Replication of Lengthened Moloney Murine Leukemia Virus Genomes Is Impaired at Multiple Stages

NAM-HEE SHIN,1 DENNIS HARTIGAN-O’CONNOR,2 JULIE K. PFEIFFER,1
AND ALICE TELESNITSKY1,4

Department of Microbiology and Immunology1 and Medical Scientist Training Program;2 University of Michigan
Medical School, Ann Arbor, Michigan 48109-0620

Received 20 August 1999/Accepted 8 December 1999

It has been assumed that RNA packaging constraints limit the size of retroviral genomes. This notion of a retroviral “headful” was tested by examining the ability of Moloney murine leukemia virus genomes lengthened by 4, 8, or 11 kb to participate in a single replication cycle. Overall, replication of these lengthened genomes was 5- to 10-fold less efficient than that of native-length genomes. When RNA expression and virion formation, RNA packaging, and early stages of replication were compared, long genomes were found to complete each step less efficiently than did normal-length genomes. To test whether short RNAs might facilitate the packaging of lengthy RNAs by heterodimerization, some experiments involved coexpression of a short packageable RNA. However, enhancement of neither long vector RNA packaging nor long vector DNA synthesis was observed in the presence of the short RNA. Most of the proviruses templated by 12 and 16 kb vectors appeared to be full length. Most products of a 19.2-kb vector contained deletions, but some integrated proviruses were around twice the native genome length. These results demonstrate that lengthy retroviral genomes can be packaged and that genome length is not strictly limited at any individual replication step. These observations also suggest that the lengthy read-through RNAs postulated to be intermediates in retroviral transduction can be packaged directly without further processing.

Regardless of whether a particular type of retrovirus encodes only Gag, Pol, and Env or if it also encodes additional proteins, most replication-competent retroviral genomes are fairly similar in length—roughly 8 to 9.5 kb long—with only a few outliers such as the 12.3-kb walleye dermal sarcoma virus (15). This suggests that there may be a selective advantage for retroviruses to maintain relatively small genomes, a notion supported by the observation that most acute transforming retroviruses acquired host-derived oncogenes at the expense of essential viral coding regions rather than by lengthening their genomes (23). It has been suggested that retroviral genome size is limited at the level of RNA packaging (10). Although some replication-competent retroviral vectors with inserted sequences are fairly stable (notably some in which the src region of Rous sarcoma virus [RSV] is replaced by heterologous sequences [26]), the observation that sequences added to other replication-competent retroviruses are rapidly deleted during virus passage has been cited as evidence for an upper size limit to packageable retroviral RNAs (10, 20). However, these later findings do not exclude the possibility that the additional sequences may have been lost simply because there was no advantage in retaining them, or that size limitations may be imposed at replication stages other than packaging. In studies that examined the RNAs encapsidated by an RSV derivative with a defective polyadenylation signal, most encapsidated RNAs were not significantly longer than standard genome length (14). Although differences in cellular and virion RNA populations were not explored in that study, these findings were interpreted as suggesting that longer read-through RNAs were specifically excluded from encapsidation (14).

Retroviruses copackage two complete copies of their genomes joined together in a noncovalent dimer linkage structure (2). Numerous genetic studies have shown that retroviruses can copackage two RNAs that differ significantly from one another if the RNAs retain compatible packaging and dimerization regions. Retroviruses can even copackage a normal genome-length RNA with a much shorter RNA (16). Experimental evidence suggesting that two RNA dimers—a total of four RNAs—can be coencapsidated when viral RNAs are short (32) is consistent with the notion that retroviruses have fairly fixed packaging capacities which can be reached in ways other than the copackaging of two same-length RNAs.

It is not clear why maintaining a two-RNA genome is advantageous for retroviruses. Genetic studies have demonstrated that whereas both retroviral RNAs can and often do participate in DNA synthesis, a single RNA is sufficient to template retroviral DNA (17). Although packaging may precede dimerization for at least some kinds of retroviruses (reference 24 and citations therein), the close genetic linkage between genome regions implicated in dimerization and in packaging has been suggested to reflect that the dimer may serve principally to provide the RNA structure required for specific RNA encapsidation (30). A corollary of this suggestion is that any RNA which could form a dimer with a viral RNA might facilitate packaging of the viral RNA. Experiments in this report tested the notion that if packaging capacity were determined by the total mass of RNA in a virion rather than by the length of individual RNAs, then the packaging maximum could be reached by copackaging two different-length RNAs, one of which could be much longer than the native genome. The findings were not consistent with this initial hypothesis but did demonstrate that lengthy RNAs can be encapsidated into murine leukemia virus (MLV) virions.
MATERIALS AND METHODS

