Moloney Murine Sarcoma Virus MuSVts110 DNA: Cloning, Nucleotide Sequence, and Gene Expression

LIN HUAI,[†] SUSANNA M. CHIOCCA,[‡] MARY A. GILBRETH, JERRY R. AINSWORTH,
LYDIA A. BISHOP, AND EDWIN C. MURPHY, JR.*

*Department of Tumor Biology, Section of Molecular Virology, The University of Texas
M. D. Anderson Cancer Center, Houston, Texas 77030*

Received 20 March 1992/Accepted 20 May 1992

We have cloned Moloney murine sarcoma virus (MuSV) MuSVts110 DNA by assembly of polymerase chain reaction (PCR)-amplified segments of integrated viral DNA from infected NRK cells (6m2 cells) and determined its complete sequence. Previously, by direct sequencing of MuSVts110 RNA transcribed in 6m2 cells, we established that the thermosensitive RNA splicing phenotype uniquely characteristic of MuSVts110 results from a deletion of 1,487 nucleotides of progenitor MuSV-124 sequences. As anticipated, the sequence obtained in this study contained precisely this same deletion. In addition, several other unexpected sequence differences were found between MuSVts110 and MuSV-124. For example, in the noncoding region upstream of the *gag* gene, MuSVts110 DNA contained a 52-nucleotide tract typical of murine leukemia virus rather than MuSV-124, suggesting that MuSVts110 originated as a MuSV-helper murine leukemia virus recombinant during reverse transcription rather than from a straightforward deletion within MuSV-124. In addition, both MuSVts110 long terminal repeats contained head-to-tail duplications of eight nucleotides in the U3 region. Finally, seven single-nucleotide substitutions were found scattered throughout MuSVts110 DNA. Three of the nucleotide substitutions were in the *gag* gene, resulting in one coding change in p15 and one in p30. All of the remaining nucleotide changes were found in the noncoding region between the 5' long terminal repeat and the *gag* gene. In NIH 3T3 cells transfected with the cloned MuSVts110 DNA, the pattern of viral RNA expression conformed with that observed in cells infected with authentic MuSVts110 virus in that viral RNA splicing was 30 to 40% efficient at growth temperatures between 28 and 33°C but reduced to trace levels above 37°C.

The Moloney murine sarcoma virus (MuSV) MuSVts110 retrovirus possesses two novel, interrelated characteristics. MuSVts110-infected cells (termed 6m2 cells) appear morphologically transformed at low growth temperatures (33°C or below); above 33°C, however, 6m2 cells rapidly revert to a normal morphology. This conditional ability to transform cells is governed by a novel growth temperature restriction on MuSVts110 RNA splicing acquired as a consequence of a 1,487-nucleotide frameshifting deletion between the *gag* and *mos* frames (18, 27, 28). This large deletion, by including the *v-mos* initiation codon and fusing the *gag* and *v-mos* genes out of register, prevented the translation of the MuSVts110 *mos* frame either independently from its own initiation codon or by read-through from the *gag* gene. However, at the same time, the deletion allowed the activation of in-frame, otherwise cryptic donor and acceptor splice sites surrounding the deletion junction (9, 27, 28). Interestingly, unlike most, if not all, other intron-containing pre-mRNAs in mammalian cells, MuSVts110 RNA is permissive for splicing only at growth temperatures of 33°C or lower. At these low growth temperatures, about 40 to 50% of MuSVts110 transcripts are spliced, but above 33°C, MuSVts110 RNA is spliced progressively less efficiently until only a trace is observed at 39°C (7, 14, 27, 28). Translation of unspliced MuSVts110 RNA produces a truncated *gag* gene product termed P58^{*gag*}. However, splicing realigns the *gag* and *v-mos* frames to allow the translation of P85^{*gag-mos*}, thus

explaining the phenotypic changes observed in 6m2 cells at 33 and 39°C.

Initial attempts to clone integrated MuSVts110 DNA from infected 6m2 cells to study the molecular basis of thermosensitive splicing proved unsuccessful despite the use of a variety of cloning strategies in phage, cosmid, and plasmid vectors (3a). To sidestep this problem, we constructed from wild-type MuSV-124 DNA a viral DNA (termed *ts32* DNA) containing precisely the same 1,487-nucleotide deletion found in authentic MuSVts110 RNA (9). Several subsequent studies indicated that all of the known characteristics of MuSVts110 (e.g., the restriction of cell transformation, viral RNA splicing, and production of P85^{*gag-mos*} to low growth temperatures) were conserved in *ts32*, arguing that the 1,487-nucleotide deletion was the major, if not sole, determinant of the novel pattern of viral RNA splicing and gene expression exhibited by MuSVts110 (7, 9, 10). Moreover, the conditional splicing of *ts32* RNA was shown to be regulated in *cis* by transcript structure (10, 11). Thus, *ts32* seemed to be a reasonable model for studies on splicing competence in MuSVts110 RNA. However, since small differences in transcript structure can have profound effects on splice site choice and splicing efficiency, it has been a concern, in the absence of a complete sequence for integrated MuSVts110 DNA, that we might not be taking into account all of the sequences that regulate MuSVts110 RNA splicing.

In this study, we amplified fragments of the MuSVts110 locus in 6m2 cells and assembled a complete copy of candidate authentic MuSVts110 DNA. After confirming that RNA transcribed from this candidate DNA was spliced in a thermodependent fashion, we determined its complete sequence, confirming the presence of the frameshifting, splice

* Corresponding author.

