EXPRESSION OF ENDOGENOUS RETROVIRUSES IN INBRED MICE: COORDINATE REGULATION AND STRUCTURE OF MULTIPLE TRANSCRIPTION UNITS

Thesis by

David E. Levy

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1985

(Submitted 28 September 1984)

ABSTRACT

The control of expression of the murine antigen Gix and of other products of endogenous retroviruses, in strain 129 mice and in its congeneic partner strain 129 Gix^-, is an example of the coordinate expression of a dispersed family of independent transcription units. In order to provide a molecular description of the Gix phenotype, evidence is presented, from DNA and RNA hybridization analyses using heterologous viral probes, indicating that this phenotype is specified by a distinct regulatory gene, defined genetically, that acts in trans to control the levels of accumulation of specific mRNA species. The steady state levels of several, structurally distinct polyadenylated RNA species are reduced in Gix^- mice, and a major reduction in transcription of these sequences accompanies this drop in abundance. Tissue specific patterns of accumulation of different sized RNA species were detected in numerous organs of the mouse, and the majority of these distinct transcripts were collectively regulated.

The isolation and characterization of cDNA copies of these endogenous retroviral transcripts demonstrated that they were derived from multiple, distinct transcription units. Differences among these RNA species were detected by S1 nuclease protection analyses, which confirmed the tissue specific patterns of RNA accumulation. The nucleotide sequences of endogenous virus cDNA clones fully documented the expression of distinct genes, the nature of the sequence heterogeneity, and the relationship of these normal cellular constituents to exogenous, infectious virus. The polymorphism was found to result from both single nucleotide changes and from deletions of different lengths of coding and non-coding information. Comparison of these sequences with exogenous virus demonstrated that the endogenous transcripts are closely related to the recombinant sequences of leukemogenic, mink cell focus forming viruses.

TABLE OF CONTENTS

FIGURES

TABLES

"We must always remember that the strength, the dignity, the delight of man is to spite and despise the shadows and stars that hide their secrets from us."

—V. Nabokov, Ada

"According to the strictly structural concept, the genome is considered as a mosaic of independent molecular blue-prints for the building of individual cellular constituents. In the execution of these plans, however, co-ordination is evidently of absolute survival value. The discovery of regulator and operator genes, and of repressive regulation of the activity of structural genes, reveals that the genome contains not only a series of blue-prints, but a co-ordinated program of protein synthesis and the means of controlling its execution."

—Jacob and Monod, 1961

INTRODUCTION

INTRODUCTION

Significance

Central to understanding the processes resulting in cellular differentiation is the elucidation of the mechanisms responsible for differential accumulation of distinct mRNA species in different cell types. There are numerous mechanisms and levels of regulation leading to the control of RNA abundance (Darnell, 1982), and much is known about the biochemistry of transcription of individual genes and about the signals involved in transcriptional initiation and RNA processing (Nevins, 1983). Clearly, differentiation must be accomplished by the simultaneous activation of batteries of genes in individual cell types (Davidson and Britten, 1979). However, the events leading to the coordinate regulation of sets of genes and the genetic elements involved in the programming of cell specific gene expression remain largely unknown. Classically, studies of gene regulation have involved the DNA tumor viruses, which can be considered a collection of coding sequences coordinately regulated in a developmental program (Ziff, 1980; Persson and Philipson, 1982). For example, there is distinct temporal control of the transcription of genes from the early region of adenovirus mediated through the action of specific viral proteins (Nevins et al., 1979). Sets of cellular genes which are coordinately expressed or developmentally regulated have also been described. These often involve families of clustered genes, such as the mammalian globin families (Efstratiadis et al., 1980) and the chorion genes of Drosophila sp. (Spradling, 1981) and silkworm (Eickbush and Kafatos, 1982). This clustered genomic organization for coordinately regulated genes may be important to the control of their expression, perhaps because of controlling mechanisms operating through effects on chromatin structure. Other gene systems respond coordinately to specific inducing signals, the best characterized example being the heat shock response in Drosophila sp. (Ashburner and Bonner, 1979). In this case, several dispersed clusters of genes are coordinately activated due to a common upstream DNA sequence (Holmgren et al., 1981; Hackett and Lis 1981; Pelham, 1982).

It is reasonable to expect that the coordinate regulation of sets of genes relies on the action of cis controlling elements. These sequence

enhancer sequences have been identified that activate immunoglobulin gene transcription (Mercola et al., 1983; Gillies et al., 1983; Banerji et al., 1983) and MHC gene expression (Gillies et al., 1984) and are apparently responsible for the tissue specific expression of these developmentally regulated gene products. It has been speculated that tissue specific enhancer elements, controlled by the presence or absence of specific binding proteins in a given cell type, are critically involved in the process of differentiation (Gillies et al., 1983). However, the identity of the agents which must act in trans in these regulatory processes has remained elusive.

One model for the requirement for secondary factors to coordinate gene expression has been demonstrated genetically and biochemically in yeast. Enzymes in several different amino acid biosynthetic pathways are derepressed during amino acid starvation. The activation of these unlinked genes involves the action of several genetically defined elements, probably acting in trans through factors binding to 5' flanking, repeated DNA sequences associated with the structural genes (Hinnebusch et al., 1983). Catabolite regulation of the expression of the metabolic enzymes for galactose utilization in yeast also is controlled by regulatory loci via trans acting, diffusable proteins which coordinately affect the transcription of the structural genes for the enzymes (Broach, 1979). The gene for a negative regulator has been cloned and characterized (Torchia et al., 1984). Similarly defined genetic elements are involved in the coordinate expression of the discoidin gene family in Dictyostelium discoideum in response to developmental signals rather than to metabolic inducers (Alexander et al., 1983). The unlinked tubulin genes of Chlamydomonas reinhardi can also be coordinately induced in response to a regulatory cue (Brunke et al., 1984).

The evolution of multigene systems may provide insights into these control mechanisms (Hood et al., 1975). The elements directing differential expression of these genes might have arisen by divergence or possibly by insertion of new signals adjacent to the transcription unit. Comparison of DNA sequence flanking the genes within a family provides an opportunity to identify potential sites that could be important in the

regulation of their expression. The study of closely related genes may provide clues to determinants of both individual and coordinate expression. Endogenous retroviruses of vertebrates can be considered such a multigene system, although in this case a number of highly related, exogenous genetic elements have entered the genome relatively recently and yet may respond to a variety of host controls.

The nature, origin, and organization of these endogenous viruses bring potentially useful characteristics to the study of gene regulation. First, because of their relatively recent incorporation into the genome from exogenous genetic elements, the endogenous class of retroviruses exhibits a high degree of sequence conservation (see Chapter 4). The coordinate regulation of their expression may therefore be a direct consequence of shared sequence elements which can be defined. Any differences between individual expressed proviruses may be responsible for the distinct patterns of expression responsible for tissue specific abundances and be recognizatble by comparison in the same manner that tropism determinants of exogenous viruses have been defined. Alternatively, the regulation of these transcription units may reflect pre-existing regulatory networks established for cellular genes since little time has elapsed for the evolution of virus specific controls.

Second, endogenous viruses are dispersed throughout the genome, as are presumably the expressed ones as well (see Chapter 5). Therefore, any mechansims controlling their collective expression must necessarily be acting in trans and therefore are distinct from cis promoter elements defined for exogenous viruses. As discussed above, this dispersed organization may also result in the individual expression of different endogenous viruses being due to their chromosomal location, and thus they can be considered probes for pre-existing regulatory environments in the genome.

Third, because endogenous viruses probably do not encode products necessary for host viability, the regulation of their expression is probably simpler than that of other cellular genes and therefore more amenable to study. It is this last characteristic, as well, which has allowed the

derivation of virus antigen negative strains, thus providing genetically defined regulatory networks for study.

Background

The genomes of endogenous retroviruses are found in multiple DNA copies in the chromosomes of many vertebrates where they are inherited vertically as Mendelian loci rather than being horizontally transmitted between animals as are infectious viruses. Because of their association with tumor incidence, the genetics and expression of endogenous retroviruses have been extensively studied (Gross, 1970). In mice, a number of distinct families of retroviruses have been identified. The murine leukemia viruses (MuLV) or type C infectious particles have close genomic counterparts as endogenous genomic sequences. Genetic evidence has identified some of these endogenous viral sequences as loci involved in the synthesis of infectious MuLV and in the induction of tumors. There are considerable strain specific differences in the number and chromosomal location of these viral loci as well as in the number, location and expression of closely related genes which do not give rise to tumor viruses. The type A defective viruses, type B mammary tumor viruses, and VL-30 virus related elements are other more distantly related families of endogenous retroviruses with patterns of expression indicating strain and tissue differences, arising at specific developmental stages and responding to host control mechansims (see Risser et al., 1983).

The expression of endogenous retrovirus related DNA sequences often does not result in production of infectious virus, and the structure of these genomic sequences is reminiscent of bacterial transposable elements and similar eukaryotic genomic elements (Temin, 1980). However, they are best understood in relation to the proviral stage of an infectious retrovirus life cycle from which they may originally have been derived (Figure 1). Introduction of infectious RNA molecules into a host cell is followed by reverse transcription into a DNA copy of the viral genome that is transported to the nucleus, where the double stranded, circular DNA molecule is integrated into the host cell genome. The resulting proviral structure is colinear with the viral RNA, flanked by long terminal repeats (LTR). The LTR U3 region in the 5' flanking region of the

provirus contains the promoter elements typical of cellular polymerase II (Pol II) transcription units, including a Goldberg/Hogness box (TATA), a CAAT-equivalent sequence, and often an enhancer element. The 5' capped and 3' polyadenylated primary RNA transcript is infectious and carries at its 3' end a copy of the 5' flanking promoter elements in order that no viral information is lost from the subsequent cycle of infection. The promoter elements from the 3' U3 region become the 5' promoter in the LTR of the next provirus, presenting the experimentally useful situation of transcriptional promoters being present in messenger RNA transcripts.

The typical infectious cycle as depicted in Figure 1 results in the transcription by cellular Pol II of genomic RNA, of gag-pol mRNA, and of subgenomic env mRNA, each transcript polyadenylated at a site in the 3' LTR. The gag-pol mRNA transcripts encode the viral structural polypeptides and, by probable readthrough of an intervening stop codon, the polymerase and associated enzymatic functions of the virus. The env mRNA encodes the viral envelope spike proteins, gp70 and p15E, as a polyprotein precursor. Endogenous proviral sequences can be expressed in the same way. Indeed, the expression of some viral loci leads to the production of infectious virus. Other related genomic sequences, however, are expressed to produce only some of the viral antigens associated with infectious virions. Furthermore, the regulation of endogenous virus expression differs from the infectious counterpart. While a provirus from an infectious virus is constitutively transcribed for the lifetime of the host cell, endogenous retroviruses appear to be under controls typical for other cellular genes, producing specific developmental and tissue distribution patterns of virus or viral antigen expression.

The developmental pattern of expression of retrovirus related antigens in mice has led to speculations concerning possible roles for non-viral, cellular counterparts of these genes in the process of differentiation (Old and Stockert, 1977). However, endogenous proviruses and experimentally introduced proviruses corresponding to infectious virus can also come under typical host controls. The expression of these loci is influenced by the chromatin structure and methylation status of the region (Groudine et al., 1981a; Hoffmann et al., 1982) and by the chromosomal location of

integration (Jaenisch et al., 1981). Also, transcription of proviruses is influenced directly by genetically defined cellular elements (Horowitz and Risser, 1982) and by biochemically detected tissue factors (Sodroski et al., 1984). The analysis of expression from recombinant constructions incorporating LTR promoter elements from infectious exogenous viruses has elucidated determinants of tissue target, of host range, and of disease specificity residing most likely in the viral enhancer element. The thymotropism of thymic disease related MuLV has been found to correlate with the particular LTR of these viruses (DesGroseillers et al., 1983). Similarly, host specific activation of transcription from feline compared with murine LTRs has been reported (Even et al., 1983). The disease inducing difference between leukemogenic and non-leukemogenic strains of some MuLV has been localized to the LTR (Lenz and Haseltine, 1983) and the induction of thymic lymphoma rather than splenic erythroleukemia depends upon the tissue specificity of transcriptional induction from the LTR of Friend and Moloney strains of MuLV (Chatis et al., 1983). Recently, it has been shown that sequences in the LTR of a human retrovirus are specifically activated by host as well as viral trans acting factors (Sodroski et al., 1984). It is reasonable to expect that similar controls will be found that regulate the expression of endogenous retroviral sequences.

Immunogenetic studies on inbred mice have shown a high degree of polymorphism in the structures of expressed endogeneous viral antigens and in the tissue and developmental patterns of this expression. Some of this polymorphism is in determinants of host range which has led to the classification of MuLV and their endogenous counterparts into four distinct types. The ecotropic viruses (named because of their ability to infect only cells of murine origin) are present in 0-6 copies per haploid genome and are non-pathogenic though possibly still associated with the induction of disease. The ecotropic host range is determined by the sequence of the gp70 envelope protein that enables it to recognize a murine cell surface receptor not found on cells of other species (Rein and Schultz, 1984). Ecotropic viruses are subdivided by strain tropism, depending upon the interaction of determinants on gag antigens with a host gene product (Pincus et al., 1971). MuLV which infect only cells of non-mouse origin

are termed xenotropic, this restriction again due to the interaction of gp70 with a specific cellular receptor found only on non–murine cells (Rein and Schultz, 1984). Previously, data from genomic hybridization analyses had been interpreted to indicate that the bulk of endogenous retroviral sequences were of the xenotropic class (Chattopadhyay et al., 1974; Buckler et al., 1982). Recent data from genomic restriction enzyme analyses and from genomic cloning experiments have indicated that xenotropic proviruses are usually present in 0–1 copies per haploid genome, while the bulk of the 20–50 copies of endogenous murine retroviruses (termed here the endogenous class) are of a distinct class, related to xenotropic virus but more closely resembling recombinant mink cell focus (MCF) forming viruses (Chattopadhyay et al., 1982; Khan et al., 1982; Quint et al., 1984). MCF viruses are recombinant viruses between ecotropic virus and a cellular sequence, encoding a new gp70 glycoprotein (Elder et al., 1977; Rommelaere et al., 1978), resulting in a dualtropic host range due to a surface receptor found on cells of both mouse and non–mouse origin (Rein and Schultz, 1984). The endogenous viral sequences related to the substituted portion of MCF viruses have never been isolated as infectious virus, even though they comprise the majority and sometimes the only class of viral sequence in the genome of particular murine strains.

In strains of mice containing no ecotropic or infectious xenotropic viral sequences, the expression of antigens related to the endogenous class of viruses can be observed without interference from a background of expression of horizontally transmitted, crossreactive antigens. A classic example of virus related antigen expression from endogenous sequences in the absence of either exogenous viral infection or the production of infectious virions is the Gix system in strain 129 mice (Stockert et al., 1971). Mice of strain 129 do not contain endogenous sequences related to ecotropic virus (Chattopadhyay et al., 1974) and cannot be induced for the production of infectious xenotropic virus (Levy, 1978; Jongstra and Moroni, 1981). Gix antigen was originally detected by immunocytotoxic assays of strain 129 thymocytes using an antiserum raised in rats against a MuLV rat leukemia. Thymocytes from different mouse strains were assayed for the ability to absorb this cytotoxic activity and were classified as either

Gix^+ or Gix^-. Genetic analysis of the Mendelian inheritance of this normal thymocyte marker originally found it in the ninth linkage group of the mouse (now known to be chromosome 17), and hence the name Gix because of its relationship to a Gross virus-induced tumor (Stockert et al., 1971). Additional breeding experiments have found complex genetic control of Gix expression and have called into question the original IX linkage group assignment (Stockert et al., 1976; Boyse, 1977).

Gix antigen has been established by immunological and biochemical analyses to be a type specific determinant of retroviral gp70 (Tung et al., 1975a; McClintock et at., 1977; Elder et al., 1977). Gix antigen has also been found to be a determinant on some but not all ecotropic viruses even though when present as a thymocyte antigen it is associated with xenotropic related gp70 (Tung et al., 1978). Oligonucleotide mapping studies of Gix^+ and Gix^- ecotropic viruses showed that this Gix phenotype correlated with a pair of allelic oligonucleotides from the env region of the viruses. The sequence of these oligonucleotides indicated that they differed by a single base change leading to an amino acid difference of aspartic acid to asparagine in Gix^+ virus compared to Gix^- virus, respectively (Donis-Keller et al., 1980). Because this amino acid change fell in a glycosylation attachment site, the suggestion was made that the Gix antigenic difference of these viruses was due to the differential glycosylation of gp70 (Rosner et al., 1980). Furthermore, it was found that some Gix^- mouse strains still expressed reasonable quantities of gp70 related antigens on thymocytes, and these glycoproteins displayed a mobility on SDS gels characteristic of viral Gix^- gp70 (Tung et al., 1975b).

The usefulness of Gix antigen as a differentiation marker for a particular stage of normal thymocyte development prompted development of congeneic inbred strains differing only in the expression of Gix (Stockert et al., 1976). These strains, termed 129 Gix^+ and 129 Gix^-, were derived by selective breeding of 129 mice with the Gix^- strain C57BL/6 (Stockert et al., 1971) and repeated backcrossing with the 129 parental strain. Surprisingly, the resulting Gix^- strain failed to express gp70 glycoproteins on thymocytes rather than derived its Gix^- phenotype by expressing a Gix^-

variant of gp70 (Tung et al., 1975b). Furthermore, the lack of gp70 expression at other tissue sites, known normally to synthesize gp70 in 129 Gix^+ mice, was demonstrated by serological assay of spleen (Strand et al., 1974) and serum (Hara et al., 1981) and by immunofluorescence monitoring of numerous adult tissues (Lerner et al., 1976). These analyses used an antiserum which recognizes a wide variety of gp70 antigenic types and is not sensitive to glycosylation, and hence these results reflect the absence or the reduction, rather than the structural alteration, of gp70 in 129 Gix^- mice. Thus, the Gix^- phenotype of this congeneic mouse strain is clearly distinct from the viral case described above. The polymorphism of gp70 glycoproteins normally expressed in various tissues of 129 Gix^+ mice (McClintock et al., 1977; Elder et al., 1977) suggests that the expression of several members of this gene family is collectively lost in 129 Gix^- mice, perhaps due to the action of a regulatory gene. The differential genetic locus Gv-1, defining the difference between the 129 Gix^+ and 129 Gix^- strains (Boyse, 1977), is a good candidate for such a regulatory gene.

* * * *

The results that are presented in the following chapters describe the analysis of the expression of retroviral antigens in inbred mice. The summary of some initial experiments presented in Chapter 1 describes immunological and protein chemical analyses of retroviral expression in transformed cells. As an approach to studying cell surface molecules which may be pertinent to normal cell interactions and which have been postulated to be derived from an intricately regulated multigene family (Hood et al., 1977), tumor specific antigens on murine fibrosarcomas were examined. Individually unique and crossreactive tumor specific transplantation antigens (TSTA) were detected on two ultraviolet light (UV) induced fibrosarcomas of C3H mice by transplantation rejection. The serological response of mice hyperimmunized with these tumors was found to be directed against cell surface glycoproteins which crossreacted with MuLV envelope proteins. Two distinct env related proteins on the cell surface of two tumor lines were distinguished by two dimensional gel electrophoresis and peptide mapping. One of the two env glycoproteins on

the tumors was associated with an infectious retrovirus, apparently activated in the transformed cell. The second protein was not associated with virions and probably corresponds to normal or possibly to tumor specific expression of endogenous, non-ecotropic, retrovirus related antigens. No differences were detected by peptide mapping between related glycoproteins from separate tumor lines. Therefore, the individually unique transplantation antigens specific for each tumor line may not be of retroviral origin.

The control of expression of the murine antigen Gix and of other products of endogenous retroviruses, in strain 129 mice and in its congeneic partner strain 129 Gix^-, is an example of the expression of retroviral antigens in normal, non-transformed tissues. As described previously, the expression of these antigens, which display homology with retroviral proteins, illustrates the action of genetically defined elements controlling the appearance of developmentally expressed gene products. The present study provides a molecular description of this phenotype by defining the structural genes responding to this control. The results presented suggest that the expression of a set of independent transcription units is coordinately regulated in trans and that some of these genes are individually regulated in distinct patterns of tissue specific expression (Figure 2).

The structural genes responding in this pathway belong to the family of endogenous murine retroviruses. Chapter 2 describes the initial characterization of the Gix phenotype at the molecular level using heterologous viral probes. Evidence is presented indicating that the Gix phenotype is specified by a distinct regulatory gene, defined genetically, acting to control the levels of accumulation of specific mRNA species. The steady state levels of retrovirus related polyadenylated RNA from the tissues of Gix^+ and Gix^- mice were examined by blot hybridization using as probes DNA fragments from cloned murine leukemia viruses (MuLV). RNA potentially encoding viral antigens was reduced or absent in Gix^- mice, even though no differences in integrated viral genomes were detected between these congeneic strains by DNA blotting. Tissue specific patterns of accumulation of different sized RNA species were detected in

brain, epididymis, liver, spleen and thymus, and the majority of these distinct RNA species were found to be collectively regulated with the Gix phenotype. Measurements of RNA synthesis suggested a major role for transcriptional control in the regulation of some retroviral transcripts. These results demonstrated that the Gix phenotype in 129 mice, which had previously been described only in serological terms, was controlled by the differential accumulation of messenger RNA, although several other alternative explanations had been suggested. Furthermore, the involvement of retroviral proviruses in this phenotype, as opposed to potential cellular genes producing crossreacting antigens, was strongly suggested by this initial characterization. It remained to be established in more precise terms, however, the relationship of these gene products to those of retroviruses and to document the involvement of multiple structural genes or of differential RNA processing in the Gix phenotype.

The isolation and characterization of cDNA copies of these endogenous retroviral transcripts allowed a more complete description of this phenotype. These results, presented in Chapter 3, establish the involvement of multiple retroviral transcription units in the response to the controlling gene, Gv-1. Thus, the product of this regulatory gene is required to be acting in trans unless the responding structural genes display a clustered organization unusual for endogenous retroviruses. Use of these cloned endogenous probes in RNA blotting experiments demonstrated the structural relationships among the regulated transcripts and their relationship to virus. All virus related RNA transcripts contained U3 elements and either pol regions, env regions, or both, identifying them as probable viral products. High stringency hybridizations did not distinguish between different transcripts containing homologous sequences suggesting a high degree of sequence conservation in this family. Differences among these RNA species could be detected, though, by the use of S1 nuclease protection analyses due to deletions of different lengths and at different positions within their coding and 3' untranslated regions. These deletions may be responsible for RNA transcripts lacking env sequences and of heterogeneous size which were identified by RNA blotting as accumulating with tissue specific abundances. These S1 analyses confirmed that these individual transcripts were indeed

the products of distinct, separate transcription units.

The nucleotide sequences of endogenous virus cDNA clones, presented in Chapter 4, fully document the expression of distinct genes, the nature of the polymorphism, and the relationship of these sequences to virus. The heterogeneity exhibited between different clones consisted of point mutation differences resulting from 3% to 6% nucleotide divergence of the different RNA transcripts. Furthermore, deletion of env coding information in some of the clones reflect the anomalous size and structure of previously detected heterogeneous RNA species which accumulate with different abundances in different cell types. An open reading frame of 360 nt from the 3' end of pol, an open reading frame of 1932 nt potentially encoding a complete env polyprotein, and two distinct U3/R regions of approximately 710 nt are presented. Comparison of these sequences with the analogous regions of infectious viruses demonstrated that these endogenous transcripts were highly related to the substituted env and pol regions of leukemogenic mink cell focus inducing viruses. The 3' third of the gp70 coding region, the p15E coding region, and the U3/R structure, however, were distinct from any known infectious virus.

These data define a molecular picture of the Gix system. Clearly, it will be of interest to pinpoint precisely the level at which Gv-1 regulation is maintained and to identify the structure and isolate the product of this gene. Some speculations on possible approaches to a definition of this controlling mechanism and some preliminary experiments are outlined in Chapter 5.

FIGURE 1

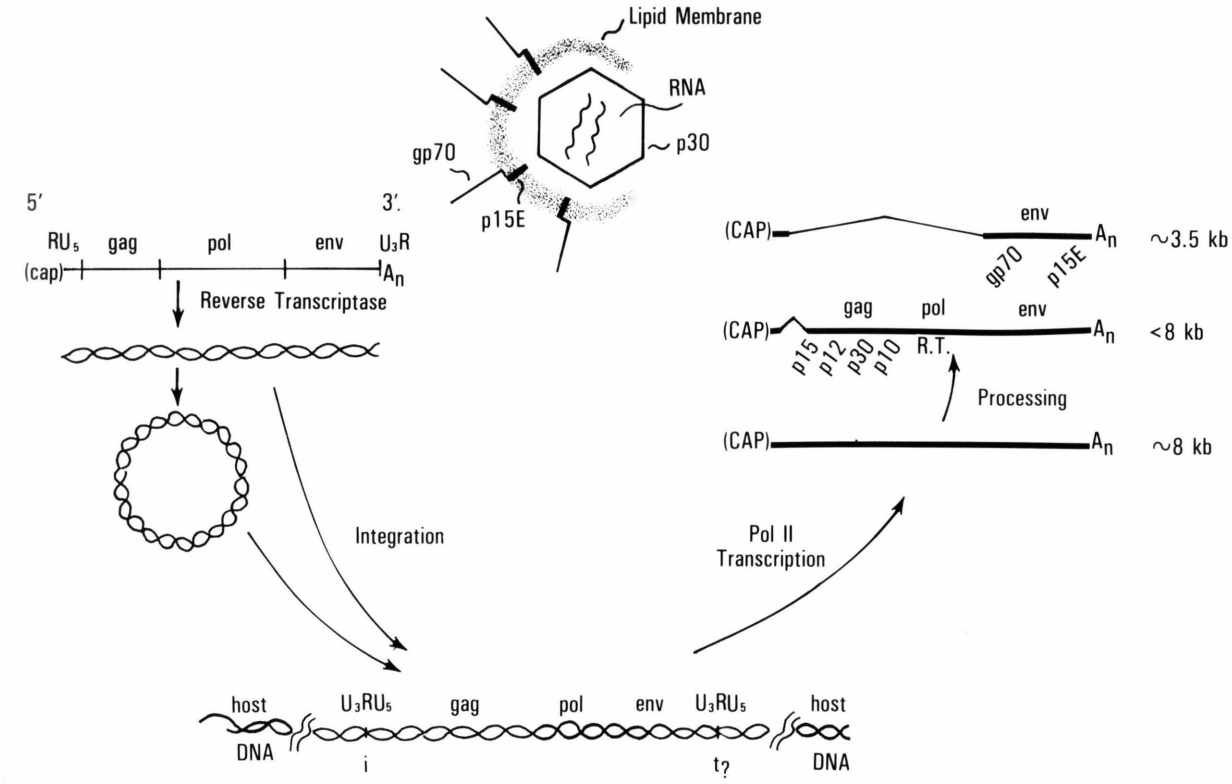

Lipid Membrane
RNA
p30
gp70
p15E
5'
3'
RU5 gag pol env U3R
(cap)
An
Reverse Transcriptase
Integration
(CAP)
env
An
~3.5 kb
gp70
p15E
(CAP)
gag pol env
An
<8 kb
p15 p12 p30 p10 R.T.
Processing
(CAP)
An
~8 kb
Pol II Transcription
host
U3RU5
gag
pol
env
U3RU5
host
DNA
i
t?
DNA

FIGURE 1

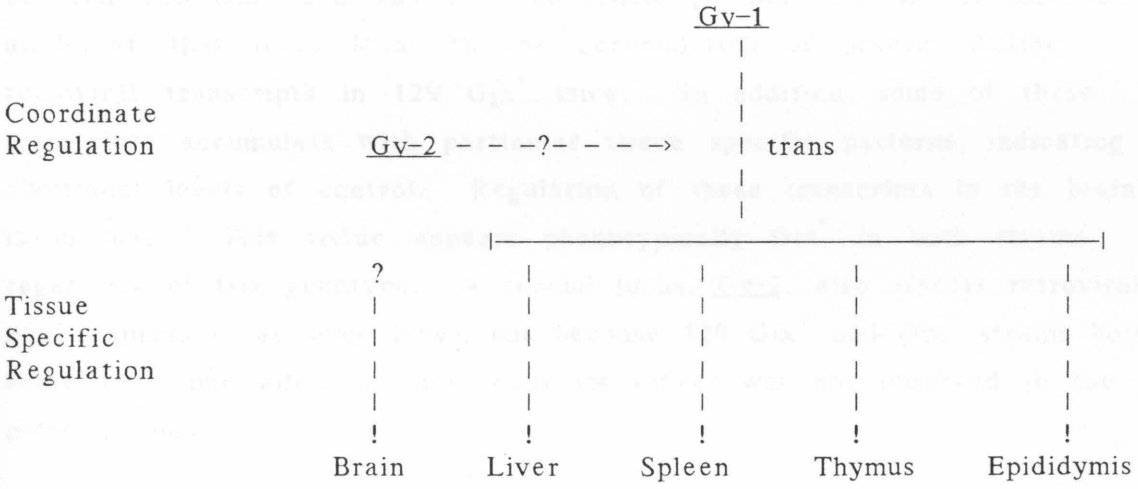
THE RETROVIRAL LIFE CYCLE

The typical retroviral life cycle is depicted in this cartoon, beginning with an intact virus. The retroviral virion contains a diploid RNA genome surrounded by viral structural proteins derived from the *gag* gene and a host cell derived lipid membrane containing the viral spike proteins gp70 and p15E encoded by the *env* gene. Infection of a permissive cell leads to release of the genomic RNA which is copied into double stranded DNA by viral reverse transcriptase. This DNA molecule or a circular intermediate is integrated into the host cell genome where it is transcribed by cellular Pol II. The initial transcript is a genome length, 5' capped and 3' polyadenylated RNA molecule from which are derived the *gag-pol* and *env* mRNA species. The *env* mRNA is produced by splicing to give a 3.5 kb transcript. The small splice depicted giving rise to the *gag-pol* mRNA is hypothetical (Shinnick et al., 1981).

FIGURE 2

Control of endogenous retroviral gene expression

in strain 129 mice



Differential abundances of several distinct transcripts

FIGURE 2

CONTROL OF ENDOGENOUS RETROVIRAL GENE EXPRESSION IN STRAIN 129 MICE

A model is depicted for the coordinate control of endogenous retroviral transcripts in 129 mice. Gv-1 is the single locus difference between 129 Gix^+ and 129 Gix^- congeneic partner strains. A positive allele at that locus leads to the accumulation of several distinct retroviral transcripts in 129 Gix^+ mice. In addition, some of these transcripts accumulate with particular tissue specific patterns, indicating additional levels of control. Regulation of these transcripts in the brain is unclear. This tissue appears phenotypically Gix^- in both strains regardless of Gix genotype. A second locus, Gv-2, also affects retroviral gene expression at some level, but because 129 Gix^+ and Gix^- strains both share the same allele at this locus its effect was not observed in the present study.

CHAPTER ONE

DETECTION AND SOLUBILIZATION OF UNIQUE AND CROSS REACTIVE TUMOR SPECIFIC TRANSPLANTATION ANTIGENS OF ULTRAVIOLET LIGHT INDUCED MURINE FIBROSARCOMAS

DETECTION AND SOLUBILIZATION OF UNIQUE AND CROSS REACTIVE TUMOR SPECIFIC TRANSPLANTATION ANTIGENS OF ULTRAVIOLET LIGHT INDUCED MURINE FIBROSARCOMAS

INTRODUCTION

It is well established that individually distinct tumor rejection antigens are associated with the neoplasms induced by chemical and physical agents in rodents (see Hellström and Brown, 1979 for review). Even when multiple tumors were derived by transformation of a cloned cell line, each malignancy displayed a unique antigenicity (Embleton and Heidelberger, 1972; Basombrio and Prehn, 1972). Though cross reactivities are readily detected by serological and in vitro immunological techniques, crossreactive tumor protection in a transplantation assay has proven to be a somewhat more rare occurrence (Basombrio, 1970). Some studies of transplantation rejection have demonstrated crossreactive specificities (Hellström et al., 1978; Leffell and Coggin, 1977), but no attempt has been made to characterize the molecules responsible for these activities.

As a means toward purifying tumor specific transplantation antigens (TSTA) from chemically induced tumors, transplantation rejection assays have been developed to demonstrate the induction of tumor protective immunity by the injection of solubilized tumor extracts (see Pellis and Kahan, 1976, for review). Such an assay is an essential step in purification since it provides an index for following the immunogenic acivity during biochemical fractionation. The present study demonstrated the presence of both individually unique and crossreactive TSTA on a pair of ultraviolet light (UV) induced fibrosarcomas by transplantation rejection. This assay has been adapted for solubilized tumor antigen and two tumor extraction methods have been shown to solubilize TSTA. The immunogenic activity was found in 3 M KCl extracts of tumor cells. It was also found to be present in 2.5% 1-butanol extracts which, unlike KCl extracts, have been shown to contain cell surface proteins essentially free from contamination by cytoplasmic proteins, nucleic acid and nucleoprotein, and by structural and integral membrane proteins (LeGrue et al., 1980).

Serological assays have also been applied to these tumor lines. By using hyperimmune syngeneic antitumor sera, cell surface proteins labeled with $Na^{125}I$ were immunoprecipitated from fibrosarcoma cell lysates. The proteins precipitated by these sera migrated on SDS polyacrylamide gels with apparent molecular weights of 70,000 to 90,000 and comigrated with proteins immunoprecipitated with heterologous antisera to murine leukemia virus (MuLV) envelope glycoproteins. These tumor antigens have been shown to be glycoproteins by their ability to bind lentil lectin affinity columns and to be specifically eluted with monosaccharides (C. Readhead, unpublished). Furthermore, they crossreact serologically with MuLV glycoproteins: goat antisera to Rauscher gp70 will preclear from cell lysates all proteins reactive with the antitumor sera.

Analyses of these tumor antigens by two dimensional (2D) gel electrophoresis revealed considerable charge and size heterogeneity. These 2D gel patterns differed for a series of six tumor lines, and these idiosyncratic patterns were stable and heritable properties of each tumor (Roman et al., 1981). Two of these tumors were tested for the presence of virus and found to be expressing an ecotropic MuLV which produced a gp70 glycoprotein showing a 2D gel pattern similar to that shown by some of the molecules immunoprecipitated from tumor cells. Other tumor antigens crossreactive with MuLV gp70 antisera did not appear to be associated with virions. Analysis of the gp70 related proteins from two tumors by 2D peptide mapping demonstrated two distinct types of cell surface molecules on each tumor. No differences were found in the structures of the homologous protein types from each tumor line which might explain the unique immunogenicities detected by tumor rejection.

RESULTS

In vivo studies

The presence of unique and crossreacting tumor transplantation specificities on two UV induced fibrosarcomas (tumor lines 861 and 2051) was demonstrated by a reciprocal immunization and challenge procedure (Table 1). Groups of mice immunized with a sublethal dose of viable tumor cells (10^4 cells per mouse) were rested 4 weeks to allow time for complete tumor rejection. At that time, mice without palpable tumors (>90% of inoculated mice) were sublethally X-irradiated (450 R) and challenged with either the immunizing tumor or with the alternative line. This radiation treatment diminishes the primary immune response of naive mice to a transplanted tumor, but does not affect the secondary response of previously immunized animals. Mice immunized and challenged with the same tumor line developed statistically fewer tumors than unimmunized controls. Furthermore, the tumors that did develop had a significantly smaller average size than those growing in the controls. Heterologously immunized groups also averaged fewer and smaller tumors than control groups, but there was a statistically significant difference between tumor growth in homologously rather than heterologously immunized mice. This distinction in the response to the two tumors is reciprocal. Of mice immunized and challenged with tumor line 861, 10% developed tumors, while 50% of mice immunized with line 2051 and challenged with tumor 861 had developed palpable tumors after 4 wk (Table 1). Likewise, all mice immunized and challenged with tumor 2051 rejected the implanted tumor, while only 60% of animals immunized with 861 could reject the 2051 tumor line challenge. This response is indicative of a common tumor transplantation antigen in combination with additional specificities unique to each tumor line.

In order to detect tumor antigens in a form compatible with protein fractionation techniques and to obviate the need for viable tumor cells in the induction of tumor protective immunity, tumors were solvent extracted and the resulting solubilized material tested for immunogenicity. The PBS soluble fraction of tumor proteins following 3 M KCl extraction was injected subcutaneously into groups of mice. A dose of tumor extract

equivalent to 10^6 tumor cells, injected 10 d prior to challenge, significantly retarded the outgrowth of tumors from an inoculation of 10^6 viable cells (Table 2). The tumors present 5 wk after animals were immunized with extracts from either tumor line were significantly smaller than tumors present in unimmunized hosts. This protection was effective against challenges from both tumors 861 and 2051. A response to the unique antigenic specificity of each tumor was not observed in the sizes of each tumor at 5 wk but could be detected by the increased latency period of tumors growing in hosts that had been immunized with the same tumor line relative to the kinetics of tumor growth in heterologously immunized mice (compare Table 3 with Table 2). When the tumor growth was observed in these animals at an earlier point in the response to a growing tumor (4 wk post challenge), there was a greater inhibition of growth after immunization with the homologous tumor than with the heterologous one. Mice immunized and challenged with line 861 or those immunized and challenged with line 2051 developed tumors that were significantly smaller at 4 wk than those of unimmunized controls, while the tumors in heterologously immunized animals were not significantly different in size from controls, as measured the the Student's t statistic. A response primarily to the tumor unique specificity was also observed by using an extended immunization protocol with KCl solubilized tumor extract involving multiple injections prior to challenge or by using a reduced, sublethal inoculation of tumor cells with the standard extract immunization procedure (data not shown).

Subcutaneous injection of mice with the PBS soluble fraction of an aqueous 2.5% butanol extract of tumor cells induced only specific immunity to the tumor from which the extract was obtained (Table 4). This extraction method routinely left greater than 90% of the cells viable, removing cell surface proteins without permanently damaging the cells. No statistically significant reduction in size or in growth rate of the crossreacting tumor was observed following immunization with butanol solubilized tumor antigen. This experiment indicates that tumor antigens capable of inducing crossprotective immunity are not solubilized in an immunogenic form by aqueous butanol extraction. This finding suggests that the individually specific TSTA are peripheral membrane molecules

while the antigens inducing crossprotective immunity may be integral membrane components.

In vitro analyses

Hyperimmune syngeneic antitumor sera have previously been shown to immunoprecipitate cell surface glycoproteins that crossreact with MuLV gp70 (Roman et al., 1981). Furthermore, these tumors were found to express an infectious MuLV with N-ecotropic host range characteristics. However, no infectious xenotropic or dualtropic virus nor any ecotropic viruses with pseudotyped xenotropic or dualtropic virion glycoproteins could be detected (M. Stone and R. Nowinski, personal communication). The expression of retroviral antigens on UV fibrosarcomas was further investigated by structural analysis. ^{125}I-labeled surface proteins were immunoprecipitated with a goat antiserum to Rauscher MuLV gp70. Heterologous antisera such as this have been found to recognize a wide variety of both endogenous and exogenous mammalian type C viruses (Gilden, 1975; Elder et al., 1977). Analysis of these immunoprecipitated glycoproteins by one-dimensional SDS polyacrylamide gel electrophoresis could not resolve different protein species with distinct molecular weights (not shown).

Analysis of immunoprecipitated tumor antigens by 2D gels, however, revealed that the diffuse patterns detected on SDS gels were due to size heterogeneity typical of glycoproteins. Figure 1 shows the 2D gel patterns for tumor antigens immunoprecipitated from tumors 861, 2051, and a series of other C3H UV fibrosarcomas. The charge heterogeneity is probably due to the extent of terminal sialic acid addition to the complex carbohydrate sidechains of these glycoproteins. Addition of terminal sialic acid residues gives the resulting glycoprotein a more acidic isoelectric point and a higher apparent molecular weight on SDS gels. This is reflected on the 2D gel autoradiogram by a shift to the right in the isoelectric focusing dimension and to higher molecular weight regions of the SDS electrophoresis dimension.

The 2D gel electrophoresis of immunoprecipitated tumor antigens resolved the glycoproteins from each tumor line into 1 to 3 distinct size

classes with considerable charge heterogeneity within each class. Each tumor showed a distinct pattern detected by 2D gel. For example, the 2D gel pattern for tumor 861 resolved two major and one minor protein bands in the 60,000 to 90,000 molecular weight range, while similar analysis of tumor 2051 revealed two size classes with molecular weights of 70,000 to 80,000 (Figure 2). The structural relationships between the glycoproteins from different tumors were further resolved by 2D peptide mapping. The regions of the 2D gels corresponding to the upper and middle bands identified on the autoradiogram from tumor 861 and to the upper and lower bands on the 2051 autoradiogram shown in Figure 2 were excised and digested with trypsin (see Experimental Procedures). The ^{125}I-labeled tryptic peptides were resolved by 2D electrophoresis/chromatography on thin layer cellulose acetate. Autoradiograms of these tryptic peptide maps are shown in Figure 3. The fingerprints of the lower molecular weight glycoproteins from tumor 861 and tumor 2051 show no significant differences from each other (Panels A and C). Likewise, the fingerprints of the large molecular weight proteins from these two tumors are also very similar (Panels B and D). However, the antigens from the two different size classes are clearly structurally distinct from each other. The ecotropic MuLV isolated from either tumor and propagated in NIH 3T3 cells showed a virion associated gp70 which gave a 2D gel profile similar to the approximately 70,000 molecular weight species from these tumors (not shown). Tryptic peptide maps of this virion gp70 appeared essentially identical to the peptide maps of the 70,000 molecular weight tumor antigens shown in Panels B and D of Figure 3 (not shown).

DISCUSSION

Two UV induced fibrosarcomas of C3H mice were tested with respect to their ability to induce transplantation immunity against themselves and against each other in syngeneic mice. Many UV induced tumors have been shown to be nontransplantable except when inoculated into immunosuppressed hosts (Kripke, 1974). At the cell doses used, the two tumors in this study were not transplantable into immunocompetent hosts. Though a higher cell dose would grow progressively in normal, syngeneic mice, an inoculation of 10^6 cells or less per mouse was consistently rejected (data not shown). However, the ability of the mice to reject these tumors is lost upon sublethal X-irradiation, thus allowing conventional transplantation rejection analyses to be carried out. Such irradiation inhibits the primary immune response to a transplanted tumor while leaving intact second set reactions to TSTA after immunization (Sjögren, 1965). Analyses such as these detect the individually unique TSTA characteristic of murine sarcomas (Prehn and Main, 1957). The finding of a weaker yet reproducible response to a crossreacting tumor specificity is similar to occasional reports of others (Hellström et al., 1978; Leffell and Coggin, 1977), though many others have reported finding only individually unique responses (see Hellström and Brown, 1979). Using a different procedure involving the specificity of UV induced suppressor T lymphocytes, Spellman and Daynes (1978) and Roberts et al. (1980) have detected crossreactive determinants on UV induced sarcomas in C3H mice. However, Kripke et al. (1979) have been unable to confirm these results in an independent study of similar tumors. The findings reported here clearly demonstrate a common tumor antigen using classical transplantation rejection assays. The reciprocal demonstration of cross protection as well as individually unique immunity, induced by immunization with live tumor cells or with KCl extracts of tumors, indicates the presence of two separate specificities rather than a quantitative difference in antigen expression. The selective solubilization of individually distinct tumor antigens by aqueous butanol extraction may indicate that the unique TSTA and the crossreactive antigens are present on distinct molecular species.

The further characterization of the TSTA of these UV induced tumors

requires the ability to detect a response to them in a form compatible with preparative isolation techniques. The extension of the transplantation rejection assay to allow immunization with solubilized tumor extracts will facilitate further fractionation of the antigen molecules (Meltzer et al., 1971; Natori et al., 1977; Sikora et al., 1979; Yamagishi et al., 1979). Both the KCl and the butanol solubilized extracts contained TSTA in immunogenic form and induced tumor protective immunity in treated mice. These extracts retained the unique specificity of the original tumors and contained sufficient antigen to engender protection against an inoculation of cells from the same tumor equal to the number used for extraction. The TSTA of these tumors are probably peripheral membrane proteins since they are easily solubilized by the aqueous butanol procedure, which leaves the cells viable and their membranes intact. There is no indication of a large pool of immunogenic TSTA in the cytosol as has been reported for another tumor system (DuBois et al., 1980) since the KCl solubilized extract, which contains membrane as well as internal cell components, was no more immunogenic in inducing individually specific tumor resistance than the exclusively membrane derived material.

The tumor rejection assay of solubilized TSTA will be central to the further characterization of antigen molecules. It will provide the means to follow antigen activity through a purification scheme as well as to address some of the pertinent questions involving the nature of these antigens. In particular, the role of viral antigens in the induction of transplantation immunity (Lennox et al., 1981) may be delineated by the direct isolation of gp70 from tumor extracts and subsequent evaluation of immunogenic activity in a tumor rejection assay.

The possibility that the crossreactive protection detected in the tumor rejection assays is directed toward carrier virus as has been suggested by Kripke et al. (1979) cannot be eliminated since tumors 861 and 2051 both express ecotropic retrovirus (M. Stone and R. Nowinski, personal communication) and have demonstrable cell surface viral gp70 (Roman et al., 1981). These UV tumors were found to express a heterogeneous set of retrovirus related antigens as resolved by 2D gel electrophoresis. Considerable heterogeneity is found, probably due to

differential glycosylation of the protein cores. Also, two distinct classes of protein molecules, represented on both tumors, were detected by peptide mapping. 2D gel electrophoresis and peptide mapping of the gp70 glycoprotein isolated from the ecotropic retrovirus expressed by these tumors indicated that the higher molecular weight antigen of tumors 861 and 2051 was contributed by this virus (not shown). The lower molecular weight antigen not associated with virions may be the result of the normal expression or of the transformation induced expression of gp70 from an endogenous retrovirus related sequence. The involvement of viral antigens in the individually specific immune response is unlikely, however, since no structural differences were detected by peptide maps of the analogous viral proteins isolated from each tumor. The TSTA of some tumors have been found to be equivalent to the serologically defined antigens (DuBois et al., 1981), and in other systems viral antigens have been implicated in the TSTA response (Lennox et al., 1981). However, the humoral and cell mediated immune responses may be directed against different sets of molecules in the system described here. The serological response is dominated by viral antigens. The antigens which mediate tumor rejection in vivo, however, particularly those molecules contributing the individually unique determinants, are not necessarily of viral origin. The distinct 2D gel patterns of the virus related antigens from each tumor may be due to consistant glycosylation differences between tumor lines. These differences or other distinctions not represented in the peptide maps may be involved in the immunigenicity of the tumors.

EXPERIMENTAL PROCEDURES

Mouse strains

C3Hf/HeN mice were obtained from Charles River (Mass.) and propagated in the Caltech vivarium by brother-sister mating. Genetic homozygosity was checked by the ability to accept intrastrain skin grafts. Mice were approximately 6 wk old at the beginning of an experiment and were randomized prior to isolating each treatment group in a separate cage. Any one experiment used mice of only one sex with initially 10 individuals per treatment group.