Plasmid construction. Dystrophin cDNA-containing gap-pol-puramycin gene (pur) expression plasmid pGPP (all other nomenclature of pGPP-133-11) has been described elsewhere (27). The version of pGPP used here (pD1012-9) differed from pNH 133-11 by the inclusion of a polylinker between Moloney MLV (M-MLV) sequences and the simian virus 40 promoter. This polylinker, which introduced EcoRI, BclI, AscI, BstNI, NcoI, and Smal/Xmal sites, was introduced by digesting a double-stranded oligonucleotide containing 5’ GGAAATTCGATACAGGGCCCGCTTCGAGG with EcoRI and MfeI and then ligating this into EcoRI-cleaved pNH 133-11. Murine dystrophin cDNA fragments 3.3 and 5.8 kb were generated by PCR using the primer set (19) which contains the intact 14-kb murine dystrophin cDNA, and inserted into the pD1012-9 oligonucleotides. The dystrophin sequences were engineered to contain Xmal sites at their termini; PCR products were digested with Xmal and inserted into the NcoI site of pD1012-9. All dystrophin PCR fragments shared a common 5’ end (starting in the fourth codon of the dystrophin coding region) and differed in their 3’ ends: the 3’ end of dystrophin sequences in pGPP-9 was 5’ AGGAAACGTTTTCAGAAAGCTTTTCC. In pGPP-11 it was 5’ AGGAAACGCTTTTCC. 110 (12) in transfection mixtures for parallel control plates and then staining for imethrine bromide; Sigma) per ml for 2 h at 37°C. Titers were determined by spectrophotometry. Transfection efficiency tests were performed by including (Gibco/BRL), purified using Qiagen plasmid maxi kits, and quantified by UV necessary. Plasmids were propagated in the Stbl2 strain of pGPP, appropriate amounts of pUC19 DNA were added as carrier DNA as

endpoint dilution. Puromycin was added directly to infected cells without cell passage. Well-separated puromycin-resistant colonies were selected and expanded for study as clonal integrants.

Purification and analysis of viral nucleic acids. (i) Cellular RNA. At 48 h posttransfection, medium was removed by aspiration and 1 ml of Trizol reagent (Gibco/BRL) was added to each 60-mm-diameter plate of cells. Subsequent steps were performed according to the manufacturer’s instructions. Samples were suspended containing 25 μl of 50 mM Tris-Cl (pH 8.5), 50 mM MgCl2, 5 mM dithiothreitol, 1.5 U of RNase-free DNase (Boehringer Mannheim Biochemicals [BMB]) per ml, and 0.5 U of RNasin (BMB) per ml, incubated 15 min at 37°C, and then stopped by addition of 25 μl of 50 mM EDTA–1 M sodium citrate (pH 4.5). The solution was then treated with 10 μl of 10 mg/ml proteinase K (Boehringer Mannheim) at 55°C for 30 min with brief mixing. RNA was then extracted with an equal-volume mixture of phenol plus chloroform-isomyl alcohol (24:1, vol/vol) and ethanol precipitated. Dried pellets were resuspended in diethyl pyrocarbonate-treated distilled H2O.

Controls to ensure that DNase treatment reduced transfected plasmid DNA in RNA preparations to below Northern blot or RNA protection assay detection levels involved processing samples from parallel experiments in which cells were transfected with plasmids that contained probe-complementary sequences but which lacked polymerase II promoters.

(ii) Viral RNA purification. Virus-containing medium was collected and filtered 12 h after 24 h posttransfection and centrifuged at 500,000 x g in a Sorvall RC M120EX ultracentrifuge for 25 min at 4°C. Pellets were resuspended in 250 μl of 50 mM Tris (pH 7.5)–10 mM EDTA–1% SDS–100 mM NaCl and the ratio of RNA/ml–100 μg of protein per ml were appropriate, each sample was then DNase treated and processed as described for viral RNA probes were prepared by cleaving pD1040-2 with EcoRI and transcribing with T3 RNA polymerase or by cleaving pB481-7 with BglII and transcribing with T7 RNA polymerase. The pD1040-2 406-base runoff transcript protects 330 bases of gag sequences when annealed to wild-type MLV RNA, while the 237-base pB481-7 runoff transcript protects 194 bases of puro RNA. Cleavage of pD1040-2 with BglII and divergent transcription with T7 RNA polymerase yielded recovery marker RNA. In vitro-transcribed RNAs were generated using standard approaches (1), with the inclusion of a 50-μg/ml concentration of GpppG to generate radiolabeled probes. Radiolabeled was omitted and higher substrate concentrations were used to produce recovery marker RNA. Recovery marker was included in experiments presented in Fig. 2A, 3A, and 3C. Recovery marker was not added to RNA samples for experiments in Table 1 and Fig. 3B, where the only required measurement was the ratio of escort and GPP-type RNAs.