[†] Present address: Department of Laboratory Medicine, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030.

[‡] Present address: Department of Carcinogenesis, Swiss Institute for Experimental Cancer Research, Lausanne, Switzerland.

FIG. 1. PCR amplification of integrated MuSVts110 DNA. For PCR amplification, three pairs of primers were used. The 5' pair (SUZ5 and 5'D) was complementary to the 5' end of U3 and MuSV-124 nucleotides 2047 to 2070, respectively. The internal pair (SUZ1 and SUZ2) was complementary to nucleotides 1944 to 1967 upstream of the *XhoI* site and nucleotides 2920 to 2944, respectively. The 3' pair (SUZ3 and SUZ6) was complementary to nucleotides 2492 to 2516 and the 3' end of U5, respectively. After PCR amplification, the 2-kb MuSVts110 5' fragment was subcloned as an *EcoRI*-*XhoI* fragment into pGEM-11Z, the 565-base internal fragment was subcloned as an *XhoI*-*AccI* fragment into pSP72, and the 1.8-kb 3' fragment was subcloned as an *AccI*-*BamHI* fragment into pSP72. The assembly of the candidate MuSVts110 DNA was as follows. The internal *XhoI*-*AccI* and *AccI*-*BamHI* fragments were first joined in the pSP72 vector, following which the joint *XhoI*-*BamHI* fragment was joined to the *EcoRI*-*XhoI* fragment in pGEM-11Z, yielding a candidate MuSVts110 DNA.

site-activating 1,487-nucleotide deletion originally found by sequencing regions of MuSVts110 RNA in infected cells (27). In addition, several other interesting sequence deviations from MuSV-124, none of which would appear to have an impact on the splicing phenotype but which instead suggest a recombinant origin for MuSVts110, were found in MuSVts110 DNA.

MATERIALS AND METHODS

Cell cultures and DNA transfection. NRK cells infected with MuSVts110 (6m2 cells) (3) were maintained at 33°C in McCoy medium containing 10% bovine serum supplemented with transferrin (HyClone). NIH 3T3 cells were maintained in Dulbecco high-glucose medium containing 10% bovine serum. Transfected NIH 3T3 cell lines were initially established in Dulbecco high-glucose medium containing 10% fetal bovine serum (HyClone) and then transferred to the same medium containing 10% bovine serum. DNA transfection was carried out by a standard calcium phosphate coprecipitation protocol (13), using linearized viral DNA and pSV2neo DNA. G418-resistant cells were selected in medium containing 400 µg of G418 (Geneticin; GIBCO) per ml.

Cellular DNA isolation. Genomic DNA was isolated from MuSVts110-infected 6m2 cells after lysis in 2% sodium dodecyl sulfate–7 M urea–0.35 M NaCl–1 mM EDTA–20 mM Tris (pH 8.0) as described previously (2).

PCR amplification. Polymerase chain reaction (PCR) amplification of cellular DNA was performed as described elsewhere (31). Cellular DNA (1 to 3 µg) was mixed with 2.5

U of *TaqI* DNA polymerase (Perkin-Elmer Cetus) in a 100-µl reaction mixture containing 100 mM Tris (pH 8.3), 100 mM KCl, 15 mM MgCl₂, 0.1% (wt/vol) gelatin, 200 mM each deoxynucleoside triphosphate, and 1 µmol of each primer. The reaction mixture was subjected to 30 cycles of amplification and then phenol-chloroform extracted, and the product was purified on a NENSORB 20 cartridge (NEN) as described previously (6). Three separate PCRs employing different sets of primers were used to amplify the entire MuSVts110 DNA (Fig. 1). The region from the beginning of the 5' long terminal repeat (LTR) through the *XhoI* site in the *gag* gene was amplified by using the primers SUZ5 (5'-GGAATTCCTGAAAGACCCACCCGTAGGT-3') and 5'D (5'-GTTTGGAGACCCGCTATGAGCAA-3'). SUZ5 contains an *EcoRI* cloning site at its 5' end and from nucleotides 9 to 30 is identical to the plus strand at the 5' end of the MuSV-124 LTR (37). 5'D is complementary to the MuSV-124 plus strand at nucleotides 2047 to 2070. The region between the *XhoI* site and the *AccI* site in the MuSVts110 *v-mos* gene was amplified by using the primers SUZ1 (5'-CACTCAACTGCCAATGAAGTCGA-3', identical to the MuSV-124 plus strand at nucleotides 1944 to 1967) and SUZ4 (3'-TAGATGACTTGGTGTAGAGTCACGT-5', complementary to the MuSV-124 plus strand at nucleotides 2920 to 2944). The region between the *AccI* site in the *v-mos* gene and the 3' LTR was amplified by using the primers SUZ3 (5'-TCTAAGCCTGTGTCGCTACCTCCCT-3') and SUZ6 (3'-TGGGTGCTGCCCCAGAAAGTGC GCCTAGGCGC-5'). SUZ3 is identical to the MuSV-124 plus strand at nucleotides 2492 to 2516; the first 23 nucleotides of SUZ6 are

complementary to the MuSV-124 plus strand at the 3' end of the LTR, while the last 9 nucleotides contain a *Bam*HI cloning site.

DNA sequencing. DNA sequencing by primer extension was carried out as previously described (9) except that Sequenase (United States Biochemical), used as recommended by the supplier, was substituted for DNA polymerase I. In areas of the sequence in which band compression was encountered, sequencing was carried out with *Taq*I polymerase (Perkin-Elmer Cetus) at 72°C according to directions provided by the supplier.

