The Search for a Distal Promoter Element in the Mouse Mammary Tumor Virus

Evelyn M. Zera
University Undergraduate Fellow, 1992-1993
Texas A&M University
Department of Biochemistry

Approved:

David O. Peterson
Fellows Advisor

Honors Director

Overview

Regulation of gene transcription occurs in a variety of ways, including the direct interaction of proteins with DNA, and the indirect effect of hormones. In addition to daily cell functions, gene regulatory mechanisms have been implicated in the development and proliferation of some diseases such as some forms of cancer. Understanding of these regulatory mechanisms has been approached by extensive biochemical studies of a few basic biological systems. The mouse mammary tumor virus (MMTV), a retrovirus that is present in many mouse cell types but specifically attacks the mammary cells, has been extensively studied for its hormonal regulation of transcription. Retroviruses are increasingly under study because of their unique system of carrying ssRNA which, upon infection of the host cell, is transcribed to dsDNA by the enzyme reverse transcriptase. The host cell genes produce the trans-acting factors which, upon integration of the proviral DNA into the host genome, modulate viral gene transcription. Understanding of transcriptional regulation of the proviral DNA thus gives valuable information which can be broadly applied to many other gene systems. Any cis-acting elements are restricted to the proviral DNA region and are thus easier to locate if not already known. MMTV transcription is positively regulated by hormones and contains several negative regulatory elements (NREs). These two control mechanisms along with the retrovirus characteristics, make it an ideal model system for the study of gene regulation. Similarities between MMTV regulation and regulation of genes in higher eukaryotes have already been found (1). However, before the overall transcription process can be fully understood, an understanding of transcription mechanisms in the absence of regulation must be elucidated. The promoter elements necessary for basal transcription must be clearly identified for an effective study to be done. This research project focuses on determining the presence of one such promoter element.

Background

The basal promoter of the MMTV, located within the long terminal repeats (LTRs) of the proviral DNA, has previously been defined as consisting of those sequences of DNA which are necessary for basal rates of transcription but are not responsive to hormones or other regulators. It has been found to extend 109 base pairs (bp) upstream from the transcription initiation start site (6, 9). Positive regulation of MMTV by

steroid hormones such as androgens (2), glucocorticoids (14, 18), and progestins (2,3), has been localized to cis-acting DNA sequences in the -80 to -200 region (5, 6, 8,13). This region, the hormone response element (HRE), has multiple binding sites for the hormone receptor complex and is considered separate from any promoter elements controlling basal transcription rates. Another regulatory element found in MMTV is one which represses basal transcription. This negative regulatory element (NRE) is found upstream of the HRE between -427 and -364. Overall transcription rates are a combination of the effects of the NRE in the absence of hormone and the HRE in the presence of hormone (4, 7, 9).

Linker-scanning mutations made across the length of the promoter from -102 to +18 along with transient transfection assays have determined the locations of four specific regions necessary for both basal and hormone induced transcription. These regions include the TATA box, NF-1 binding site, a 10bp repeat containing three related octamer sequences, and a sequence near the transcription initiation start site between bp +2 and +10 (12, 17). Mutations in these regions resulted in decreased levels of both basal and hormone induced transcription. However, mutations made in one of the hormone receptor binding sites (located at -94 to -99) decreased hormone induced as well as basal level transcription (17) (Figure 1). This was unexpected as basal transcription levels should be unaffected. One possibility for this result is the presence of an additional promoter element located within or closely linked to the receptor binding site. This research project is designed to explore the possibility of the presence of a distal promoter element by constructing MMTV promoters with random mutations in the region -103 to -87 and determining whether the effects on hormone-induced and basal transcription are separable. This can be accomplished by using PCR primers with random mutations to generate mutated promoters. The mutated promoters can then be separated, cloned, and tested in a transient transfection assay using the chloramphenicol acetyltransferase (CAT) reporter gene.