(iii) RNA protection assays. Viral and cellular RNA and radiolabeled RNA probes were prepared as indicated above, and the same amount of RNA probe, which exceeded the total amount of sample to be tested by a molar factor of 5 to 50, was used for all samples within each experiment. Samples were hybridized, digested with RNAses A and T, processed, and separated on 5% polyacrylamide–8 M urea gels by standard techniques (1). Dried gels were exposed to film and/or analyzed with a Molecular Dynamics PhosphorImager and ImageQuant software. In experiments where multiple RNA probes were used to produce recovery marker RNA, recovered marker RNA was included in experiments presented in Fig. 2A, 3A, and 3C. Recovery marker was not added to RNA samples for experiments in Table 1 and Fig. 3B, where the only required measurement was the ratio of escort and GPP-type RNAs.

Virion proteins were quantified by performing quantitative reverse transcription–PCR analysis (qPCR) as described elsewhere (31). All values were confirmed by assaying more than one concentration of enzyme, and all samples used were at least 10-fold above uninfected control medium values and were within the linear range of the assay. All virions in this study should be formed of identical Gag-Pol transcribed from the 5’ end of the cytomegalovirus promoter-driven expression plasmid pCH110 (13). (32) by transient transfecting them with pGPP and then testing the resulting virus for the ability to confer puromycin resistance to 293T or D17 cells. The cell clone that produced the highest puromycin-resistant titer was used in several individual clones were functionally tested for the ability to supply GaLV gag-pol sequences when annealed to wild-type MLV RNA, while the 237-base pB481-7 runoff transcript protects 194 bases of puro RNA. Cleavage of pD1040-2 with BglII and divergent transcription with T7 RNA polymerase yielded recovery marker RNA. In vitro-transcribed RNAs were generated using standard approaches (1), with the inclusion of a 50-μg/ml concentration of GpppG to generate radiolabeled probes. Radiolabeled was omitted and higher substrate concentrations were used to produce recovery marker RNA. Recovery marker was included in experiments presented in Fig. 2A, 3A, and 3C. Recovery marker was not added to RNA samples for experiments in Table 1 and Fig. 3B, where the only required measurement was the ratio of escort and GPP-type RNAs.

Virion RNA was purified by a random-primed DNA labeling kit from BMB or Redi-Prime kit from Amersham. EcoRI-linearized pD1040-2 was used to make RNA blot probes, while puromycin resistance or dystrophin gene fragments were used to make probes for DNA blots.

For viral RNA blotts, two 6-cm-diameter plates of ET cells which stably expressed escort RNA were transfected with pGPP or derivatives and RNA was processed as above. RNA was separated on 0.7% formaldehyde-agarose gels by standard techniques (1). Dried gels were exposed to film and/or analyzed with a Molecular Dynamics PhosphorImager and ImageQuant software. In experiments where multiple RNA probes were used to produce recovery marker RNA, recovered marker RNA was included in experiments presented in Fig. 2A, 3A, and 3C. Recovery marker was not added to RNA samples for experiments in Table 1 and Fig. 3B, where the only required measurement was the ratio of escort and GPP-type RNAs.

Viral and cellular RNA and radiolabeled RNA probes were prepared as indicated above, and the same amount of RNA probe, which exceeded the total amount of sample to be tested by a molar factor of 5 to 50, was used for all samples within each experiment. Samples were hybridized, digested with RNAses A and T, processed, and separated on 5% polyacrylamide–8 M urea gels by standard techniques (1). Dried gels were exposed to film and/or analyzed with a Molecular Dynamics PhosphorImager and ImageQuant software. In experiments where multiple RNA probes were used to produce recovery marker RNA, recovered marker RNA was included in experiments presented in Fig. 2A, 3A, and 3C. Recovery marker was not added to RNA samples for experiments in Table 1 and Fig. 3B, where the only required measurement was the ratio of escort and GPP-type RNAs.

Viral and cellular RNA and radiolabeled RNA probes were prepared as indicated above, and the same amount of RNA probe, which exceeded the total amount of sample to be tested by a molar factor of 5 to 50, was used for all samples within each experiment. Samples were hybridized, digested with RNases A and T, processed, and separated on 5% polyacrylamide–8 M urea gels by standard techniques (1). Dried gels were exposed to film and/or analyzed with a Molecular Dynamics PhosphorImager and ImageQuant software. In experiments where multiple RNA probes were used to produce recovery marker RNA, recovered marker RNA was included in experiments presented in Fig. 2A, 3A, and 3C. Recovery marker was not added to RNA samples for experiments in Table 1 and Fig. 3B, where the only required measurement was the ratio of escort and GPP-type RNAs.