Cellular RNA isolation and S1 analysis. In order to measure Mu SVts110 RNA splicing at various growth temperatures, we followed a standard protocol. Cells were first grown at 33°C to about 75% confluency and then shifted overnight to either 28, 37, or 39°C. Total cellular RNA was isolated by guanidine thiocyanate extraction and centrifugation through a cesium chloride cushion as described previously (2). S1 analyses using a 5'-end-labeled probe capable of measuring ligation at the 3' splice site in the MuSVts110 *v-mos* gene were carried out as described previously (11, 27, 28).

Quantitation of viral gene expression. The relative levels of unspliced and spliced viral RNA in a given amount (usually 10 to 50 µg) of cellular RNA were quantitated by determining the radioactivity in the S1 analysis gels, using a Betagen Betascope.

Nucleotide sequence accession number. The nucleotide sequence determined in this study has been submitted to GenBank under accession number M96854.

RESULTS

PCR amplification and assembly of integrated MuSVts110 DNA. We have previously determined that MuSVts110 DNA is present at the single-copy level in the 6m2 genome and generated a partial restriction map of this cellular locus (2, 8). By taking advantage of the absence of the absence of *Bam*HI and *Eco*RI sites either within or near MuSVts110 DNA in the 6m2 genome, genomic DNA from MuSVts110-infected 6m2 cells was amplified by using three pairs of overlapping MuSV-124-specific primers (Fig. 1). As noted above, for convenience in later subcloning, the primers complementary to the 5' and 3' LTRs contained *Eco*RI and *Bam*HI sites, respectively, at their outside termini. After amplification, the overlapping 6m2 DNA amplification products were digested with either *Eco*RI-*Xho*I, *Xho*I-*Acc*I, or *Acc*I-*Bam*HI, yielding the expected 5', interior, and 3' viral DNA fragments of approximately 2 kb, 555 nucleotides, and 1.9 kb, respectively (not shown). A candidate MuSVts110 DNA was assembled in pGEM-11Z (Promega) from the subcloned PCR products as outlined in Fig. 1.

Candidate MuSVts110 gene expression in transfected cells. As a test of the authenticity of the candidate MuSVts110 DNA, we examined whether RNA transcribed from this DNA exhibited the MuSVts110-characteristic growth temperature-restricted splicing phenotype. It has been well documented that in MuSVts110-infected 6m2 cells (7, 14-16, 27, 28) and in cells transfected with the MuSVts110 equivalent (*ts32*) constructed from MuSV-124 (7, 9), MuSVts110 RNA is spliced efficiently only at low growth temperatures. In an S1 nuclease assay specific for MuSVts110 RNA splicing (27, 28), ligation at the proper 3' splice site was shown to be efficient at 33°C in NIH 3T3 cells transfected with the candidate MuSVts110 DNA (Fig. 2, lane 4) and essentially absent at 39°C (lane 5). The growth temperature

FIG. 2. Splicing phenotype of RNA transcribed from transfected candidate MuSVts110 DNA. Cellular RNA from both control 6m2 cells and NIH 3T3 cells transfected with the candidate MuSVts110 DNA (3T3/ts110 cells) grown at either 28 or 39°C was hybridized to the pBacc probe (diagrammed at the bottom). After S1 nuclease digestion, the protected fragments were analyzed on a 5% polyacrylamide gel containing 8 M urea. Lanes: 1, probe-alone control; 2, 6m2 RNA, 28°C; 3, 6m2 RNA, 39°C; 4, 3T3/ts110 RNA, 28°C; 5, 3T3/ts110 RNA, 39°C.

restriction on splicing was essentially indistinguishable from that of authentic MuSVts110 RNA in 6m2 cells (lanes 2 and 3), although MuSVts110 RNA splicing in the transfected cultures was somewhat less efficient than in infected cells. We have noted this difference previously in describing the splicing phenotype of *ts32* RNA in transfected NIH 3T3 cells (9). There is no clear rationale for this effect, although it might be explained by degree of methylation. The LTRs of transfected retroviral DNAs are substantially more methylated than those of retroviruses introduced into a cell by infection (17), a condition which may explain the lower rate of viral transcription usually observed in transfected cells. In addition, methylation patterns have also been shown to affect the splicing of both retroviral and adenoviral pre-mRNAs (5, 19, 35), and a model has been proposed in which methylation of adenosine residues in branch point consensus sequences might protect against undesired splicing (19). Since both the 5' and 3' splice sites used in splicing the MuSVts110 RNA are ordinarily cryptic, and thus presumably undesirable, splicing between these sites may be selected against by methylation in transfected cells.

Sequencing of MuSVts110 DNA. The results of the S1 analyses established that the candidate MuSVts110 DNA corresponded to native MuSVts110 with respect to its most characteristic biological property, that of growth temperature-restricted RNA splicing. Thus, the candidate MuSVts110 DNA was completely sequenced by using a series of synthetic oligonucleotide primers and compared with MuSV-124 DNA. The sequence obtained was 4,382 nucleotides in length (Fig. 3), 36 nucleotides longer than would be predicted if MuSVts110 were a straightforward 1,487-nucleotide deletion mutant of MuSV-124 (5,833 nucleotides [37]) as our earlier limited sequencing data had suggested.