Tumors

Fibrosarcomas 861, 2051, 2009, 97, 1024, and 1053 were originally induced in C3Hf/Sm mice by exposure to high intensity (3.5×10^5 ergs/cm^2/sec) ultraviolet radiation (Kripke, 1974). They were maintained by subcutaneous passage from frozen stocks into the dorsal side of shaved mice which had been previously exposed to low dose ultraviolet light to render them tumor susceptible (Daynes et al., 1977). Tumors of transplant generations 5-8 were used for all in vivo experiments. Single cell suspensions of tumors were prepared by a modification of the procedure of Woodward and Daynes (1978). In brief, tumor bearing mice were killed by cervical dislocation and their subcutaneous tumors were removed into cold Dulbecco's modified Eagle's medium (GIBCO) buffered with 10 mM HEPES (Sigma). Necrotic tissue was carefully removed and the remainder of the tumor was minced with scissors into approximately 1 mm^3 fragments. These fragments were added to medium containing 150 u/ml of collagenase (Sigma) and 0.8 u/ml of Dispase (Boehringer Mannheim). Approximately 500 ml of enzyme mixture were used per 10 g of wet tissue. The tumor tissue was warmed to 37°C and stirred continuously for 1.5 h at which time the remaining fragments were removed by filtration through Nitex cloth (Tetko). Cells were washed 3 times with medium by low speed centrifugation, and the number of viable cells collected was estimated from the erythrosin B excluding cells counted in a hemocytometer.

Tumor extracts

Hypertonic salt extracts of tumor cells were prepared by the method

of Reisfeld and Kahan (1970). A single cell suspension of tumor cells was pelleted from medium by centrifugation and resuspended in cold 3 M KCl, using approximately 10 ml per 2.5 x 10^8 cells, and incubated at $4^{\circ}C$ for 16 h. The extract was subsequently centrifuged at 100,000 x g for 60 min, dialyzed 3 times against a 100 fold excess of phosphate buffered saline conatining 0.01% phenylmethylsulfonylfluoride (PBS) at $4^{\circ}C$, and finally centrifuged at 100,000 x g for 60 min.

Single phase butanol extracts were prepared by the method of Le Grue et al. (1980). Tumor cells were washed 3 times in Dulbecco's phosphate buffered saline (DPBS), resuspended in 2.5% 1-butanol (Mallinckrodt) in DPBS at an approximate concentration of 2.5 x 10^7 cells per ml, and incubated for 5 min at room temperature with gentle agitation. The cells were then pelleted at 400 x g for 5 min and the resulting supernatant was clarified by centrifugation at 100,000 x g for 60 min. Butanol was removed by dialysis 3 times against a 100 fold excess of PBS at $4^{\circ}C$.

Protein concentrations of the tumor extracts were determined by the Bio-Rad dye binding assay. Extracts were stored at $-70^{\circ}C$ until used and were frozen and thawed one time only.

Tumor rejection assay

An in vivo assay for TSTA was performed by a modification of the method of Kripke (1974). Groups of 10 C3H mice were immunized by subcutaneous injection of viable tumor cells or tumor cell extracts 4 wk or 10 d prior to challenge, respectively. Challenges with the immunizing tumor or with the alternative tumor line were made by subcutaneous injection of a lethal dose of tumor cells in Dulbecco's medium 1 d following sublethal whole body X-irradiation (450 R) of the immunized mice. Unimmunized, irradiated mice were included in each experiment as controls. Tumor growth was scored weekly beginning 2 wk after challenge. Mice with palpable tumors were considered positive and tumor diameters were determined by the average of two perpendicular measurements made with vernier calipers by a person ignorant of the treatments received by the mice.

Statistics

The mean tumor diameter + the standard error of the mean (SEM) was calculated for each treatment group at a time point when the maximum number of tumors had developed in control groups (3–5 wk post challenge). The statistical significance of the differences between means was assessed on the basis of Student's t statistic. Tumor incidence per group (the number of mice bearing tumors divided by the number inoculated) was compared between groups by using Fisher's exact probability test (Siegel, 1956).

Antisera

Syngeneic antitumor sera were obtained after hyperimmunization with tumor fragments implanted into normal mice at weekly to monthly intervals over periods of 6 mo to 1 yr. Anti-Rauscher leukemia virus gp70 antiserum made in goats was obtained from the Division of Cancer, Cause, and Prevention of the National Cancer Institute.

In vitro analysis

In vitro cultured fibrosarcomas were collected in DPBS and 2–5 x 10^6 cells were labeled with 5 mCi $Na^{125}I$ (Amersham) by lactoperoxidase (Hubbard and Cohn, 1972) or chololglycoluril (Markwell and Fox, 1978) iodination methods. Labeled cells were washed 4 times in PBS, lysed in 0.5% Nonidet P–40 (NP40), centrifuged at 100,000 x g, and the solubilized supernatant was stored at $-70^{O}C$.

Immunoprecipitation was preceded by absorbing labeled tumor cell lysates with fixed Staphylococcus aureus organisms (Pansorbin, Calbiochem-Behring). Lysates were subsequently incubated with 100 ul of antitumor sera or 10 ul of antiviral sera for 1 hr at $4^{O}C$, incubated with Pansorbin for 1 hr at $4^{O}C$, centrifuged, and the resulting bacterial pellets were washed 3 times in Tris buffered saline containing 0.3% NP40. Precipitates were removed from Pansorbin pellets by incubating with a solution of 9.5 M urea, 2% NP40, and 75 mM dithiothreitol (DTT).

Immunoprecipitated products were analyzed by one- and two-

dimensional polyacrylamide gel electrophoresis as previously described (O'Farrell, 1975). For peptide fingerprinting, labeled material was sliced from fixed and dried 2D gels. Dried gel slices were heated to 65°C overnight in a vacuum dessicator to insure removal of acetic acid and methanol from the gel fixative, minced into small pieces with razor blades, and incubated at 37°C in 50 mM NH_4HCO_3 containing TPCK-trypsin (205 u/mg, Worthington) at 50 ug/ml for 2 hr. Labeled peptides were eluted with 4 washes of 200 ul each of distilled water, concentrated by lyophilization, and resuspended in electrophoresis buffer (acetic acid: formic acid:H_2O in the ratio 15:5:8, v/v). Approximately 5000 cpm for each sample were applied to cellulose acetate thin layer plates (Eastman Kodak) and electrophoresed in electrophoresis buffer at 550 V for 45 min. Dried plates were subsequently chromatographed with a vertically ascending solvent of H_2O:butanol:pyridine:acetic acid in the ratio 24:30:20:6 for approximately 4 h. Plates were then air dried and autoradiographed with presensitized Kodak RP film with intensifying screens at -70°C for 1-4 wk.

TABLE 1

IDENTIFICATION OF BOTH UNIQUE AND COMMON TSTA
IN UV INDUCED FIBROSARCOMAS

Immunization[1]	Challenge	Tumor size[2]	p — Unique[3]	p — Common[4]	Tumor incidence	p — Unique[5]	p — Common[6]
861	861	1.5±0.0	<0.005		1/10	0.003	
				<0.01			0.07
2051	861	2.6±1.5	<0.005		5/10	0.18	
2051	2051	0.0±0.0	<0.01		0/10	0.003	
				<0.005			0.04
861	2051	2.4±0.5	<0.025		4/10	0.24	
None	861	7.1±3.1			8/10		
None	2051	8.3±4.6			6/9		

[1]Injection of 10^4 viable tumor cells subcutaneously 4 wk before challenge.—[2]Average tumor diameter (mm)±SEM from all mice bearing progressive tumors at 3 wk after challenge.—[3]Statistical significance of the difference in average tumor size compared to untreated control calculated from Student's t distribution.—[4]Statistical significance of the difference in average tumor size of homologously compared to heterologously immunized mice calculated from Student's t distribution.—[5]Statistical significance of the difference in tumor frequency compared to untreated control calculated from Fisher's exact probability test.—[6]Statistical significance of the difference in tumor frequency of homologously compared to heterologously immunized mice calculated from Fisher's exact probability test.

TABLE 2

DEMONSTRATION OF TUMOR PROTECTIVE IMMUNITY INDUCED WITH SOLUBILIZED TUMOR EXTRACT

Immunization[1]	Challenge	Mean tumor diameter ± SEM (mm)[2]	Frac.[3]	p[4]
861	861	3.6±2.5	9/10	‹0.005
2051	861	5.8±3.8	8/9	‹0.05
2051	2051	5.2±4.4	3/10	‹0.1
861	2051	3.5±1.0	3/5	‹0.01
None	861	11.1±5.6	6/6	
None	2051	9.1±2.8	6/8	

[1]Subcutaneous injection of 3 M KCl-solubilized proteins from 10^6 tumor cells 10 d before challenge. — [2]Measured at 5 wk after challenge. — [4]Fraction of challenged animals with palpable tumors at 5 wk. — [3]Statistical significance of the difference in mean tumor diameter compared to unimmunized controls calculated from Student's t distribution. Each treatment group initially contained 10 animals.

TABLE 3

INDUCTION OF UNIQUE TUMOR IMMUNITY WITH
3 M KCl SOLUBILIZED CELLULAR PROTEINS

Immunization[1]	Challenge	Mean tumor diameter ± SEM (mm)[2]	Frac.[3]	p[4]
861	861	3.3±2.3	10/10	<0.01
2051	861	5.3±2.8	8/9	n.s.
2051	2051	4.6±3.4	4/10	<0.05
861	2051	7.5±2.3	3/5	n.s.
None	861	6.9±2.9	6/6	
None	2051	8.8±3.0	6/8	

[1]Subcutaneous injection of 3 M KCl solubilized proteins from 10^6 tumor cells 10 d before challenge. — [2]Measured 4 wk after challenge. — [3]Fraction of challenged animals with palpable tumors at 4 wk. — [4]Statistical significance of the difference in mean tumor diameter compared to unimmunized controls calculated from Student's t distribution; n.s.= not significant. Each treatment group initially contained 10 animals.

TABLE 4

INDUCTION OF UNIQUE TUMOR IMMUNITY WITH BUTANOL SOLUBILIZED CELL SURFACE PROTEINS

Immunization[1]	Challenge	Mean tumor diameter ± SEM (mm)	Frac.[2]	p[3]
861	861	5.4±3.9[4]	7/10	<0.05
861	2051	7.2±1.0	3/7	n.s.
None	861	11.1±5.6	6/6	
None	2051	9.1±2.8	6/8	
2051	2051	2.9±2.4[5]	5/8	<0.01
2051	861	4.8±2.2	9/9	n.s.
None	2051	9.6±4.8	7/8	
None	861	5.1±1.8	8/8	

[1]Subcutaneous injection of 2.5% 1-butanol-solubilized proteins from 10^6 tumor cells 10 d before challenge. — [2]Fraction of challenged animals with palpable tumors at the time tumor sizes were measured. — [3]Statistical significance of the difference in mean tumor diameter compared to unimmunized controls calculated from Student's t distribution; n.s. = not significant. — [4]Measured 5 wk after challenge. — [5]Measured 4 wk after challenge. Each treatment group initially contained 10 animals.

FIGURE 1

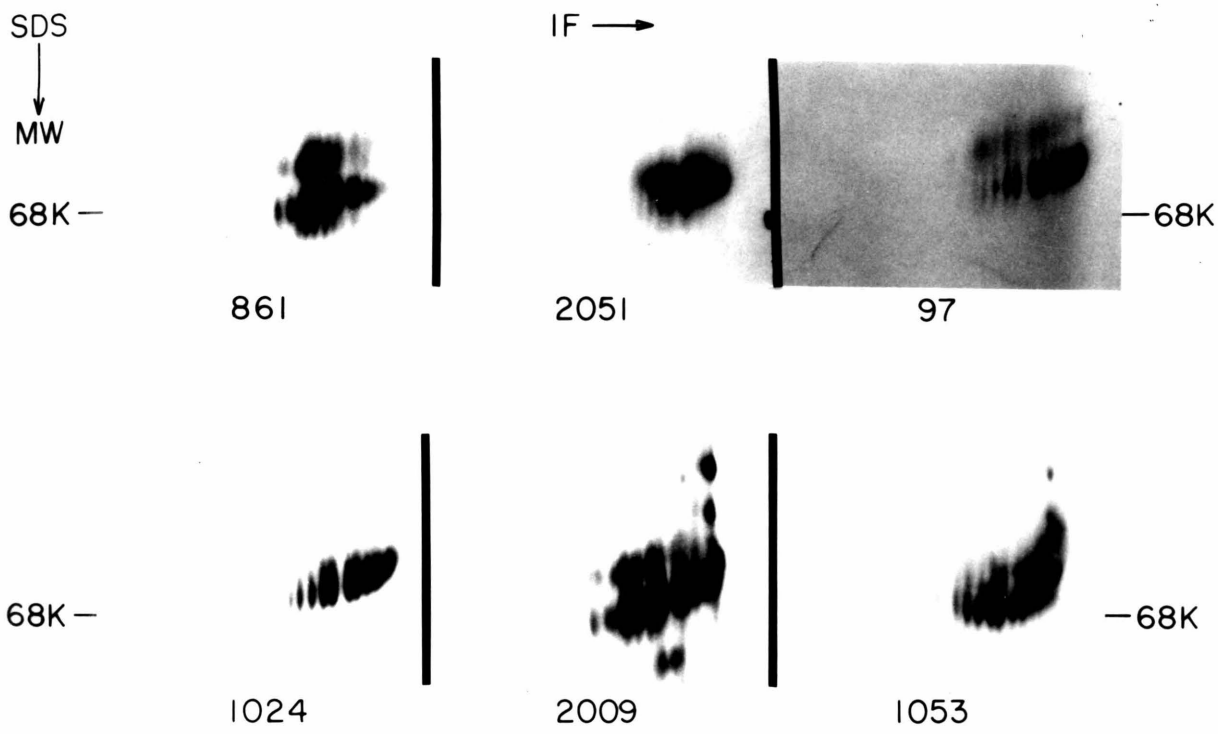


FIGURE 1

2D GEL ELECTROPHORESIS ANALYSIS OF TUMOR RETROVIRAL ANTIGENS

Cell surface, ^{125}I-labeled proteins were immunoprecipitated from tumor cell lysates of six different UV fibrosarcomas of C3H mice by using goat antisera directed against purified Rauscher gp70 protein. Antigen/antibody complexes were collected on fixed S. aureus cells, eluted with urea, NP-40, and SDS and analyzed by 2D gel electrophoresis using a 7.5% acrylamide second dimension. Gels were fixed, stained, and dried, and subsequently exposed presensitized Kodak XR film with an intensifying film. Molecular weights were estimated by co-electrophoresis of standard proteins of known size. Only the 68,000 molecular weight region of the resulting autoradiograms is shown. The number of the tumor cell line is indicated beneath each panel.

FIGURE 2

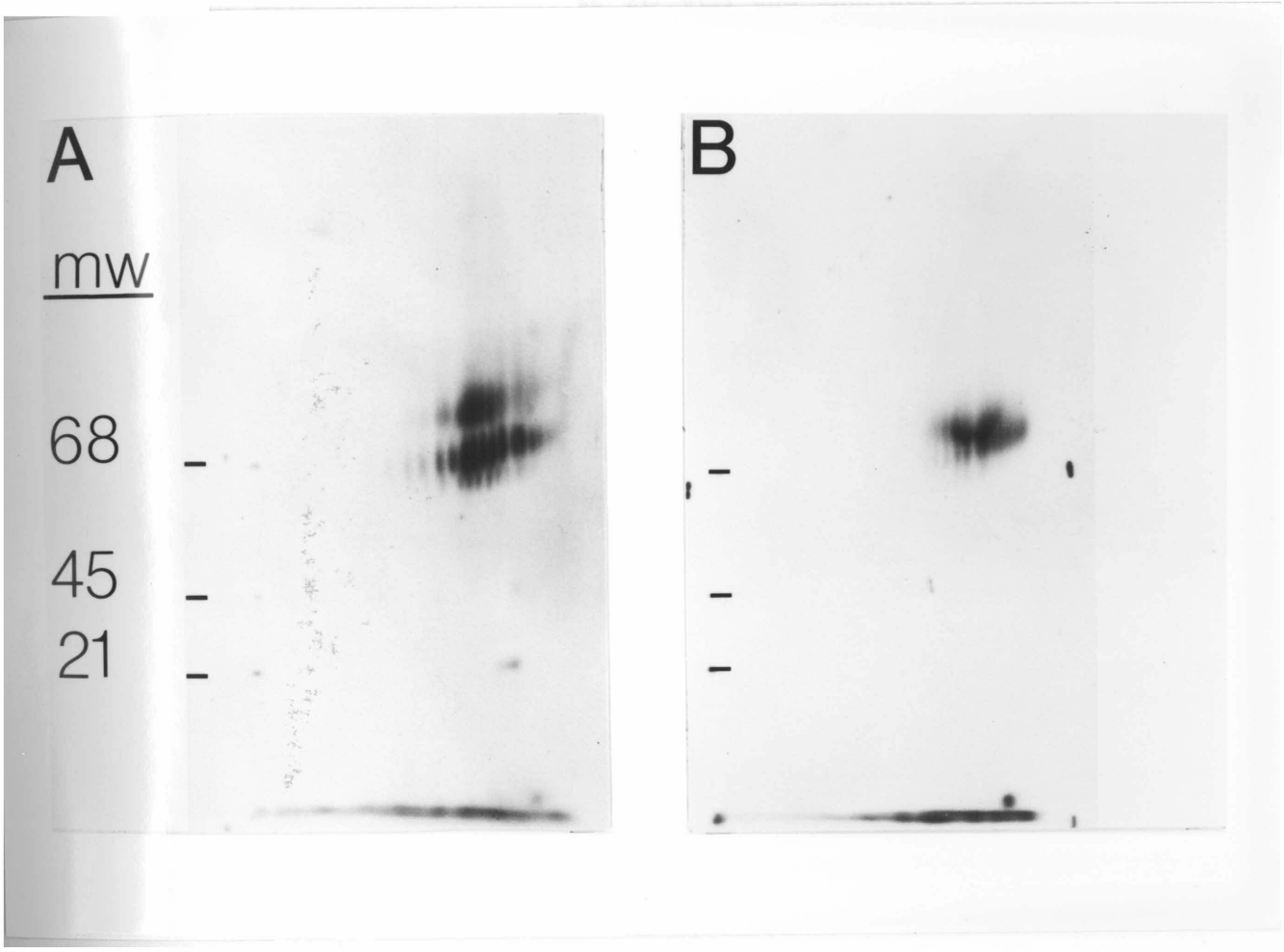


FIGURE 2

PREPARATIVE 2D GEL ELECTROPHORESIS OF TUMOR ANTIGENS

The complete autoradiograms of 2D gels of retroviral antigens from tumor lines 861 and 2051 analyzed in Figure 1 are shown. The upper and middle bands of line 861 (A) and the upper and lower bands from line 2051 (B) were excised from these 2D gels and analyzed by tryptic peptide fingerprinting (Figure 3). The two bands in panel A are easily discerned; the autoradiogram of line 2051 was interpreted as a more intense, slightly "U" shaped upper band and a less intense, slanted lower band. The full autoradiogram indicates that no other detectable proteins were precipitated by this antiserum.

FIGURE 3

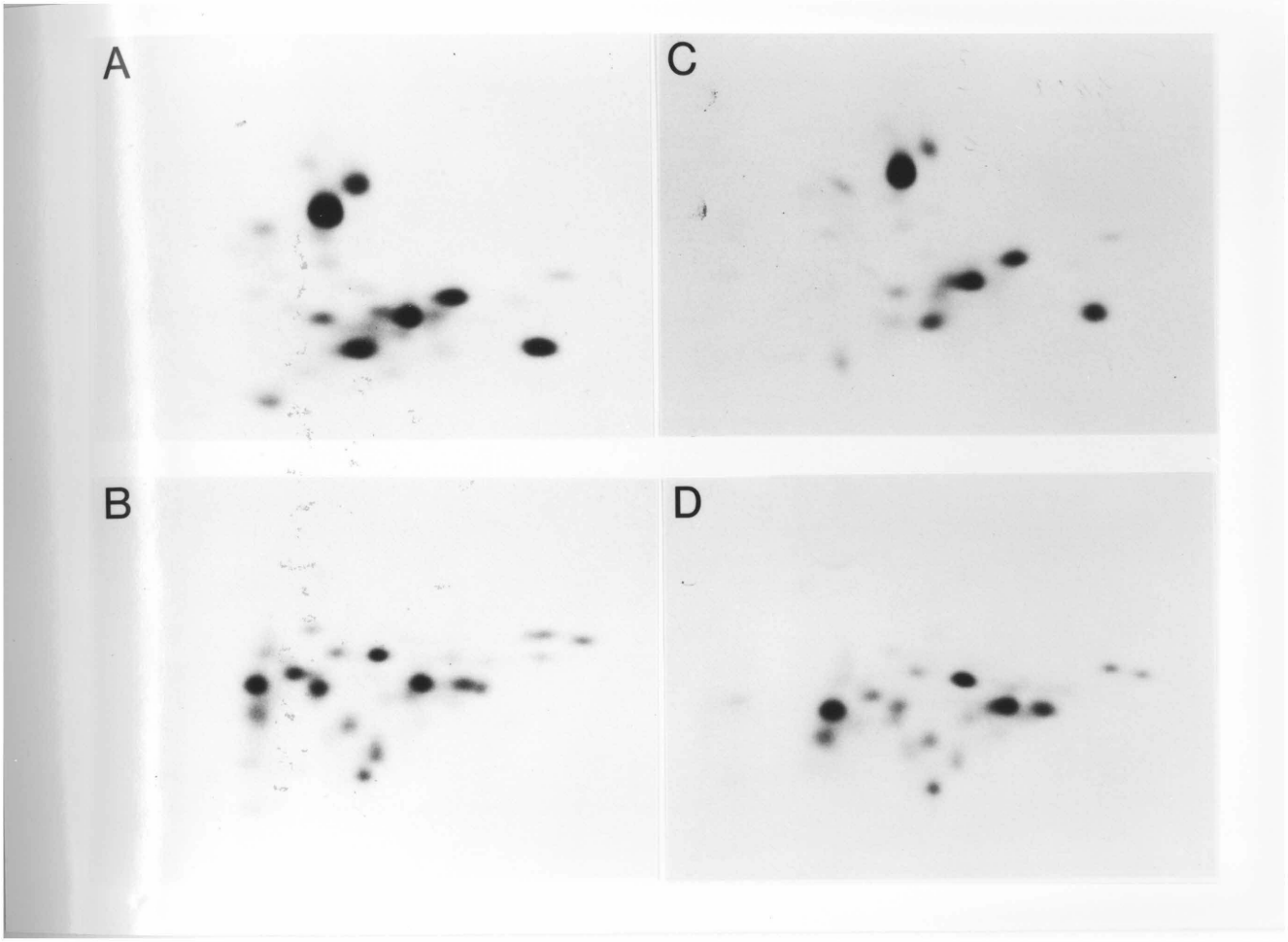


FIGURE 3

2D TRYPTIC PEPTIDE MAPS OF TUMOR ANTIGENS

The ^{125}I-labeled proteins isolated by immunoprecipitation and 2D gel electrophoresis (Figure 2) were digested with trypsin, eluted from the gel slices, and analyzed by electrophoresis and chromatography on thin layer cellulose acetate (see Experimental Procedures). The thin layer plates were air dried and exposed to presensitized Kodak RP film with intensifying screens at -70°C for 1 month. Electrophoresis occurred along the horozontal dimension from left to right; chromotography front developed from top to bottom. (A) 861 middle band. (B) 861 upper band. (C) 2051 lower band. (D) 2051 upper band.

CHAPTER TWO

ANALYSIS OF RETROVIRUS RELATED TRANSCRIPTS USING HETEROLOGOUS VIRAL PROBES

ANALYSIS OF RETROVIRUS RELATED TRANSCRIPTS USING HETEROLOGOUS VIRAL PROBES

INTRODUCTION

Genomes of laboratory mice contain multiple copies of DNA sequences homologous with murine leukemia virus (MuLV). The expression of these viral genes is subject to undefined host controls resulting in strain specific patterns of viral production (Rowe, 1978). Unlike many strains, the 129 mouse does not produce infectious virions from endogenous information (Levy, 1978). Only the major envelope glycoprotein (gp70) and major core protein (p30) of MuLV have been detected in 129 tissues (Strand et al., 1974). The first demonstration of a retroviral gene product in these mice was the detection of Gix (Stockert et al., 1971) which was found to be a type specific antigenic determinant of gp70 (Obata et al., 1975; Tung et al., 1975a; Tung et al., 1975b). Classically, the Gix antigenic determinant is considered characteristic of the thymic stage of T cell differentiation in many murine strains (Stockert et al., 1971). Other gp70 proteins have been demonstrated in a variety of murine tissues of other differentiation pathways including spleen, liver, epididymis, serum, and bone marrow (Strand et al., 1974; Lerner et al., 1976; McClintock et al., 1977; Del Villano et al., 1975; Del Villano and Lerner, 1976) and in a number of non-viral tumors (Lennox et al., 1981; Roman et al., 1981). The Gix determinant is inherited as a Mendelian trait (Stockert et al., 1971; Stockert et al., 1976), a property which allowed the derivation of a congeneic strain which lacks this thymic determinant (Stockert et al., 1975).

The loss of the expressed retroviral genome of 129 mice (i.e., the structural genes for gp70 and p30) as has been suggested (Strand et al., 1974; Tung et al., 1982) could result in a Gix^- phenotype. However, several observations suggest instead that a trans acting positive regulatory gene which promotes gp70 expression in 129 Gix^+ mice has been replaced by a corresponding negative allele in the congeneic 129 Gix^- strain. First, genetic analysis has identified two unlinked genes (Gv-1 and Gv-2), one dominant and one co-dominant, that are required for Gix expression.

Only one of these (Gv-1) is absent from Gix$^-$ mice, the strains being identical at Gv-2 and at all other loci tested (Stockert et al., 1976). Second, the gp70 molecules found at different tissue sites of 129 Gix$^+$ but absent or reduced in 129 Gix$^-$ are serologically distinct (McClintock et al., 1977) and have distinct tryptic peptide fingerprints (Elder et al., 1977), suggesting that multiple structural genes are being affected. For example, another gp70 antigenic determinant (GERLD), which may represent a distinct gp70 molecule, is also lost in the 129 Gix$^-$ mouse (Obata et al., 1981). Third, although 129 Gix$^-$ mice are often described as completely lacking gp70, low but detectable quantities of this antigen have been reported (Lerner et al., 1976). Until recently, the analysis of this genetic regulation was limited to such serological and biochemical studies of the protein antigens. With the availability of cloned retroviral DNA probes, it has become possible to examine directly the genetic structure and events responsible for these traits. This chapter describes blot hybridization analyses which support the conclusion that 129 Gix$^+$ and Gix$^-$ mice contain the same complement of retroviral genes. Furthermore, the Gix phenotype correlates with the differential accumulation of specific mRNA molecules homologous with retroviral DNA, implicating a regulatory locus which coordinately controls the expression of a number of distinct retroviral transcripts. We propose a model in which the Gv-1^a allele of 129 Gix$^+$ mice encodes a product which acts in trans to promote the specific expression of endogenous proviruses.

RESULTS

Gix$^+$ and Gix$^-$ mice are congeneic for endogenous retroviral sequences

Although no infectious retroviral particles have been isolated from mouse strain 129, viral complementary DNA is detectable in 129 cellular DNA (Rowe, 1978; Chan et al., 1980; Buckler et al., 1982). To determine if the difference in the presence of viral proteins detected between strain 129 and its congeneic 129 Gix$^-$ partner strain represented a change in viral structural genes, the retrovirus homologous sequences in the chromosomal DNA of these strains were examined by DNA blot hybridization (Southern, 1975). Retroviral sequences were detected in the germline and the somatic cell DNA of both 129 Gix$^+$ and the congeneic Gix$^-$ mice. These sequences are present in high molecular weight DNA (Figure 2A) and therefore are unlikely to reflect the presence of unintegrated viral cDNA. Furthermore, no differences were detected between the germline and somatic DNA in the number or location of viral sequences (Figure 2B). This result indicates that no evident infection of the somatic tissue with exogenous viruses or rearrangements of endogenous sequences has taken place.

A minimal estimate of the number of separately integrated retroviruses can be derived from an Eco RI digestion pattern since cleavage at unique sites within flanking DNA generates a distinct DNA fragment for each integrated MuLV (Lowy et al., 1980; Chattopadhyay et al., 1980). However, a comparison of 129 Gix$^+$ and Gix$^-$ DNA homologous with MoMuLV (Figure 2B) reveals that the lack of viral antigens in Gix$^-$ mice is probably not reflected by a corresponding difference in viral structural genes. All Eco RI generated viral DNA fragments of antigen expressing 129 Gix$^+$ mice are also present in the DNA of the non-expressing 129 Gix$^-$ mice.

Retroviral RNA levels reflect Gix phenotype

The expression of endogenous retroviral genes was examined in the epididymis of Gix$^+$ and Gix$^-$ 129 mice since the male genital tract is the major site of accumulation of gp70 (Del Villano and Lerner, 1976). To

measure steady state levels of mRNA potentially coding for gp70, total poly(A) containing RNA was fractionated by molecular weight on a denaturing agarose gel and transferred to nitrocellulose. Hybridization with MoMuLV DNA (Figure 3A) demonstrated that the reduced expression of the Gix antigen in 129 Gix^- epididymis correlates with a reduced accumulation of retroviral RNA. Whereas the 8.0, 7.2, and 3.5 kilobase (kb) RNA species present in a viral infected cell (Lane 1) are also present in the normal 129 Gix^+ epididymis (Lane 2), a greatly reduced amount of these molecules is observed in the 129 Gix^- epididymis (Lane 3).

Hybridization of the Gix^+ and Gix^- epididymal RNA preparations with the gp70-coding region of MoMuLV revealed that the RNA species potentially encoding the gp70 protein is reduced in Gix^- mice (Figure 3B). During normal viral replication, a 21 S subgenomic RNA is translated into the envelope polyprotein precursor (Faller et al., 1978). This molecule probably corresponds to the 3.5 kb band observed with RNA from 60A cells and epididymis which contains Mo^{env} homologous sequences (Figure 3B) and sediments in a sucrose gradient at approximately 22 S (data not shown). In Gix^- RNA, the 3.5 kb species is virtually undetectable with a probe to the entire MoMuLV but can be visualized in low amounts with a probe specific for the _env_ gene (compare Figures 3A and 3B). The low but detectable amount of _env_ related RNA in Gix^- epididymis supports the conclusion that the Gix^- phenotype is not a result of the absence of the structural gene for gp70 but is some alternative genetic event which affects the cellular accumulation of viral RNA.

To further characterize the retroviral RNA species regulated by _Gv-1_, epididymal RNA from 129 Gix^+ mice was hybridized with probes specific for the _pol_ region of MoMuLV (not shown) and the _gag-pol_ coding region of cloned AKV (Figure 3C). The 8.0 kb and 7.2 kb bands which are present in reduced amounts in the Gix^- epididymis blot have sequences homologous to all the subregions represented in the viral probes and probably correspond to full length viral transcripts. The 3.5 kb band which is also reduced in the Gix^- blot is homologous only to the _env_ coding region of MoMuLV, specifically missing sequences homologous to the

gag-pol region. This is consistent with this RNA being the spliced envelope message analogous to the subgenomic message produced in a typical viral replicative cycle. The 6.0 kb band which is absent from the Gix^- blot and the 5.2 kb band which is present in Gix^- at approximately the same level as Gix^+ (compare to Figure 3A, Lanes 2 and 3) are both homologous only to the gag-pol region of the virus, specifically missing any potentially envelope coding information. No RNA transcripts with this structure are produced in the course of a normal productive viral infection (see Introduction).

Although the 129 mouse strain is viral free in the sense that no infectious retroviral particles can be detected, it is interesting to note that all the major high molecular weight viral RNA species of a productively infected cell are also present in normal tissues of these mice, including the 8.0 kb, presumably full length, viral RNA. It has been suggested that 129 endogenous viruses lack the capacity to express the polymerase function (Jongstra and Moroni, 1981). If this is true, the defect responsible for this lack of expression is too small to be detected by this analysis and does not involve a major deletion. The 7.2 kb RNA detected in both 129 mice and in infected cells also appears to contain all viral structural information. It has been suggested that viral genomic RNA and gag-pol mRNA exist in non-overlapping pools (Levin and Rosensak, 1976) and that the mRNA is created by a small splice at the 5' end. The 7.2 kb RNA detected here may represent this spliced mRNA, although its length is less than would have been predicted based upon the size of the proposed intervening sequence in MoMuLV (Shinnick et al, 1981). Alternatively, since such a presumptive intron has not been detected in viral RNA, this 7.2 kb RNA may be a deleted transcript which still retains sufficient sequence to hybridize all the probes tested in this analysis.

Multiple retroviral RNA species are collectively expressed

Serological analyses of a wide variety of tissues of the 129 Gix^+ mouse have detected gp70 molecules (Lerner et al., 1976) with distinct antigenic characteristics (McClintock et al., 1977) and tryptic peptide maps (Elder et al., 1977). The expression of viral RNA in several tissues of

129 mice was investigated by RNA blotting, revealing a tissue-specific pattern of RNA accumulation (Figure 4A). The polyadenylated RNA of whole murine brain contains barely detectable levels of endogenous retroviral RNA. Liver, spleen, thymus, and epididymis all share the 8.0 kb and 7.2 kb full-length transcripts and 3.5 kb envelope-specific subgenomic transcript. In addition, each tissue reveals specific qualitative and quantitative differences in endogenous viral transcripts. Liver has a doublet of RNA transcripts of length approximately 4.1 kb and greatly reduced amounts of the 5.2 kb RNA species. Thymus shows reduced amounts of the full-length transcripts but accumulates a large abundance of 5.0-5.2 kb molecules. Spleen accumulates primarily the 5.2 kb transcript. All these tissue specific RNA transcripts contain sequences homologous with the viral gag pol region but lack any detectable homology with env (data not shown).

A similar analysis of retroviral RNA in the tissues of 129 Gix^- mice revealed that all retroviral homologous transcripts in all tissues are reduced or absent in Gix^- mice with the exception of the 5.2 kb transcript in the epididymis (Figure 4B). Visualization of RNA species homologous with two constitutively transcribed nonviral genes by hybridization with probes to two housekeeping functions (Figure 4C) provides a measure of the mass of RNA derived from each tissue. Although these results cannot confirm the finding that **distinct** tissue specific gp70 proteins are regulated in concert, these tissue specific RNA species which may be transcribed from separate retroviral structural genes are each under the control of Gv-1.

Transcriptional regulation of retroviral RNA expression

Transcriptional activities of RNA sequences homologous with MoMuLV were measured in vitro by labeling nascent RNA transcripts in isolated nuclei and detecting MoMuLV related transcripts by hybridization in DNA excess. This method has been shown to provide an accurate reflection of transcriptional activities without the influence of differential processing rates and to yield an estimate similar to that obtained from pulse labeled cells in vivo (Evans et al., 1977; Weber et al., 1977; Derman et al., 1981). In these experiments, the nuclei of liver (a gp70 positive tissue in

the Gix^+ strain) and brain (gp70 negative) tissue were used to provide transcription rate measurements because they are relatively easy to prepare and are highly active in vitro. Similar results were obtained with Gix^+ and Gix^- spleen nuclei, but at much reduced levels of total UTP incorporation (not shown). The transcriptional activities of retroviral sequences in Gix^+ and Gix^- mice were compared by using a probe to promotor proximal sequences at the 5' end of the viral genome. The cDNA recombinant plasmid pCHO B was included as control to provide an internal standard of labeling and relative hybridization efficiency.

Comparison of the results obtained from liver nuclei from Gix^+ and Gix^- mice (Table 1) shows that the relative transcriptional activity for gag-pol sequences is reduced more than 5 fold in the Gix^- liver as compared with Gix^+, in agreement with the reduction in mRNA observed at steady state. Moreover, transcription of retroviral sequences in brain of either Gix^+ or Gix^- mice is at the level of Gix^- liver, consistent with the low levels of RNA and protein detected in these tissues. The transcriptional activity of sequences homologous with pCHO B remains constant regardless of the choice of tissue and strain of mouse, consistent with the results obtained previously (Derman et al., 1981; Soreq et al., 1980).

The nature of the residual transcriptional activity of retroviral sequences in tissues showing little accumulation of $poly(A)^+$ retroviral RNA was studied in Gix^- liver. The transcription of gag-pol related sequences was measured in Gix^- liver nuclei in the presence of alpha-amanitin, a potent inhibitor of RNA polymerase II (Lindell et al., 1970). Treatment of nuclei with this drug during labeling reduced the total incorporation of UTP to 68% of the incorporation into untreated nuclei (Table 2), consistent with a significant inhibition of RNA polymerase II activity. Transcription of gag-pol sequences was reduced by 77%, showing that the majority of this synthesis is directed by RNA polymerase II. In a parallel hybridization with pCHO B DNA, labeled RNA was detected reflecting a similar transcriptional inhibition by amanitin treatment. A similar analysis using a 3' retroviral probe for env related transcription was not possible because, in this hybridization, greater than 80% of

homologous RNA hybridizing the probe was not synthesized by Pol II as determined by sensitivity to amanitin inhibition (not shown).

DISCUSSION

The mechanisms controlling the expression of specific gene products, as illustrated by animal viral systems, can occur at many steps during mRNA biogenesis including at the level of transcription and post transcriptional processing (for review, see Darnell et al., 1980). However, the genetic loci and the gene products which dictate these mechanisms of eukaryotic gene regulation remain undefined. The murine strain 129 congeneic pair is an example of the control of a particular phenotype by a single genetic locus. The Gix phenotype in 129 mice is dependent upon the presence of a positive allele at Gv-1 (Stockert et al., 1971). Contrary to some suggestions (Strand et al., 1974; Tung et al., 1982), analysis of the genomic DNA of these mice indicates that Gv-1 is not a viral structural gene. No differences in the number or location of the numerous endogenous retroviral genomes were seen that would account for the Gix^- phenotype, nor were differences in the two strains observed in the structure of the polyadenylated RNA produced by these endogenous sequences. Instead, quantitative differences in the accumulation of different sized mRNA species were found to correlate with the Gix phenotype. Although models involving an altered genomic structure or missing provirus in 129 Gix^- mice or invoking translational or post translational defects could have been envisaged, the lack (or scarcity) of gp70 and p30 in Gix^- mice is most likely the result of a substantial reduction in the accumulation of multiple, endogenous retroviral mRNA transcripts of distinct primary structure.

The tissue specific polymorphism exhibited by the products of endogenous retroviral sequences suggests that more than one transcription unit is affected to produce the Gix phenotype. This polymorphism is detected in two manners. First, serological (McClintock et al., 1977) and structural (Elder et al., 1977) studies of gp70 proteins have demonstrated that distinct molecular species are deposited in different tissues. Although the analyses reported here cannot detect this polymorphism due to the heterologous nature of the probe and the not unexpected similarity in size of env mRNA transcripts, all potential gp70 coding mRNA species are reduced in all tissues of the Gix^- strain. Second, the tissue specific

distribution of distinct polyadenylated RNA species indicates further the activity of multiple gag-pol transcription units. The presence of multiple copies of endogenous retroviral DNA sequences integrated within the murine genome has been well documented (Rowe, 1978; Chan et al., 1980; Buckler et al., 1982) as seen in Figure 2. Analysis by restriction enzyme mapping suggests that many of these sequences may reflect partially deleted or otherwise altered proviruses. Tissue specific transcription of such distinct retroviral sequences would result in the observed heterogeneity of gag-pol polyadenylated RNA species. Transcriptional control by Gv-1 of these distinct, multiple retroviral genomes would result in the observed down regulation of all viral RNA. In this model, the epididymal specific 5.2 kb gag-pol RNA would result from an unusual transcription unit which is independent of the Gv-1 regulatory element and does not contain env sequences. An alternative interpretation is that the tissue specific pattern of gag-pol RNA is the result of tissue specific post transcriptional processing, either by choice of alternative splicing or polyadenylation sites, of transcripts originating from a single Gv-1 regulated transcription unit.

Nascent RNA labeling experiments in isolated nuclei suggest that a significant mechanism involved in the regulation of endogenous retroviral expression is at the level of transcription. In the liver of Gix^- mice as well as in the brain of either Gix^+ or Gix^- strains, the transcription of gag-pol sequences is significantly reduced in agreement with the lack of accumulated mature polyadenylated mRNA. Residual gag-pol RNA transcription in nuclei from Gix^- mice is probably also the result of RNA polymerase II activity and thus could result in stable mRNA. The presumably non-Pol II transcription of env related RNA detected in liver nuclei from Gix^- mice (not shown) is probably the result of RNA polymerase III activity in the in vitro transcription reaction. This transcription probably would not give rise to stable, capped and polyadenylated RNA nor encode protein products because these are believed to be exclusive features of RNA transcribed by Pol II (Chambon, 1975), the enzyme that synthesizes typical retroviral transcripts (Dinowitz, 1975; Fuhrman et al., 1981). It is possible that this transcription that is resistant to amanitin originates from retroviral sequences adjacent to one

or more active polymerase III transcription units. A similar finding has been observed for rat growth hormone transcription (R. Evans, personal communication) and possibly for the nuclear transcripts of some rat brain specific genes (Sutcliffe et al., 1984).

In summary, examination of nucleic acid sequences in 129 Gix^+ mice and in the congeneic 129 Gix^- strain has demonstrated a correlation between the Gix serological phenotype and the abundance of specific retroviral mRNA species. Rather than a change in the number or structure of chromosomal proviruses between Gix^+ and Gix^- mice, the Gix phenotype is probably due to a trans acting mechanism encoded by Gv-1 that regulates the transcription or the accumulation of specific RNA species.

EXPERIMENTAL PROCEDURES

Mouse strains and cell lines

129 Gix^+ and 129 Gix^- congeneic mouse stocks were obtained from the Congeneic Mouse Production Facility of Memorial Sloan-Kettering Cancer Center through the generosity of E. A. Boyce and maintained at the SCRF vivarium. A lymphoblastoid cell line chronically infected with a murine leukemia virus (SCRF 60A) was obtained from F. Jensen and maintained as a suspension culture in RPMI 1640 medium supplemented with 10% bovine serum.

Isolation of chromosomal DNA and restriction analysis

Liver, thymus and sperm were obtained from 4-6 wk 129 Gix^+ and 129 Gix^- male mice. Nuclei were isolated from thymus and liver by homogenization in a solution containing 17.1% sucrose, 15 mM Tris-HCl, pH 7.5, 0.5 mM spermidine, 0.15 mM spermine, 1 mM EDTA, 1 mM EGTA, 25 mM KCl, and 0.3% Triton X-100 (van der Putten et al., 1982). DNA was extracted from nuclei or sperm after treatment with Proteinase K and SDS at 100 ug/ml and 2%, respectively, at 56^oC for 2 h, and extracted with phenol and chloroform. DNA was recovered by spooling from 70% ethanol, resuspended in TE (10 mM Tris-HCl, pH 8; 0.1 mM EDTA) and precipitated from 0.62 M NaCl and 5.5% PEG-8000 after incubating at 4^oC for 15 min. DNA was analyzed by the method of Southern (1975) and hybridized with ^{32}P-labeled nick translated probes (Rigby et al., 1977) at a specific activity of 10^8 to 10^9 cpm/ug.

RNA isolation and blot hybridization

Mouse tissues were obtained from 5 to 6 wk old males, frozen immediately in liquid nitrogen, and stored at -70^oC until use. Frozen tissues were pulverized in a stainless steel press at -70^oC, blended with Proteinase K (250 ug/ml) and 1% SDS in isotonic buffer, and the total nucleic acid was extracted with phenol/chloroform and precipitated with ethanol as previously described (Levy et al., 1982). Cytoplasmic RNA was extracted from tissue culture cell lines after lysis in 0.25% NP40 in isotonic buffer (Wilson and Darnell, 1981). $Poly(A)^+$ RNA was isolated by oligo(dT) cellulose affinity chromatography (Aviv and Leder, 1972),

fractionated on denaturing formaldehyde agarose gels (Rave et al., 1979), transferred to nitrocellulose (Thomas, 1980) and hybridized with nick translated probes in 50% formamide and 0.75 M NaCl at 42°C as previously described (Nevins and Wilson, 1981).

Recombinant DNA clones

A recombinant DNA clone of Moloney leukemia virus in the plasmid pBR322 (pMLV-1) was obtained from I. Verma (Berns et al., 1980). Subgenomic fragments and subclones were prepared using standard procedures utilizing a restriction map deduced from the complete nucleotide sequence (Shinnick et al., 1981) and are illustrated in Figure 1. Clone pAKV$^{gag-pol}$ of the gag-pol region of the endogenous ecotropic virus, AKV-1, of AKR mice was kindly provided by D. Lowy. Probes for constitutively synthesized mRNA transcripts (pCHO B and pCHO C) were obtained through the generosity of M. Harpold (Harpold et al., 1979). The recombinant plasmid DNA was purified essentially as described by Clewell and Helinski (1970).

In vitro transcription in isolated nuclei

In vitro transcription reactions were performed by a modification of published procedures (Evans et al., 1977; McKnight and Palmiter, 1979; Groudine et al., 1981). Nuclei were prepared by homogenization in a solution containing 10 mM Tris-HCl, pH 8, 2.5 mM $MgCl_2$, 5 mM DTT, 0.25% Triton X-100, and 0.3 M sucrose in a motor driven Potter homogenizer at 4°C. Nuclei were filtered through Nitex cloth, and recovered by centrifugation at 200 x g for 5 min. Nuclei were resuspended in the 0.3 M sucrose solution, mixed with an equal volume of 2.1 M sucrose in the same solution, and placed in a SW40 centrifuge tube under a cushion of 2.1 M sucrose. Nuclei were pelleted at 30,000 rpm for 1 h, resuspended at approximately 2 x 10^9 nuclei per ml in a solution containing 50 mM Tris-HCl, pH 8, 5 mM $MgCl_2$, 40% glycerol, 0.5 mM DTT, and 0.1 mM EDTA, and frozen at -70°C.

Transcription reactions contained 2 x 10^7 nuclei in 200 ul of a solution containing 10 mM Tris-HCl, pH 8, 16% glycerol, 150 mM KCl, 5 mM $MgCl_2$, 1 mM $MnCl_2$, 14 mM 2-mercaptoethanol, 1 mM each of

GTP, CTP, and ATP, and 0.67 mCi of [^{32}P]UTP (3000 Ci/mmol). Some reactions included amanitin at 1 ug/ml. Reactions were incubated at 26^{o}C for 45 min, digested with 50 ug/ml iodoacetate treated DNase I containing 1 mM $CaCl_2$, followed by lysis of the nuclei with 1% SDS. RNA was recovered by TCA precipitation onto nitrocellulose filters, equilibrated with buffer, and incubated in a solution of 0.1 M NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$ and 10 mM Tris-HCl, pH 8, containing 50 ug/ml DNase I. RNA was eluted from the filters in a solution of 1% SDS, 1 mM EDTA, and 10 mM Tris-HCl, pH 8, and recovered by ethanol precipitation.

Labeled RNA, broken with alkali by incubating in a solution of 0.2 M NaOH for 20 min on ice, was hybridized with plasmid DNA, immobilized on nitrocellulose filters, in a solution of 0.3 M NaCl, 20 mM TES, 10 mM EDTA, and 0.2% SDS at 67^{o}C for 40 h. Filters were washed in 0.2 x SSC at 60^{o}C followed by incubation with 5 ug/ml pancreatic RNase A at 37^{o}C in 2 x SSC, and were counted in liquid scintillation fluid. Relative transcription was calculated from the total radioactivity bound to the filter, minus a plasmid control, per 10^6 input cpm. This number reflects the relative transcription rate for each sequence.