Viral and cellular RNA and radiolabeled RNA probes were prepared as indicated above, and the same amount of RNA probe, which exceeded the total amount of sample to be tested by a molar factor of 5 to 50, was used for all samples within each experiment. Samples were hybridized, digested with RNases A and T, processed, and separated on 5% polyacrylamide–8 M urea gels by standard techniques (1). Dried gels were exposed to film and/or analyzed with a Molecular Dynamics PhosphorImager and ImageQuant software. In experiments where multiple RNA probes were used to produce recovery marker RNA, recovered marker RNA was included in experiments presented in Fig. 2A, 3A, and 3C. Recovery marker was not added to RNA samples for experiments in Table 1 and Fig. 3B, where the only required measurement was the ratio of escort and GPP-type RNAs.
Construction of large RNA genomes and small escort RNAs.

The goal of this study was to assess effects of retroviral genome length on RNA packaging and other replication steps. In the approach used, both the encapsidated RNAs and all proteins necessary for the production of virions were encoded by single expression plasmids. The parental expression plasmid, pGPP, was an M-MLV provirus plasmid in which the splice donor site was inactivated and the env coding region was replaced with a puromycin resistance expression cassette (27) (Fig. 1). This vector encodes an 8.1-kb viral RNA, which is essentially the same length as the native MLV genome (8.2 kb).

We generated lengthened derivatives of pGPP (Fig. 1) that contained fragments of murine dystrophin cDNA inserted between pol and the puromycin resistance gene cassette. Inserts of 3.9, 7.9, and 11.1 kb were made to yield pGPP+4, pGPP+8, and pGPP+11, which are predicted to produce viral RNAs 12, 16, and 19.2 kb in length, respectively. pC1hyg, a plasmid that encodes a short packageable RNA, was also constructed to test the ability of short RNAs to heterodimerize with long viral RNAs and facilitate their encapsidation (Fig. 1). Reflecting this postulated role of C1hyg RNA, this short vector is herein referred to as an escort RNA. pψhyg is a splice donor-defective derivative of an MLV provirus plasmid which contains a hygromycin B phosphotransferase expression cassette in place of the viral coding sequences, followed by the herpes simplex virus polyadenylation signal. Blotting analysis revealed that expression of pψhyg yielded intracellular RNAs of the predicted ~3.1-kb length (data not shown).

The rationale for using a single construct to express both virion proteins and genomic RNA was as follows. If an alteration to GPP did not affect packaging, then when that GPP derivative was expressed in mammalian cells, the relative amounts of RNA which would be directed to mRNA and to encapsidated genome functions would remain the same as for the parental GPP vector. In contrast, if changes introduced into a GPP derivative caused RNA encapsidation defects, then the amount of RNA per unit of viral protein in extracellular virions would decrease. Hence, comparing the molar amount of vector RNA per unit extracellular virion protein of the GPP vector to values for its derivatives would provide a way of examining effects on encapsidation.

Relative replication efficiency of large and small genomes.

As an initial test of these vectors’ ability to support a single replication cycle, pGPP and its derivatives were transfected into ET cells, which are human 293T-derived cells that stably express ecotropic Env (27), and the resulting virion titers were determined. Equimolar amounts of each pGPP-type plasmid were transfected into ET cells. Parallel experiments aimed at examining the effects of the C1hyg escort RNA were also performed by transfecting secondary plates of ET cells with the same amounts of pGPP-type plasmids supplemented with equimolar pψhyg. Fresh murine cells were then infected with an equal volume of each virus stock, and puromycin-resistant colony titers were determined.

Results from two such experiments are presented in Fig. 2A and B. Titers are presented as puromycin-resistant CFU per volume of virus-containing medium, standardized to the titer for the parental GPP expressed in the absence of C1hyg. The titer of the parental vector was roughly 10⁴ puromycin-resistant CFU/ml. These initial trials demonstrated that product titers were around 5- to 10-fold lower for large retroviral vectors than for native-length ones generated in experiments where
virus was produced from transiently transfected cells. A comparison of the data in Fig. 2A (no pΨ+hyg) and 2B (cotransfection with pΨ+hyg) indicates that pΨ+hyg did not increase the puromycin-resistant CFU of long vectors and may have decreased titers slightly. This finding appears to be inconsistent with the initial hypothesis that pΨ+hyg might enhance packaging and the generation of proviral products when coexpressed with lengthy viral RNAs. Note that although the lengthened genomes all yielded lower titers than native-length GPP, there were no consistent size-related trends in apparent replication efficiency among lengthened genomes (GPP+4, GPP+8, and GPP+11). Titers normalized to the amount of virion protein in the culture medium are presented in Fig. 2C and D. As assessed by levels of virus protein in the culture medium, the amount of virus produced by cells transfected with larger vector plasmids was slightly (less than twofold) lower than the amount produced by cells transfected with the parental pGPP. Whether this nominal but reproducible decrease in virion protein release from transfected cells resulted from reduced transfection efficiency for large plasmids or an effect on viral replication was not determined.

