
Nucleotide sequence of the 5' noncoding region and part of the *gag* gene of mouse mammary tumor virus; identification of the 5' splicing site for subgenomic mRNAs

N.Fasel, E.Buetti, J.Firzlaff, K.Pearson and H.Diggelmann

Swiss Institute of Experimental Cancer Research, 1066 Epalinges, Switzerland

Received 5 September 1983; Revised and Accepted 5 October 1983

ABSTRACT

We have determined the sequence of the first 1371 nucleotides at the 5' end of the genome of mouse mammary tumor virus using molecularly cloned proviral DNA of the GR virus strain. The most likely initiation codon used for the *gag* gene of mouse mammary tumor virus is the first one, located 312 nucleotides from the 5' end of the viral RNA. The 5' splicing site for the subgenomic mRNA's is located \sim 288 nucleotides downstream from the 5' end of the viral RNA. From the DNA sequence the amino acid sequence of the N-terminal half of the *gag* precursor protein, including p10 and p21, was deduced (353 amino acids).

INTRODUCTION

Several interesting regions of the genome of mouse mammary tumor virus (MMTV) have been sequenced recently. These include the long terminal repeats (LTR) of exogenous and endogenous proviruses of two different strains, C3H and GR. (1-4). The LTR contains the signals for initiation of viral RNA synthesis and for polyadenylation. The 5' terminus of MMTV RNA has been mapped (5-8) and the 3' end has been sequenced (9). The LTR also contains the DNA sequences necessary for the regulation of viral RNA synthesis by glucocorticoids (3, 5, 10, 11, 11a, 11b). The sequences have been mapped more precisely between \sim 100 and \sim 200 nucleotides upstream from the initiation site of RNA synthesis (8). One or more binding sites for the glucocorticoid receptor complex have been localized in the LTR (12-17). Furthermore the *env* region including the 3' splicing site for the *env* mRNA has been sequenced (18, 18a). Most recently the 5' (19) and 3' (19,20) splicing sites for the 1.7 kb mRNA potentially coding for the LTR gene product have been identified. Information on the 5' splicing site of *env* mRNA and on the *gag* region was so far not available. Many laboratories have encountered considerable difficulties in molecular cloning of this region of the MMTV genome. Recombinant plasmids or bacteriophages containing DNA

fragments from this area seem to be generally underrepresented in gene libraries and are difficult to propagate in different host cells (7, 18a, 21).

In this paper we report the nucleotide sequence of the 1.1 kb Pst I fragment covering the 5' untranslated region and about half of the gag gene of the exogenous GR-MMTV. Using the S1 nuclease mapping technique the 5' splicing site for the subgenomic mRNAs has been identified. From the sequence data presented here and the analysis of the gag precursor protein labeled with myristic acid (22) it seems most likely that the first AUG in position 313-315 is used for the synthesis of the gag precursor polypeptide.

MATERIALS AND METHODS

Generation of MMTV subclones for sequencing

From the pBR 322 recombinant clones containing either the 1.1 kb Pst I fragment or the 3.8 kb Bam HI - Xho I fragment of MMTV (3, 8), the MMTV sequences were cut out by the appropriate restriction enzyme and isolated by electroelution from normal agarose or by phenol extraction from low gelling temperature agarose gels. The isolated fragments were digested with the restriction enzymes (Biolabs Inc.) Rsa I or Hae III, treated with E. coli DNA polymerase (Klenow enzyme A, Boehringer) and cloned into the Hinc II restriction site of M13mp7. One of the clones was obtained by digestion of a recombinant M13mp8 phage with DNAase I according to the procedure of Hong (23).

DNA sequencing

The sequence of the M13 recombinant phages was determined by the chain termination method of Sanger (24). To obtain the sequence in the region where no recombinant phages covering both DNA strands were available, the chemical degradation method of Maxam and Gilbert (25) was used.

Preparation of RNA

Total cellular RNA was extracted from tissue culture cells by the hot-phenol procedure (26). Poly(A)-containing RNA was isolated by oligo(dT)-cellulose column chromatography.

S1 nuclease mapping

The 2.7 kb Pvu II fragment of plasmid pBX (8) was purified by electrophoresis on a low-melting agarose gel. 1.3 μ g of this DNA were 3'end labeled in the second nucleotide by incubation with 10 μ Ci of

(α - ^{32}P) dATP (specific activity ~ 400 Ci/mmol; Radiochemical Center, Amersham), 0.1 mM dCTP and Klenow DNA polymerase for 5 hrs at 16°C (24). The DNA probe was coprecipitated with the RNA, and the S1 analysis was performed according to Favaloro *et al.* (27), with some modifications (8). Gel electrophoresis on neutral and alkaline agarose gels was done as described by Favaloro *et al.* (27); 6% polyacrylamide-7M urea sequencing gels were performed as described by Sanger *et al.* (24). Size marker DNA fragments were labeled at their 5' ends with (γ - ^{32}P) ATP and T4 polynucleotide kinase (Boehringer). The dried gels were exposed at -70°C with Kodak X-Omat AR films and intensifying screens.