FIGURE 1

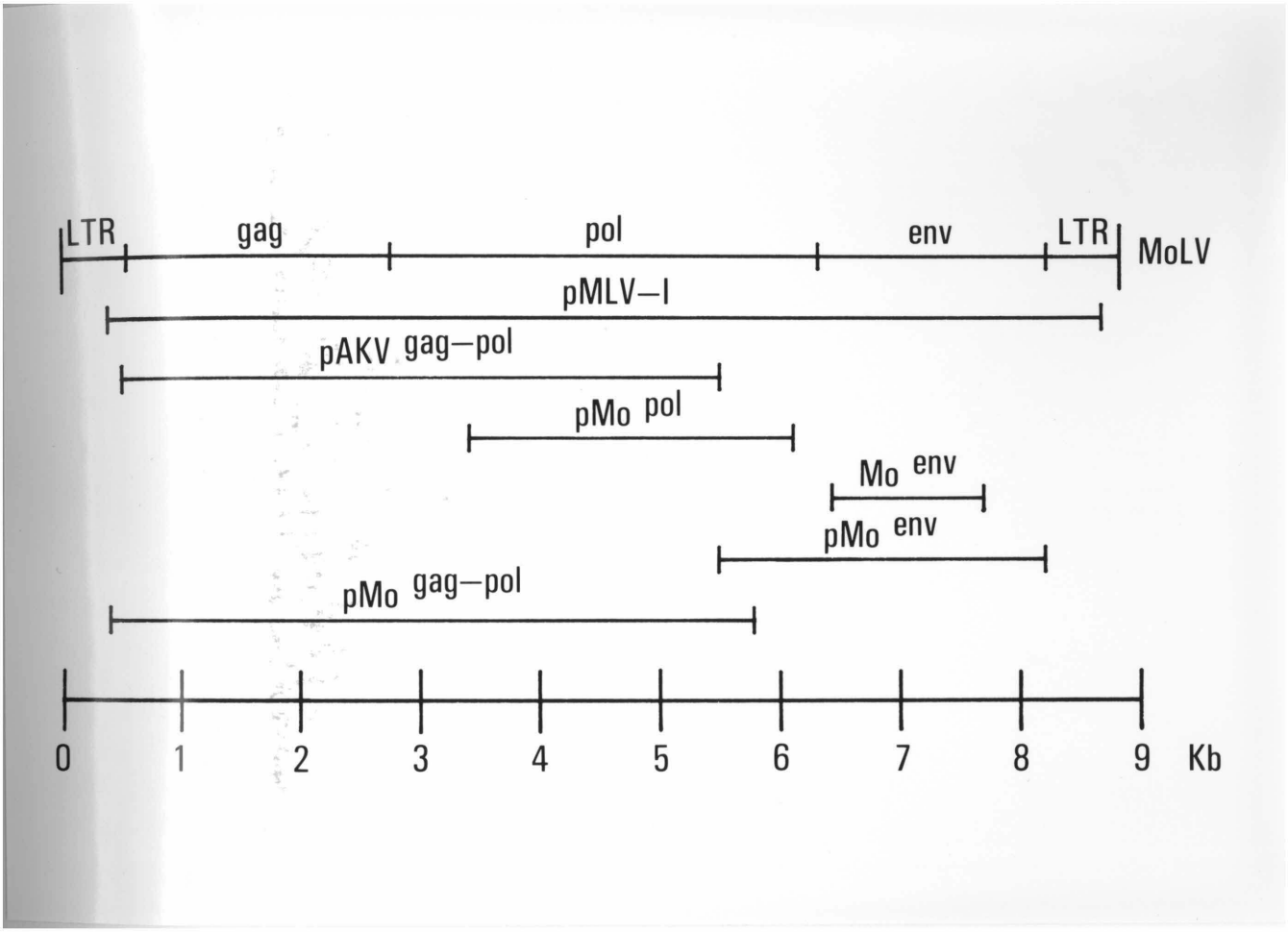

LTR
gag
pol
env
LTR
MoLV
pMLV—I
pAKV gag—pol
pMo pol
Mo env
pMo env
pMo gag—pol
0
1
2
3
4
5
6
7
8
9
Kb

FIGURE 1

LINEAR MAP OF MURINE LEUKEMIA VIRUS RECOMBINANT DNA PROBES

pMLV-1, a copy of the unintegrated form of MoMuLV cloned into the Hind III site of pBR322 (Berns et al., 1980), is presented corresponding to the gene order of viral genomic RNA. pMo^{pol} was derived from pMLV-1 by Kpn I digestion and religation, pMo^{env} by Pvu II digestion and religation, and $pMo^{gag-pol}$ by Xba I digestion and religation. Mo^{env} representing the gp70 coding region of MoMuLV was purified by sucrose gradient centrifugation from pMLV-1 digested with Hpa I. $pAKV^{gag-pol}$ is a subclone in pBR322 of a portion of AKR ecotropic virus and was kindly provided by D. Lowy.

FIGURE 2

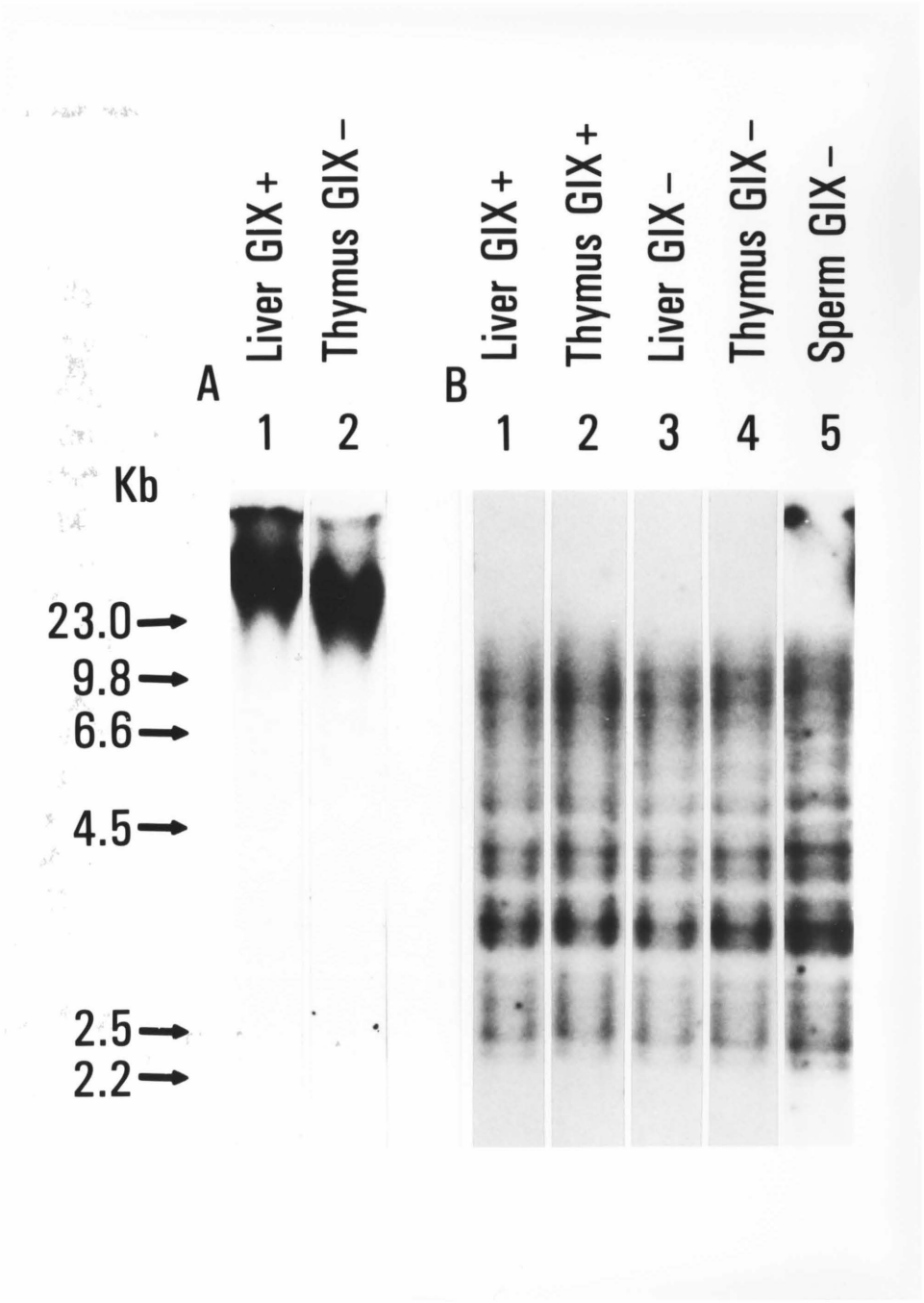

A
Liver GIX+
Thymus GIX−
1
2
B
Liver GIX+
Thymus GIX+
Liver GIX−
Thymus GIX−
Sperm GIX−
1
2
3
4
5
Kb
23.0
9.8
6.6
4.5
2.5
2.2

FIGURE 2

ENDOGENOUS MoMuLV env HOMOLOGOUS SEQUENCES IN STRAIN 129 MICE

High molecular weight DNA from liver and thymus of Gix^+ and Gix^- 129 mice and from sperm of Gix^- mice were separated on a 0.6% agarose gel untreated (A) and after difestion with Eco RI (B). After transfer to nitrocellulose, the DNA was annealed to nick translated Mo^{env} DNA fragment at 67°C in 0.75 M NaCl and the blot was washed at 50°C with 30 mM NaCl and exposed to presensitized Kodak XAR film with an intensifying screen. The mobility of Hind III restriction fragments of lambda DNA is indicated. (A) Lanes: 1, Gix^+ liver; 2, Gix^- thymus. (B) Lanes: 1, Gix^+ liver; 2, Gix^+ thymus; 3, Gix^- liver; 4, Gix^- thymus; 5, Gix^- sperm.

FIGURE 3

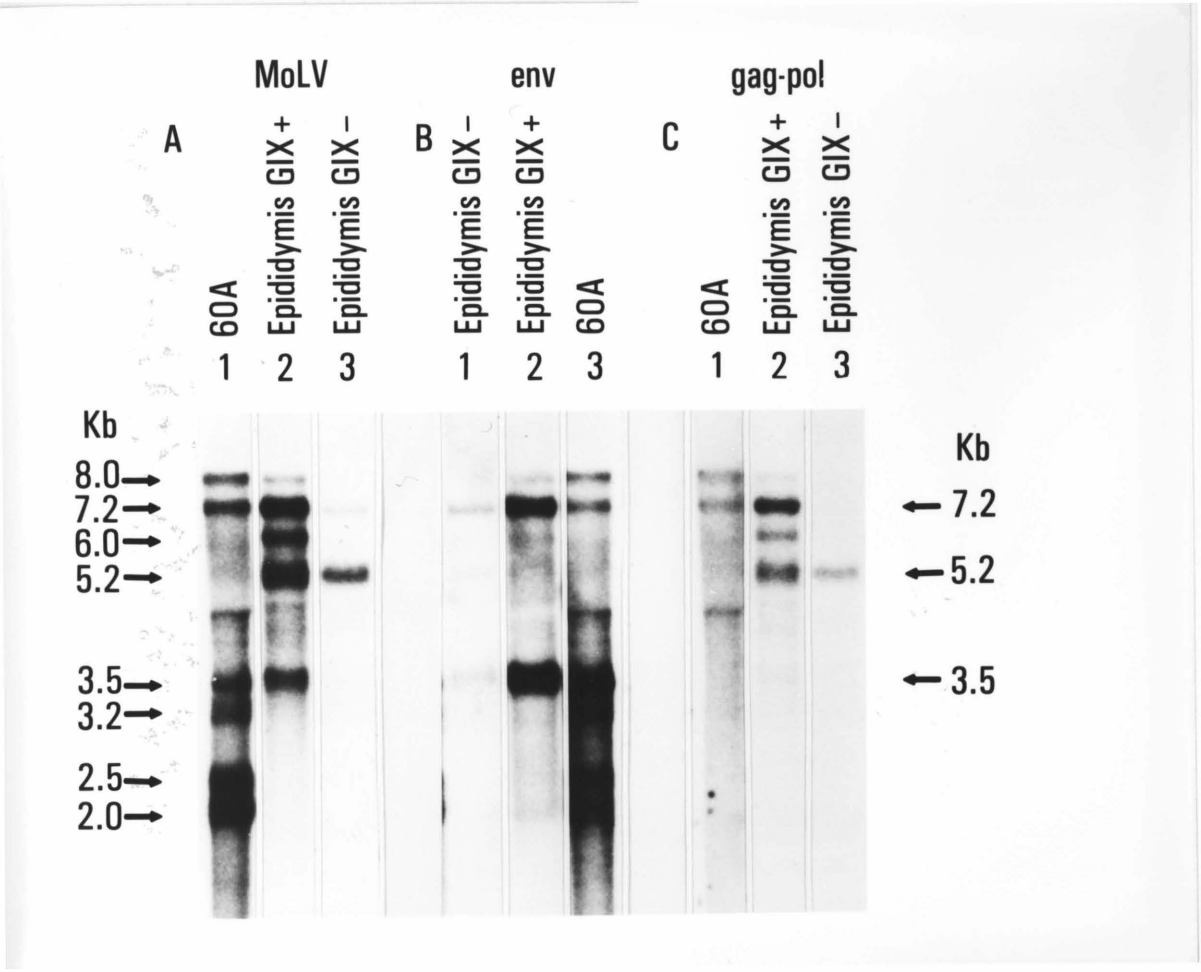

MoLV
env
gag-pol
A
B
C
60A
Epididymis GIX+
Epididymis GIX−
1
2
3
Kb
8.0
7.2
6.0
5.2
3.5
3.2
2.5
2.0
Kb
7.2
5.2
3.5

FIGURE 3

HYBRIDIZATION OF POLYADENYLATED RNA WITH MoMuLV PROBES

Poly(A)$^+$ RNA (5ug) from SCRF 60A cells and from 129 Gix$^+$ and 129 Gix$^-$ epididymis was electrophoresed in triplicate lanes of a 1% agarose/formaldehyde gel, transferred to nitrocellulose, and hybridized with the nick translated recombinant DNA probes (2-4 x 10^5 cpm/ng) pMLV-1 (a), Mo^{env} (B), and pAKV$^{gag-pol}$ (C). Sizes of individual bands were estimated from the mobility of denatured DNA fragments of known size generated by restriction endonuclease digestion of pMLV-1. Preliminary data (not shown) indicate that the 2.5 kb and 2.0 kb RNA in 60A cells contain gp70 coding information and the U3 polyadenylation site but lack the intervening p15E sequences and may thus encode a secreted form of gp70 since p15E normally functions as the membrane anchor for gp70 in infected cells and viruses (Pinter et al., 1978). (A and C) Lanes: 1, 60A; 2, Gix$^+$; 3, Gix$^-$. (B) Lanes: 1, Gix$^-$; 2, Gix$^+$; 3, 60A.

FIGURE 4 (A)

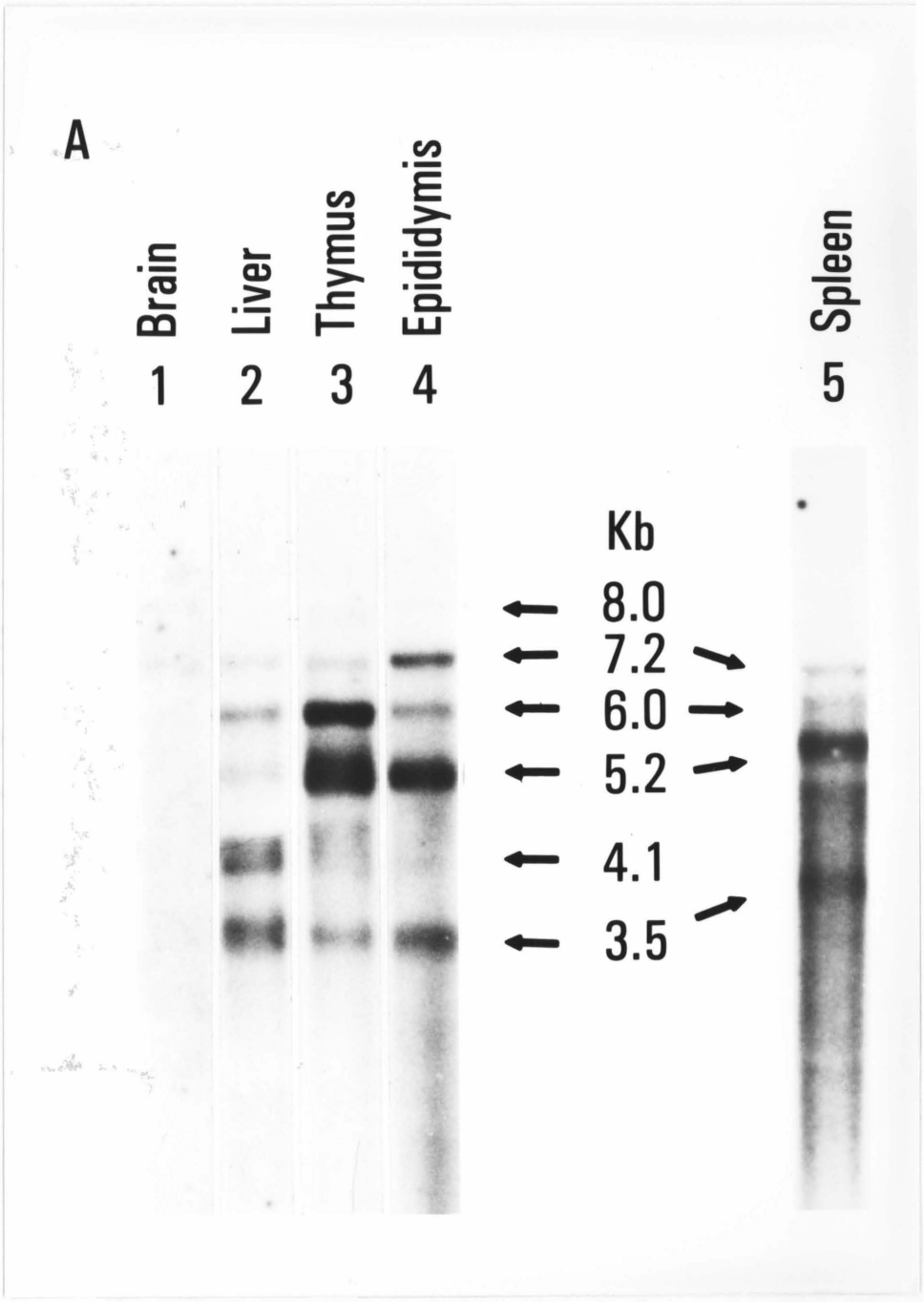

A
Brain
Liver
Thymus
Epididymis
Spleen
1
2
3
4
5
Kb
8.0
7.2
6.0
5.2
4.1
3.5

FIGURE 4 (B & C)

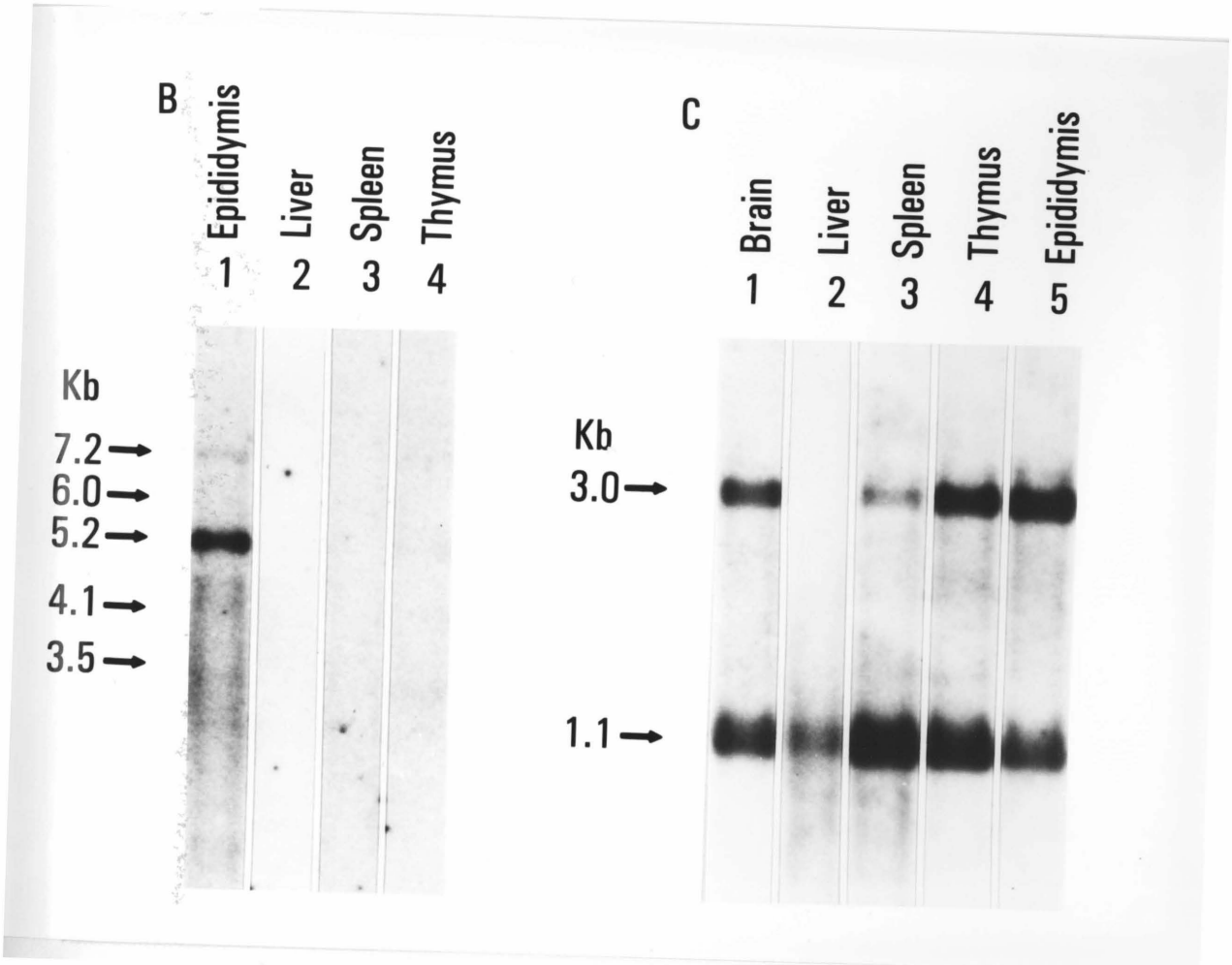

B
Epididymis
Liver
Spleen
Thymus
1
2
3
4
Kb
7.2
6.0
5.2
4.1
3.5
C
Brain
Liver
Spleen
Thymus
Epididymis
1
2
3
4
5
Kb
3.0
1.1

FIGURE 4

HYBRIDIZATION OF POLYADENYLATED RNA FROM TISSUES OF CONGENEIC 129 MICE

Poly(A)$^+$ RNA was analyzed as described in Figure 3. RNA from Gix$^+$ mice (A) and from Gix$^-$ mice (B) was hybridized with nick translated pMLV-1. RNA from brain of Gix$^+$ mice and from liver, spleen thymus and epididymis from Gix$^-$ mice (C) was hybridized with nick translated pCHO B and pCHO C. (A) Lanes: 1, brain; 2, liver; 3, thymus; 4, epididymis; 5, spleen. (B) Lanes: 1, brain; 2, liver;, 3, spleen; 4, thymus; 5, epididymis.

TABLE 1

TRANSCRIPTION OF MoMuLV HOMOLOGOUS RNA IN Gix^+ AND Gix^- LIVER AND BRAIN NUCLEI

Clone	Input, cpm x 10^{-6}		Hybridization, cpm		Relative transcription, no. x 10^6	%Gix^+ liver activity
	Exp.1	Exp.2	Exp.1	Exp.2		
			Gix^+			
			Liver			
$pMo^{gag-pol}$	11.5	11.5	1,356	1,706	133.1	100
pCHO B	11.5	11.5	260	223	21.0	100
			Brain			
$pMO^{gag-pol}$	3.5	1.8	72	44	22.4	17
pCHO B	3.5	1.8	33	40	15.7	75
			Gix^-			
			Liver			
$pMO^{gag-pol}$	11.9	9.2	297	213	24.0	18
pCHO B	11.9	9.2	271	147	19.4	92
			Brain			
$pMO^{gag-pol}$	6.7	1.4	54	40	17.9	13
pCHO B	6.7	1.4	84	42	20.8	99

Table 1. Sucrose purified nuclei from the livers or brains of 4 to 6 week old male 129 Gix^+ and 129 Gix^- mice were incubated with [^{32}P]UTP as described (Evans et al., 1977; Groudine et al., 1981). Relative transcriptional activities represent arithmetic means for two separate nuclei preparations of the amount of radioactivity bound to duplicate filters, as pancreatic ribonuclease A resistant RNA/DNA hybrids, minus background radioactivity bound to filters containing only plasmid vector DNA, per 10^6 input cpm. For comparative purposes, the transcriptional activities in Gix^+ liver nuclei were arbitrarily considered as 100%. All results reported were more than twice background level of radioactivity bound to vector DNA alone. The eukaryotic DNA sequence of $pMo^{gag-pol}$ is 5.5 kb long and that of pCHO B is 0.7 kb long.

TABLE 2

EFFECT OF AMANITIN ON MoMuLV HOMOLOGOUS TRANSCRIPTION IN Gix$^-$ LIVER NUCLEI

	Hybridization (cpm)		
Clone	Control	Amanitin	% inhibition
Incorporation:	9.2×10^6	6.3×10^6	32
$pMo^{gag-pol}$	213	49	77
pCHO B	147	42	71

Transcriptional activities were measured as described in Table 1 except for the inclusion of alpha-amanitin at 1 ug/ml in parallel samples during nuclei labeling. These data were derived from a single experiment.

CHAPTER THREE

Gv-1 COORDINATELY REGULATES TRANSCRIPTION FROM INDEPENDENT ENDOGENOUS RETROVIRUSES

Gv-1 COORDINATELY REGULATES TRANSCRIPTION FROM INDEPENDENT ENDOGENOUS RETROVIRUSES

INTRODUCTION

The 129 Gix$^-$ strain was selected only for the absence of thymic cell surface gp70. The Gix$^-$ phenotype, however, appears more complex than simply the loss of a single expressed gp70 structural gene or of an active retroviral provirus as was originally suggested (Strand et al., 1974; Tung et al., 1982). 129 Gix$^-$ mice are often described as completely lacking viral antigens. However, low but detectable quantities of gp70 and another viral protein, p30, have been reported (Lerner et al., 1976) indicating that the structural genes are still present. To pinpoint the action of the Gv-1 gene and the level at which this regulation is maintained, and to fully document the involvement of multiple structural genes in the Gix phenotype, an investigation was initiated into the extent of retroviral gene expression in strain 129 mice. As described in the last chapter, a negative allele at Gv-1 correlated with reduced steady state levels of retroviral RNA. Furthermore, transcription measurements in isolated nuclei demonstrated that the Gv-1 locus is influencing the transcriptional activity of the structural genes for endogenous retroviruses. The observation that the abundances of multiple sized RNA species correlated with the Gv-1 allele leads us to hypothesize that multiple independent transcription units are being affected. If multiple, dispersed genes are coordinately regulated, Gv-1 is required to be acting in trans, perhaps encoding a transcription factor required for activity of these endogenous proviruses. By this model, the Gix$^-$ phenotype results from a lesion in Gv-1, leading to reduced expression from proviral structural genes.

The demonstration that the Gix phenotype results from multiple transcription units responding to a common regulatory signal requires that the products of individual structural genes be examined. Potential structural heterogeneity of expressed retroviral transcripts can be determined from cDNA clones and the nature of this heterogeneity may define the minimum number of independent transcription units involved in

the phenotype. However, since cDNA cloning necessarily examines only individuals, the results obtained from this procedure must be related back to the RNA population as a whole. In order to define the phenotype controlled by Gv-1, cDNA clones derived from cytoplasmic RNA species homologous to Moloney murine leukemia virus (MoMuLV) have been isolated. The structures of these clones indicated that transcripts from multiple, distinct proviral genomic sequences are contributing to the Gix phenotype. Blot hybridization results using these cloned probes indicated that transcripts homologous to these polymorphic clones are collectively regulated by Gv-1. However, due to the high degree of sequence conservation of endogenous virus related transcripts, these probes did not discriminate between individual RNA species by blot hybridization. To establish that polymorphic transcripts corresponding to these probes were indeed regulated, however, required that this heterogeneity be directly assayed in the RNA population. It was formally possible that the cDNA clones were copies of transcripts present at a low level in both 129 Gix^+ and 129 Gix^- mice and not under any differential regulation. Such cDNA clones, although themselves heterogeneous, might hybridize a uniform population of regulated transcripts, thus giving the impression that polymorphic transcripts were being differentially expressed. In order to distinguish between these two possibilities, the cloned retrovirus cDNA sequences were used as probes in S1 nuclease protection hybridizations. Results from these analyses demonstrated that the population of RNA transcripts that is differentially expressed between Gix^+ and Gix^- mice are indeed polymorphic and that the cDNA clones accurately reflect the abundant transcripts in this population. Furthermore, some of these transcripts are regulated in a tissue specific manner showing independent control of distinct transcription units which are are also coordinately regulated by the action of Gv-1. These observations support the postulated model of trans regulation through products encoded by Gv-1 and Gv-2 and may lay the basis for further efforts aimed at characterizing and isolating the gene products of these regulatory loci.

RESULTS

Cloning Endogenous Retroviral Related mRNA

The analysis of cDNA clones of endogenous retrovirus related transcripts has demonstrated the presence of multiple, distinct retroviral RNA sequences in 129 mice. Clone libraries were prepared from the polyadenylated RNA of epididymis, liver, spleen and thymus of Gix^+ mice and screened at reduced stringency for retroviral related sequences with a $MoMuLV^{env}$ probe (see Experimental Procedures). Positive colonies were detected in the epididymis cDNA library at the level of 0.5% and in liver, spleen and thymus in the range of 0.01%–0.05%. Their plasmid DNAs were aligned with the viral map by hybridization with restriction fragments of MoMuLV (Figure 1). The retrovirus related clones selected from these libraries were used to probe RNA blots prepared from poly(A)-enriched RNA from tissues of 129 Gix^+ and 129 Gix^- mice to identify virus related mRNA transcripts. All endogenously expressed retroviral transcripts that annealed with these probes had been detected in our previous study, which depended upon the limited cross hybridization of heterologous probes (Levy et al., 1982).

Figure 2 shows the hybridization pattern of RNA homologous to the pol-env cDNA probe, clone E1. Several points can be noted from this blot. First, this pattern is strikingly similar to that observed previously with ecotropic MoMuLV probes. The 7.2 kb and 3.5 kb RNAs have structures typical for viral transcripts synthesized during a productive retroviral infectious cycle. The large RNA transcript, containing gag, pol, and env sequences, matches full length genomic RNA and gag-pol messenger RNA, while the 3.5 kb RNA conforms with spliced, env encoding mRNA. An 8.0 kb RNA present at variable abundance also contains all viral structural elements (not shown). It may be a natural variant transcript produced during retroviral replication or may be viral genomic RNA. A similar RNA has been observed in virally infected cells (Levy et al., 1982). The RNA species in the 4–6 kb range, previously shown to contain sequences homologous only to the gag-pol genes of MoMuLV, are also detected by the endogenous probes. Two additional transcripts of 2.0–2.5 kb were previously detected only in tumor cells as RNA species

containing gp70 coding sequences but lacking p15E coding capacity (Levy et al., 1982). The increased sensitivity achieved using homologous probes allows detection of these minor species in variable amounts in normal tissue (see Figure 3). These species have also been detected in thymus of 129 Gix^+ using *env*-specific probes derived from spleen focus forming virus (Boccara et al., 1983).

Retroviral RNA accumulates with characteristic tissue specific abundances. When corrected for the mass of RNA applied to the gel, total hybridization to epididymal poly(A)$^+$ RNA is approximately 10 fold greater than to liver, thymus, or spleen RNA preparations, while hybridization to brain RNA is 10 fold lower yet (densitometry not shown). In addition, the *gag-pol* containing RNA species in the 4-6 kb range accumulate with characteristic, individual patterns for each tissue. For example, epididymis contains retroviral RNAs of 5.2 kb and 6.0 kb while liver shows RNA at 4.0 kb and 4.2 kb. Thymus shares the epididymal pattern except the 6.0 kb RNA is at higher relative abundance and thymus expresses an additional 5.0 kb RNA. We have been able to show that these RNA species, which are not typical in size or structure for normal retroviral replication, derive from separate transcription units (see below).

The RNA blot shown in Figure 2 exemplifies the effect of *Gv-1* regulation. With the exception of epididymis (discussed below), the use of MoMuLV probes does not detect endogenous retrovirus related RNA in tissues of the Gix^- strain even when larger amounts of RNA are loaded on the gel (Chapter 2; Levy et al., 1982). However, the increased sensitivity of hybridization with endogenous probes does enable detection of some retroviral RNA expression in Gix^- mice. Most clearly demonstrated by the lengthened exposure of the thymus lane (Figure 2, far left lane), the RNA species expressed in Gix^+ are also expressed by Gix^-. Quantitative densitometry of these autoradiograms (not shown) indicates a 10-12 fold reduction of retroviral RNA in Gix^- thymus (e.g., the 7.2 kb and 5.1 kb RNAs). Interestingly, each distinct RNA species exhibits an idiosyncratic response to *Gv-1* regulation. Whereas the 7.2 kb and 5.2 kb RNAs are approximately 10 fold less abundant, the 8.0 kb, 6.0 kb, and 3.5 kb RNA

bands are further reduced in intensity. The hybridization depicted in this figure was obtained using a cDNA clone derived from epididymis. Due to the high degree of retroviral sequence conservation among expressed endogenous retroviral transcription units (see Chapter 4), similar results are obtained using cDNA probes derived from other tissues (data not shown).

The detection in Gix^- tissue of low levels of retroviral RNA comigrating with Gix^+ species is a strong indication that the Gix phenotype results not from a defect in proviruses directly (e.g., proviral deletion as suggested by Strand et al., 1974 and Tung et al., 1982) but rather from a reduced level of transcription of competent genes. This is consistent with our previous findings of no proviral deletions in Gix^- genomic DNA and of a reduced yet measurable transcription rate for retrovirus related sequences in Gix^- spleen and liver (Levy et al., 1982).

Retroviral Transcripts are Derived from Independent Proviruses

The U3/R regions of different retroviruses, containing promoter and enhancer elements in the 5' untranscribed region of the proviruses, are often distinct in structure, perhaps reflecting species and tissue tropisms (Varmus, 1982). As discussed in the Introduction, this region is repeated at the 3' end of viral RNA transcripts where it contributes the poly(A) addition signal. We considered whether the transcripts shown in Figure 2 similarly generated their 3' termini and whether the different RNA species could be distinguished by virtue of their U3 regions which presumably are copies of the respective proviral promoters. A cDNA restriction fragment containing only U3 sequence was produced by digesting clone E2 with the endonuclease Pst I and isolating the small, 3' terminal fragment (see Figure 4). Figure 3 depicts the hybridization of this probe to $poly(A)^+$ RNA of Gix^+ tissues. All retroviral transcripts identified by homology with *pol* and *env* sequences also hybridized U3, with this probe detecting autoradiographic signals of similar relative intensities as those detected by clone E1 (compare with Figure 2). This indicates that all these RNA species may have typical retroviral 3' termini, generated by cleavage and polyadenylation within this U3 structure, even the atypical *gag-pol* transcripts which lack *env* sequence and might not be expected to be contiguous with U3. However, it is possible that the sequences hybridizing

the U3 probe are internal in some RNA transcripts rather than being 3' terminal. This altenative cannot be distinguished by these analyses.

To determine whether different RNA species include different U3 sequences, a Pst I–Msp I fragment from the U3 region of clone E2 was 3'–end labeled at the Pst I site and used in an S1 nuclease protection mapping experiment (Figure 4). In all tissues, two major protected products were resolved. The 450 nt fragment was protected by hybridization of an RNA transcript fully complementary to the probe. A major subpopulation of transcripts protects only the 5' 120 nt of the probe, demonstrating a break in homology at the beginning of the presumed enhancer region (point **a** on the U3 map of Figure 4). Two other, minor populations of RNA molecules are detected by this analysis: a 185 nt protected fragment and a 400 nt fragment which map points of non–homology within U3, indicated as points **b** and **c**, respectively. In this and in all other S1 analyses reported here, control hybridizations using yeast RNA or rat brain RNA in place of mouse RNA gave no nuclease protected fragments. Furthermore, in most cases the labeled probe also contained plasmid sequences so that the size of self–annealed probe could be readily distinguished from full length protection of the cDNA sequence (see Experimental Procedures).

The different RNA populations identified by these distinct U3 regions are present in all tissues examined of both Gix^+ and Gix^- animals. However, increased amounts of input RNA were required for brain and Gix^- tissues, commensurate with the different steady state levels of RNA detected by blot hybridizations. Furthermore, molar ratios of fragments protected by RNA of different tissues indicated that individual RNA species bearing different U3 regions are separately regulated. For example, the sequence generating the 120 nt protected fragment is under–represented relative to the U3 structure represented in clone E2 when RNA from Gix^- tissues is compared with Gix^+. Tissue specific regulation is apparent, moreover, in the RNA species with this U3 structure that are low in Gix^- epididymis and liver and undetectable in Gix^- thymus. These data indicate that the expression of RNA transcripts with different U3 structures are affected differentially by Gix regulation in each tissue.

The precise nature of the observed U3 heterogeneity was examined directly by sequencing several representative U3 containing cDNA clones. Figure 5 presents the sequences of a Pst I–Kpn I fragment from the U3 regions of two endogenous transcripts. The multiple nucleotide changes and insertion/deletion differences in the sequences of this region of clones E2 and L62 lead us to conclude that these endogenous transcripts are derived from separate genomic proviruses rather than from a common precursor through a somatic process such as RNA processing. The positions of deletions in clone L62 relative to E2 are consistent with the sizes of S1 protected fragments of Figure 4. A 5 nt deletion (sequence position 127–131) is present 125 nt 3' from the labeled nucleotide. Likewise, a 42 nt deletion (187–228) is located 185 nt downstream from the end label used in the S1 experiment. RNA species with these deletions would generate the 120 nt and 185 nt fragments protected from S1 digestion depicted in Figure 4 when hybridized with the E2 Pst I–Msp I fragment. In clone L62, both deletions exist on the same molecule and thus only the 5 nt, 5' most deletion would be expected to be detected by S1 analysis. The 185 nt fragment in Figure 4 may thus represent a minor population of transcripts resembling L62 by missing nucleotides 187–228 and yet containing nucleotides 127–131 of clone E2. Alternatively, the 185 nt fragment may have been generated by incomplete cleavage of S1 at the first deletion point. The 400 nt protected fragment is not explained by the sequences of these two clones. Presumably, therefore, another population of transcripts is present in Gix^+ tissues which is homologous with E2 from the Pst I site until approximately nucleotide 400. Thus, at least two and perhaps four distinct U3 structures are present in Gix^+ RNA transcripts.

The major difference between the two endogenous U3 sequences is the deletion of the two sequences described above, falling within the presumed enhancer region of these promoter elements as defined by analogy with exogenous infectious viruses such as AKV or AKR MCF 247 virus (see Kelly et al., 1983). This is also a region which distinguishes different viral LTRs from each other and may be important for the tissue and species tropism of infectious retroviruses (Varmus, 1982). The most

striking feature distinguishing the 129 Gix^+ endogenous U3 related regions from the U3 sequences of the exogenous viruses, however, is the presence in the endogenous transcripts of an approximately 200 nt insertion between nucleotides 240 and 440 of sequence not found in infectious virus (Khan and Martin, 1983; see Chapter 4). This insertion further separates the presumed enhancer regions from other promoter features (e.g., the 5'-CAAT-3' equivalent element and the TATA box at positions 483-487 and 539-546, respectively). The U3 sequence in the 3' LTR of a retrovirus contributes the 5' U3 of the provirus on subsequent rounds of viral replication and integration and thus provides the promoter activity (see Varmus, 1982). The features specific to U3 noted here may thus also be represented in the promoters for these transcripts.

The genomic organization of U3 sequences was found to be consistent with this idea. Total genomic DNA from 129 Gix^+ mice was digested with the restriction endonucleases Pst I and Kpn I, gel fractionated, and analyzed by filter hybridization using U3 probes (Figure 6). The sequences presented in Figure 5 predict a difference of 50 bp between the Pst I-Kpn I fragments of the two U3 structures. Cleavage at these conserved restriction sites (Chattapadhyay et al., 1982) recruits the multiple copies of endogenous LTR into two major fragment sizes of 550 bp and 600 bp detected by hybridization with the 253 bp Pst I-Msp I probe **A** (Figure 6, lane **A**). The great majority of LTR structures thus fall into one of these two classes, defined by the deletions in clone L62, and presumably represent both 5' and 3' LTR structures. The sizes of these genomic restriction fragments predict the presence of the 200 bp insertion which distinguishes LTRs from infectious virus; this is confirmed by hybridization with the 167 bp Alu I-Bgl II probe **B** (Figure 6, lane **B**) from within this insert region (nucleotides 253-419 in Figure 5). Alignment of the sequence of this region with the sequence of the transposon-like sequence reported by Wirth et al. (1983, 1984) reveals 96% similarity between these endogenous LTRs and this family of middle repetitive DNA (Chapter 4) which explains the smear of hybridization to high molecular weight restriction fragments which do not hybridize probe **A**. The several distinguishing features mapping to the enhancer and promoter elements suggest a possible role in the specificity of endogenous viral transcription.

The biologic activities of these DNA sequences are being investigated (D.E.L., R. McKinnon, and M.C.W., unpublished; see Chapter 5).

Tissue Specific RNA Transcripts are Transcribed from Independent Proviruses

The analyses presented above strongly suggested that multiple, distinct proviral species are regulated by Gv-1 and give rise to multiple RNA transcripts. We next examined coding regions for evidence of structural heterogeneity to explain the distinct sized, 4-6 kb transcripts which accumulate with apparent tissue specificity. The transcripts hybridize gag-pol probes and LTR probes but not envelope probes (Figures 2 and 3; Levy et al., 1982). Therefore, we probed the structure of pol-env regions by S1 analysis using a cDNA restriction fragment cleaved from clone E1 and 3' end labeled at an Xba I site approximately 900 nt 5' to the pol-env junction. Clone E1 is probably a partial copy of a full length transcript; this probe therefore detects the first position of distinguishing non-homology 3' to the labeled site in aberrant transcripts. RNA from Gix^+ tissues protects four major fragments which identify cleavage sites clustering around the BamH I site at the pol-env junction (Figure 7). The sizes and relative intensities of the protected fragments reflect the abundances of the 4-6 kb gag-pol messages detected by RNA blotting. Fragments of 1050 nt protected by spleen and thymus RNA from Gix^+ animals and epididymal RNA from Gix^- animals correlate with the approximately 5.2 kb RNA species detected in spleen, thymus and Gix^- epididymis but absent from liver. The minor mobility differences between these fragments protected by the spleen and thymus samples may reflect the slight size heterogeneity of the 5.2 kb transcripts observed by RNA blot hybridizations (see Figures 1 and 2). Likewise, the 1100 nt fragment protected by thymus RNA may derive from the 6.0 kb thymus RNA transcript. Thus, different RNA transcripts contain sequence differences (nucleotide changes or deletions) in the region homologous to the pol-env junction of the cloned probe.

Clearly, the heterogeneity about the divergence point assayed in this S1 experiment does not explain completely the size heterogeneity of the transcripts. Presumably, the differences in the transcripts causing the loss

of homology with the probe mapping to this point result in molecules of different lengths. For example, different sized deletions whose 5' breakpoints all map to this region would result in similarly sized S1 fragments and yet different length transcripts. However, these differences do not fully explain the size differences detected by RNA blotting. Further deletions are probably exhibited by the gag-pol regions which are beyond the scope of this probe.

The analyses of the 4-6 kb tissue specific RNA species reported above are most consistent with structures containing some gag-pol sequences fused to U3, eliminating the normally present env sequences. To focus on the nature of the env deletions, we compared the RNA populations from different tissues with probes to the 5' end and 3' end of the env coding sequence. Figure 8 depicts the pattern of protected fragments produced using a DNA probe to the 5' two-thirds of the gp70 coding region, 3' end-labeled at the BamH I site at the extreme 3' end of the pol gene sequence derived from clone E1. The 1000 nt fragment protected by RNA from all four tissues is detecting RNA species colinear with the probe, either genomic RNA or spliced env messenger, since the labeled end of the probe is 3' to the presumed splice acceptor site for the gp70 message (Shinnick et al., 1981). The fragments of 65 nt and 100 nt, however, map different RNA species whose sequences diverge from that of the DNA probe at points near the 5' end of the gp70 coding region. These different RNA species are present at different relative abundances in different tissues, as seen in the ratios of the intensities of the shorter fragments compared with full length. The 180 nt fragment is under represented when RNA from liver is used for the hybridization, but substantially higher using RNA from thymus. Conversely, the 65 nt fragment is protected with liver and thymus RNA while it is undetectable using kidney RNA. These fragments correspond to cleavage at sites **a** and **b** indicated on the map, which are equivalent to **c** and **d** of Figure 7. They probably result from hybridization of the 5.2 kb and 6.0 kb gag-pol transcripts which accumulate differentially in Gix^+ tissues. The size difference of these transcripts is partially due to this deletion heterogeneity as well as presumably to additional deletions in gag-pol sequences.

These tissue specific transcripts diverge from a typical retroviral sequence at the 5' end of env and yet also end with a typical viral U3 sequence at their 3' ends. To map the precise site of the presumed env deletion, nuclease protection analysis was done to define the point where these transcripts return to homology with intact sequences. Using a probe to p15E sequences derived from clone L510 (see Figure 1), end labeled at an Rsa I site within U3 (nucleotides 34–37 in Figure 5), an abundant fragment of 230 nt from this 825 nt probe was protected in all tissues. This mapped the point of non-homology to the extreme 3' end of the env gene demonstrating that essentially all env sequences were deleted in some RNA transcripts in these tissues. Other protected fragments are also depicted in this figure which correspond to other deletion endpoints. Interestingly, a unique 210 nt fragment is protected by hybridization with RNA from Gix^- epididymis. This probably corresponds to the 5.2 kb transcript detected on RNA blots of Gix^- epididymis which is not responsive to Gv-1 dependent regulation (Figure 2). As this fragment is not protected by RNA from any Gix^+ tissue, it is consistent with being a transcript from a single provirus contributed by the non-129 parental strain used as the source of the Gv-1 null allele during the construction of these congenic inbred lines (Stockert et al., 1971).

To identify precisely the endpoints of the env deletions in particular RNA molecules, suggested by the preceding experiments, we have sequenced portions of two cDNA plasmid inserts from clones S13 and E2 which correspond to transcripts with pol and LTR regions but which lack env (see Chapter 4). The regions identified by hybridization where these potential deletions occur, two pol-p15E junctions, are shown in Figure 9. As predicted by the S1 results, the 5' breakpoints of these deletions both map to the pol-env junction while the 3' breakpoints fall at the 3' end of the env gene in the p15E coding region, thus eliminating much of the env gene. The sequence across these breakpoints presented in Figure 9 are aligned with sequences for the corresponding regions of full length transcripts (see Chapter 4). Interestingly, in both cases these deletion endpoints map precisely to two different 7 nt direct repeats present in the sequences of the full length transcripts. In del env-1 of clone S13,

the sequence 5'-GGACCCT-3' is found at the junction of pol sequence with p15E coding sequence. This sequence is present twice in full length transcripts, once at the 3' end of pol and once within p15E, precisely at the ends of the env deletion of del env-1. Similarly for del env-2, the sequence 5'-GGTCCAG-3' marks the point of deletion and is present in two copies in the full length transcripts precisely at the deletion breakpoints. In addition, there are scattered single nucleotide differences between these transcripts and the putative full length RNA species (see Chapter 4). These data are most consistent with the 4-6 kb RNA transcripts being derived from deleted proviral structures rather than through the processing of a single precursor.

