JOURNAL OF VIROLOGY, Feb. 1983, p. 607–617
0022-538X/83/020607-11$02.00/0

Vol. 45, No. 2

Recombinational Junctions of Variants of Moloney Murine Sarcoma Virus: Generation and Divergence of a Mammalian Transforming Gene

DANIEL J. DONOGHUE[1,2]* AND TONY HUNTER[2]

Department of Chemistry, University of California, San Diego, La Jolla, California 92093,[1] and Tumor Virology Laboratory, The Salk Institute, San Diego, California 92138[2]

Received 11 August 1982/Accepted 25 October 1982

Different variants of Moloney murine sarcoma virus (MSV) were examined by nucleotide sequencing to compare the junctions between the acquired cellular sequence, v-*mos*, and the adjacent virus-derived sequences. These variants included 124-MSV, m1-MSV, and HT1-MSV and also the purportedly independent isolate Gazdar MSV. These four strains have an identical 5′ junction between the murine leukemia virus *env* gene and the v-*mos* gene. This junction lies within the sixth codon of the chimeric *env-mos* coding region that encodes the transforming gene product. In contrast, at the 3′ junction between the v-*mos* gene and the murine leukemia virus *env* gene, the three variants examined here were all different. A small deletion was found in the COOH-terminal portion of the m1-MSV *env-mos* coding region, indicating that the COOH terminus of this transforming gene product must be different from that of 124-MSV or HT1-MSV. The data presented here are consistent with the thesis that a virus closely related to HT1-MSV was the primordial Moloney MSV, and that all other related strains evolved from it by deletion or rearrangement. The variability observed in the Moloney MSV family is discussed in terms of possible mechanisms for the initial capture of *mos* sequences by the parental retrovirus and also in comparison with other transforming retrovirus families, such as Abelson murine leukemia virus and Rous sarcoma virus.

Moloney murine sarcoma virus (Mo-MSV) is a replication-defective transforming retrovirus which arose by recombination between Moloney murine leukemia virus (Mo-MLV) and the proto-oncogene c-*mos*, present in the BALB/c mouse genome. Mo-MSV was initially isolated from a rhabdomyosarcoma of an Mo-MLV-infected mouse (28). From the initial Mo-MLV/Mo-MSV tumor, many variants of Mo-MSV were isolated (1–5, 8, 15, 16, 19, 24, 39) (Fig. 1).

Among these, the *mos* gene of 124-MSV, termed v-*mos*^{124}, has been extensively characterized. The nucleotide sequence of v-*mos*^{124} is known (11, 36, 37, 40, 41) and reveals the presence of an open reading frame encoding a gene product of M_r 41,000. Both parental nucleotide sequences also are known, and comparison of the c-*mos* (40, 41) and Mo-MLV (38) sequences with the v-*mos*^{124} sequence allowed identification of the apparent recombination points which generated v-*mos*^{124}. Indeed, the v-*mos*^{124} coding region constitutes part of a chimeric coding region, consisting of the NH_2-terminal coding region of the envelope (*env*) gene of Mo-MLV fused to v-*mos*^{124}. Therefore, although the coding region is commonly referred to simply as v-*mos*, it is more correctly termed an *env-mos* fusion gene.

The coding potential of strain 124-MSV was first examined by in vitro translation of virion RNA in the rabbit reticulocyte lysate system (9, 32). By the nomenclature of Papkoff et al. (32), four overlapping *mos*-specific proteins of observed M_r's of 37,000, 33,000, 24,000, and 18,000 were identified. These proteins share a common COOH terminus but differ at their NH_2 termini. The alignment of these in vitro translation products with the nucleotide sequence of the v-*mos*^{124} coding region was confirmed by various methods (33), such as hybrid arrest translation, comparison of the NH_2-terminal tryptic peptides of the M_r 37,000 and 24,000 products with synthetic peptides, hydroxylamine cleavage of the in vitro translation products at a unique Asn-Gly bond, and finally, immunoprecipitation with an antiserum against a synthetic peptide corresponding to the predicted v-*mos*^{124} COOH terminus. The entire *env-mos* coding region was also fused to the coding region of the small t antigen of polyoma virus and expressed in a strain of *Escherichia coli* as a polyoma small t:*mos* fusion protein of the antici-