Assessing packaging biases between large and small genomes. The stoichiometries of pΨ+hyg and GPP-type RNAs within cotransfected cells were compared to their ratios in virions. The value which described the molar amount of large GPP-type vector RNA per pΨ+hyg RNA in the presence and absence of pΨ+hyg escort RNA, (G and H) Puromycin-resistant CFU per unit of virion RNA in the presence and absence of pΨ+hyg escort RNA. The values in panel B were normalized to the GPP value in panel A. For all other panels, values presented were normalized to values obtained for the parental GPP vector in that panel and were determined from data such as those in Fig. 3, using approaches described in Materials and Methods. In panels E through H, “encapsidated RNA” refers to GPP, GPP derivative, or pΨ+MLV RNA and does not include pΨ+hyg escort RNA.

FIG. 2. Replication efficiencies of native-length and lengthened genomes. (A and B) Puromycin-resistant CFU per unit volume of virus-containing culture medium harvested from ET cells transiently transfected with equimolar amounts of vector expression plasmids, normalized to the value obtained for the parental pGPP plasmid. In experiments that included the pΨ+hyg escort plasmid, equimolar pΨ+hyg was added in place of some of the carrier DNA used in the no-pΨ+hyg transfections. (C and D) Puromycin-resistant CFU per unit of virion protein in the culture medium. (E and F) Packaging of viral RNA per unit virion protein in the presence and absence of pΨ+hyg escort RNA. (G and H) Packaging of viral RNA per unit virion RNA in the presence and absence of pΨ+hyg escort RNA. The values in panel B were normalized to the GPP value in panel A. For all other panels, values presented were normalized to values obtained for the parental GPP vector in that panel and were determined from data such as those in Fig. 3, using approaches described in Materials and Methods. In panels E through H, “encapsidated RNA” refers to GPP, GPP derivative, or pΨ+MLV RNA and does not include pΨ+hyg escort RNA.
defective in packaging, then it would be less prevalent relative to \( \Psi^+ \) hyg in virions than in producer cells. Although no attempt was made to determine whether proteins were limiting or RNA was expressed at saturating concentrations, these measurements were interpreted as addressing the ability of the lengthened RNAs to compete with short RNAs for the packaging machinery.

The data were generated by cotransfecting ET cells with equimolar amounts of \( \Psi^+ \) hyg and pGPP or its derivatives and then determining RNA ratios in both cells and virions by RNase protection assays. The RNase protection probe spanned the junction of sequences that the GPP vectors share with the \( \Psi^+ \) hyg escort and sequences in which they differ (Fig. 1, pD1040-2). Thus, a single probe could be used to detect both long vector and short escort RNAs as protected products of distinct sizes, and the molar ratios of the two RNAs could be determined by the ratios of these protected bands. In experiments that addressed RNA packaging in the absence of \( \Psi^+ \) hyg, packaging biases were examined by determining how much RNA was encapsidated per unit of extracellular virion protein. Because quantifying virion RNAs required several purification steps, viral RNA recovery was monitored by introducing an RNA recovery marker early during virion RNA purification and then normalizing values based on the amount of recovery marker in the final product. The recovery marker was a nonradiolabeled in vitro transcript complementary to a short portion of the same RNase protection probe used to detect \( \Psi^+ \) hyg and GPP RNAs.

Figure 3 shows the results of RNase protection assays performed on cellular and virion RNA harvested from producer cells transfected with pGPP or related plasmid in the presence or absence of \( \Psi^+ \) hyg. The intracellular ratios of \( \Psi^+ \) hyg and each larger coexpressed RNA, the ratios of these two types of RNAs within virions, and the ratios of these two values (the packaging factor) are presented in Table 1. These values were not corrected for virion numbers but instead describe the relative amounts of the indicated RNAs. Controls were performed in which a wild-type MLV provirus plasmid (pNCA) (6) or a packaging-defective provirus plasmid (\( \Psi^- \) MLV [27]) was cotransfected with \( \Psi^+ \) hyg. As anticipated, encapsidation of \( \Psi^- \) packaging mutant RNAs was strongly disfavored relative to \( \Psi^+ \) hyg RNAs (Table 1, \( \Psi^- \) MLV). The relative prevalence of the parental GPP RNA and the \( \Psi^+ \) hyg escort in virions was very similar to that observed intracellularly, suggesting that GPP and \( \Psi^+ \) hyg were equally fit for encapsidation. In contrast, modest biases against packaging were observed for the dystrophin-containing lengthened GPP derivative RNAs, with each of these long RNAs around half as prevalent in virions as in cells relative to the amounts of \( \Psi^- \) hyg (Table 1).