Proviral Hypomethylation Corresponds to Gv-1 Regulation

We have previously presented evidence suggesting that a major component of the Gv-1-directed regulation is acting by affecting transcription rates of endogenous proviruses (Levy et al., 1982). However, because of the high degree of sequence conservation among endogenous proviruses, we have been unable to produce a probe which distinguishes active from inactive genomic copies directly. To distinguish proviral genes predicted to be active in different tissues, the extent of cytosine methylation of retroviral DNA sequences was examined in different tissues, a property often associated with the activity of a gene (Doerfler, 1983, 1984). In general, transcriptionally active proviruses have been found to be partially unmethylated, whereas the bulk of endogenous retroviral sequences are integrated in regions of methylated DNA and presumably transcriptionally silent (Groudine et al., 1981; Montandon et al., 1982; van der Putten et al., 1982; Hoffman et al., 1982). Here we have used the restriction endonuclease Sma I as a probe for the methylation status of proviruses in genomic DNA from liver and thymus of Gix^+ and Gix^- mice (Figure 10). High molecular weight DNA was digested with EcoR I or with EcoR I plus Sma I and fractionated on an agarose gel. Since Sma I cleaves the sequence 5'-CCCGGG-3' only when the internal cytosine residue is unmethylated, the release of a new fragment in the EcoR I digestion pattern is indicative of a demethylated site. Hybridization with a retrovirus related probe reveals relatively few EcoR I proviral restriction fragments containing unmethylated Sma I sites. However,

fragments of approximately 2.2 kb and 3.5 kb are released upon Sma I digestion of Gix^+ and Gix^- DNA (marked by stars in Figure 10). Furthermore, there are fragments of 1.35 kb and 5.0 kb released by Sma I in liver DNA of Gix^+ mice that are not present in liver DNA of Gix^- mice or in thymus DNA of either Gix^+ or Gix^- mice. Likewise, there is a fragment of 6.0 kb found in thymus DNA from Gix^+ mice that is not present in thymus DNA of Gix^- mice or in liver DNA of either Gix^+ or Gix^- mice (marked with arrowheads). These data indicate that there are specific C-p-G residues in proviral DNA which become demethylated only in Gix^+ mice and only in certain tissues. Probe B in Figure 10, which hybridizes only to sequences on the 5' side of the conserved EcoR I site in endogenous proviruses, visualizes these demethylated restriction fragments, indicating that the demethylation is occurring at the 5' end of the proviruses and may be the Sma I site within the 5' LTR of the promoter region.

DISCUSSION

The analysis of the normal expression of endogenous retroviral sequences in 129 Gix^+ and Gix^- mouse strains has demonstrated the action of a regulatory element, encoded by Gv-1, which functions in trans to control the expression of this multigene family. Classically, Gv-1 had been defined genetically as controlling the level of gp70 protein on thymocytes (Stockert et al., 1971; 1975) and was initially considered to be the structural gene for a transcriptionally active provirus (Strand et al., 1974). Further experiments, however, indicated that the synthesis of several, distinct gp70 proteins (Elder et al., 1977) and of several distinct retroviral transcripts (Levy et al., 1982) was coordinately regulated by the Gv-1 locus, suggesting that Gv-1 is distinct from the structural genes for retroviral proteins. However, it could not be formally excluded that a single provirus, producing multiple transcripts through alternative processing pathways, could account for the observed phenotype. For example, the 7.2 kb transcript corresponding to viral genomic RNA and the 3.5 kb envelope mRNA can both be derived from a single provirus. However, we report here DNA sequence and S1 protection data which clearly demonstrate that a number of retroviral RNA transcripts normally expressed in 129 Gix^+ mice are derived from distinct proviruses, that they are expressed in a tissue specific manner, and that they are coordinately regulated by the Gv-1 locus.

The differences that demonstrate expression of multiple proviral genes coordinately regulated under Gv-1 control fall into two categories. When the homologous regions of different clones are compared, scattered single nucleotide changes are apparent resulting in overall sequence differences of 3% to 6% (Figure 5 and Chapter 4). These nucleotide differences are greater than would be expected due to reverse transcriptase infidelity during in vitro cDNA synthesis (Battula and Loeb, 1974, 1975). Furthermore, nucleotide changes resulting in restriction site differences between cDNA clones have been found to be represented in genomic copies of retroviral sequences. The second category of divergence between retroviral transcripts are deletions of segments of coding sequence. Like the mRNA species containing single nucleotide changes,

the 4-6 kb transcripts which contain these deletions appear to be transcribed from independent proviruses rather than resulting from common precursors by an RNA processing mechanism. That splicing is not a likely mechanism generating these deleted transcripts is strongly suggested by the nature of the deletions. The deletions in the sequences presented in Figure 5 are shorter (Wieringa et al., 1984) and of different sequence (Mount, 1982) than would be expected for eukaryotic introns. In addition, the two cases presented in Figure 9 show the breakpoint of the deletions terminating precisely in 7 bp direct repeats. The involvement of short, directly repeated DNA sequences in the generation of deletions is common in prokaryotes (Farabaugh et al., 1978; Albertini et al., 1982) and in eukaryotes (Efstratiadis et al., 1980). It has also recently been described for infectious retroviruses (Omer et al., 1983; Mark and Rapp, 1984). The mechanism generating these deletions is thought to be slipped mispairing during DNA replication (Streisinger et al., 1966) in which one member of a repeat base-pairs with the complementary copy of the other member of the repeat on the newly synthesized daughter strand causing the region between to loop out. A similar mechanism could also operate during either reverse transcription or replacement synthesis during viral replication, the resulting deleted proviral DNA becoming integrated into the host germ line. During reverse transcription, this would be analogous to the production of "weak-stop" DNA by stalling of reverse transcriptase, with these DNAs subsequently serving as primers for continued transcription after annealing at a different location. Use of the repeats as primers annealed at secondary sites coupled with exonuclease digestion to create a 3' end flush to the point of homology for continued DNA replication would explain the precise inclusion of exactly one copy of the direct repeat in the deleted provirus. The breakpoints of other proposed eukaryotic examples of direct repeat mediated deletions have not usually fallen precisely at the repeat. An alternative mechanism potentially creating these deletions is homologous recombination mediated by the short repeated sequences. The position of both deletions examined in Figure 9 as well as the presumed positions of others based on the structure of the 4-6 kb transcripts fall within a region of the viral genome thought to be involved in recombination events generating mink cell focus-forming viruses such as AKR MCF 247 (Holland et al., 1983), Moloney MCF (Bosselman et al.,

1982), and C3H MCF CI-1 (Mark and Rapp, 1984). In fact, the 5' recombinant junction identified in the latter case contains the 5' copy of the direct repeat involved in deletion 1 shown in Figure 9. Thus, this region of the provirus may be a recombinational hot spot.

The extensive sequence conservation of endogenous proviruses appears to preclude the isolation of unique hybridization probes for RNA blotting procedures. To evaluate the contribution of any particular cloned transcript to the RNA population of a particular tissue, S1 protection analyses were required. From this procedure, it is apparent that the same complement of viral transcripts expressed in Gix^+ mice is present, at much lower levels, in Gix^- mice. This analysis also revealed a Gix^--specific transcript which is not under Gv-1 control but which is also not present in 129 Gix^+ mice. Furthermore, these S1 protection results confirm the differential accumulation of individual transcripts in different tissues. Since these cDNA clones as well as the exogenous viral probes appear to hybridize the same RNA transcripts and many of the fragments derived in S1 protection analyses can be explained by the different cDNA clones described, a reasonably complete picture of the abundant Gix^+ retroviral transcripts has probably been obtained. This allows the general conclusion to be drawn that Gv-1 controls the expression of all active 129 Gix^+ endogenous proviruses.

The mechanism of action of Gv-1 is still unclear. We have presented evidence that this locus exerts control over a number of independent transcription units to decrease the steady state level of poly(A)-containing retroviral RNA. Furthermore, this reduction in RNA abundance corresponds to greater than a five fold reduction in the rate of transcription of the retroviral genes (Levy et al., 1982; Chapter 2). A further indication of a transcriptional component in the control of retroviral gene expression by the Gv-1 locus comes from analysis of cytosine methylation levels. Proviral DNA in 129 mice presents an interesting example of specific methylation patterns. Hypomethylation of several Sma I sites correlates with the Gix phenotype in a tissue specific manner. Although the methylation of relatively few C-p-G residues were assessed by the Sma I digestion patterns presented in Figure 9, these data

support again differential transcription of some proviral loci. It should be noted that the only hypomethylation detected which correlated with the Gix phenotype was also tissue specific. The absence of detectable hypomethylation patterns for common, Gv-1 regulated transcription units is surprising. Furthermore, the tissue specific hypomethylated DNA fragments give more intense hybridization signals than would be expected for single copy genes. This may indicate that several proviruses containing similarly spaced EcoR I and Sma I restriction cleavage sites are collectively hypomethylated in each tissue.

The data presented in this chapter suppport our model of the Gv-1 locus encoding a trans-acting regulatory factor which specifically controls the transcription of the active endogenous proviruses of 129 mice. This regulation may be similar to the genetically defined trans regulators of yeast such as the general amino acid control (Wolfner et al., 1975) or galactose metabolism (Douglas and Hawthorn, 1966) enzymatic pathways. These regulatory loci were defined by mutations which rendered the mutant cells unable to respond to specific inducers. Operationally, the Gix phenotype can be considered in a similar light, although in this case the expression is tissue specific rather than inducible, presumably responding to developmentally regulated signals. The complete control of this phenotype involves multiple complementation groups: a series of structural genes and the regulatory loci Gv-1 and Gv-2. Genetic analyses of the control of viral antigen production indicate that Gv-2 is a dominant regulator while the penetrance of Gv-1 is semi-dominant. These data suggest the following model for proviral regulation.

The $Gv\text{-}1^{b}$ allele giving the Gix^{-} phenotype may result in a lesion in a positive activator required for efficient transcription of proviruses. The Gv-2 gene product is a second factor which interacts with the Gv-1 product to form a functional transcription factor. In 129 Gix^{+} mice, the Gv-1 and Gv-2 gene products are produced, resulting in a high consitutive level of retroviral transcription. In 129 Gix^{-} mice, the Gv-2 product is synthesized, but the Gv-1 product is defective, resulting in defective transcription complexes and only low levels of proviral transcription. In (Gix^{+} x Gix^{-})F1 animals, a full complement of the Gv-2 product is

synthesized while half complements of wild-type and defective Gv-1 products are made. Half the resulting transcription complexes will thus be defective. If these defective complexes retain the ability to bind to proviral structural genes but fail to activate transcription, they will form competitive inhibitors for functional complexes, giving rise to the intermediate levels of antigen production described for F1 hybrid animals (Stockert et al., 1971). Tissue specific proviral transcription and the differential response of individual transcription units to the action of Gv-1 indicate additional complexitites to the control of endogenous retroviral gene expression.

EXPERIMENTAL PROCEDURES

Construction of cDNA libraries

Polyadenylated RNA from epididymis, liver, thymus, and spleen of 4–6 wk old male 129 Gix^+ mice was isolated as described in Chapter 2. The generation of cDNA recombinants was performed essentially as described by Efstratiadis et al. (1976) and Wickens et al. (1978). First strand cDNA synthesis was conducted, after denaturation of the RNA with 2.5 mM CH_3HgOH (Bailey and Davidson, 1976), by oligo(dT) priming and synthesis with AMV reverse transcriptase (James Beard, Life Sciences) in the presence of 30 ug/ml actinomycin D and 1 mM deoxynucleotide triphosphates at pH 8.4 and 42°C for 1 h. The complementary strand was synthesized, after boiling the RNA/cDNA hybrids in 0.1 mM EDTA, by self priming using the large fragment of E. coli DNA polymerase I in a solution containing 0.25 mM deoxynucleotide triphosphates at 15°C for 30 min, followed by incubation at 37°C for 1 h. Double stranded cDNA was digested with S1 nuclease, size fractionated on agarose gels, recovered by binding to glass powder after melting the agarose in NaI (Vogelstein and Gillespie, 1979), and inserted into the Pst I site of pBR322 by homopolymer tailing (Villa-Komaroff et al., 1978; Rowekamp and Firtel, 1980; Michelson and Orkin, 1982). Recombinant cDNA was used to transform E. coli strain MC1061 (Casadaban and Cohen, 1980) using $CaCl_2$ essentially as described by Dagert and Ehrlich (1979), and tetracycline resistant colonies were screened for MuLV related sequences by the method of Hanahan and Meselson (1980). Positively hybridizing colonies were rescreened at low density, and plasmid DNA was isolated by the procedure of Birnboim and Doly (1979). Clones were initially characterized by restriction enzyme mapping and by blot hybridization (Southern, 1975) or in situ hybridization in agarose gels (Shinnick et al., 1975) to known restriction fragments from MoMuLV.

S1 nuclease protection

Nuclease protection analysis of RNA transcripts was performed by the method of Berk and Sharp (1977) as modified by Weaver and Weissman (1979). In brief, restriction fragments from cDNA clones were end labeled at 5' termini by using T4 polynucleotide kinase and [^{32}P]rATP or at 3'

ends by using T4 DNA polymerase and the appropriate ^{32}P-labeled deoxynucleotide triphosphates. Labeled fragments were recut with another restriction enzyme, and the appropriate fragment was recovered following agarose gel electrophoresis in order to generate DNA fragments labeled on only one strand (Maxam and Gilbert, 1980). All labeled fragments, with the exception of the probe used for the experiment depicted in Figure 8 B, contained sufficient plasmid sequence at their non-labeled end to produce a discernable gel mobility difference between any self annealed probe and the full cDNA insert protected by heteroduplex.

Labeled DNA fragments (10^5 cpm) and eukaryotic RNA (1-5 ug of polyA$^+$ or 50 ug of total cytoplasmic RNA) were mixed together with enough yeast RNA to give a final mass of 50 ug of nucleic acid, denatured by heating, and hybridized in a solution containing 400 mM NaCl, 40 mM PIPES, 1 mM EDTA, and 80% formamide (Casey and Davidson, 1977), in a volume of 30 ul, for 3-5 h at a temperature empirically determined for each probe to give maximal amounts of heteroduplex formation with minimal DNA duplex formation (generally 45^{o}C-55^{o}C). Hybrids were digested at 37^{o}C for 30 min with S1 nuclease at a concentration of 100 units per ml in a solution of 0.28 M NaCl, 50 mM sodium acetate, 4.5 mM Zn_2SO_4, pH 4.6, containing 20 ug/ml single stranded DNA from salmon sperm. Protected DNA fragments were resolved on denaturing polyacrylamide gels, and sizes were estimated from the mobility of end labeled restriction fragments of known size from the plasmid pBR322.

DNA sequencing

For sequencing, cDNA restriciton fragments were isolated from agarose gels and subcloned into the single stranded phage vector, M13 (Messing and Vieira, 1982; Sanger et al., 1980) and sequenced by the base specific chain termination method of Sanger et al. (1977) by standard procedures using the large fragment of E. coli Pol I, dideoxynucleotide triphosphates, [^{32}P]dATP or [^{32}P]dCTP, and incubations at 37^{o}C for 15 min. Labeled fragments were resolved on thin, denaturing polyacrylamide gels (Sanger and Coulson, 1978) employing buffer gradients (Biggin et al., 1983). In general, a 50 cm x 20 cm x 0.35 cm gradient gel would

resolve a sequence ladder of approximately 300 nt from six M13 clones.

FIGURE 1

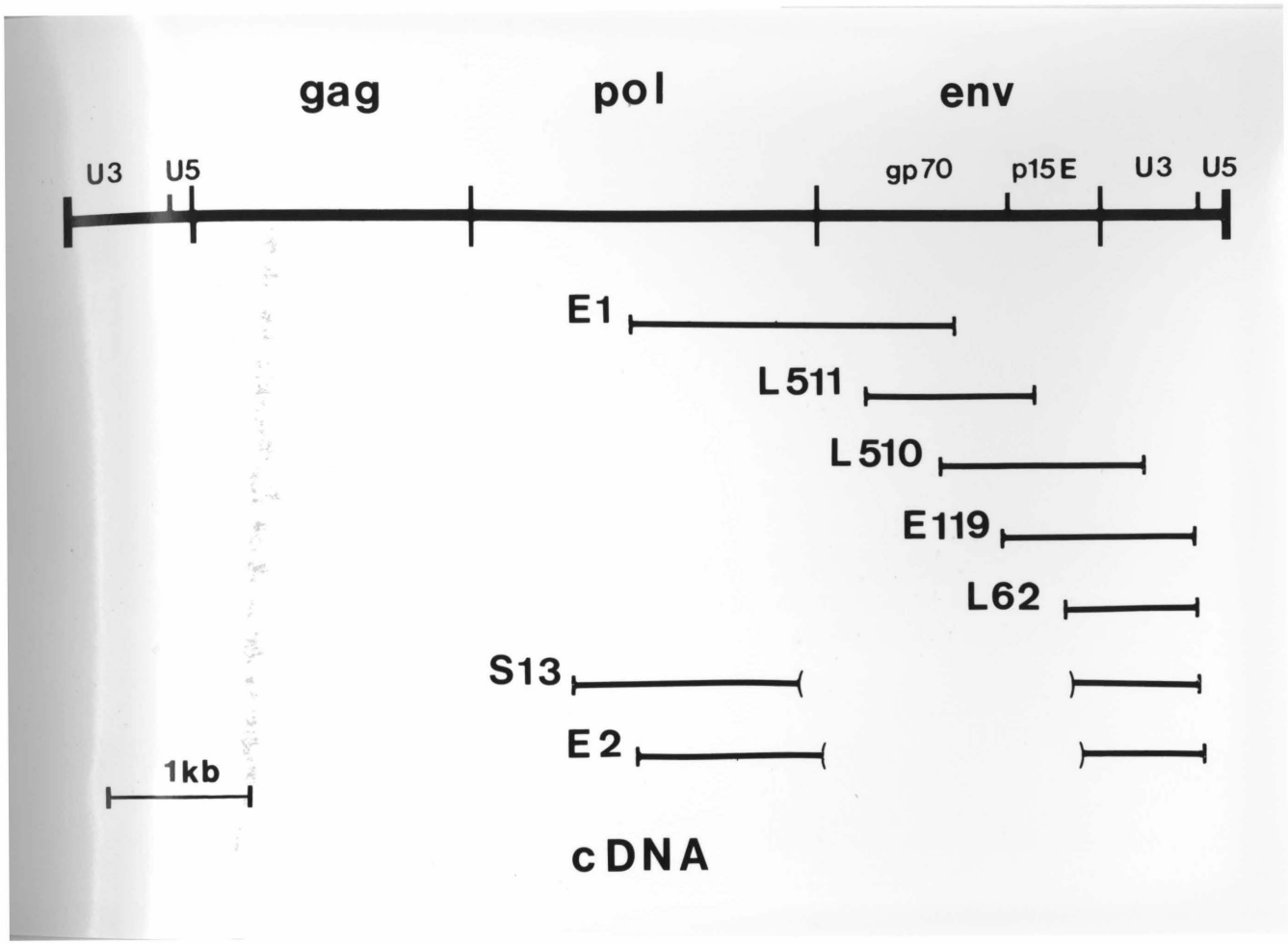

gag
pol
env
U3
U5
gp70
p15E
U3
U5
E1
L511
L510
E119
L62
S13
E2
1kb
cDNA

FIGURE 1

STRUCTURE OF RETROVIRAL cDNA CLONES

The cDNA inserts of retroviral clones derived from 129 Gix^+ polyadenylated RNA were aligned with a map of MoMuLV provirus by restriciton enzyme mapping, hybridization and DNA sequencing. Vertical bars indicate positions of junctions of eukaryotic sequences with plasmid DNA. Parentheses enclose regions with deletions of retroviral sequence. Clone S13 exhibits del env-1, and clone E2 exhibits del env-2 (see text). Nucleotide sequences from clones E1, L511, L510, L62, S13, and E2 are presented in Chapter 4. The designation of each clone includes an initial letter indicating the tissue from which the template RNA was derived. E, epididymis; L, liver; S, spleen.

FIGURE 2

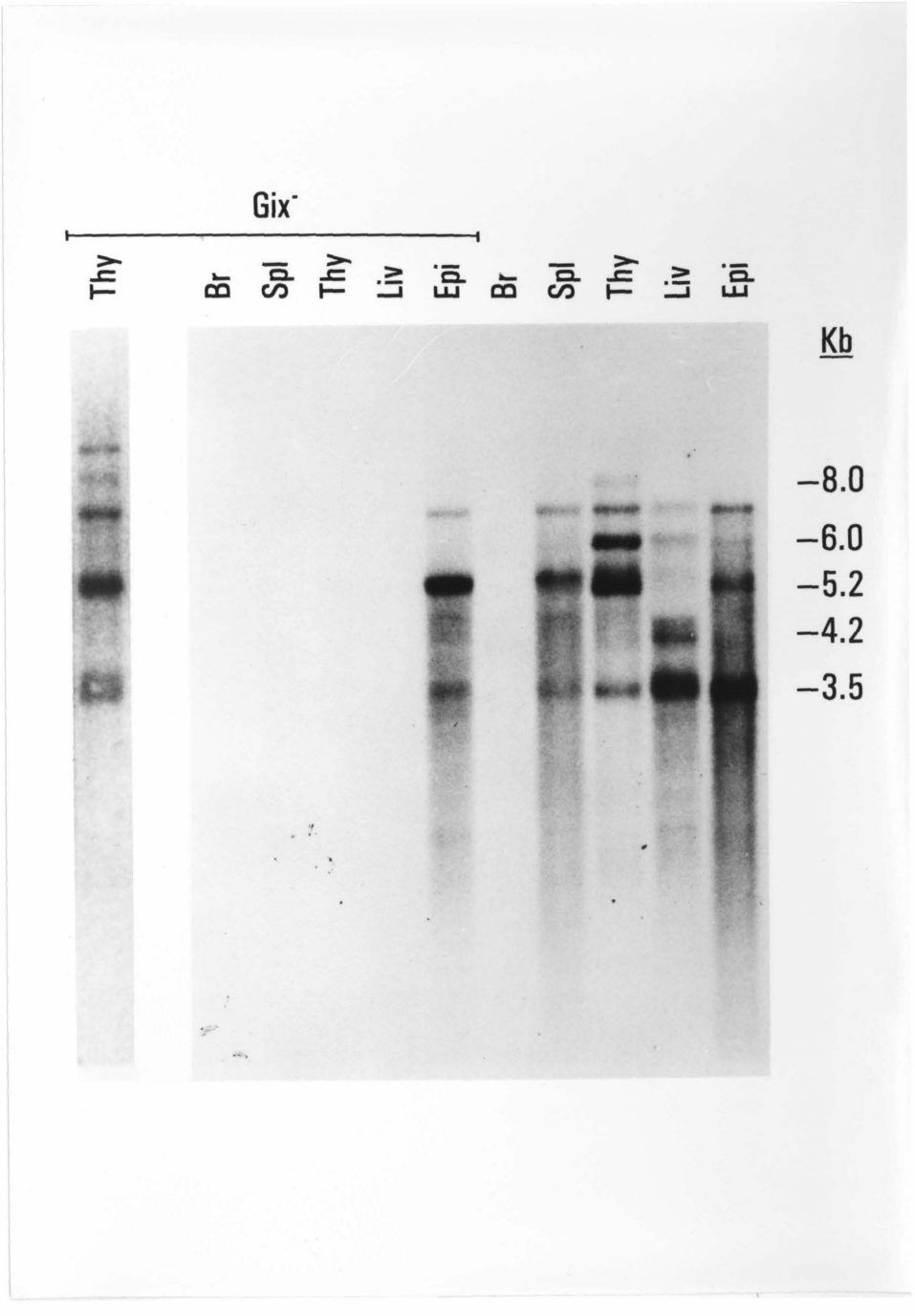


FIGURE 2

BLOT HYBRIDIZATION OF Gix^+ AND Gix^- RNA

Polyadenylated RNA derived from the indicated tissues of 129 Gix^- (left 6 lanes) or 129 Gix^+ (right 5 lanes) animals was electrophoresed on a 0.9% denaturing agarose gel, transfered to nitrocellulose, and hybridized with the cDNA insert fragment derived from clone E1. 5 ug of Gix^- RNA and of Gix^+ brain RNA were loaded per lane, while 2.5 ug of Gix^+ spleen, thymus and liver RNA and 1 ug of Gix^+ epididymis RNA were used. The thymus lane at the far left was exposed for autoradiography 10 fold longer than the other lanes. Br, brain; Spl, spleen; Thy, thymus; Liv, liver; Epi, epididymis.

FIGURE 3

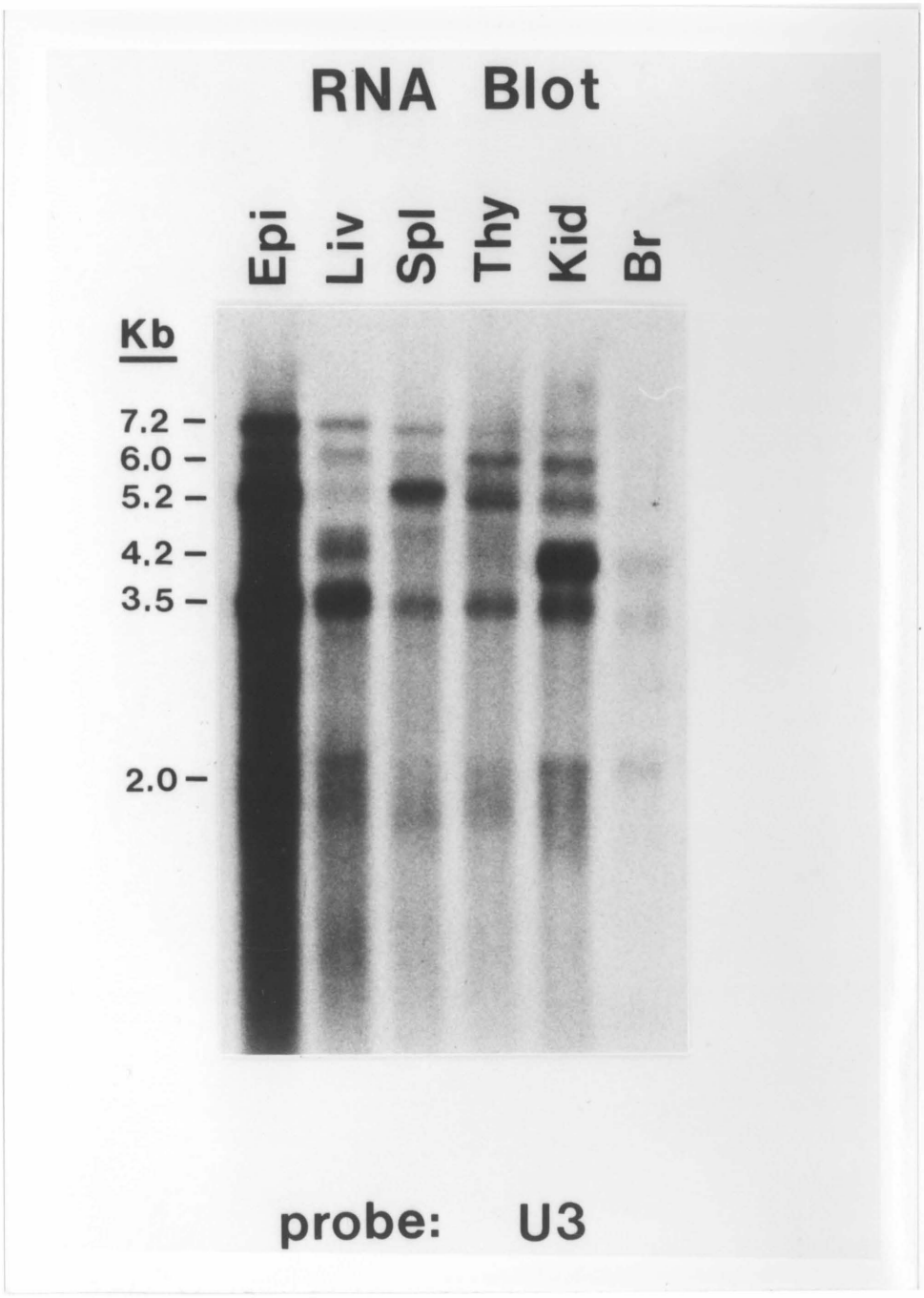


FIGURE 3

ENDOGENOUS RETROVIRUS RELATED TRANSCRIPTS CONTAIN U3 REGIONS

Polyadenylated RNA from 129 Gix^+ mice (2.5 ug/lane) was fractionated on a denaturing 1.1% agarose gel and analyzed by blot hybridization as described in Figure 2. The probe used was derived from the U3 region of clone E2 (the 600 nt Pst I to Sma I fragment presented in Figure 5).

FIGURE 4

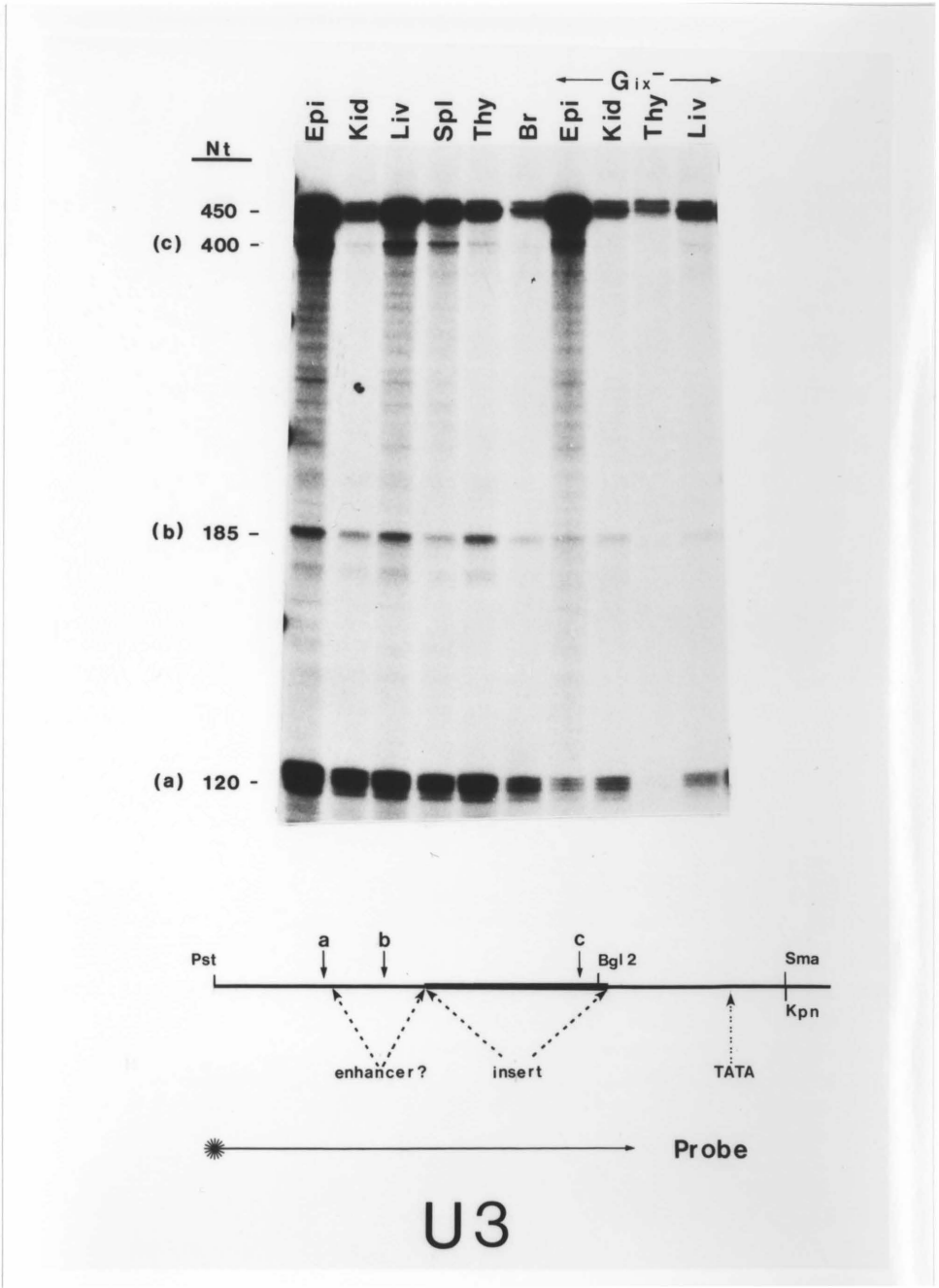

Epi
Kid
Liv
Spl
Thy
Br
Epi
Kid
Thy
Liv
Gix-
Nt
450 -
(c) 400 -
(b) 185 -
(a) 120 -
Pst
a
b
c
Bgl 2
Sma
Kpn
enhancer?
insert
TATA
Probe
U3

FIGURE 4

S1 NUCLEASE PROTECTION DEMONSTRATES U3 POLYMORPHISM

Polyadenylated RNA from 129 Gix^+ and 129 Gix^- animals was analyzed by S1 nuclease protection using a U3 probe derived from clone E2. 1 ug of Gix^+ epididymis RNA, 2.5 ug of Gix^+ kidney, liver, spleen, and thymus RNA, and 5 ug of Gix^+ brain and Gix^- epididymis, kidney, thymus, and liver RNA were hybridized with the end labeled probe and digested with S1. The resulting labeled fragments were displayed on a 6% acrylamide-7 M urea denaturing gel. The probe was a 450 nt Pst I to Msp I partial digestion product from the U3 region of clone E2 subcloned into a plasmid vector. It was end labeled on the complementary strand at the Pst I site by using T4 DNA polymerase and [^{32}P]dATP and [^{32}P]dCTP. The actual probe was 500 nt long due to the retention of plasmid sequences at the non-labeled end. Therefore, the 450 nt fragment on the gel is due to a protected heteroduplex of the entire cDNA sequence rather than to DNA duplex formation. The approximate locations of the inferred cleavage sites, producing the observed protected fragments, are indicated on the map diagram along with other features of the U3 region (see text).

FIGURE 5

```
              10         20         30         40         50
E2    CTGCAGTAAC GCCATTTTGC AAGGCATGAA AAAGTACCAG AGCTGAGTTC
L62   .......... .......... .......... .......... ....A.....

              60         70         80         90        100
E2    TCAAAAGTTA CAAGGAAGTT CAGTTATAGA TTAACAGTTA AAGATCAAGG
L62   .......... ....A..... ......A... .......... .....T....

             110        120        130        140        150
E2    CTGAATAGCA CTGGGACAGG GGCCGAACAG GATATCGGTG GTCAAGCACC
L62   .......AT. .......... ....A.---- -......... ..........

             160        170        180        190        200
E2    TGGGCCCCGG CTCAGGGCCA AGAACAGATG GTTCTCAGAT AAAGCGGAAC
L62   .......... .......... .......... ......---- ----------

             210        220        230        240        250
E2    CAGCAACAGA CACACAGAAA GCCCCGATAG ACGTCA--GT TAGCAGAACT
L62   ---------- ---------- --------.. ......GT.. ..........

             260        270        280        290        300
E2    AGCTTCACTG ATTTAGAAAA ATAGAGGTG- CACAGTGCTC TGGTCACTCC
L62   .......... .......... .........T .......... ...C......

             310        320        330        340        350
E2    TTGAACCTGT GTGTCTGCCA ATGTTCTGAC CAGATATGTG CCCATTGCTG
L62   .......... .......... .......... ...G.G.... .......T..

             360        370        380        390        400
E2    AACCTTCATT AGACTCTTTC CTTGTACCCC TCCCCTACCC ATTTCTTGAA
L62   .......... ....C..C.. ..C....... ....A..... .........G

             410        420        430        440        450
E2    AATAGACATT GTTTAGATCT AAAAAGTCCC ACCTCAGTTT CCCCAAATGA
L62   .......... .......A.. .......... .......... ..........

             460        470        480        490        500
E2    CCGGGAAATA CCCCAAACCT TATTCGAACT AACCAACCAG CTCGCTTCTC
L62   ...A...... ......G... .......... .......... ..........

             510        520        530        540        550
E2    GCTTCTGTAA CCGCGCTTTT TGCCCCCAGC CCTAGCCCTA TAAAAAGGGT
L62   .......... .......... ...T...... ..C....... ..........

             560        570        580        590        600
E2    AAAAACTCCA CACTCGGCGC GCCAGTCCTC CGATAGACTG AGTCGCCCGG
L62   .......... .......... .......... .......... C.........

E2    GTACC
L62   .....
```

FIGURE 5

SEQUENCES OF TWO DIFFERENT ENDOGENOUS U3 REGIONS

Sequences from the cDNA inserts of clones E2 and L62 were determined by the method of Sanger et al. (1977) after subcloning into M13 phage vectors. The sequence presented represents a portion of U3 from a Pst I site near the U3-p15E coding region junction to a Kpn I site within R, the viral terminal redundancy (see Chapter 4). The sequence of clone E2 is displayed. Nucleotide differences in the L62 sequence are indicated below the E2 sequence, indentical nucleotides are indicated by dots (.), and gaps introduced to increase the alignment are indicated by dashes (-).

FIGURE 6

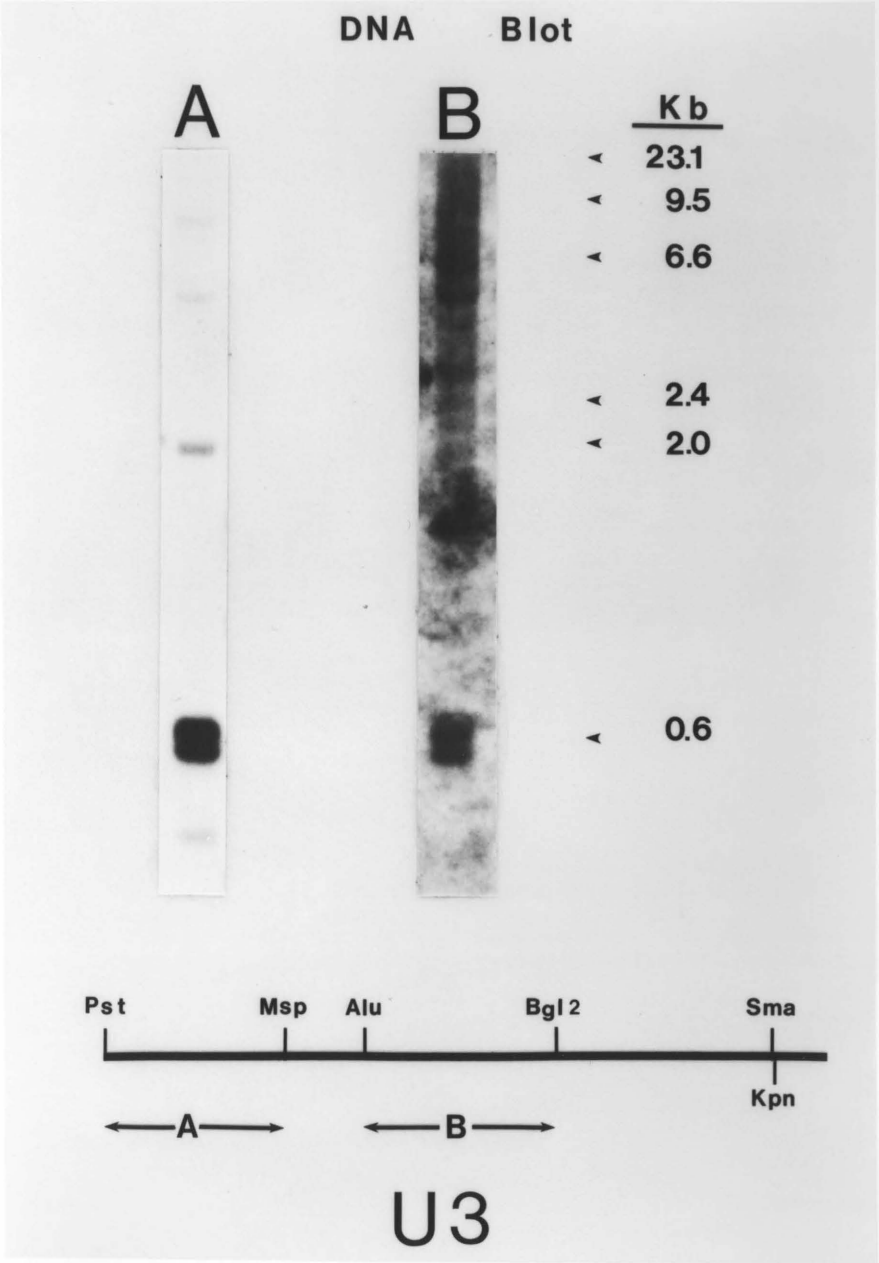


FIGURE 6

TWO MAJOR CLASSES OF U3 HOMOLOGOUS SEQUENCES IN THE 129 GENOME

High molecular weight genomic DNA (5 ug) from liver of 129 Gix^+ mice was digested with the restriction endonucleases Pst I and Kpn I, fractionated on a 1.2% agarose gel, and analyzed by blot hybridization (Southern, 1975). Lane A was hybridized with the 160 nt Pst I to Msp I fragment and lane B with the 170 nt Alu I to Bgl II fragment from clone E2 indicated in the diagram. These fragments represent regions 1–160 and 251–419 in the E2 sequence presented in Figure 5.

FIGURE 7

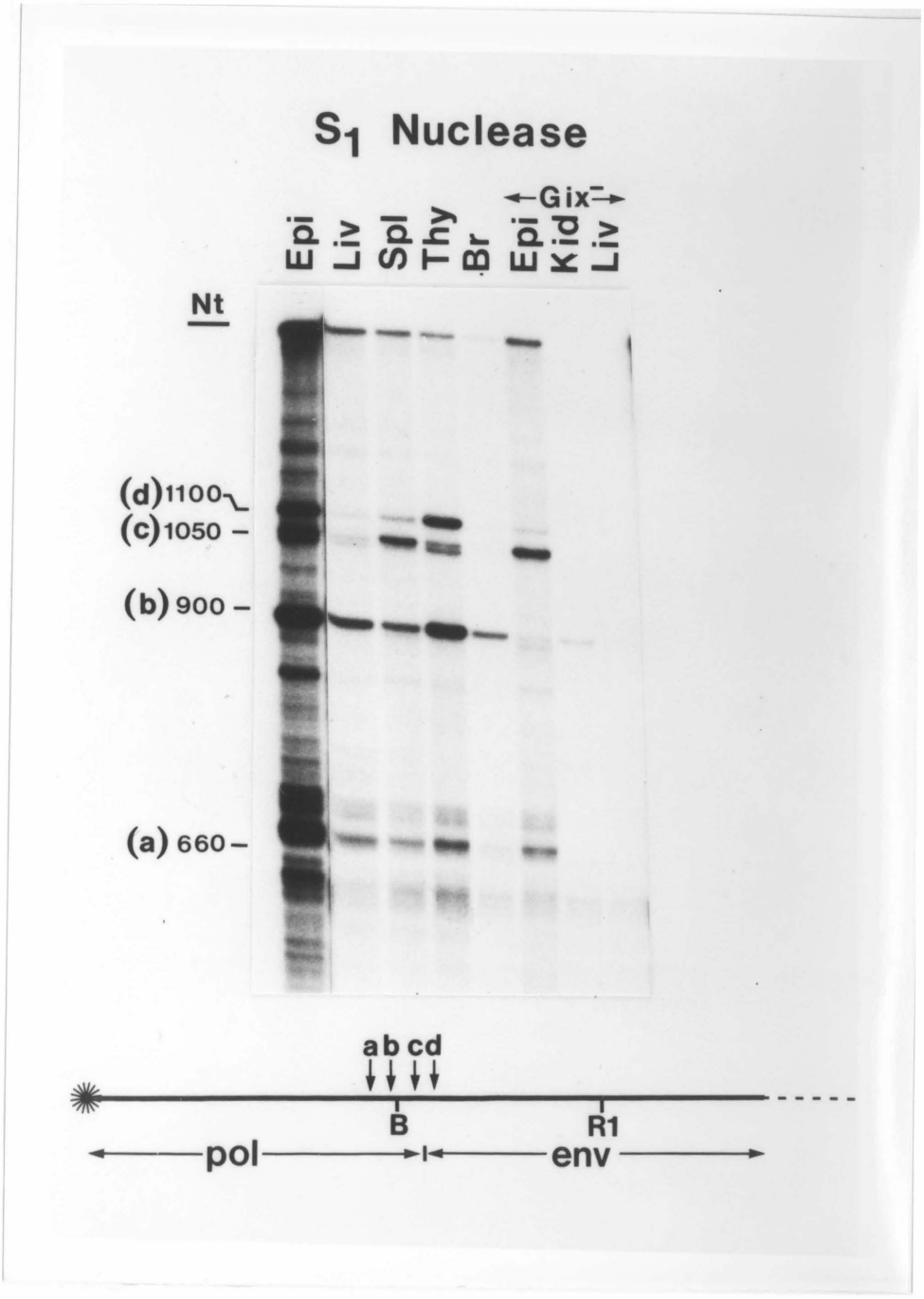

S1 Nuclease
Epi
Liv
Spl
Thy
Br
Epi
Kid
Liv
Gix
Nt
(d) 1100
(c) 1050
(b) 900
(a) 660
a b c d
B
R1
pol
env

FIGURE 7

TRANSCRIPTS FROM DIFFERENT TISSUES EXHIBIT DIFFERENT DELETION ENDPOINTS AT THE pol-env JUNCTION

S1 nuclease protection was performed as described for Figure 3. The probe contained 2 kb of cDNA sequence from clone E2, labeled at an Xba I site in the pol region, plus 750 nt of vector sequence beyond the env region. The diagram indicates the approximate positions of cleavages that could result in the observed protected fragments. Also shown are two restriction sites of clone E1. B, BamH I; R1, EcoR I.

