VL30: A MOUSE RETROVIRUS-LIKE FAMILY OF REPETITIVE DNA ELEMENTS

Andrew T. Carter

B.tech (Hons) Brunel University

Submitted for the degree of

Doctor of Philosophy

Department of Biological Sciences

University of Warwick

Coventry

March 1985

TO MY PARENTS

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. John D. Norton for his patience during the two years in which he trained me as his research technician, for persuading me to undertake this work, and for his continued supervision, together with that of Dr. Roger J. Avery, throughout the period of study and assembly of this thesis. Thanks are also due to Dr. Nigel Dimmock for accepting official responsibility for me during the writing of this manuscript; to members of the Tumour Virus Laboratory, now disbanded, also to past and present members of the Department of Biological Sciences for providing a friendly, cooperative and stimulating environment in which to work. Grateful acknowledgements to Miss Candice O'Connor for supplying Apodemus and Clethrionomys DNA, to Dr. John Norton for human lymphocyte DNA and many other materials, to Miss Yvonne Gibson and Dr. John Norton for permission to include in this thesis the results of their work on VL30 RNA (section R-6.4) and to Peter Roberts for guidance in agarose:formamide gel electrophoresis (figure R-14). Sincere thanks to Mrs. Susan Easton for typing this manuscript rapidly and accurately, and to Dr. Andrew Easton for his critical advice on the Methods Section.

Finally, biggest thanks of all to my wife, Anne-Marie, for her moral support and unceasing patience over the last few years.

DECLARATION

I declare that the work described in this thesis is my own, except where indicated in the text and in the acknowledgements already made. No part of this thesis has been used in any previous application for a degree.

A.T. Carter.

A.T. Carter

SUMMARY

Mouse and rat cells encode an abundant 30S RNA which shares many structural properties with retroviral genomic RNA. This VL30 RNA can be efficiently packaged into retrovirus particles. Mouse cells recently infected with a MuLV (VL30) pseudotype were shown to contain full-length, reverse-transcribed DNA copies of both RNA species. VL30 DNA could also be synthesized in quantity using the endogenous reverse transcriptase activity of detergent-disrupted MuLV (VL30) particles. This DNA was found to be identical to that produced *in vivo*. Several 4.6-4.9kbp molecular clones (NVL clones) of VL30 cDNA were obtained. The retrovirus-like LTRs of each clone displayed a moderate restriction enzyme site heterogeneity, but NVL unique sequence was identical in each case.

Southern blotting experiments using NVL probes showed that (a) most of the 100-200 NIH-3T3 DNA mouse VL30 elements were organized into provirus-like structures with a high degree of sequence conservation, and (b) the majority of these elements were hypermethylated and transcriptionally inactive, whereas an expressed NVL-like sub-class could account for no more than 5% of mouse VL30 genes.

NVL-related sequences in rodent DNAs other than the mouse were markedly less abundant and showed a greater sequence divergence. This was in contrast to MuLV-related sequences whose copy number and homology to a cloned MuLV probe decreased more gradually with phylogenetic distance from the mouse. Sub-genomic NVL probes showed that two rodent species had each conserved a different block of NVL-like sequence. These data indicate that each family has exhibited a different rate of sequence divergence during rodent evolution.

Finally, a MuLV (VL30)-infected rat fibroblast line was shown to have received 1-2 copies per cell of a transcriptionally active NVL-like element. This suggests the possibility that evolution of each rodent VL30 family has been influenced by retrovirus-mediated transmission across the species barrier.

ABBREVIATIONS

Ab-	Abelson
5-AC	5-azacytidine
AEV	avian erythroblastosis virus
ALV	avian leukosis virus
amp^R	ampicillin-resistant
AMV	avian myeloblastosis virus
bp, kbp	base pair, thousand base pairs
BrdU	bromodeoxyuridine
BSA	bovine serum albumen
ccc	covalently closed circular
cDNA	complementary DNA
CIAP	calf intestinal alkaline phosphatase
c-*onc*	cellular oncogene
cpm	counts per minute
CTAB	cetyltrimethylammonium bromide
DEAE-	diethylaminoethyl-
DEPC	diethyl pyrocarbonate
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNAase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
ds	double-stranded
DTT	dithiothreitol
EDTA	ethylenediaminotetraacetate
env	genic region coding for virion envelope proteins
ev-	avian endogenous retrovirus locus
FI, FII, FIII, DNA	Form I, II, III DNA (ccc, open circular and linear DNA)
Fr-	Friend-
gag	genic region coding for virion structural proteins
Ha-	Harvey-
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid
IAP	intracisternal A particle
IdU	iododeoxyuridine
IS	insertion sequence

Ki-	Kirsten-
LINE	long interspersed repetitive element
LTR	long terminal repeat
MCF virus	mink cell focus-forming virus
MMTV	mouse mammary tumour virus
Mo-	Moloney-
mRNA	messenger RNA
MuLV	murine leukaemia virus
MuSV	murine sarcoma virus
NBCS	newborn calf serum
NVL	NIH 3T3 VL30 element
<u>onc</u>	oncogene
PBS(-), (+)	primer binding site for -ve, +ve strand DNA synthesis
PDGF	platelet-derived growth factor
pol	polymerase
<u>pol</u>	genic region coding for reverse transcriptase
psi	per square inch
R	terminally redundant region of retrovirus genome
Ra-	Rasheed-
RNA	ribonucleic acid
RNAase	ribonuclease
rpm	revolutions per minute
RSV	Rous sarcoma virus
S	Svedberg unit (10^{-13}S)
SDS	sodium dodecyl sulphate
SFFV	spleen focus-forming virus
SINE	short interspersed repetitive element
ss	single-stranded
SSC	standard saline citrate
SSV	simian sarcoma virus
TCF	tissue culture fluid
TEMED	N,N,N',N'-tetramethyl-ethylenediamine
<u>tet</u>^{s}	tetracycline-sensitive
Tris	Tris (hydroxymethyl) aminomethane, hydrochloride salt buffer
tRNA	transfer RNA
U_3, U_5	3'- and 5'-unique sequence of retrovirus genome
VLP	virus-like particle

v-<u>onc</u>	viral oncogene
WoLV	woolly monkey leukaemia virus (= SSAV, simian sarcoma-associated virus)
X-MuLV	xenotropic MuLV

GENERAL INTRODUCTION

GENERAL INTRODUCTION

The work described in this thesis concerns the VL30 family of mouse genetic elements. These are structurally similar to retrovirus proviruses, but do not encode retrovirus particles. The 30S RNA which is encoded by this gene family was first detected by its ability to be selectively packaged by C-type retroviruses that had been propagated in cultured mouse cells (Duesberg & Scolnick, 1977). B-type retroviruses, however, package VL30 RNA with very poor efficiency (Howk *et al.*, 1978; Sherwin *et al.*, 1978).

Many other properties of VL30 DNA and RNA are paralleled in various retrovirus systems. In order to draw on these analogies it is first necessary to have a basic knowledge of retrovirus morphology and replication, and to understand some of the complex terminology which has arisen.

To this end, the beginning of the introduction will include (a) a short history outlining the major discoveries in retrovirology (section I-1), (b) the morphology of retroviruses and the classification system which is based on this (section I-2) and (c) the biological properties of the Retroviridae and some of the factors which determine their host range (section I-3). The latter half of the introduction is more directly relevant to the present studies, and deals with retrovirus replication (section I-4), links between retroviruses and cancer (section I-5) and finally the different classes of endogenous mouse retroviruses and retrovirus-like elements (section I-6).

I-1 Historical background

I-1.1 Filterable viruses

In the mid-nineteenth century special filters were developed at Pasteur's institute in Paris which were able to retain bacterial cells. These filters were often used to demonstrate the presence of microorganisms in an infectious fluid by virtue of the fact that a filtrate was no longer able to produce infection.

In 1892, Ivanowsky used a similar filter to test an extract from tobacco plants infected with mosaic disease. To his surprise the filtrate retained full infectivity. During the following 20-30 years it was realized that many other important diseases were caused by similar, filterable agents. The latter were thought to represent merely a smaller version of the recognised classes of bacteria, protozoa or fungi, and came to be termed first, "filterable viruses" ("virus" is the Latin for "poisonous liquid") and later just "viruses".

Beijerinck (1899) was the first to demonstrate that viruses were different from any known living entity when he showed that TMV could be precipitated by alcohol, or diffused through agar, but still retain infectivity. He concluded that the virus was not alive, but rather a "fluid infectious principle".

I-1.2 Viruses and cancer

Almost any tissue of a complex eukaryotic organism may, as a rare event, undergo abnormal growth. Some of its constituent cells then fail to respond normally to the mechanisms which govern growth and differentiation, and continue to divide in an uncontrolled fashion. If a malignant tumour results, fatal damage to the host occurs as a result of its invasive growth into surrounding organs and, commonly, its dissemination via blood or lymph systems to form metastases throughout the body.

The first evidence of a relationship between viruses and cancer was reported by Ellerman and Bang (1908), who were able to transmit chicken leukaemia by filtered extracts of leukaemic cells and blood serum. Leukaemia was not then recognized as a type of cancer, however, and these findings were not given the attention they deserved. Two years later Peyton Rous used a similar method to transmit a solid tumour (sarcoma) in closely related Plymouth Rock chickens (Rous, 1910, 1911). Even this discovery was not well received; over 20 years were to pass before the concept of a virus as a tumour agent gained any credence. This change of attitude stemmed from Shope's work on the viral origin of rabbit papillomas (Shope & Hurst, 1933).

Oncogenic viruses offer investigators of cancer the opportunity to use a simple model with which to explore the mechanisms by which a cell becomes tumourigenic. They have since been classified into two main taxonomic groups; those containing a DNA genome within the virion, and those possessing an RNA genome. Shope's papilloma virus contains DNA and is an example of the Papovaviridae; other DNA tumour viruses are members of the Adenoviridae, the Herpesviridae and the Hepadnaviridae families. All RNA tumour viruses, however, belong to a single family, the Retroviridae, by virtue of shared biochemical characteristics which will be discussed later.

The avian leukosis virus (ALV) detected by Ellerman & Bang (1908) and its relative, the Rous sarcoma virus (RSV) are both examples of retroviruses. Partly as a result of their early isolation, but also due to the relative ease with which chickens may be used for laboratory research, the avian RNA tumour viruses have become one of the best-characterized retrovirus systems. Others include those infecting mice, cats, cattle and monkeys. Since the work presented in this thesis concerns the study of a mouse retrovirus-like family of genetic elements, this historical background will continue by briefly describing

the isolation of those mouse retroviruses which have been most widely used in laboratory research.

I-1.3 Mouse retroviruses

Most studies of mouse retroviruses have relied heavily upon the use of inbred strains of mice. These were originally selected for their high or low tumour incidence and were established over many generations of brother-sister matings (for review, see Little, 1947). C3H mice were bred for their high incidence of mammary cancer (Strong, 1935) and were found to transmit this disease through the mother's milk (Bittner, 1936). Furthermore tumours occurred at high frequency only in animals containing high levels of oestrogens in the bloodstream. Bittner's "milk factor" was later shown to be a virus (MMTV, the prototypic example of a "B-type" retrovirus) by its ability to pass through bacterial filters (Bittner, 1942).

Other inbred mouse strains, e.g. AKR, C58 suffered from a high frequency of spontaneous leukaemia. Attempts to transmit leukaemia using cell-free extracts repeatedly failed until Ludwig Gross (1951) decided to use 1-day old suckling mice as recipients. By using a cell-free extract from a spontaneous AKR mouse lymphoma, he was able to induce leukaemia in over 50% of inoculated mice. It was of interest that the mouse strain chosen to receive the extract (on the basis of its low [1-2%] incidence of spontaneous leukaemia) was C3H, initially bred for its 90% incidence of mammary cancer. This was the first murine leukaemia virus (MuLV) to be reported. The Gross passage A MuLV stock used in much of the present day research was derived by serial passage of the original isolate, which was phenotypically variable, through many generations of suckling mice. The stock eventually achieved a more reproducible leukaemogenic potential (Gross, 1975) although biological cloning experiments have since demonstrated that it contains both

ecotropic and dualtropic (MCF) MuLV classes (see later for definitions) (Famulari et al., 1982).

Since Gross's pioneering work, many strains of MuLV have been discovered. The earliest ones derive from extracts of spontaneous, transplantable tumours and are the most leukaemogenic (e.g. the Moloney, Rauscher, Friend and Kirsten isolates), perhaps as a consequence of the stringent in vivo assays which were then routinely used (Gross, 1970). With the advent of sensitive tissue culture techniques (Hartley et al., 1969) many more MuLVs were isolated from mouse cells. Some of these are of low leukaemogenicity, or are even non-leukaemogenic, but have been classified as MuLVs on the basis of their similar genome structure and nucleotide sequence relatedness.

After passaging the Moloney strain of MuLV (Mo-MuLV) in rats, Jennifer Harvey obtained a virus preparation containing particles that produced solid connective tissue tumours, or sarcomas (Harvey, 1964). This was the first isolation of a murine sarcoma virus (Ha-MuSV) and was closely followed by others. Thus Moloney (1966) infected newborn BALB/c mice with Mo-MuLV to obtain Mo-MuSV, which caused rhabdomyosarcomas, and Kirsten & Mayer (1967) passaged mouse erythroblastosis virus (Ki-MuLV) in rats to obtain Ki-MuSV.

A characteristic of all MuSV isolates is that, unlike MuLV, they can also transform tissue culture fibroblasts. MuSV are replication-defective, requiring the presence of a non-defective MuLV ("helper virus") for propagation (see Aaronson & Weaver, 1971). Studies of the genome organization of Ki-MuSV and Ha-MuSV have shown that this increased oncogenicity arose as a result of the recombination between their respective progenitor MuLV sequences and genetic information from the rats in which they were passaged. Their replication defectiveness resulted from the acquisition of these cellular sequences at the expense of genes required for viral replication (Scolnick et al., 1973; Anderson

& Robbins, 1976). This has since been demonstrated for all MuSV, and the characterization of the newly acquired genetic material in each case led to the discovery of a family of cellular genes which are of fundamental importance to the understanding of the many mechanisms by which cells become cancerous. The molecular biology of these "cellular oncogenes" will be discussed more fully in section I-5.