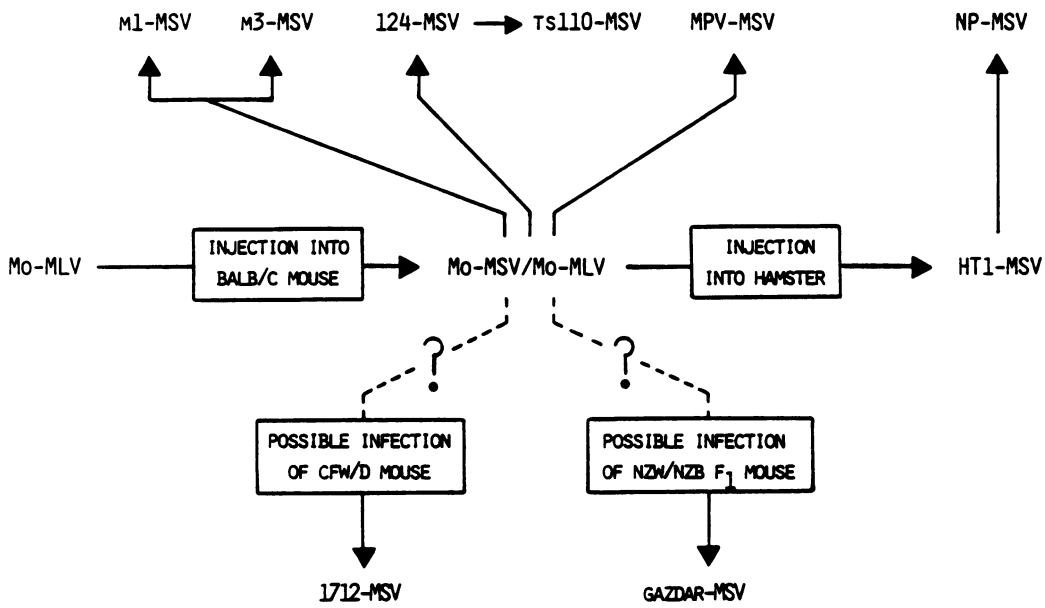

FIG. 1. Genealogy of the Mo-MSV family. The m1-MSV and m3-MSV variants were isolated by soft agar cloning of mouse 3T3FL cells infected with Mo-MSV/Mo-MLV (4, 15, 16). The variant 124-MSV was isolated several years later from mouse TB cells infected with Mo-MSV/Mo-MLV (3). By electron microscopy (14), the virus:*mos* junctions of m3-MSV are indistinguishable from 124-MSV and m1-MSV, and only the latter two were included in the present study. The temperature-sensitive mutant *ts*110-MSV was derived from a subclone of 124-MSV, designated 349-MSV, by mutagenesis and selection (5). Cells infected with *ts*110-MSV have been shown to contain two defective RNA genomes in addition to the helper Mo-MLV (23). As it is not known which genome is responsible for the transformed phenotype, *ts*110-MSV was not included in this study. The myeloproliferative virus MPV-MSV was also isolated from Mo-MSV/Mo-MLV, but unlike 124-MSV it can induce erythroleukemia and myeloid leukemia in adult mice (8, 35). MPV-MSV has not yet been molecularly cloned and was not examined here. The HT-1 line of hamster cells was isolated as a nonproducer transformed cell line after injection of Mo-MSV/Mo-MLV into hamsters (24). NP-MSV was derived by rescue of a transforming activity from HT-1 cells by cocultivation with cells producing various strains of MLV (1, 39). By electron microscopy (14), the virus:*mos* junctions of HT1-MSV and NP-MSV are indistinguishable, and only HT1-MSV was examined in this study. Gz-MSV was reported as an independent isolate from an NZW/NZB F_1 mouse (19, 30). Analysis of its 5'-*env*:*mos* junction suggests, however, that it may have arisen from Mo-MSV after infection of an NZW/NZB F_1 mouse (13). Strain 1712-MSV was reported as an independent isolate from a CFW/D mouse injected with MLV (2, 10), but is indistinguishable from m3-MSV by electron microscopy (14). Therefore, 1712-MSV probably arose from Mo-MSV present as a contaminant in the MLV stock which was used to inject the CFW/D mouse and was not included in this study. It should be noted that after Moloney (28) isolated the initial Mo-MLV/Mo-MSV tumor, cell-free extracts were passaged six additional times in BALB/c mice, at which time the uncloned virus was used to inoculate hamsters. A single resulting tumor was put into tissue culture and is referred to as the HT1 line (24). Other variants of Mo-MSV were isolated subsequently from uncloned stocks which were passaged at least an additional 150 times in BALB/c mice. This passage history may have some bearing upon the finding (see the text) that HT1-MSV most closely resembles a primordial Mo-MSV.

pated size, which could be immunoprecipitated with either a polyoma anti-tumor serum or the aforementioned anti-COOH-terminal *mos* peptide serum (12). These various approaches confirmed that the in vitro translation product of observed M_r 37,000 represents the full-length product of the *env-mos* coding region, as predicted by the nucleotide sequences of the cloned plasmids pDD0 (11) and pMSV-1 (40, 41). The calculated M_r of 41,000 is in fairly good agreement with the observed M_r of 37,000. A protein very similar to the largest in vitro translation product, designated $p37^{mos}$, is also present in cells transformed by 124-MSV (34). This suggests that the longest product of the *env-mos* coding region is likely to be the bona fide transforming gene product of 124-MSV.

In a previous study by electron microscopy (14), the genomic RNAs of different variants of Mo-MSV were compared to elucidate the arrangement of the v-*mos* sequences with regard to the flanking Mo-MLV-derived sequences. This earlier study compared the genomic organization of some of the Mo-MSV variants, as summarized in Fig. 2. Briefly, although all strains contained indistinguishable v-*mos* se-

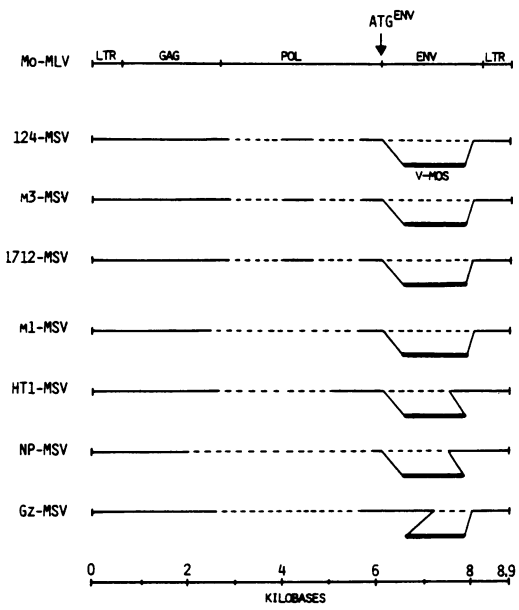

FIG. 2. Electron microscopic comparison of the Mo-MSV family. The DNA genomes are shown for a number of strains of the Mo-MSV family and compared with the parental virus, Mo-MLV. Deletions with respect to Mo-MLV are shown by dashed lines. The v-*mos* substitution is indicated in each genome by the bold line underneath the *env* gene deletion. This figure is based upon earlier electron microscopic measurements of heteroduplexes formed between the Mo-MSV genomic RNAs and either Mo-MLV cDNA or 124-MSV cDNA (14). The m1-MSV and HT1-MSV genomes also have been examined in the electron microscope by Vande Woude et al. (43) with similar results.

quences, differences were noted in the amount of the *env* gene retained both upstream and downstream of v-*mos*. At the 5′ junction between the Mo-MLV *env* gene and v-*mos* (referred to as the 5′-*env*:*mos* junction), all of the genomes appear very similar except for Gazdar MSV (Gz-MSV), which contains about 1.2 kilobases (kb) more of the *env* gene upstream of v-mos^{Gz}. At the 3′ junction between v-*mos* and the Mo-MLV *env* gene (referred to as the 3′-*mos*:*env* junction), all of the genomes appear very similar except for NP-MSV and HT1-MSV, which contain an additional 0.6 kb of the *env* gene downstream of v-*mos*.

Given the detailed knowledge of the v-mos^{124} nucleotide sequence and the unique structure of its chimeric *env-mos* gene product, it was of interest to reexamine the apparent recombination points of these related variants with the precision of nucleotide sequencing. We wished to understand first, the recombinational events which led to diversity at the junctions between the *env* gene and v-*mos* previously observed by electron microscopy, and second, whether the predicted gene products of other strains would be similar to the v-mos^{124} gene product.

MATERIALS AND METHODS

Plasmids. Plasmids pM13, pHT10, and pHT13 were obtained from the laboratory of George Vande Woude (National Institutes of Health). These plasmids represent subclones of the cloned m1-MSV genome (Fig. 3A) or of the cloned HT1-MSV genome (Fig. 3B) (6, 42). Plasmid pMTx-1, which contains the c-*mos* gene on a single large *Eco*RI fragment, was described previously (40). Plasmid pDD81 (Fig. 3A) was constructed in this work and is a subclone of the 3′-*mos*:*env* junction of m1-MSV.

Nucleotide sequencing. The fragments of m1-MSV and HT1-MSV prepared for nucleotide sequencing are shown diagrammatically in Fig. 3A and B. Most of the c-*mos* sequence which is shown in Fig. 7 was previously published (40, 41) except for that portion which overlaps the 3′-*mos*:*env* junction of HT1-MSV. The additional sequence was provided by C. Van Beveren of the Salk Institute (personal communication). We also reconfirmed this sequence, which is contained in the *Hin*dIII-*Bgl*II fragment downstream of the c-*mos* gene. All fragments, whether 5′ or 3′ labeled, were subjected to the chemical sequencing reactions described by Maxam and Gilbert (27). Sequencing products were fractionated on either 6, 8, or 24% polyacrylamide sequencing gels by standard protocols.

RESULTS

NH_2-terminal coding region of *mos*. Figure 4 shows the NH_2-terminal nucleotide sequences of the *mos* gene of strains 124-MSV, m1-MSV, and HT1-MSV in comparison with their parental sequences. These three variants of Mo-MSV all had an identical 5′-*env*:*mos* junction, which lay within the sixth codon of the chimeric *env-mos* reading frame.

Although the junctions of the three variants were identical, their sequences differed in other respects. (i) The third codon of the presumptive parental Mo-MLV *env* gene was CGT (arginine). In strain 124-MSV, a single base change (G to A) altered the third codon to CAT (histidine). (ii) Strain 124-MSV had 192 nucleotides deleted in the *pol* gene immediately upstream of the *env-mos* coding region. In both respects, strains m1-MSV and HT1-MSV were more closely related to the parental sequence than to 124-MSV. Whether these alterations in 124-MSV have any biological function is not known.

