTP (818 rpm/μmol), 0.1% Nonidet P40, and 0.072 unit of enzyme. Individual components were omitted as indicated. MgCl\(_2\) was added at 0.5 mM concentration. Experiment B, complete reaction mixture was as described in Experiment A. pB6B15.23 reverse transcriptase was diluted (1:4000) in Buffer M plus 0.2 mM NaCl and treated as described prior to the reverse transcriptase assay. 20 μl of enzyme was digested with proteinase K were omitted as indicated. MgCl\(_2\) was added at 0.5 mM concentration. Dithiothreitol was performed on ice for 15 min.

<table>
<thead>
<tr>
<th>A. Additions</th>
<th>pmol [(^{32})P]TMP incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>73</td>
</tr>
<tr>
<td>Omit DTT</td>
<td>80</td>
</tr>
<tr>
<td>Omit MnCl(_2)</td>
<td>&lt;0.2</td>
</tr>
<tr>
<td>Omit MnCl(_2) + MgCl(_2)</td>
<td>1.5</td>
</tr>
<tr>
<td>Add MgCl(_2)</td>
<td>70</td>
</tr>
<tr>
<td>Omit oligo(dT)</td>
<td>1.3</td>
</tr>
<tr>
<td>Omit poly(rA)</td>
<td>&lt;0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Preincubation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>50</td>
</tr>
<tr>
<td>Heated 37 °C, 15 min</td>
<td>46</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>10 mM N-ethylmaleimide, plus 50 mM DTT</td>
<td>0.5</td>
</tr>
<tr>
<td>50 mM DTT, plus 10 mM N-ethylmaleimide</td>
<td>43</td>
</tr>
<tr>
<td>Heated 70 °C, 15 min</td>
<td>&lt;0.1</td>
</tr>
</tbody>
</table>

were found when the concentration of free Mn\(^{2+}\) was maintained at 0.5 mM or at 1.0 mM in the reaction mixtures. Omission of DTT from the reaction mixture appeared to slightly stimulate the reaction; however, the enzyme fraction contained 5 mM DTT, which may have partially compensated for this omission. The enzyme was almost completely inhibited in the presence of the sulphydryl antagonist, N-ethylmaleimide (Table IIB). This inhibition was completely prevented by the prior addition of dithiothreitol.

The final fraction was sensitive to preincubation with proteinase K as well as heating at 70 °C for 15 min. Heating at 42 °C for 15 min resulted in a 50% loss of activity (data not shown).

Contaminating Activities—The purified enzyme preparation was assayed for ATPase activity in the absence and presence of single- and double-stranded DNA, and of poly(rA)oligo(dT), using [α-\(^{32}\)P]ATP. No released [\(^{32}\)P]inorganic phosphate could be detected.

Fidelity of DNA Synthesis—The assay measuring incorporation on the oligo(dT) and poly(rA) substrates was found to be totally dependent on template and primer (see above). To determine the fidelity of the template-directed synthesis, the incorporation of various [α-\(^{32}\)P]dNTPs was measured. The levels of incorporation of [α-\(^{32}\)P]dATP, [α-\(^{32}\)P]dCTP, and [α-\(^{32}\)P]dGTP were less than 0.1% that of [α-\(^{32}\)P]dTTP. Since the authentic M-MuLV reverse transcriptase is capable of DNA-dependent DNA synthesis using an RNA primer, the possibility existed for the incorporation of dAMP; this would result from the use of the oligo(dT) as a template and the poly(rA) as the primer. No significant incorporation of [α-\(^{32}\)P]dATP was detected even with the addition of unlabeled dTTP, suggesting that second strand synthesis was limited under the conditions of the assay.

Synthesis of Long cDNA Products—In the viral life cycle, reverse transcriptase must synthesize double-stranded DNA products over 8 kilobase pairs in length (1, 24, 25). The size and nature of the cDNA products of pB6B15.23 reverse transcriptase were determined using various poly(A)+ mRNA preparations primed with oligo(dT). [α-\(^{32}\)P]dNTPs were incorporated, and the DNA products were analyzed by electrophoresis through alkaline agarose gels and autoradiography. A titration of the enzyme on 60 ng of poly(A)+ RNA from HOPEC myelomas, under conditions described for the avian reverse transcriptase, showed that the size of the DNA product depended dramatically on the amount of enzyme added (data not shown). Little or no synthesis was detectable in the presence of 0.02–0.29 unit of enzyme. In the presence of 2.89 units of enzyme, the average DNA product was only 300 bases long. When the amount of enzyme was increased 10-fold, the DNA products were between 315 and 1900 bases in length.

