The EMBO Journal, Vol. 1 No. 1, pp.3–7, 1982

The region of mouse mammary tumor virus DNA containing the long terminal repeat includes a long coding sequence and signals for hormonally regulated transcription

N.Fasel, K.Pearson, E.Buetti, and H.Diggelmann*

Swiss Institute for Experimental Cancer Research, 1066 Epalinges, Lausanne, Switzerland

Communicated by H.Diggelmann
Received on 29 October 1981

Starting from a biologically active recombinant DNA clone of exogenous unintegrated GR mouse mammary tumor virus, we have generated three subclones of PstI fragments of 1.45, 1.1, and 2.0 kb in the plasmid vector pBR322. The nucleotide sequence has been determined for the clone of 1.45 kb which includes almost the complete region of the long terminal repeat (LTR) plus an adjacent stretch of unique sequence DNA. A short region of the 2.0 kb clone, containing the beginning of the LTR, has also been sequenced. Starting with the A of an initiation codon outside the LTR, we detected an open reading frame of 960 nucleotides, potentially coding for a protein of 320 amino acids (36 K). Two hundred nucleotides downstream from the termination codon, and ∿25 nucleotides upstream from the presumptive initiation site of viral RNA synthesis, we found a promotor-like sequence. The sequence AGTAAA was detected ∿15–20 nucleotides upstream from the 3′ end of virion RNA and probably serves as a polyadenylation signal. The 1.45 kb PstI fragment has been transfected into Ltk$^-$ cells together with a plasmid containing the thymidine kinase gene of herpes simplex virus. The virus-specific RNA synthesis detected in a Tk$^+$ cell clone was strongly stimulated by the addition of dexamethasone.

Key words: mouse mammary tumor virus/long terminal repeat/coding capacity/glucocorticoid response/transfection

Introduction

Among retroviruses, mouse mammary tumor virus (MMTV) has two properties that make it particularly interesting to study.

Firstly, it causes mammary carcinomas instead of sarcomas or leukemias as caused by most other retroviruses. No transforming gene has been identified so far for MMTV. However, Dickson and Peters (1981) have shown that polyadenylated 3′ end fragments of viral RNA can be translated *in vitro* into a polypeptide of 36 K which shows no resemblance to any of the structural viral proteins. Very recently the authors were also able to synthesize related polypeptides using complementary RNA made from a cloned PstI fragment containing the long terminal repeat (LTR) of C3H-MMTV DNA (Dickson *et al.*, 1981). It was therefore postulated that this DNA fragment contains a large open reading frame of ∿1000 nucleotides permitting the synthesis of a 36 K protein. In addition, sequence analysis of the LTR of C3H-MMTV DNA (Donehower *et al.*, 1981) suggests an open reading frame of 594 nucleotides sufficient to code for a protein of 20–25 K. So far there is no evidence for the existence of any of these polypeptides in MMTV infected cells or mammary tumor cells.

Secondly, MMTV gene expression is hormonally regulated both in mammary tumors and in cultured tumor cells or infected heterologous cells (Parks *et al.*, 1974; Ringold *et al.* 1975; Scolnick *et al.*, 1976). The hormonal stimulation seems to affect the rate of viral RNA synthesis even in the absence of protein synthesis (Ringold *et al.*, 1975; Scolnick *et al.*, 1976). We have recently shown that cloned MMTV DNA derived from unintegrated circular DNA is biologically active in mouse cells (Buetti and Diggelmann, 1981) and mink lung cells (D.Owen and H.Diggelmann, unpublished data) and that its expression is stimulated by glucocorticoid hormones (Buetti and Diggelmann, 1981). These experiments strongly suggested that the sequences responsible for the hormone response were located on the viral DNA itself.

The goal of the experiments presented in this paper was to determine the DNA sequence of the regions of GR-MMTV DNA containing the LTR in order to examine potential signals for transcriptional control and its coding capacity. Furthermore, we decided to investigate its biological activity after transfection into mouse cells.

Our data demonstrate that, in addition to the expected signals for initiation of viral transcription and polyadenylation, the 3′ end of the MMTV genome contains an open reading frame large enough to code for the 36 K protein evidenced in the *in vitro* translation experiments. Our transfection studies suggest that the sequences necessary for the response to glucocorticoid hormones are located within the 1.45 kb PstI fragment containing almost the complete LTR in addition to a few nucleotides of unique sequence DNA from the 5′ end of the viral genome.