To address the contribution of \( \Psi^- \) hyg to this fairly efficient encapsidation of long RNAs, levels of large RNAs in virions produced both in the presence and in the absence of \( \Psi^- \) hyg were compared. An RNase protection assay of virion RNAs produced in the absence of \( \Psi^- \) hyg is shown in Fig. 3C. Amounts of viral RNA per unit virion proteins in the presence and in the absence of \( \Psi^- \) hyg escort are compared in Fig. 2E.

### Table 1. Relative prevalence and packaging of large and escort RNAs

<table>
<thead>
<tr>
<th>Cotransfected large vector plasmid</th>
<th>Ratio, large/( \Psi^- ) hyg (mean ± SD)</th>
<th>Packaging factor&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cellular RNA</td>
<td>Virion RNA</td>
</tr>
<tr>
<td>pGPP</td>
<td>0.78 ± 0.01</td>
<td>0.73 ± 0.02</td>
</tr>
<tr>
<td>pGPP +4</td>
<td>0.65 ± 0.11</td>
<td>0.36 ± 0.06</td>
</tr>
<tr>
<td>pGPP +8</td>
<td>0.59 ± 0.01</td>
<td>0.29 ± 0.00</td>
</tr>
<tr>
<td>pGPP +11</td>
<td>0.55 ± 0.01</td>
<td>0.33 ± 0.02</td>
</tr>
<tr>
<td>pNCA (M-MLV)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td>( \Psi^- ) MLV&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.79 ± 0.01</td>
<td>0.07 ± 0.01</td>
</tr>
</tbody>
</table>

<sup>a</sup> Values for wild-type M-MLV (pNCA) were determined only once.

<sup>b</sup> Defined as the ratio of virion RNA large/\( \Psi^- \) hyg divided by cellular RNA large/\( \Psi^- \) hyg.
and F. These data suggest that coexpression of the shortΨ+hyg escort RNA did not facilitate long vector packaging.

**Examining viral RNAs designed to exceed wild-type length.**

As assayed above, long vector packaging appeared to be reduced less than twofold relative to the native-length vector, while titers per unit virion protein were decreased roughly four- to fivefold. It seemed possible that some component of the decreased titer per unit RNA for the lengthened GPP derivatives might have resulted if more of the RNAs were incomplete in long vector virions than in those of native-length vectors. For the data shown in Fig. 2E through H, RNAs were quantified by RNase protection assays that used a probe complementary to a short portion of RNA near 5′ ends of the genomes, and no attempt was made to determine the amount of 3′ RNA sequences. Thus, to assess whether virions may have differed in full-length RNA content, RNase protection assays were performed on virion RNAs, using excess amounts of two different probes in each reaction. One probe was complementary to sequences near the 5′ ends of the RNAs, and the other was complementary to puromycin gene sequences near the 3′ ends.

The results (Fig. 4) were quantified by a PhosphorImager. The molar amount of each product was determined as described in Materials and Methods, and then the molar ratios of the 5′ and 3′ portions of the RNAs were compared. For GPP, the determined molar ratio of 5′ to 3′ ends was 1.01; for GPP+4, the ratio was 0.98; for GPP+8, it was 1.16; and for GPP+11, the ratio was 1.29. Thus, the molar ratio of 5′ and 3′ ends for native-length GPP did not differ greatly from values for its lengthened derivatives. Similar results were obtained when a single probe derived from the long terminal repeat which protected both 5′ and 3′ sequences, was used.

Because results in Fig. 4 did not differentiate between whether full-length RNAs were packaged or if virions contained shortened RNAs generated by the use of clpI splice sites or other unintended genome modifications, Northern blots were used to assess viral RNA lengths both in virus-producing cells and in extracellular virions. Figure 5A shows a blot of total cellular RNA from cells expressing the GPP derivatives probed with a fragment of gag. To assess RNA lengths, samples were heat denatured and separated on denaturing gels. Using the mobilities ofΨ+hyg and wild-type MLV RNA as size standards, this blot shows that RNAs of roughly the predicted lengths were generated in cells expressing GPP derivatives. Data for virion RNAs are shown in Fig. 5B. At least some of the encapsidated RNAs for each vector appeared to be of the full intended length, but significant smears of faster-mobility material were also detected. There were no detectable discrete-length RNAs among these shorter products, suggesting that the virion RNA was not dominated by specific unintended spliced RNAs.

**Effects of genome length on efficiency of postpackaging steps.** Having obtained data suggesting that packaging defects were fairly modest, it seemed possible that some of the observed titer decreases for long genomes might reflect defects in postpackaging steps of the viral replication cycle. Size-related defects might include disrupting maturation of viral particles or RNA (9), altering the stoichiometry of putative accessory factors such as nucleocapsid (7), perturbing the higher-order structure required for reverse transcription (17), or interfering with the sustained synapsing of viral DNA ends required for integration (22).