Materials and Methods

Oligonucleotide synthesis. Two oligonucleotides were synthesized with an Applied Biosystems DNA synthesizer by the DNA Technologies Laboratory, Departments of Biology and Biochemistry & Biophysics, Texas A&M University. DNA sequences synthesized were identical to the MMTV

promoter beginning at base -115, encompassing an SstI recognition site, and continuing downstream until the 17 bp target region (-103 to -87) was reached. The construction of mutant one (mut1) was continued by generating random mutations within the first 10 bases of this region and wild type sequences for the remainder of the region, while mut2 contained random mutations within the last 10 bases of the target region and wildtype sequences in the first seven bases. Each oligonucleotide was extended with the wild type sequence for seven bases past its mutated region (Figure 2). Each base in the mutated region is the result of its addition from a mixture of 50% of the wild type nucleotide, and 12.5% of each of the four nucleotides, thus giving the wildtype nucleotide a 0.625 probability of addition and each of the other three nucleotides a 0.125 probability. The oligonucleotide preparations were resuspended in 200µl ddH_2O and concentrations determined from OD_{260} readings defining one OD_{260} unit as being equal to 33µg DNA per ml.

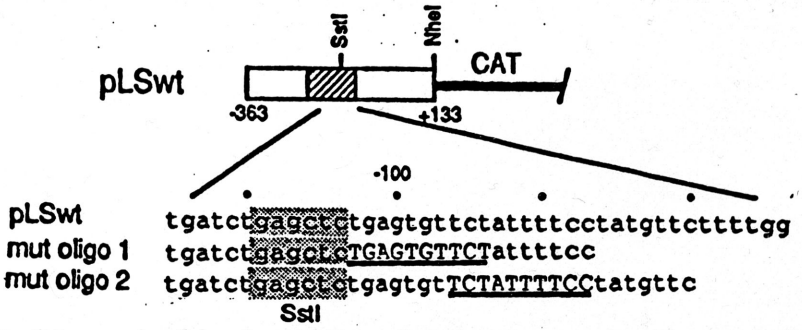

Figure 2. Oligonucleotides for PCR-mediated mutagenesis. The underlined bases in upper case letters indicate positions where the oligonucleotide has a 0.625 probability of containing the wild-type base and a 0.125 probability of containing each of the other three.

Polymerase chain reaction (PCR). The two mutant oligonucleotide preparations were individually subjected to PCR using a Precision Scientific GTC-1 Genetic Thermal Cycler. The PCR reaction used 20ng of pLSwt as a template, 40ng of the mutant oligo as one primer, 40ng of CAT primer which anneals at +195 to +219, 20nmol of dNTPs, Taq buffer and 2.5 units Taq DNA polymerase (both from Promega) brought to a volume of 100µl with ddH_2O. Paraffin oil (100µl) was used to cover and thereby prevent dehydration of the reaction mixture. The mixture was initially heated to 94°C for one minute and thirty seconds in order to allow the template to denature. It was then cooled to 37°C for two minutes to allow the primers to anneal to the individual strands of template DNA. The annealed primer was then extended along the template for three minutes at 72°C. Additional denaturing steps were performed at 91°C for one

minute. The cycle was performed 34 times with a final extension at 72°C for seven minutes. Products were recovered by ethanol precipitation and were subsequently digested with the restriction enzymes SstI (Gibco PRL) and NheI (Promega). Digested products were separated from DNA fragments on a 6% polyacrylamide gel (Figure 3) and recovered by incubating the gel fragment containing the DNA with an equal volume of 0.5 M ammonium acetate and 1mM EDTA solution at 37°C overnight. The DNA was then ethanol precipitated from the supernatant. Recovery was verified with agarose gel electrophoresis and ethidium bromide (EtBr) staining.

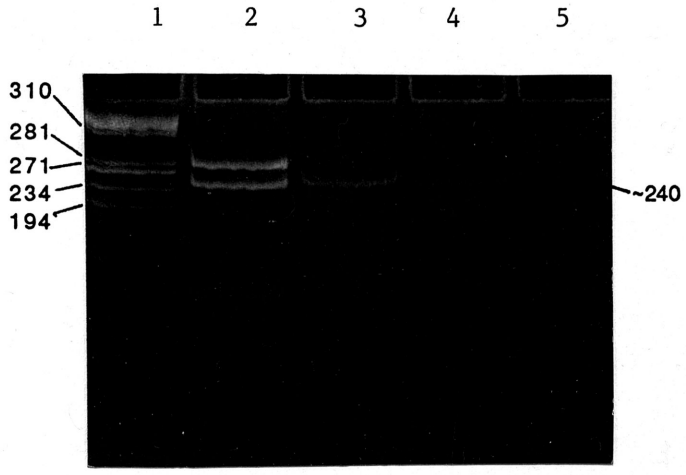

Figure 3. Polyacrylamide gel analysis of PCR product after digestion with SstI and NheI. Lane 1 is a marker of HaeIII digested ϕX174. Lane 2 is a control PCR product which used PBRRI and CAT primers while lanes 3-5 are digested PCR products using mut1 oligonucleotide as one primer and the CAT primer as the other. The indicated bands at ~240bp containing the mutated promoter regions were eluted from the gel and ligated with the pLS template.