Results

Subcloning of PstI fragments of GR-MMTV DNA in pBR322 and sequence analysis

The 9 kb insert DNA of a λ recombinant phage containing the complete information of GR-MMTV (Buetti and Diggelmann, 1981) was digested with the restriction enzyme PstI and the resulting fragments of 1.1 kb, 1.45 kb, and 2.0 kb were inserted into the PstI site of pBR322. For the plasmid containing the 1.45 kb fragment which spans almost all of the LTR (Majors and Varmus, 1981), we first established the restriction enzyme map by the method of Smith and Birnstiel (Figure 1). The complete base sequence of both DNA strands was then determined using restriction fragments generated with the enzymes AluI, HhaI, HinfI, HaeIII, RsaI, and TaqI, according to the method of Maxam and Gilbert (1980). To sequence the 5′ end of the LTR, situated in the 2.0 kb PstI fragment, the following strategy was used: the large BglII fragment including the LTR was isolated from the complete 9 kb insert and the base sequence was determined using the method of Maxam and Gilbert starting at the BglII site towards the LTR across the PstI site. Both ends of the 1.45 kb fragment were sequenced by the method of Sanger (Sanger *et al.*, 1980) after recloning the fragment into the HincII site of bacteriophage M13mp7 (Messing *et al.*, 1981).

The complete sequence of the 1.45 kb PstI fragment and

*To whom reprint requests should be sent.

 0261-4189/82/0101-0003$2.00/0

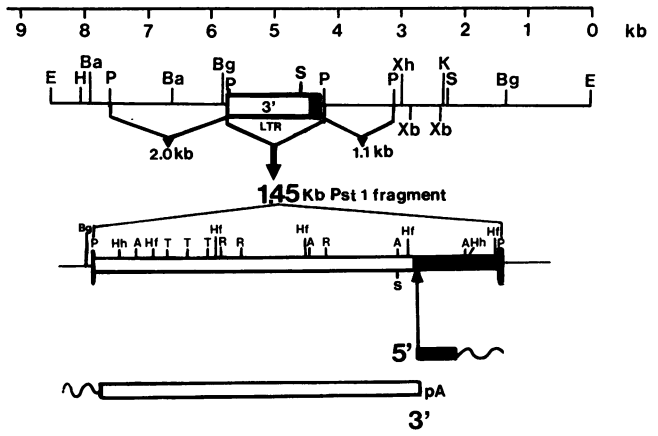

Fig. 1. Restriction endonuclease map of the cloned complete, permuted MMTV DNA (clone H) and the subcloned 1.45 kb PstI fragment. Ba, BamHI; BG, BglII; E, EcoRI; H, HindIII; K, KpnI; P, PstI; S, SstI; Xb, XbaI; Xh, XhoI; Hh, HhaI; A, AluI; Hf, HinfI; T, TaqI; R, RsaI. The bars at the bottom show the 5′ and 3′ end of viral RNA giving rise to the LTR. The wavy lines represent adjacent viral RNA sequences.

the adjacent 40 nucleotides of the 2.0 kb PstI fragment of GR-MMTV DNA is shown in Figure 2. For comparison, the corresponding region of C3H-MMTV DNA sequenced by Donehower *et al.*, (1981) is indicated on the line below (only the bases differing from the GR-MMTV sequence are indicated). The beginning and the end of the LTR have been determined previously (Majors and Varmus, 1981; Donehower *et al.*, 1981) by sequencing several MMTV-DNA fragments of integrated proviruses. The approximate end of viral RNA (∇) has been located by alignment with the sequence of the cDNA generated from the 3′ end of virion RNA (Klemenz *et al.*, 1981). The DNA sequence analysis demonstrates the presence of an AT-rich region, or Hogness box, (ATAATATAAAA in position 1192–1202) 23–33 nucleotides upstream from the proposed cap site (in position 1226, L.D.Johnson, personal communication). A polyadenylation signal (AGTAAA) is present in position 1219–1224 ∿15–20 nucleotides before the 3′ end of the virion RNA. A direct repeat of 10 base pairs (CTTATGTAAA) precedes the Hogness box by a few nucleotides. The DNA sequence obtained was examined for open reading frames for the synthesis of viral polypeptides. As shown in Figure 3, a reading frame starting with an ATG in position 32–34 is open for 960 nucleotides (until nucleotide 991). In addition, the region between nucleotides 1240–1390 is open in all three reading frames. The results of the sequence analysis are summarized in Figure 4.