Gz-MSV was isolated from a spontaneously occurring tumor in an NZW/NZB F_1 mouse and thus was considered to be distinct from the Mo-MSV family (19, 30). When examined by electron microscopy (Fig. 2), Gz-MSV showed the same v-*mos* substitution as was observed in Mo-MSV and had about 1.2 kb of the *env* gene directly upstream of v-mos^{Gz}. Given that other

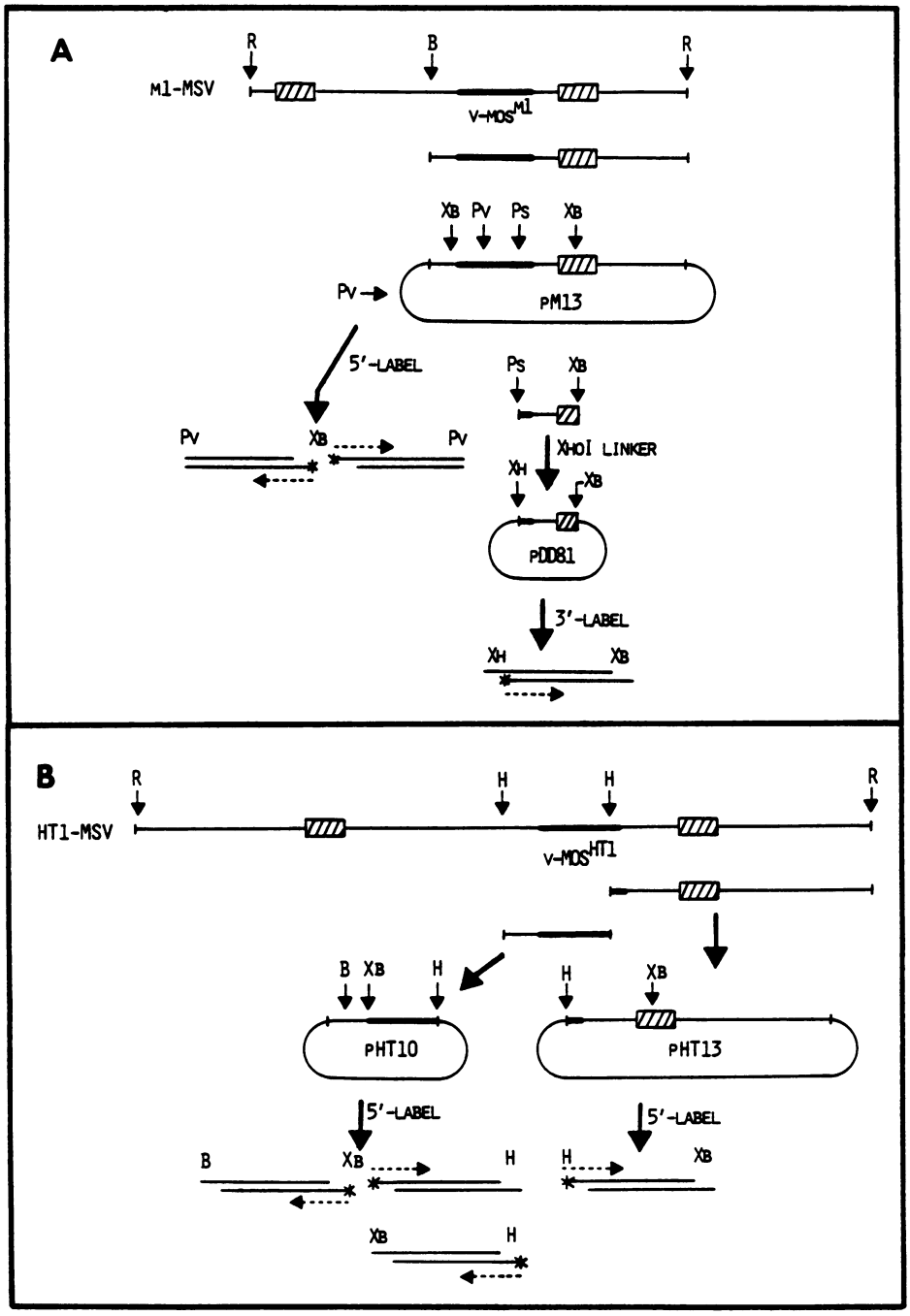

FIG. 3. Nucleotide sequencing strategy for the junctions of m1-MSV and HT1-MSV. The maps of m1-MSV and HT1-MSV are shown in panels A and B, respectively (6). The bold line represents v-*mos*, and the hatched boxes represent the long terminal repeat sequences. In each of the fragments used for sequencing, shown as double-stranded molecules, the ^{32}P-labeled terminus is designated with an asterisk. The dashed arrow represents that portion of each fragment which was sequenced. Abbreviations: B, *Bgl*II; H, *Hin*dIII; Ps, *Pst*I; Pv, *Pvu*II; R, *Eco*RI; Xb, *Xba*I; Xh, *Xho*I.

variants of Mo-MSV encode a fused *env-mos* gene product, it would be expected that strain Gz-MSV would encode an *env-mos* gene product which would be correspondingly longer due to the additional *env* gene sequences at its NH_2 terminus. We recently reported the nucleotide sequence of Gz-MSV across the 5'-*env*:*mos* junction, as determined by primer extension nucleotide sequencing of the virion RNA of Gz-MSV (13). These results are shown (Fig. 5) in comparison with the parental sequences. Surprisingly, v-mos^{Gz} had the same 5'-*env*:*mos* junction present in other variants of Mo-MSV, occurring within the sixth codon of the chimeric *env-mos* coding region. However, the entire *env-mos* coding region was juxtaposed 1.1 kb downstream in the *env* gene, presumably by a second recombinational event with helper MLV. As previously discussed (13), this genomic structure can be explained if Gz-MSV resulted from infection of an NZW/NZB F_1 mouse with Mo-MSV, which subsequently underwent rearrangement by recombination with the helper MLV.

It is interesting to note that in the third codon of the fused *env-mos* coding region, the Gz-MSV sequence was CAT (histidine) and thus was more closely related to 124-MSV than to the parental *env* gene sequence, which was CGT (arginine).

COOH-terminal coding region of *mos*. Figure 6 shows the COOH-terminal coding region and 3'-*mos*:*env* junction of strains 124-MSV and m1-MSV in comparison with the parental nucleotide sequences. The open reading frame in 124-MSV, which encodes the *env-mos* gene product, $p37^{mos}$, has been determined by a variety of techniques to terminate as shown in Fig. 6 (12, 33). Termination of $p37^{mos}$ occurs at a TGA codon in the c-*mos*-derived portion, which lies about 50 nucleotides upstream of the 3'-*env*:*mos* junction. In this region, the v-mos^{124} sequence differed from the parental c-*mos* sequence in several minor respects. (i) There was a single nucleotide change (G to A), which altered the third amino acid before the COOH terminus from alanine in c-*mos* to threonine in 124-MSV. (ii) There were three other single nucleotide substitutions and a single nucleotide deletion in the 3' untranslated sequence.

The COOH-terminal coding region of the m1-MSV *mos* gene (Fig. 6) was significantly different from that of 124-MSV; a deletion of 13 nucleotides has occurred, starting within the sixth codon from the COOH terminus of c-*mos*. This deletion caused a shift in the reading frame of v-mos^{m1}, so that the terminator codon of the parental c-*mos* gene was translated out of phase. Termination of v-mos^{m1} occurred about 20 nucleotides further downstream at a TAG codon. The occurrence of a single nucleotide deletion immediately before the v-mos^{m1} terminator codon should also be noted.