To address whether length-related limitations were imposed during early stages of retroviral replication, the titer of puromycin-resistant proviruses formed per unit of encapsidated RNA were compared for parental GPP and for GPP+4, GPP+8, and GPP+11. The results (Fig. 2G and H) indicate that all of the lengthened vectors generated roughly threefold-fewer colonies per mole of encapsidated RNA than did parental GPP.

**Proviral DNAs templated by lengthened vectors.** Clonal integrated proviruses generated from GPP-type vectors were examined by Southern blotting. The restriction digests used to analyze these proviruses are indicated at the top of Fig. 6, and some of the blots used to analyze clonal GPP+11 integrants are displayed below. Figure 6A shows a blot of ClaI digests of the indicated lengths in kilobases. The identity of the faster-migrating band in lanes 1 and 2 of panel A was not determined.
genomic DNA from several individual GPP+11 proviral clones which contained shorter-than-anticipated dystrophin segments. As indicated at the top of Fig. 6, an intact GPP+11 provirus would be predicted to yield a 10.5-kb band detectable on these blots. Note that ClaI fragments detected for the GPP+11 integrants analyzed in Fig. 6A varied in length, with several appearing to have products of the same altered mobility (Fig. 6A, lanes 4 to 6). The frequent occurrence of similar-sized products suggests that these may have had a common origin such as a 2.7-kb deletion which resulted from unintended splicing or from a hot spot for recombination during reverse transcription. Genomic rearrangements such as deletions after a single round of replication. Overall, these results suggest that most products of GPP+11 contained deletions after a single round of replication.

Because GPP vectors contain gag and pol sequences, a second round of replication was possible when virions shed by cells transduced with GPP proviruses were provided Env. This was accomplished by infecting 3T3 cells that expressed GaLV Env as the endpoint of the first replication round (see Materials and Methods) and then infecting fresh D17 cells with virions produced by these cells. Integrated products of some GPP+11 proviruses generated during a second round of replication were examined by Southern blotting (Fig. 6C). Among the GPP+11 products, proviruses which were 12.8 and 14.3 kb long after the first replication round generated progeny which appeared to maintain these lengths on secondary passage (Fig. 6C, lanes 3, 4, and 6), but the tested product of a full-length 19.8 kb provirus was reduced to an apparent length of 16.0 kb during a second replication round (Fig. 6C, lane 5). Similar approaches were used to examine clonal GPP+4 and GPP+8 integrants (data not shown). All four GPP+4 progeny proviruses tested appeared to be full length, as did two of four GPP+8 integrants. In addition to the two GPP+8 integrants that yielded the restriction patterns predicted for unrearranged

FIG. 6. Southern blot analysis of individual proviral products of the GPP+11 vector. (Top) Schematic drawing of a GPP+11 provirus showing the locations of restriction sites used to assess provirus sizes. Because of the template switches during reverse transcription, MLV-based vector DNAs should be 0.6 kb longer than their RNA templates (11). ClaI restriction sites and the size (10.5 kb) of the ClaI restriction fragment predicted for an unrearranged GPP+11 provirus are indicated, as are BstEII site locations and the sizes predicted for BstEII digestion (2.7, 3.2, and 6.3 kb). Drawing is not to scale. For abbreviations, see the legend to Fig. 1. (Bottom) Southern blot analyses of integrated vector proviruses. (A) GPP+11 clonal integrants that contain dystrophin region deletions digested with ClaI. Lanes: 1, size standards; 2, mock-infected cell DNA; 3 to 6, individual clonal integrants. (B) Candidate full-length GPP+11 GM integrant DNA analyzed with BstEII. Lanes: 1, size standards; 2, mock-infected cell DNA; 3, clone 4-8 DNA. (C) Restriction analysis of GPP+11 vector product sizes after a second round of replication. D17 clone DNAs were digested with ClaI. Lanes: 1, marker; 2, mock-infected D17 cellular DNA; 3, second-round product generated by clone 16-6; 4 to 6, second-round products from 4-4, 4-8, and 4-14, respectively; 7, size standards. Mobilities of DNA size standards are indicated at the left. For each panel, arrows indicate provirus-specific products and Δ indicates host-derived products detectable in both mock-infected and transduced cells. First-replication-round products were analyzed in 3T3 cells that express GaLV Env; second-round products were analyzed in D17 cells.
proviruses, one yielded a ClaI fragment approximately 2 kb longer than predicted for a full-length GPP+8 product, and the fourth contained a ClaI fragment about 2.3 kb shorter than predicted.