I-1.4 Endogenous retroviruses and proviruses

Many naturally occurring cancers of vertebrate species do not seem to behave as though they were the result of infectious diseases. Darlington (1948) may have been the first to suggest that some cancers were induced by viruses which arose from cellular genetic elements, which he called "proviruses". It is now known that all retroviruses indeed replicate via a DNA intermediate (the provirus) which becomes integrated into the genome of the infected cell. Infection of a germ-line cell, moreover, may result in a provirus which can be vertically transmitted as a stably inherited genetic locus. It has been more difficult to prove that any tumour in a normal population of animals can arise as the result of the expression of a germ line-derived retrovirus provirus.

Evidence for the presence of endogenous retroviruses as components of normal cell genomes was obtained by parallel studies of the avian and murine experimental systems. Genetic studies by Muhlbock (1965) showed that female GR mice developed mammary tumours even after foster nursing on females of MMTV-free strains. This phenomenon was later suggested by Bentvelzen (1968) and Bentvelzen *et al*. (1970) to be due to the presence of a "germinal provirus" at a single genetic locus. Their thoughts were probably guided by several previously reported results. For instance Gross (1958) and Lieberman & Kaplan (1959) used X-irradiation of mice to induce lymphomas which contained infectious MuLV particles (RadLV).

Lwoff (1960) and Latarjet & Duplan (1962) suggested that this treatment, in addition to those with other carginogens, might be activating previously dormant viral genomes in a similar fashion to the induction of temperate bacteriophages in lysogenic bacteria.

Further evidence for the existence of endogenous retroviruses was obtained from the chicken system, when leukosis-free embryos were found to express the group-specific (gs) antigen of avian retroviruses (Dougherty et al., 1967). As with the MMTV expressed by GR mice, this expression was later shown to be determined by a single Mendelianly inherited, genetic locus (Payne & Chubb, 1968). Mouse retrovirologists quickly responded; Aaronson et al., (1969) found that long-term in vitro culture of several mouse cell lines eventually resulted in the spontaneous shedding of MuLV particles.

These findings led Heubner & Todaro (1969) to propose their "oncogene hypothesis". This predicted that spontaneous tumours were a consequence of events occurring within the cellular genomic DNA, as it was suggested that "virogenes" (retrovirus proviruses) each containing an "oncogene" were present in most vertebrate cells. These could be vertically transmitted from parent to offspring via the germ-line, such that every somatic cell of the latter would contain identical viral genetic information. The action of carcinogens was proposed to involve the induction of expression of these otherwise cryptic genes (Huebner & Todaro, 1969; Todaro & Huebner, 1972).

This hypothesis was not concerned with events which occurred during horizontal transmission of a retrovirus. Temin (1961, 1962) had observed that cells infected by RSV maintained a remarkably stable transformed phenotype, and suggested that the retrovirus provirus was passed, as an intracellular form, from parent to daughter cells at mitosis. He implied that the provirus was a DNA molecule which integrated into the cellular DNA following infection of the cell by a

retrovirus, and provided indirect evidence for this by demonstrating that actinomycin D blocked retrovirus replication (Temin, 1963, 1964).

The first direct evidence that retroviruses could integrate a DNA copy of their genome into that of their host came from the work of Baluda (1972) who used radiolabelled RSV DNA to detect additional viral sequences in the DNA of infected rodent cells. Further strong support for the provirus hypothesis emerged from early transfection studies. It was found that the DNA of RSV-infected rat or hamster cells was able to induce focus formation when transfected onto chick fibroblasts, using DEAE-dextran (Hill & Hillova, 1972a, 1972b). These foci produced RSV particles of the same genotype as those used to infect the rodent cells.

Biochemical proof for the possibility of proviral DNA synthesis was finally obtained with the discovery of RNA-directed DNA polymerase activity in retrovirus particles (Temin & Mizutani, 1970; Temin & Baltimore, 1972).

In the ensuing years viral DNA has been purified from infected cells, has been physically characterized by use of restriction endonucleases, and has been studied at the nucleotide level following its amplification by molecular cloning in bacteria (Varmus & Swanstrom, 1982). Using the same technology, endogenous retroviruses have been shown to closely resemble these exogenously acquired proviruses (Coffin, 1982) and have been detected in the genomes of all vertebrate species, including man (Martin et al., 1981). Infectious retroviruses are known to cause several commercially important diseases, as well as being strongly implicated in human T-cell leukaemia and acquired immunodeficiency syndrome (AIDS) (Weiss, 1984). The role of endogenous retroviruses in tumourigenesis may be restricted to highly viraemic inbred mice such as the AKR (MuLV) and GR (MMTV) strains. There is a possibility, however, that some may be a benefit to their host, for instance by preventing or modulating the infection by a more pathogenic

retrovirus (Rovigatti & Astrin, 1983). The ubiquity of these viruses, nevertheless, is not evidence that they are essential, eg. for ontogeny, since Astrin et al (1979) have described a flock of healthy chickens, members of which lack ALV-related sequences.

I-2 Morphological and biochemical features of retrovirus particles

I-2.1 Genomic RNA

Retrovirus genomes are comparatively small (8-10kb) molecules of RNA which contain a number of unique features. Perhaps the most unusual is that each virion contains a 60-70S dimer composed of two identical, single-stranded subunits. Dimer structure has been examined by electron microscopic (EM) analysis of partially denatured 60-70S complexes. By tagging the 3' poly(A) tails of each subunit with poly(dT)-containing circles of SV40 DNA, Bender & Davidson (1976) demonstrated that only the 5' ends were involved in the dimer linkage. The point of linkage is visualized under the EM as a characteristic Y-shaped structure in the centre of the complex, and has been calculated to be approximately 300 nucleotides from the 5' end of each subunit (for review, see Chien et al., 1980). All models of the exact nature of the linkage arrangement are highly speculative, and usually assume a degree of inverted repeat-mediated base-pairing, with or without the involvement of the tRNA primer molecule (see later), which may be capable of binding to both RNA strands (Haseltine et al., 1977a).

60-70S RNA has not been reported in the cytoplasm of infected cells, and there is some evidence which suggests that dimer formation occurs shortly after virion shedding, presumably as a result of interaction with an unidentified virion component (Canaani et al., 1973). The significance of a physically (if not genetically) diploid genome in viral replication will be discussed later (section I-4).

Gentle denaturation of 60-70S RNA irreversibly reduces its sedimentation coefficient to 34-38S, commensurate with an RNA molecule of 2.5-3 x 10^6 Da, or 7.5-9.3kb (Duesberg, 1968). These single-stranded RNA molecules resemble eukaryotic mRNA in structure as they are positive

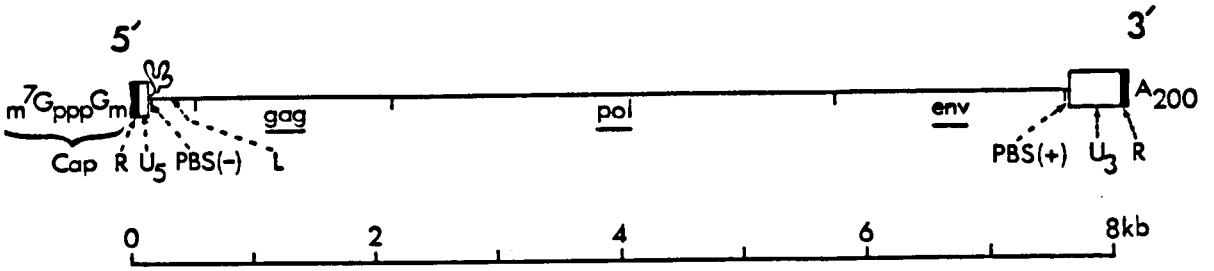


Figure I-1 Structure of a nondefective MuLV genome. The boxes indicate the positions of the terminal R, U_5 and U_3 domains. PBS(-), PBS(+), sites for priming of minus- and plus-sense DNA synthesis, respectively, during reverse transcription. L, non-coding region between PBS(-) and the initiation codon for gag. These domains, and the coding regions gag, pol and env are defined in the text. [Drawn to scale using the complete sequence of Mo-MuLV (Hughes, 1982)].

sense, possess approximately 200 adenylic acid residues at the 3' terminus, and are modified at the 5' end by the addition of a $m^7G^{5'}ppp$ group to the 5' terminal (virus-coded) G residue (which is methylated at the ribose 2' position) via a 5'-5' triphosphate linkage (Coffin, 1982) (see figure I-1). This cap structure is believed to constitute a signal for the initiation of translation, by acting as an agent for ribosome binding. RSV RNA has another eukaryotic mRNA feature as it possesses 10-12 N^6-methyladenosine residues which are clustered towards the 3' end of the genome (Beemon & Keith, 1976). Unlike the genomic RNA of the families Picornaviridae and Togaviridae, retrovirus RNA is non-infectious, requiring the presence of reverse transcriptase which is carried within the virion.

The primer for retrovirus DNA synthesis (see later) is a tRNA molecule that is tightly bound to each RNA subunit about 100-200 nucleotides from the 5' end. This association is mediated by base-pairing of the 3' terminal 16-19 nucleotides of the tRNA molecule to a complementary sequence of the virus RNA [termed the minus-sense primer binding site, or PBS(-)] (Cordell et al., 1976; Peters et al., 1977). It is interesting to observe that different classes of retrovirus use different tRNA molecules. Most avian retroviruses use $tRNA^{Trp}$, MuLV/MuSV use $tRNA^{Pro}$, MMTV uses $tRNA^{Lys}$ (Coffin, 1982), intracisternal A-particle genes (IAP genes) use $tRNA^{Phe}$ (Ono & Ohishi, 1983; Burt et al., 1984) and those VL30 elements known to be transcriptionally active use $tRNA^{Gly}$ (Norton et al., 1984b).

Virion-packaged tRNAs are a non-random subset of the population found in the cytoplasm of the host cell. This is probably the result of preferential selection for the appropriate tRNA primers, a mechanism which apparently involves their binding by reverse transcriptase (Sawyer & Hanafusa, 1979; Peters & Hu, 1980).

Location of the tRNA primer at the 5' end of the virus genome led

to the formation of models for reverse transcription which predicted the need for a short repeated sequence at each end of this RNA. One of the many experimental proofs of this hypothesis was the demonstration that strong-stop DNA (see later) of ALV and MuLV could hybridize to both ends of the respective genomic RNA (Coffin & Haseltine, 1977; Coffin et al., 1978). The precise length of this redundant (R) sequence differs from virus to virus; in ALV it is 17-21 nucleotides, in MuLV it is 60-70 nucleotides long (Coffin, 1982).

Other features of the RNA genome are the U_5 and U_3 regions, representing unique 5' and 3' sequences, respectively, and the L region. As detailed in the virus replication section, the U_5 region has no known functional role, apart perhaps from physically separating the R sequence and the PBS(-). The U_3 region (which in RSV was originally referred to as the c region) lies immediately 5' of the 3' R sequence and contains sequences for control of virus transcription (see later). The L region is an untranslated stretch of RNA extending from the 3' end of the PBS(-) to the initiation codon for gag. It varies in length from 60-300 nucleotides, according to the virus. This region contains the signal for packaging of genomic RNA into virions (Shank & Linial, 1980; Mann et al., 1983), the 5' donor splicing site for the production of subgenomic mRNAs, and may also participate in the dimerization of virion RNA subunits (Coffin, 1982).

I-2.2 Virus-encoded proteins

I-2.2.1 Nomenclature

This has been standardized for all retroviruses following the proposals of August et al. (1974) with subsequent additions (Hughes, 1982). Each protein is designated (1) by its molecular weight ($X10^{-3}$Da), (2) by a descriptive prefix, thus:-

p = protein

gp = glycoprotein

pp = phosphoprotein

Pr = precursor protein

P = polyprotein (eg the product of a fusion of a virus structural gene with that of a transduced v-onc gene),

(3) by a superscript used to indicate the coding region from which the protein was derived. Thus examples of MuLV proteins are: $p30^{gag}$; $gp70^{env}$; $pp12^{gag}$; $Pr65^{gag}$; $Pr180^{gag-pol}$; $P120^{gag-abl}$.

I-2.2.2 The genetic map

All non-defective retrovirus genomes encode at least three primary protein products. These are designated (after Baltimore, 1975); gag, which codes for four internally located, structural virion components (providing the group-specific viral antigens); pol, which codes for the RNA-dependent DNA polymerase (reverse transcriptase); and env, which encodes the envelope components. The latter include the type-specific antigenic determinants, and are important in the determination of host range specificities.

The order of these genes, which is apparently invariant for all retroviruses, was first determined for RSV by deletion mapping of T_1-resistant oligoribonucleotides (reviewed by Beeman, 1978) and is 5'-gag-pol-env-3'. RSV is unusual since its highly oncogenic variants contain the v-src oncogene to the 3' of env, without a concomitant loss of replication competence. Other acutely oncogenic retroviruses which have acquired an oncogene have done so at the expense of virus information, thus losing the ability to replicate (see section I-5).

The mechanisms by which retroviruses synthesize and process their viral proteins appear to be a mixture of those employed by DNA viruses and other RNA viruses. For example retroviruses are similar to DNA viruses in using a splicing mechanism to produce different mRNAs.

However, as with picornaviruses and togaviruses, they synthesize the structural proteins via polyprotein precursors. This will be discussed more fully in section I-4.

I-2.3 Biochemical content of retrovirus particles

The chemical composition of all retroviruses is about 60-70% protein, 30-40% lipid, 2-3% carbohydrate and 1% RNA (for review, see Duesberg, 1970). Very small amounts of DNA have been detected in some preparations of virus, but these are probably of cellular origin (Levinson *et al.*, 1972).

Most of the virion lipid derives from the cell-surface membrane, and is located in the unit-membrane virus envelope. Retroviruses are budded from particular areas of the cell surface and this may affect the lipid composition. For instance RSV particles bud from areas of the plasma membrane rich in sphingomyelin but poor in phosphatidylcholine (Quigley *et al.*, 1971).