Plasmid construction. pLSwt, previously described in Toohey, *et al.* (17), (Figure 4), was used as the template for mutant promoter construction. It was digested with SstI and NheI and run on an agarose gel to separate out the endogenous promoter region. The template was recovered from the gel by cutting out the appropriate band, mashing the gel slice, adding 100µl phenol, vortexing for 10 seconds, and freezing at -70°C for 15 minutes. The sample was then centrifuged for 15 minutes, and the supernatant was removed and extracted first with a phenol/chloroform mixture and then with chloroform. The DNA was then ethanol precipitated and resuspended in 20µl ddH_2O. Ligation with the digested PCR products followed (Ligase from Promega). Ligated plasmids

were transformed into *E. coli* HB101 competent cells using the classic $CaCl_2$ method (15) and plated out on selective media containing 50mM carbenicillin. Colonies were selected and propagated separately.

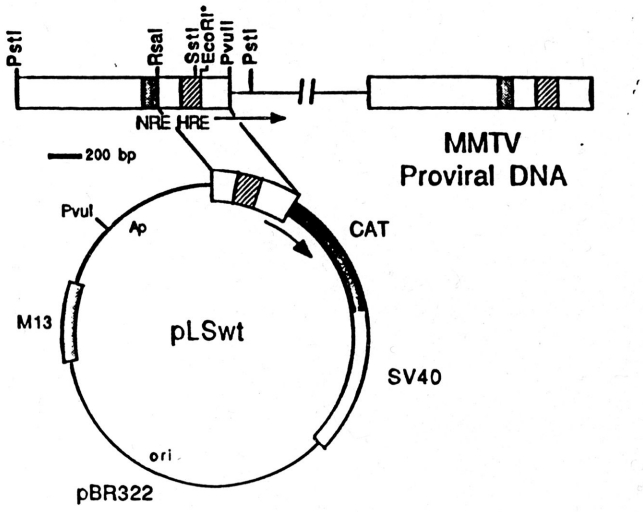

Figure 4. Structure of MMTV LTR and pLSwt. Plasmid pLSwt (5.5 kilobases) contains MMTV LTR sequences from position -363 (RsaI) to +133 (PvuII) with respect to the transcription initiation start site. The positions of the HRE (▨), NRE (▩), coding sequences for CAT (■), and simian virus 40 sequences that specify poly(A) addition (□) are indicated.

DNA sequencing. Individual colonies were grown up overnight in 5ml Luria Broth (LB) containing 50mM carbenicillin. DNA mini-preps were obtained through lysis by boiling and phenol and chloroform extractions (15). Sequencing was performed by the dideoxy chain termination technique using a primer that hybridizes at +114, [α-^{35}S] dATP (Dupont, NEN Research Products) and Sequenase Version 2.0 DNA Sequencing Kit from United States Biochemical Corporation.

Large scale plasmid preps. Selected mutant plasmids were purified by alkaline extraction and CsCl-EtBr equilibrium density centrifugation followed by extraction with NaCl saturated isobutanol (15). The resulting plasmid was further purified on a Biogel A-5m Econocolumn. Plasmid concentrations were determined by OD_{260} readings and verified by agarose gel electrophoresis and EtBr staining.

Results

To determine the optimal length of the target region to be mutated

in the synthesis of oligonucleotide primers according to the specifications in the Materials and Methods section, the probability, Q, of a number of mutations, i, arising in a target size of n bases synthesized, was calculated for a number of different target sizes according to the binomial distribution

$$Q = \binom{n}{i} p^i (1-p)^{n-i}$$

Results of these calculations are shown in Figure 5. A primer with a mutated region of 10bp was selected as it gives maximum probabilities for 3 and 4 mutations, and previous results using this method showed a slight downward shift from the theoretical probabilities when actually sequenced (11).

i	n=6	n=7	n=8	n=9	n=10
0	6.0	3.7	2.3	1.5	0.9
1	21.5	15.6	11.2	7.9	5.5
2	32.2	28.2	23.5	18.9	14.7
3	25.7	28.2	28.2	26.4	23.6
4	11.6	16.9	21.1	23.8	24.8
5	2.8	6.1	10.1	14.3	17.8
6	0.3	1.2	3.0	5.7	8.9
7		0.1	0.5	1.5	3.1
8			0.0	0.2	0.7
9				0.0	0.1
10					0.0