FIGURE 8

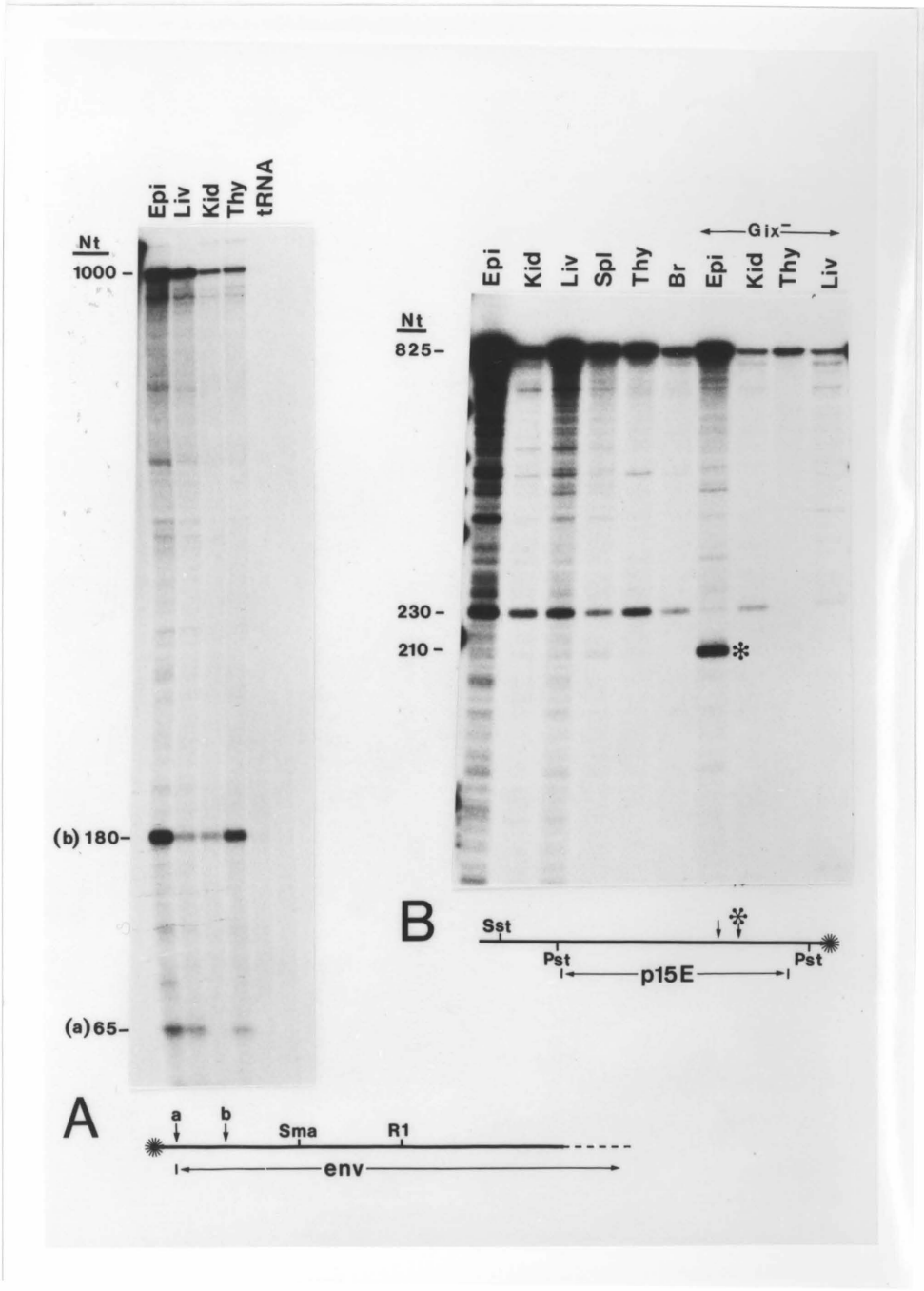

Epi
Liv
Kid
Thy
tRNA
Nt
1000 -
(b) 180-
(a) 65-
A
a
b
Sma
R1
env
Epi
Kid
Liv
Spl
Thy
Br
Gix⁻
Epi
Kid
Thy
Liv
Nt
825-
230 -
210 -
B
Sst
Pst
p15E
Pst

FIGURE 8

DELETION ENDPOINTS NEAR THE pol-env AND env-U3 JUNCTIONS

S1 nuclease protection experiments were performed as described. The hybridization probe used in Panel A contained 1000 nt of cDNA sequence from clone L511, labeled at the unique BamH I in this insert (see Chapter 4), and retained 100 nt of vector sequence at the unlabeled end (indicated by dashes on the diagram). The probe used in Panel B was an 825 nt Rsa I fragment, derived from clone L510, and end labeled by using T4 polynucleotide kinase and [^{32}P]rATP. The location of a potential cleavage to produce the 230 nt protected fragments is indicated by an arrow. The location of a cleavage site to produce the protected fragment seen only in the Gix$^-$ epididymis lane is indicated by a star. This probe did not retain any flanking plasmid DNA, but no labeled fragments were observed in the control hybridization containg only yeast RNA (not shown).

FIGURE 9

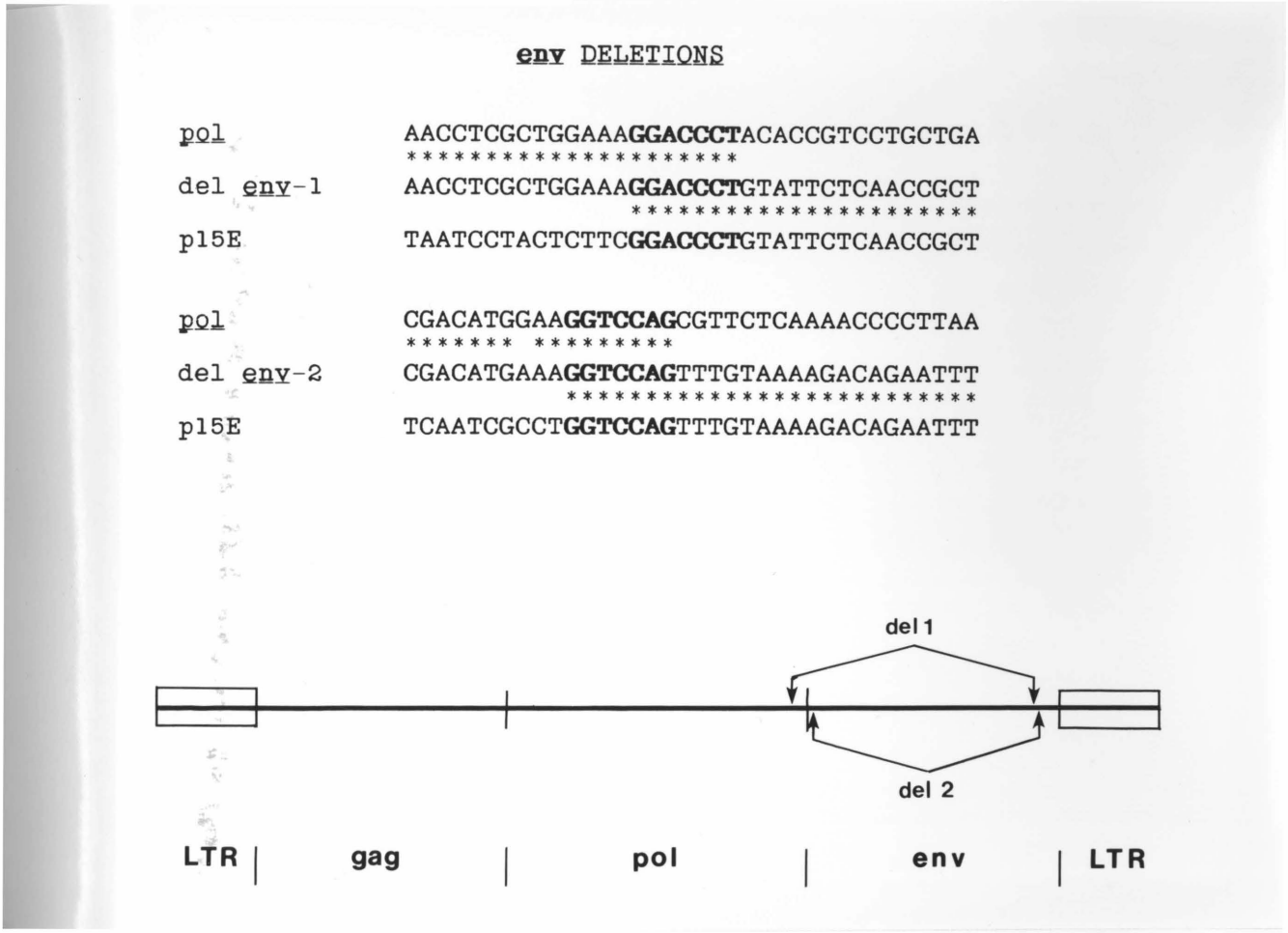

env DELETIONS
pol AACCTCGCTGGAAAGGACCCTACACCGTCCTGCTGA
del env-1 AACCTCGCTGGAAAGGACCCTGTATTCTCAACCGCT
p15E TAATCCTACTCTTCGGACCCTGTATTCTCAACCGCT
pol CGACATGGAAGGTCCAGCGTTCTCAAAACCCCTTAA
del env-2 CGACATGAAAGGTCCAGTTTGTAAAAGACAGAATTT
p15E TCAATCGCCTGGTCCAGTTTGTAAAAGACAGAATTT
del 1
del 2
LTR
gag
pol
env
LTR

FIGURE 9

DELETIONS MAP TO 7 bp DIRECT REPEATS

The DNA sequences at the points where pol sequence was found with p15E coding sequence are presented for del env-1 (clone S13) and del env-2 (clone E2). The corresponding pol sequences and p15E coding sequences form cDNA clones of undeleted transcripts (clones E1 and L510, respectively) are also presented. The regions of the undeleted transcripts that are retained in the deleted ones are indicated by stars (*). The 7 bp direct repeats are in bold face. The single nucleotide mismatch between del env-2 and pol is due to the sequence polymorphism of endogenous retroviral transcripts (see Chapter 4 for more complete sequences). The positions of these deletions are indicated on a schematic provirus map.

FIGURE 10

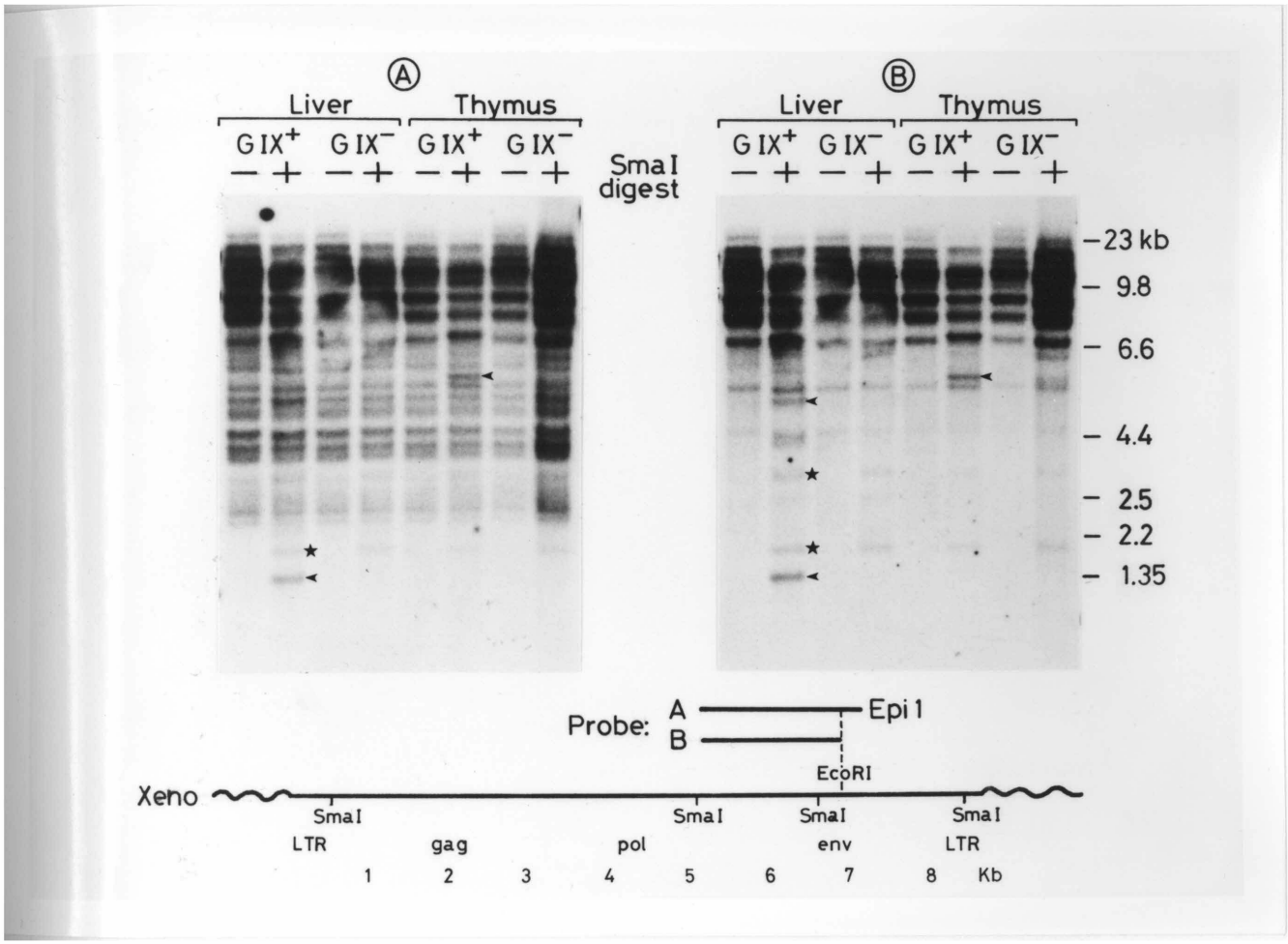

A
B
Liver
Thymus
GIX+
GIX−
SmaI digest
23 kb
9.8
6.6
4.4
2.5
2.2
1.35
Probe:
A
B
Epi 1
EcoRI
Xeno
SmaI
LTR
gag
pol
env
LTR
1
2
3
4
5
6
7
8
Kb

FIGURE 10

TISSUE SPECIFIC PATTERNS OF HYPOMETHYLATION CORRELATE WITH THE Gix PHENOTYPE

High molecular weight chromosomal DNA from liver and thymus of 129 Gix^+ and Gix^- mice was digested with the restriction endonuclease EcoR I alone or with EcoR I in combination with Sma I. The resulting DNA fragments were fractionated on a 0.8% agarose gel and transfered to nitrocellulose. Both panels represent hybridizations with the cDNA insert fragment from clone E1. The probe in Panel A was the entire 2.2 kb insert, while in Panel B, a subfragment of this clone retaining only sequence 5' to the EcoR I site was used. These probes are indicated above a diagram of a xenotropic virus restriction map (Chattopadhyay et al., 1980). The hybridizations in both panels used the same nitrocellulose blot.

CHAPTER 4

NUCLEOTIDE SEQUENCE OF ENDOGENOUS RETROVIRAL TRANSCRIPTS NORMALLY EXPRESSED IN UNINFECTED STRAIN 129 MICE

NUCLEOTIDE SEQUENCE OF ENDOGENOUS RETROVIRAL TRANSCRIPTS NORMALLY EXPRESSED IN UNINFECTED STRAIN 129 MICE

INTRODUCTION

Endogenous murine leukemia viruses are type C mammalian retroviruses integrated into the murine germline which are vertically transmitted by Mendelian inheritance and serve as the genetic pool for exogenous infectious virus. These proviruses, bounded by long terminal repeats (LTR), contain a genome of approximately 8 kb, colinear with viral genomic RNA, containing three distinct genes in the order 5'-gag-pol-env-3' which encode the viral structural polypeptides, enzymatic functions, and envelope spike proteins of the virus, respectively (see Bishop, 1978). In inbred mice these retroviruses display three distinct host ranges, mediated through their respective viral envelope glycoproteins by receptor mediated recognition. Ecotropic, xenotropic and dualtropic viral isolates have been classified due to their ability to infect only mouse cells, only non-mouse cells, or cells of both mouse and non-mouse origin, respectively. The third class, dualtropic viruses, are also referred to as mink cell focus (MCF) inducing because they are often capable of leading to characteristic alterations in the cellular morphology of infected mink cells. Ecotropic and xenotropic viruses can be induced directly from proviruses integrated into the genome of various mouse strains. Dualtropic (MCF) viruses, however, are formed by recombination between exogenous, infectious ecotropic virus and unknown, endogenous viral sequences related to xenotropic virus, resulting in the replacement of variable portions of the 3' end of the ecotropic genome with endogenous viral information encoding altered envelope proteins. It is viruses of this type which are generally associated with the development of leukemia (see Risser et al., 1983).

Sequences homologous with all three viral types have been detected in the murine genome, and the pattern of spontaneous expression of retroviral sequences also varies widely among different mouse strains. The inducibility of retroviruses in inbred mice is governed by a variety of

genetically defined loci including the proviral structural genes as well as regulatory genes distinct from the induced virus (Kozak and Rowe, 1980, 1982; Horowitz and Risser, 1982). Similarly, the expression of viral antigens in the absence of infectious virus is strain specific, perhaps also as a result of regulatory gene control (see Chapters 2 and 3). Considerable heterogeneity exists among different inbred strains in the levels and in the structure of these normal products. Serologically and structurally distinct molecules as well as membrane bound and secreted forms of viral proteins are found in thymus, spleen, liver, and epididymis of normal animals (Old and Stockert, 1977).

In the study of endogenous retrovirus expression in 129 mice, heterogeneity in the size and structure of retrovirus related products have been detected (see Chapter 3). Genetic analysis of this system, facilitated by use of the congenic partner strain 129 Gix^- that is characterized by the absence or reduction of these virus related gene products (Stockert et al., 1971, 1975), has suggested the coordinate regulation of the expression of these polymorphic products. In order to determine the structural basis for this heterogeneity, to document the number of independent transcription units that are encoding these transcripts, and to define the relationship between these normal cellular components and exogenous viruses, the nucleotide sequence of cDNA clones corresponding to endogenous viral sequences expressed in 129 Gix^+ mice were determined. Polymorphism previously detected in the RNA population was found to be due to scattered single nucleotide differences between different transcripts as well as deletions of varying lengths of retroviral sequences. These findings strongly implicate multiple genomic transcription units in the origin of these regulated RNA species. Furthermore, the nucleotide sequences establish that these normal gene products are most likely encoded by endogenous retroviral genomes rather than by homologous cellular genes, such as is the case for the normal oncogene counterparts of acutely transforming retroviruses (Bishop, 1983). Comparisons with known viral sequences has revealed a striking similarity between these normally synthesized RNA species and the sequences of the substituted region of leukemogenic MCF dualtropic viruses, suggesting a possible relationship between these transcriptionally active proviruses and the

generation of recombinant virus.

RESULTS

Identification of retroviral cDNA clones

Libraries of cDNA clones were generated by oligo(dT) primed DNA synthesis from total, polyadenylated RNA from tissues of 129 Gix^+ mice. Retroviral sequences were selected by colony hybridization using nick translated DNA probes. Initially, a library derived from epididymal mRNA, previously shown to have a high abundance of polyadenylated transcripts homologous to retroviral probes (Levy et al., 1982), was screened with probes from the pol and env regions of Moloney murine leukemia virus (MoMuLV; see Chapter 2). Clones E1 and E2 (Figure 1) identified from this screen were characterized by restriction enzyme mapping and by hybridization to a molecular clone of MoMuLV (Berns et al, 1980) and found to contain sequence homologous to pol-env regions and pol-U3 regions, respectively, of MoMuLV. The eukaryotic insert sequence of these plasmids was purified and subsequently used to re-screen the epididymis cDNA library and to screen additional libraries prepared from oligo(dT)-generated cDNA from thymus, spleen, and liver RNA from 129 Gix^+ animals. These clones were assigned names with the initial letter indicating the tissue of origin (see legend to Figure 1).

Several clones selected in this manner were characterized, and their cDNA insert sequences are illustrated in Figure 1. Because of the potential heterogeneity of retroviral transcripts, overlapping cDNA clones cannot automatically be assumed to represent a contiguous mRNA species. Although no single clone contained a cDNA sequence representing an entire pol gene or env gene open reading frame, sufficient overlaps of identical sequence existed between clones to allow us to construct a reasonable estimate of a cumulative structure for the 3' end of these endogenous viral transcripts. Portions of these cDNA inserts were subjected to DNA sequence analysis by the base-specific chain termination method (Sanger et al, 1977) after subcloning into the single stranded bacteriophage, M13 (Messing and Vieira, 1982; Sanger et al, 1980). The sequences obtained for these clones were compared with each other and with the corresponding sequences of AKR endogenous viral clone A-12 (Khan, 1984), AKR MCF 247 virus (Kelley et al, 1983; Holland et al, 1983), Moloney

MCF virus (Bosselman et al, 1982), and AKV virus (Lenz et al, 1982; Herr, 1984) by using the computer program ALIGN (Orcutt et al., 1982). AKV is an infectious ecotropic virus from AKR mice and is similar in structure to ecotropic viruses from many strains of mice. Both AKR MCF 247 virus and MoMCF virus are dualtropic, leukemogenic viruses isolated from thymomas. They were derived by recombinations between ecotropic viruses (AKV and MoMuLV, respectively) and undefined endogenous sequences, replacing a 3' portion of the pol gene and the 5' portion of the gp70 coding region of the ecotropic virus (see Figure 1). MCF 247 also derived its extreme 3' end from non-ecotropic information (Kelley et al., 1983). Clone A-12 is a retrovirus related genomic sequence from a lambda phage recombinant library of Eco RI genomic restriction fragments from AKR mice and may represent the endogenous parent for MCF viruses such as 247 and MoMCF (Kahn, 1984).

The pol gene

Initial restriction analysis indicated that clones E1, E2, and S13 contained sequences hybridizing pol region restriction fragments of MoMuLV and shared restriction enzyme cleavage patterns indicating a close homology to each other (not shown). However, they all three clearly contained unique sequence as well. For example, although cDNA inserts E1 and E2 shared an Xba I site near their 5' ends and a BamH I site at the 3' end of the pol region, clone E2 contained cleavage sites for Pvu II and Hind III within pol not present in the corresponding region of E1. These two clones also differed in the pattern of Sau3A I restriction fragments (data not shown). This indicated that they might reflect RNA transcript heterogeneity within a synonymous region of the viral genome.

The nucleotide sequence of 360 nt from the 3' end of the pol gene of clone E1 and the sequences from corresponding regions of clones E2 and S13 are presented in Figure 2 along with the predicted amino acid sequence of clone E1. The sequence heterogeneity of the endogenous retroviral transcripts of 129 Gix^{+} mice is clearly illustrated by these clones. First, sequence heterogeneity is seen within the overlapping regions of these clones. Of the 315 nucleotides aligned between clones E1 and E2, there are 15 mismatches and one gap of 2 nt (position 57-58),

or 4.8% difference. Between clones E1 and S13, there are 3 mismatches and between clones E2 and S13, there are 5 differences, resulting in differences of 3.2% and 5.3%, respectively. Further, clones E2 and S13 exhibit large deletions which eliminate the 3' portion of the pol gene represented here. These deletions extend through the majority of the env gene resulting in the fusion of pol sequence with the 3' end of the p15E coding sequence (see Figure 4). The actual endpoints for these deletions are ambiguous due to 7 bp direct repeats at these deletion points in the undeleted sequence (see Chapter 3, Figure 9). The deletion in S13 extends from the sequence GGACCCT at position 178-184 until the second copy of this sequence found at position 478-484 in p15E (Figure 4), retaining a single copy of the repeat and thereby deleting 1977 nt. A similar deletion in E2 begins at nucleotide 316 in Figure 2, following the sequence GGTCCAG, and likewise extends through the second copy of this repeat at position 501-507 of p15E, thus eliminating 1817 nt. These deleted cDNA clones are not likely to be artifacts of in vitro cDNA synthesis since analogously deleted transcripts have been detected by S1 protection analysis of the RNA population (see Chapter 3).

For comparison, viral pol region sequences from an AKR endogenous genomic clone (A12), AKR MCF 247 virus (247), Moloney MCF virus (MoMCF) and AKV ecotropic virus (AKV) are also shown in Figure 2. This region of clone E1 predicts an open reading frame of 120 amino acids, ending with an ochre codon immediately following nucleotide 360 (nucleotides 59-61 of Figure 3). This open reading frame overlaps the env open reading frame for 58 nt encoding 19 amino acids of pol and 19 amino acids of gp70 in different triplet reading frames (see Figure 3). The methionine initiation codon for env begins with nucleotide 303 in Figure 2. Such sequence overlap at the pol-env junction has also been found for all sequenced murine type C viruses (e.g., MoMuLV, AKV, and MoMCF). The sequences of this region also differ with respect to the viral sequences presented in Figure 2 (see Table 1). Clone E1 differs from the AKR sequences A12 and 247 (which are identical in this region) by 3 nucleotides out of 360 nt, leading to a single predicted amino acid change (glutamine to leucine at position 85-87). E1 differs from MoMCF at 36 of 360 nt resulting in 7 predicted amino acid changes and differs

from AKV at 53 of 348 nt, requiring the introduction of a gap of 4 codons to increase sequence similarity. Thus, the E1 sequence differs by 10% from MoMCF and by 15% from AKV. It should be noted that MoMCF contains both ecotropic and endogenous information in this region of pol since the recombination probably took place at position 210 in Figure 2 (Bosselman et al., 1982). All the differences between E1 and MoMCF occur 5' to this probable recombination point between MoMCF and the parental MoMuLV. Thus, the substituted regions of pol in both of these recombinant MCF viruses are virtually identical with the E1 sequence, even though these viruses arose in different mouse strains (AKR and Balb/c) and were derived from distinct ecotropic parental viruses (AKV and MoMuLV).

The potential 3' splice acceptor sequence (CACTTACAG/G) (Mount, 1982) is conserved in all six sequences at position 16-25 (Figure 2). Its use could result in the 3.5 kb spliced env coding mRNA. Transcripts with a structure consistent with such a splice have been detected by RNA blotting and S1 nuclease protection (Chapters 2 and 3). In experiments using a probe derived from the insert from clone E1, end labeled at the BamH I site at position 239-244 in Figure 2, a fragment of approximately 225 nt was protected after annealing with RNA from epididymis, thymus, and liver from 129 Gix^+ animals (not shown). This is consistent with a significant number of transcripts utilizing a splice site in this region. Significantly, however, the sequences at the deletion endpoints of clones E2 and S13 poorly match splice consensus sequences. The RNA species responsible for these clones were probably not generated by RNA splicing but rather more likely were a direct result of deletions mediated by the 7 nt direct repeats (see Chapter 3).

The env gene

No single cDNA clone was isolated which spanned the entire 3.5 kb env transcript nor the entire env coding portion of a spliced or of a genomic RNA species. However, the combination of the overlapping clones E1, L511, and L510 allowed us to construct a contiguous picture of this region. Because these clones share extensive regions of sequence overlap with 100% nucleotide identity over the sequenced regions, we are

reasonably confident that the sequence presented reflects that of a single, continuous RNA transcript. Two other, independent clones from a portion of this region, E119 and L62, also share 100% identity with the sequence presented over the 600 nt and 200 nt, respectively, for which their sequences were determined (not shown). This sequence contains an open triplet reading frame of 1932 nt predicting a potential protein sequence of 644 amino acids, representing the gp70-p15E polyprotein precursor.

Structure of gp70. The sequence presented in Figure 3 was derived from clones E1 and L511, which share an identical sequence for 599 overlapping nucleotides, and may thus represent a single RNA transcript. The env precursor polyprotein is assumed to be cleaved into gp70 and p15E by a process analogous to that thought to occur for MoMuLV (Shinnick et al., 1981), predicting a trypsin-like cleavage following the lysine-arginine doublet at the end of the presented sequence. Thus, the gp70 region would encode a 441 amino acid polypeptide. No confirming protein sequence data are available for the carboxyl terminus of endogenous gp70. The actual mature protein may be somewhat shorter than shown due to carboxypeptidase-like activity often associated with dibasic cleavage sites (Steiner et al., 1980). The carboxy terminal residue of mature Rauscher MuLV ecotropic gp71 protein has been found to be tyrosine (Henderson et al., 1978) which would place the carboxy terminus at codon 1315-1317 in Figure 3 if the endogenous protein is similarly processed. The complete amino acid sequence determined for Friend MuLV ecotropic gp71A protein, however, indicated that the carboxy terminus of this protein is Lys-Arg (Chen, 1982).

The amino terminus of the mature protein probably also is generated by proteolytic cleavage (Shinnick et al, 1981). The first potential initiator methionine codon is within the last 19 codons of the pol open reading frame, in a reading frame shifted with respect to the pol frame. However, the mature protein likely begins with the valine residue at position 97-99 since the 25 amino acids beginning at that position match 24 of the first 25 amino acids determined for the amino terminus of Rauscher MCF gp69 protein (Schultz et al, 1983). This portion of Rauscher MCF gp70 was derived, following recombination, from endogenous

viral information. The 31 predicted amino acids between the initiator methionine and this potential amino terminus could form a typical membrane protein signal peptide (Blobel and Dobberstein, 1975; Engelman and Steitz, 1981) containing a stretch of 14 uncharged, mostly hydrophobic amino acids bounded by charged residues. The predicted size of 44,200 MW predicted for the mature protein agrees well with the mobility of deglycosylated endogenous gp70 on SDS polyacrylamide gels (J. Elder, personal communication).

There are six potential attachment sites for N-linked glycosylation of the cannonical sequence Asn-X-Thr/Ser (Marshall, 1974), two at the amino end and four clustered in the carboxyl third of the molecule, similar to viral gp70 (Rosner et al, 1980). These sites are all conserved in the two MCF sequences presented. However, the single amino acid which does not match the 129 endogenous sequence presented here out of the 25 amino terminal residues that have been determined for Rauscher MCF (Schultz et al., 1983) lies within a potential glycosylation site and thus eliminates one site in the Rauscher MCF protein. Still, the MCF proteins are clearly highly related to the 129 endogenous transcript.

Another region which is potentially glycosylated in some viruses is encoded at position 1216-1224. An oligonucleotide containing this region in ecotropic virus is predictive for the Gix antigenic determinant of gp70 which depends upon differential glycosylation. Gix^+ viruses can differ from Gix^- viruses by as little as a single nucleotide at position 1216 where a guanine to adenine transition changes the encoded aspartic acid residue to asparagine, thus creating a new glycosylation site (Donis-Keller et al, 1980) for attachment of the added carbohydrate typical of Gix^- viruses (Rosner et al, 1980). The 129 endogenous transcript contains a GAC aspartic acid codon at position 1216-1218 (Figure 3) giving the sequence **Asp**-Leu-Thr rather than the carbohydrate attachment sequence of **Asn**-Leu-Thr typical of Gix^- gp70 proteins (Donis-Keller et al., 1980). Interestingly, the 129 endogenous transcript sequence differs from Gix^+ viruses such as AKV or 247 by 7 of the 30 nt Gix oligonucleotide of AKV virus, and the amino acid sequence for this and the surrounding regions differs markedly from that of Gix^+ AKV (see Figure 6). The endogenous

protein, however, presumably still retains reactivity with Gix antisera marking it as a Gix^+ gp70 (Stockert et al, 1971). These results suggest that the Gix antigenic determinant is highly dependent upon tertiary structure of the protein rather than upon linear sequence. This requisite structure must be disrupted by the added glycosylation of gp70 of Gix^- viruses.

Homology relationships. The sequence of the 129 endogenous gp70 coding region more closely resembles the gp70 coding regions of recombinant viruses than those of ecotropic or xenotropic virus. Two recombinant MCF viral sequences which show considerable similarity to this RNA sequence are presented in Figure 3; two MCF viruses induced from C3H mouse cell lines (Mark and Rapp, 1984) as well as the sequences of two recombinant spleen focus forming viruses, Friend SFFV (Amanuma et al., 1983) and Rauscher SFFV (Bestwick et al., 1984), show similarly high degrees of homology (not shown). However, in order to align the predicted endogenous gp70 sequence with that of ecotropic MoMuLV (Shinnick et al., 1981), for example, four insertions and five deletions are required to juxtapose similar regions. Similarly, alignment with the available nucleotide sequence of the 5' one-third of a xenotropic gp70 coding region (Repaske et al., 1983) requires the insertion of four codons and the deletion of one codon.

The endogenous sequence very closely resembles the AKR endogenous clone A12, which probably represents the endogenous parent of MCF viruses (Kahn, 1984). The alignment in Figure 3 shows only 4 nucleotide differences out of the 564 nt of A12 sequence available, which results in only one amino acid difference. There are 74 nt changes required when aligning the 129 endogenous transcript sequence with the 1320 nt of the 247 virus gp70 region, with 29 of those changes leading to differences in 18 amino acid codons (resulting in 95% nucleotide and 96% amino acid similarity). Furthermore, introduction of a one-codon gap is necessary to increase the alignment, indicating a 129 endogenous transcript specific amino acid of threonine at position 952-954. However, these differences which distinquish the MCF virus sequences from the endogenous sequence cluster in the 3' third of the sequence (see Figure 6), past the probable

recombination point in 247 at position 856 (Holland et al, 1983), and thus within the region derived from the ecotropic virus parent (see Figure 1). Interestingly, of the differences between A12 and clones E1 and L511, only one is shared with 247 so that within this region in which 247 is derived from endogenous information, it is more similar to the 129 sequence than to the AKR endogenous sequence. MoMCF shares all A12 differences from the endogenous transcript in this region and has three additional nucleotide differences, two of which change codons as well, for a total of five codon changes in this region. Overall, MoMCF shows 79 differences in 1320 nt (94% similarity), with 48 of them leading to 29 non-synonymous codon differences (93% amino acid similarity), as well as requiring the same one-codon gap inserted in the 247 sequence. Again, however, most change between the aligned sequences cluster at the 3' end of the sequence, past the presumed recombination point, within the ecotropic-derived region (see Figure 1).

The clustering of these differences are relevant to the mapping of potential sites of the recombinational events that presumably gave rise to MoMCF virus. Bosselman et al (1982) indicated this recombination at position 1073 in Figure 3 on the basis of similarity with MoMuLV. The data presented here, however, imply that recombination more likely took place further 5', closer to the region implicated for 247 recombination (position 856). The region 3' to this potential recombination point, therefore, may actually be ecotropic in origin, the differences between MoMCF and MoMuLV in this region having been introduced by mutation rather than by recombination. Interestingly, of the differences between MoMCF and the 129 sequence, 31 are silent while 48 lead to codon changes; in 247, 46 changes are silent while 28 lead to amino acid differences. This preponderance of coding changes may be indicative of a selective pressure favoring an altered protein sequence for MoMCF.

p15E coding region. The beginning of p15E as presented in Figure 4 is defined by analogy with Rauscher MuLV with which it shares amino terminal protein sequence. The sequence presented is from clone L510 which contains the entire p15E open reading frame 606 nt (202 amino acids predicting a protein of 22,000 MW). This clone overlaps clone

L511, and, for the approximately 400 nt determined, their sequences are identical. Furthermore, clone E119 also shares sequence with L510 for approximately 350 nt of sequence determined (data not shown). However, the overlapping portion of the clone S13 contains 4 differences out of 122 nt, and that of clone E2 shows 6 differences out of 99 nt, two of which are shared with S13. Further sequence heterogeneity was detected in a p15E coding region clone from thymus which showed 5 nt differences in the 3' half of p15E, leading to two amino acid changes (Gln to Leu at position 535-537 and Lys to Glu at position 588-600).

The 129 endogenous transcript sequences would encode a protein one amino acid longer than the p15E from 247 virus due to a codon insertion at position 364-366 resulting in the addition of a serine residue. This endogenous transcript specific codon was also present in clone E119 and in the thymus clone sequences. The endogenous transcript is three codons longer than the p15E region of MoMCF due to a second insertion, this of two codons at position 580-585 (Asp-Pro). This insertion is also present in 247 virus as well as in the feline retrovirus FeLV (Elder and Mullins, 1983). Overall, 247 virus shows 112 nt differences with 34 of them leading to changes in 19 codons (82% nucleic acid and 91% amino acid similarity) and MoMCF has 142 differences with 38 leading to changes in 24 codons (76% and 88% nucleic acid and coding similarity, respectively). AKV is almost identical with 247 over this region until the 3' terminus where 247 virus is closer to the 129 sequence and was probably again derived from a non-ecotropic parent due to another recombinational event at at about position 500 (Kelley et al., 1983; see Figure 1).

There is a single Asn-X-Thr/Ser potential glycosylation site at position 400-408 in the endogenous p15E sequence and in infectious viruses. However, it is unlikely that this attachment site is used because viral p15E is not known to be glycosylated (Rosner et al, 1980). Furthermore, this site may be buried in the cell membrane since it is at the border of a stretch of 30 uncharged, mostly hydrophobic amino acids (position 406-495) bounded by positively charged residues which probably span the membrane (Green et al., 1981; Lenz et al, 1982). The amino acid sequence of this potential membrane spanning domain (position

406-495) shows complete identity with 247 and AKV although there are still 17 nucleotide differences out of 90 nt (81% nucleotide similarity). The potential cleavage sequence (Ala-Leu/Val-Leu) identified in MoMuLV as producing the R-peptide (Green et al., 1981; Shinnick et al, 1981) is also conserved in all three sequences presented; however, the sequence of the predicted R-peptide itself (position 544-606) is highly variable in both length and sequence. This is a region previously identified as evolving rapidly in viruses (Shinnick et al, 1981; Lenz et al, 1982). Comparison of two 129 endogenous p15E coding region sequences (clone L510 and a clone derived from thymus, not shown) found that one of the two amino acid differences between these two sequences fell within the R-peptide region.

The U3 region

The two sequences presented in Figure 5 begin 1 nt 3' to the ochre codon ending the p15E open reading frame. These sequences again reveal the extent of divergence among 129 Gix^+ endogenous retroviral transcript (see Chapter 3). The two endogenous U3 regions shown differ by (i) 29 single nucleotide changes out of 658 corresponding positions (4.4% divergence), (ii) two deletions of 5 nt and 42 nt, respectively, in clone L62 relative to E2, and (iii) two deletions of 2 nt and 1 nt in E2 relative to L62. These clones represent the 3' terminal portion of the respective transcripts since both plasmids contained A/T homopolymer tracts derived from the poly(A) of the mRNA. L62 is polyadenylated precisely at the 3' end of the viral terminal redundancy, while E2 is polyadenylated after 4 further nucleotides which are identical to the first 4 nt of MoMuLV U5 region.

These two endogenous sequences contain several features typical for viral U3. By analogy to AKV (Herr, 1984), positions 1-32 contain the origin of plus-strand DNA synthesis (Gilboa et al, 1979; Sutcliffe et al., 1980), U3 extends for nucleotides 33-637, and position 638-706 contains the viral RNA terminal redundancy. Within these regions are the "CAT box" equivalent sequence CAAC at 553-556, the TATA promoter sequence 55 nt further 3', and the AATAAA polyadenylation signal at position 684-689. Despite these overall similarities, these sequences differ substantially from the corresponding regions of AKV and 247. There are

38 differences compared with 247 and 82 compared with AKV out of approximately 485 aligned nucleotides. Furthermore, numerous insertions were required to increase sequence alignment. These include the presence of a 192 nt insert (position 308-500) in the endogenous sequences which is not present in either virus. This insert, which was found to be typical for LTR regions of endogenous Balb/c provirus (Khan and Martin, 1983), is bounded by 6 nt direct repeats (interrupted in L62 by a 2 nt insertion) which are present in 247 as tandem repeats and of which only a single member of the set is present in AKV. This pattern of direct repeats is reminiscent of the action of transposable genetic elements which may be involved in the evolution of retroviral LTRs (Kahn and Martin, 1983). Furthermore, the 192 nt insertion bears close similarity to a region of a middle repetitive DNA element (IS in Figure 5), most copies of which in the 129 mouse genome are not associated with retrovirus and show evidence of movement (Wirth et al, 1983, 1984). For the approximately 210 nt which can be aligned well with E2 and L62 (Figure 5), this repetitive element shares approximately 92% sequence similarity with the endogenous transcripts.

The endogenous transcripts also differ from AKV and 247 by a 14 nt tandem duplication (position 138-165 in Figure 5), only one member of the set being present in the viruses. This duplication is just prior to the potential enhancer regions of the viruses (Varmus, 1982), a region of considerable sequence divergence (position 170-300). In particular, the sequence TGGAAAGTCCC associated with viral enhancers and considered a core sequence in the SV40 enhancer (Weiher et al, 1983) is present in both AKV (169-179) and 247 (282-292) but is absent from the two endogenous sequences. It should be noted that the large deletion in L62 relative to E2 is within this potential enhancer region. In preliminary experiments to evaluate potential promoter activity of these cloned U3 regions (see Chapter 5), we have found that E2 will replace the SV40 promoter to drive the neomycin resistance gene of the eukaryotic vector of pSV2neo (Southern and Berg, 1982). However, in an initial experiment this construct was found to result in a substantial proportion of RNA transcripts with a structure inconsistent with the use of the legitimate initiation site.

DISCUSSION

Genomes of laboratory mice contain multiple copies of DNA sequences which hybridize probes derived from murine leukemia virus (Chan et al., 1980; Jenkins et al., 1982; Chattopadhyay et al., 1982). Depending upon the strain of mouse, these sequences can be expressed to give rise to full length infectious virus (Chattopadhyay et al., 1974), to viral antigens in the absence of detectable virions (Stockert et al., 1971; Old et al., 1977; Elder et al., 1977), or to virus related RNA transcripts which may or may not encode proteins (Levy et al., 1982; Boccara et al., 1983). There is considerable sequence heterogeneity among the multiple copies of integrated endogenous proviral sequences as well as among the different RNA transcripts expressed in different strains. In our analysis of retroviral expression in 129 Gix^+ and 129 Gix^- mice, it became clear that a heterogeneous set of transcripts was being produced and that this pattern of transcription was distinct for each tissue examined (Levy et al., 1982). In order to more fully examine the regulated expression of these transcripts, it became of interest to understand the structural basis for the heterogeneity, to determine the source of the variability, and to define the relationship of these sequences to known viral phenotypes. Here we report the cloning of cDNA copies and the determination of the nucleic acid sequences of the 3' end of the endogenous viral transcripts from 129 Gix^+ mice.

Through the use of overlapping cDNA clones, we have constructed a map representing the 3' end of what is proposed to be the major full length transcript normally expressed in these mice. All portions of this sequence which have been analyzed by nuclease S1 protection (for example, see Chapter 3) were found to reflect an abundant RNA transcript. This sequence contains 360 nt of the 3' end of the pol triplet open reading frame, a 1932 nt open reading frame for the env gene product, and approximately 710 nt of two distinct 3' end U3/R structures which appear to be represented in the RNA population in about equal proportion (see Chapter 3). In addition to this U3 polymorphism, clones were obtained which provide the structural basis for previously described transcript polymorphism in 129 mice (Levy et al., 1982). This

polymorphism consists of limited sequence differences within the pol and p15E coding regions, as well as substantial deletions of the env coding capacity. This limited sequence heterogeneity signifies a highly conserved family of genes is being transcribed, indicating the recent evolutionary dispersal of these genetic elements, while the frequency of deletion events implied by these clones and the set of transcripts they represent may indicate a high propensity for homologous recombination or perhaps for replication errors during possible RNA mediated transposition events.

The single nucleotide differences observed between individual cDNA clones from the pol, p15E, and U3 regions establish that several distinct retroviral transcription units are active in 129 Gix^+ mice. The overlapping regions of pol range from 3.2% to 5.3% divergence while the differences between these clones in p15E are 3.3% to 6.1%. Likewise, the two sequences presented from U3 vary by 4.4% single nucleotide changes. These differences most likely reflect sequence polymorphism of the expressed transcripts rather than cDNA synthesis errors. Reverse transcriptase has been estimated to incorporate an inappropriate nucleotide at approximately 1 in 500 nucleotides in vitro (Battula and Loeb, 1974, 1975) which is substantially lower than the difference frequency we have detected in these transcripts. Furthermore, much of the presented sequence has been confirmed on independently derived clones. In addition, sequence divergence resulting in restriction enzyme cleavage site polymorphisms are reflected in integrated viral sequences in the 129 genome. For example, a unique Bgl II site present in one of the two U3 sequences presented in Figure 5 was also found represented in some U3 hybridizing genomic restriction fragments (not shown).

The clearest form of sequence divergence demonstrated by these clones is found in the deletions apparent when corresponding regions of different clones are compared. These deletions accurately reflect transcripts in the RNA population previously identified by RNA blot (Levy et al., 1982) and S1 analysis (Chapter 3). In addition to the single nucleotide differences, several lines of evidence led us to conclude that these deleted transcripts were derived from distinct transcription units and not from a common precursor through alternative processing pathways.

First, proviruses with similar deleted structures which could have given rise to the observed transcripts have been detected in the genomes of laboratory mice (Khan et al., 1983). Second, the sequences at the junctions of the deletions, both in the deleted clones and in the region of the undeleted examples that would be considered introns assuming a splicing model, do not match presently accepted splicing consensus donor and acceptor sequences (Mount, 1982). Therefore, even if these deleted transcripts were indeed spliced, which we do not believe, they would probably not have been derived from the same sequence represented in the undeleted example. Originating from either mechanism, these transcripts would be the products of different transcription units. Finally, the lengths of several of the deletions are shorter than appears to be a minimal length for introns to be efficiently removed (Wieringa et al., 1984), perhaps due to the physical constraints of the splicing mechanism. It is interesting that in the case of the large env deletions, the endpoints map to recombination points identified as responsible for the creation of MCF viruses. The 5' breakpoint of S13 (Figure 2) corresponds to the 5' recombination point of MCF CI-3 virus (Mark and Rapp, 1984) while the 3' breakpoint of E2 (Figure 4) corresponds to the recombination point in MCF 247 p15E (Kelley et al., 1983). These results have a clear implication for the regulated expression of retroviral transcripts. Since these several distinct sequences, which must represent RNA transcibed from separate genomic transcription units, all hybridize transcripts that are absent or reduced in 129 Gix^- mice (Levy et al., 1982; Chapter 3), several proviral genomes are being coordinately regulated in response to the action of the Gv-1 locus. The dispersed configuration of proviral organization within the mouse genome (Blatt et al., 1983; Wejman et al., 1984) would require Gv-1 to be acting from a distance in trans to regulate the abundances of these independently transcribed RNA species.

Comparison of these cDNA clones with the sequences of exogenous infectious viruses may be instructive concerning both the nature of these transcripts and their possible involvement in pathogenic processes. It is clear from the data presented here that the endogenous retroviral transcripts normally expressed in uninfected 129 Gix^+ mice are closely related to the MCF class of murine leukemia virus. These transcripts,

although heterogeneous with respect to each other, all show substantial sequence similarity with the substituted, non-ecotropic portions of MCF 247 and MoMCF viruses. Endogenous proviruses of this type have previously been classified as xenotropic-like (Buckler et al., 1982) because they show much less homology with ecotropic viruses, such as AKV (Figure 6); however, they also cannot be as simply aligned with xenotropic virus such as NFS-Th-1 (Repaske et al., 1983; Khan, 1984) as they can with the MCF sequences. The assumption that the bulk of endogenous viruses were xenotropic has presented a conundrum concerning the origin and dispersal of these sequences. Since xenotropic virus does not infect mouse cells, it has been unclear how Mus species acquired the large number of proviruses presently in the genome. However, mouse cells possess specific receptors for MCF viruses (Rein and Schultz, 1984), and therefore proviruses of this type could have entered the genome independently as distinct infections.

Nonetheless, the following aspects of these sequences lead us to propose that the majority of endogenous proviruses of this type are the result of recent genomic amplification and transpostion events rather than of multiple exogenous infections of the germline. First, these endogenous sequences, although somewhat polymorphic, display considerable sequence conservation when compared with the divergence common between different viral strains. The sequences of the 129 endogenous transcripts as well as that of the AKR genomic clone A12 and of the non-ecotropic portions of AKR MCF 247 and MoMCF viruses and other MCF viruses (Amanuma et al., 1983; Schultz et al., 1983; Mark and Rapp, 1984) are all closely related. This is especially significant since these recombinant viruses arose in unrelated strains of mice by recombining with structurally distinct ecotropic viruses. The nucleotide differences between these sequences can mostly be accounted for by purine to purine and pyrimidine to pyrimidine transition mutations, a change typical of common mutation pathways (Miller, 1983). For example, in the 5' two-thirds of the gp70 coding region, approximately 90% of the nucleotide changes between MCF 247 and the 129 transcript are transitions, whereas in the 3' one-third, which contains the contribution of the ecotropic parent, only 78% of the changes are transitions; also, a codon has been deleted. Sequence comparisons of the gp70 regions that are substituted by recombination into

the MCF viruses indicate less than 2% variation among these sequences of endogenous origin. Such variability could result from the fixation of changes due to the random drift of genomic sequences (Wilson et al., 1977). The analogous regions of AKV, MoMuLV, and Friend MuLV, however, vary by approximately 25% and require gaps to align their sequences with each other. This rapid evolution is probably typical for horizontally transmitted viruses, due to strong selective pressures for rapid growth, and is in sharp contrast to the sequence conservation exhibited by endogenous retroviruses. It thus seems unlikely that the 25-50 copies of endogenous viruses in the mouse genome could have been derived from a population of infectious viruses by multiple, independent infections of the germline and still retain sequence homogeneity.