The carbohydrate content of the virion also co-purifies with viral envelopes and comprises the sugar residues of the virus glycoproteins. Protease digestion of intact virus particles will remove these surface glycoproteins with a subsequent loss of virus infectivity (Cardiff *et al.*, 1974). All retroviruses contain at least one major glycoprotein, but avian type C viruses and murine type B viruses possess, in addition, a smaller glycoprotein. This is more hydrophobic, and is therefore believed to act as a trans-membrane anchor to stabilize the lipid bi-layer of the envelope (Dickson *et al.*, 1982).

I-2.4 Morphological classification

Retroviruses are non-lytic and acquire their outer membrane when the virion core is budded out through the plasma membrane of the host cell. The mature particle is 60-120nm in diameter, and is roughly

spherical. Glycoproteins (mentioned above) are embedded in the virus envelope and form projecting peplomers, or spikes, the prominence of which may be used as a diagnostic feature. Many of the earliest EM studies using thin-sectioning and negative staining techniques were performed by Bernhard and his colleagues (Bernhard et al., 1958). These workers classified retroviruses into four categories, the A-, B-, C- and D-type particles (for review, see Bernhard, 1958).

I-2.4.1 A-type particles

These are retrovirus-like particles found only within the cell. They possess no infectivity, even when purified in large quantities from disrupted cells. Particles range from 60-90nm in diameter, and possess a toroidal core (nucleoid) surrounded by a double shell. Two forms have been identified.

(a) Intracytoplasmic

These are present in mouse tumour cells producing MMTV, and in EM studies are often found associated with cytoplasmic membranes. They are now believed to represent the intracellular stage of the preformed core structure of MMTV. These bud from the cell to give "enveloped A-type particles", which then mature to generate infectious B-type particles (Bernhard, 1958). Maturation is believed to involve the proteolytic cleavage of precursor gag polypeptides into their final forms (Tanaka, 1977; see Dickson et al., 1982). Morphologically similar particles have been detected within cultured Drosophila melanogaster cells (Heine et al., 1980).

(b) Intracisternal (IAPs)

These are abundant retrovirus-like structures present in early mouse embryos and in a variety of mouse tumour cells. They have also been observed in hamsters (Sobis & Vandeputte, 1978) and guinea pigs (Evans et al., 1978). The particles bud from the endoplasmic reticulum

and remain within the cisternae. Although they contain reverse transcriptase (Wilson & Kuff, 1972) and several discrete poly(A)-containing RNA molecules ranging from 4 to 7 kilobases long (Kuff et al., 1981) they are not infectious when injected into mice (Hall et al., 1968; Kuff et al., 1968). IAPs appear at the two-cell stage in mouse embryos, and disappear as differentiation proceeds. This has suggested a linkage of IAPs with specific stages of embryogenesis (Biczysko et al., 1973).

I-2.4.2 B-type particles

The prototype member of this group is MMTV. As indicated above, there are several diagnostic features pertaining to the development of MMTV particles. These include (a) well-developed intracytoplasmic particles with a toroidal nucleoid. These acquire characteristically prominent (5-10nm) glycoprotein spikes upon budding. (b) Maturation of "extracellular A-type particles" to infectious B-types. These have a more electron dense core that is eccentrically located within a 130nm diameter envelope (Calafat & Hageman, 1969).

MMTV particles are expressed in response to ovarian hormones. Cultured mouse mammary tumour cells remain responsive to glucocorticoids such as dexamethasone (Ringold, 1979) probably through an interaction of the glucocorticoid-receptor complex with a sequence located within the LTR of the integrated MMTV provirus (Huang et al., 1981; Govindan et al., 1982) (see section I-4.3.1).

B-type particles containing reverse transcriptase activity have been identified in human milk samples (Schlom et al., 1971) but their involvement in human breast cancer remains controversial (Calafat & Hageman, 1973).

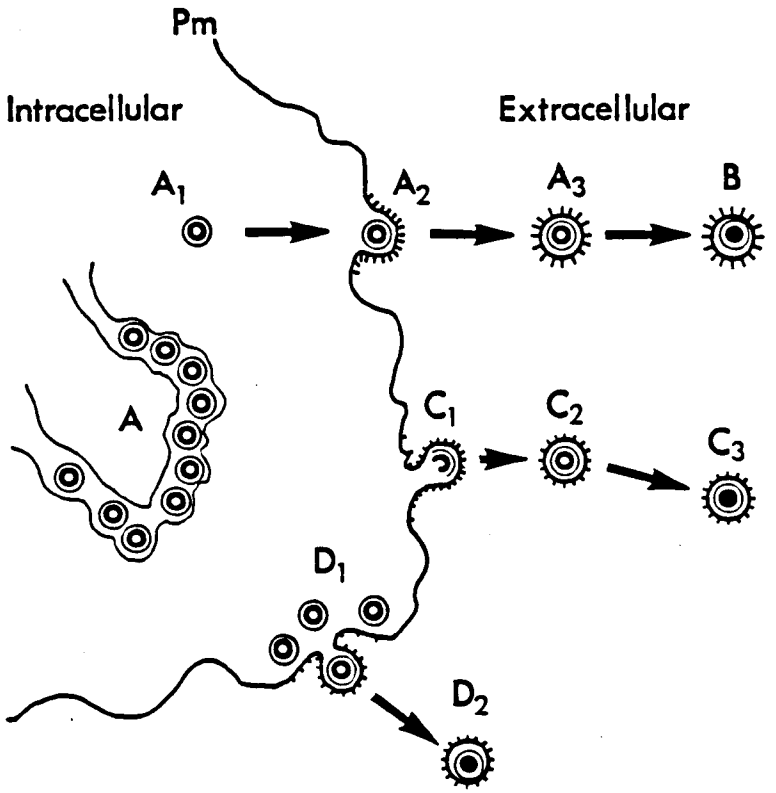


Figure I-2 Morphogenesis of retroviruses. Schematic illustrating the morphological differences between A-, B-, C- and D-type particles (not drawn to scale). A, intracisternal, and A_1, intracytoplasmic A-type particles, both 60-90nm diameter; A_2, budding A-type; A_3, 75nm diameter enveloped A-particle (immature B-type), long spikes; B, mature B-type, electron-dense core eccentrically placed within 125-130nm diameter envelope; C_1, crescent-shaped immature core of budding C-type; C_2, immature extracellular C-type with electron-lucent core; C_3, 80-110nm diameter mature C-type, centrally located core, spikes not always discernible, D_1, 60-95nm diameter intracytoplasmic D-types ("immature D-types") grouping near plasmamembrane (Pm); D_2, mature D-type, eccentrically placed core in 100-120nm diameter envelope, shorter spikes than B-type.

I-2.4.3 C-type particles

The majority of retroviruses isolated from vertebrate species are of this class. Unlike other types, no spherical intracytoplasmic particle can be distinguished until budding has started at the cell surface. At this stage the virion core is visualized as an electron-dense crescent beneath a bulge in the plasma membrane, which may bear glycoprotein spikes. These are not as prominent as those of B-types. As the core is extruded, the cell membrane pinches off around a now spherical core to give the 80-110 diameter extracellular particle. The viral core is extremely complex (Nermut et al., 1972). It has an icosahedral symmetry and consists of a core membrane about 3nm thick which is covered by a layer of hexagonally arranged, ring-like subunits about 6nm in diameter. This membrane surrounds the ribonucleoprotein complex which varies in appearance from a toroid to a more densely-packed structure. The virion core lies centrally within the virus envelope. The latter is mainly lipid, about 8nm thick and contains globular peplomers (spikes) about 8nm in diameter (the major envelope glycoproteins).

I-2.4.4 D-type particles

A relatively new addition to the retrovirus morphological classification scheme, the prototype of this class is the Mason-Pfizer monkey virus (MPMV), first discovered by EM analysis of a spontaneous mammary adenocarcinoma of a female rhesus monkey (Chopra & Mason, 1970). All subsequent D-type isolates have been made from primate species, the most well-characterized isolates being the squirrel monkey retrovirus (SMRV) (Heberling et al., 1977; Smith et al., 1977), the langur virus (PO-1-Lu) (Todaro et al., 1978) and the well-publicized agent causing AIDS in humans, HTLV III (Weiss, 1984).

The virion matures by the budding of MMTV-like intracytoplasmic

A-type particles. Extracellular forms are 100-200nm in diameter with a centrally located, electron-dense nucleoid. The envelope spikes are much shorter than those of MMTV.

The morphogenesis of A-, B-, C- and D-type particles is schematically represented in figure I-2.

I-3 Biological properties of the Retroviridae

I-3.1 Sub-families

The family Retroviridae is divided into three sub-families on the basis of their characteristic pathologies. The Oncovirinae include all the oncogenic members and their structurally similar relatives, and have been the subject of most study. The Lentivirinae cause slowly progressive neurological diseases, and the Spumavirinae are associated with symptomless chronic viraemias (Fenner, 1976).

I-3.1.1 Oncovirinae

These are widely distributed in normal animal populations and cause various types of leukaemia, carcinoma and sarcoma (see table I-1). Study of virion morphogenesis has enabled their classification into A-, B-, C- or D-types, as discussed above.

Historically the Oncovirinae have been sub-divided into two pathological classes - those causing solid tumours (sarcoma viruses) and those which produce latent infections that may result in the induction of neoplasia of the lymphoid or haematopoietic tissues. Well-characterized examples of these leukaemia viruses are the avian leukosis viruses (ALVs) and the murine leukaemia viruses (MuLVs).

As a general rule, leukaemia viruses are able to replicate in the appropriate line of tissue culture fibroblast cells, but are unable to induce a phenotypic transformation. However some of the acute defective leukaemia viruses such as AEV and Abelson murine leukaemia virus (Ab-MuLV) can transform fibroblasts, whereas others, such as AMV or Friend spleen focus-forming virus (Fr-SFFV) can not (Weiss, 1982).

Acutely oncogenic sarcoma viruses are usually able to transform fibroblasts in culture. Most of these isolates are defective, having transduced a cellular oncogene at the expense of virus genes needed for replication (see section I-5). Again there is an exception; most

Virus	Species of Isolation	Disease
avian erythroblastosis (AEV)	chicken	E,S
avian myeloblastosis (AMV)	chicken	L
avian sarcoma (ASV)	chicken	S
myelocytomatosis (MC29)	chicken	S,C,L
reticuloendotheliosis-related (REAV)	turkey	various
reticuloendotheliosis (REV)	birds	various
murine leukaemia (MuLV)	mouse	L,A,E,N
murine sarcoma (MuSV)	mouse	S
rat sarcoma (RaSV)	rat	S
feline leukaemia (FeLV)	cat	L
feline sarcoma (FeSV)	cat	S
bovine leukaemia (BLV)	cow	L
equine infectious anaemia (EIAV)	horse	A
gibbon ape leukaemia (GALV)	gibbon	L
simian sarcoma-associated (SSAV)	woolly monkey	L
simian sarcoma (SSV)	woolly monkey	S

Table I-1 C-type retrovirus-induced disease. E, erythroleukaemia, S, sarcoma; L, leukaemia (of various sorts); C, carcinoma; A, anaemia; N, neurological disorders. From Teich (1982).

strains of RSV can transform fibroblasts by virtue of their src oncogene, but, unusually, are non-defective as v-src is located 3' to the RSV env gene and thereby does not interfere with viral replication.

I-3.1.2 Lentivirinae

This subfamily consists of five natural isolates. Four of these are from sheep, and may be variants of a single virus. These are visna (the prototype lentivirus), maedi, zwoegerziekte and progressive pneumonia virus (PPV). The fifth isolate is from goats, and is now called goat leukoencephalitis virus (GLV) (Teich, 1982).

In their natural hosts these viruses cause non-neoplastic diseases with a long incubation period (up to several years) which is followed by a protracted symptomatic phase. The latter involves the production of lesions which are caused by a slowly progressive inflammation of the target tissues. The term "slow virus infection" was first used by Sigurdsson (1954) to describe these symptoms. Its continued usage reflects not the kinetics of virus replication, which are similar to those of other retroviruses (Harter et al., 1968) but the fact that the overall production of infectious virus is extremely limited (Haase et al., 1977). This restriction occurs at least partly at the transcriptional level (Brahic et al., 1981) and probably accounts for the length of time required before recognizable lesions occur. The virus successfully eludes the host's immune system over this extended period, but when they do appear the lesions themselves are the result of the interaction between the virus and the latter, since immunosuppressive treatment prevents this inflammatory tissue destruction (Teich et al., 1982).

Lentiviruses have several unique characteristics. Those propagated in vitro synthesize 50-100 copies of DNA per cell, followed by a secondary amplification to 200-300 copies per cell (Clements et al.,

1979). Unusually, 80-90% of this DNA persists in the nucleus in an unintegrated state (Clements et al., 1979). In addition, although this is obligatory for oncoviruses, cellular DNA synthesis and/or mitosis is not required for lentivirus replication. Finally, this subfamily is again atypical as its members produce lytic infections. If visna virus, for example, is grown in cell lines derived from sheep coroid plexus it will form plaques, an observation which has been used to develop a quantitative assay system (Sigurdsson et al., 1960).

I-3.1.3 Spumavirinae

This is the only group of retroviruses whose members have not been associated directly with a disease of their natural host. They may, however, be propagated readily in cultured fibroblast and epithelial cells, where they eventually give rise to a characteristic vacuolated or foamy syncytial degeneration. They have been isolated from monkeys (simian foamy viruses, SFV types 1-9), cattle (bovine syncytial virus, BSV), cats (feline syncytium-forming virus, FSFV), hamsters and humans (see Teich, 1982). It is unclear whether the human isolates are the result of natural infection or are zoonoses from monkeys (Nemo et al., 1978; Muller et al., 1980).