The 3'-*mos*:*env* junction was also different in m1-MSV than in 124-MSV. The m1-MSV genome retained about 25 nucleotides less of the presumptive c-*mos* 3' untranslated region. This observation is consistent with electron microscopic evidence (43) and also explains the absence (6, 43) in m1-MSV of an *Hin*dIII restriction site that is present in the c-*mos* gene and also in strains 124-MSV and HT1-MSV. It should also be noted that the m1-MSV genome contained about 20 nucleotides more of the *env* gene (positions 7,597 to 7,617 of Mo-MLV) than did 124-MSV. Thus, m1-MSV differed significantly from 124-MSV both in the COOH-terminal coding region and also at the 3'-*mos*:*env* junction.

The COOH-terminal coding region of the HT1-MSV *mos* gene (Fig. 7) was identical with that of its parental c-*mos* gene. Strain HT1-MSV was also unique in that it retained more of the presumptive 3' untranslated region of c-*mos* than did 124-MSV. Whereas 124-MSV retained only about 50 nucleotides, HT1-MSV retained about 120 nucleotides from the presumptive 3' untranslated region of c-*mos*. Based on the earlier electron microscopic study (14) (Fig. 2), it was also known that HT1-MSV retains about 600 nucleotides more of the *env* gene downstream of v-mos^{HT1} than does 124-MSV. This was confirmed by the nucleotide sequence shown in Fig. 7. The 3' junction between v-mos^{HT1} and the *env* gene occurred at position 6,975 in Mo-MLV, compared with the 3'-junction between v-mos^{124} and the *env* gene, which occurred at position 7,618 in Mo-MLV.

The presumptive 3' untranslated region of *mos* that was retained in strain HT1-MSV differed from the parental c-*mos* sequence in several respects. (i) There was one single nucleotide deletion and one single nucleotide substitution. (ii) There was a two-nucleotide -TA- insertion, which occurred after a tract of seven T residues in the parental c-*mos* sequence. Curiously, the position of this small insertion coincided with the *mos* endpoint of the 3'-*mos*:*env* junction of strain 124-MSV.

In the sequence of the *env* gene retained downstream of v-mos^{HT1}, 11 nucleotide changes were found in comparison with the Mo-MLV *env* gene (Fig. 7). In fact, the *env* gene sequence retained in strain HT1-MSV was more closely related to the *env* gene of Akv virus (26). This observation emphasizes the difficulty in correctly identifying the parental retroviruses for either Mo-MSV or Mo-MLV, as unknown recombinational events could have occurred with endogenous murine retroviruses during the passage

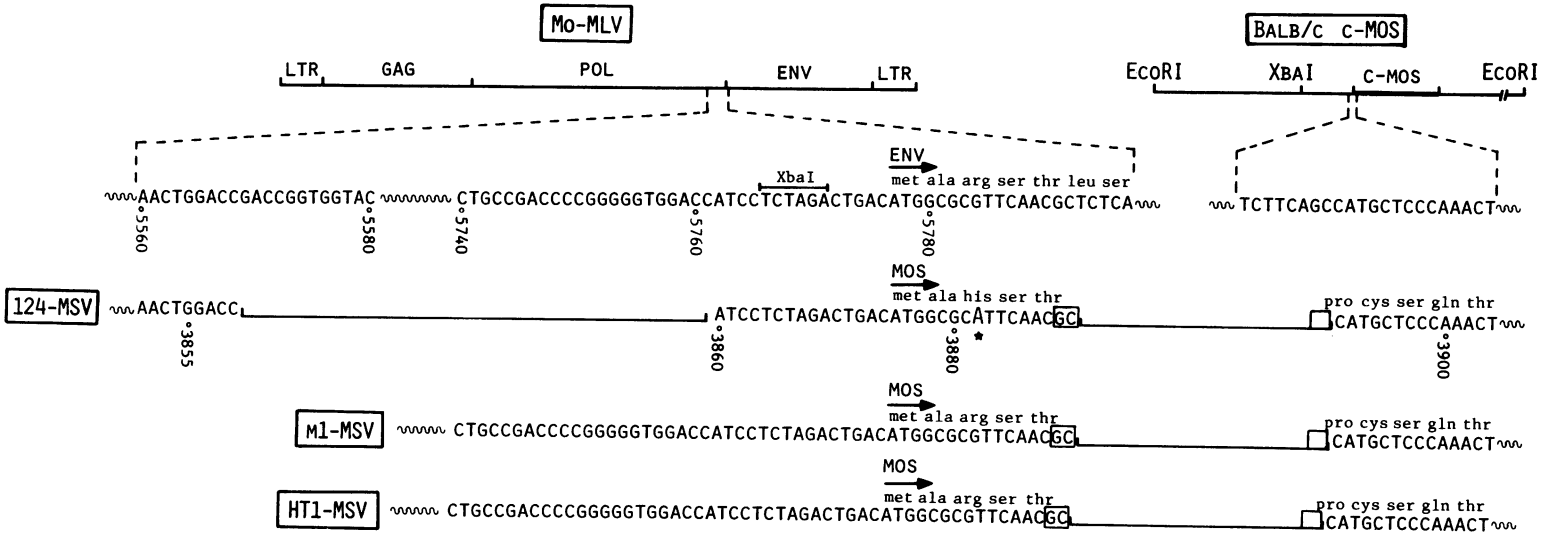

FIG. 4. Sequences of the 5′-*env*:*mos* junctions of 124-MSV, ml-MSV, and HT1-MSV in comparison with the parental sequences. Note that all the variants of Mo-MSV shown here had an identical 5′-*env*:*mos* junction. The position of a single base change in 124-MSV is indicated by an asterisk. At the 5′-*env*:*mos* junction there are only two nucleotides of homology, -GC- (small boxes). The Mo-MLV sequence is from the data of Shinnick et al. (38). The BALB/c c-*mos* sequence is from Van Beveren et al. (40, 41). The 124-MSV sequence is from Donoghue (11) and Van Beveren et al. (40) and is shown with the numbering system for the complete 124-MSV genome (41).