Transfection of the 1.45 kb PstI fragment into Ltk⁻ cells

To study the biological activity of the LTR region, the agarose gel-purified 1.45 kb fragment was transfected into Ltk⁻ cells (Graham *et al.*, 1980) in the presence of a plasmid containing the thymidine kinase *(tk)* gene of herpes simplex virus (Wilkie *et al.*, 1979). Cell clones that had acquired the *tk* gene were selected in "HAT" medium (Littlefield, 1964). The presence of new MMTV sequences in the DNA of Tk⁺ cell clones was investigated by the Southern transfer technique (Southern, 1975). Figure 5 shows MMTV-specific DNA present in high molecular weight DNA of transfected cells (lane 3) or in EcoRI fragments generated from the same DNA (lane 2) and from Ltk⁻ cell DNA (lane 1) for comparison. No MMTV-specific DNA band was found in the 1.45 kb region of the undigested DNA suggesting integration of the transfected DNA into larger structures. In addition to the endogenous MMTV DNA bands present in L cell DNA, a single new EcoRI restriction fragment of 7.5 kb was found in the transfected cell DNA suggesting the integration of one copy of the 1.45 kb DNA.

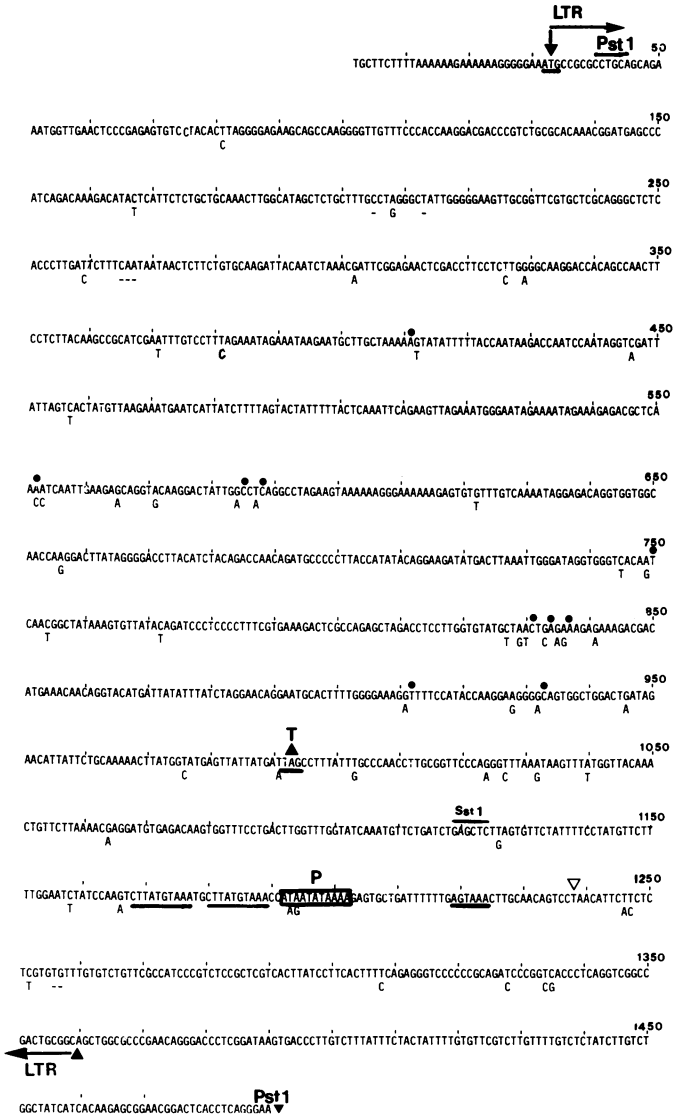

Fig. 2. Nucleotide sequence of the 1.45 kb PstI fragment and of the last 40 nucleotides of the neighbouring 2.0 kb PstI fragment of GR-MMTV DNA. The upper line indicates the GR-MMTV sequence. The nucleotides differing in the C3H-MMTV sequence (Donehower *et al.*, 1981) are indicated on the lower line: ∇, approximate 3′ end of virion RNA; ●, amino acid substitutions expected between the polypeptides postulated for GR- and C3H-MMTV; ▼ ▲, initiation and termination sites of translation; P, promotor-like sequence (Hogness box), LTR = long terminal repeat.