As these viruses are apparently non-pathogenic in vivo, they have not been studied very intensively and consequently little is known of their biology and biochemistry. The morphology of their intracytoplasmic and extracellular viral particles closely resembles an intermediate between a B-type and a C-type oncovirus particle (Malmquist et al., 1969). FSFV has been isolated from primary embryonic cells but it is not known whether this is indicative of the expression of an endogenous provirus (Hackett & Manning, 1971). Viral DNA is synthesized upon infection, but has not been conclusively demonstrated to integrate (Chiswell & Pringle, 1977, 1979b). Little is known of viral RNA, and of

the proteins only the reverse transcriptase has been studied in detail (Chiswell & Pringle, 1979a). This is serologically unrelated to that of several well-characterized oncoviruses (Liu et al., 1977).

Spumaviruses are a good model system for chronic viral disease as they persist for long periods despite high levels of circulating antibody, and can replicate in many tissue types (Hooks et al., 1972). Some can depress the cell-mediated immune response, and are thus indirectly deleterious to a persistently infected host (Hooks & Detrick-Hooks, 1979).

I-3.2 Assay of retroviruses

Assays are designed to exploit various characteristics of both the transforming and the non-transforming varieties. A brief description of the more common assays is useful to this discussion as it serves to underline the salient biological features of the Retroviridae.

I-3.2.1 Immunological assays

These are used for non-cytopathic leukaemia viruses, and include complement fixation tests, which are still used for the assay of ALV in chicken eggs. Complement fixation tests have now been mostly superseded by enzyme-linked immunosorbent assays (ELISAs) which are more sensitive (Clark & Dougherty, 1980). A sensitive immunocytological assay for MuLV has been developed by Nexo (1977). Here, infected centres of susceptible mouse cells in a monolayer previously inoculated with MuLV are detected by staining with peroxidase-coupled anti-MuLV serum. The number of detected centres of infection increases linearly with added virus inoculum.

I-3.2.2 Plaque assays

Cytopathic plaques in the appropriate tissue culture cells are

generated by visna virus, avian reticuloendotheliosis-associated virus (REAV) and some ALVs (see Weiss, 1982). The most widely used "plaque" assay for MuLVs is the XC assay developed by Rowe et al., (1970). Only certain types of ecotropic MuLVs may be titrated by this method which involves the formation of syncytia in rat XC cells. These are co-cultivated with lightly UV-irradiated, MuLV-infected mouse cells. Further details of the XC assay are given in Methods, section M-4.1. When working properly this assay is a sensitive one, and can be used to biologically clone MuLV.

I-3.2.3 S^+L^- Focus assay

This is useful for titrating strains of MuLV which do not create syncytia in rat XC cells. It uses a non-producer mouse cell line which has been infected with Mo-MuSV (ie sarcoma$^+$, leukaemia$^-$). During the original culture of this line, spontaneous reversion occurred to a phenotypically non-transformed morphology (Bassin et al., 1970). If a monolayer of these S^+L^- cells is inoculated with MuLV, infected cells re-transform to a rounded morphology. Groups of these detach from the monolayer to leave plaques which can be enumerated (Bassin et al., 1974).

I-3.2.4 Assays of viral particles

These methods are used to detect intracellular retroviruses after their purification from disrupted cells, or extracellular particles released into the tissue culture fluid. Particles can be visualized by EM, or their proteins detected by immunological assay (see above) but the most commonly used test is the reverse transcriptase assay (Baltimore, 1970; Temin & Mizutani, 1970). The specificity of this may be altered by the appropriate choice of added template:primer. For screening many samples, the general template:primer, poly(rA):oligo(dT)

may be used, where a positive result is measured by the ability of a sample to synthesize a poly(dT) molecule (using radiolabelled TTP) on the poly(rA) template. A more highly specific template primer, such as poly(rCm):oligo(dG) (Gerard et al., 1974) has been used to further characterize positives from a primary screen (Heine et al., 1980).

I-3.2.5 Transformation assays

(a) In vivo

The chorioallantoic membrane of embryonated chicken eggs has been used as an assay system for RSV (Keogh, 1938). The virus induces the formation of "pocks" or small tumours, the number of which is linearly related to the concentration of virus in the inoculum.

Murine erythroblastosis viruses such as the defective Friend-SFFV and Rauscher-SFFV induce foci on the surface of the spleen of an infected mouse. These macroscopically visible foci are centres of proliferating erythroid cells, and are formed quite rapidly; the SFFV component of the polycythaemic strain of Fr-MuLV induces spleen foci 9 days after injection into susceptible mice (Axelrad & Steeves, 1964). One virus particle is required to produce each spleen focus (Steeves et al., 1971).

(b) In vitro

Most sarcoma viruses induce the formation of foci in monolayers of cultured fibroblast cells, and an assay based on this phenomenon is widely used. The morphology of the foci and the time required before their appearance varies with the system used, but in each case the number of foci produced is linearly dependent on the dose of virus inoculated (Temin & Rubin, 1958; Aaronson et al., 1970). Further details of a MuSV focus assay are given in Methods, section M-4.2.

Also used to titrate RSV and MuSV is the anchorage-independence assay. Normal fibroblasts will only grow in culture if provided with a

solid substrate. However sarcoma virus-transformed cells possess the ability to form colonies in a soft agar gel. The number of colonies produced is linearly proportional to the concentration of virus in the inoculum (Wyke & Linial, 1973; Zavada & Macpherson, 1970). This method is also very useful for the biological cloning of new retrovirus mutants (Wyke, 1973).

I-3.3 Pseudotype retroviruses

Many stocks of retroviruses are a mixture of replication competent viruses and their defective mutant progeny. Most acutely transforming retroviruses are defective, and can only replicate by using virion structural proteins supplied by a related non-defective, or "helper" virus. This phenomenon is called phenotypic mixing, and the particle produced is termed a pseudotype (Rubin, 1965).

Pseudotypes have many uses in retrovirology. Non-producer cell lines are those which have been infected by a defective retrovirus, and either shed non-infectious particles (Bassin *et al.*, 1971b) or no particles at all (Aaronson & Rowe, 1970). Superinfection with a competent retrovirus results in the "rescue" of the defective virus RNA into pseudotype particles. Not only defective viruses may form pseudotypes; the host range of a MuLV may be extended by incorporation of its genome into a xenotropic MuLV particle (see later). Those of avian retroviruses have been similarly extended to allow them to infect mammalian cells (Quade, 1979).

A well-known example of phenotypic mixing is the vesicular stomatitis virus (VSV)/retrovirus system (Zavada, 1972). Here, retrovirus envelope glycoproteins may be incorporated into the envelope of the rhabdovirus following a mixed infection. These VSV pseudotypes can be used to study retrovirus host range specificities that are determined at the cell surface, and are also a sensitive way to detect a

latent retrovirus infection in cells where only env glycoproteins are expressed (Weiss, 1982).

I-3.4 Host range

Factors governing this have been studied most intensively in the avian and murine retrovirus systems. A major factor in deciding a cell's resistance or susceptibility to infection by a retrovirus is the presence or absence of cell-surface receptors which are specific for glycoproteins of the virion envelope. Other host cell restrictions may operate to prevent provirus integration, or to block the subsequent maturation of new virus particles.

Avian and murine retroviruses differ in the degree to which they are restricted by different types of host cell. ALVs are divided into 5 subgroups, A-E, according to the types of cell which can be penetrated by the virus. Viruses from each group will adsorb to the surface of an avian cell regardless of the presence of receptors; however efficient penetration will only occur if the cell bears receptors for the specific type of envelope gp85 molecule. According to the particular combination of cell receptor and gp85 molecule, restriction of penetration may be absolute, or may be 10^3- to 10^6-fold (Crittenden & Motta, 1975; Duff & Vogt, 1969; Weiss, 1982). Host cell resistance may be overcome by introducing the viral genome as a pseudotype within the appropriate viral envelope (Hanafusa, 1965) or by modifying the cell plasma membrane with polyethyleneglycol (Rohde et al., 1978) or with inactivated Sendai or Newcastle disease virus (Robinson, 1967; Weiss, 1969).

The host range of different MuLV isolates is also defined by cell surface and intracellular interactions. As shown in figure I-3, each host range subgroup is termed a "tropism". Receptor restriction operates via the envelope glycoproteins, and has been classified into (1) ecotropism, in which the MuLV replicate in mouse cells but not in

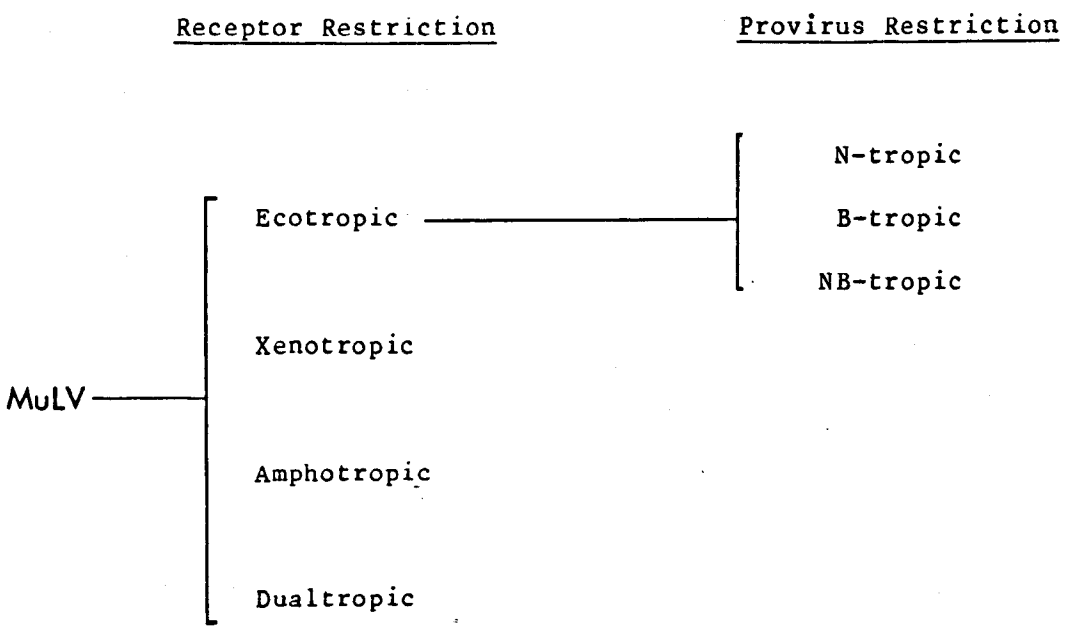


Figure I-3. Host-range "tropisms" of murine leukaemia viruses. From Weiss (1982).

those of unrelated species; (2) xenotropism, in which the MuLV grow on cells of, for example, mink, rabbit, duck and human, but do not grow on mouse cells (Levy, 1973, 1975); (3) amphotropism and (4) dualtropism in which the MuLV possess both eco- and xenotropic host ranges. Amphotropic viruses have been recovered only from certain wild mice trapped in California, and must use a novel cell-surface receptor as they are not neutralized by anti-ecotropic or anti-xenotropic antisera (Hartley & Rowe, 1976; Rasheed et al, 1976). Dualtropic MuLV, also called polytropic or mink cell focus-forming (MCF) MuLV were discovered by Hartley et al. (1977) in cells from preleukaemic thymuses. These are neutralized by antibodies to both eco- and xenotropic MuLV, and appear to be env gene recombinants between endogenous forms of the latter two viruses (Elder et al., 1977; Rommelaere et al., 1978).

Intracellular restriction of MuLV replication is controlled by a single dominant autosomal gene, Fv-1, so named as it was first studied with respect to its influence on Fr-MuLV replication (Lilly & Pincus, 1973). "N-tropic" MuLV are ecotropic MuLV which can replicate in NIH-Swiss cells (containing the Fv-1^n allele), and "B-tropic" MuLV are ecotropic MuLV which can replicate in BALB/c cells (containing the Fv-1^b allele). "NB-tropic" MuLV plate equally well on both cell types. The product of the Fv-1 gene inhibits the formation of unintegrated circular DNA, either by specifically preventing circularization or by causing the synthesis of linear DNA which is not capable of this function. The level of linear DNA synthesis itself is not grossly affected (Jolicoeur & Rassart, 1980; Yang et al., 1980).

I-4 Retrovirus replication

Upon entry of a susceptible host cell the retrovirus commences synthesis of viral DNA in the cytoplasm. Several copies of viral reverse transcriptase are available within the virion; one of these is used to produce a linear, double-stranded DNA copy of the viral RNA genome which is then transported to the cell nucleus. In most retrovirus systems much of this DNA circularizes; soon afterwards integration takes place. To complete the replication cycle, the newly integrated provirus (of which there may be one to several copies) is transcribed by the cellular RNA polymerase II into genomic RNA, which is packaged into virion cores, and into messenger RNA, which is used to synthesize the virion structural proteins. New virions bud from the cell membrane as described in section I-2.

I-4.1 Reverse transcription

Much of the supporting evidence for the following model has been provided by studies of viral DNA synthesis in detergent-activated virions (see section R-2). Under optimum conditions these "endogenous" reactions generate products which are identical to the viral DNA that can be recovered from recently infected cells (Dina & Benz, 1980; Gilboa *et al.*, 1979b). Most of the current knowledge stems from experiments using avian and murine oncoviruses, in which the strategy for reverse transcription is very similar (see Coffin, 1979; Varmus & Swanstrom, 1982 and Hughes, 1983 for reviews in which experimental evidence for the model is discussed more fully).

As discussed in section I-2.1, each retrovirus genome bears its own, specific class of tRNA molecule which is hydrogen-bonded at about 100-200 bases from the 5' end of the viral RNA. This molecule provides the 3'-OH group which is needed to prime the synthesis of DNA, a requirement shared by all known DNA polymerases (Wells *et al.*, 1972).

The 100-200 bases of RNA at the 5' end of the genome are now used as template for synthesis of a short strand of DNA which is termed (-) strand strong-stop. [(-) as the DNA is complementary to plus-sense genomic RNA, "strong-stop" as synthesis is almost immediately halted by lack of template] - see figure I-4(A,B).