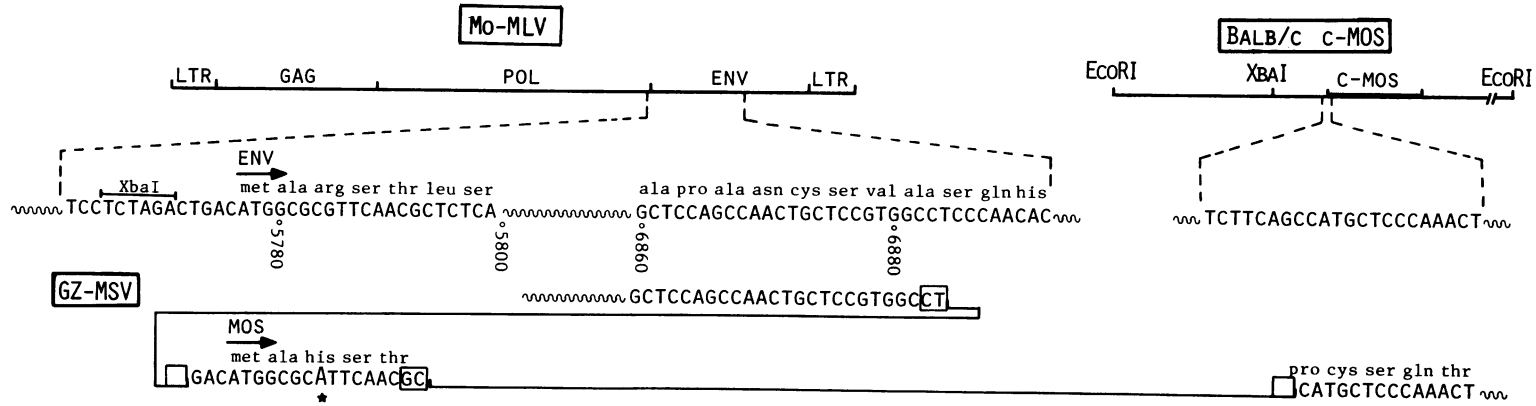

FIG. 5. Sequence of the 5′-*env*:*mos* junction of Gz-MSV in comparison with the parental sequences. Note that two recombinational events have occurred to generate Gz-MSV. The 5′-*env*:*mos* junction is identical with that of other variants of Mo-MSV (Fig. 4). A second *env*:*env* junction also is observed, arising from the juxtaposition of the *env-mos* coding region 1.1 kb downstream in the *env* gene. The position of a single base change in Gz-MSV is indicated by an asterisk. At the 5′-*env*:*mos* junction there are only two nucleotides of homology, -GC- (boxes), and at the *env*:*env* junction there are also two nucleotides of homology, -CT-. The Mo-MLV sequence is from Shinnick et al. (38). The BALB/c c-*mos* sequence is from Van Beveren et al. (40, 41). The Gz-MSV sequence is based on the primer extension nucleotide sequence data of Donoghue and Hunter (13).

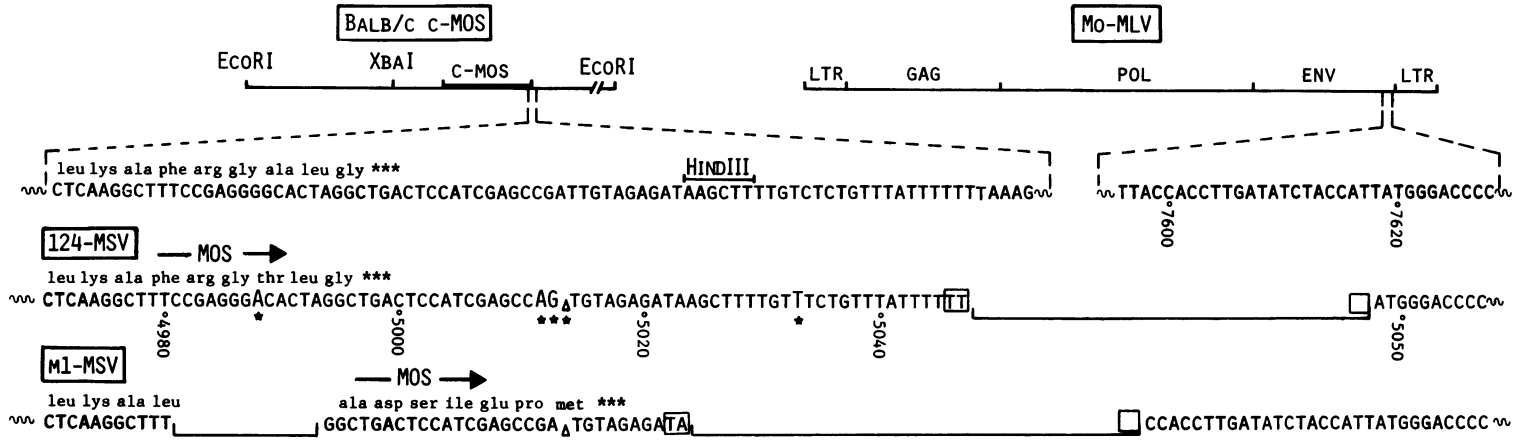

FIG. 6. Sequences of the 3′-*mos*:*env* junctions of 124-MSV and m1-MSV in comparison with the parental sequences. Note the deletion of 13 nucleotides in the COOH-terminal coding region of m1-MSV. Also, note that 124-MSV and m1-MSV have very different 3′-*mos*:*env* junctions. At the 3′-*mos*:*env* junction of 124-MSV, there are two nucleotides of homology, -TT- (small boxes). At the 3′-*mos*:*env* junction of m1-MSV, there are two nucleotides of homology, -TA-. Single nucleotide deletions or alterations with respect to the parental sequences are indicated by asterisks beneath the sequence. The BALB/c c-*mos* nucleotide sequence and presumptive amino acid sequence are from Van Beveren et al. (40, 41). The Mo-MLV sequence is from Shinnick et al. (38).

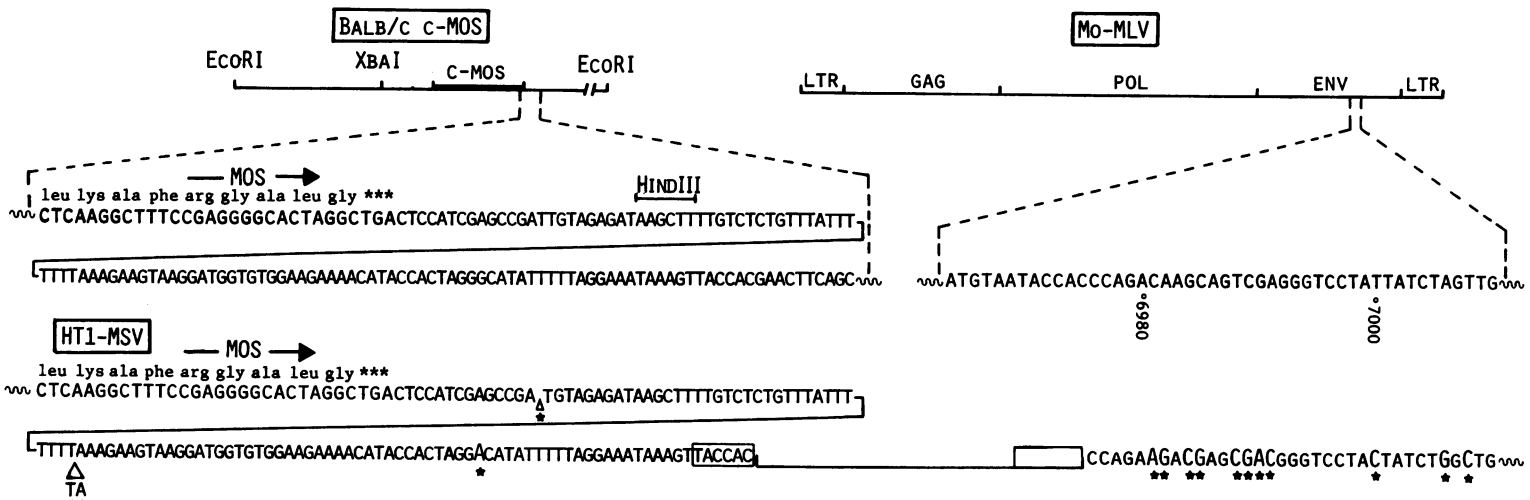

FIG. 7. Sequence of the 3′-*mos*:*env* junction of HT1-MSV in comparison with the parental sequences. Note that this junction retains more information from both parental sequences than did the variants shown in Fig. 6. As discussed in the text, this junction may represent the recombination point of the primordial strain of Mo-MSV. At this junction, there are six nucleotides of homology, -TACCAC- (small boxes). Single nucleotide deletions or alterations with respect to the parental sequences are indicated by asterisks beneath the sequence. The BALB/c c-*mos* nucleotide sequence and presumptive amino acid sequence are partially from Van Beveren et al. (40, 41) and also from this work. The Mo-MLV sequence is from Shinnick et al. (38).

history of Mo-MSV/Mo-MLV. As discussed in the legend to Fig. 1, the HT1-MSV genome was immortalized in the hamster tumor cell line HT-1 early after the original Mo-MSV/Mo-MLV tumor was isolated. The molecularly cloned HT1-MSV was derived by rescue with feline leukemia virus (42) and immediately used to transform mink cells. Any recombination event between HT1-MSV and a virus related to Akv virus must have occurred during the brief Mo-MSV passage in BALB/c mice or, possibly, before its progenitor became tumorigenic.