Second, the LTR of these endogenous sequences contains characteristic features which indicated substantial conservation, although this region usually undergoes rapid change during the evolution of different viral strains (Varmus, 1982) and even between different isolates of the same infectious virus (Shinnick et al., 1981; Holland et al., 1983; Villemur et al, 1983). For example, the two 129 endogenous transcripts as well as the LTR genomic clones sequenced by Khan and Martin (1983) contain the same tandem duplications not found in infectious LTRs. Also, they all contain the characteristic 192 bp insert not found in any infectious viruses. This insert is related to a family of repetitive elements (LTR-IS) which show characteristic structural features of transposons (Wirth et al, 1983) and an organization indicative of movement within the mouse genome (Wirth et al., 1984). Indeed, the sequence duplication flanking this insert within the LTR may have resulted from transpostion of the LTR-IS sequence into an integrated retroviral LTR. Since this insertion is always in an identical location, this potential transposition probably occurred only once in a primordial LTR from which all present endogenous viruses are derived. It is interesting to note the presence of tandem repeats of these flanking duplications in the LTR sequence from MCF 247 without the 192 bp insertion. This could indicate the excision of a transposable element in the generation of MCF 247, leaving behind only the tandem duplication of the insertion target site, similar to the action of Ds elements in maize (Sutton et al., 1984).

Furthermore, no infectious virus has been isolated corresponding to these endogenous sequences, especially none which utilizes these LTR sequences. Therefore, if endogenous retroviruses were derived from multiple infections from a pool of exogenous viruses, these viruses probably no longer exist in the population. An alternative origin for these genomic sequences, which we find more appealing to consider, is that a single infectious agent integrated into a murine ancestral species, became defective perhaps through the inactivation of its LTR or some other necessary function, and then became amplified through cellular processes. This is in contrast to a scenario requiring the members of a dispersed family of viruses, which entered the genome independently as competant viruses, all to lose infectious activity independently. Other examples of transposon-like behavior of retroviruses have been described (Temin, 1980; Jenkins et al., 1981). The coordinate regulation of the expression of these sequences may therefore be the result of the amplification of a regulatory sequence along with the dispersal of the viral structural genes. Analysis of the organization of retrovirus related sequences of cats (Soe et al., 1983) and of humans (Steele et al., 1984) has revealed evidence of the duplication of flanking sequences along with the possible amplification of the viral sequences.

The sequence of an endogenous env region presented here allows a precise definition of the substituted sequences acquired by pathogenic MCF viruses. From the histograms presented in Figure 6, it is clear that MCF 247 acquired endogenous sequence in the 3' portion of pol, the 5' two-thirds of gp70, and the extreme 3' terminal region of p15E while retaining ecotropic sequences for the remaining regions between these substitutions. Sequence comparisons in the U3 region (Figure 5) indicate that the 247 sequence is neither ecotropic in origin nor of the two types expressed in 129 mice. Greater sequence similarity is observed between MCF 247 U3 and analogous xenotropic sequences (Kahn and Martin, 1983).

The major difference between the pol sequences of the ecotropic virus and the leukemogenic MCF viruses is the deletion of four codons at the 3' end, accompanied by a clustering of amino acid differences. These

changes are probably within a function of the pol gene product that is essential for the establishment of a productive viral infection (Schwartzberg et al., 1984). Interestingly, similar changes occur in this region of MoMuLV (Shinnick et al., 1981), which is also a leukemogenic virus. This is the region of overlap between pol and env, particularly in AKV which may have an extended env leader peptide originating from a potential initiating methionine at position 237-239 in Figure 2 (Lenz et al., 1982). Because of this overlap, a selective advantage influencing a recombinant virus to acquire one of the two potential protein products of these two reading frames might result in the acquisition of the other gene as an advantitious event. It is therefore difficult to draw a firm conclusion from these data concerning potential leukemogenic determinants.

The acquisition of endogenous sequences encoding the amino terminal portion of gp70 is presumably responsible for the expanded host range of MCF viruses and may influence their tissue tropism (Devare et al., 1978). The reason for the return to ecotropic sequence for the carboxy portion of the protein and for the bulk of p15E is less clear, although this feature has been found to be a major determinant of leukemogenesis in recombinant viruses (Lung et al., 1983). The major differences between the endogenous sequence and that of MCF 247 and AKV fall into five regions separated by stretches of relative amino acid conservation although the nucleic acid sequences are still divergent (Figure 6 B and C). The first change involves the insertion of a single amino acid (Thr) accompanied by two amino acid changes (Ser to Ala and Tyr to Leu) making the endogenous protein slightly more hydrophobic. The second region of change in this portion of gp70 involves a carboxy terminal stretch of 28 noncharged amino acids which becomes substantially more hydrophobic, potentially changing the secondary structure of the predicted endogenous protein due to the substitution of two prolines for threonines and the replacement of a another threonine residue by isoleucine. The intervening region is relatively conserved, is charged and hydrophilic, and contains potential attachment sites for carbohydrate addition. This region could be important in the noncovalent interaction between gp70 and p15E. Similarly, the region of nucleotides 169-303 of p15E retains amino acid identity, though marked by nucleotide divergence (Figure 4) and also is

hydrophilic and highly charged. Flanking this site in p15E, however, are two regions of divergence. The amino terminal 43 amino acids of p15E in AKV ecotropic virus contains a stretch of 42 noncharged amino acids that has been suggested to possibly span the membrane (Lenz et al., 1982) or to contain a membrane fusing activity by analogy with the fusion proteins of other viruses (White et al., 1983). In the endogenous sequence, this stretch is interrupted by three charged residues (Asp, Lys, and Glu), perhaps modifying any potential helical structure.

The last two divergent regions between endogenous and MCF or AKV env flank the carboxy terminal, presumptive transmembrane region of p15E. The first involves the deletion of a single codon accompanied by several amino acid changes. The second involves the carboxy terminal R peptide, which is thought to be involved in virus maturation and assembly (Green et al., 1981). It should be noted that a probable second recombination in the formation of MCF 247 brings endogenous information to encode this region, clearly distinct from the ecotropic sequence. The significance of this change may involve possible cytoplasmic interactions of virion proteins, or, alternatively, may simply be a consequence of the required recombinational substitution of the ecotropic LTR for one of possible xenotropic origin (Quint et al., 1984).

Clearly, proteins synthesized from these normally expressed retroviral transcripts would closely resemble the MCF envelope proteins of recombinant viruses but would also have a distinct structure due to the portion not acquired by the recombinant viruses. The biological significance of the normal expression of MCF-like envelope proteins may be the blockage of the specific MCF cell surface receptor (Rein and Schultz, 1984) leading to resistance to MCF viral infection (Bassin et al., 1982). Since these MCF viruses are strongly implicated in the incidence of leukemia, such resistance may impart a significant selective advantage to the animal. The potentially altered structure of these proteins with respect to infectious viral proteins, particularly in the membrane anchor protein p15E, may result in the release of endogenous gp70 in the secreted form found in serum and epididymis (Hara et al., 1982).

EXPERIMENTAL PROCEDURES

DNA Sequence analysis

DNA sequences were obtained by the method of Sanger et al. (1977) using the single stranded vectors M13 mp8, mp9, mp10, and mp11 (Messing and Vieira, 1982) as described in Chapter 3. Sequence data were compiled using the computer programs of Staden (1982) and compared with published sequences using the computer program ALIGN (Orcutt et al., 1982).

FIGURE 1

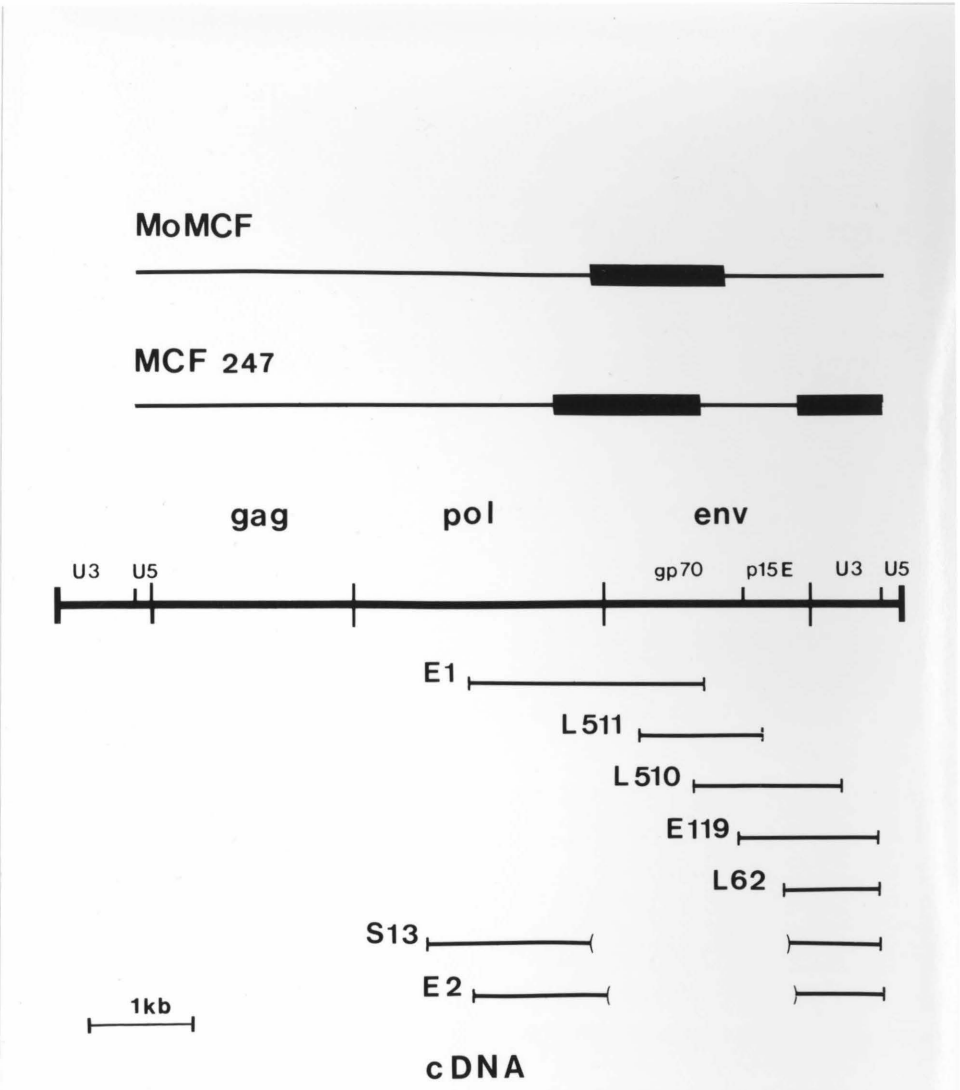
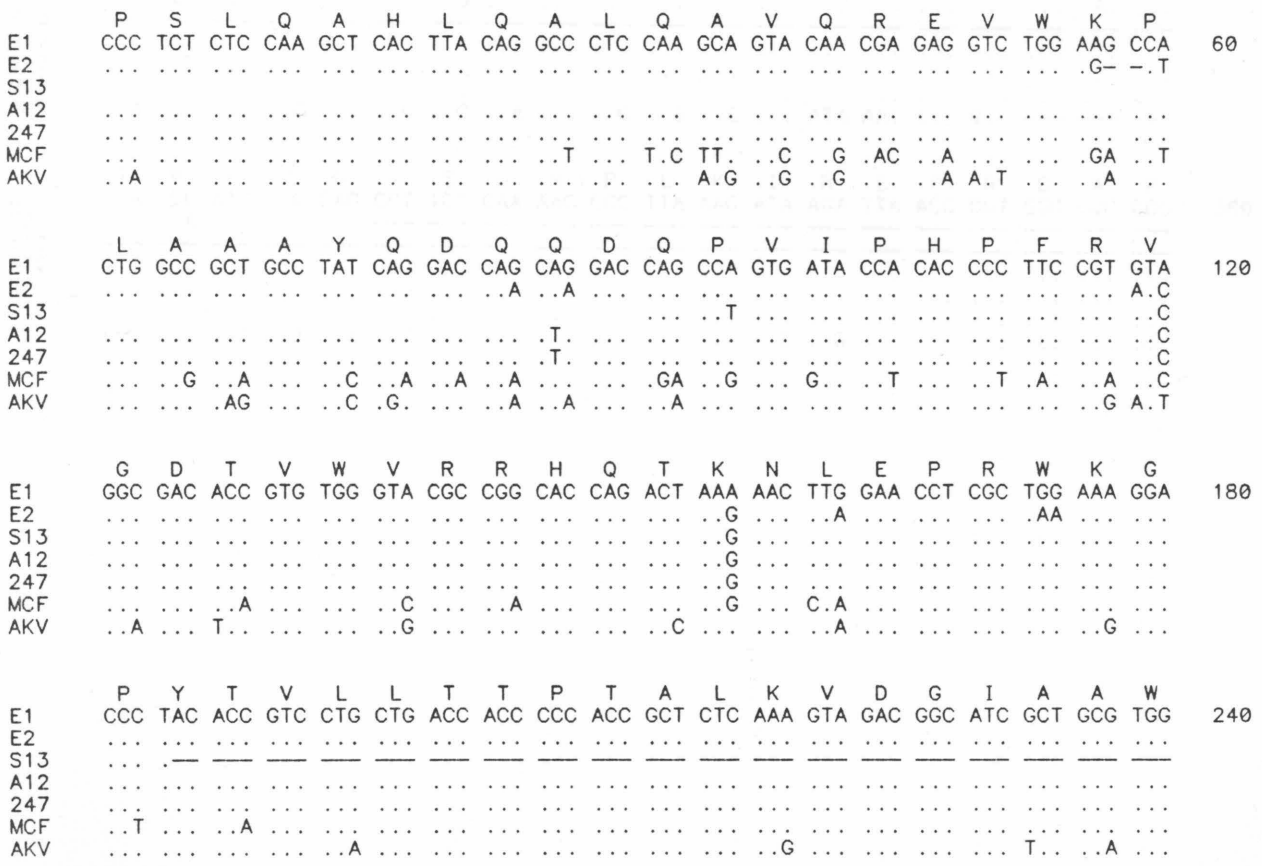

MoMCF
MCF 247
gag
pol
env
U3
U5
gp70
p15E
U3
U5
E1
L511
L510
E119
L62
S13
E2
1kb
cDNA

FIGURE 1

STRUCTURE OF RETROVIRAL cDNA CLONES

The cDNA inserts of retroviral clones derived from 129 Gix^+ polyadenylated RNA were aligned with MoMuLV proviral structure (middle) by restriction enzyme mapping, hybridization and DNA sequencing. Vertical bars in the cDNA clones indicate positions of junctions of eukaryotic sequences with plasmid DNA. Parentheses enclose deleted regions of retroviral sequence. The designation of each clone includes an initial letter indicating the tissue from which the RNA was derived. E, epididymis; L, liver; S, spleen. At the top are diagrams of MoMCF and MCF 247 viral genomic RNA. The thin lines represent ecotropic sequences while the thickened portions represent regions of non-ecotropic information substituted into the genome by recombination. The MCF 247 diagram is based on the restriciton maps of Chattopadhyay et al. (1981) and the sequences of Kelley et al. (1983) and Holland et al. (1983). The diagram of MoMCF is after Bosselman et al. (1982).

FIGURE 2


P S L Q A H L Q A L Q A V Q R E V W K P
E1 CCC TCT CTC CAA GCT CAC TTA CAG GCC CTC CAA GCA GTA CAA CGA GAG GTC TGG AAG CCA 60
E2 ... ... .G- -.T
S13
A12 ... ...
247 ... ...
MCF ... ...T ... T.C TT. ..C ..G .AC ..A ... ... .GA ..T
AKV ..A ... ... A.G ..G ..G ... ..A A.T ... ..A ...
L A A A Y Q D Q Q D Q P V I P H P F R V
E1 CTG GCC GCT GCC TAT CAG GAC CAG CAG GAC CAG CCA GTG ATA CCA CAC CCC TTC CGT GTA 120
E2 ... ...A ..A ... ... A.C
S13 ... ..T C
A12 ... ... .T. C
247 ... ... .T. C
MCF ... ..G ..A ... ..C ..A ..A ..A ... ... .GA ..G ... G.. ..T ... ..T .A. ..A ..C
AKV ... ... .AG ... ..C .G. ... ..A ..A ... ..A ... ...G A.T
G D T V W V R R H Q T K N L E P R W K G
E1 GGC GAC ACC GTG TGG GTA CGC CGG CAC CAG ACT AAA AAC TTG GAA CCT CGC TGG AAA GGA 180
E2 ... ...G ... ..A ... ... .AA ... ...
S13 ... ...G ... ...
A12 ... ...G ... ...
247 ... ...G ... ...
MCF ... .A ... .C ... ..A ... ...G ... C.A ... ...
AKV ..A ... T.. ... .G ... ...C ... .A ... ...G ...
P Y T V L L T T P T A L K V D G I A A W
E1 CCC TAC ACC GTC CTG CTG ACC ACC CCC ACC GCT CTC AAA GTA GAC GGC ATC GCT GCG TGG 240
E2 ... ...
S13 ... .— — — — — — — — — — — — — — — — — — — —
A12 ... ...
247 ... ...
MCF ..T ... ..A ... ...
AKV ... ...A ... ...G ... ... T.. ..A ...

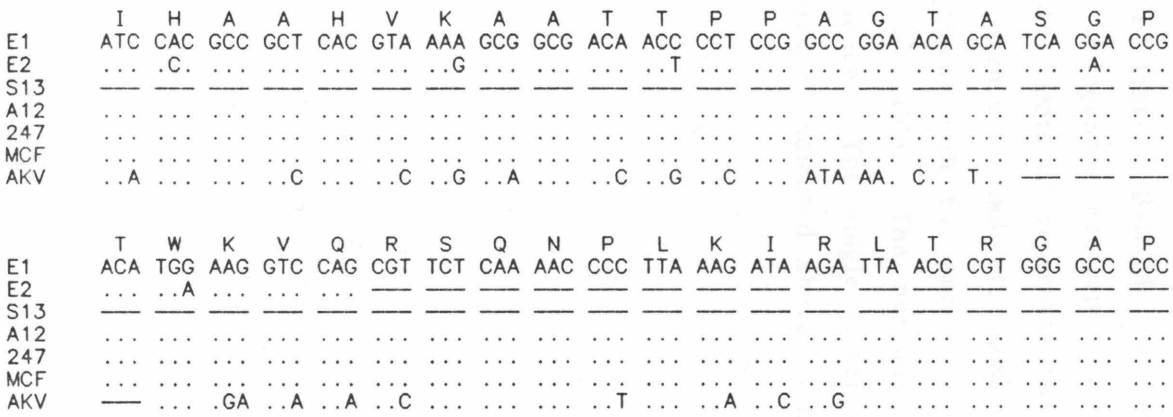
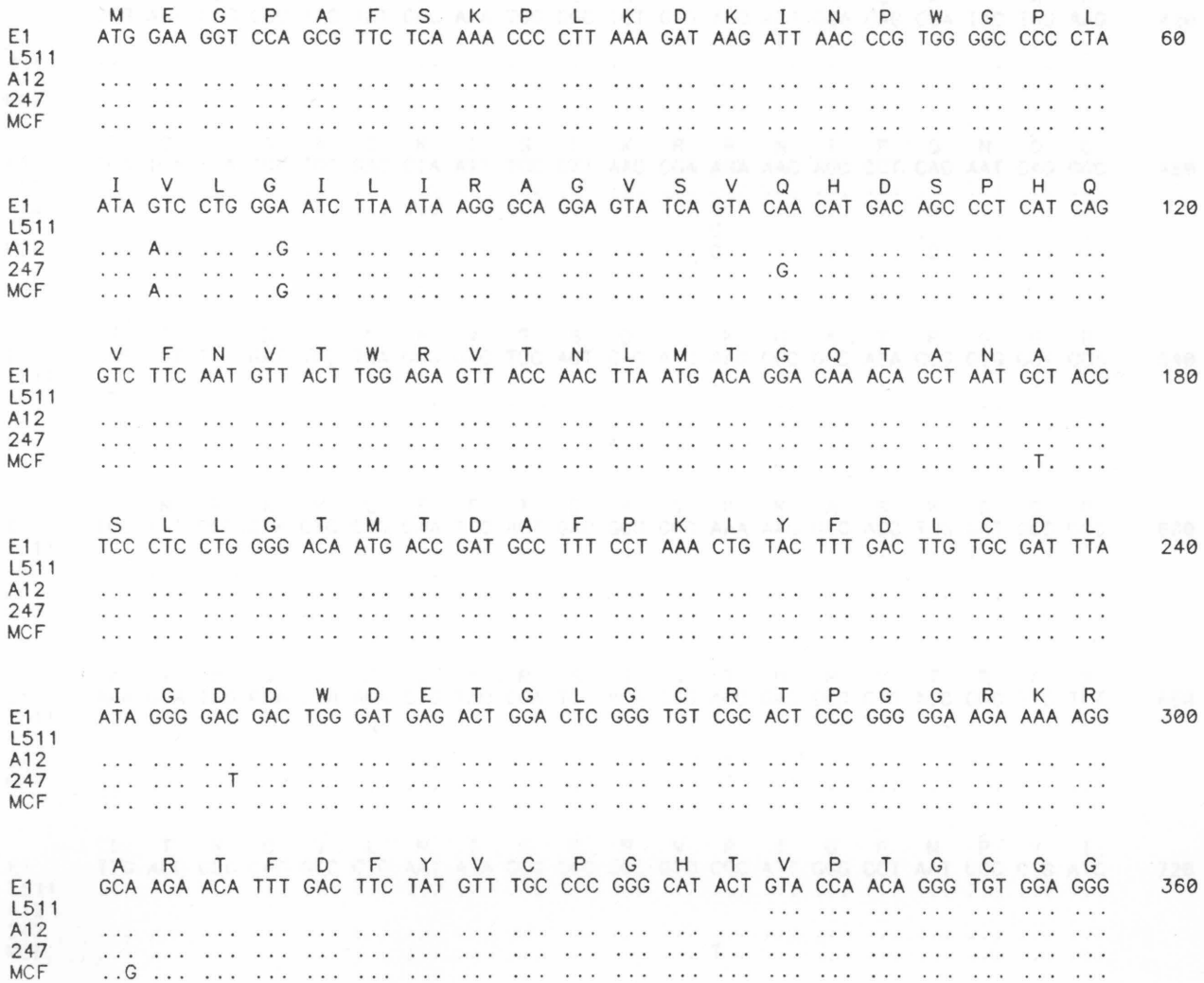

I H A A H V K A A T T P P A G T A S G P
E1 ATC CAC GCC GCT CAC GTA AAA GCG GCG ACA ACC CCT CCG GCC GGA ACA GCA TCA GGA CCG 300
E2 ... .C. ... ...G ... ...T ... ... .A. ...
S13 ---
A12 ... ...
247 ... ...
MCF ... ...
AKV ..A ... .C ... ..C ..G ..A ... ..C ..G ..C ... ATA AA. C.. T.. --- --- ---
T W K V Q R S Q N P L K I R L T R G A P
E1 ACA TGG AAG GTC CAG CGT TCT CAA AAC CCC TTA AAG ATA AGA TTA ACC CGT GGG GCC CCC 360
E2 ... ..A ... ... --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
S13 ---
A12 ... ...
247 ... ...
MCF ... ...
AKV --- ... .GA ..A ..A ..C ... ...T ... ..A ..C ..G ... ...

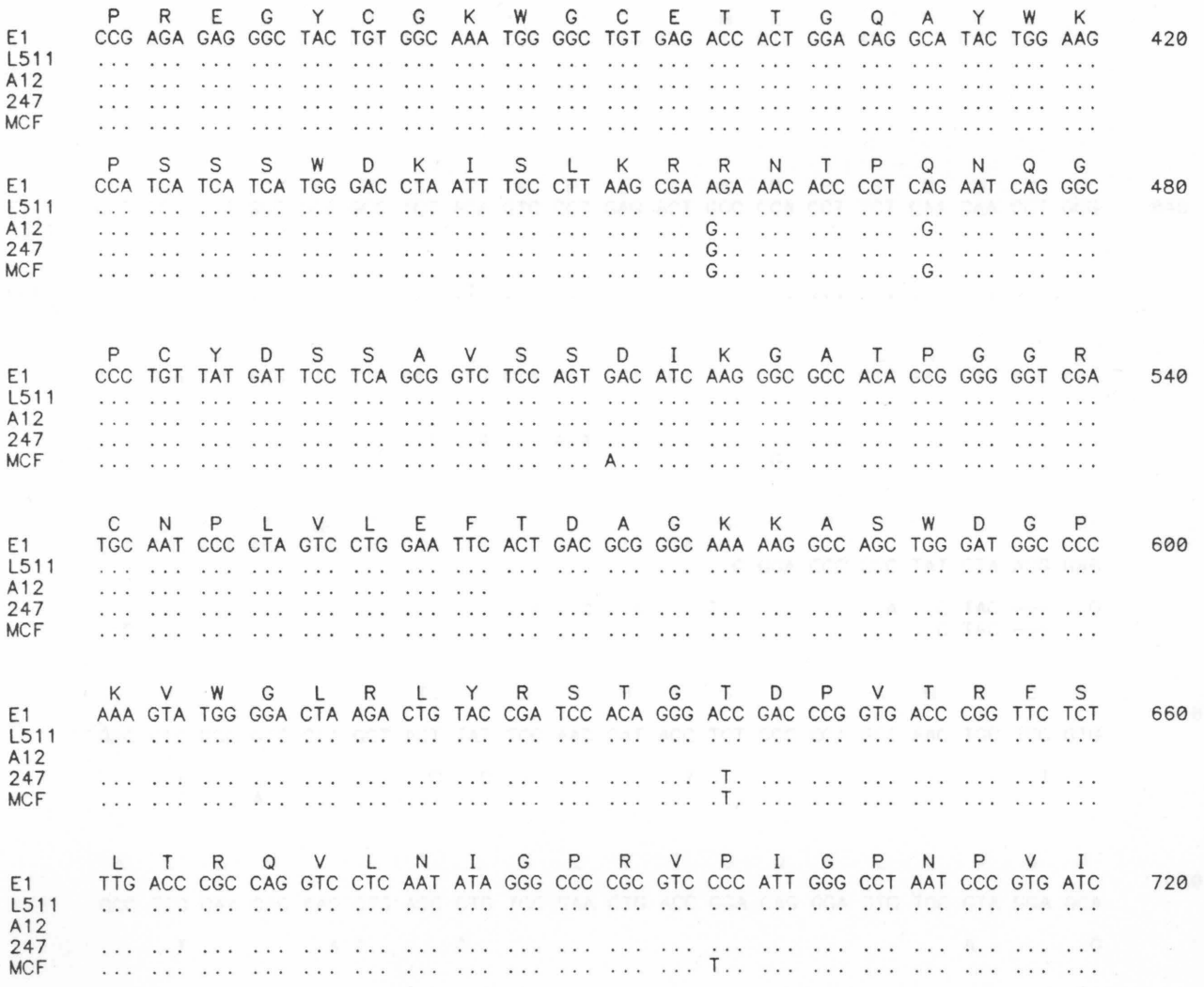
FIGURE 2

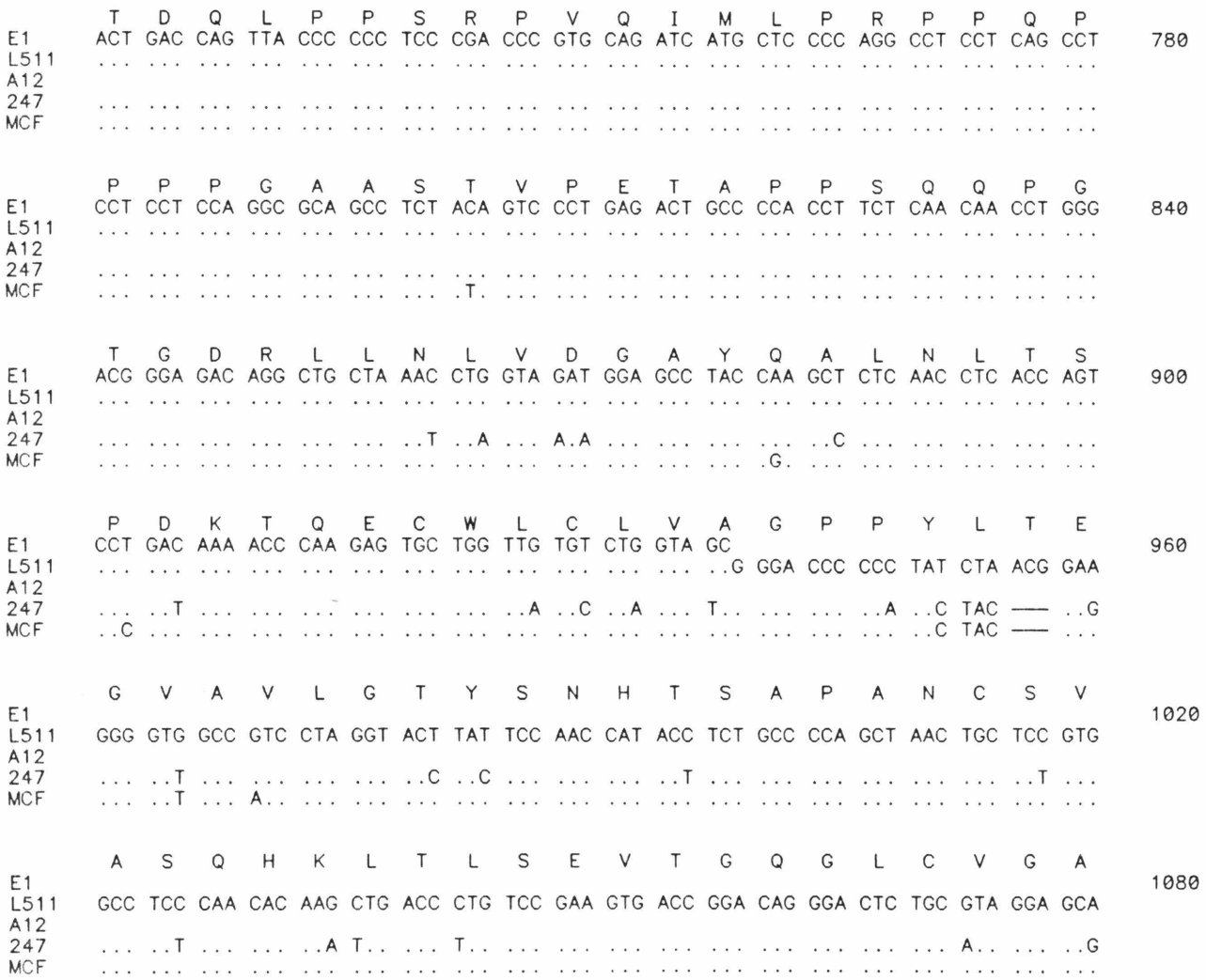
SEQUENCE COMPARISON OF THE pol REGION

The DNA sequence and predicted amino acid sequence for 360 nt from the 3' end of the pol gene open reading frame of endogenous transcript clone E1 are presented. This nucleotide sequence is compared with sequences of clones E2 and S13 from 129 Gix^+ mice, the proviral genomic clone A12 from AKR mice (Khan, 1984), AKR MCF 247 virus (247) from an AKR thymoma (Holland et al., 1983), Moloney MCF virus (MCF) from a Balb/Mo thymoma (Bosselman et al., 1982), and AKV virus, the endogenous ecotropic virus of AKR mice (Lenz et al., 1983; Herr, 1984). Dots (.) represent nucleotides identical with the sequence of clone E1, dashes (–) indicate gaps introduced in the sequences to increase the alignment with E1, and blank spaces are present where no sequence data are available. The numbers to the right count nucleotides of the E1 sequence. The single letter code is used for the translated amino acid sequence, indicated above each codon.

FIGURE 3


M E G P A F S K P L K D K I N P W G P L
E1 ATG GAA GGT CCA GCG TTC TCA AAA CCC CTT AAA GAT AAG ATT AAC CCG TGG GGC CCC CTA 60
L511
A12 ... ...
247 ... ...
MCF ... ...
I V L G I L I R A G V S V Q H D S P H Q
E1 ATA GTC CTG GGA ATC TTA ATA AGG GCA GGA GTA TCA GTA CAA CAT GAC AGC CCT CAT CAG 120
L511
A12 ... A.. ... ..G ... ...
247 ... ... .G. ... ...
MCF ... A.. ... ..G ... ...
V F N V T W R V T N L M T G Q T A N A T
E1 GTC TTC AAT GTT ACT TGG AGA GTT ACC AAC TTA ATG ACA GGA CAA ACA GCT AAT GCT ACC 180
L511
A12 ... ...
247 ... ...
MCF ... ... .T. ...
S L L G T M T D A F P K L Y F D L C D L
E1 TCC CTC CTG GGG ACA ATG ACC GAT GCC TTT CCT AAA CTG TAC TTT GAC TTG TGC GAT TTA 240
L511
A12 ... ...
247 ... ...
MCF ... ...
I G D D W D E T G L G C R T P G G R K R
E1 ATA GGG GAC GAC TGG GAT GAG ACT GGA CTC GGG TGT CGC ACT CCC GGG GGA AGA AAA AGG 300
L511
A12 ... ...
247 ... .T ... ...
MCF ... ...
A R T F D F Y V C P G H T V P T G C G G
E1 GCA AGA ACA TTT GAC TTC TAT GTT TGC CCC GGG CAT ACT GTA CCA ACA GGG TGT GGA GGG 360
L511 ... ...
A12 ... ...
247 ... ...
MCF ..G ... ...


P R E G Y C G K W G C E T T G Q A Y W K
E1 CCG AGA GAG GGC TAC TGT GGC AAA TGG GGC TGT GAG ACC ACT GGA CAG GCA TAC TGG AAG 420
L511 ... ...
A12 ... ...
247 ... ...
MCF ... ...
P S S S W D K I S L K R R N T P Q N Q G
E1 CCA TCA TCA TCA TGG GAC CTA ATT TCC CTT AAG CGA AGA AAC ACC CCT CAG AAT CAG GGC 480
L511 ... ...
A12 ... ... G.. ... ... .G. ... ...
247 ... ... G.. ... ...
MCF ... ... G.. ... ... .G. ... ...
P C Y D S S A V S S D I K G A T P G G R
E1 CCC TGT TAT GAT TCC TCA GCG GTC TCC AGT GAC ATC AAG GGC GCC ACA CCG GGG GGT CGA 540
L511 ... ...
A12 ... ...
247 ... ...
MCF ... ... A.. ... ...
C N P L V L E F T D A G K K A S W D G P
E1 TGC AAT CCC CTA GTC CTG GAA TTC ACT GAC GCG GGC AAA AAG GCC AGC TGG GAT GGC CCC 600
L511 ... ...
A12 ... ...
247 ... ...
MCF ... ...
K V W G L R L Y R S T G T D P V T R F S
E1 AAA GTA TGG GGA CTA AGA CTG TAC CGA TCC ACA GGG ACC GAC CCG GTG ACC CGG TTC TCT 660
L511 ... ...
A12
247 ... ... .T. ... ...
MCF ... ... .T. ... ...
L T R Q V L N I G P R V P I G P N P V I
E1 TTG ACC CGC CAG GTC CTC AAT ATA GGG CCC CGC GTC CCC ATT GGG CCT AAT CCC GTG ATC 720
L511 ... ...
A12
247 ... ...
MCF ... ... T.. ... ...


T D Q L P P S R P V Q I M L P R P P Q P
E1 ACT GAC CAG TTA CCC CCC TCC CGA CCC GTG CAG ATC ATG CTC CCC AGG CCT CCT CAG CCT 780
L511 ... ...
A12
247 ... ...
MCF ... ...
P P P G A A S T V P E T A P P S Q Q P G
E1 CCT CCT CCA GGC GCA GCC TCT ACA GTC CCT GAG ACT GCC CCA CCT TCT CAA CAA CCT GGG 840
L511 ... ...
A12
247 ... ...
MCF ... ... .T. ... ...
T G D R L L N L V D G A Y Q A L N L T S
E1 ACG GGA GAC AGG CTG CTA AAC CTG GTA GAT GGA GCC TAC CAA GCT CTC AAC CTC ACC AGT 900
L511 ... ...
A12
247 ... ...T ..A ... A.A ... ...C ... ...
MCF ... ... .G. ... ...
P D K T Q E C W L C L V A G P P Y L T E
E1 CCT GAC AAA ACC CAA GAG TGC TGG TTG TGT CTG GTA GC 960
L511 ... ...G GGA CCC CCC TAT CTA ACG GAA
A12
247 ... ..T ... ...A ..C ..A ... T.. ... .A ..C TAC —— ..G
MCF ..C ... ...C TAC —— ...
G V A V L G T Y S N H T S A P A N C S V
E1 1020
L511 GGG GTG GCC GTC CTA GGT ACT TAT TCC AAC CAT ACC TCT GCC CCA GCT AAC TGC TCC GTG
A12
247 ... ..T ... ...C ..C ... ...T ... ...T ...
MCF ... ..T ... A.. ... ...
A S Q H K L T L S E V T G Q G L C V G A
E1 1080
L511 GCC TCC CAA CAC AAG CTG ACC CTG TCC GAA GTG ACC GGA CAG GGA CTC TGC GTA GGA GCA
A12
247 ... ..T ... .A T.. ... T.. ... ... A.. ... ..G
MCF ... ...

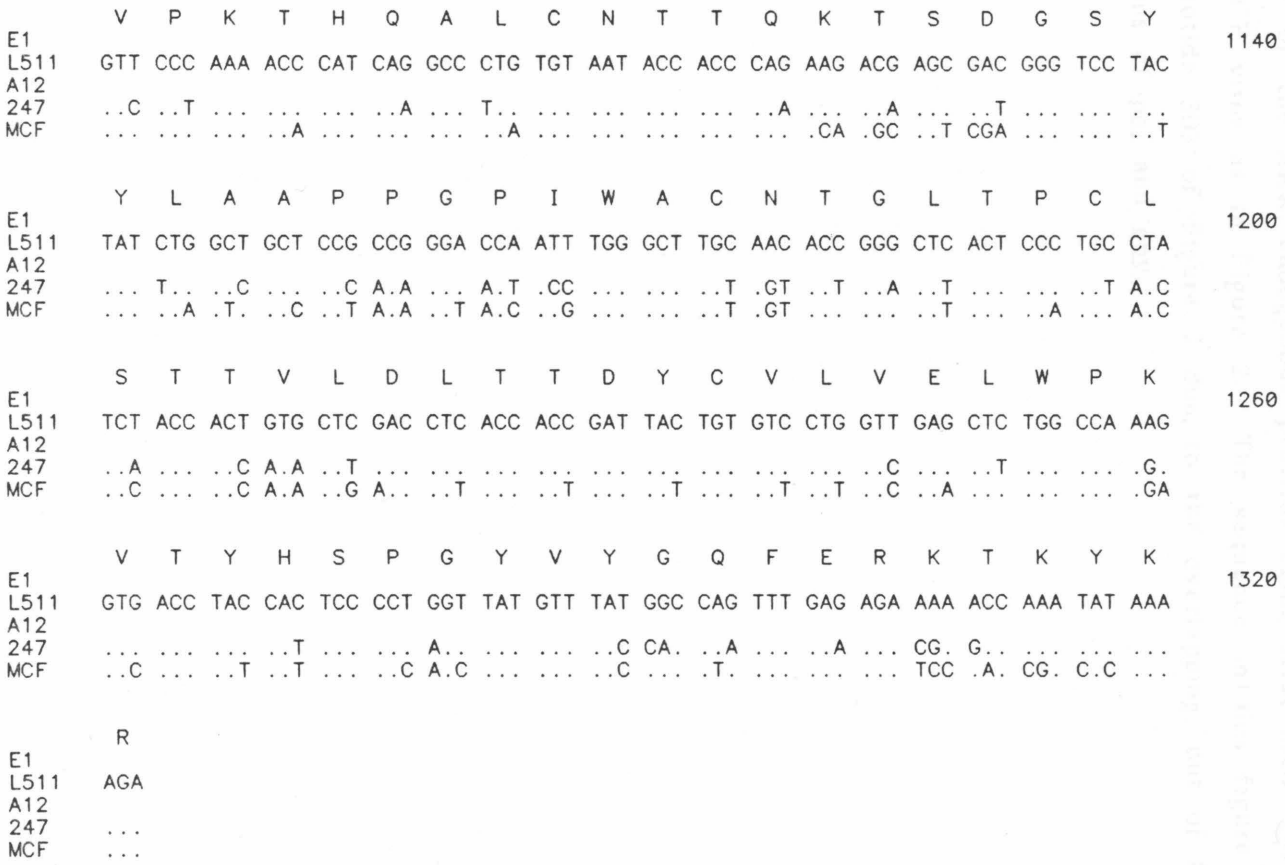
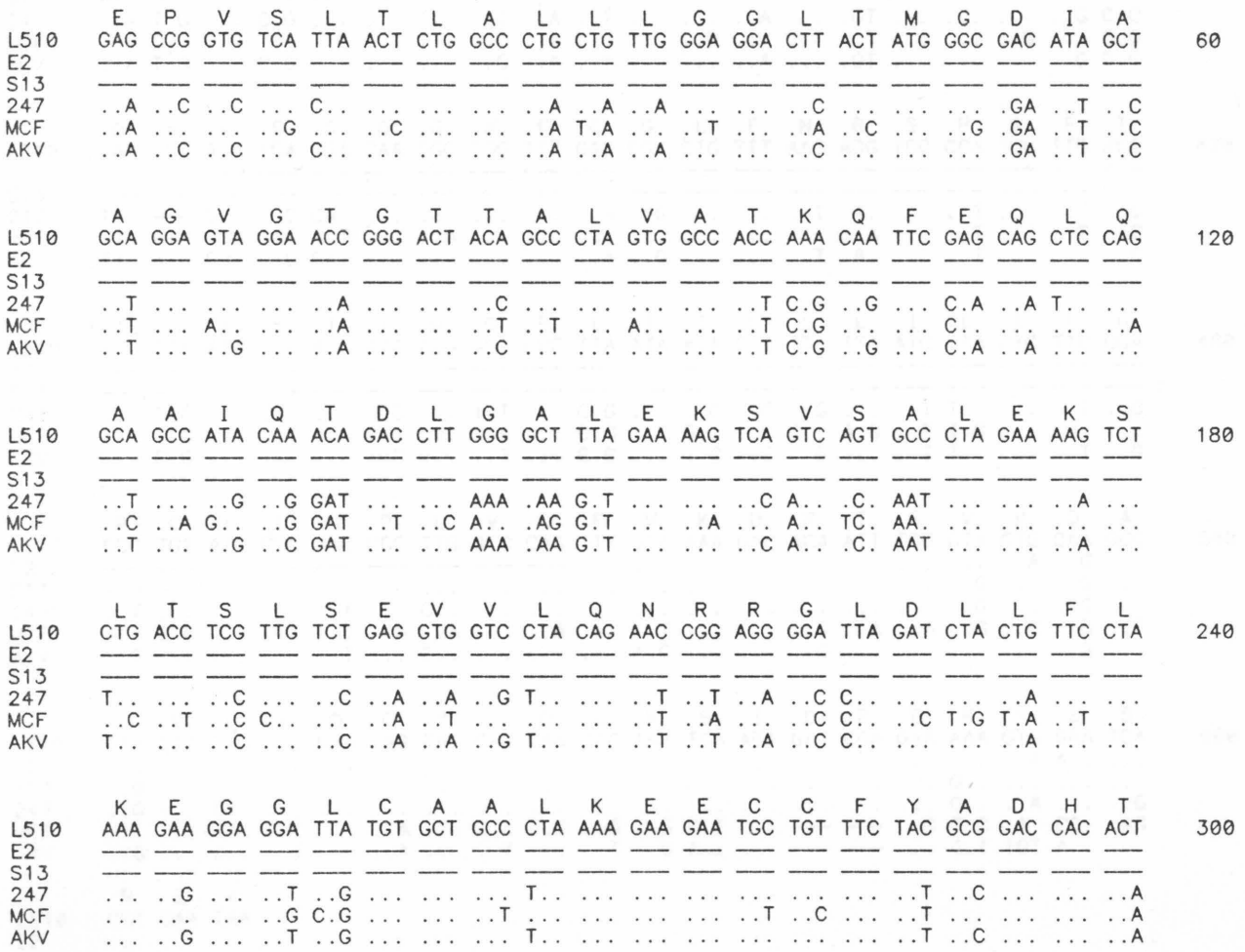

V P K T H Q A L C N T T Q K T S D G S Y
E1
L511 GTT CCC AAA ACC CAT CAG GCC CTG TGT AAT ACC ACC CAG AAG ACG AGC GAC GGG TCC TAC
A12
247 ..C ..T ... ...A ... T.. ... ...A ... ..A ... ..T ... ...
MCF ... ...A ... ...A ... ... .CA .GC ..T CGA ... .T
1140
Y L A A P P G P I W A C N T G L T P C L
E1
L511 TAT CTG GCT GCT CCG CCG GGA CCA ATT TGG GCT TGC AAC ACC GGG CTC ACT CCC TGC CTA
A12
247 ... T.. ..C ... ..C A.A ... A.T .CC ... .T .GT ..T ..A ..T ... .T A.C
MCF ... ..A .T. ..C ..T A.A ..T A.C ..G ... .T .GT ... .T ... ..A ... A.C
1200
S T T V L D L T T D Y C V L V E L W P K
E1
L511 TCT ACC ACT GTG CTC GAC CTC ACC ACC GAT TAC TGT GTC CTG GTT GAG CTC TGG CCA AAG
A12
247 ..A ... ..C A.A ..T ... ...C ... ..T ... ... .G.
MCF ..C ... ..C A.A ..G A.. ..T ... ..T ... ..T ... ..T ..T ..C ..A ... ... .GA
1260
V T Y H S P G Y V Y G Q F E R K T K Y K
E1
L511 GTG ACC TAC CAC TCC CCT GGT TAT GTT TAT GGC CAG TTT GAG AGA AAA ACC AAA TAT AAA
A12
247 ... ...T ... ... A.. ... .C CA. ..A ... ..A ... CG. G.. ... ...
MCF ..C ... ..T ..T ... ..C A.C ... .C ... .T. ... ... TCC .A. CG. C.C ...
1320
R
E1
L511 AGA
A12
247 ...
MCF ...

FIGURE 3

SEQUENCE COMPARISON OF THE gp70 CODING REGION

The DNA sequence and predicted amino acid sequence for the gp70 coding region of 129 Gix^+ endogenous transcripts was derived from two overlapping cDNA clones E1 and L511. This sequence is compared with those of the AKR endogenous proviral clone A12, MCF 247 virus, and MoMCF virus as in Figure 2. The sequence in this figure begins with nucleotide 303 of Figure 2 due to the overlapping, out of frame coding regions of pol and env.