MuSVts110 LTRs. The MuSVts110 LTRs were unequivocally

FIG. 3. Complete sequence of MuSVts110 DNA. The sequence of MuSVts110 DNA was determined by using a series of synthetic oligonucleotide primers and compared with the established MuSV-124 sequence. MuSV-124 nucleotides identical to the MuSVts110 sequence are indicated with an asterisk except in the short region of the MuSV-124 *v-mos* gene (marked with vertical bars) which is part of the MuSVts110 P58^{gag} reading frame, where the homologous nucleotides are indicated by vertical bars. The duplications (nucleotides 48 to 63 and 3824 to 3829) in the MuSVts110 LTRs are underlined in the MuSVts110 sequence and marked with a gap in the MuSV-124 sequence. The MuLV-like 52-nucleotide substitution (nucleotides 861 to 912) in the region upstream of the MuSVts110 *gag* gene is underlined, while the cognate MuSV-124 32-nucleotide tract is marked in bold. Nucleotide substitutions (at MuSVts110 nucleotides 806, 832, 1081, 1133, and 1788) are marked in bold at the appropriate place in MuSV-124. The position of the 1,487-nucleotide deletion of MuSV-124 sequences (at nucleotide 2426 in MuSVts110) is marked by a transition from asterisks to vertical bars. The sequence of the spliced MuSVts110 RNA (without converting T to U) and its translation product is also indicated starting just below the 5' splice site in the MuSVts110 sequence and continuing through the *v-mos* reading frame.

FIG. 3—Continued.

As noted, by direct sequencing of MuSVts110 viral RNA in 6m2 cells, we had previously determined that a 1,487-nucleotide deletion joining MuSV-124 nucleotides 2404 and

3892 had occurred in MuSVts110 (27). This deletion occurred between direct repeats of 6 nucleotides (-AACGCC-) located at nucleotides 2399 to 2404 and 3976 to 3981 in the MuSV-124 *gag* and *v-mos* genes, respectively. The deletion included one of the repeats and the sequence between the repeats. In our original assignment of the deletion junction to nucleotides 2404 and 3892, we arbitrarily decided that the -AACGCC- sequence in the *gag* gene had been retained, while the copy in the *v-mos* gene had been deleted. While the sequence obtained for integrated MuSVts110 in this study confirmed that the 1,487-nucleotide deletion occurred precisely as previously reported (27), we shifted the exact nucleotide assignment of the deletion boundaries to accommodate two considerations. First, as a combined effect of the 8-nucleotide duplication in the 5' LTR and a net gain of 20 nucleotides caused by a substitution in the region just upstream of the *gag* gene (discussed below), MuSVts110 is 28 nucleotides longer than MuSV-124 upstream of the 1,487-nucleotide deletion. Without any other adjustments, this 28-nucleotide increase in length would shift the deletion junction from nucleotides 2404 and 3892 to nucleotides 2432 and 3920. However, we have made one further adjustment, since it seems most likely that the deletion occurred as a result of reverse transcriptase template switching after copying of the -AACGCC- repeat in the *v-mos* gene during reverse transcription of the minus strand of viral DNA. In this case, the repeat in the *v-mos* gene would be retained (underlined in Fig. 3) while the copy in the *gag* gene would be lost. Thus, we have shifted the assignment of the deletion junction 6 nucleotides further upstream from *gag* gene nucleotide 2432 to nucleotide 2426.

As noted previously, the effect of the 1,487-nucleotide deletion is to join the *gag* and *v-mos* genes out of frame such that translation of MuSVts110 RNA will terminate 38 nucleotides downstream of the *gag-mos* junction at a TAG stop codon positioned right at the 3' splice site (reinforced by an in-frame TGA stop codon four codons further downstream) (Fig. 3). At the same time, the deletion activated MuSVts110 RNA for splicing by reducing the distance between the cryptic 5' and 3' splice sites in the *gag* and *v-mos* genes, respectively, in the process removing the 3' splice site used in MuLV for splicing of the *env* gene mRNA (11). The spliced MuSVts110 RNA can be translated into P85^{*gag-mos*}, a serine kinase capable of cell transformation (21, 22). The sequence of the spliced MuSVts110 RNA from the splice junction through the *v-mos* termination codon is shown in italics in Fig. 3 starting at nucleotide 2045.

MuSVts110 *v-mos* gene. As shown above, the 1,487-nucleotide deletion sustained by MuSVts110 included 11 nucleotides at the 5' end of the MuSV-124 *v-mos* reading frame. With this exception, we found no sequence differences between the MuSVts110 and MuSV-124 *v-mos* reading frames (Fig. 3).

Substitutions and additions. The most notable deviation in MuSVts110 from the MuSV-124 sequence, apart from the 1,487-nucleotide deletion, was a 52-nucleotide tract (MuSVts110 nucleotides 861 to 912) which was substituted for an unrelated 32-nucleotide region in MuSV-124 (nucleotides 853 to 884) (Fig. 3). As this substitution is located between the 5' LTR and the *gag* gene, it has no effect on the viral reading frames. Interestingly, the sequence of these 52 nucleotides in MuSVts110, with the exception of an A-to-G transition at nucleotide 864, a T-to-C transition at nucleotide 878, and an additional C at nucleotide 903, was identical to that of a 51-nucleotide tract found in murine leukemia virus (MuLV) in the same region between the 5' LTR and the *gag* gene

(33). It is assumed that these 51 nucleotides were lost in the generation of MuSV-124 from MuLV, most probably by reverse transcriptase template switching onto a cellular RNA (30), and replaced with 32 nucleotides of unknown origin (37). In MuSVts110, the MuLV-characteristic sequence was regained, and the MuSV-124 32 nucleotide tract was lost. Since the most likely way for this sequence to have found its way back into MuSVts110 is during reverse transcription, MuSVts110, rather than being a straightforward deletion mutant of MuSV-124, is very probably an MuSV-MuLV recombinant.