**DISCUSSION**

In this report, the packaging efficiencies of native-length retroviral RNAs were compared to those of RNAs significantly longer than wild type. The ability of lengthy RNAs to compete with short RNAs for the MLV packaging machinery and the ability of long genomes to participate in subsequent replication steps were examined. Overall, replication of lengthened genomes was reduced 5- to 10-fold relative to native-length genomes. This reduction was the product of modest defects in virion production and in packaging, and somewhat greater defects in subsequent stages, with large vector replication not absolutely restricted at any one stage. RNAs twice the natural genome length could become encapsidated and could template integration-competent DNA. Although many of the resulting proviruses contained detectable defects such as deletions, in some cases, progeny proviruses were able to form infectious particles that could complete a subsequent round of replication. These findings demonstrate that large genome replication is impaired at several stages but that there are no absolute size restrictions against genomes twice the normal length at any stage of the MLV replication cycle.

A premise of this study had been that dimers of lengthy RNAs would be excluded from packaging. Therefore, in some experiments, a short packageable RNA, \( \Psi \) hyg, was coexpressed with the lengthy vectors to test whether short RNA might facilitate long RNA packaging by heterodimerization. Surprisingly, long RNAs were packaged fairly well even in the absence of the \( \Psi \) hyg. No augmentation of lengthy RNA packaging was observed in the presence of the \( \Psi \) hyg escort.

Some of the observed titer decrease for long genomes may have reflected a decrease in the amount of intact template RNA for lengthy vectors. Most lengthy vector virion RNAs did not appear intact on Northern blots, but it is not uncommon for virion RNA to appear degraded (5). This apparently is the case for lengthened RNAs like those analyzed in this report, because several groups have published data showing fairly discrete bands for nonadenatured retroviral RNA but smears when the same virion-associated RNAs were denatured (8, 18, 24). Note, however, that the presence of sequences encompassing the full length of the RNAs would be required to template full-length proviral DNAs. Thus, the observation here that some full-length proviruses were generated implies that at least some virions encapsidated the entire length of the RNAs, even for the longest genomes.

How can the finding that lengthy genomes can complete a replication cycle be reconciled with previous findings regarding the rapid loss of insertions in replication-competent retroviruses? Most likely, the key difference is that the present experiments examined a single replication cycle and hence selective advantages were not assessed. As retroviruses replicate, individual viruses with relatively modest survival advantages can come to dominate populations rapidly (4). In previously reported experiments, the inserted sequences did not benefit virus replication (10, 34) and may even have been lost for the simple reason that they increased genome length and thus the time required for replication. Many types of viruses generate deletion-containing defective interfering particles during undilated virus passage, and coding region-deleted genomes are detectable in retroviral cultures during passage at high multiplicity of infection (3, 33, 39, 40). It has long been recognized that retroviral sequences that do not provide a selective advantage, as is the case for src in culture-adapted RSV, are prone to deletion even with a low multiplicity of infection (38). It seems likely that the nonessential dystrophin insertions studied here would be lost rapidly if the vectors replicated freely.

Regardless of their rates during sustained replication, this demonstration of the packageability of long RNAs has important ramifications for rare retroviral replication events. For example, some models for how retroviruses transduce cellular oncogenes envision initial capture of the host gene as resulting from proviral insertion near or within a host proto-oncogene (35, 36, 42). To conform with prevailing notions of packaging limits, transduction models generally go on to postulate either a DNA-level deletion of the downstream long terminal repeat or aberrant splicing of a read-through RNA to prune a chimeric virus-host RNA to a packageable length. The finding in this report of the encapsidation of RNAs more than twice the native length demonstrates that lengthy read-through RNAs could be packaged directly without further processing.

A possible practical application of encapsidating large retroviral genomes would be the formation of longer integrated proviruses from retroviral vectors. For such applications, only a single replication cycle is generally required, and so competitive disadvantage would not affect the outcome. However, the significant rate of errors observed among product DNAs suggests that such large vectors would be useful only in protocols that allowed screening of several integrants to identify ones that had maintained integrity.

Many questions remain unanswered. Were the lengthy vectors encapsidated as homodimers? As monomers? Might even the “unsorted” vectors have become encapsidated in partnership with fortuitously 3’ truncated RNAs? Do coexpressed RNAs randomly copackate, or do biases for or against copackage of nonhomologous RNAs exist? With the 19.2-kb vectors, has the upper size limit of packageable RNAs been reached or can even longer RNAs be encapsidated? Further study is required to address these and related questions.

**ACKNOWLEDGMENTS**

We thank Maribeth Eiden for providing GaLV env, Jeffrey Chambers for providing dystrophin cDNA, and Ovetta Fuller and Thomas Glaser for other reagents. We thank Nancy Sullivan and M. J. Wierland for critical reading of the manuscript and the anonymous reviewers whose comments improved the final version. We also thank David Rekosh and Alan Engleman for advice and Heidi Ackerby, Deanna Kulpa Storn, and the Horace H. Rackham School of Graduate Studies for assistance and support in early stages of this project. This research was supported by NIH grant CA69300 and by the Searle Scholars program of the Chicago Community Trust.

**REFERENCES**

8. Fu, W., R. J. Gorelick, and A. Rein. 1994. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-