Total RNA was extracted from Ltk⁺ cell clones grown in the presence or absence of 10^{-6} M dexamethasone. The RNA was bound to nitrocellulose filters and the amount of MMTV-specific RNA was estimated by hybridization to an excess of nick-translated [^{32}P]MMTV DNA. As shown in Figure 6, for one of the clones, a small amount of virus-specific RNA was detected in the absence of dexamethasone. The amount of MMTV-specific RNA was strongly increased in hormone-treated cells. In several other cell clones we were unable to demonstrate MMTV-specific RNA synthesis. The state of the DNA of these clones is under investigation. L cells transfected with the *tk* plasmid alone contained undetectable

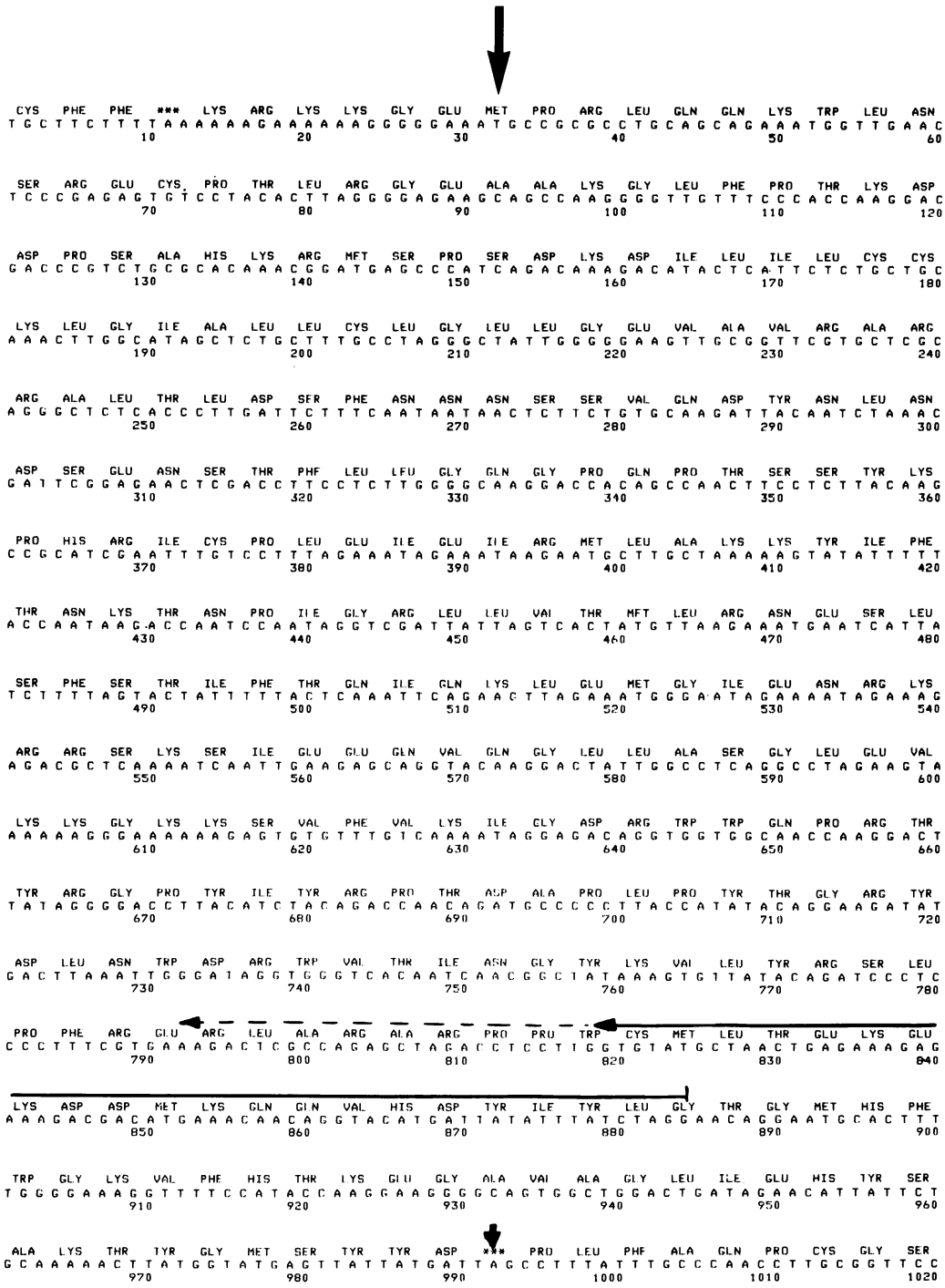

Fig. 3. Proposed amino acid sequence of the 36 K protein as derived from the DNA sequence in Figure 2. —► ---► Peptide of 23 and 32 amino acids synthesized by solid phase peptide synthesis by B.Gutte according to Merrifield (1969).

levels of MMTV-specific RNA.

Discussion

Comparison of the base sequence of the GR strain of MMTV presented here with the one obtained from the C3H strain (Donehower *et al.*, 1981; J.Majors, personal communication) revealed a considerable number (4%) of base differences. Within the potential coding region the observed base variations generally occurred in the third base of a codon and led to either no change in the amino acid or conservative amino acid substitutions. The open reading frame observed by Donehower *et al.