DISCUSSION

Junction points: recombination or deletion? Every transforming retrovirus that bears a cell-derived sequence has two junctions between viral information and cell-derived information. In Mo-MSV, these are referred to as the 5'-*env*:*mos* and the 3'-*mos*:*env* junctions. Unfortunately, knowledge of the nucleotide sequence at the junctions is not sufficient to define the recombination events that resulted in the original capture of the cellular sequence by the retrovirus. The junctions which were observed may be identical to the original recombination points. Alternatively, if deletions occurred subsequently that spanned the original recombination points, then the observed junctions will bear no obvious relationship to the recombination points. We will attempt to discern between recombinational junctions and deletion junctions in the following discussion.

In all but one of the junctions presented in this work there was some homology between the parental sequences, ranging from two to as many as six nucleotides of homology. In these instances, the precise nucleotide of junction obviously could not be identified. Small boxes have been drawn around the nucleotides at the junctions that are present in both parental sequences in Fig. 4 through 7.

Conservation of the 5'-*env*:*mos* junctions. The variant strains 124-MSV, m1-MSV, HT1-MSV, and Gz-MSV all had an identical 5'-*env*:*mos* junction. This invariability suggests that this junction may represent the original recombination point that gave rise to the transforming virus.

The initiation codon for the *env-mos* gene product was thus the same as that used to initiate translation of $Pr80^{env}$, the Mo-MLV *env* gene product. We have no way of knowing whether the *env*-derived NH_2 terminus constitutes part of a functional domain of the *mos* gene product or whether its only role is to initiate translation. One argument, however, suggests the latter interpretation. Strain *ts*110-MSV was derived from a subclone of strain 124-MSV and encodes a *gag-mos* fusion protein of M_r 85,000, designated P85, which contains peptides derived from *gag* and from *mos* (5, 23). Presumably, the *ts*110-MSV genome arose by a deletion which fused the retained *gag* coding region with the *mos* coding region, so that P85 possesses a *gag*-derived NH_2 terminus. Unfortunately, this interpretation is complicated by the presence of two different defective RNA genomes in *ts*110-MSV-infected cells (23). At present it is not known which defective RNA species is responsible for the temperature-sensitive phenotype nor whether P85 is the sole *mos*-specific protein present in infected cells.

Variability of the 3'-*mos*:*env* junction. The three isolates 124-MSV, m1-MSV, and HT1-MSV all had different 3'-*mos*:*env* junctions. Both the *mos* and the *env* endpoints differed in all three strains. HT1-MSV retained the greatest amount of c-*mos* information and also the greatest amount of *env* information. Thus, starting with HT1-MSV, it would be possible to derive 124-MSV or m1-MSV by simple deletions. It would not be possible, however, to derive the 124-MSV 3'-*mos*:*env* junction by deletion or rearrangement of m1-MSV, nor vice versa. This suggests that the 3'-*mos*:*env* junction of HT1-MSV may represent the original recombination point that gave rise to the transforming virus and, therefore, that HT1-MSV may be closely related to the primordial Mo-MSV. This suggestion is further supported by the absence of single nucleotide changes between v-mos^{HT1} and c-*mos* in those portions of their coding sequences presented here. In this regard, it is interesting to note that the stock of Mo-MSV used to create the hamster tumor from which HT1-MSV derives was passaged many fewer times than were the stocks used to isolate the other variants of Mo-MSV (see the legend to Fig. 1).

Even though the virus:*mos* junctions of HT1-MSV may be ancestral to the other variants, this was not true of the rest of the HT1-MSV genome. For instance, HT1-MSV was more deleted in the *pol* gene than 124-MSV, m3-MSV, and 1712-MSV (Fig. 2). This is not surprising, as there is probably no correlation between the retention of *gag* and *pol* sequences and the transforming potential of a given variant. Except for the initial acquisition of *mos* sequences by Mo-MSV, the other deletions and alterations observed in the different variants are probably temporally unrelated.

COOH-terminal coding region. The data presented here suggest that the COOH termini of the 124-MSV and HT1-MSV gene products should be very similar. In fact, the *mos* gene products of both variants can be immunoprecipitated with an antiserum against a synthetic peptide, 12 residues in length, corresponding to the predicted v-mos^{124} COOH terminus (34). As the

COOH-terminal coding region of v-*mos*^{HT1} was identical to the predicted c-*mos* sequence, this also suggests that the antiserum should recognize the c-*mos* gene product. Preliminary data indicate that in cell lines transformed by the murine c-*mos* gene (7) activated by a viral long terminal repeat sequence, it is in fact possible to immunoprecipitate a *mos*-specific gene product with the anti-COOH-terminal serum (J. Papkoff and T. Hunter, unpublished data).

We did not determine the sequence of the COOH-terminal coding region of Gz-MSV in this work. It is possible, however, to immunoprecipitate a protein of observed M_r 37,000 from the in vitro translation products of Gz-MSV virion RNA with the anti-COOH-terminal v-*mos*^{124} serum (31). This indicates that the COOH terminus of the v-*mos*^{Gz} gene product is closely related to that of v-*mos*^{124} and v-*mos*^{HT1}.

In contrast, the data presented here reveal that the transforming gene product of m1-MSV should have an altered COOH terminus due to a 13-nucleotide deletion which shifts the reading frame. The net result of this change would be the replacement of the six COOH-terminal residues, as seen in v-*mos*^{124}, with a new COOH terminus eight residues long. Consistent with this prediction, it was impossible to specifically immunoprecipitate a v-*mos*^{m1} gene product from m1-MSV-transformed cells with the aforementioned antiserum, which is directed against the COOH terminus of the v-*mos*^{124} gene product (J. Papkoff and T. Hunter, unpublished data).

An alternative sequence has been presented by Reddy et al. (36, 37) for the COOH-terminal coding region of v-*mos*^{124}. In the most recent version (37) of their sequence determination, one additional nucleotide shifts the reading frame such that it would terminate downstream of the 3′-*mos*:*env* junction. The net result of this change would be the replacement of the 17 COOH-terminal residues of the v-*mos*^{124} gene product, as predicted by the nucleotide sequences of the cloned plasmids pDD0 (11) and pMSV-1 (40, 41), with an altered COOH terminus 52 residues long. The example of m1-MSV presented above suggests that the *mos* transforming gene product can tolerate some variability in its COOH terminus. It should be noted that the COOH-terminal reading frame of v-*mos*^{m1} was different from that proposed by Reddy et al. (36, 37) for v-*mos*^{124}. Conceivably the particular clone of 124-MSV sequenced by Reddy and coworkers may have acquired a single nucleotide insertion and, in fact, may differ from the clones of 124-MSV sequenced by other laboratories. Perhaps the COOH terminus of the *mos* transforming gene product can be encoded by any of the three possible reading frames, which would suggest that it can tolerate great variability without loss of biological transforming activity. A similar interpretation also has been offered by Watson et al. (44).