It is postulated that a transfer of DNA synthesis must now occur to the 3' end of the RNA genome. This may be the same template or possibly the second of the two copies of viral RNA which are packaged into the virion. The 5' and 3' ends of retrovirus RNA contain a redundant region (R) of 16-21 bases (avian) or 50-60 bases (murine) which facilitates this template transfer. To allow base-pairing, the 5' end of the viral RNA may be displaced or may be degraded by RNAase H activity (fig. I-4, C). The latter is physically inseparable from the reverse transcriptase molecule (Moelling et al., 1971) and will degrade the RNA strand of a DNA:RNA hybrid, but will ignore either free- or double-stranded RNA.

Once this transfer has taken place (fig. I-4, D), extension of the (-) DNA strand continues along the 3' U_3 RNA region and starts to use as template a region of the genome which is required for priming the synthesis of the second, (+) DNA strand (E,F). This region consists of a polypurine tract of ribonucleotides, the exact sequence of which varies slightly between different retroviruses (Varmus, 1982). For RSV (Resnick et al., 1984) and for Mo-MuLV (Finston & Champoux, 1984) it has been recently shown that the molecule used for priming is a short stretch of this purine-rich RNA, which is excised from the viral genome by RNAase H. Initiation of (+) strand DNA synthesis occurs before completion of the growing (-) DNA strand (Gilboa et al., 1979a) and the site at which this occurs defines the unique region/U_3 region boundary (Varmus & Swanstrom, 1982) (fig. I-4, F).

The remaining 3' viral RNA is displaced or degraded to allow the (+) strand DNA to extend, using (-) DNA as template (Mitra et al.,

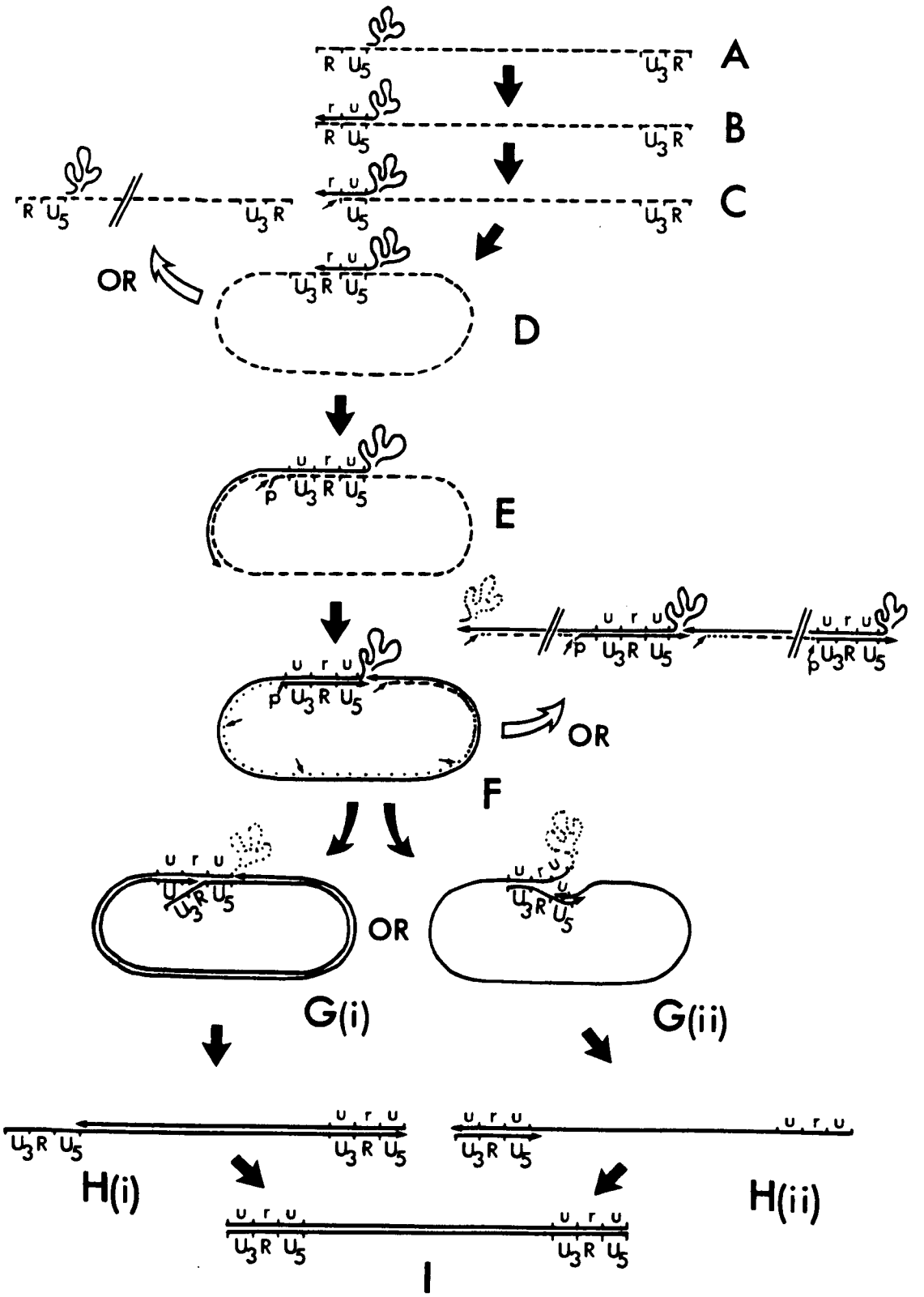

A
B
C
D
E
F
G(i)
G(ii)
H(i)
H(ii)
I
OR
OR
OR
R
U5
U3R
r
u
p

Figure I-4 Alternative mechanisms for reverse transcription of retrovirus genomic RNA. All models begin with tRNA - primed synthesis of a short minus (-) strand of "strong-stop" DNA at the 5' end of the genome (A-C). Following degradation of the 5' R region of RNA, template may be transferred to the 3' end of either the same molecule (D) or to that of another genomic RNA (arrowed). Synthesis of (-) DNA proceeds until just after the tRNA binding site is reached (E,F). Meanwhile a short region of RNA (p) is excised from the genomic RNA and used to prime synthesis of plus (+)-strand "strong-stop" DNA (E). Synthesis of this halts after using part of the tRNA as template (F). The second template transfer may also occur intra-(F) or intermolecularly (arrowed), using complementary tRNA binding site sequences. According to Dina & Benz (1980), synthesis of (+)-DNA finally displaces the (+)-strand strong-stop DNA [G(i)] to form the linear molecule H(i). The model of Gilboa et al. (1979a) proposes that it is the (-)-DNA which invades the short (+)-strand strong-stop duplex [G(ii)] to generate the molecule H(ii). Continued synthesis of either (-)-DNA [H(i)] or (+)-DNA [H(ii)] generates a flush-ended duplex DNA with the termini as shown (I). , tRNA primer; U_3, R, U_5, (+)-strand terminal regions of RNA or DNA, where appropriate; u, r, u, the respective, complementary (-)-strand DNA regions. Dashed lines, RNA; dotted lines, degraded RNA; unbroken lines, DNA; small arrows, sites of RNA degradation by RNAase H.

1979). After copying most of the tRNA primer binding site [PBS (-)], synthesis halts. By analogy with (-) strand synthesis, this product has been called (+) strand strong-stop DNA (Gilboa et al., 1979a) (fig. I-4, F).

At this stage a second transcriptional jump is necessary for the synthesis of two complete LTR sequences. It is now generally accepted that the jump takes place from an RNA template to a DNA template (Shank et al., 1978b). The elongating (-) strand DNA stops before it reaches the 5' end of the viral RNA, apparently within the PBS (-). Thus it contains PBS (-) sequences at its 3' end which are complementary to those situated at the 3' end of the (+) strand strong-stop DNA. Hydrogen bonding between these two short sequences may form a circular replicative intermediate (see figure I-4,F), or it may join two separate (linear) molecules (arrowed).

At this stage, Dina & Benz (1980) propose that the (+) strand strong-stop DNA, using (-) DNA as template, elongates around a circular intermediate, finally displacing its own 5' end [G(i)]. The circle is caused to linearize by this action, and the displaced (+) strong-stop DNA is then used as template to complete the final LTR. This is the 5' LTR (on the left in conventional drawings of viral DNA) (fig. I-4, H(i) and I).

Gilboa et al. (1979b) suggest that following the second transcriptional jump, the (-) DNA strand proceeds to invade the duplex formed by its 5' end and the (+) strand strong-stop DNA. In this model, therefore, the LTR destined to lie at the 5' end of the provirus is synthesized first. Completion of this LTR causes the circular intermediate to linearize, whereupon the (+) strand strong-stop DNA elongates to complete the viral DNA duplex (fig. I-4, G(ii), H(ii) and I).

Dina & Benz (1980) noted that a high proportion of replicative

intermediates which were isolated during the latter stages of *in vitro* synthesis reactions lacked a complete 5' LTR. Their version may thus be the correct one, but the possibility still exists that similar incomplete molecules are generated by aberrant reverse transcription along the lines suggested by Gilboa *et al*. (1979a). As described later (section R-3) it was of interest to discover that one of the VL30 cDNA clones also possessed a truncated 5' LTR.

I-4.2 Integrated proviral DNA

The final product of reverse transcription is a blunt-ended, linear DNA duplex. *In vivo* these molecules are completed within 3-4 hours post infection, whereupon they begin to migrate into the cell nucleus. One to three hours later, circular DNA forms are produced from the linear precursors (Shank & Varmus, 1978; Yang *et al*., 1980). There are always two types of circle; one contains two copies of the LTR, and the smaller contains only one copy (Shank *et al*., 1978a; Hsu *et al*., 1978, Norton *et al*., 1982). The exact method of circle formation is unknown, but it is generally assumed that the larger version occurs by ligation of the two ends of linear DNA, and that the smaller version loses an LTR by circularizing via homologous recombination between its LTRs (Varmus & Swanstrom, 1982; Hughes, 1983).

It is still not known which of the three forms of viral DNA is the immediate precursor to integration. Since the circular forms accumulate in the nucleus at about the same time at which the first integrations take place, they are favourites for this role. Panganiban & Temin (1984) constructed a retrovirus-derived vector which contained the junction of two flush-ended LTRs that had been ligated together (ie as found in the large DNA circles). Efficient integration into cell DNA was observed to have occurred at this artificial junction site.

Visna virus may also provide a model system to explore the

integration of the oncogenic retroviruses. A recent study of visna virus DNA grown in sheep cells revealed that successful replication of the virus involves only unintegrated DNA (Harris et al., 1984). This remains in the nucleus and is closely associated with cell DNA, but most significantly it does not circularize. This observation was taken as evidence to favour the circular DNA intermediates of oncoviruses as the precursors to integration, and may also account for the inability of these workers to find endogenous visna proviruses in cellular DNA (Harris et al., 1984).

For the majority of retroviruses, integration is necessary for replication. There is good evidence that viral DNA only integrates into cellular DNA which is replicating, or which has recently done so (Varmus et al., 1979). Proviruses seem to favour no particular site for integration in the host cell DNA; if there is a specific target site, then large numbers of these must be evenly distributed throughout the genomes of many birds and mammals (Shimotohno & Temin, 1980; Majors & Varmus, 1981; Norton & Avery, 1984). The baboon endogenous virus may prove to be the exception to this rule. Although this integrates into several different sites in baboon DNA (Cohen et al., 1981) in human cells it invariably integrates into the short arm of chromosome 6 (Lemons et al., 1978). Even so, within this region, the exact site of insertion varies (Hughes, 1983).

As with the transposable elements found in Drosophila (Dunsmuir et al., 1980), yeast (Farabaugh & Fink, 1980) and bacteria (Calos & Miller, 1980) the insertion of a retrovirus provirus causes the duplication of a short stretch of host cell DNA (Majors & Varmus, 1981). The size of this duplication is specific for the type of retrovirus, regardless of the cell in which it is propagated. MMTV and ASV/ALV are associated with 6bp repeats, and MuLV with 4bp repeats (Majors & Varmus, 1981; Hughes et al., 1981a; Dhar et al., 1980). The repeats which flank the

provirus are direct, suggesting that the host DNA is subject to a staggered cleavage at the point of insertion, followed by repair of each single-stranded region after integration has taken place (Shapiro, 1979).

Although no cell DNA sequence is sacrificed, proviral integration always results in the loss of 2bp from each end of the provirus, leaving the 5' end with the bases TG- and the 3' end with the bases -CA. These are also found at the ends of eukaryotic transposable elements (Temin, 1982). These features imply a rigidly determined mode of integration, which may be contrasted with that demonstrated by DNA tumour viruses such as papova- and adenoviruses. These integrate using any site of the viral DNA, and unlike retrovirus proviruses, which are always colinear with genomic RNA, these DNA viruses are often rearranged upon integration (Weinberg, 1980). Interestingly, retrovirus DNA introduced into cells by microinjection or by transfection behaves like any piece of DNA; the integration sites within each LTR are ignored, and any segment of the introduced DNA may integrate (Copeland et al., 1981). These results strongly suggest a requirement for a virus-coded protein to facilitate integration.

I-4.3 Transcription and translation

The integrated provirus is the template for transcription of both genomic and viral mRNA by the host cell RNA polymerase II (pol II) (Dinowitz, 1975). Both transcripts are initiated in the 5' LTR (the RNA cap site defines the 5' border of the R region) and are terminated in the 3' LTR (the poly(A) addition site defines the 3' border of the R region). By analogy with functional eukaryotic genes, several other features of the LTR sequence are thought to be important for transcription (Figure I-5).