FIGURE 4


E P V S L T L A L L L G G L T M G D I A
L510 GAG CCG GTG TCA TTA ACT CTG GCC CTG CTG TTG GGA GGA CTT ACT ATG GGC GAC ATA GCT 60
E2
S13
247 ..A ..C ..C ... C.. ... ...A ..A ..A ... .C ... ... .GA ..T ..C
MCF ..A ... .G ... ..C ... .A T.A ... ..T ... ..A ..C ... ..G .GA ..T ..C
AKV ..A ..C ..C ... C.. ... ...A ..A ..A ... .C ... ... .GA ..T ..C
A G V G T G T T A L V A T K Q F E Q L Q
L510 GCA GGA GTA GGA ACC GGG ACT ACA GCC CTA GTG GCC ACC AAA CAA TTC GAG CAG CTC CAG 120
E2
S13
247 ..T ... ...A ... .C ... ...T C.G ..G ... C.A ..A T.. ...
MCF ..T ... A.. ... ..A ... .T ..T ... A.. ... ..T C.G ... ... C.. ... .A
AKV ..T ... ..G ... ..A ... .C ... ...T C.G ..G ... C.A ..A ... ...
A A I Q T D L G A L E K S V S A L E K S
L510 GCA GCC ATA CAA ACA GAC CTT GGG GCT TTA GAA AAG TCA GTC AGT GCC CTA GAA AAG TCT 180
E2
S13
247 ..T ... ..G ..G GAT ... ... AAA .AA G.T ... .C A.. .C. AAT ... .A ...
MCF ..C ..A G.. ..G GAT ..T ..C A.. .AG G.T ... ..A ... A.. TC. AA. ... ...
AKV ..T ... ..G ..C GAT ... ... AAA .AA G.T ... .C A.. .C. AAT ... .A ...
L T S L S E V V L Q N R R G L D L L F L
L510 CTG ACC TCG TTG TCT GAG GTG GTC CTA CAG AAC CGG AGG GGA TTA GAT CTA CTG TTC CTA 240
E2
S13
247 T.. ... ..C ... ..C ..A ..A ..G T.. ... ..T ..T ..A ..C C.. ... .A ... ...
MCF ..C ..T ..C C.. ... ..A ..T ... ...T ..A ... ..C C.. ..C T.G T.A ..T ...
AKV T.. ... ..C ... ..C ..A ..A ..G T.. ... ..T ..T ..A ..C C.. ... .A ... ...
K E G G L C A A L K E E C C F Y A D H T
L510 AAA GAA GGA GGA TTA TGT GCT GCC CTA AAA GAA GAA TGC TGT TTC TAC GCG GAC CAC ACT 300
E2
S13
247 ... ..G ... ..T ..G ... ... T.. ... ...T ..C ... .A
MCF ... ...G C.G ... .T ... ...T ..C ... ..T ... ...A
AKV ... ..G ... ..T ..G ... ... T.. ... ...T ..C ... .A

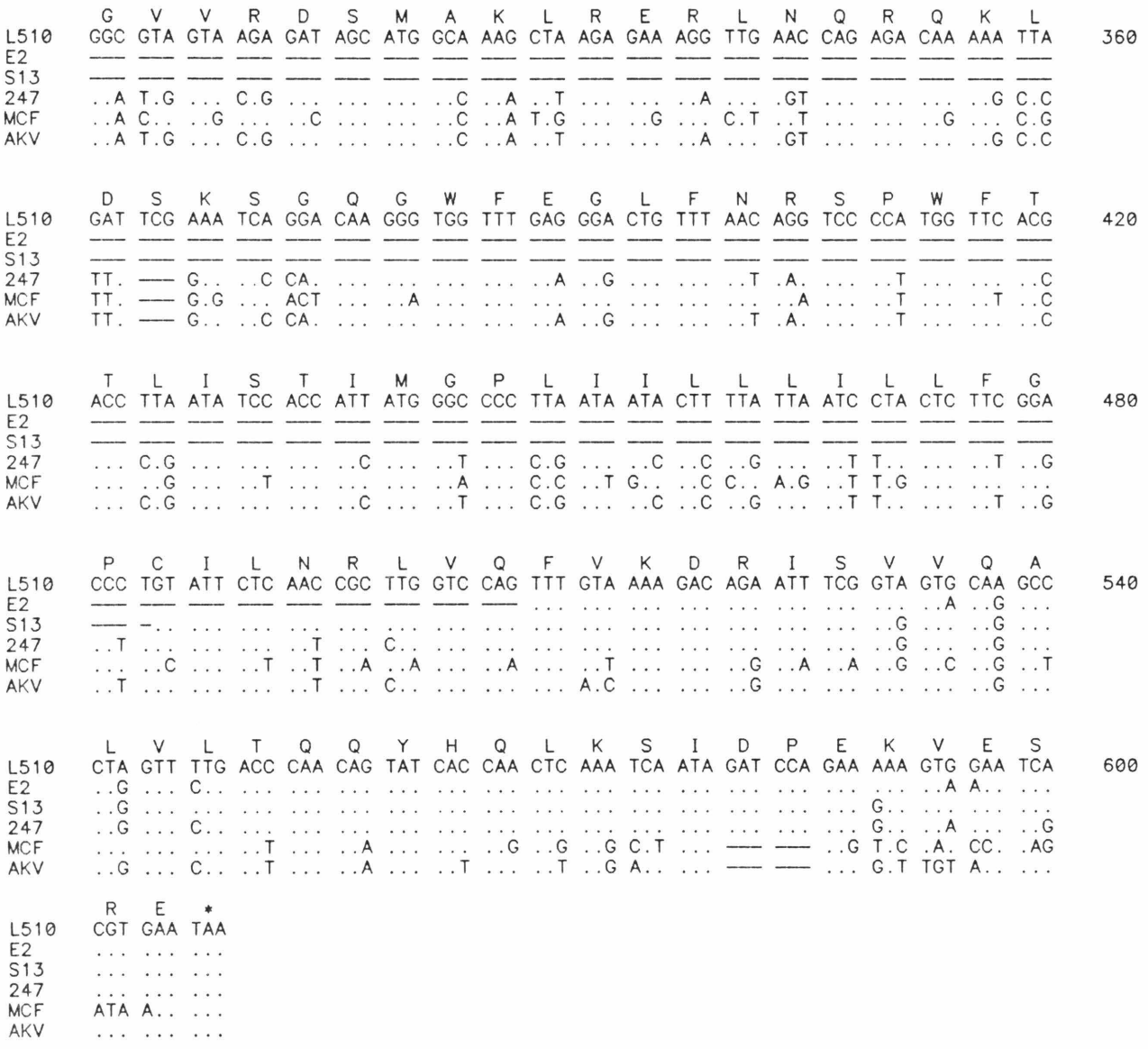

G V V R D S M A K L R E R L N Q R Q K L
L510 GGC GTA GTA AGA GAT AGC ATG GCA AAG CTA AGA GAA AGG TTG AAC CAG AGA CAA AAA TTA 360
E2
S13
247 ..A T.G ... C.G ... ...C ..A ..T ... .A ... .GT ... ...G C.C
MCF ..A C.. ..G ... ..C ... .C ..A T.G ... ..G ... C.T ..T ... .G ... C.G
AKV ..A T.G ... C.G ... ...C ..A ..T ... .A ... .GT ... ...G C.C
D S K S G Q G W F E G L F N R S P W F T
L510 GAT TCG AAA TCA GGA CAA GGG TGG TTT GAG GGA CTG TTT AAC AGG TCC CCA TGG TTC ACG 420
E2
S13
247 TT. G.. ..C CA. ... ...A ..G ... .T .A. ... ..T C
MCF TT. G.G ... ACT ... ..A ... ...A ... ..T ... ..T ..C
AKV TT. G.. ..C CA. ... ...A ..G ... .T .A. ... ..T C
T L I S T I M G P L I I L L L I L L F G
L510 ACC TTA ATA TCC ACC ATT ATG GGC CCC TTA ATA ATA CTT TTA TTA ATC CTA CTC TTC GGA 480
E2
S13
247 ... C.G ... ...C ... ..T ... C.G ... ..C ..C ..G ... ..T T.. ... ..T ..G
MCF ... ..G ... ..T ... ...A ... C.C ..T G.. ..C C.. A.G ..T T.G ... ...
AKV ... C.G ... ...C ... ..T ... C.G ... ..C ..C ..G ... ..T T.. ... ..T ..G
P C I L N R L V Q F V K D R I S V V Q A
L510 CCC TGT ATT CTC AAC CGC TTG GTC CAG TTT GTA AAA GAC AGA ATT TCG GTA GTG CAA GCC 540
E2 ... ...A ..G ...
S13 -.. ... ...G ... ..G ...
247 ..T ... ...T ... C.. ... ...G ... ..G ...
MCF ... ..C ... ..T ..T ..A ..A ... ..A ... ..T ... .G ..A ..A ..G ..C ..G ..T
AKV ..T ... ...T ... C.. ... ... A.C ... .G ... ...G ...
L V L T Q Q Y H Q L K S I D P E K V E S
L510 CTA GTT TTG ACC CAA CAG TAT CAC CAA CTC AAA TCA ATA GAT CCA GAA AAA GTG GAA TCA 600
E2 ..G ... C.. ... ...A A.. ...
S13 ..G ... ... G.. ... ...
247 ..G ... C.. ... ... G.. ..A ... ..G
MCF ... ...T ... ..A ... .G ..G ..G C.T ... -- -- ..G T.C .A. CC. .AG
AKV ..G ... C.. ..T ... ..A ... ..T ... ..T ..G A.. ... -- -- ... G.T TGT A.. ...
R E *
L510 CGT GAA TAA
E2 ... ...
S13 ... ...
247 ... ...
MCF ATA A.. ...
AKV ... ...

FIGURE 4

SEQUENCE COMPARISON OF THE p15E CODING REGION

The DNA sequence and predicted amino acid sequence for the p15E coding region of 129 Gix^+ cDNA clone L510 is presented and compared with the corresponding sequences of cDNA clones E2 and S13 and of the infectious viruses MCF 247 (Kelley et al., 1983), MoMCF, and AKV as in Figure 2. The first nucleotide in this figure is one nucleotide 3' to the last nucleotide in Figure 3.

FIGURE 5

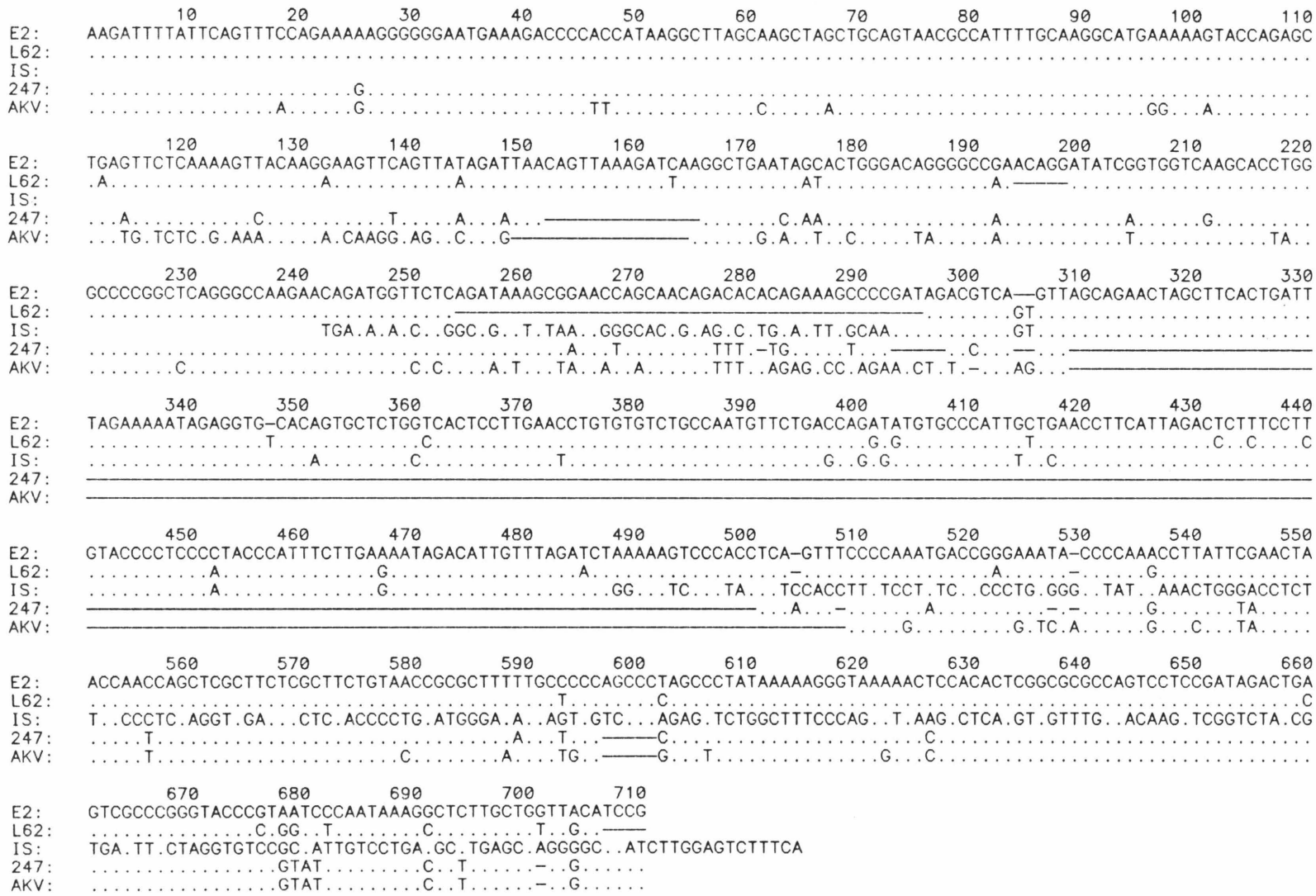
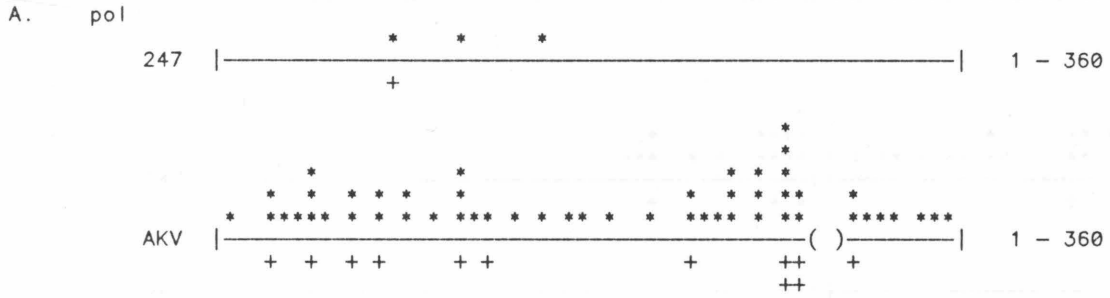

10 20 30 40 50 60 70 80 90 100 110
E2: AAGATTTTATTCAGTTTCCAGAAAAAGGGGGGAATGAAAGACCCCACCATAAGGCTTAGCAAGCTAGCTGCAGTAACGCCATTTTGCAAGGCATGAAAAAGTACCAGAGC
L62:
IS:
247:
AKV:
120 130 140 150 160 170 180 190 200 210 220
E2: TGAGTTCTCAAAAGTTACAAGGAAGTTCAGTTATAGATTAACAGTTAAAGATCAAGGCTGAATAGCACTGGGACAGGGGCCGAACAGGATATCGGTGGTCAAGCACCTGG
L62:
IS:
247:
AKV:
230 240 250 260 270 280 290 300 310 320 330
E2: GCCCCGGCTCAGGGCCAAGAACAGATGGTTCTCAGATAAAGCGGAACCAGCAACAGACACACAGAAAGCCCCGATAGACGTCA—GTTAGCAGAACTAGCTTCACTGATT
L62:
IS:
247:
AKV:
340 350 360 370 380 390 400 410 420 430 440
E2: TAGAAAAATAGAGGTG-CACAGTGCTCTGGTCACTCCTTGAACCTGTGTGTCTGCCAATGTTCTGACCAGATATGTGCCCATTGCTGAACCTTCATTAGACTCTTTCCTT
L62:
IS:
247:
AKV:
450 460 470 480 490 500 510 520 530 540 550
E2: GTACCCCTCCCCTACCCATTTCTTGAAAATAGACATTGTTTAGATCTAAAAAGTCCCACCTCA-GTTTCCCCAAATGACCGGGAAATA-CCCCAAACCTTATTCGAACTA
L62:
IS:
247:
AKV:
560 570 580 590 600 610 620 630 640 650 660
E2: ACCAACCAGCTCGCTTCTCGCTTCTGTAACCGCGCTTTTTGCCCCCAGCCCTAGCCCTATAAAAAGGGTAAAAACTCCACACTCGGCGCGCCAGTCCTCCGATAGACTGA
L62:
IS:
247:
AKV:
670 680 690 700 710
E2: GTCGCCCGGGTACCCGTAATCCCAATAAAGGCTCTTGCTGGTTACATCCG
L62:
IS:
247:
AKV:

FIGURE 5

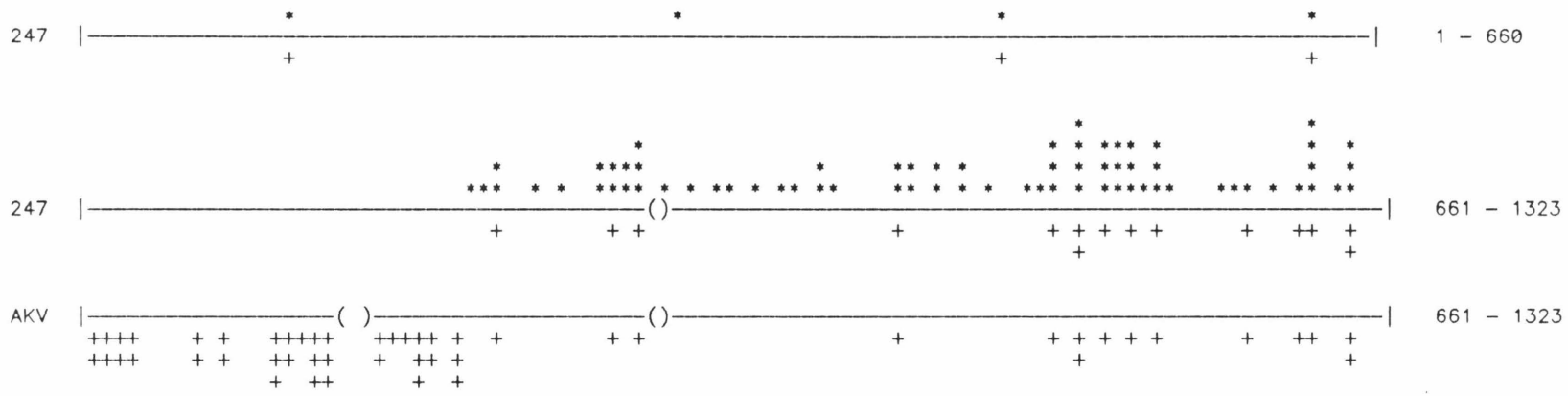
SEQUENCE COMPARISON OF THE U3/R PORTION OF THE LTR

The DNA sequences of two U3/R regions from cDNA clones from 129 Gix^+ mice (E2 and L62) are compared with sequences from the middle repetitive genomic DNA element LTR-IS B1 (IS) derived from a genomic library of 129 Gix^+ DNA (Wirth et al., 1983), from the LTR of MCF 247 virus (Kelley et al., 1983), and from the LTR of AKV (Herr, 1984) as in Figure 2. The numbers above the sequences number a composite sequence including gaps and therefore do not directly reflect the nucleotide positions in any single sequence.

FIGURE 6



B. gp70



C. p15E

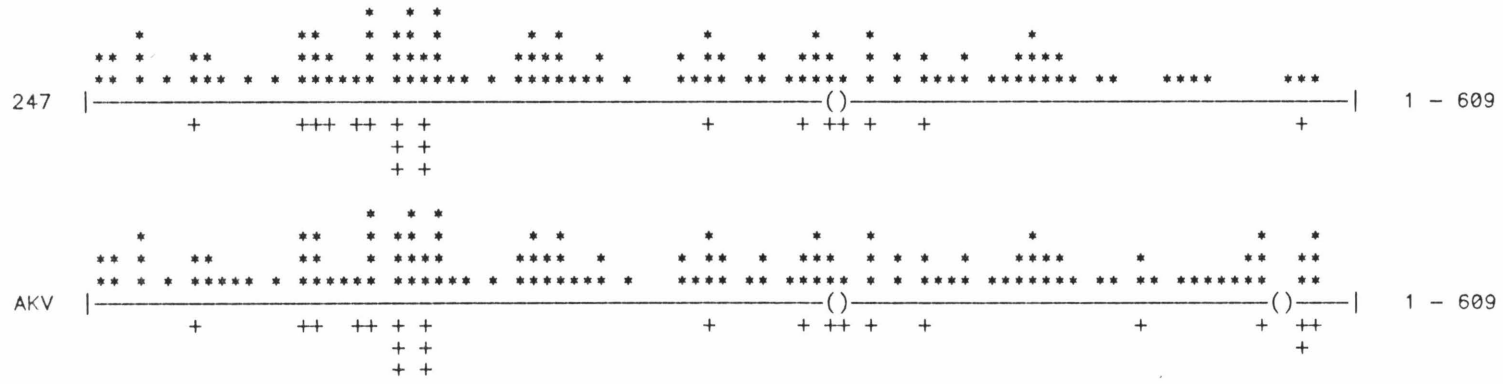


FIGURE 6

HISTOGRAM DISPLAY OF SEQUENCE SIMILARITIES

The histograms represent nucleotide and amino acid differences between the indicated sequence and the composite sequence derived for the 129 Gix^+ endogenous transcript (see text). The line represents identitiy with the 129 endogenous sequence, while asterisks (*) above the line and pluses (+) below the line indicate the approximate positions of nucleotide and amino acid differences, respectively, in the sequence indicated at the left. Parentheses represents gaps introduced in the indicated sequence to increase alignment with the 129 edogenous sequence. (A) pol region from Figure 2; (B) gp70 open reading frame from Figure 3; (C) p15e open reading frame from Figure 4. For simplicity, only the second half of the gp70 region from AKV is represented in B.

TABLE 1

SUMMARY OF HOMOLOGY RELATIONSHIPS
BETWEEN 129 ENDOGENOUS
CONTIGUOUS SEQUENCE AND VIRAL SEQUENCES

A. pol (1-360)

Sequence	Silent[1]	Coding[2]	% nuc.[3]	% a.a.[4]	No.[5]	Gap[6]
E12	15		4.8		313	1
S13	3		3.2		94	0
A12	2	1	0.8	0.8	360	0
247	2	1	0.8	0.8	360	0
MoMCF	24	12	10.0	5.8	360	0
AKV	33	20	15.2	10.3	348	1

B. gp70 (1-801)

A12	1	3	0.7	1.6	564	0
247	1	3	0.5	1.1	801	0
MoMCF	2	7	1.1	2.6	801	0

C. gp70 (802-1323)

247	44	26	13.5	8.7	519	1
MoMCF	29	40	13.3	12.7	519	1

D. p15E (1-543)

Sequence	Silent	Coding	% nuc.	% a.a.	No.	Gap
E2		3	5.1		59	0
S13		3	8.3		36	0
247	75	33	20.0	10.0	540	1
MoMCF	94	30	23.0	10.0	540	1
AKV	76	34	20.1	10.0	540	1

E. R-peptide (544-606)

E2		3	4.8		63	0
S13		1	1.6		63	0
247	3	1	6.3	4.8	63	0
MoMCF	7	12	33.3	31.6	57	1
AKV	6	7	22.8	21.1	57	1

A contiguous sequence for the 129 endogenous transcript was assumed using the longest sequences present in Tables 2-4. The table is divided into regions with the nucleotide position from the appropriate table indicated in parentheses after the region name. [1]Number of nucleotide differences in the given sequence compared to the 129 sequence that fall within synonymous codons. — [2]Number of nucleotide differences within different coding codons. (The deleted endogenous sequences E2 and S13 do not contain open reading frames; therefore, total nucleotide differences are reported.) — [3]% nucleotide difference. — [4]% predicted amino acid difference. — [5]Number of aligned nucleotides. — [6]Number of gaps introduced to create alignments.

CHAPTER FIVE

SPECULATIONS AND PROSPECTS

SPECULATIONS AND PROSPECTS

The results presented from these studies define Gv-1 as a locus, acting in trans largely at a transcriptional level, that regulates the abundance of retroviral RNA species in 129 mice. Genetic evidence has defined two loci controlling the Gix phenotype. Gv-1^a and Gv-1^b are codominant alleles at the Gv-1 locus and it is this allelic difference that defines the strains 129 Gix$^+$ and 129 Gix$^-$. Gix$^+$ and Gix$^-$ 129 strains both have a dominant allele at a second locus, Gv-2, which is required for Gix expression. A model for a mechanistic explanation of these data was presented in Chapter 3. It assumed that the Gix$^+$ phenotype resulted from a positive allele at Gv-1, producing a factor that, in combination with the Gv-2 gene product, potentiates transcription from the affected retroviral transcription units. By this model, Gv-2 encodes a DNA binding protein which specifically recognizes a sequence adjacent to retroviral promoters, perhaps within the U3 enhancer region. The Gv-1 product, in turn, binds either the Gv-2 protein at the promoter or another regulatory DNA sequence directly, and the action of this bound complex potentiates transcription from the associated viral locus. This increased transcription could be from an altered chromatin configuration in the region induced by Gv-1, allowing increased entry of Pol II. Alternatively, the Gv-1 gene product may contain a binding site for Pol II, and this serves to facilitate the placement of the polymerase molecules for transcriptional initiation. According to this model, then, the Gix$^-$ phenotype results from the production of a faulty Gv-1 product which still binds the Gv-2 protein at the promoter but fails to stimulate transcription. In the case of (129 Gix$^+$ x 129 Gix$^-$)F_1 hybrid mice, the Gv-2 DNA binding protein is produced in normal amounts as before and recognizes retroviral promoters. A mixture of wildtype Gv-1^a product and defective, Gv-1^b product would be synthesized in these mice, however. Our model requires that both the wildtype and the defective Gv-1 products must bind the Gv-2 factor at the promoter site with equal affinities. The subsequent stimulation of transcription, however, could only involve those complexes containing a wildtype, Gv-1^a product. Complexes of Gv-1^b/Gv-2, while still occupying a regulatory site at the retroviral promoter, would fail to enhance transcription. If equal amounts of wildtype, Gv-1^a and defective, Gv-1^b

products were synthesized, the resulting competition for promoter binding would produce the half levels of retroviral antigen observed in F_1 hybrid animals (Stockert et al., 1971).

The model as described here only addresses the collective regulation exhibited by Gv-1. The observed tissue specific patterns of expression could arise by a number of possible mechanisms. Additional transcription factors specifically produced in different tissue and required for transcription from individual promoters could be produced. Alternatively, tissue specific transcription may result by virtue of the integration site of each particular retroviral transcription unit. If alteration of chromatin structure or demethylation of cytosine residues are prerequisites for transcription, individual retroviruses may become available for transcription at different times and in different tissues dependent upon the activity of surrounding, cellular genes. Such dependence of transcription upon chromosomal integration site has been invoked as an explanation for the differential activities and tissue specific transcription of experimentally introduced, endogenous retroviruses (Jaenisch et al., 1981).

We feel that the above model of positive transcription factors binding at independent retroviral promoters in the mouse is the more appealing model to explain the coordinate regulation of the Gix phenotype. However, before designing experimental approaches to test this hypothesis, it is important to consider an alternative to this model. The Gix phenotype could be considered in terms of repressors rather than activators. By this alternative model, the Gix^- phenotype would result from the production of an efficacious repressor molecule encoded by $Gv\text{-}1^b$. This repressor would bind the Gv-2 transcription factor, inhibiting its effect at retroviral promoters. A defective gene sequence, $Gv\text{-}1^a$, encoding either no Gv-1 product or one unable to interact with the Gv-2 product or with DNA, would allow the Gv-2 factor to stimulate transcription at retroviral promoters and the Gix^+ phenotype would be observed. To explain the half levels of antigen produced in F_1 hybrid animals, one must propose that the Gv-1 repressor product is synthesized in limiting amounts so that a half gene dose results in half repression. Alternatively, a model of competition between wildtype and defective Gv-1

repressors for the Gv-2 product could be proposed, again with both types of Gv-1 products able to bind but with only the $Gv\text{-}1^{b}$ repressor effectively inhibiting transcription. Unfortunately, available genetic data do not distinguish repression from activation models.

Evidence of the interaction between host gene products and transposable genetic elements has been observed in other systems. In Drosophila melanogaster, a number of spontaneous mutations are apparently the result of the insertion of transposable elements in the vicinity of the affected gene (see Rubin et al., 1982). Some of these mutations are suppressed in response to recessive mutations in unlinked host genes. For example, a number of mutations suppressed by suppressor of Hairy-wing [su(Hw)] correlate with the insertion of a gypsy transposable element (Modelell et al., 1983). These mutations do not appear to result from disruption of structural genes directly by the insertion of the transposon (Freund and Meselson, 1984) but rather derive their mutant phenotype as a result of the disruption of transcription at the structural gene due to an interaction between a product of the wildtype copy of the separate suppressor gene, $su(Hw)^{+}$, and the transposon. Mutations at the suppressor locus, in turn, eliminate this interaction between host factor and transposon, thus restoring wildtype expression at the affected structural gene. Homozygous $su(Hw)^{-}$ mutant flies show a five fold reduction in gypsy transcription (Jackson, 1984). A model for suppressor mutation action consistent with these data is that stimulation of transcription from the gypsy transposon promoter, dependent upon the host $su(Hw)^{+}$ gene product, disrupts transcription from nearby cellular genes. Suppression of the mutant phenotype by the second site mutation results from inactivation of the gene for the transcription factor required by gypsy, thus allowing normal transcription of the affected cellular genes to ensue.

Data from other examples of suppressible mutations fit this general model of such mutations resulting from the transcription of a transposable element, dependent upon a host transcription factor, disrupting transcription from affected loci. The white-apricot phenotype is suppressible and is caused by the insertion of the transposon copia within the transcription unilt but non-coding region of the gene for eye color (Bingham et al.,

1981). In yeast, mutations at the HIS4 locus have been described which result from insertion of the Ty element (Chaleff and Fink, 1980). The his4-912 mutation, which results from Ty insertion into the non-transcribed flanking region of the HIS4 gene, is suppressible by unlinked host genes which affect transcription of the inserted transposon (Silverman and Fink, 1984). A coat color mutation (dilute) of DBA/2 mice is caused by insertion of an ecotropic retrovirus at the agouti locus (Jenkins et al., 1981), presumably into non-coding or non-transcribed regions of the coat color gene (Copeland et al., 1983). Although direct involvement of transcription from this retrovirus has not been implicated in the dilute mutation, this virus is probably expressed in DBA/2 mice (Risser et al., 1983). Importantly, the dilute phenotype is suppressible by a mutation at a second locus, the unlinked host gene, dsu (Sweet, 1983). The wildtype products of these suppressor genes described here, from a variety of species, may be similar to Gv-1: transcription factors that act in trans to stimulate expression of endogenous transposable elements.

Further definition of the Gix phenotype must concentrate on understanding the molecular mechanisms of regulation by the action of the Gv-1 and Gv-2 encoded products. One aspect of this phenotype which should be addressed more fully would be an accurate description of the organization of the responding structural genes. Throughout this study, it has been assumed that the expressed proviruses of 129 Gix^+ mice are dispersed throughout the genome since this is the general case of endogenous retroviruses. If, on the other hand, it were found that these active transcription units display an unexpected clustered organization, the regulation directed by Gv-1 could be fundamentally different from the presented model. One approach to mapping the organization of expressed proviruses would involve recombinant inbred mice. A recombinant inbred series derived from 129 Gix^+ mice and a second strain that expresses structurally distinct virus related transcripts should show independent inheritance of the expressed loci, if they are indeed unlinked in the genome. An analysis of this type might also reveal the chromosomal locations of expressed loci, relative to other, previously mapped, genetic loci. In this regard, we have determined that NZB and BALB/c mouse strains express structurally distinct viral transcripts from those expressed

by 129 (not shown). Five (NZB x 129) recombinant inbred strains have been derived at Jackson Labs. Unfortunately, this is an insufficient number for adequate linkage analysis (Taylor, 1978), so this study must await the derivation of further R.I. strains.

An initial step to a fuller description of the regulation described here would involve cloning and characterizing the DNA sequences at the Gv-1 locus, which encode the regulatory element, and at the affected retroviral transcription units, which respond to this regulation. Evaluation of flanking sequences which might potentially respond to Gv-1 regulation requires an assay system that reflects a change in transcription mediated by Gv-1. Cloning and characterizing the Gv-1 gene requires a selection system that applies a differential growth ability under the control of Gv-1. These two problems are interrelated and are also dependent upon the correctness of the model as outlined above.

The regulation of the Gix phenotype by the Gv-1 locus could be dependent upon either activator or repressor molecules. Thus, any selection schemes designed to facilitate cloning the Gv-1 locus or any test assays for measuring the response to Gv-1 must take both possibilities into consideration. It is therefore necessary to design systems that will distinguish between a positive allele at Gv-1 enhancing transcription from endogenous retroviral promoters and such an allele inhibiting retroviral transcripton. Either action by Gv-1 could be exploited in selection schemes for further characterization of this locus. For example, assuming that Gv-1 encodes a positive activating transcription factor, an assay system based upon transforming cells from Gix^- phenotype to Gix^+ phenotype could be devised. Conversely, if Gv-1 encodes a repressor molecule, then selection schemes must choose Gix^- cells. In either case, it will be necessary to obtain cell lines which reflect the Gix phenotype. Mouse embryo fibroblast cells from 129 Gix^+ and 129 Gix^- mice could be established and evaluated by RNA blotting and immunofluorescence for expression of endogenous retroviral products. Also, hepatocyte cultures from liver or hybridomas from B cells or T cells might also be considered.

One possibility for selecting cells rendered Gix^+ could use dominant selectable markers placed under Gv-1 control. For example, if transcription of the Tn5 phosphotransferase gene from the plasmid pSV2neo (Southern and Berg, 1982) were transfered from the control of the SV40 early region promoter to the control of a Gv-1 responsive promoter, such a construct could be used to confer a selective growth advantage on Gix^+ cells by conferring $G418^r$. This construct could be used to create a test system for cloning Gv-1. If a Gix^- cell line were cotransformed with a Gv-1 responsive, $G418^r$ marker and Gix^+ genomic DNA, only cells receiving the Gv-1 locus would survive G418 selection. The use of marked, Gix^+ genomic DNA (e.g., previously cloned into a lambda vector) would allow the transferred genomic DNA in the resulting $G418^r$ clones to be identified and characterized.

Preliminary experiments to evaluate this scheme have been initiated. Due to the terminal redundancy of retroviral transcription units, potential promoter sequences contained in U3 regions can be recovered in cDNA from the 3' end of retroviral mRNA transcripts (see Chapter 4). One such sequence from clone E2 has been used to replace the SV40 promoter of the plasmid pSV2neo. This construct has been found to direct the synthesis of a Tn5 related transcript when introduced into mouse L cells (Figure 1). Once Gix^+ and Gix^- cell lines are available, such a construct could be evaluated for susceptibility to Gv-1 control. However, this may not be the best test system. The use of cDNA as a source of promoter elements necessarily evaluates only the 3' end U3 region, a promoter that is active in vivo in Gix^+ mice cannot, therefore, be obtained from cDNA. In some cases, comparison of the U3 regions from 3' and 5' ends of retrovirus related proviruses has found them to be different (Steele et al., 1984). Furthermore, the use of a 3' U3 may be limiting since only the promoter element itself can be evaluated without contribution from the flanking DNA milieu which may contain regulatory elements. For example, the transfected human and rabbit globin genes require sequences both 5' and 3' to the transcriptional initiation site for faithful, regulated expression (Charnay et al., 1984; Wright et al., 1984). Also, the specificity of expression of endogenous retroviruses may depend upon chromosomal location (Jaenisch et al., 1981). In some cases, differences

in expression between different endogenous retroviruses have been found not to correlate with sequence differences in the LTR (Scholl et al., 1983). For these reasons, it would probably be more advantageous to obtain 5' retroviral promoter and associated flanking regions for these constructs.

Another potential problem with the scheme outlined above is that it necessates isolating the test gene from its normal environment, re-introducing it into a new cell, and requiring it to respond correctly. A selection process for genomic DNA sequences that does not depend upon identifying and isolating Gv-1 responsive transcription units could exploit instead insertion mutagenesis to identify Gv-1. This process would involve transfecting Gix^+ cells with a vector DNA sequence which could then randomly integrate in the genome. Cells selected for the loss of Gix expression might contain the foreign DNA inserted at the Gv-1 locus in such a way that it disrupts transcription of this gene. Other Gix^- cells resulting from this strategy could contain insertion mutations in previously active, retroviral structural genes. This scheme is similar to the effect of the Mov13 locus, in which an inserted retrovirus disrupts a collagen gene (Breindl et al., 1984). A protocol such as this would be facilitated by using a cell line derived from $(Gix^+ \times Gix^-)F_1$ hybirds so that the recipient cells will contain only a single dose of the Gv-1 gene, thereby enabling it to be disrupted by a single-hit event. Selection for the elimination of the Gix phenotype requires a growth advantage for Gix^- cells. This might be experimentally achieved using a cytotoxic, anti-gp70 antiserum such as was used initially for the characterization of the Gix phenotype (Stockert et al., 1971).

Another possible protocol that exploits endogenous retroviral transcription in their normal chromosomal environment could directly select for the conversion of Gix^- cells to Gix^+ by the production of cell surface gp70. By this scheme, a Gix^- cell line would be co-transfected with Gix^+ genomic DNA and a dominant selectable marker under independent control. Cells transformed for the selectable marker could then be evaluated for cell surface expression of retroviral antigens by using a fluorescent antibody reagent coupled with selection through a cell sorter. Such a

protocol has been successfully employed in several recent cloning ventures (Newman et al., 1983; Kavathas and Herzenberg, 1983; Kühn et al., 1984).

The selection strategies outlined above all assume that the Gix^+ phenotype results from a positive activator stimulating transcription from retroviral transcription units. Each of these protocols could be adapted to identify a repressor gene in case Gv-1 encodes a negative transcription factor. For example, Gix^- cells transfected with vector DNA could be evaluated for insertion mutagenesis of a repressor gene by selection of cells initiating retroviral antigen expression by using fluorescent antibodies and the cell sorter. Similarly, Gix^+ cells transfected with Gix^- genomic DNA could be selected for the transfer of a repressor locus by eliminating the remaining Gix^+ cells with a cytotoxic antiserum.

None of these procedures for selecting for either a positive or negative regulatory factor encoded by the Gv-1 locus will be trivial. They rely upon many difficult steps such as establishment of cell lines that accurately reflect the Gix phenotype, identification of responsive transcription units out of the many copies of endogenous retrovirus, the establishment of effective selection strategies, and the cloning of a single gene out of the mouse genome. However, the identification and characterization of the Gv-1 encoded product and the localization of the sequences upon which it acts are the necessary steps toward a fuller understanding of this regulatory process.

FIGURE 1

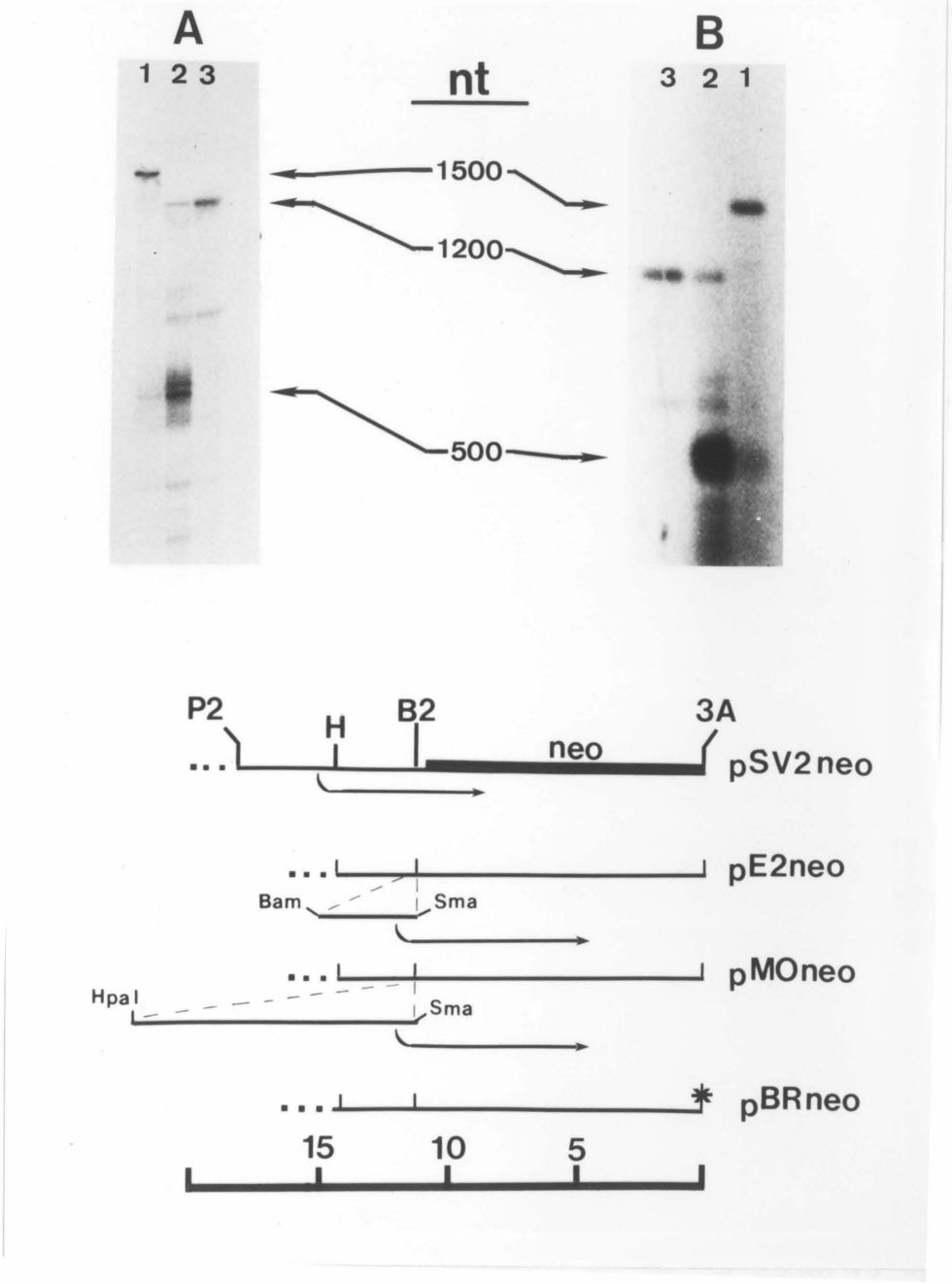

A
1 2 3
nt
1500
1200
500
B
3 2 1
P2
H
B2
neo
3A
pSV2neo
pE2neo
Bam
Sma
pMOneo
HpaI
Sma
pBRneo
15
10
5

FIGURE 1

S1 ANALYSIS OF *neo* TRANSCRIPTS

Mouse L cells were transfected by the method of Graham and Van der Eb, 1973 with the plasmid DNA constructs illustrated at the bottom of the figure. pSV2neo (Southern and Berg, 1982) was the parent plasmid in the constructions. pE2neo and pMOneo contain the *neo* gene from pSV2neo without the SV40 origin and early promoter region. These plasmids contain potential promoter regions from the 129 cDNA plasmid, E2, and the MoMuLV clone, pMLV-1 (Berns et al., 1980), respectively. The cDNA sequence of pE2neo is a Bam HI to Sma I fragment covering the entire U3 region and a portion of p15E coding sequence. The viral sequence of pMOneo is a Hpa I to Sma I fragment from nucleotides 7200 to 8300 in the MoMuLV sequence of Shinnick et al. (1981) containing p15E coding sequences and the viral U3 region. These potential promoter regions were cloned into the unique Bgl II site of the *neo* gene, just 5' to the protein translational start site. The SV40 origin was elimiated by re-cloning a Hind III to Eco RI fragment into the plasmid, pBR322. L cells transfected with these constructs were grown in the presence of G418, and resistant cells were harvested and used for RNA isolation. The RNA from $G418^r$ cells was analyzed for *neo* transcripts by S1 nuclease protection using as a probe pBRneo, a plasmid containing the bacterial *neo* gene (Southern and Berg, 1982). This plasmid was end labeled at the unique Bam HI (Sau 3A) site by using $[^{32}P]$rATP and T4 polynucleotide kinase. As indicated in the diagram, transcripts originating from the normal SV40 origin of pSV2neo or from the predicted start sites of pE2neo and pMOneo will protect labeled fragments of approximately 1.5 kb (pSV2neo) and 1.2 kb (pE2neo and pMOneo). However, due to the complementarity of the probe to the plasmid constructs, this analysis cannot distinguish normally initiated transcripts from potential ones initiated 5' to the promoter regions since all such transcripts would fully protect the *neo* gene from the Hind III site (pSV2neo) or the Bgl II site (retroviral constructs). The major fragment of 500 nt protected by RNA from pE2neo (Lane 2), however, is consistent with a transcriptional initiation within the 3' portion of the *neo* gene.

Lanes 1, pSV2neo transformed cells; lanes 2, pE2neo transformed cells; lanes 3, pMOneo transformed cells. Panel A depicts an S1 analysis on a 4% acrylamide-urea gel; Panel B depicts the same samples analyzed on a 1.5% alkaline agarose gel. In the diagram of the recombinant constructs, the restriction sites indicated for pSV2neo are shared by all four plasmids. P2, Pvu II; H, Hind III; B2, Bgl II; 3A, Sau 3A. The scale at the bottom is in hundreds of nucleotides.

REFERENCES

Albertini, A. M., Hofer, M., Calos, M. P., and Miller, J. H. (1982). On the formation of spontaneious deletions: the importance of short sequence homologies in the generation of large deletions. Cell **29**: 319-328.

Alexander, S., Shinnick, T. M., and Lerner, R. A. (1983). Mutants of Dictyostelium discoideum blocked in expression of all members of the developmentally regulated discoidin multigene family. Cell **34**: 467-475.

Amanuma, H., Katori, A., Obata, M., Sagata, N., and Ikawa, Y. (1983). Complete nucleotide sequence of the gene for the specific glycoprotein (gp55) of Friend spleen focus-forming virus. Proc. Natl. Acad. Sci. USA **80**: 3913-3917.

Ashburner, M. and Bonner, J. J. (1979). The induction of gene activity in Drosophila by heat shock. Cell **17**: 241-254.

Aviv, H., and Leder, P. (1972). Purification of biologically active globin messenger RNA by chromotography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**: 1408-1412.

Bailey, J. M., and Davidson, N. (1976). Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. Anal. Biochem. **70**: 75-85.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell **33**: 729-740.

Barbacid, M., Stephenson, J. R., and Aaronson, S. A. (1977). Evolutionary relationships between gag gene-coded proteins of murine and primate endogenous type-C RNA viruses. Cell **10**: 641-648.

Basombrio, M. A. (1970). Search for common antigenicities among twenty-five sarcomas induced by methycholanthrene. Cancer Res. **30**: 2458-2462.

Basombrio, M. A., and Prehn, R. T. (1972). Antigenic diversity of tumors chemically induced within the progeny of a single cell. Int. J. Cancer **10**: 1-8.