Capture of *mos* by Mo-MSV. As discussed above, HT1-MSV is probably most closely related to the primordial Mo-MSV from which all isolates of Mo-MSV are derived. At the apparent 5′-*env*:*mos* recombination point, there were only two nucleotides of homology, -GC-, between the parental sequences. There was somewhat greater homology at the apparent 3′-*mos*:*env* recombination point of HT1-MSV, consisting of the six nucleotides -TACCAC-. Assuming that these junctions do represent the true recombination points, it would seem that the recombinational events do not require extensive nucleotide homology between the parental sequences.

It would be interesting to know whether the capture of *mos* proceeded by means of a DNA or an mRNA intermediate. It has been postulated that 3′ recombination points containing an oligodeoxyadenylate tract may arise by recombinational events involving mRNA intermediates (22, 25, 45; I. Lemischka and P. A. Sharp, personal communication). At the apparent 3′-*mos*:*env* recombination point in HT1-MSV, however, there was no such tract.

At present, expression of the c-*mos* gene has not been detected (17, 18), and so no information is available on the structure of the c-*mos* mRNA or protein. Thus, the possibility of introns in c-*mos* cannot be excluded; nor is it possible to identify the authentic initiation site or termination site for the c-*mos* protein. In considering the capture of *mos* by Mo-MSV, it is therefore impossible to know whether the recombination points lie within the c-*mos* coding region in introns, or possibly in flanking untranslated sequences. For these reasons, it may be difficult to decide between various models for *mos* capture until data are obtained on the expression of c-*mos*.

Other families of transforming retroviruses. For several other transforming oncogenes besides *mos*, multiple variants or isolates have been described. Although most of these familial relationships have not been analyzed at the molecular level, some information is available for the Abelson MLV and Rous sarcoma virus (RSV) families.

Abelson MLV arose by recombination between Mo-MLV and the cellular proto-oncogene c-*abl*. The primordial strain is probably P160 (20), which encodes a *gag-abl* fusion protein of M_r 160,000. Another commonly studied strain, P120, is related to the P160 strain by a single internal deletion (20). Several other variants of Abelson MLV have been derived from either strain P160 or P120, including the strains desig-

nated P106, P105, P100, P95, and P90 (21, 46). Characterization of these variants suggests that portions of the P160 *gag-abl* proteins are dispensable for fibroblast transformation but not for efficient lymphoid cell transformation (20, 46). The virus:*abl* junctions have been sequenced for one strain of Abelson MLV (20). These junction sequences indicate that the v-*abl* endpoints are contained within the mRNA transcripts of c-*abl*. Unfortunately, this information is not sufficient to distinguish between recombinational models involving DNA versus mRNA intermediates. Curiously, the viral endpoint of the 3′-*abl*:virus junction is three nucleotides from the viral endpoint of the 3′-*mos*:*env* junction of strain 124-MSV. Given that 124-MSV probably arose by deletion of a virus similar to HT1-MSV (see above), the proximity of the 3′-*mos*:virus and 3′-*abl*:virus junctions is almost surely coincidental.

In RSV, the *src*:virus junctions of two different isolates have been compared by nucleotide sequencing (R. Swanstrom, R. Parker, and J. M. Bishop, personal communication), the Prague C and Schmidt-Ruppin A strains of RSV. Comparison of these sequences with the c-*src* sequence indicates that at the 5′-virus:*src* junctions of both the Prague C and Schmidt-Ruppin A strains, the *src* endpoint lies within an intron of c-*src*, upstream of the ATG which initiates $pp60^{src}$. This strongly suggests that some of the recombinational events which generated the RSV family occurred by means of DNA intermediates rather than mRNA intermediates, which would be devoid of introns.

ACKNOWLEDGMENTS

We thank George Vande Woude for providing us with the HT1-MSV and m1-MSV plasmid subclones used in this work. We also thank Jon Cooper, Jackie Papkoff, and Bart Sefton for valuable criticism of the manuscript, Richard Parker for communicating the results for virus:*src* junctions before publication, and Chuck Van Beveren for communicating unpublished c-*mos* sequence data.

D.J.D. gratefully acknowledges a fellowship from the Helen Hay Whitney Foundation. This work was supported by Public Health Service grants CA-14195, CA-17096, and CA-28458 from the National Cancer Institute.

LITERATURE CITED

1. **Aaronson, S. A., R. H. Bassin, and C. Weaver.** 1972. Comparison of murine sarcoma viruses in nonproducer and S^+L^- cells. J. Virol. **9**:701–704.
2. **Ball, J. K., D. Harvey, and J. A. McCarter.** 1973. Evidence for naturally occurring murine sarcoma virus. Nature (London) **241**:272–274.
3. **Ball, J. K., J. A. McCarter, and S. M. Sunderland.** 1973. Evidence for helper independent murine sarcoma virus. I. Segregation of replication-defective and transformation-defective viruses. Virology **56**:268–284.
4. **Bassin, R. J., L. A. Phillips, M. J. Kramer, D. K. Haapala, P. T. Peebles, S. Nomura, and P. J. Fischinger.** 1971. Transformation of mouse 3T3 cells by murine sarcoma virus: release of virus-like particles in the absence of replicating murine leukemia helper virus. Proc. Natl. Acad. Sci. U.S.A. **68**:1520–1524.
5. **Blair, D. G., M. A. Hull, and E. A. Finch.** 1979. The isolation and preliminary characterization of temperature-sensitive transformation mutants of Moloney sarcoma virus. Virology **95**:303–316.
6. **Blair, D. G., W. L. McClements, M. K. Oskarsson, P. J. Fischinger, and G. F. Vande Woude.** 1980. Biological activity of cloned Moloney sarcoma virus DNA: terminally redundant sequences may enhance transformation efficiency. Proc. Natl. Acad. Sci. U.S.A. **77**:3504–3508.
7. **Blair, D. G., M. K. Oskarsson, T. G. Wood, W. L. McClements, P. J. Fischinger, and G. F. Vande Woude.** 1981. Activation of the transforming potential of a normal cell sequence: a molecular model for oncogenesis. Science **212**:941–943.
8. **Chirigos, M. A., D. Scott, W. Turner, and K. Perk.** 1968. Biological, pathological and physical characterization of a possible variant of a murine sarcoma virus (Moloney). Int. J. Cancer **3**:223–237.
9. **Cremer, K., E. P. Reddy, and S. A. Aaronson.** 1981. Translational products of Moloney murine sarcoma virus RNA: identification of proteins encoded by the murine sarcoma virus *src* gene. J. Virol. **38**:704–711.
10. **Deng, C., and E. Wimmer.** 1978. Two different murine sarcoma virus isolates have homologous genome sequences: implications for their origin. Virology **89**:309–313.
11. **Donoghue, D. J.** 1982. Demonstration of biological activity and nucleotide sequence of an in vitro synthesized clone of the Moloney murine sarcoma virus *mos* gene. J. Virol. **42**:538–546.
12. **Donoghue, D. J., and T. Hunter.** 1982. Expression of the transforming region of Moloney murine sarcoma virus in *E. coli* as a fusion protein with the small tumor antigen of polyoma virus. Proc. Natl. Acad. Sci. U.S.A. **79**:800–804.
13. **Donoghue, D. J., and T. Hunter.** 1982. A generalized method of subcloning DNA fragments by restriction site reconstruction: application to sequencing the amino-terminal coding region of the transforming gene of Gazdar murine sarcoma virus. Nucleic Acids Res. **10**:2549–2563.
14. **Donoghue, D. J., P. A. Sharp, and R. A. Weinberg.** 1979. Comparative study of different isolates of murine sarcoma virus. J. Virol. **32**:1015–1027.
15. **Fischinger, P. J., S. Nomura, P. T. Peebles, D. K. Haapala, and R. H. Bassin.** 1972. Reversion of murine sarcoma virus-transformed mouse cells: variants without a rescuable sarcoma virus. Science **176**:1033–1035.
16. **Fischinger, P. J., S. Nomura, N. Tuttle-Fuller, and K. J. Dunn.** 1974. Revertants of mouse cells transformed by murine sarcoma virus. III. Metastable expression of virus functions in revertants retransformed by murine sarcoma virus. Virology **59**:217–227.
17. **Frankel, A. E., and P. J. Fischinger.** 1976. Nucleotide sequences in mouse DNA and RNA specific for Moloney sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **73**:3705–3709.
18. **Gattoni, S., P. Kirschmeier, I. B. Weinstein, J. Escobedo, and D. Dina.** 1982. Cellular Moloney murine sarcoma (c-*mos*) sequences are hypermethylated and transcriptionally silent in normal and transformed rodent cells. Mol. Cell Biol. **2**:42–51.
19. **Gazdar, A. F., H. C. Chopra, and P. S. Sarma.** 1972. Properties of a murine sarcoma virus isolated from a tumor arising in an NZW/NZB F_1 hybrid mouse. I. Isolation and pathology of tumor induced in rodents. Int. J. Cancer **9**:219–233.
20. **Goff, S. P., and D. Baltimore.** 1982. The cellular oncogene of the Abelson murine leukemia virus genome. Adv. Viral Oncol. **1**:127–139.
21. **Goff, S. P., O. N. Witte, E. Gilboa, N. Rosenberg, and D. Baltimore.** 1981. Genome structure of Abelson murine leukemia virus variants: proviruses in fibroblasts and lymphoid cells. J. Virol. **38**:460–468.
22. **Hollis, G. F., P. A. Hieter, O. W. McBride, D. Swan, and P. Leder.** 1982. Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type