* was much shorter than the one reported here, due to the presence of two additional bases (CT) between nucleotides 323 and 324 which lead to a change in reading frame. Unpublished data by L.Donehower and J.Majors (personal communication) now also suggest a 960 bp open reading frame in C3H-MMTV. The corresponding protein of about 36 K has not yet been detected in MMTV infected or mammary tumor cells, possibly because of the lack of an antiserum. However, it seems extremely unlikely that such a large coding region exists without being used. On the basis of the proposed amino acid sequence,

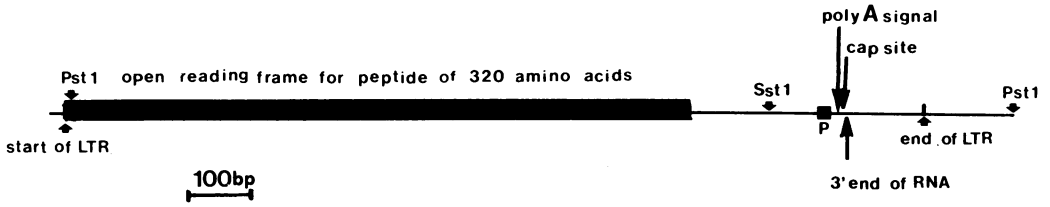

Fig. 4. Regulatory sequences and coding capacity of the DNA region sequenced.

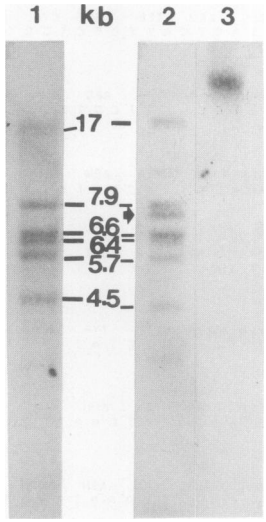

Fig. 5. Southern blot analysis of DNA from L cells transfected with the 1.45 kb PstI fragment. Ten micrograms of cellular DNA were electrophoresed in a 0.8% agarose gel without digestion (lane **3**) or after digestion with the restriction endonuclease EcoRI (lanes **1** and **2**). The DNA was blotted onto a nitrocellulose filter and hybridized to cloned [^{32}P]MMTV DNA. Lane **1**, DNA extracted from Ltk$^-$ cells; Lane **2** and **3**, DNA from transfected cell clone. The molecular weights were calculated from λ Charon 4A EcoRI and λ wt HindIII DNA fragments run in a parallel slot.

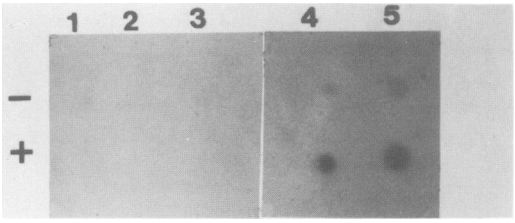

Fig. 6. MMTV specific RNA in cells transfected with the 1.45 kb PstI fragment. Total RNA extracted from a Ltk$^+$ cell clone transfected either with a *tk* plasmid (1 = 10 μg; 2 = 20 μg; 3 = 30 μg) or with a *tk* plasmid plus the 1.45 kb PstI fragment (4 = 10 μg; 5 = 20 μg). RNA extracted from cells grown in the absence (−) or presence (+) of 10^{-6} M dexamethasone was applied in 1−3 μl to a nitrocellulose filter. MMTV-specific RNA was detected by hybridization to cloned [^{32}P]MMTV DNA and autoradiography.