RESULTS

Sequencing strategy

The MMTV DNA subclones used in this study were derived from a lambda recombinant phage containing an insert of 9 kb (lambda MMTV-GR clone H, 28) comprising the complete genetic information of MMTV and one LTR in a permuted arrangement. The two relevant Pst I fragments of 1.45 kb containing the LTR and of 1.1 kb containing the adjacent gag region were subcloned in the plasmid pBR 322. The nucleotide sequence of the 1.45 kb Pst I fragment has been reported previously (3). At its 3' end it contains 268 nucleotides of untranslated leader sequence (7, 8). Subclones of the 1.1 kb Pst I fragment were generated in the bacteriophage M13 for sequencing. The sequencing strategy is shown in Figure 1. All sequences indicated by an arrow were obtained by the method of Sanger (24) using different M13 vectors. The two sequences marked with a dot were obtained using the technique of Maxam and Gilbert (25). As shown in Figure 1 the whole fragment has been sequenced on both strands and on overlapping clones.

The DNA sequence of the 1.1. kb fragment

Figure 2 shows the nucleotide sequence of the 3' end of the 1.45 kb Pst I fragment and of the total 1.1 kb Pst I fragment, starting at the unique site of initiation of viral RNA synthesis described previously (8). The terminally repeated sequence (R) beginning at the cap site (position 1) (GCAACAGTCCT.....) was previously found at the 3' end of viral RNA (9). The viral LTR ends within the Pvu II recognition site at position 133. The sequence complementary to the 3' end of tRNA^{Lys} starts in position 136. This sequence allows the binding of this tRNA to be used as a primer in the reverse transcriptase reaction. In several

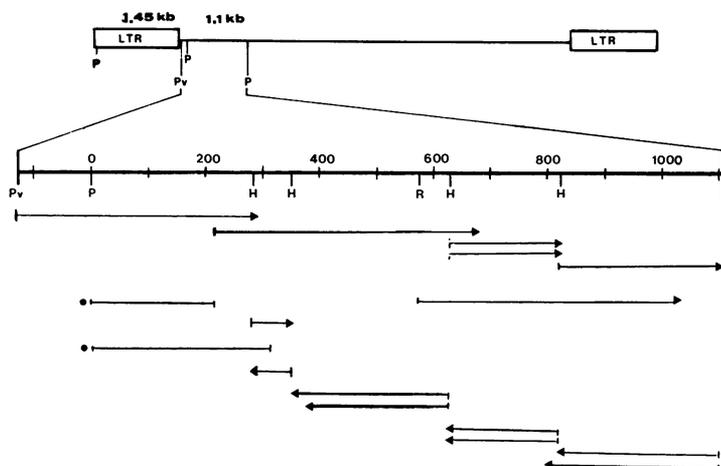


Figure 1. Partial restriction map and sequencing strategy of the MMTV DNA. The upper part of the figure shows the position of the sequenced DNA within the MMTV provirus. Indicated as well are the restriction sites used in generating the cloned DNA fragments for sequencing and the number of nucleotides sequenced downstream of the Pst I restriction site next to the left LTR. The lower part shows the sequenced DNA fragments. Generally the sequence was determined by the dideoxy chain termination method (24) except the ones denoted by a dot, where the sequence was obtained with the method of Maxam and Gilbert (25).

LTR = Long terminal repeat

H = Hae III, P = Pst I, Pv = Pvu II, R = Rsa I

retroviruses pyrimidine-rich repetitive sequences have been implicated in the dimer linkage structure found between two viral RNA molecules close to the 5' end (29, 30). Such a structure is found between nucleotides 435-442 and 445-452 from the 5' end of the viral RNA.

The analysis of the sequence for its coding capacity revealed a single open reading frame starting with the first possible initiation codon (ATG) 313 nucleotides from the initiation site of viral transcription and running until the end of the fragment (353 amino acids) therefore covering approximately half of the gag gene. The second possible initiation codon in this frame is located in position 502 of the sequence. Initiation codons present in the two other reading frames are followed by frequent termination signals and are therefore probably not functional.