The Hogness or "TATAA box" is a eukaryotic promoter for pol II,

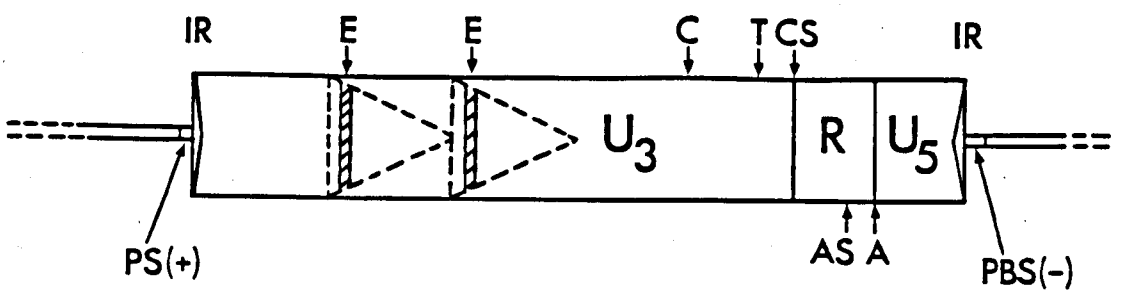


Figure I-5 Important features of a murine leukaemia virus LTR. The indicated positions of regulatory sequence elements for both reverse transcription and transcription are taken from the complete sequence of Akv MuLV (Etzerodt et al., 1984), but are similarly located in many C-type LTRs. PS(+), polypurine tract used to prime synthesis of (+)-strand DNA; IR, inverted repeat; E, consensus enhancer sequence, often found within long direct repeats, as shown; C, "CAT" box; T, "TATA" box; CS, 5' RNA cap site; AS, polyadenylation signal; A, polyadenylation site; PBS(-), binding site for the tRNA molecule used to prime synthesis of (-)-strand DNA.

analogous to the prokaryotic Pribnow box promoter (Breathnach & Chambon, 1981). Sequences closely resembling this element are found 20-30bp upstream from the retroviral cap site. Also found in the LTR, and implicated in the promotion of eukaryotic gene transcription is the "CAAT box", with the consensus sequence 5'-GG$^{C}_{T}$CCAATCT-3'. This is positioned 70-80bp upstream from the start site (Benoist et al., 1980; Corden et al., 1980). Although each LTR contains the signals for both initiation and termination of transcription, the 5' LTR is favoured for the former function, and the 3' LTR for the latter. The consensus signal for addition of poly(A), 5'-AATAAA-3' (Proudfoot & Brownlee, 1974) is located 16-22bp upstream of the poly(A) addition site and may be found within the U_3 region (RSV, MMTV) or within the R region (spleen necrosis virus, MuLV/MuSV). It is theoretically possible for the latter viruses to synthesize a short transcript which originates and terminates in the same R region, but this is not favoured (Benz et al., 1980). The LTRs also contain enhancer sequences which are able to modulate the transcriptional activity of host cell genes upstream or downstream of the provirus; these will be discussed elsewhere (section D-6). In some instances the 3' LTR of a provirus has been shown to promote the transcription of downstream host DNA sequences; mechanisms for this unusual behaviour and its implications for oncogenesis will be discussed in section I-5.

Retroviruses employ two different strategies to express their gene products. The gag and pol genes are encoded on genome-sized mRNA which is translated into a polyprotein that is subsequently cleaved into individual gene products (Jamjoom et al., 1977). The other method is more reminiscent of adeno- and papovaviruses (Tooze, 1980) as it involves RNA splicing. This is used for genes at the 3' end of the viral genome, for example the env gene of murine and avian retroviruses (Rothenberg et al., 1978; Dudley & Varmus, 1981; Weiss et al., 1977) and

the v-src gene of RSV (Weiss et al., 1981). In each case a short (100-300 nucleotide) leader RNA from the extreme 5' end of the genome is spliced to the main body of the subgenomic mRNA. Donor and acceptor splice sites have been identified by comparative studies of eukaryotic genes and bear the respective consensus sequences 5'-AGGTAAGT-3' (Seif et al., 1979) and 5'-PyNPyPyPyNCAG-3' (Sharp, 1981), where Py = pyrimidine and N = any nucleotide. Using deletion analysis of Mo-MuLV-derived expression vectors, Hwang et al. (1984) identified two short regions of the 5,185 nucleotide-long intron which were required for the efficient formation of spliced env mRNA. One of these extended at least 85 nucleotides upstream of the 3' acceptor splice site, and contained the sequence 5'-TACTAAC-3', which is also found in yeast pol II introns (Langford & Gallwitz, 1983). The other region required for splicing was located within a 560 nucleotide-long sequence at the centre of the intron (Hwang et al., 1984).

Although the mRNA for the gag and pol gene products is structurally indistinguishable from genomic RNA, the two species appear to form two distinct pools within the cell. When actinomycin D is added to inhibit viral RNA synthesis, viral proteins continue to be synthesized but the subsequent mature particles package only cellular RNA (Levin & Rosenak, 1976). The greater half-life of the mRNA which is implied by this observation may be the result of its association with polyribosomes (Varmus & Swanstrom, 1982).

Figure I-6 summarizes retrovirus replication and virion assembly, showing establishment of the provirus, mRNAs produced, protein products and the location of these within the mature virion. As indicated, the gag gene polyprotein is cleaved to yield the proteins which assemble to form the virion core (Vogt et al., 1975) - see section I-4.4. The $p15^{gag}$ of avian retroviruses has been identified as the protease which cleaves the other three gag proteins from the polyprotein; presumably

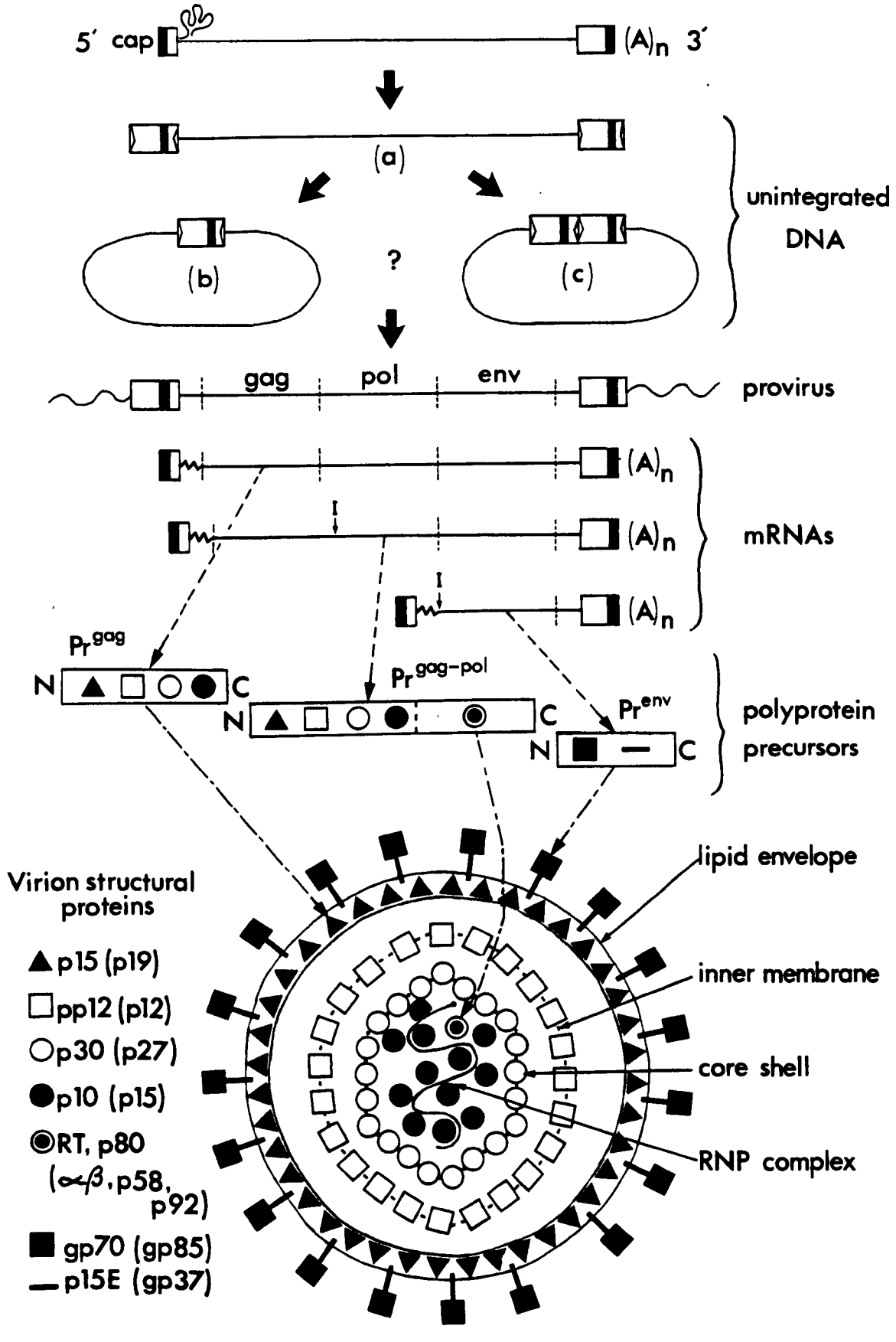

5′ cap
(A)n 3′
(a)
(b)
?
(c)
unintegrated DNA
gag
pol
env
provirus
(A)n
(A)n
(A)n
mRNAs
Prgag
N
C
Prgag-pol
N
C
Prenv
N
C
polyprotein precursors
Virion structural proteins
p15 (p19)
pp12 (p12)
p30 (p27)
p10 (p15)
RT, p80 (α/β, p58, p92)
gp70 (gp85)
p15E (gp37)
lipid envelope
inner membrane
core shell
RNP complex

Figure I-6 Retrovirus replication. The virion RNA genome (top, see fig. I-1 for explanation) is reverse-transcribed to generate a linear duplex DNA (fig. I-4) with LTRs as indicated (a). This migrates to the nucleus, where circular DNA forms (b) and (c) accumulate. It is unknown which of forms (a)-(c) integrates to generate the provirus. The mRNA molecules transcribed from the provirus and encoding the three structural genes are indicated, showing the leader sequence of R-U_5 and some adjacent unique sequence (zigzag lines). Vertical arrows (I) on the mRNA molecules indicate the point where a region is spliced out to create a subgenomic-sized mRNA. In some cases, the pol mRNA is thought to contain a small splice near the gag-pol junction (see text). The molecular weights ($x10^{-3}$Da) of the structural proteins are indicated for MuLV, and in brackets, for ASLV. Their arrangement within the precursor polyproteins and in the mature virion is indicated schematically. N, NH_2 terminus; C, COOH terminus; RNP, ribonucleoprotein. After Varmus & Swanstrom (1982) and Dickson et al. (1982).

$p15^{gag}$ itself is released from Pr 76^{gag} by a cellular protease (Vogt et al., 1979). No equivalent virus-coded protease has been identified in MuLV.

The precursor to reverse transcriptase is a 160-180kDA gag - pol polyprotein. Its existence suggests the presence of a weak terminator at the junction between the gag and pol regions, that may sometimes allow read-through of pol. Alternatively, the gag termination codon is simply removed from the mRNA by splicing (Varmus & Swanstrom, 1982). Final processing of the protein to the active form of the enzyme occurs only in extracellular particles, a mechanism that may serve to prevent free enzyme from using cellular mRNAs as templates (Witte & Baltimore, 1978). Interestingly the avian reverse transcriptase is an $\alpha\beta$ dimer, where the α chain is merely a truncated version of the β chain, but the murine enzyme exists as a 70-80kDA monomer. The properties of the latter more closely resemble the avian α subunit (Verma, 1977).

Unlike gag-pol mRNA, env mRNA is translated on membrane-bound polyribosomes (Lee et al., 1979). Nascent polypeptide chains possess a signal sequence at their amino terminus which ensures their insertion into the membrane of the rough endoplasmic reticulum. Unusually, the 5' terminus of the RNA encoding this env signal sequence overlaps with the 3' terminus of the pol gene, using different reading frames (see Hughes, 1982). Once inserted into the membrane, the signal sequence is removed and mannose-rich oligosaccharides are attached. Further glycosylation may occur as the polypeptide is cleaved into its final glycoprotein products (see Dickson et al., 1982).

Not all viral proteins are incorporated into the virion. The MuLV envelope glycoprotein, gp70, is often expressed in tissues of mice which fail to express other viral proteins (McClintock et al., 1977). Its abundance in the male genital tract and in the seminal fluid suggests that this protein might be involved in differentiation (Lerner et al.,

1976).

Gag proteins are normally located in the mature virion core. However all MuLV-infected cells contain larger, glycosylated versions of the gag polyprotein ($gP85^{gag}$ and $gP95^{gag}$) which are present at the cell surface and in the extracellular matrices (Ledbetter et al., 1978). An analogous protein is seen in avian retrovirus-infected cells (Buetti & Diggelmann, 1980) and it is possible that these non-virion proteins play a part in the maturation of virus particles (Edwards & Fan, 1981).

I-4.3.1 Control of transcription

Although the LTR contains the sequences necessary for promotion of transcription, there is considerable evidence that control of this transcription is under the influence of the host cell. It is thought that integration into a transcriptionally silent part of a chromosome will prevent expression of proviral RNA, whilst integration into an active region of chromatin encourages efficient viral expression. As observed with other eukaryotic genes, control of expression may be achieved by specific patterns of DNA methylation. These issues are discussed more fully in section I-6 (endogenous retrovirus expression).

As mentioned in section I-2.4, the expression of MMTV proviruses may sometimes be controlled by ovarian hormones. In cell lines derived from mouse mammary adenocarcinomas, induction of MMTV RNA expression by glucocorticoids occurs within 15-30 minutes, and does not require RNA or DNA synthesis (Scolnick et al., 1976b). However in all cases a basal level of viral RNA is detectable before hormone treatment, so the hormone may not be capable of inducing de novo expression (Ringold, 1979). It is unknown whether glucocorticoids can induce the expression of endogenous MMTV proviruses, which are usually hypermethylated, or whether they can act only on those proviruses acquired by milk-borne infection (Cohen, 1980). It is now known that the activated cell-

surface glucocorticoid receptor is internalized and interacts with MMTV LTR sequences (Huang et al., 1981). Within a 120bp region responsible both for receptor binding and for glucocorticoid regulation of transcription, four sites are reproducibly protected from DNase I digestion (Scheidereit et al., 1983). These share the "core" sequence 5'-TGTTCT-3', and it is thought that receptor binding at these positions alters the DNA helix structure in adjacent regions of the LTR. One of these adjacent regions bears strong homology to the enhancer "core" sequence of DNA tumour viruses, suggesting the possibility that both elements cooperate to induce tissue-specific expression (Scheidereit & Beato, 1984).