Two additional nucleotide substitutions (noted in bold in Fig. 3) were found at nucleotides 806 and 832 in the noncoding sequence between the 5' LTR and the *gag* gene. As discussed above, three nucleotide substitutions occurred in the *gag* gene, two of which changed the amino acids coded (all of these substitutions are noted in bold in Fig. 3).

DISCUSSION

The work presented here establishes the complete sequence for integrated MuSVts110 DNA in 6m2 cells. The PCR amplification and assembly approach was taken since MuSVts110 DNA had previously eluded cloning by a variety of other methods. An analysis of the open reading frames showed that full-length RNA transcribed from this DNA will be translated into P58^{*gag*} and that the spliced transcript can be translated into P85^{*gag-mos*}. Thermosensitive splicing of RNA transcribed from the candidate MuSVts110 DNA was confirmed. In its most important aspects, the MuSVts110 sequence obtained in this study confirmed the relevant differences established between wild-type MuSV-124 and MuSVts110 RNA by earlier studies. In addition, several other deviations from the wild-type MuSV-124 sequence not detected in earlier studies were found. These differences suggested a recombinant origin for MuSVts110.

In establishing the sequence obtained as a valid one for integrated MuSVts110 DNA, possible *TaqI* errors had to be considered. *TaqI* DNA polymerase is estimated to incorporate incorrect nucleotides about once every 9,000 bases (36). To assess whether the seven single-nucleotide substitutions observed in MuSVts110 could be *TaqI* errors, we performed reverse transcriptase sequencing of the appropriate regions of MuSVts110 RNA transcribed in 6m2 cells (not shown). The same substitutions were found and hence are unlikely to be *TaqI* misincorporations. Similarly, even though duplications very rarely occur during reverse transcription of retroviral genomes (29), the eight-nucleotide duplications in the U3 regions of both MuSVts110 LTRs are not likely to be artifactual, since the likelihood of the same artifact occurring at the same position in DNA fragments amplified separately by using different sets of primers is minimal. Thus, it is likely that a duplication occurred in U3 during reverse transcription of the minus strand of MuSVts110 DNA, which was then propagated into both MuSVts110 LTRs during plus-strand DNA synthesis. Moreover, these duplications in the LTRs provided a fortuitous signature strongly suggesting that the PCR amplification products thought to represent the 5' and 3' segments of MuSVts110 DNA were, in fact, derived from the same viral DNA, since it would be extremely unlikely that the 3' LTR of MuSVts110 and the 5' LTR of a separate, MuLV-related DNA in 6m2 cells would contain the same duplication. In fact, others have noted that in recombinant retroviral DNAs with mutations in U3, the variation is identical in both the 5' and 3' LTRs (32).

Our earlier studies established that the crucial difference

between MuSVts110 and wild-type MuSV-124 DNA is a 1,487-nucleotide frameshifting deletion fusing the MuSV ts110 *gag* and *v-mos* coding sequences out of frame. The effect of this deletion on viral gene expression is to prohibit expression of the *v-mos* gene and activate previously cryptic splice sites in the *gag* and *v-mos* genes for growth temperature-restricted splicing (27, 28). We have previously established that this deletion is the major, if not sole, determinant of the unusual phenotype exhibited by MuSVts110 (9, 10). This was most clearly shown by assessing gene expression in an MuSV-124 mutant (termed *ts32*) engineered to contain no other deviations from the wild-type sequence except a 1,487-nucleotide deletion in precisely the same location as found in MuSVts110 RNA. Cells transfected with *ts32* DNA expressed accurately spliced viral RNA at 28 to 33°C; however, no viral RNA splicing was observed at 39°C (9). Moreover, a P85^{*gag-mos*} serine kinase was produced in the *ts32*-transfected cells at low growth temperatures only (9). In these crucial particulars, RNA splicing and P85^{*gag-mos*} production was virtually identical in *ts32*-containing cells and in 6m2 cells containing authentic MuSVts110. The sequence determined for integrated MuSVts110 DNA shows the presence of the 1,487-nucleotide deletion between precisely the same nucleotides as containing authentic MuSVts110. The sequence determined for integrated MuSVts110 DNA shows the presence of the 1,487-nucleotide deletion between precisely the same nucleotides as predicted from the earlier MuSVts110 RNA sequencing studies, although, as noted above, the nucleotide position of the deletion boundary has been changed to accommodate both previously unrecognized sequence variations in MuSVts110 as well as a model for its origin. As would be predicted, RNA transcribed from transfected MuSVts110 DNA is spliced only at a low growth temperature and is translated into a P85^{*gag-mos*} kinase.

One impetus for determining the complete MuSVts110 sequence was to determine whether there might be other mutations beyond the 1,487-nucleotide deletion that could exert an effect on its splicing phenotype. Although this possibility was not tested rigorously, we consider it very unlikely that the other sequence differences observed, i.e., the duplications in the LTRs, the substitution of an MuLV-characteristic tract upstream of the *gag* gene, and the scattered single-nucleotide substitutions, would be likely to exert any effect on the MuSVts110 splicing phenotype. First, none of these sequence differences appear to be near or affect any conserved splicing signals. Second, no sequence changes were observed in the two regions demonstrated to influence the thermosensitivity of MuSVts110 RNA splicing, the MuSVts110 branch point (6, 10) and the 3' half of the *v-mos* exon (34). In addition, since we have previously shown that translation of the MuSVts110 *gag* gene is not required to maintain the MuSVts110 splicing phenotype (11), it would follow that neither of the substitutions observed in the *gag* gene would be expected to affect splicing. Finally, in an MuSVts110 mutant in which the region containing the 52-nucleotide substitution and all but one of the single-nucleotide substitutions outside the *gag* gene were removed, the splicing phenotype was essentially unaffected (11). Thus, it appears very unlikely that the recently recognized differences between MuSVts110 and MuSV-124 will prove relevant to the regulation of its splicing phenotype.