B.Gutte (Univeristy of Zürich) synthesized a peptide of 23 amino acids (marked with a solid line in Figure 3) by solid phase peptide synthesis (Merrifield, 1969). This oligopeptide has been injected into rabbits in order to produce antibodies. A larger peptide, indicated with a dotted line in Figure 3, is also being made. We have chosen this area of the peptide sequence because it contains both a highly conserved region between GR- and C3H-MMTV strains (nucleotides 842−886) and a highly diverged region (nucleotides 827−842), which might allow the generation of type-common and type-specific antibodies.

Our data are in agreement with the predictions of the *in vitro* translation data, as the molecular weight of the proposed polypeptide was calculated to be 36.9 K, corresponding to the molecular weight of the products obtained by Dickson and Peters (1981) and Dickson *et al.* (1981). Since the plasmid used in the latter study for the synthesis of RNA lacks the first nine bases, one would have to postulate that the observed 36 K protein was not initiated with the correct codon, but within the β-lactamase gene to which the fragment had been fused. The shorter peptides observed could be due to initiation of translation on internal AUG's. The calculated size of such peptides is 32.7 K, 23.3 K, 21 K, 18.6 K, 6.5 K, 3.8 K, and 0.7 K. Several bands corresponding to the predicted ones have been observed in the *in vitro* translation experiments of Dickson *et al.* (1981). The fact that the A of the initiation codon is located outside the LTR within the viral genome indicates that it can only be functional in the right LTR. It is possible, however, that after the integration event cellular sequences provide the left LTR with the necessary A in addition to upstream regulatory sequences for transcription and ribosome binding. The postulated protein might play a role in tumor formation or could be involved in the integration of viral DNA or possibly in the hormone-stimulated transcription.

Signals for initiation and termination of transcription were found in the 3′ portion of the LTR, thus confirming previous data (Klemenz *et al.*, 1981). The AT rich sequence of GR-MMTV DNA is 11 nucleotides long, compared to 7 nucleotides in C3H-MMTV (Donehower *et al.*, 1981; Majors and Varmus, 1981). In all these studies the polyadenylation signal AGTAAA was found instead of the usual AATAAA found in most eucaryotic mRNA's. This same sequence (AGTAAA) plus a GT rich region downstream from it is also found in the middle of the LTR in position 598−603. A direct repeat of 10 nucleotides is located two basepairs upstream from the Hogness box. Its functional significance is unknown.

To determine the biological activity of the DNA fragment sequenced, the purified 1.45 kb insert was transfected into Ltk$^-$ mouse cells together with a plasmid containing the *tk* gene of herpes simplex virus. From the Southern blot analysis of DNA isolated from an Ltk$^+$ cell clone three months after transfection, we conclude that a single 1.45 kb fragment had been joined to large molecular weight DNA and was probably stably integrated into cellular DNA. This shows that the left end of the LTR (the last 7 nucleotides) is not indispensable for stable transfection of L cells.

Our data demonstrate that the number of virus-specific transcripts is significantly increased after treatment of the cells with dexamethasone. No virus-specific transcripts were found in cell clones transfected with the *tk* plasmid only. It is therefore very likely that the virus-specific transcription is initiated on the 1.45 kb transfected fragment and that the hormone-responsive sequence is located within this fragment on the LTR or within the first 120 nucleotides of unique sequence DNA. However, although we only found hormone-responsive transcription when the transfecting DNA included

sequences from the 1.45 kb fragment, it cannot be excluded that the incoming MMTV DNA fragment became associated with a cellular hormone-responsive site that produced the observed effect. We are presently trying to locate the hormone-responsive region of the DNA more precisely by using mutilated DNA fragments in transfection experiments.