Figure 5. Theoretical percentages of the number of mutations (i) arising from a mutated region of length n using the binomial distribution $Q = \binom{n}{i} p^i (1-p)^{n-i}$

To determine if mutated promoters were indeed formed, plasmids were isolated and sequenced from overnight cultures of individual colonies from the transformed *E. coli* cells. A sequencing gel comparing the wild type promoter sequence to a mutated promoter sequence is shown in Figure 6. The results of the sequencing of the mutated promoters are shown in Figure 7. Promoters contained single point mutations, many single point mutations spread across the target region, and small clusters of point mutations. The percentages of the number of mutations found for the sequenced promoters are shown in Figure 8. These results show a definite downward shift from the theoretical values calculated in Figure 5, showing an observed majority of only two mutations per promoter as opposed to the theoretical majority value of four mutations per promoter.

Mutants having two mutations clustered together were selected for further study in CAT assays. The selected mutants are 1I, 1J, 1N, 1O, and 1W.

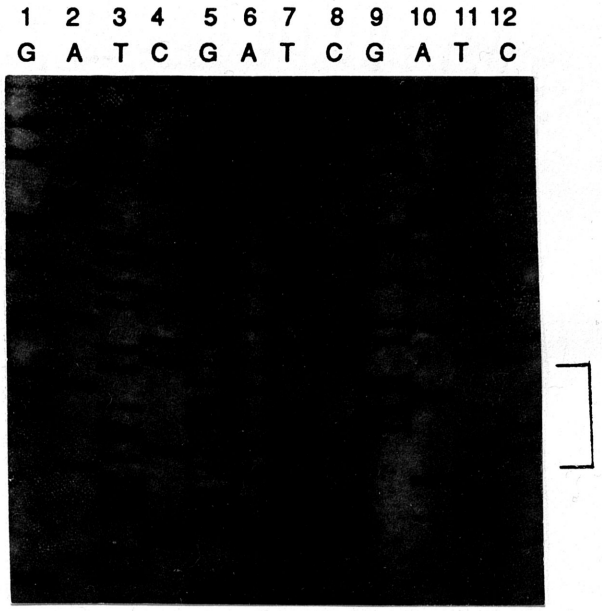

Figure 6. Sequencing gel of MMTV promoter mut1 plasmids. Lanes 1-4 show the wild-type sequence, TGAGT GTTCT, in the region of interest (-103 to -94). Lanes 5-8 give the sequence for mut1R, and lanes 9-12 contain the mut1Q sequence.

		-100		-90
wt	agctc	tgagt	gttct	atttt
mut1E	agctc	Agagt	gAtct	atttt
mut1I	agctc	tgaCt	Tttct	atttt
mut1J	agctc	tCGgt	gttct	atttt
mut1K	agctc	tgagt	gtGct	atttt
mut1L	agctc	tTaAt	CAtct	atttt
mut1M	agctc	tgTgA	gttAt	atttt
mut1N	agctc	tgagt	gCtAt	atttt
mut1O	agctc	GgTgt	gttct	atttt
mut1P	agctc	Cgagt	Tttct	atttt
mut1Q	agctc	GgagG	gCtct	atttt
mut1R	agctc	Cgagt	gttct	atttt
mut1S	agctc	tgagt	gttct	atttt
mut1T	agctc	tgagt	gttct	atttt
mut1V	agctc	tgagt	gttct	atttt
mut1W	agctc	tgagt	CAtct	atttt
mut1X	agctc	tgagt	gttTt	atttt

Figure 7. Mutations in the promoter target region. The relevant portion of each plasmid is shown with the changed bases in capital letters. Each plasmid contains MMTV sequences from -363 to +133 with -115 to +219 being generated by the polymerase chain reaction.

i	calculated	observed
0	0.9	18.8
1	5.5	18.8
2	14.7	43.8
3	23.6	12.5
4	24.8	6.3
5	17.8	0.0
6	8.9	0.0
7	3.1	0.0
8	0.7	0.0
9	0.1	0.0
10	0.0	0.0

Figure 8. A comparison of the theoretical percentages for a number of mutations, i, to be found in a region of length 10 with the actual percentages of mutations found. Sixteen plasmid promoters containing a ten base mutated region were sequenced. Three had no mutations, three had one mutation, seven had two mutations, two had three mutations, and one had four mutations.