Determination of the 5' splicing site

It has been shown previously that the subgenomic mRNA coding for

```

ALA THR VAL LEU THR PHE PHE SER ARG VAL PHE VAL SER VAL ARG HIS PRO VAL SER ALA
GCAACAGTCTTAACATTCCTCTCTCGTGTGTTTGTCTCTCGCCATCCCGTCTCCGCT
+1
      10      20      30      40      50      60
ARG HIS LEU SER PHE THR PHE GLN ARG VAL PRO PRO GLN ILE PRO VAL THR LEU ARG SER
CGTCACATYATCCTTCACTTTCACAGGGTCCCCCGCAGATCCCGTACCCCTCAGGTCC
      70      80      90     100     110     120
ALA ASP CYS GLY SER TRP ARG PRO ASN ARG ASP PRO ARG ILE SER ASP PRO CYS LEU TYR
GCCGACTCGCGCCAGCTCGCGCCCGAACAGGGACCCCTCGGATAAGTGACCCCTGTCTTTAT
      130     PvII     140     150     160     170     180
PHE TYR TYR PHE VAL PHE VAL LEU PHE CYS LEU TYR LEU VAL TRP LEU SER SER GLN GLU
TTCCTACTATTTTGTGTTCTCTTCTTTTGTCTCTATCTTCTCTGGCTATCATCACAAAG
      190     200     210     220     230     240
ARG ASN GLY LEU THR THR GLY ASN CYS SER LEU ALA TYR GLY GLU GLU VAL GLY TYR GLY
CGCAACGGACTCACCACAGCGGAAGCTGCAGTCTCGCCTACGGAGAGAGAGCTAGGTTACGGT
      250     260     PatI 270     280     290     300
VAL PRO LEU ASP NET GLY VAL SER GLY SER LYS GLY GLN LYS LEU PHE VAL SER VAL LEU
GTGCCATTGGATATCGGGCTCTCGCGCTCAAAGGGCAGAACTCTTTGTTTCTGTTTAA
      310     320     330     340     350     360
GLN ARG LEU LEU SER GLU ARG GLY LEU HIS VAL LYS GLU SER SER THR ILE GLU PHE TYR
CAAAGGCTCCTCTCAGACAGCGGCTCTTCATGTGAAGGAGACAGCACAAATACAGCTTTTAT
      370     380     390     400     410     420
GLN PHE LEU ILE LYS VAL SER LEU GLY PHE PRO LYS LYS GLU ASP LEU ASN LEU GLN ASP
CAGTTTCTGATAAAGCTTCCCTTCTGCTTCCCGAAGAGAGGAGGACTTGAATTTACAAGAT
      430     440     450     460     470     480
TRP LYS ARG VAL GLY ARG GLU MET LYS LYS TYR ALA ALA ASP ASP GLY THR ASP SER ILE
TGGAAAAGCGGTAGCAAGACAAATGAAGAAATATGCAGCGGACCATGGGACACATAGCATA
      490     500     510     520     530     540
PRO LYS GLN ALA TYR PRO ILE TRP LEU GLN LEU ARG GLU ILE LEU THR GLU GLN SER ASP
CCAAAGCAGCCCTACCCCATTTGGCTTCACTTAAGAGAGATATTAACAGAACACTCAGAT
      550     560     570     580     590     600
LEU VAL LEU LEU SER ALA GLU ALA LYS SER VAL THR GLU GLU GLU LEU GLU GLU GLY LEU
TTGCTTACTATCCGCTCAGGCCAAAGTCTGCTGACTGAAGAAAGACTTAGAAGCAAGCTTTA
      610     620     630     640     650     660
THR GLY LEU LEU SER ALA SER SER GLN GLU LYS THR TYR GLY THR ARG GLY THR ALA TYR
ACCGGGCTACTATCAGCAAGTTCAACAAGAAAAAAGCTTATGGGACCCAGGGGAACCCATAT
      670     680     690     700     710     720
ALA GLU ILE ASP THR GLU ALA ASP LYS LEU SER GLU HIS ILE TYR ASP GLU PRO TYR GLU
GCAGAAATACACACAGCGCACAAATTCCTGAAACACATTTATGATCAACCATATGAA
      730     740     750     760     770     780
GLU LYS GLU LYS ALA ASP LYS ASN GLU GLU LYS ASP HIS VAL ARG LYS VAL LYS LYS ILE
GAAAGCGCAAGCCAGATAAAATGAGCAAAAGGACCATGTTAGAAAGGCTAAGAAAATA
      790     800     810     820     830     840
VAL GLN ARG LYS GLU ASN SER GLU HIS LYS ARG LYS GLU LYS ASP GLN LYS ALA PHE LEU
GTACAAAGCAAAATAAGTCAACATAAGCAAAAGACAAAGCATCAAAGCCCTTTTAA
      850     860     870     880     890     900