Materials and methods

Preparation of PstI clones and restriction mapping

The purified 9 kb insert of the bacteriophage λ carrying the complete MMTV-DNA (clone H) (Buetti and Diggelmann, 1981) was cut with the restriction enzyme PstI. The three fragments of 1.1, 1.45, and 2.0 kb were purified on low melting agarose gels and inserted into the plasmid vector pBR322 previously cut with PstI and treated with alkaline phosphatase. The fragments were ligated to the vector using T4 DNA ligase in a 4 day reaction at 0°C. The ligated DNA was used to transform competent *Escherichia coli* HB 101 cells. Amp^s Tet^r transformed colonies were tested for the presence of MMTV DNA by the colony hybridization technique (Grunstein and Hogness, 1975). The DNA of positive colonies was extracted, cut with PstI, and analyzed by Southern blot hybridization (Southern, 1975). The restriction map of the 1.45 kb fragment was established by partial digestion with AluI, RSAI, HaeIII, HhaI, HinfI, and TaqI (Smith and Birnstiel, 1976). A large-scale preparation of the plasmid containing the 1.45 kb PstI insert was done according to the procedure of McMaster (McMaster *et al.*, 1980).

(All restriction enzymes were from BRL Inc. or Biolabs Inc. except for PstI which was a gift from G.Fey).

Sequencing of DNA fragments

The 1.45 kb insert was cut out from the plasmid with PstI and separated on a 1.2% preparative low melting agarose gel (Sigma). The eluted insert was cut with a restriction enzyme, labeled either with T4 polynucleotide kinase (P-L Biochemicals or Boehringer) and [γ^{32}P]ATP (Radiochemical Centre, Amersham) or with *E. coli* DNA polymerase (Klenow Enzyme A, Boehringer) and the appropriate [α^{32}P]dNTP. Labeled DNA fragments were cut with another restriction enzyme, separated on a 5% polyacrylamide gel, and electroeluted into a dialysis bag. DNA sequencing was performed according to the procedure of Maxam and Gilbert (1980).

The 1.45 kb and 2.0 kb PstI inserts were treated with DNA polymerase (Klenow enzyme) in order to obtain blunt ends and cloned into the HincII site of the replicative form of bacteriophage M13mp7 (Messing *et al.*, 1981). The single-stranded DNA was sequenced using the method of Sanger *et al.* (1977; 1980). To sequence across the left PstI site in the 1.45 kb fragment, the 9 kb insert of the complete MMTV DNA (clone H, Buetti and Diggelmann, 1981) was cut with BglII, labeled at the 5′ ends by T4 polynucleotide kinase and [γ^{32}P]ATP. The large BglII fragment was isolated on a low melting agarose gel, cut with XhoI and the two resulting fragments were separated by agarose gel electrophoresis. The larger fragment was sequenced using the technique of Maxam and Gilbert. This sequence was confirmed by the method of Sanger (1980) using the 2.0 kb PstI insert cloned in bacteriophage M13mp7.

Transfection

Ltk^- $aprt^-$ cells provided by M.Perucho and M.Wigler via N.Hynes were transfected with 0.3 μg of purified 1.45 kb insert DNA, 50 ng of a pBR322 plasmid containing the herpes virus *tk* gene (HSC-1-*tk* plasmid; Wilkie *et al.*, 1979, provided by C.Weissmann) and 15 μg of carrier DNA (salmon sperm)/60 mm petri dish using procedures previously described (Graham *et al.*, 1980; Buetti and Diggelmann, 1981).

Extraction of nucleic acids and blot analysis

DNA was extracted from Tk^+ cells after 2–3 months of growth in HAT medium following the procedure of Groner and Hynes (1980). Total cellular RNA was extracted with phenol at 65°C (Scherrer and Darnell, 1962). DNA blots were performed according to Southern (1975) as previously described (Buetti and Diggelmann, 1981). For the spot blot analysis 10, 20, or 30 μg of total RNA (in 1–3 μl) were applied to a nitrocellulose filter and baked *in vacuo* for 2 h at 80°C. The filter was washed and hybridized, as previously described for filters containing DNA fragments (Buetti and Diggelmann, 1981).

Acknowledgements

We thank A.Vessaz, F.Bezencon, and D.Petrelli-Sem for expert technical assistance. O.Hagenbüchle and R.Bovey were very helpful during the initial phase of the sequencing work and O.Hagenbüchle and L.Bucher made the computer program available to us. G.Fey provided us with the PstI enzyme. We also thank P.Beard and T.Graf for critical reading of the manuscript. This work was supported by grants from the Swiss National Science Foundation and the Swiss Cancer League.