Optimal conditions for synthesis with the pB6B15.23 reverse transcriptase were found by titrating the various components of the reaction. Both MgCl\(_2\) and MnCl\(_2\) could fulfill the divalent cation requirement at optimal concentrations of 10 and 6 mM, respectively. Maximal size and DNA synthesis occurred with a mixture of 8 mM MgCl\(_2\) and 4 mM MnCl\(_2\). At this level of divalent cations, the addition of 20 mM NaCl inhibited the DNA synthesis (data not shown). Efficient DNA synthesis occurred in the presence of high concentrations of deoxynucleoside triphosphates; reaction mixtures contained 2 mM of each dNTP.

Using the optimal conditions, cDNA was synthesized with an oligo(dT) primer on poly(A)+ RNA isolated from human fetal muscle tissue. This RNA preparation was chosen because it is presumably enriched for the very large (7 kilobases) myosin heavy chain mRNA; the size distribution of the product would therefore not be limited by the size of the RNA templates. Using a commercial preparation of the avian viral reverse transcriptase (20.8 units), the average size of the product was 1200–3300 bases (Fig. 7, lane 1). The product produced with 28.9, 86.7, and 173 units of pB6B15.23 reverse transcriptase is shown in lanes 2–4. DNA synthesis saturated with 86.7 units of enzyme, and the majority of the DNA products were between 1.3 and 9.9 kilobases in length. The size distribution was not changed by the presence of excess enzyme, indicating that RNase contamination was negligible. DNA synthesis was not detected in the absence of RNA (lane 5) or reverse transcriptase (lane 6).

To determine whether the products were single-stranded copies of the mRNAs or double-stranded molecules resulting from hairpin loopbacks serving as primers, cDNA synthesis was performed on poly(A)+ mRNA isolated from human reticulocyte lysates (Fig. 8). The predominant species in these preparations are mRNAs of about 570 and 640 nucleotides, encoding the α- and β-globins. Synthesis with low levels of avian reverse transcriptase yielded a single major product approximately 600 bases in length, corresponding to the full-length single-stranded cDNA species (lane 1). Synthesis at high levels yielded predominantly two DNA species approximately 600 and 1200 bases in length, corresponding to the single- and double-stranded cDNAs (lane 2); the major product was the smaller, single-stranded cDNA copy. Synthesis with low levels of the purified pB6B15.23 reverse transcriptase (lane 3) also yielded the single-stranded product. With high levels of the enzyme, the 1200-base products were the predominant species, and the single strand products were not detected as a discrete species, indicating that after the completion of the full-length first strand, the DNA was efficiently looped back and used as a primer for the second strand synthesis. To confirm that the 1200-base products were in fact the result of second strand synthesis, DNA synthesis was performed in...
the presence of actinomycin D. Actinomycin D binds preferentially to double-stranded nucleic acids and therefore inhibits the second strand synthesis. The major products of cDNA synthesis of reticulocyte poly(A)+ RNA with pB6B15.23 reverse transcriptase in the presence of actinomycin D were the 600-base species (lane 6). No products larger than this species could be detected. As before, no products were synthesized in the absence of RNA (lane 7) or enzyme (lane 8).

**DISCUSSION**

These experiments show that a gene fusion containing a portion of the bacterial trpE gene and the central portion of the M-MuLV pol gene can induce the synthesis of a stable protein with high levels of reverse transcriptase activity. A critical step in the successful expression of this activity was the screening of numerous variants of our initial gene con-
Purification of M-MuLV Reverse Transcriptase from E. coli

structs for the formation of maximum levels of stable, soluble protein. Although our initial clone did induce considerable reverse transcriptase activity (9), the fusion protein was exceedingly unstable, and the large primary translation product was reproducibly cleaved into several distinct proteolytic products. The distribution of those species into soluble and insoluble fractions showed that the smaller products were more soluble than the larger ones. The approach taken to counter these problems, therefore, was to generate deletions in the DNA which removed unnecessary codons, and to screen the variants for maximal activity. This procedure resulted in the isolation of a construct that overproduced an extremely stable, soluble, and active fusion protein. This approach may be of general use in maximizing soluble activity of a variety of proteins expressed in E. coli.

The reverse transcriptase fusion protein required a 22-fold purification to yield a nearly homogeneous enzyme preparation. The purification scheme involved multiple column chromatography steps, including polyribocytidylic acid-agarose. This resin has been used for the rapid method of purification of RNA dependent-DNA polymerase from avian myeloblastosis virions (26). The results described in this paper showed that this affinity column was useful as well for the rapid purification of reverse transcriptase activities expressed in E. coli.