```

```

ALA THR ASP TRP ASN ASP ASP ASP LEU SER PRO GLU ASP TRP ASP ASN LEU GLU GLU GLN
GCCACAGATTGGAAACGATGATGACCTGTCCCTGACGATTGGGATAAATTTGGAGGAACAA
          910                920                930                940                950                960

ALA ALA HIS TYR HIS ASP ASP ASP GLU LEU ILE LEU PRO VAL LYS ARG LYS VAL VAL LYS
GGCGCACATTATCATGATGATGATGATGAGCTTGCATCCTTCCAGTAAAAAGGAAGGTGCTTAAG
          970                980                990                1000               1010               1020

LYS LYS PRO LEU ALA LEU ARG ARG LYS PRO LEU PRO PRO VAL GLY PHE ALA GLY ALA MET
AAGAAACCTCTGGCACTCAGCAAGCAAACCCCTGCCCTCCGGTAGGTTTTGCCAGGAGCAATG
          1030               1040               1050               1060               1070               1080

ALA GLU ALA ARG GLU LYS GLY ASP LEU THR PHE THR PHE PRO VAL VAL PHE MET GLY GLU
GCAGAGGCCAGGCAAAAGGACAGATTTGACTTTTACTCTTCTCTGTACTTTTTATGGGAGAA
          1090               1100               1110               1120               1130               1140

SER ASP ASP ASP ASP THR PRO VAL TRP GLU PRO LEU PRO LEU LYS THR LEU LYS GLU LEU
AGTGTGATGATGATGATACACCTGTTTGGCAACCCCTGCCATTGAAAACCTTAAAGAACTA
          1150               1160               1170               1180               1190               1200

GLN SER ALA VAL ARG THR MET GLY PRO SER ALA PRO TYR THR LEU GLU VAL VAL ASP MET
CAATCGGCAGTTAGGACTATGGGACCATCTGCTCCATACACCTTGGAGGTGGTAGACATG
          1210               1220               1230               1240               1250               1260

VAL ALA SER GLN TRP LEU THR PRO SER ASP TRP HIS GLN THR ALA ARG ALA THR LEU SER
GTAGCTAGTCAGTGGCTCACCCCGAGTGATTGGCCAAACAGCCAGAGCTACCTTATCC
          1270               1280               1290               1300               1310               1320

PRO GLY ASP TYR VAL LEU TRP ARG THR GLU TYR GLU GLU LYS SER LYS GLU
CCTGGGGATTATGTTTTATCGGAGAACTGAATATGAAGAAAAGTAAGAG
          1330               1340               1350               1360               1370
    
```

9

Figure 2. Nucleotide sequence of the first 1371 nucleotides of the 5' end of the MMTV genome and deduced amino acid sequence.

the envelope glycoprotein hybridizes only very weakly with the 1.1 kb Pst I fragment sequenced, while it is initiated within the left LTR, 268 nucleotides upstream of the Pst I site (7, 8). It is therefore very likely that the 5' splicing site is located within the sequenced regions in the vicinity of this Pst I site.

This was demonstrated by the S1 mapping method (31) using an end-labeled probe (32). This technique allows one to measure directly the distance of the splicing site from a labeled restriction site in the DNA. We used as a probe a 2.7 kb DNA fragment isolated after digestion with Pvu II of the plasmid pBX (8). It contains MMTV DNA sequences from the Pvu II site (located 134 nucleotides from the cap, 8) to the Xho I site located ~ 1.4 kb downstream, plus pBR 322 sequences. The MMTV portion was labeled with (alpha-³²P) dATP at the 3' end of its minus strand (Fig. 3) and annealed with RNA from infected or transfected cells. After digestion of the hybrids with S1 nuclease, the protected duplex fragments were analysed by electrophoresis on a 6% polyacrylamide sequencing

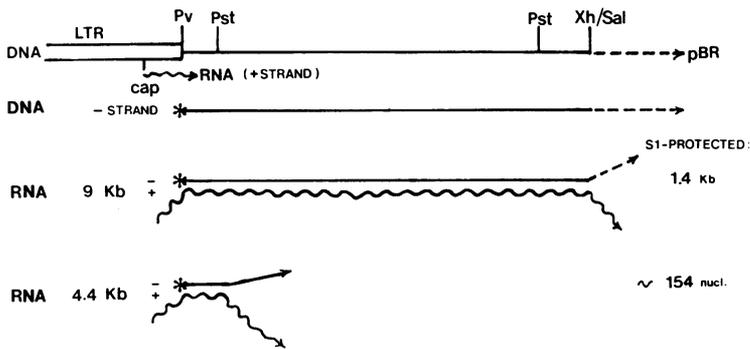


Figure 3. Experimental design for determination of the 5' splicing site. $3'$ - ^{32}P -labeled probe was a 2.7 kb Pvu II fragment of the plasmid pBX (8), where the MMTV DNA was labeled in the minus strand at the Pvu II site (the other $3'$ -end label was in plasmid DNA).