References

Buetti,E., and Diggelmann,H. (1981) *Cell*, **23**, 335-345.
Dickson,C., and Peters,G. (1981) *J. Virol.*, **37**, 36-47.
Dickson,C., Smith,R., and Peters,G. (1981) *Nature*, **291**, 511-513.
Donehower,L., Huang,A., and Hager,G.L. (1981) *J. Virol.*, **37**, 226-238.
Graham,F.L., Bacchetti,S., McKinnon,R., Stanners,C., Cordell,B., and Goodman,H.M., (1980) in Baserga,R., Croce,C., and Rovera,G. (eds.) *Introduction of Macromolecules into Viable Mammalian Cells*, Alan R. Liss, Inc. New York, pp. 3-25.
Groner,B., and Hynes,N.E. (1980) *J. Virol.*, **33**, 1013-1025.
Grunstein,M., and Hogness,D. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 3961-3965.
Klemenz,R., Reinhardt,M., and Diggelmann,H. (1981) *Mol. Biol. Rep.*, **7**, 123-126.
Littlefield,J.W. (1964) *Science (Wash.)*, **145**, 709-710.
Majors,J.E., and Varmus,H.E. (1981) *Nature*, **289**, 253-258.
Maxam,A., and Gilbert,W. (1980) in Moldave,K., and Grossman,L. (eds.) *Methods in Enzymology* **65**, New York Academic Press, pp. 499-560.
McMaster,G.K., Samulski,R.J., Stein,J.L., and Stein,G.S. (1980) *Anal. Biochem.*, **109**, 47-54.
Merrifield,R.B. (1969) *Adv. Enzymol.*, **32**, 221-296.
Messing,J., Crea,R., and Seeburg,P.H. (1981) *Nucleic Acids Res.*, **9**, 309-321.
Parks,W.P., Scolnick,E.M., and Kozikowski,E.H. (1974) *Science (Wash.)*, **183**, 158-160.
Ringold,G.M., Yamamoto,K.R., Tomkins,G.M., Bishop,J.M., and Varmus,H.E. (1975) *Cell*, **6**, 299-305.
Sanger,F., Coulson,A.R., Barrell,B.F., Smith,A.J.H., and Roe,B. (1980) *J. Mol. Biol.*, **143**, 161-178.
Sanger,F., Nicklen,S., and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Scherrer,K., and Darnell,J.E., (1962) *Biochem. Biophys. Res. Commun.*, **7**, 486-490.
Scolnick,E.M., Young,H.A., and Parks,W.P. (1976) *Virology*, **69**, 148-156.
Smith,H.O., and Birnstiel,M.L. (1976) *Nucleic Acids Res.*, **3**, 2387-2398.
Southern,E. (1975) *J. Mol. Biol.*, **98**, 503-517.
Wilkie,N.M., Clements,J.B., Boll,W., Mantei,N., Londsdale,D., and Weissmann,C. (1979) *Nucleic Acids Res.*, **7**, 859-877.

Note added in proof

The size of MMTV-specific transcripts in the LTk^+ cell clone transfected with the 1.45-kb PstI fragment was analyzed by blot hybridization. A band of 1.3 kb and possibly some smaller RNA was detected (Figure 7, lane 3).

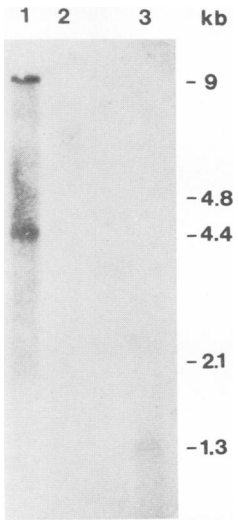

Fig. 7. Size determination of MMTV-specific RNA in transfected cells. Total poly(A)-containing RNA from cells grown in the presence of 10^{-6} M dexamethasone was electrophoresed on a denaturing agarose gel, transferred to nitrocellulose and hybridized with [^{32}P]MMTV cDNA as described (Buetti and Diggelmann, 1981). Transfected cell RNA (10 μg), lane **3**. As controls, 1 μg GR cell RNA (lane **1**) and 3 μg LTk^- RNA (lane **2**) are also shown. The positions of GR RNA and of rRNA from an adjacent slot are indicated as size markers.