Origin of MuSVts110. The MuSVts110 retrovirus was originally obtained by Blair et al. (3) from NRK cultures infected with a UV-irradiated mixture of MuSV-349 and

helper MuLV virions selected for the ability to grow as transformed colonies in soft agar at 33 but not at 39°C. Although it has been neither extensively studied nor completely sequenced, the parental MuSV-349 virus should be essentially identical to the well-characterized and completely sequenced MuSV-124 (37), as the original MuSV-349 cell line was obtained by infection of mouse TB cells (a mixture of fetal thymus and bone marrow cells from CFW/D mice) with MuSV-124-helper MuLV virions from a cloned cell line (1). However, the helper MuLV associated with MuSV-124 was XC negative, whereas the helper MuLV shed from the MuSV-349-infected line was XC positive (1). These inconsistent data might be explained by an observation by other investigators who found that a noninfectious, XC-negative clone of MuLV DNA (pMLV-1) contained single-nucleotide mutations in the p30 and protease regions, whereas another cloned MuLV DNA (pMLV-48) which was both infectious and XC positive did not contain these mutations (24). Thus, the helper MuLV in the original MuSV (MuLV) stock might have been heterogeneous with respect to these mutations and have given rise to MuSV-MuLV-infected cell lines whose helper MuLV was either XC negative (MuSV-124) or positive (MuSV-349).

The complete sequence determined here for MuSVts110 suggests that the virus arose through recombination between MuSV-349 and helper MuLV during reverse transcription in NRK cells (Fig. 4). Whether a critical role was played by prior UV irradiation of the virions is unknown. In DNA, UV light induces pyrimidine dimers and cyclobutane adducts (20). The effect of UV on RNA is far less well known; however, exposure to UV can cause RNA-RNA and RNA-protein cross-links (4). This effect has some specificity for pyrimidines, as evidenced by the fact that exposure to UV has been shown to cross-link proteins specifically to an -AUUA- sequence found at the 3' ends of many cytokine RNAs (12). In this connection, a pyrimidine cluster is found in MuSV-124 just upstream of the *v-mos* gene AACGCC (-AUUCAACGCC-), while upstream of the *gag* gene copy of the repeat there is a 48-nucleotide AG run containing only four scattered pyrimidines (33, 37), making the *v-mos* repeat a more likely target for UV damage. In prokaryotes at least, pyrimidine dimers can inhibit DNA synthesis one nucleotide prior to the site of damage (25, 26). Thus, it is plausible that UV treatment could have caused damage which forced a stoppage in retroviral DNA synthesis just after copying of the -ACCGC- in the MuSV-124 *v-mos* sequence, setting the stage for a copy choice template switch (23, 29, 30). In this model, it is assumed that an RNA-RNA or RNA-viral protein cross-link centered around the -AUUC- sequence upstream of the MuSV-124 *v-mos* -AACGCC- sequence halted reverse transcription and allowed RNase H-mediated digestion of the template RNA, leaving the nascent minus-strand DNA free to jump to the other copy of -AACGCC- in the *gag* gene and continue reverse transcription. The net effect would be the deletion of one of the repeats and the intervening sequence.

While this model explains the 1,487-nucleotide deletion, it does not explain how MuSVts110 acquired the 52 nucleotides typical of helper MuLV upstream of the *gag* gene. Acquisition of this sequence in MuSVts110 can be explained by assuming that the NRK progenitor of the 6m2 cell line was infected with an MuSV-MuLV heterozygote or with both MuSV-349 and helper MuLV viruses and that the template switch instead occurred between the -AACGCC- in the MuSV-124 *v-mos* gene and the -AACGCC- in the helper MuLV *gag* gene. After this jump, continued reverse tran-

FIG. 4. Possible derivation of MuSVts110. (A) Diagrammatic structure of MuSV-349 (MuSV-124) and MuLV, the viruses present in the harvests from MuSV-349-infected mouse TB cells. The cryptic 5' and 3' splice sites and -AACGCC- repeats are marked. The 51-nucleotide tract in MuLV is indicated with a diamond, and the -GTAACGCC- sequence in U3 is marked with a single bar. (B) Generation of MuSVts110 by recombination between MuSV-349 (MuSV-124) and MuLV during reverse transcription. The model assumes that the cell was infected with either two viral particles, one of which was MuSV-349 (MuSV-124), the other of which was MuLV, or with heterozygotes carrying one MuSV-349 (MuSV-124) and one MuLV genome. Reverse transcription of minus-strand MuSVts110 DNA was initiated on the MuSV-349 (MuSV-124) RNA and transferred to the 3' end of another copy of the same RNA, accounting for the fact that MuSVts110 lacks the *Xba*I site found in the MuLV LTR and contains an enhancer region in U3 identical to that in MuSV-124 and distinct from that in MuLV. Head-to-tail duplication of the sequence -GTAACGCC- very probably occurred during this step (indicated by a switch from a single bar to a double bar). Continuing reverse transcription of minus-strand MuSVts110 DNA is postulated to have encountered a break or UV-induced damage at the sequence -AACGCC- in the MuSV-349 (MuSV-124) *v-mos* gene, at which time the nascent DNA jumped to the copy of -AACGCC- in the *gag* gene of helper MuLV RNA present in the same cell. The completion of the minus-strand DNA would allow the incorporation of the 51-nucleotide tract in MuLV into MuSVts110. Initiation and synthesis of the plus strand of MuSVts110 DNA from nearly complete minus strand would allow the transfer of the duplication in U3 to the 5' LTR. The completed MuSVts110 DNA would then contain a duplication in the U3 region of each LTR, a nearly perfect copy of the MuLV 51-nucleotide tract, and a deletion between the -AACGCC- repeats in the *gag* and *v-mos* genes.