gel (Fig. 4). For reference $5'$ - ^{32}P end-labeled SV40 DNA fragments were electrophoresed in a parallel lane (1). A major fragment of ~ 154 nucleotides was protected by RNA of L cells that had acquired cloned MMTV DNA by transfection (28) (lane 2). The same fragment was also protected by RNA of the GR tumor cell line (lanes 3 and 4). Its abundance was much less than in transfected cells, in agreement with our previous observation (28, 33) that the relative amounts of subgenomic mRNA, as compared with full-length 9 kb RNA, are higher in transfected cell clones than in GR cells. This result locates the 5' splicing site of the major subgenomic RNA at ~ 154 nucleotides downstream of the T of the Pvu II site, at position ~ 289 from the cap. An analysis of the nucleotide sequence of this area (Fig. 2) reveals the presence of a possible 5' splicing consensus sequence (34) AG/GTAGG, where cleavage would occur after the G at position 288. A possible alternative site is located 4 nucleotides downstream from this site. The heavily labeled band near the origin in Figure 4 is the fragment of about 1.4 kb that is protected by the 9 kb MMTV RNA, as shown more clearly in Figure 5. RNA from untransfected L cells did not protect any detectable DNA fragment (lane 5). The multiple fainter bands in slot 2 comigrating with bands in slots 3 and 4 probably represent additional S1-sensitive sites in the RNA-DNA duplexes. Inspection of the sequence data (Fig. 2) in the relevant area suggests that at least some of them may be due to local denaturation of the duplex in (A+T) rich regions leading to S1 sensitivity under the

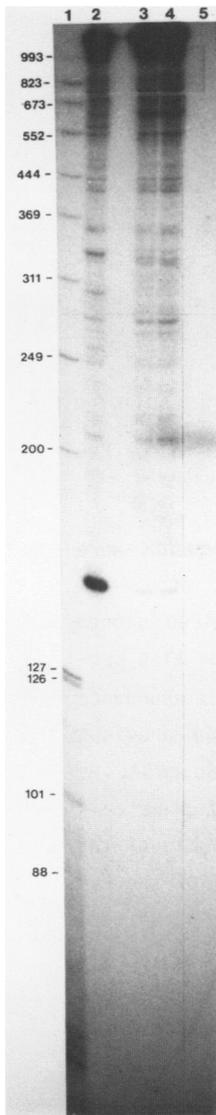


Figure 4. Determination of the 5' splicing site. The experimental design and the 3'-³²P-labeled DNA probe are described in the legend to Figure 3. Approximately 20 ng of DNA probe per 25 µg total RNA were incubated for annealing. Hybrids were digested with S1 nuclease and analyzed on a 6% sequencing gel as described in Materials and Methods. Hybridization reactions included the following RNA's. Lane 2 : 25 µg of total RNA from an L cell clone (H2-3) transfected with the 9 kb exogenous GR-MMTV insert (28). Lanes 3 and 4 : 5 µg of poly (A)-containing RNA plus 45 µg of poly(A)-minus RNA from the GR tumor cell line. Lane 5 : 50 µg of Ltk-total RNA. All cells were grown for 20 hrs in medium containing 10⁻⁶ M dexamethasone before RNA extraction. Lane 1 : 5'-³²P-labeled marker fragments of SV40 DNA digested with Eco RII. One single film exposure is shown. The shade above nucleotide 200 in lane 5 is an artefact of the exposure.

conditions of digestion used (1'000 units/ml, 35 min at 37°C). Likewise the fragment of ~ 250 nucleotides (position 390) might be due to S1 digestion between a 5 bp inverted repeat when a potential cruciform structure could be formed. In other cases we cannot exclude the existence of alternative minor 5' splicing sites, since the sequence data reveal the presence of variants of a consensus sequence (34). In any

case, infected and transfected cells show a coincidence of minor S1-sensitive sites, except for one. This observation supports the notion of colinearity between viral RNA and cloned MMTV DNA (see below). It is noteworthy that the only discrepancy (fragment of ~ 305 nucleotides, position 443) maps in the middle of an 8 bp direct repeat separated by 2 T's, which had been tentatively identified as the dimer linkage structure (see above).

Colinearity of the cloned DNA with RNA from infected cells

Other laboratories reported the observation that attempts to clone MMTV DNAs of different origins very often yielded recombinant DNA molecules containing rearrangements and deletions in the region of the 1.1 kb Pst I fragment (7, 21). The 9 kb insert of recombinant lambda DNA from which the subclones used for sequencing were derived was considered to be a complete copy of MMTV DNA based on the following evidence. In transfected cells we found MMTV RNA and precursor proteins that comigrated electrophoretically with the respective viral products isolated from MMTV-infected cells (28, 35, 36). However this assay for biological activity could not detect silent deletions or rearrangements in the 5' noncoding region or small in frame deletions in the gag coding part of the DNA.