processing. Nature (London) 296:321–322.
23. Horn, J. P., T. G. Wood, E. C. Murphy, 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.
24. Huebner, R. J., J. W. Hartley, W. P. Rowe, W. T. Lane, and W. I. Capps. 1966. Rescue of the defective genome of Moloney sarcoma virus from a noninfectious hamster tumor and the production of pseudotype sarcoma viruses with various murine leukemia viruses. Proc. Natl. Acad. Sci. U.S.A. 56:1164–1169.
25. Jagadeeswaran, P., B. G. Forget, and S. M. Weissman. 1981. Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA PolIII transcripts? Cell 26:141–142.
26. Lenz, J., R. Crowther, A. Straceski, and W. Haseltine. 1982. Nucleotide sequence of the Akv *env* gene. J. Virol. 42:519–529.
27. Maxam, A., and W. Gilbert. 1980. Sequencing with base-specific chemical cleavages. Methods Enzymol. 65:499–560.
28. Moloney, J. B. 1966. A virus induced rhabdomyosarcoma of mice. Natl. Cancer Inst. Monogr. 22:139–142.
29. Oskarsson, M., W. L. McClements, D. G. Blair, J. V. Maizel, and G. F. Vande Woude. 1980. Properties of a normal mouse cell DNA sequence (*sarc*) homologous to the *src* sequence of Moloney sarcoma virus. Science 207:1222–1224.
30. Pang, R. H. L., L. A. Phillips, and D. K. Haapala. 1977. Characterization of Gazdar murine sarcoma virus by nucleic acid hybridization and analysis of viral expression in cells. J. Virol. 24:551–556.
31. Papkoff, J., and T. Hunter. 1982. Identification of proteins encoded by the Gazdar murine sarcoma virus genome by in vitro translation and comparison with Moloney murine sarcoma virus 124. J. Virol. 43:533–543.
32. Papkoff, J., T. Hunter, and K. Beemon. 1980. In vitro translation of virion RNA from Moloney murine sarcoma virus. Virology 101:91–103.
33. Papkoff, J., M. H.-T. Lai, T. Hunter, and I. M. Verma. 1981. Analysis of transforming gene products from Moloney murine sarcoma virus. Cell 27:109–119.
34. Papkoff, J., I. M. Verma, and T. Hunter. 1982. Detection of a transforming gene product in cells transformed by Moloney murine sarcoma virus. Cell 29:417–426.
35. Pragnell, I. B., A. Fusco, C. Arbuthnott, F. Smadja-Joffe, B. Klein, C. Jasmin, and W. Ostertag. 1981. Analysis of the myeloproliferative sarcoma virus genome: limited changes in the prototype lead to altered target cell specificity. J. Virol. 38:952–957.
36. Reddy, E. P., M. J. Smith, and S. A. Aaronson. 1981. Complete nucleotide sequence and organization of the Moloney murine sarcoma virus genome. Science 214:445–450.
37. Reddy, E. P., M. J. Smith, E. Canaani, K. C. Robbins, S. R. Tronick, S. Zain, and S. A. Aaronson. 1980. Nucleotide sequence analysis of the transforming region and large terminal redundancies of Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. 72:4650–4654.
38. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) 293:543–548.
39. Todaro, J. G., and S. A. Aaronson. 1969. Properties of clonal lines of murine sarcoma virus transformed BALB/3T3 cells. Virology 38:174–202.
40. Van Beveren, C., J. A. Galleshaw, V. Jonas, A. J. M. Berns, R. F. Doolittle, D. J. Donoghue, and I. M. Verma. 1981. Nucleotide sequence and formation of the transforming gene of a mouse sarcoma virus. Nature (London) 289:258–262.
41. Van Beveren, C., F. van Straaten, J. A. Galleshaw, and I. M. Verma. 1981. Nucleotide sequence of the genome of a murine sarcoma virus. Cell 27:97–108.
42. Vande Woude, G. F., M. Oskarsson, L. W. Enquist, S. Nomura, M. Sullivan, and P. J. Fischinger. 1979. Cloning of integrated Moloney sarcoma proviral DNA sequences in bacteriophage lambda. Proc. Natl. Acad. Sci. U.S.A. 76:4464–4468.
43. Vande Woude, G. F., M. Oskarsson, W. L. McClements, L. Enquist, D. G. Blair, P. J. Fischinger, J. V. Maizel, and M. Sullivan. 1979. Characterization of integrated Moloney sarcoma proviruses and flanking host sequences cloned in bacteriophage lambda. Cold Spring Harbor Symp. Quant. Biol. 44:735–745.
44. Watson, R., M. Oskarsson, and G. F. Vande Woude. 1982. Human DNA sequence homologous to the transforming gene (*mos*) of Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. 79:4078–4082.
45. Wilde, C. D., C. E. Crowther, T. P. Cripe, M. G. Lee, and N. J. Cowan. 1982. Evidence that a human beta-tubulin pseudogene is derived from its corresponding mRNA. Nature (London) 297:83–84.
46. Ziegler, S. F., C. A. Whitlock, S. P. Goff, A. Gifford, and O. N. Witte. 1981. Lethal effect of the Abelson murine leukemia virus transforming gene product. Cell 27:477–486.