Interestingly, a similar close association between a glucocorticoid response element and an enhancer sequence has recently been identified within the LTR of an IAP element that has integrated downstream of a mouse renin gene (Burt et al., 1984). Whether or not these IAP LTR elements are responsible for the observed enhanced and androgen responsive expression of this gene is under investigation.

I-4.4 Virion assembly

Little is known of the precise mechanism by which retrovirus particles are assembled. Mutant particles have been isolated which lack envelope glycoproteins (Weiss, 1969) or genomic RNA (Linial et al., 1978) but no particles are produced if the processing of gag (core) proteins is altered in any way (Eisenman et al., 1975).

Only about half of retrovirus particle RNA is represented by the genomic RNA. Other species which are packaged during virion assembly include 28S and 18S rRNAs, as well as 5S and 7S RNAs (see Taylor, 1977; Bishop, 1978). Virions also contain about 125 molecules of 4S RNA (tRNA) of which about 10 molecules are loosely associated with the 60-70S genome complex (Canaani & Duesberg, 1972; Faras et al., 1973). At

least with avian retroviruses it appears that the spectrum of tRNA molecules included in the virion (which favours the primer tRNAs) is partly controlled by their association with the reverse transcriptase (Sawyer & Hanafusa, 1979; Peters & Hu, 1980).

Genomic RNA is probably packaged by its interaction with the gag protein p19 (ASV/ALV) or p15 (MuSV/MuLV). In the virion these proteins are associated with both the lipid envelope and viral RNA. The avian p19 can bind to double-stranded RNA, and may therefore also mediate the splicing of mRNA (Leis et al., 1978). As discussed elsewhere (section D-6) a specific sequence within the genome is involved in packaging. This has been mapped to a region lying between the 5' donor splice site and the initiation codon for the gag precursor protein (Mann et al., 1983).

A model for the assembly of retrovirus particles has been proposed by Bolognesi et al. (1978), based on several lines of indirect evidence. In this, gag and gag-pol precursor polyproteins are organized under the plasma membrane by interacting at their amino termini with membrane-located envelope glycoproteins, and at their carboxyl termini with genomic RNA. Processing of the polyproteins may be triggered by the cell surface gag glycoproteins mentioned above (section I-4.3) which in turn probably activates virion morphogenesis. Avian retroviruses appear to assemble and bud almost simultaneously with final cleavage of the polyprotein, whilst extracellular MuLV particles may still contain gag precursors (Jamjoom et al., 1975). During budding, viral particles acquire an outer envelope (section I-2), and soon afterwards mature by final processing of the reverse transcriptase (section I-4.3) and the genomic RNA (section I-2).

I-5 Retroviruses and the study of cancer

Studies of experimentally induced cancers in animals have suggested that carcinogenesis is a process involving multiple, independent steps, and that early on in this chain of events distinct alterations of genotype occur within those cells destined to form the tumour mass (Cairns, 1981).

Many of these genetic lesions are specific for a particular type of cancer. For example 90% of chronic myeloid leukaemias contain the "Philadelphia" chromosome number 22. This is characteristically shortened due to translocation of some of its material to another chromosome. Often a reciprocal exchange occurs between chromosome numbers 22 and 9 (Rowley, 1983).

Study of retroviruses has led to the discovery of a class of cellular DNA genes which may play an important role in many cancers, including those associated with the type of gross chromosome rearrangement outlined above (Bishop, 1983a). As might be expected, there is evidence that these "cellular oncogenes" are more usually concerned with processes that control normal cell growth and division (Bishop, 1983b).

I-5.1 Retroviral oncogenes

Strongly oncogenic retroviruses (section I-1) may be generated, as a rare event, by passage of the progenitor leukaemia virus in animals or, under certain conditions (Rapp & Todaro, 1980) in cultured cells. The first sarcoma virus to be isolated and studied in detail was RSV, and the properties of its viral oncogene, v-*src*, are perhaps the best understood.

I-5.1.1 v-*src*

Localization of v-*src* in the RSV genome was aided by the fact that

passage of RSV routinely generates transformation-defective (td) forms lacking all or part of this oncogene. By comparing oligonucleotide fingerprints of wild-type and td nucleic acid, v-src was mapped to the 3' end of the viral genome (Wang et al., 1975).

In RSV-transformed cells v-src is transcribed as a spliced, sub-genomic mRNA in much the same way as the retrovirus env genes. The 60kDA phosphoprotein product, $pp60^{v-src}$, is synthesized on soluble polyribosomes and is transported to the plasma membrane. Here its amino terminus attaches to the inner face of the membrane, and the carboxyl end attaches to a component of the cytoskeleton (see Bishop & Varmus, 1982 and references therein).

DNA fragments encoding only v-src were able to transform cultured fibroblasts (Copeland et al., 1980). Since cell transformation involves many phenotypic changes, it was thought that $pp60^{v-src}$ might possess a pleiotropic activity in the cell. Phosphorylation is recognized as a common method by which the activity of many proteins may be regulated (Rubin & Rosen, 1975), so kinase activity was tested for, and discovered. The fact that this seemed to be specific for tyrosine residues was considered to be highly significant for the oncogenic activity of $pp60^{v-src}$, as phosphotyrosine is very rare in normal cells, constituting about 0.01% of total protein phosphoamino acids (see Hunter & Sefton, 1980).

Prompted by the oncogene hypothesis, a src-specific probe was used to examine chicken genomic DNA for homologous sequences. These genes (c-src genes) were detected not only in chicken DNA, but in that of fish, mammals and even Drosophila (Stehelin et al., 1976; Spector et al., 1978; Shilo & Weinberg, 1981). This strong evolutionary conservation implies that the product of c-src genes is required by normal cells for an important function.

Virus	Viral oncogene(s)	Protein product	Properties
Avian	(v-)		
Rous sarcoma	src*	$pp60^{src}$	A (i)
Fujinami sarcoma	fps*	P130 or $P140^{gag-fps}$	A + B (i)
Yamaguchi 73 sarcoma	yes*	$P90^{gag-yes(env)}$	A (i)
Esh sarcoma	yes*	$P80^{gag-yes}$	G (i)
Rochester-2-sarcoma (UR2)	ros*	$P68^{gag-ros}$	G (i)
Myelocytomatosis MC29	myc	$P110^{gag-myc}$	F (iv)
Carcinoma MH2	mil*,myc	$P100^{gag-mil}$: $p48^{myc}$	B (iv); F?
Avian erythroblastosis	erb-A, erb-B*	$P75^{gag-erb-A}$; $gp74^{erb-B(env)}$	B (iv); H + D (i)
Avian myeloblastosis	myb	$p48^{myb-env}$	E (iv)
E26	myb,ets	$P135^{gag-myb-ets}$	E (iv)
Reticuloendotheliosis strain T	rel	$p55^{env-rel(env)}$	G (iv)
Mammalian			
Moloney murine sarcoma-124	mos*	$p37^{env-mos}$	B (iv)
Moloney murine sarcoma-ts 110	mos*	$P85^{gag-mos}$	G (ii)
Harvey & BALB murine sarcoma	Ha-ras,bas	$p21^{Ha-ras}$	A (ii)
Kirsten murine sarcoma	Ki-ras	$p21^{Ki-ras}$	A (ii)
Rasheed rat sarcoma	Ra-ras	$P29^{gag-ras}$	A (ii)
Murine sarcoma-3611	raf*	$P90^{gag-raf}$	G (iv)
Abelson murine leukaemia	abl*	$P120^{gag-abl}$	H (i)
FBJ murine sarcoma	fos	$p55^{fos}$	E (iv)
FBR murine sarcoma	fos,fox	$P75^{gag-fos-fox}$	E (iv)
ST feline sarcoma	fes*	$P85^{gag-fes}$	A + B (i)
GA feline sarcoma	fes*	$P95^{gag-fes}$	A + B (i)
SM feline sarcoma	fms*	$gP180^{gag-fms}$	C+D+H (iv)
GR feline sarcoma	fgr*	$p70^{gag-fgr-env}$	C? (i)
PI feline sarcoma	sis*	$p76^{gag-sis}$	G (iv)
Simian sarcoma	sis*	$p28^{env-sis}$	H? (iii)

Table I-2. Oncogenes transduced by retroviruses and their protein products.

A plasma membrane (inner face)
B cytoplasm
C cytoskeleton
D cytoplasmic membranes
E nucleus and/or perinucleus
F nuclear matrix
G unknown
H cell surface

(i) protein tyrosine kinase
(ii) protein threonine autokinase
(iii) secretion of PDGF-like protein?
(iv) unknown

Bishop & Varmus (1982); Bishop (1983b); Donoghue & Hunter (1983); Duesberg (1983); Jansen et al. (1983a,b); Kloetzer et al. (1983); Rapp et al. (1983); Reddy et al. (1983); Garrett et al. (1984); Hunter (1984); Klempnauer et al. (1984); Naharro et al. (1984); Roussel et al. (1984); Van Beveren et al. (1984); Weinmaster et al. (1984) and Wilhelmsen et al. (1984).

* The predicted amino acid sequences of these genes indicate varying degrees of homology. (Galibert et al., 1984; Besmer et al., 1983; and above references).

I-5.2 Structure of viral oncogenes

So far over twenty sequences connected with the oncogenic properties of sarcoma viruses have been discovered and their protein products identified (table I-2). Oncogenes have been reviewed extensively, several of the more recent reviews including: Bishop & Varmus (1982); Bishop (1983a and b); Duesberg (1983); Cooper & Lane (1984) and Hunter (1984).

In all cases except for RSV, the new genetic information detected in the viruses listed in table I-2 has been transduced at the expense of sequences required for replication. Commonly the oncogene occupies space normally reserved for parts of the gag and pol genic regions. In virus-infected cells these oncogenes are transcribed as a genome-length mRNA, and often the protein product is a gag-onc fusion polyprotein.

In fewer examples the v-onc is located within the env region, and is transcribed as a spliced, sub-genomic mRNA. RSV is unusual as v-src is located to the 3' of the env gene, enabling the virus to retain infectivity.

Perhaps most interesting are the tripartite recombinant retroviruses. Four retroviruses are known which contain two different oncogenes; MH2, E26, AEV and FBR-MuSV. As indicated in the table, in two viruses each oncogene is expressed independently of the other, and in the other two they form part of a triple fusion protein.

The Harvey and Kirsten MuSVs, each generated following passage of MuLV in rats, have each transduced a different member of a small family of cellular oncogenes (c-Ha-ras and c-Ki-ras, respectively). Still more striking is the fact that each v-ras gene is embedded in sequences derived from members of the rat VL30 gene family. These number about 200 copies per haploid genome but possess no known intrinsic oncogenic activity (Ellis et al., 1981).

Another interesting tripartite genome is packaged by GR-FeSV

particles. Amino acid sequence predictions of the oncogene product indicate that it is encoded by two unrelated domains. The putative oncogenic portion, v-fgr, is related to v-src and several other oncogenes (see table) but the first 128 amino acids share extensive homology to eukaryotic cytoskeletal actin (Naharro et al., 1984).

I-5.3 Transduction

It is generally accepted that the new regions of sequence detected in the genomes of sarcoma viruses are cell-derived (Bishop, 1983b). The homologous genes in chromosomal DNA are always located in a constant position on the same chromosome within a species, they share extensive sequence homology with genes of species belonging to distantly related phyla, and most are interrupted by several introns. In contrast, v-oncs contain no introns. The most popular model for retroviral transduction of c-oncs that is concordant with several features observed from sequence comparisons of v-oncs and their cellular DNA homologues, is outlined diagrammatically in Figure I-7.

Both versions of the model propose that a provirus integrates upstream of a c-onc, and that this event is followed by a local rearrangement of chromosomal DNA. It is possible that part of the c-onc is transposed directly into the provirus (part A) but this is thought unlikely (Bishop, 1983b). In the favoured version (B) (Swanstrom et al., 1983) the initial rearrangement deletes 3' proviral sequences and 5' c-onc sequences to create a novel transcriptional unit. The resulting spliced transcript of this hybrid lacks a 3' U_3-R region, but this is replaced following recombination with a non-defective viral RNA genome. A method whereby co-packaged heterodimer RNA templates can cooperate to produce a hybrid reverse transcript has been described (Coffin, 1979).