The S1 nuclease assay described above was used to test the extent of colinearity in the 5'-proximal ~ 1.5 kb of the genome between the cloned DNA and RNA from infected cells. After annealing of the RNAs with the 2.7 kb Pvu II fragment carrying a 3' end label in the minus strand (Fig. 3) and digestion with an excess of S1 nuclease at 37°C, the protected duplex fragments were analyzed by electrophoresis on neutral or alkaline agarose gels (27). Figure 5 shows that DNA fragments ~ 1.4 kb long, corresponding to the entire MMTV DNA portion of the probe, were protected by both RNAs, whether from cells transfected with the 9 kb cloned DNA (lanes 1 and 1') or from the virally-infected GR tumor cell line (lanes 2 + 3 and 2' + 3'). RNA from untransfected cells did not protect any distinct DNA bands (lanes 4 and 4'). Under the conditions used, discontinuities like insertion or deletion loops, as well as more complex rearrangements, would lead to S1-sensitive sites in both strands (31) in at least a fraction of the molecules, and therefore to DNA fragments shorter than 1.4 kb. The only distinct fast-migrating band in Figure 5 is the DNA fragment of ~ 154 nucleotides protected by spliced RNA (see Fig. 4), which is particularly abundant in transfected cells.

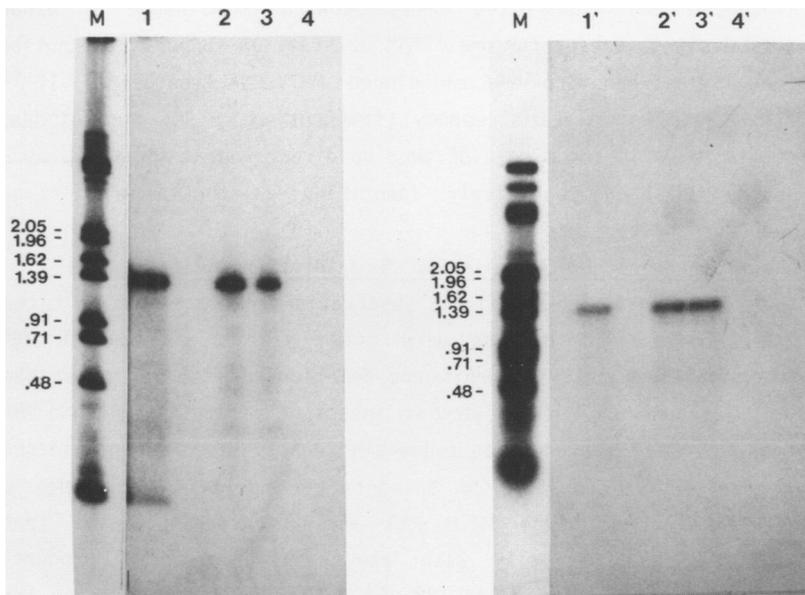


Figure 5. Colinearity of the cloned DNA with RNA from infected cells. Approximately 20 ng of probe was annealed with 25 μ g of RNA; hybrids were digested with S1 nuclease and aliquots were electrophoresed in neutral (samples 1-4) or alkaline (samples 1'-4') 1% agarose gels as described in Materials and Methods. Lanes 1 and 1' : RNA from an L cell clone (H2-3) transfected with the 9 kb exogenous GR-MMTV insert (28). Lanes 2, 2' and 3, 3' : RNA from the GR tumor cell line. Lanes 4 and 4' : RNA from Ltk⁻ cells. All cells were grown for 20 hrs in medium containing 10^{-6} M dexamethasone before RNA extraction. Lane M : 5'-³²P-labeled marker fragments of λ wt DNA digested with Eco RI and Hind III.

We therefore conclude that cells transfected with the 9 kb cloned MMTV DNA insert synthesize MMTV RNA that is largely colinear with GR RNA.

DISCUSSION

Starting from a biologically active lambda recombinant phage carrying the complete genetic information of GR-MMTV we have subcloned and sequenced the region corresponding to the first 1371 nucleotides at the 5' end of the viral genome. This is the region of MMTV DNA which has been difficult to clone in several laboratories and which has been proposed to contain a "poison" sequence (21, 18a). Also in our laboratory a certain number of difficulties arose in the handling and subcloning of the subgenomic DNA fragments containing this area of the viral genome. In M13mp7 phages containing only the first \sim 400 nucleotides of the Pst

I fragment (0-400, Fig. 1) the insert was always present in only one orientation. When sequences to the left or the right of this area were also present the fragments were clonable in both orientations. Frequent deletions of discrete length (~ 400 bp) have also been observed in the course of the experiments. Despite these difficulties we ended up with a series of plasmids and phages covering the critical region.