scription would yield an appropriately deleted viral DNA also containing the MuLV 52-nucleotide tract.

The origin of the duplications in each MuSVts110 LTR is unknown. Since the duplicated sequence is not flanked by direct repeats in either the MuSV-124 or MuLV U3 region, it seems unlikely that a slippage mechanism can be proposed. Others have noted that duplications occurring during reverse transcription are very rare (29); however, the presence of the duplication in both LTRs is evidence that the duplication occurred during reverse transcription of the minus strand of MuSVts110 DNA. We think it interesting but of unknown significance that these duplications (-GTAACGCC- to -GTAACGCCGTAACGCC-) each contain a copy of the

-AACGCC- repeat involved in the 1,487-nucleotide central deletion.

ACKNOWLEDGMENTS

The contributions of the first two authors were equivalent, and the order of appearance is arbitrary.

We thank Sandra A. Kinney and James J. Syrewicz for excellent technical assistance and Brenda Cykala for manuscript preparation and editing. We are indebted to numerous colleagues, in particular Bill L. Brizzard, Elizabeth A. Putnam, Paul E. Cizdziel, Michael A. Nash, and Jade L. Wong for their efforts and suggestions in earlier, less successful stages of this project. We are grateful to Donald G. Blair for the provision of materials and many helpful suggestions.

This work was supported in part by grants from the National Cancer Institute (CA-34734), the National Institute of Environmental Health Services (ES-05323), the Robert A. Welch Foundation (G-854), and Reproductive Hazards in the Workplace, Home, Community and Environment research grant 15-146 from the March of Dimes Birth Defects Foundation. Synthetic oligonucleotides were provided by the UTMDACC Macromolecular Synthesis Facility, which is supported by core grant CA-16672 from the National Institutes of Health. L.H. was supported in part by funds provided by the Chinese government, S.M.C. was supported by a predoctoral fellowship from the Robert A. Welch Foundation, M.A.G. was supported by the UTMDACC College Student Summer Program, J.R.A. was supported by the UTHSC Graduate School of Biomedical Sciences, and L.A.B. was supported by the University of Texas Medical School (Houston) M.D./Ph.D. program and Achievement Rewards for College Scientists.