Bassin, R. H., Ruscetti, S., Ali, I., Haapala, D., and Rein, A. (1982). Normal DBA/2 mouse cells synthesize a glycoprotein which interferes with MCF virus infection. Virology **123**: 139-151.

Battula, N., and Loeb, L. A. (1974). The infidelity of avian myeloblastosis virus deoxyribonucleic acid polymerase in polynucleotide replication. J. Biol. Chem. **249**: 4086-4093.

Battula, N., and Loeb, L. A. (1975). On the fidelity of DNA replication. Characterization of polynucleotides with errors in base-pairing synthesized by avian myeloblastosis virus deoxyribonucleic acid polymerase. J. Biol. Chem. **250**: 4405-4409.

Berk, A. J., and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**: 721-732.

Berns, A. J. M., Lai, M. H. T., Bosselman, R. A., McKennett, M. A., Bacheler, L. T., Fan, H., Maandag, E. C. R., van der Putten, H., and Verma, I. M. (1980). Molecular cloning of unintegrated and a portion of integrated Moloney murine leukemia viral DNA in bacteriophage lambda. J. Virol. **36**: 254-263.

Bestwick, R. K., Boswell, B. A., and Kabat, D. (1984). Molecular cloning of biologically active Rauscher spleen focus-forming virus and the sequences of its env gene and long terminal repeat. J. Virol. **51**: 695-705.

Biggin, M., ^{35}D., Gibson, T. J., and Hong, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**: 3963-3965.

Bingham, P. M., and Judd, B. H. (1981). A copy of the copia transposable element is very tightly linked to the w^a allele at the white locus of D. melanogaster. Cell **25**: 705-711.

Birnboim, H. L., and Doly, J. (1979). A rapid alkaline extraction procedure of screening recombinant plasmid DNA. Nucleic Acids Res. **7**: 1513-1523.

Bishop, J. M. (1978). Retroviruses. Ann. Rev. Biochem. **47**: 35-88.

Bishop, J. M. (1983). Cellular oncogenes and retroviruses. Ann. Rev. Biochem. **52**: 301-354.

Blatt, C., Mileham, K., Haas, M., Nesbitt, M., N., Harper, M. E., and Simon, M. I. (1983). Chromosomal mapping of the mink cell focus-inducing and xenotropic env gene family in the mouse. Proc. Natl. Acad. Sci. USA **80**: 6298-6302.

Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes. J. Cell Biol. **67**: 835-851.

Boccara, M., Souyri, M., Maragian, C., Stavnezer, E., and Fleissner, E. (1983). Evidence for a new form of retroviral env transcript in leukemic and normal mouse lymphoid cells. J. Virol. **48**: 102-109.

Bosselman R. A., van Straaten, F., Van Beveren C., Verma, I. M., and Vogt, M. (1982). Analysis of the env gene of a molecularly cloned and biologically active Moloney mink cell focus-forming proviral DNA. J. Virol. **44**: 19-31.

Boyse, E. A. (1977). The Gix system in relation to C-type viruses and heredity. Immunol. Rev. **33**: 125-145.

Breindl, M., Harbers, K., and Jaenisch, R. (1984). Retrovirus-induced lethal mutation in collagen I gene of mice is associated with an altered

chromatin structure. Cell **38**: 9-16.

Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. New facts regarding the organization of the genome provide clues to the nature of gene regulation. Science **165**: 349-357.

Broach, J. R. (1979). Galactose regulation in Saccharomyces cerevisiae. The enzymes encoded by the GAL7,10,1 cluster are co-ordinately controlled and separately translated. J. Mol. Biol. **131**: 41-53.

Brunke, K. J., Anthony, J. G., Sternberg, E. J., and Weeks, D. P. (1984). Repeated consensus sequence and pseudopromoters in the four coordinately regulated tubulin genes of Chlamydomonas reinhardi. Mol. Cell. Biol. **4**: 1115-1124.

Buckler, C. E., Hoggan, M. S., Chan, H. W., Sears, J. F., Khan, A. S., Moore, J. L., Hartley, J. W., Rowe, W. P., and Martin, M. A. (1982). Cloning and characterization of an envelope-specific probe from xenotropic murine leukemia proviral DNA. J. Virol. **41**: 228-236.

Casadaban, M. J., and Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. J. Mol. Biol. **138**: 179-207.

Casey, J., and Davidson, N. (1977). Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucleic Acids Res. **4**: 1539-1552.

Chaleff, D. T., and Fink, G. R. (1980). Genetic events associated with an insertion mutation in yeast. Cell **21**: 227-237.

Chambon, P. (1975). Eucaryotic nuclear RNA polymerases. Ann. Rev. Biochem. **44**: 613-638.

Chan, H. W., Bryan, T., Moore, J. L., Staal, S. P., Rose, W. P., and Martin, M. A. (1980). Structure of endogenous murine leukemia virus DNA in mouse genomes. Proc. Natl. Acad. Sci. USA **77**: 5774-5778.

Charnay, P., Treisman, R., Mellon, P., Chao, M., Axel, R., and Maniatis, T. (1984). Differences in human a- and b-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell **38**: 251-263.

Chatis, P. A., Holland, C. A., Hartley, J. W., Rowe, W. P., and Hopkins, N. (1983). Role for the 3' end of the genome in determining disease specificity of Friend and Moloney murine leukemia viruses. Proc. Hatl. Acad. Sci. USA **80**: 4408-4411.

Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S., and Rowe, W. P. (1974). Qualitative and quantitative studies of AKR-type murine leukemia virus sequences in mouse DNA. Cold Spring Harbor Symp. Quant. Biol. **39**: 1085-1101.

Chattopadhyay, S. K., Lander, M. R., Rands, E., and Lowy, D. R. (1980). Structure of endogenous murine leukemia virus DNA in mouse genomes.

Proc. Natl. Acad. Sci. USA **77**: 5774-5778.

Chattopadhyay, S. K., Lander, M. R., Gupta, S., Rands, E., and Lowy, D. R. (1981). Origin of mink cytopathic focus-forming (MCF) viruses: comparison with ecotropic and xenotropic murine leukemia virus genomes. Virology **113**: 465-483.

Chattopadhyay, S. K., Cloyd, M.W., Linemeyer, D. L., Lander, M. R., Rands, E., and Lowy, D. R. (1982). Cellular origin and role of mink cell focus-forming ciruses in murine thymic lymphomas. Nature (London) **295**: 24-31.

Chen, R. (1982). Complete amino acid sequence and glycosylation sites of glycoprotein gp71A of Friend murine leukemia virus. Proc. Natl. Acad. Sci. USA **79**: 5788-5792.

Chiu, I.-M., Callahan, R., Tronick, S. R., Schlom, J., and Aaronson, S. A. (1984). Major pol gene progenitors in the evolution of oncoviruses. Science **223**: 364-370.

Chiu, I.-M., Callahan, R., Tronick, S. R., Schlom, J., and Aaronson, S. A. (1984). Major pol gene progenitors in the evolution of oncoviruses. Science **223**: 364-370.

Clewell, D. B., and Helinski, D. R. (1970). Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event. Biochemistry **9**: 4428-4440.

Copeland, N. G., Hutchison, K. W., and Jenkins, N. A. (1983). Excision of the DBA ecotropic provirus in dilute coat color revertants of mice occurs by homologous recombination involving the viral LTRs. Cell **33**: 379-387.

Dagert, M., and Ehrlich, S. D. (1979). Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene **6**: 23-28.

Darnell, J. E. (1982). Variety in the level of gene control in eukaryotic cells. Nature (London) **297**: 365-371.

Darnell, J. E., Jr., Ziff, E., Nevins, J. R., and Wilson, M. (1980). Variety in the mechanisms of adenovirus gene control. Ann. N. Y. Acad. Sci. **354**: 466-471.

Davidson, E. H., and Britten, R. J. (1979). Regulation of gene expression: possible role of repetitive sequences. Science **204**: 1052-1059.

Davidson, E. H., Jacobs, H. T., and Britten, R. J. (1983). Very short repeats and coordinate induction of genes. Nature **301**: 468-470.

Daynes, R. A., Spellman, C. W., Woodward, J. G., and Stewart, D. A. (1977). Studies into the transplantation biology of ultraviolet light-induced tumors. Transplantation **23**: 343-348.

Del Villano, B. C., Nave, B., Croker, B. P., Lerner, R. A., and Dixon, F. J. (1975). The oncornavirus glycoprotein gp69/71: a constituent of the surface of normal and malignant thymocytes. J. Exp. Med. **141**: 172–187.

Del Villano, B. C., and Lerner, R. A. (1976). Relationship between the oncornavirus gene product gp70 and a major protein secretion of the mouse genital tract. Nature (London) **259**: 497–499.

Derman, E., Krauter, K., Walling, L., Weinberger, C., Ray, M., and Darnell, J. E., Jr. (1981). Transcriptional control in the production of liver-specific mRNAs. Cell **23**: 731–739.

DesGroseillers, L., Tassart, E., and Jolicoeur, P. (1983). Thymotropism of murine leukemia virus is conferred by its long terminal repeat. Proc. Natl. Acad. Sci. USA **80**: 4203–4207.

Devare, S. G., Rapp, U. R., Todaro, G. J., and Stephenson, J. R. (1978). Acquisition of oncogenicity by endogenous mouse type C viruses: effects of variations in env and gag genes. J. Virol. **28**: 457–465.

Dinowitz, M. (1975). Inhibition of Rous sarcoma virus by alpha-amanitin: possible role of cell DNA-dependent RNA polymerase form II. Virology **66**: 1–9.

Doerfler, W. (1983). DNA methylation and gene activity. Ann. Rev. Biochem. **52**: 93–124.

Doerfler, W. (1984). DNA methylation and its functional significance: studies on the adenovirus system. Curr. Top. Microbiol. Immunol. **108**: 79–98.

Donis-Keller, H., Rommelaere, J., Ellis, R. W., and Hopkins, N. (1980). Nucleotide sequences associated with differences in electrophoretic mobility of envelope glycoprotein gp70 and with Gix antigen phenotype of certain murine leukemia viruses. Proc. Natl. Acad. Sci. USA **77**: 1642–1645.

Douglas, H. C., and Hawthorne, D. C. (1966). Regulation of genes controlling synthesis of the galactose pathway enzymes in yeast. Genetics **54**: 911–916.

DuBois, G. C., Apella, E., Law, L. W., DeLeo, A. B., and Old, L. J. (1980). Immunogenic properties of soluble cytosol fractions of Meth A sarcoma. Cancer Res. **40**: 4204–4208.

Dynan, W. S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**: 79–87.

Efstratiadis, A., Kafatos, F. C., Maxam, A. M., and Maniatis, T. (1976). Enzymatic in vitro synthesis of globin genes. Cell **7**: 279–288.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. D., Smithies, O., Baralle, F. E., Shoulders,

C. C., and Proudfoot, N. J. (1980). The structure and evolution of the human b-globin gene family. Cell **21**: 653-668.

Eickbush, T. H., and Kafatos, F. C. (1982). A walk in the chorion locus of Bombyx mori. Cell **29**: 633-643.

Elder, J. H., Gautsch, J. W., Jensen, F., Lerner, R. A., Hartley, J. W., and Rowe, W. P. (1977). Biochemical evidence that MCF murine leukemia viruses are envelope (env) gene recombinants. Proc. Natl. Acad. Sci. USA **74**: 4676-4680.

Elder, J. H., Jensen, F., Bryand, M., and Lerner, R. A. (1977). Polymorphism of the major envelope glycoprotein (gp70) of murine C-type viruses: virion associated and differentiation antigens encoded by a multi-gene family. Nature (London) **267**: 23-28.

Elder, J. H., and Mullins, J. I. (1983). Nucleotide sequence of the envelope gene of Gardner-Arnstein feline leukemia virus B reveals unique sequence homologies with a murine mink cell focus-forming virus. J. Virol. **46**: 871-880.

Embleton, M. J., and Heidelberger, C. (1972). Antigenicity of clones of mouse prostate cells transformed in vitro. Int. J. Cancer **9**: 8-18.

Engelman, D. M., and Steitz, T. A. (1981). The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell **23**: 411-422.

Evans, R., Fraser, N., Ziff, E., Weber, J., Wilson, M., and Darnell, J. E. (1977). The initiation sites for RNA transcription in Ad2 DNA. Cell **12**: 733-739.

Even, J., Anderson, S. J., Hampe, A., Galibert, F., Lowy, D., Khoury, G., and Sherr, C. J. (1983). Mutant feline sarcoma proviruses containing the viral oncogene (v-fes) and either feline or murine control elements. J. Virol. **45**: 1004-1016.

Faller, D. V., Rommelaere, J., and Hopkins, N. (1978). Large T1 oligonucleotides of Moloney leukemia virus missing in an env gene recombinant, HIX, are present on an intracellular 21S Moloney viral RNA species. Proc. Natl. Acad. Sci. USA **75**: 2964-2968.

Farabaugh, P. J., Schmeissner, U., Hofer, M., and Miller, J. H. (1978). Genetic studies of the lac repressor. VII. On the molecular nature of spontaneous hotspots in the lacI gene of Escherichia coli. J. Mol. Biol. **126**: 847-863.

Freund, R. and Meselson, M (1984). Long terminal repeat nucleotide sequence and specific insertion of the gypsy transposon. Proc. Natl. Acad. Sci. USA **81**: 4462-4464.

Fuhrman, S. A., Van Beveren, C., and Verma, I. M. (1981). Identification of a RNA polymerase II initiation site in the long terminal repeat of Moloney murine leukemia viral DNA. Proc. Natl. Acad. Sci. USA **78**:

5411-5415.

Fujimura, F. K., Deininger, P. L., Friedmann, T., and Linney, E. (1981). Mutation near the polyoma DNA replication origin permits productive infection of F9 embryonal carcinoma cells. Cell **23**: 809-814.

Gilboa, E., Mitra, S. W., Goff, S., and Baltimore, D. (1979). A detailed model of reverse transcription and tests of crucial aspects. Cell **18**: 93-100.

Gilden, R. V. (1975). Interrelationships among RNA tumor viruses and host cells. Adv. Cancer Res. **22**: 157-202.

Gillies, S. D., Morrison, S. L., Oi, V. T., and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell **33**: 717-728.

Gillies, S. D., Folson, V., and Tonegawa, S. (1984). Cell type-specific enhancer element associated with a mouse MHC gene, E_B. Nature (London) **310**: 594-597.

Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**: 1044-1051.

Graham, F. L., and Van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**: 456-467.

Green, N., Shinnick, T. M., Witte, O., Ponticelli, A., Sutcliffe, J. G., and Lerner, R. A. (1981). Sequence-specific antibodies show that maturation of Moloney leukemia virus envelope polyprotein involves removal of a COOH-terminal peptide. Proc. Natl. Acad. Sci. USA **78**: 6023-6027.

Green, M. R., Treisman, R., and Maniatis, T. (1983). Transcriptional activation of cloned human β-globin genes by viral immediate-early gene products. Cell **35**: 137-148.

Gross, L. (1970). Oncogenic Viruses. Pergamon, New York: 1-991.

Groudine, M., Eisenman, R., and Weintraub, H. (1981). Chromatin structure of endogenous retroviral genomes and activation by an inhibitor of DNA methylation. Nature (London) **292**: 311-317.

Groudine, M., Peretz. M., and Weintraub, H. (1981b). Transcriptional regulation of hemoglobin switching in chicken embryos. Mol. Cell. Biol. **1**: 281-288.

Gruss, P., Dhar, R., and Khoury, G. (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. USA **78**: 943-947.

Hackett, R. W., and Lis, J. T. (1981). DNA sequence analysis reveals extensive homologies of regions preceding *hsp70* and *ab* heat shock genes in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **78**: 6196-6200.

Hanahan, D., and Meselson, M. (1980). Plasmid screening at high colony density. Gene **10**: 63-67.

Hara, I., Shozo, I., and Dixon, F. J. (1982). Murine serum glycoprotein gp70 behaves as an acute phase reactant. J. Exp. Med. **155**: 345-357.

Harpold, M. M., Evans, R. M., Salditt-Georgieff, M., and Darnell, J. E. (1979). Production of mRNA in CHinese hamster cells: relationship of the rate of synthesis to the cytoplasmic concentration of nine specific mRNA sequences. Cell **17**: 1025-1035.

Hellström, K. E., Hellström, I., and Brown, J. P. (1978). Unique and common tumor-specific transplantation antigens of chemically induced mouse sarcomas. Int. J. Cancer **21**: 317-322.

Hellström, J. E., and Brown, J. P. (1979). Tumor antigens. The Antigens **5**: 1-82.

Henderson. L. E., Copeland, T. D., Smythers, G. W., Marquardt, H., and Oroszlan, S. (1978). Amino-terminal amino acid sequence and carboxy-terminal analysis of Rauscher murine leukemia virus glycoproteins. Virology **85**: 319-322.

Herr, W. (1984). Nucleotide sequence of AKV murine leukemia virus. J. Virol. **49**: 471-478.

Hinnebusch, A. G., and Fink, G. R. (1983). Repeated DNA sequences upstream from HIS1 also occur at several other co-regulated genes in Saccharomyces cerevisiae. J. Biol. Chem. **258**: 5238-5247.

Hoffmann, J. W., Steffen, D., Gusella, J., Tabin, C., Bird, S., Cowing, D., and Weinberg, R. A. (1982). DNA methylation affecting the expression of murine leukemia proviruses. J. Virol. **44**: 144-157.

Hoggan, M. D., Buckler, C. E., Sears, J. F., Rowe, W. P., and Martin, M. A. (1983). Organization and stability of endogenous xenotropic murine leukemia virus proviral DNA in mouse genomes. J. Virol. **45**: 473-477.

Hokama, L. H., Devi, B. G., Mullins, J. I., and Roy-Burman, P. (1983). Molecular cloning and characterization of endogenous feline leukemia virus sequences from a cat genomic library. J. Virol. **46**: 829-840.

Holland, C. A., Wozney, J., and Hopkins, N. (1983). Nucleotide sequence of the gp70 gene of murine retrovirus MCF 247. J. Virol. **47**: 413-420.

Holmgren, R., Corces, V., Morimoto, R., Blackman, R., and Meselson, M. (1981). Sequence homologies in the 5' regions of four Drosophila heat-shock genes. Proc. Natl. Acad. Sci. USA **78**: 3775-3778.

Hood, L., Campbell, J. H., and Elgin, S. C. R. (1975). The organization, expression, and evolution of antibody genes and other multigene families. Ann. Rev. Genet. **9**: 305-353.

Hood, L., Huang, H. V., and Dreyer, W. J. (1977). The area-code hypothesis: the immune system provides clues to understanding the genetic and molecular basis of cell recognition during development. J. Supramol. Struct. **7**: 531-559.

Horowitz, J., and Risser, R. (1982). A locus that enhances the induction of endogenous ecotropic murine leukemia viruses is distinct from genome-length ecotropic proviruses. J. Virol. **44**: 950-957.

Hubbard, A. L., and Cohn, Z. A. (1972). The enzymatic iodination of the red cell membrane. J. Cell Biol. **55**: 390-405.

Imperiale. M. J., Feldman, L. T., and Nevins, J. R. (1983). Activation of gene expression by adenovirus and herpesvirus regulatory genes acting in trans and by a cis-acting adenovirus enhancer element. Cell **35**: 127-136.

Imperiale, M. J., Kao, H.-T., Feldman, L. T., Nevins, J. R., and Strickland, S. (1984). Common control of the heat shock gene and early adenovirus genes: evidence for a cellular E1A-like activity. Mol. Cell. Biol. **4**: 867-874.

Imperiale, M. J., and Nevins, J. R. (1984). Adenovirus 5 E2 transcription unit: an E1A-inducible promoter with an essential element that functions independently of position or orientation. Mol. Cell. Biol. **4**: 875-882.

Jackson, I. J. (1984). Transposable elements and suppressor genes. Nature (London) **309**: 751-752.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. **3**: 318-356.

Jaenisch, R., Jähner, D., Nobis, P., Simon, I., Löhler, J., Harbers, K., and Grotkopp, D. (1981). Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell **24**: 519-529.

Jenkins, N. A., Copeland, N. G., Taylor, B. A., and Lee, B. K. (1981). Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. Nature (London) **293**: 370-374.

Jenkins, N. A., Copeland, N. G., Taylor, B. A., and Lee, B. K. (1982). Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of Mus musculus. J. Virol. **43**: 26-36.

Jongstra, J., and Moroni, C. (1981). Lipopolysaccharide induces retroviral antigen expression in 129/J mouse lymphocytes: evidence for assembly of a defective viral particle. J. Virol. **37**: 1044-1050.

Kavathas, P., and Herzenberg, L. A. (1983). Stable transformation of mouse L cells for human membrane T-cell differentiation antigens, HLA and b_2-microglobulin: selection by fluorescence-activated cell sorting. Proc. Natl. Acad. Sci. USA **80**: 524-528.

Kelley, M., Holland, C. A., Lung, M. L., Chattopadhyay, S. K., Lowy, D. R., and Hopkins, N. H. (1983). Nucleotide sequence of the 3' end of MCF 247 murine leukemia virus. J. Virol. **45**: 291-298.

Khan, A. S. (1984). Nucleotide sequence analysis establishes the role of endogenous murine leukemia virus DNA segments in formation of recombinant mink cell focus-forming murine leukemia viruses. J. Virol. **50**: 864-871.

Khan, A. S., and Martin, M. A. (1983). Endogenous murine leukemia proviral long terminal repeats contain a unique 190-base-pair insert. Proc. Natl. Acad. Sci. USA **80**: 2699-2703.

Khan, A. S., Rowe, W. P., and Martin, M. A. (1982). Cloning of endogenous murine leukemia virus-related sequences from chromosomal DNA of BALB/c and AKR/J mice: identification of an env proggenitor of AKR-247 mink cell focus-forming proviral DNA. J. Virol. **44**: 625-636.

Koch, W., Hunsmann, G., and Friedrich, R. (1983). Nucleotide sequence of the envelope gene of Friend murine leukemia virus. J. Virol. **45**: 1-9.

Kozak, C. A., and Rowe, W. P. (1980). Genetic mapping of xenotropic murine leukemia virus-inducing loci in five mouse strains. J. Exp. Med. **152**: 1419-1423.

Kozak, C. A., and Rowe, W. P. (1982). Genetic mapping of ecotropic murine leukemia virus-inducing loci in six inbred strains. J. Exp. Med. **155**: 524-534.

Kripke, M. (1974). Antigenicity of murine skin tumors induced by ultraviolet light. J. Natl. Cancer Inst. **53**: 1333-1336.

Kripke, M. L., Thorn, R. M., Lill, P. H., Civin, C. I., Pazmino, N. H., and Fisher, M. S. (1979). Further characterization of immunological unresponsiveness induced in mice by ultraviolet radiation. Transplantation **28**: 212-217.

Kühn, L. C., McClelland, A., and Ruddle, F. H. (1984). Gene transfer, expression, and molecular cloning of the human transferrin receptor gene. Cell **37**: 95-103.

Laimins, L. A., Khoury, G., Gorman, C., Howard, B., and Gruss, P. (1982). Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA **79**: 6453-6457.

Leff, T., Elkaim, R., Goding, C. R., Jalinot, P., Sassone-Corsi, P., Perricaudet, M., Kédinger, C., and Chambon, P. (1984). Individual products of the adenovirus 12S and 13S E1a mRNAs stimulate viral EIIa and EIII expression at the transcriptional level. Proc. Natl. Acad. Sci. USA **81**: 4381-4385.

Leffell, M. S., and Coggin, J. H., Jr. (1977). Common transplantation antigens on methylcholanthrene-induced murine sarcomas detected by three

assays of tumor rejection. Cancer Res. **37**: 4112-4119.

LeGrue, S. J., Kahan, B. D., and Pellis, N. R. (1980). Extraction of a murine tumor-specific transplantation antigen with 1-butanol. I. Partial purification by isoelectric focusing. J. Natl. Cancer Inst. **55**: 191-196.

Lennox, E. S., Lowe, A. D., Cohn, J., and Evan, G. (1981). Specific antigens on methylcholanthrene-induced tumors of mice. Transplant. Proc. **13**: 1759-1761.

Lenz, J., Crowther, R., Straceski, A., and Haseltine, W. (1982). Nucleotide sequence of the Akv env gene. J. Virol. **42**: 519-529.

Lenz, J., and Haseltine, W. A. (1983). Localization of the leukemogenic determinants of SL3-3, an ecotropic, XC-positive murine leukemia virus of AKR mouse origin. J. Virol. **47**: 317-328.

Lerner, R. A., Wilson, C. B., Del Villano, B. C., McConahey, P. J., and Dixon, F. J. (1976). Endogenous oncornaviral gene expression in adult and fetal mice: qualitative, histologic, and physiologic studies of the major viral glycoprotein, gp70. J. Exp. Med. **143**: 151-166.

Levin, J. G., and Rosenak, M. J. (1976). Synthesis of murine leukemia virus proteins associated with virions assembled in actinomycin D-treated cells: evidence for persistence of viral messenger RNA. Proc. Natl. Acad. Sci. USA **73**: 1154-1158.

Levy, D. E., Lerner, R. A., and Wilson, M. C. (1982). A genetic locus regulates the expression of tissue specific mRNAs from multiple transcription units. Proc. Natl. Acad. Sci. USA **79**: 5823-5827.

Levy, J. A. (1978). Xenotropic type C viruses. Curr. Top. Microbiol. Immunol. **79**: 109-213.

Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G., and Rutter, W. J. (1970). Specific inhibition of nuclear RNA polymerase II by a-amanitin. Science **170**: 447-449.

Lowy, D. R., Rands, E., Chattopadhyay, S. K., Garon, C. F., and Hager, G. L. (1980). Molecular cloning of infectious integrated murine leukemia virus DNA from infected mouse cells. Proc. Natl. Acad. Sci. USA **77**: 614-618.

Lung, M. L., Hartley, J. W., Rowe, W. P., and Hopkins, N. H. (1983). Large RNase T1-resistant oligonucleotides encoding p15E and the U3 region of the long terminal repeat distinguish two biological classes of mink cell focus-forming type C viruses of inbred mice. J. Virol. **45**: 275-290.

Majors, J. E., and Varmus, H, E. (1981). Nucleotide sequence at host-proviral junctions for mouse mammary tumour virus. Nature (London) **289**: 253-258.

Mark, G. E., and Rapp, U. R. (1984). Envelope gene sequence of two in vitro-generated mink cell focus-forming murine leukemia viruses which

contain the entire gp70 sequence of the endogenous nonecotropic parent. J. Virol. **49**: 530-539.

Markwell, M. A. K., and Fox, F. F. (1978). Surface-specific iodination of membrane proteins of viruses and eukaryotic cells using 1,3,4,6-tetrachloro-3a,6a-diphenylglycouril. Biochemistry **17**: 4807-4817.

Marshall, R. D. (1974). The nature and metabolism of the carbohydrate-protein linkages of glycoproteins. Biochem. Soc. Symp. **40**: 17-26.

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**: 499-560.

McClintock, P. R., Ihle, J. N., and Joseph, D. R. (1977). Expression of AKR murine leukemia virus gp71-like and BALB(X) gp71-like antigens in normal mouse tissues in the absence of overt virus expression. J. Exp. Med. **146**: 422-434.

McKnight, G. S., and Palmiter, R. (1979). Transcriptional regulation of the ovalbumin and conalbumin genes by steroid hormones in chick oviduct. J. Biol. Chem. **254**: 9050-9058.

Meltzer, M. S., Leonard, E. J., Rapp, H. J., and Borsos, T. (1971). Tumor-specific antigen solubilized by hypertonic potassium chloride. J. Natl. Cancer Inst. **47**: 703-709.

Mercola, M., Wang, X.-f., Olsen, J., and Calame, K. (1983). Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. Science **221**: 663-665.

Messing, J. and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**: 269-276.

Michelson, A. M., and Orkin, S. H. (1982). Characterization of the homopolymer tailing reaction catalyzed by terminal deoxynucleotidyl transferase. Implications for the cloning of cDNA. J. Biol. Chem. **257**: 14773-14782.

Miller, J. H. (1983). Mutational specificity in bacteria. Ann. Rev. Genet. **17**: 215-238.

Modolell, J., Bender, W., and Meselson, M. (1983). Drosophila melanogaster mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3-kilobase mobile element. Proc. Natl. Acad. Sci. USA **80**: 1678-1682.

Montandon, P.-E., Montandon, F., and Fan, H. (1982). Methylation state and DNase I sensitivity of chromatin containing Moloney murine leukemia virus DNA in exogenously infected mouse cells. J. Virol. **44**: 475-486.

Morse, H. C. III, Chused, T. M., Boehm-truitt, M., Mathieson, B. J., Sharrow, S. O., and Hartley, J. W. (1979). XenCSA: cell surface antigens related to the major glycoproteins (gp70) of xenotropic murine leukemia

viruses. J. Immunol. **122**: 443-454.

Mount, S. M. (1982). A catalogue of splice junction sequences. Nucleic Acids Res. **10**: 459-472.

Natori, T., Law, L. W., and Apella, E. (1977). Biological and biochemical properties of Nonidet P40-solubilized and partially purified tumor-specific antigens of the transplantation type from plasma membranes of a methyl-cholanthrene-induced sarcoma. Cancer Res. **37**: 3406-3413.

Nevins, J. R. (1983). The pathway of eukaryotic mRNA formation. Ann. Rev. Biochem. **52**: 441-466.

Nevins, J. R., Ginsberg, H. S., Blanchard, J.-M., Wilson, M. C., and Darnell, J. E. (1979). Regulation of the primary expression of the early adenovirus transcription units. J. Virol. **32**: 727-733.

Nevins, J. R. and Wilson, M. C. (1981). Regulation of adenovirus 2 gene expression at the level of transcriptional termination and RNA processing. Nature (London) **290**: 113-118.

Newman, R., Domingo, D., Trotter, J., and Trowbridge, I. (1983). Selection and properties of a mouse L-cell transformant expressing human transferrin receptor. Nature **304**: 643-645.

Obata, Y., Ikeda, H., Stockert, E., and Boyse, E. A. (1975). Relation of Gix antigen of thymocytes to envelope glycoprotein of murine leukemia virus. J. Exp. Med. **141**: 188-197.

Obata, Y., Stockert, E., DeLeo, A. B., O'Donnell, P. V., Snyder, H. W., and Old, L. J. (1981). G(ERLD): a cell surface antigen of the mouse related to xenotropic MuLV defined by naturally occurring antibody and monoclonal antibody. Relation to Gix, G(RADA1), G(AKSL2) systems of MuLV-related antigens. J. Exp. Med. **154**: 659-675.

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250**: 4007-4021.

Old, L. J., and Stockert, E. (1977). Immunogenetics of cell surface antigens of mouse leukemia. Ann. Rev. Genet. **11**: 127-160.

Omer, C. A., Pogue-Geile, K., Guntaka, R., Staskus, K. A., and Faras, A. J. (1983). Involvement of directly repeated sequences in the generation of deletions of the avian sarcoma virus src gene. J. Virol. **47**: 380-382.

Orcutt, B. C., Dayhoff, M. O., and Barker, W. C. (1982). User's guide for the alignment score program. National Biomedical Research Foundation, Washington. NBR Report: 820501-08710.

Osborne, T. F., Arvidson, D. N., Tyau, E. S., Dunsworth-Browne, M., and Berk, A. J. (1984). Transcription control region within the protein-coding portion of adenovirus E1A genes. Mol. Cell. Biol. **4**: 1293-1305.

Payvar, F., DeFranco, D., Firestone, G. L., Edgar, B., Wrange, O., Okert,

S., Gustafsson, J.-A., and Yamamoto, K. R. (1983). Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. Cell **35**: 381-392.

Pelham, H. R. B. (1982). A regulatory upstream promoter element in the Drosophila Hsp 70 heat-shock gene. Cell **30**: 517-528.

Pellis, N. R., and Kahan, B. D. (1976). Methods to demonstrate the immunogenicity of solubilized tumor antigens: I. The immunoprophylaxis assay. Meth. Cancer Res. **13**: 291-330.

Persson, H., and Philipson, L. (1982). Regulation of Adenovirus gene expression. Curr. Top. Microbiol. Immunol. **97**: 157-203.

Pincus, T., Rowe, W. P., and Lilly, F. (1971). A major genetic locus affecting resistance to infection with murine leukemia viruses. II. Apparent identity to a major locus described for resistance to Friend murine lejkemia virus. J. Exp. Med. **133**: 1234-1241.

Pinter A., Lieman-Hurwitz, J., and Fleissner, E. (1978). The nature of the association between the murine leukemia virus envelope proteins. Virology **91**: 345-351.

Prehn, R. T., and Main, J. M. (1957). Immunity to methylcholanthrene-induced sarcomas. J. Natl. Cancer Inst. **18**: 769-778.

Quint, W., Boelens, W., van Wezenbeek, P., Cuypers, T., Maandag, E. R., Selten, G., and Berns, A. (1984). Generation of AKR mink cell focus-forming viruses: a conserved single-copy xenotropic-like provirus provides recombinant long terminal repeat sequences. J. Virol. **50**: 432-438.

Rapp, U. R., and Barbacid, M. (1983). Activation of a type C virus particle in cells from the inbred mouse strain 129/J: antigenic relationship with horizontally transmitted type C viruses of primates. Arch. Virol. **76**: 373-379.

Rave, N., Crkvenjakov, R., and Boedtker, H. (1979). Identification of procollagen mRNAs transferred to diazobenzyloxymethyl paper from formaldehyde agarose gels. Nucleic Acids Res. **6**: 3559-3567.

Razin, A. and Riggs, A. D. (1980). DNA methylation and gene function. Science **210**: 604-610.

Rein, A. (1982). A distinct receptor for the MCF-recombinant viruses in mouse cells. Virology **20**: 251-257.

Rein, A., and Schultz, A. (1984). Different recombinant murine leukemia viruses use different cell surface receptors. Virology **136**: 144-152.

Reisfeld, R. A., and Kahan, B. D. (1970). Biological and chemical characterization of human histocompatibility antigens. Fed. Proc. **29**: 2034-2040.

Repaske, R., O'Neill, R. R., Khan, A. S., and Martin, M. A. (1983).

Nucleotide sequence of the env-specific segment of NFS-TH-1 xenotropic murine leukemia virus. J. Virol. **46**: 204-211.

Rigby, P. W. J., Dreckmann, M., Rhodes, C., and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase. J. Mol. Biol. **113**: 237-251.

Risser, R., Horowitz, J. M., and NcCubrey, J. (1983). Endogenous mouse leukemia viruses. Ann. Rev. Genet. **17**: 85-121.

Roberts, L. K., Spellman, C. W., and Daynes, R. A. (1980). Modulation of immunoregulatory responses directed toward various tumor antigens within hosts possessing distinct immunologic potentials. J. Immunol. **125**: 663-672.

Roman, J. M., Hirsch, J., Readhead, C., Levy, D., DeOgny, L., and Dreyer, W. J. (1981). Heritable differences among gp70-like molecules on C3H ultraviolet light-induced fibrosarcomas. Transplant. Proc. **13**: 1782-1786.

Rommelaere, J., Faller, D. V., and Hopkins, N. (1978). Characterization and mapping of RNase T1-resistant oligonucleotides derived from the genomes of Akv and MCF murine leukemia viruses. Proc. Natl. Acad. Sci. USA **75**: 495-499.

Rosner, M. R., Tung, J.-S., Hopkins, N., and Robbins, P. W. (1980). Relationship of Gix antigen expression to the glycosylation of murine leukemia virus glycoprotein. Proc. Natl. Acad. Sci. **77**: 6420-6424.

Rowe, W. P. (1978). Leukemia virus genomes in the chromosomal DNA of the mouse. Harvey Lect. **71**: 173-192.

Rowekamp, W., and Firtel, R. A. (1980). Isolation of developmentally regulated genes from Dictyostelium. Dev. Biol. **79**: 409-418.

Rubin, G. M., Kidwell, M. G., and Bingham, P. M. (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell **29**: 987-994.

Sanger, F., and Coulson, A. R. (1978). The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**: 107-110.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463-5476.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980). Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**: 161-178.

Scholl, D. R., Kahn, S., Malavarca, R., Astrin, S., and Skalka, A. M. (1983). Nucleotide sequence of the long terminal repeat and flanking cellular sequences of avian endogenous retrovirus ev-2: variation in Rous-associated virus-0 expression cannot be explained by differences in primary sequence. J. Virol. **45**: 868-871.

Schultz, A., Rein, A., Henderson, L., and Oroszlan, S. (1983). Biological, chemical, and immunological studies of Rauscher ecotropic and mink cell focus-forming viruses from JLS-V9 cells. J. Virol. **45**: 995-1003.

Schwartzberg, P., Colicelli, J., and Goff, S. P. (1984). Construction and analysis of deletion mutations in the pol gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell **37**: 1043-1052.

Shinnick, T. M., Lerner, R. A., and Sutcliffe, J. G. (1981). Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**: 543-548.

Shinnick, T. M., Lund, E., Smithies, O., and Blattner, F. R. (1975). Hybridization of labeled RNA to DNA in agarose gels. Nucleic Acids Res. **2**: 1911-1929.

Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. McGraw Hill Book Co., Inc., New York.

Sikora, K., Koch, G., Brenner, S., and Lennox, E. (1979). Partial purification of tumour-specific transplantation antigens from methyl-cholanthrene-induced murine sarcomas by immobilized lectins. Br. J. Cancer **40**: 831-838.

Silverman, S. J. and Fink, G. R. (1984). Effects of Ty insertions on HIS4 transcription in Saccharomyces cerevisiae. Mol. Cell. Biol. **4**: 1246-1251.

Sjögren, H. O. (1965). Transplantation methods as a tool for detection of tumor specific antigens. Progr. Exp. Tumor. Res. **6**: 289-322.

Sodroski, J. C., Rosen, C. A., and Haseltine, W. S. (1984). Trans-acting transcriptional activation of the long terminal repeat of human T lymphotropic viruses in infected cells. Science **225**: 381-385.

Soe, L. H., Devi, B. G., Mullins, J. I., and Roy-Burman, P. (1983). Molecular cloning and characterization of endogenous feline leukemia virus sequences from a cat genomic library. J. Virol. **46**: 829-840.

Soreq, H., Harpold, M., Wilson, M., and Darnell, J. E., Jr. (1980). Rate of synthesis and concentration of specific mRNA sequences in cultured Chinese hamster ovary cells compared to liver cells. Biochem. Biophys. Res. Commun. **92**: 485-491.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**: 503-517.

Southern, P. J., and Berg, P. (1982). Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**: 327-341.

Spellman, C. W., and Daynes, R. A. (1978). Ultraviolet light induced murine suppressor lymphocytes dictate specificity of anti-ultraviolet tumor immune responses. Cell. Immunol. **38**: 25-34.

Spradling, A. C. (1981). The organization and amplification of two chromosomal domains containing Drosophila chorion genes. Cell **27**: 193-201.

Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10**: 4731-4751.

Steele, P. E., Rabson, A. B., Bryan, T., and Martin, M. A. (1984). Distinctive termini characterize two families of human endogenous retroviral sequences. Science **225**: 943-947.

Steiner, D. F., Quinn, P. S., Chan, S. J., Marsh, J., and Tager, H. S. (1980). Processing mechanisms in the biosynthesis of proteins. Ann. N.Y. Acad. Sci. **343**: 1-16.

Stockert, E., Old, L. J., and Boyse, E. A. (1971). The Gix system. A cell surface allo-antigen associated with murine leukemia virus; implications regarding chromosomal integration of the viral genome. J. Exp. Med. **149**: 200-215.

Stockert, E., Boyse, E. A., Obata, Y., Ikeda, H., Sarker, N. H., and Hoffman, H. A. (1975). New mutant and congenic mouse stocks expressing the murine leukemia virus-associated thymocyte surface antigen Gix. J. Exp. Med. 142: 512-517.

Stockert, E., Boyse, E. A., Sato, H., and Itakura, K. (1976). Heredity of the Gix thymocyte antigen associated with murine leukemia virus: segregation data simulating genetic linkage. Proc. Natl. Acad. Sci. USA **73**: 2077-2081.

Strand, M., Lilly, F., and August, J. T. (1974). Host control of endogenous murine leukemia virus gene expression: concentrations of viral proteins in high and low leukemia mouse strains. Proc. Natl. Acad. Sci. USA **71**: 3682-3686.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1966). Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31**: 77-84.

Sutcliffe, J. G., Shinnick, T. M., Lerner, R. A., Johnson, P., and Verma, I. M. (1980). Nucleotide sequence of a cloned murine leukemia virus DNA fragment. Cold Spring Harbor Symp. Quant. Biol. **44**: 1275-1279.

Sutcliffe, J. G., Shinnick, T. M., Verma, I. M., and Lerner, R. A. (1980). Nucleotide sequence of Moloney leukemia virus: 3' end reveals details of replication, analogy to bacterial transposons, and an unexpected gene. Proc. Natl. Acad. Sci. USA **77**: 3302-3306.

Sutcliffe, J. G., Milner, R. J., Gottesfeld, J. M., and Lerner, R. A. (1984). Identifier sequences are transcribed specifically in brain. Nature (London) **308**: 237-241.

Sutton, W. D., Gerlach, W. L., Schwartz, D., and Peacock, W. J. (1984). Molecular analysis of Ds controlling element mutations at the Adh1 locus of Maize. Science **223**: 1265-1268.

Svensson C., and Akusjärvi, G. (1984). Adenovirus 2 early region 1A stimulates expression of both viral and cellular genes. EMBO J. **3**: 789-794.

Sweet, H. O. (1983). Dilute suppressor, a new gene in the house mouse. J. Hered. **74**: 305-306.

Taylor, B. A. (1978). Recombinant inbred strains: use in gene mapping. In H. C. Morse III (ed.), Origins of inbred mice. Academic Press, Inc., New York: pp. 423-438.

Temin, H. M. (1980). Origin of retroviruses from cellular moveable genetic elements. Cell **21**: 599-600.

Thomas, P. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose.. Proc. Natl. Acad. Sci. USA **77**: 5201-5205.

Torchia, T. E., Hamilton, R. W., Cano, C. L., and Hopper, J. E. (1984). Disruption of regulatory gene GAL80 in Saccharomyces cerevisiae: effects on carbon-controlled regulation of the galactose/melibiose pathway genes. Mol. Cell. Biol. **4**: 1521-1527.

Tung, J.-S., Vitetta, E. S., Fleissner, E., and Boyse, E. A. (1975a). Biochemical evidence linking the Gix thymocyte surface antigen to the gp69/71 envelope glycoprotein of murine leukemia virus. J. Exp. Med. **141**: 198-205.

Tung, J.-S., Fleissner, E., Vitetta, E. S., and Boyse, E. A. (1975b). Expression of murine leukemia virus envelope glycoprotein gp69/71 on mouse thymocytes. Evidence for two structural variants distinguished by presence vs. absence of Gix antigen. J. Exp. Med. **142**: 518-523.

Tung, J.-S., O'Donnell, P. V., Fleissner, E., and Boyse, E. A. (1978). Relationships of gp70 of MuLV envelopes to gp70 components of mouse lymphocyte plasma membranes. J. Exp. Med. **147**: 1280-1284.

Tung, J.-S., Shen, F.-W., Viamontes, G., Palladino, M., and Fleissner, E. (1982). The same genetic locus directs differentiation-linked expression of endogenous retrovirus gp70 on thymocytes and spleen cells in the mouse. Immunogenetics **15**: 103-108.

van der Putten, H., Quint, W., Verma, I. M., and Berns, A. (1982). Moloney murine leukemia virus-induced tumors: recombinant proviruses in active chromatin regions. Nucleic Acids Res. **10**: 577-592.

Varmus, H. E. (1982). Form and function of retroviral proviruses. Science **216**: 812-820.

Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R.,

Naker, S. P., Chick, W. L., and Gilbert, W. (1978). A bacterial clone synthesizing proinsulin. Proc. Natl. Acad. Sci. USA **75**: 3727-3731.

Villemur, R., Rassart, E., DesGroseillers, L., and Jolicoeur, P. (1983). Molecular cloning of viral DNA from leukemogenic Gross passage A murine leukemia virus and nucleotide sequence of its long terminal repeat. J. Virol. **45**: 539-546.

Vogelstein, B., and Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76**: 615-619.

Weaver, R. F., and Weissmann, C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15 S b-globin mRNA precursor and mature 10 S b-globin mRNA have identical map coordinates. Nucleic Acids Res. **7**: 1175-1193.

Weber, J., Jelinek, W., and Darnell, J. E., Jr. (1977). The definition of a large viral transcription unit late in Ad2 infection of HeLa cells: mapping of nascent RNA molecules labeled in isolated nuclei. Cell **10**: 611-616.

Weiher, H., König, M., and Gruss, P. (1983). Multiple point mutations affecting the simian virus 40 enhancer. Science **219**: 626-631.

Wejman, J. C., Taylor, B. A., Jenkins., N. A., and Copeland, N. G. (1984). Endogenous xenotropic murine leukemia virus-related sequences map to chromosomal regions encoding mouse lymphocyte antigens. J. Virol. **50**: 237-247.

White, J., Kielian, M., and Helenius, A. (1983). Membrane fusion proteins of enveloped animal viruses. Quart. Rev. Biophys. **16**: 151-195.

Wickens, M. P., Buell, G. N., and Schimke, R. T. (1978). Synthesis of double-stranded DNA complementary to lysozyme, ovomucoid and ovalbumin mRNAs. J. Biol. Chem. **253**: 2483-2495.

Wieringa, B., Hofer, E., and Weissmann, C. (1984). A minimal intron length but no specific internal sequence is required for splicing the large rabbit b-globin intron. Cell **37**: 915-925.

Wilson, A. C., Carlson, S. S., and White, T. J. (1977). Biochemical evolution. Ann. Rev. Biochem. **46**: 573-639.

Wilson, M. C., and Darnell, J. E., Jr. (1981). Control of messenger RNA concentration by differential cytoplasmic half-life. Adenovirus messenger RNAs from transcription units 1A and 1B. J. Mol. Biol. **148**: 231-251.

Wirth, T., Glöggler, K., Baumruker, T., Schmidt, M., and Horak, I. (1983). Family of middle repetitive DNA sequences in the mouse genome with structural features of solitary retroviral long terminal repeats. Proc. Natl. Acad. Sci. USA **80**: 3327-3330.

Wirth, T., Schmidt, M., Baumruker, T., and Horak, I. (1984). Evidence for mobility of a new family of mouse middle repetitive DNA elements

(LTR-IS). Nucleic Acids Res. **12**: 3603-3610.

Wolfner, M., Yep, D., Messenguy, F., and Fink, G. R. (1975). Integration of amino acid biosynthesis into the cell cycle of Saccharomyces cerevisiae. J. Mol. Biol. **96**: 273-291.

Woodward, J. G., and Daynes, R. A. (1978). Cell-mediated immune response to syngeneic UV induced tumors. I. The presence of tumor associated macrophages and their possible role in the in vitro generation of cytotoxic lymphocytes. Cell. Immunol. **41**: 304-319.

Wright, S., Rosenthal, A., Flavell, R., and Grosveld, F. (1984). DNA sequences required for regulated expression of b-globin genes in murine erythroleukemia cells. Cell **38**: 265-273.

Yamagishi, H., Pellis, N. R., and Kahan, B. D. (1979). Tumor-protective and -facilitating antigens from 3 M KCl-solubilized tumor extracts. J. Surg. Res. **26**: 393-399.

Ziff, E. B. (1980). Transcription and RNA processing by the DNA tumour viruses. Nature (London) **287**: 491-499.