A large open reading frame extends from the TATA box (not shown) over the whole 5' region of the viral RNA. Starting at the AUG in position 313-315 from the 5' end, the first 353 amino acids of the gag precursor polypeptide can be predicted. As the N-terminal end of this protein has not been sequenced the assignment of the gag initiation codon remains to be determined. However it has been recently shown that the gag precursor protein of MMTV is myristylated (22). The acceptor site for this protein-modification has been shown to be a glycine residue and myristylation has been found at the N-termini of retroviral gag proteins (37). It is therefore very likely that the N-terminal sequence of the pr76 precursor molecule is Met-Gly, corresponding to the two codons found in position 313-318 of the viral RNA. The next initiation codon in position 502-504 is not followed by a glycine residue and could not be modified. For these reasons we assume that the first possible initiation codon is used for the translation of the gag gene product.

In the derived amino acid sequence three possible cleavage sites for a trypsin-like protease are observed. The first one is found after amino acid 59, the second after amino acid 187 and the third one after amino acid 232. If gag-precursor cleavage is conservative as has been observed frequently (38) and if only two of these sites were used, peptides of 59 and 128 or 173 amino acids would be generated. These are slightly shorter than the molecular weights of p10 and p21 determined. In the processing of the N-terminal part of the murine leukemia virus gag-precursor cleavage sites between an aromatic residue and proline are used (38). Two such sequences are found at positions 81 and 270. If these processing sites were used peptides of 81 and 188 amino acids would be created. These values correspond better to the ones observed for p10 and p21.

Several functional features common to retroviruses were located in the sequenced fragments. These include the terminal repeat of 11 nucleotides at the 5' and 3' end of viral RNA (9) and a pyrimidine-rich repetitive sequence (435-452 nucleotides from the 5' end) which might be

involved in the dimer linkage structure found between two viral RNA molecules (29, 30).

The 5' splicing site for the subgenomic mRNA's, most of which is 4.4 kb env mRNA (33), was located at nucleotide 288 from the 5' end of the viral RNA. The same 5' splicing site has been identified for the 1.7 kb mRNA coding for the putative LTR protein (19). The subgenomic mRNAs do not include the AUG of the gag gene and contain their own initiation codon (1-4, 18). They also miss the hypothetical dimer linkage structure and should therefore not be efficiently packaged into virus particles.

The colinearity of the cloned DNA with RNA from infected cells was also examined, using the S1 nuclease assay under stringent conditions. This method should allow the detection of small deletions or rearrangements in the cloned DNA. We were unable to detect such defects and therefore conclude that the complete DNA clone used in our previous transfection experiments and the subgenomic DNA fragments sequenced here are largely colinear with viral RNA isolated from infected cells. This is in agreement with our previous observation that precursor polypeptides of the correct length are synthesized in cells transfected with cloned MMTV DNA (28, 35, 36).

ACKNOWLEDGEMENTS

We thank M. Bernard for expert technical assistance, R. Sahli for his advice, L. Bucher for help with the computer, J.L. Darlix for his interest and suggestions and P. Beard for critical reading of the manuscript. This work was supported by the Swiss National Science Foundation and a grant from the Swiss Cancer League. J. Firzloff was a recipient of a fellowship of the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Donehower, L.A., Huang, A.L. and Hager, G.L. (1981) *J. Virol.*, **37**, 226-238.
2. Donehower, L.A., Fleurdelys, B. and Hager, G.L. (1983) *J. Virol.*, **45**, 941-949.
3. Fasel, N., Pearson, K., Buetti, E. and Diggelmann, H. (1982). *The EMBO Journal* **1**, 3-7.
4. Kennedy, N., Knedlitschek, G., Groner, B., Hynes, N.E., Herrlich, P., Michalides, R. and van Ooyen, A.J.J. (1982). *Nature* **295**, 622-624.
5. Huang, A.L., Ostrowski, M.C., Berard, D. and Hager, G.L. (1981) *Cell* **27**, 245-255.
6. Ucker, D.S., Ross, S.R. and Yamamoto, K.R. (1981) *Cell* **27**, 257-266.
7. Ucker, D.S., Firestone, G.L. and Yamamoto, K.R. (1983) *Mol. Cell Biol.* **3**, 551-561.
8. Buetti, E. and Diggelmann, H. (1983) *The EMBO Journal* **8**, 1423-1429.
9. Klemenz, R., Reinhardt, M. and Diggelmann, H. (1981) *Mol. Biol.*