New information has been obtained about the various functional domains of the pol gene through its expression in E. coli. The expression of the pol gene from nucleotide position 2574-4588 (10) confirms that both the reverse transcriptase and RNase H activities are encoded by this region and can coexist in a single protein species of Mr = 71,000. Neither activity required the exact viral termini, since the N terminus of the fusion protein is encoded by the trpE gene and the C terminus contains nine random amino acids encoded by pBR322 sequences. Analysis of clones showed that a deletion 140 base pairs larger than that in pB6B15.23 (up to the BglII site) still did not abolish reverse transcriptase activity.

Due to the difficulty in obtaining large quantities of the authentic M-MuLV reverse transcriptase, direct comparison between the pB6B15.23 enzyme and the authentic enzyme could not be made. Data previously published on the viral enzyme indicates that the pB6B15.23 reverse transcriptase was identical to its viral counterpart in its optimal conditions for synthesis on poly(rA):oligo(dT) (7, 15, 27). The sedimentation coefficients determined by glycerol gradient centrifugation also showed that both the viral and the cloned enzyme were monomers (7, 8). The major difference detected between the viral protein and pB6B15.23 reverse transcriptase was the low activity of the pB6B15.23 enzyme on poly(rA):oligo(dT) in the presence of Mg²⁺. The ratio of dTMP incorporation in the presence of Mn²⁺ versus Mg²⁺ for the authentic viral protein was reported to be 3.5:1 (7), whereas for pB6B15.23 reverse transcriptase, this ratio was 49:1.

The structure of the reverse transcriptase from avian retroviruses is quite different from that of the murine viruses. The predominant functional form of the avian enzyme is a heterodimer of two subunits (28), α and β; the larger β subunit is cleaved in the virion to form the smaller α subunit and a third protein, pp32, exhibiting DNA endonuclease activity (29-31). The enzymatic properties of the avian enzyme is also quite different from those of the pB6B15.23 enzyme. Published protocols, for example, have suggested that the addition of NaPP, (32) and synthesis at elevated temperatures (33) were suitable for the formation of full-length products using the avian enzyme; we found that the bacterial enzyme, in contrast, was very sensitive to NaPP, and lost 50% of its activity in 15 min when incubated at 42 °C. An additional difference between the avian and the pB6B15.23 enzyme was that the avian enzyme formed double-stranded cDNAs only poorly, while the bacterial reverse transcriptase was found to efficiently catalyze hairpin synthesis on DNA to form double-stranded DNA products. Actinomycin D inhibited double-stranded DNA synthesis and limited synthesis to the first strand, as with other known DNA polymerases.

On a natural RNA template, the size of the DNA product was found to increase with increasing concentration of the cloned M-MuLV reverse transcriptase; under optimal conditions, the enzyme could synthesize cDNAs up to 9.9 kilobases long. Maximal DNA synthesis using the poly(A)⁺ mRNA from human fetal muscle tissue occurred when the protein was present at a 4.5-fold excess over the RNA (w/w), or approximately one molecule of protein every 48 nucleotides; it is not known, however, what fraction of the reverse transcriptase molecules are active.

Further studies in this laboratory will focus on generating temperature-sensitive mutants of the reverse transcriptase and RNase H activities. These mutants will be isolated by mutagenesis of the cloned MuLV reverse transcriptase; the effects of the mutations will be analyzed after transfer of the altered DNA back into the viral genome and recovery of virus. We hope that analysis of such mutants will result in a better understanding of the role of the activities of the enzyme in the viral life cycle and the interactions of the protein with the other viral gene products.

REFERENCES

Purification of M-MuLV Reverse Transcriptase from E. coli


Purification of M-MuLV Reverse Transcriptase from E. coli

**Purification Protocol**

Extracts for laboratoiy precipitation were prepared as described for screening of deletion mutants (see above) with the following modifications. 