REFERENCES

- Ball, J. K., J. A. McCarter, and S. M. Sunderland. 1973. Evidence for helper-independent murine sarcoma virus. *Virology* 56:268-284.
- Biggart, N. W., G. Gallick, and E. C. Murphy, Jr. 1987. Nickel-induced heritable alterations in retroviral transforming gene expression. *J. Virol.* 61:2378-2388.
- Blair, D. G., M. A. Hull, and E. A. Finch. 1979. The isolation and preliminary characterization of temperature-sensitive transforming mutants of Moloney murine sarcoma virus. *Virology* 95:303-316.
- Brizzard, B. L., E. A. Putnam, and E. C. Murphy, Jr. Unpublished data.
- Budowsky, E. I., and G. G. Abdurashkova. 1989. Polynucleotide crosslinks induced by ultraviolet light and their use for the structural investigation of nucleoproteins. *Prog. Nucleic Acid Res. Mol. Biol.* 37:1-65.
- Chen-Kiang, S., J. R. Nevins, and J. E. Darnell. 1979. N⁶-methyladenosine in adenovirus type 2 nuclear RNA is conserved in the formation of messenger RNA. *J. Mol. Biol.* 135:733-752.
- Chiocca, S. M., D. A. Sterner, N. W. Biggart, and E. C. Murphy, Jr. 1991. Nickel mutagenesis: alteration of the MuSVts110 thermosensitive splicing phenotype by a nickel-induced duplication of the 3' splice site. *Mol. Carcinog.* 4:61-71.
- Cizdziel, P. E., M. de Mars, and E. C. Murphy, Jr. 1988. Exploitation of a thermosensitive splicing event to study pre-mRNA splicing in vivo. *Mol. Cell. Biol.* 8:1558-1569.
- Cizdziel, P. E., M. A. Nash, D. G. Blair, and E. C. Murphy, Jr. 1986. Molecular basis underlying phenotypic revertants of Moloney murine sarcoma virus MuSVts110. *J. Virol.* 57:310-317.
- de Mars, M., P. E. Cizdziel, and E. C. Murphy, Jr. 1988. Activation of thermosensitive RNA splicing and production of a heat-labile P85 *gag-mos* kinase by the introduction of a specific deletion in murine sarcoma virus-124 DNA. *J. Virol.* 62:1907-1916.
- de Mars, M., P. E. Cizdziel, and E. C. Murphy, Jr. 1990. Activation of cryptic splice sites in MuSV-124 mutants. *J. Virol.* 64:5260-5269.
- de Mars, M., D. A. Sterner, S. M. Chiocca, N. W. Biggart, and E. C. Murphy, Jr. 1990. Regulation of RNA splicing in *gag*-deficient mutants of Moloney murine sarcoma virus MuSV ts110. *J. Virol.* 64:1421-1428.
- Gillis, P., and J. S. Malter. 1991. The adenosine-uridine binding factor recognizes the AU-rich elements of cytokine, lymphokine, and oncogene mRNAs. *J. Biol. Chem.* 266:3172-3177.
- Graham, R., and A. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* 52:456-467.
- Hamelin, R. B., B. L. Brizzard, M. A. Nash, and E. C. Murphy, Jr. 1985. Temperature-sensitive viral RNA expression in Moloney murine sarcoma virus ts110-infected cells. *J. Virol.* 53:616-623.
- Hamelin, R. B., K. Kabat, D. G. Blair, and R. B. Arlinghaus. 1986. Temperature-sensitive splicing defect of ts110 Moloney murine sarcoma virus is virus encoded. *J. Virol.* 57:301-309.
- Horn, J. P., T. Wood, E. C. Murphy, Jr., D. G. Blair, and R. B. Arlinghaus. 1981. A selective temperature-sensitive defect in viral RNA expression in cells infected with a ts transformation mutant of murine sarcoma virus. *Cell* 25:37-46.
- Huang, L. H.-S., and E. Gilboa. 1984. Expression of genes introduced into cells by retroviral infection is more efficient than that of genes introduced into cells by DNA transfection. *J. Virol.* 50:417-424.
- Junghans, R. P., E. C. Murphy, Jr., and R. B. Arlinghaus. 1982. Electron microscopic analysis of ts110 Moloney mouse sarcoma virus, a variant of wild-type virus with two RNAs containing large deletions. *J. Mol. Biol.* 161:229-255.
- Kane, S. E., and K. Beemon. 1985. Precise location of m⁶A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Mol. Cell. Biol.* 5:2298-2306.
- Kittler, L., and G. Löber. 1977. Photochemistry of the nucleic acids. *Photochem. Photobiol. Rev.* 2:39-65.
- Kloetzer, W., S. A. Maxwell, and R. B. Arlinghaus. 1983. P85^{gag-mos} encoded by ts110 Moloney murine sarcoma virus has an associated protein kinase activity. *Proc. Natl. Acad. Sci. USA* 80:412-416.
- Kloetzer, W. S., S. A. Maxwell, and R. B. Arlinghaus. 1984. Further characterization of the P85^{gag-mos}-associated protein kinase activity. *Virology* 138:143-155.
- Luo, G., and J. Taylor. 1990. Template switching by reverse transcriptase during DNA synthesis. *J. Virol.* 64:4321-4328.
- Miller, A. D., and I. Verma. 1984. Two-base changes restore infectivity to a noninfectious molecular clone of Moloney murine leukemia virus (pMLV-1). *J. Virol.* 49:214-222.
- Moore, P., and B. S. Strauss. 1979. Sites of inhibition of *in vitro* DNA synthesis in carcinogen- and UV-treated DNA. *Nature (London)* 278:664-666.
- Moore, P. D., K. K. Bose, S. D. Rabkin, and B. S. Strauss. 1981. Sites of termination of *in vitro* DNA synthesis on ultraviolet- and N-acetylaminofluorene-treated ϕ X-174 templates by prokaryotic and eukaryotic DNA polymerases. *Proc. Natl. Acad. Sci. USA* 78:110-114.
- Nash, M., B. Brizzard, J. L. Wong, and E. C. Murphy, Jr. 1985. Murine sarcoma virus ts110 RNA transcripts: origin from a single proviral DNA and the sequence of the *gag/mos* junctions in both precursor and spliced viral RNAs. *J. Virol.* 53:624-633.
- Nash, M., N. V. Brown, J. L. Wong, R. B. Arlinghaus, and E. C. Murphy, Jr. 1984. S1 nuclease analysis mapping of viral RNAs from a temperature-sensitive transformation mutant of murine sarcoma virus. *J. Virol.* 50:478-488.
- Pathak, V. K., and H. M. Temin. 1990. Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hot-spots in a retroviral shuttle vector after a single replication cycle: deletions with deletions with insertions. *Proc. Natl. Acad. Sci. USA* 87:6024-6028.
- Pulsinelli, G. A., and H. Temin. 1991. Characterization of large deletions occurring during a single round of retrovirus vector replication: novel deletion mechanism involving errors in strand transfer. *J. Virol.* 65:4786-4797.
- Saiki, R. K., S. Scharf, F. Faloona, K. B. et al. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Shimotohno, K., and H. Temin. 1982. Spontaneous variation and synthesis in the U3 region of the long terminal repeat of an avian retrovirus. *J. Virol.* 41:163-171.
- Shinnick, T., R. Lerner, and J. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature (London)* 293:543-548.
- Sterner, D. A., and E. C. Murphy, Jr. Submitted for publication.
- Stoltzfus, C. M., and R. W. Dane. 1982. Accumulation of spliced avian retrovirus RNA is inhibited in S-adenosylmethionine-depleted chicken embryo fibroblasts. *J. Virol.* 42:918-931.
- Tindall, K. R., and T. A. Kunkel. 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* 27:6008-6018.
- Van Beveren, C., F. van Straaten, J. Galleghaw, and I. Verma. 1981. Nucleotide sequence of the genome of a murine sarcoma virus. *Cell* 27:97-108.