- Rep. 7, 123-126.
10. Lee, F., Mulligan, R., Berg, P. and Ringold, G. (1981) *Nature* 294, 228-232.
 11. Groner, B., Kennedy, N., Rahmsdorf, U., Herrlich, P., van Ooyen, A. and Hynes, N.E. (1981) in Dumont, J.E., Nunez, J., and Schultz, G. (eds.). *Hormones and cell regulation* 6, Elsevier North-Holland, Amsterdam, pp. 217-228.
 - 11a. Hynes, N., Van Ooyen, A.J.J., Kennedy, N., Herrlich, P., Ponta, H. and Groner, B. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3637-3641.
 - 11b. Chandler, V.L., Maler, B.A. and Yamamoto, K. (1983) *Cell* 33, 489-499.
 12. Payvar, F., Wrangle, O., Carlstedt-Duke, J., Okret, S., Gustafsson, J.A. and Yamamoto, K.R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6628-6632.
 13. Payvar, F., Firestone, G.L., Ross, S.R., Chandler, V.L., Wrangle, O., Carlstedt-Duke, J., Gustafsson, J.A. and Yamamoto, K.R. (1982) *J. Cell. Biochem.* 19, 241-247.
 14. Geisse, S., Scheidereit, C., Westphal, H.M., Hynes, N.E., Groner, B. and Beato, M. (1982) *The EMBO Journal* 1, 1613-1619.
 15. Govindan, M.V., Spiess, E. and Majors, J. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5157-5161.
 16. Pfahl, M. (1982) *Cell* 31, 475-482.
 17. Yamamoto, K.R., Payvar, F., Firestone, G.L., Maler, B., Wrangle, O., Carlstedt-Duke, J., Gustafsson, J.A. and Chandler, V.L. (1983) *Cold Spring Harbor Symp. Quant. Biol.* 47, 977-984.
 18. Redmond, S.M.S. and Dickson, C. (1983) *The EMBO Journal* 2, 125-131.
 - 18a. Majors, J.E. and Varmus, H.E. (1983) *J. Virol.* 47, 495-504.
 19. Van Ooyen, A.J.J., Michalides, R.J.A.M. and Nusse, R. (1983) *J. Virol.* 46, 362-370.
 20. Wheeler, D.A., Butel, J.S., Medina, D., Cardiff, R.D. and Hager, G. L. (1983) *J. Virol.* 46, 42-49.
 21. Majors, J.E. and Varmus, H.E. (1981) *Nature* 289, 253-257.
 22. Schultz, A.M. and Oroszlan, S. (1983) *J. Virol.* 46, 355-361.
 23. Hong, G.F. (1982) *J. Mol. Biol.* 158, 539-549.
 24. Sanger, F., Coulson, A.R., Barrell, B.F., Smith, A.J.H. and Roe, B. (1980) *J. Mol. Biol.* 143, 161-178.
 25. Maxam, A. and Gilbert, W. (1980) in Moldave, K. and Grossman, L. (eds.) *Methods in Enzymology* 65, Academic Press, New York, pp. 499-560.
 26. Scherrer, K. (1969) in Habel, K. and Salzman, N.P. (eds.). *Fundamental Techniques in Virology*, Academic Press, New York, pp. 413-432.
 27. Favalaro, J., Treisman, R. and Kamen, R. (1980) in Grossman, L., and Moldave, K. (eds.). *Methods in Enzymology* 65, Academic Press, New York, pp. 718-749.
 28. Buetti, E. and Diggelmann, H. (1981) *Cell* 23, 335-345.
 29. Darlix, J.L. and Spahr, P.F. (1982) *J. Mol. Biol.* 160, 147-161.
 30. Murti, K., Bonadurant, M. and Tereba, A. (1981) *J. Virol.* 37, 411-419.
 31. Berk, A.J. and Sharp, P.A. (1977) *Cell* 12, 721-732.
 32. Weaver, R.F. and Weissmann, C. (1979) *Nucleic Acids Res.* 7, 1175-1193.
 33. Diggelmann, H., Buetti, E., Fasel, N., Owen, D. and Vessaz, A.L. (1983) In *Understanding Breast Cancer: Chemical and Laboratory Concepts*. M. Rich and P. Furmanski, (eds.), Marcel Dekker, in press.
 34. Sharp, P.A. (1981) *Cell* 23, 643-646.
 35. Owen, D. and Diggelmann, H. (1983) *J. Virol.* 45, 148-154.
 36. Diggelmann, H., Vessaz, A.L. and Buetti, E. (1982) *Virology* 122, 332-341.
 37. Henderson, L.E., Krutzsch, H.C. and Oroszlan, S. (1983) *Proc. Natl. Acad. Sci. USA* 80, 339-343.
 38. Oroszlan, S., Henderson, L.E., Stephenson, J.R., Copeland, T.D., Long, C.W., Ihle, J.N. and Gilden, R.V. (1978) *Proc. Natl. Acad. Sci. USA* 75, 1404-1408.