- [15]methionine was added to 40 μM/ml (7.5 ml) cultures at the time of tetracycline acid addition. Extracted cytosolic extracts of pBB15.13 were grown in the presence of supplemented M9 medium plus 10 μg/ml ampicillin and 200 μg/ml tetracycline throughout. Induced tetracycline acid was added. Cells lacking plasmids were grown in the absence of ampicillin. After digestion with lysozyme, the solution was made 15 mM glucose, 0.15 M sodium deoxycholate, 0.15 M sodium dodecyl sulfate, 10 μM NAD, pH 7.5, and 0.1 M NaCl (50 phosphorylase buffer). The mixture was kept on ice for 10 min and 0.1 ml of 10% phosphatase buffer (100 mM NaH2PO4, 100 mM Na2HPO4, 0.1% sodium dodecyl sulfate, 0.1% Triton X-100, 10 mM EDTA) and 0.01% bromophenol blue was added. The mixture was then centrifuged at 10,000 rpm for 90 min, and 200 μl aliquots of the supernatant were incubated with each antibody (15 μl) overnight. The complexes were adsorbed to Pansorbin for 1 h on ice, and collected by centrifugation. Pellets were washed twice in 500 μl of 1X phosphorolysis buffer and resuspended in 0.125 M Tris-HCl buffer, pH 6.8, 0.25 M sodium dodecyl sulfate, 10% glycerol, 0.01% bromophenol blue, 0.1% EDTA, and 0.5% mercaptoethanol (50 μl). Samples were boiled for 10 min and aliquots (50 μl) were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (15% polyacrylamide separating gel, 4% polyacrylamide stacking gel). The gel was washed 3 times for 5 min in 8 M urea and were for 30 min in 1 M sodium periodate, digested, and subjected to autoradiography.

**Alpase assay**

Reaction mixtures (25 μl) contained 50 mM Tris-Cl, pH 8.3, 2 mM GTP, 2 mM MgCl2, 1 mM dATP (2 μM/ml), and levels of enzyme varying between 20 and 75 units. The mixtures also contained either no DNA, Phage λ DNA, single stranded circles (100 μg), phage λ DNA digested with HindIII (100 μg), or EcoRI (100 μg), or 0.5 μg

**Purification of Reverse Transcriptase Activity**

Seven grams (wet weight) of H6011 cells containing pBB15.23 were disrupted and made 30 μg/ml Tris-HCl, pH 7.5, 0.1% sodium deoxycholate, 0.15 M NaCl, 1 mM EDTA, and 1 mM PMF to a final w/v ratio of 1:4. The enzyme (1 mg/ml final concentration) was added and the suspension was kept on ice for 20 min. Mobilin PSO was then added from a final concentration of 0.15. The mixture was incubated an additional 5 min, was made 100 mM NaCl, and centrifuged at 40,000 rpm for 1 h. The supernatant was dialyzed for 1 h against Buffer M, and then diluted with Buffer M to a conductivity equivalent to that of buffer H plus 75 mM NaCl (total volume 2X ml). The fraction was loaded onto a DEAE column (6 x 9.5 cm) (Dowex), which was equilibrated with Buffer H plus 75 mM NaCl. The column was washed with the same buffer. Most of the reverse transcriptase activity was not retained by the column (see Table 1). The flow through fraction was dialyzed directly into a phenol-chloroform:isoamyl alcohol column (2.5 x 28 cm) (Dowex) equilibrated with Buffer H plus 75 mM NaCl. The column was washed with the same buffer and eluted with a 60 ml linear gradient of 15-200 mM NaCl in Buffer H. Fractions (10 μl) were collected and assayed for reverse transcriptase and RNAse activity. The reverse transcriptase activity was completely retained on the column and eluted as a single peak between 0.25-0.60 M NaCl (see Fig. 2). The predominant RNAse activity was coincident with the reverse transcriptase activity. Phenol-chloroform fractions 27-30 were pooled (40 μl). Alkyl (10 μl) were diluted in Buffer H to a conductivity equivalent to 150 mM NaCl in Buffer H and loaded individually onto an agarose-polyacrylamide gel column (1.5 x 8.5 cm) equilibrated in Buffer H containing 50 mM NaCl. The column was washed with 1 column volume of the same buffer and was loaded individually onto a 15-30% glycerol gradient of 50 to 300 mM NaCl in Buffer H. The reverse transcriptase activity, the RNAse H activity, and the reverse transcriptase activity (RNAse H activity was coincident) were eluted in 0.5 column volumes at 0.35-0.50 M NaCl (see Fig. 4). Fractions 24-34 were pooled and concentrated by ultraviolet light chromatography (see Fig. 2). The fraction was then applied to a hydrophobic column chromatography (2.5 x 3 cm) equilibrated in Buffer H containing 0.2 M NaCl, washed with 1 column volume of the same buffer, and eluted with Buffer H containing 0.2 M NaCl and 100 mM NaF, pH 7.0. Fractions (1 μl) were collected, those containing reverse transcriptase activity were dialyzed for 1 h against storage buffer and kept at -70°C in aliquots.