The CAT assay is an indirect measure of MMTV transcription because the CAT enzyme activity in extracts of transfected cells is proportional background activity is present in mammalian cells. ^{14}C labeled chloramphenicol is used and the substrate and product are separated by thin layer chromatography. However, in order for this procedure to be used and to be reproducible, plasmid preparations must be quite pure. Large scale plasmid preps that are chromatographed on gel filtration to remove the RNA are sufficient in this regard, producing results that are consistent with other preparations of the same plasmid (10). A large scale plasmid prep has been completed for mut1I.

Discussion

As seen in Figure 8, the percentages of the number of mutations in the promoters sequenced is downshifted from a majority of three and four mutants in the theoretically calculated numbers of mutations to an actual majority of two mutations per mutant promoter. This is most likely a result of the PCR procedure in which the mutated primers must anneal to the pLSwt template. Wild type or near wild type sequences can anneal at a greater number of sites and thus have a greater chance of doing so and being replicated. Primers with many mutations do not have as many sites that match and thus may not anneal well enough to be extended along the

template.

Selection of the mutants to be used in the CAT assay is dependent upon the number and location of mutations. Preliminary studies will focus on promoters containing two or three grouped mutations such as mutations 1I, 1J, 1N, 1O, and1W. These few grouped mutations are spread along the target region and were selected in order to help specify the exact region that is responsible for the decreased basal transcription. Sequences identified to have an adverse affect on basal transcription will be further studied using promoters with single point mutations in the region of interest. In this way, the exact sequences necessary for basal transcription may be elucidated.

The next step in characterizing this region of the MMTV promoter would be to look for a protein that binds to the identified sequences. This can be partially accomplished with DNase I footprinting which can identify the sequences to which a protein binds. Any binding protein must then be isolated and assessed for its ability to act as a trans- factor affecting transcription.

Literature Cited

1. **Argentin, S., Sun, Y.L., and Lihrmann, I.** (1991) *J. Biol.Chem.* **266**, 23315-22.
2. **Cato, A.C.B., Henderson, D., and Ponta, H.** (1987) *EMBO J.* **6**, 363-368.
3. **Cato, A.C.B., Miksicek, R., Schutz, G., Arnemann, J., and Beato, M.** (1986) *EMBO J.* **5**, 2237-2240.
4. **Hsu, C.-L., Fabriatis, C., and Dudley, J.** (1988) *J. Virol.* **62**, 4644-4652.
5. **Hynes, N., Van Ooyen, A.J.J., Kennedy, N., Herrlich, P., Ponta, H., and Groner, B.** (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3637-3641.
6. **Lee, F.C., Hall, C.V., Ringold, G.M., Dobson, D.E., Luh, J., and Jacob, P.E.** (1984) *Nucleic Acids Res.* **12**, 4191-4206.
7. **Lee, J. W., Moffitt, P.G., Morley, K. L., and Peterson, D.O.** (1991) *J. Biol. Chem.* **266**, 24101-24108.
8. **Majors, J., and Varmus, H.E.** (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5866-5870.
9. **Morley, K.L., Toohey, M.H., and Peterson, D.O.** (1987) *Nucleic Acids Res.* **15**, 6973-6989.
10. **Peterson, D.O.** Unpublished results.
11. **Pierce, J.** (1991) Dissertation, Texas A&M University.
12. **Pierce, J., Fee, B.E., Toohey, M.G., and Peterson, D.O.** (1993) *J. Virol.* **67**, 415-424.
13. **Ponta, H., Kennedy, N., Skroch, P., Hynes, N.E., and Groner, B.** (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1020-1024.
14. **Ringold, G.M., Yamamoto, K.R., Tomkins, G.M., Bishop, J.M., and Varmus, H.E.** (1975) *Cell* **6**, 299-305.
15. **Sambrook, J., Fritsch, E.F., and Maniatis, T.** *Molecular Cloning*, 2nd Edition, Cold Springs Harbor Laboratory Press (1989).
16. **Toohey, M.G., Morley, K.L., and Peterson, D.O.** (1986) *Mol. Cell Biol.* **6**, 4526-4538.
17. **Toohey, M.G., Lee, J.W., Huang, M., and Peterson, D.O.** (1990) *J. Virol.* **64**, 4477-4488.
18. **Ucker, D.S., and Yamamoto, K.R.** (1984) *J. Biol. Chem.* **259**, 7416-7420.