Most available evidence supports the latter model (Figure I-7, part

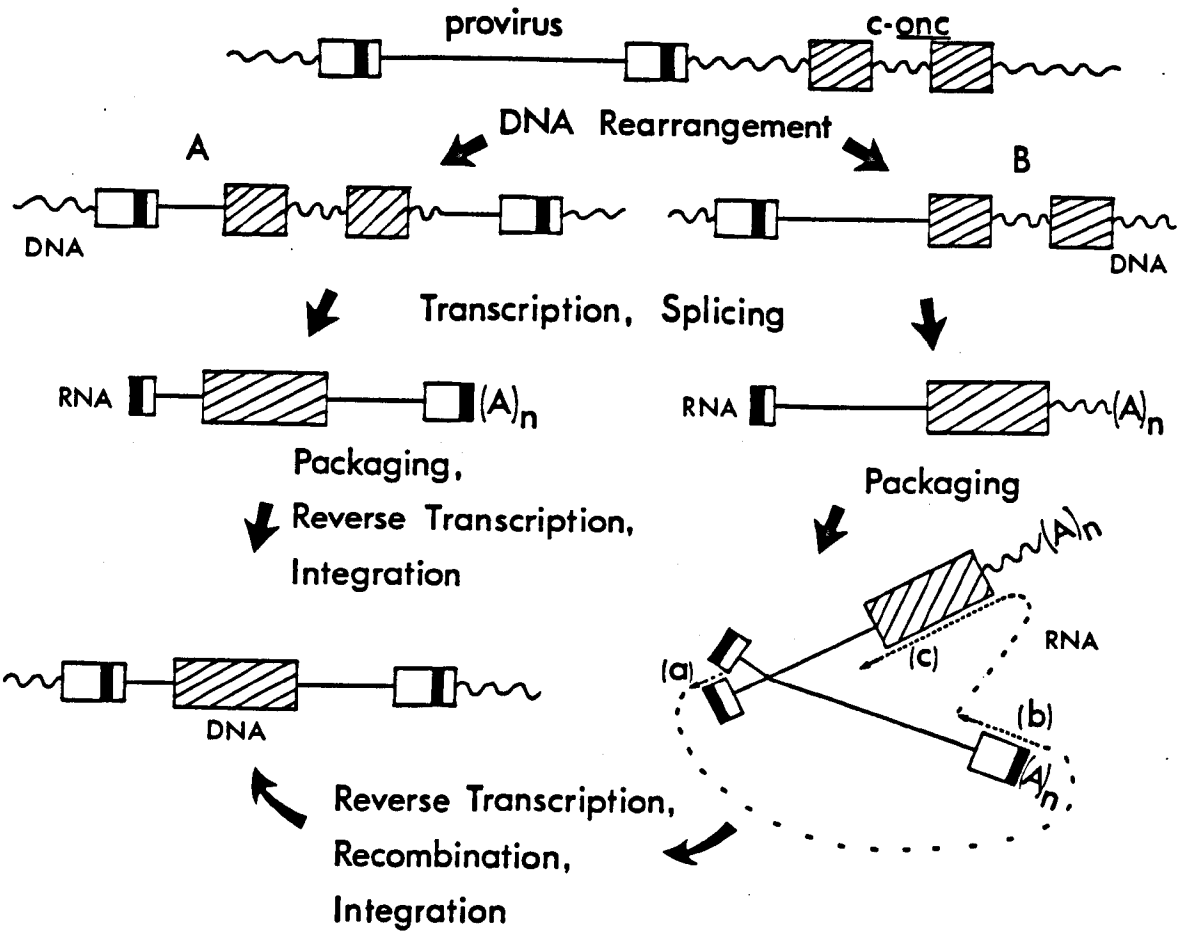


Figure I-7 Models for transduction of cellular oncogenes by retroviruses. Both models begin with an intact provirus integrated upstream from a cellular proto- oncogene. The DNA rearrangement postulated for each version could take either of two forms; A, direct transposition of the c-onc into the provirus, or B, deletion of viral 3'- and cellular flanking sequences to generate a hybrid viral-oncogene transcriptional unit. Events thereafter are as illustrated (see text for details). Part B, (a)-(c) indicate different stages in the route of reverse-transcriptase during the formation of a recombinant viral genome. Black and white boxes, retrovirus LTRs; shaded boxes, exons of the cellular proto-oncogene; straight lines, viral DNA/RNA, wavy lines, cellular DNA/RNA. After Bishop (1983b) and Swanstrom et al. (1983).

B). Both c-myb and c-src have been transduced by AMV and RSV, respectively such that the leftward recombination has occurred within an intron; introns within the transduced portion of each c-onc have been precisely spliced out, and the rightward recombination point (proposed to occur between RNA molecules) has occurred within an exon in each case (Klempnauer et al., 1982; Swanstrom et al., 1983).

Tripartite retroviruses are proposed to follow similar mechanisms of alternating rounds of recombination at the DNA and RNA levels, with at least two different co-packaging events being likely: see Norton et al (1984c) for a model to explain the formation of Ki-MuSV from components of Ki-MuLV, rat VL30 and c-Ki-ras.

I-5.4 How do oncogenes cause cancer?

A growing body of evidence suggests that cellular oncogenes, or c-oncs, are expressed in normal tissues and that their protein products control the processes of cell growth, division and differentiation (Cooper & Lane, 1984; Hunter, 1984). Transduction of all or part of its coding sequences by a retrovirus may be just one way by which a c-onc becomes abnormally expressed to initiate and/or promote tumour formation. There are four proposed mechanisms by which a c-onc could become "activated": (1) the dosage hypothesis, ie overexpression; (2) ectopic expression, ie in an inappropriate cell type; (3) unscheduled expression, ie at the wrong time during the cell cycle or during the development of a cell lineage, and (4) mutation of the normal c-onc gene product. Examples favouring each mechanism have been described; some favour more than one mechanism, as these are not mutually exclusive. The review by Hunter (1984) covers this ground in detail so only the major points will be discussed herein.

I-5.4.1 Dosage

Cloned c-onc genes have been linked to a retrovirus LTR (containing a promoter and enhancer) and transfected into NIH 3T3 cells. Both mouse c-mos and rat c-Ha-ras genes, thus treated, will induce transformed foci, (Oskarsson et al., 1980; De Feo et al., 1981) but (a) human c-mos, chicken c-src and chicken c-myc genes will not (Bishop, 1983b) and (b) these LTR-activated c-onc genes fail to transform primary embryo cells or to be oncogenic in vivo (Bishop, 1983b). This may be due to the need for two different c-oncs to be activated for tumour initiation/promotion (see later).

I-5.4.2 Ectopic expression

The c-sis gene product bears extensive homology to one of the two subunits of platelet-derived growth factor (Waterfield et al., 1983) and may indeed exhibit the same functional properties. PDGF is probably only produced by the megakaryocytes from which blood platelets are derived. Those types of cell which possess receptors for PDGF include those which participate in wound healing (Stiles, 1983) and it is from these cell types that some SSV-induced tumours arise (Theilen et al., 1971). (SSV carries the viral oncogene, v-sis). Thus an important stage of transformation is probably chronic production of a PDGF-like molecule by cells also possessing PDGF receptors. This autocrine growth stimulation could possibly be achieved by other means than the introduction of an activated, v-sis gene by a virus. For example any DNA-damaging mutagen might disrupt the regulatory mechanism of c-sis.

I-5.4.3 Unscheduled expression

C-onc RNA may be expressed constantly, or it may fluctuate according to the cell cycle and/or state of differentiation of a cell (Slamon & Cline, 1984). For example c-myc is not expressed in resting

cells, but is abundantly expressed early in the G1 phase (Kelly et al., 1983). Any unscheduled expression of c-myc, the product of which is a DNA-binding protein, is likely to upset cell growth and division. Numerous examples of tumours in which c-myc de-regulation occurs have been described. Thus in avian B-cell lymphomas, integration of ALV close to c-myc may be an early event in tumourigenesis (Hayward et al., 1981). In mouse (Shen-Ong et al., 1982) or in human B-cell neoplasms (see Leder et al., 1983) the myc gene is often partly disrupted by its translocation into one of several possible immunoglobulin loci. It is thought that removal or extensive mutation of the 5' noncoding c-myc exon in these instances is responsible for transcriptional activation (Taub et al., 1984).

I-5.4.4 Mutated structure

All viral oncogenes so far examined are different from their cellular DNA precursors. Perhaps as a reflection of the error-prone viral reverse transcription process, all v-oncs contain a scattering of novel bases not found in the appropriate c-onc exon. Those of v-src amount to 1-2% of its sequence, and in several cases are the cause of amino acid substitutions in the protein product (Lee et al., 1981; Takeya & Hanafusa, 1983). In this example these substitutions may not be the direct cause of cell transformation by $pp60^{v-src}$, since they are different in each $pp60^{v-src}$ protein encoded by different strains of RSV. It is probable that the novel carboxy terminal of all $pp60^{v-src}$ proteins is the crucial factor. This is encoded by cellular sequence normally located about 1kb further downstream from the usual c-src 3' coding sequences (Parker et al., 1984).

The most quoted example of oncogene mutation in this respect must be that of the activated ras genes known to be associated with several human cancers. These genes, detected by the 3T3 transfection assay (see

Cooper & Lane, 1984) encode two types of amino acid substitution in the $p21^{c-ras}$ product. One is a change in the normal glycine at position 12 to one of several amino acids, and the second type bears a new amino acid at position 61, instead of glutamine. Interestingly both types of v-ras gene also contain base substitutions in codon 12, and it has been shown by in vitro recombination experiments that this mutation can be responsible for oncogenesis (see Hunter, 1984).

I-5.5 Multi-step oncogenesis

As alluded to earlier, tumourigenesis is considered to require a series of pre-neoplastic and neoplastic changes. Why then are some activated oncogenes capable of inducing a stable, transformed phenotype in the NIH 3T3 transfection assay, apparently in a single step? The explanation lies in the fact that, during establishment of the NIH 3T3 line, the cells have become immortalized and may be considered to be pre-neoplastic rather than normal. Some oncogenes require that this condition be fulfilled before their abnormal expression completes the process to bring about transformation (Newbold & Overell, 1983). Recently it has been shown that in order to transform primary embryo cells, more than one type of activated oncogene is required (reviewed by Cairns & Logan, 1983). Studies using co-transfected oncogenes have revealed different complementation classes of oncogenes (see table I-3). As indicated, these oncogenes include not only those sometimes transduced by retroviruses, but also those required for DNA tumour virus replication, implying that common pathways are sometimes used by these two oncogenic agents. It is interesting to note that the Ela gene product shares some amino acid homology with that of the myc gene, and that both complement c-Ha-ras-1 when this is activated (Hunter, 1984). The p53 protein has been detected in a range of neoplastic cell types, in addition to those of normal tissues growing exponentially in culture

"Immortalizing" genes	"Transforming" genes
myc	Ha-*ras*-1
polyoma large T	Ha-*ras*-1
adenovirus-2 Ela	Ha-*ras*-1
adenovirus-2 Ela	polyoma middle T
p53	Ha-*ras*-1
p53	SV40 large T
p53	adenovirus-2 ElB

Table I-3. Complementary oncogenes. Pairs of transforming genes which, upon co-transfection, transform primary embryo fibroblasts.

References: Land *et al*. (1983); Ruley (1983); Parada *et al*. (1984); Eliyahu *et al*., (1984); Jenkins *et al*. (1984).

(Dippold et al., 1981). In primary cells p53 has a very short half-life, but in cells transformed by SV40 or adenovirus the protein binds to the large T antigen or the Elb protein, respectively, and is stabilized (Lane, 1984). The products of c-myc, polyoma large T, Ela and the p53 protein are all located in the nucleus of the cell, and are all implicated with "establishment" of cells. In other words they mimic the immortalization demonstrated by NIH 3T3 cells (Land et al., 1984; Parada et al., 1984). Conversely the products of ras genes, of polyoma middle T and of adenovirus Elb are associated with morphological alteration and anchorage independence (see Lane, 1984). These observations imply that activation of different types of oncogene in tumours corresponds with the recognized stages of progression in tumour formation. Thus it can be postulated that activation of a gene such as myc leads to the establishment of a pre-neoplastic cell population. If no further damage occurred, perhaps this abnormal state could be controlled. However activation of an oncogene of the second class (eg ras) in a cell belonging to this proliferating population would lead to cancer (Cooper & Lane, 1984).

This mechanism of multi-step oncogenesis may be sufficient to explain the induction of tumours by retroviruses which carry no oncogenes. ALV-induced chicken bursal lymphomas contain activated c-myc genes as a result of adjacent insertion of the provirus ("promoter insertion"; Hayward et al., 1981; Payne et al., 1982), but transfection of this tumour DNA into 3T3 cells has revealed the presence of yet another activated oncogene, designated Blym-1 (Cooper & Neiman, 1981). The mechanism of activation of this second oncogene is unknown, but its homologue in the human genome has been similarly detected in 6 out of 6 Burkitt's lymphomas investigated (Diamond et al., 1983), which also contain translocated c-myc sequences.

Carcinomas induced by MMTV frequently contain activated genes, designated int-1 and int-2 that are also adjacent to an integrated provirus (Nusse & Varmus, 1982; Dickson et al., 1984). Likewise, a putative cellular oncogene designated MLVI-1 may be occasionally activated by MuLV integration, since several T-cell lymphomas contain proviruses integrated in this specific region of chromosomal DNA (Tsichlis et al., 1983). It has been postulated that in each case a second cellular oncogene may also be involved in tumourigenesis (Cooper & Lane, 1984). Thus MLVI-1 may complement Tlym-1, and MMTV int-1 or int-2 may complement ts-4. Each of these genes has been detected in the appropriate tumour tissue DNA by the 3T3 transfection assay (Cooper & Lane, 1984).

Although further investigation will be required to test the verity of this model, it certainly represents a major advance in the theory of human cancer, and underlines the need to understand more of the normal processes of cell division and differentiation.

I-6 Endogenous retrovirus-like elements

These are provirus-like sequences that are transmitted from parent to offspring as a result of their incorporation into germline DNA (see Rovigatti & Astrin, 1983; Risser *et al.*, 1983). In the mouse there are four classes; MuLVs, MMTVs, IAPs and VL30 sequences. Each class shares little, if any, sequence homology with the others, but all are united by their obvious structural similarities to replication competent retroviruses (see sections I-2 and I-4). Assuming 50 copies of MuLV proviruses per mouse haploid genome, 5 copies of MMTV, 1000 IAP genes and 150 VL30 units, there are at least 8000kbp of retrovirus-like sequences within a genome of 3×10^9bp (ie about 0.3%). These elements may thus be considered as members of a large multi-gene "superfamily" of moderately repetitive DNA sequences. Analysis of other repetitive sequence elements residing in the eukaryotic genome has provided evidence suggesting that many classes either share a common origin, or that they exhibit signs of convergent evolution.

I-6.1 General organization of eukaryotic chromosomal DNA

How are repetitive elements organized within the genome? This can be assessed by denaturing chromosomal DNA and by studying the kinetics of its reassociation under standard conditions (see Britten & Kohne, 1968). Renaturation of complementary strands depends on the initial concentration of DNA (Co) and on time (t). Measurement of the $Cot_{\frac{1}{2}}$ value, at which half the DNA has renatured, enables estimation of the length of the basic repeating unit of the haploid genome. In the case of *E. coli* this value is approximately the same as the length of the genome, indicating that most sequences are uniquely represented.

Eukaryotic DNAs are more complex, containing highly repeated short sequences, which reassociate very rapidly, moderately repeated sequences which reassociate less rapidly, and finally single copy DNA, which

reassociates very slowly (Britten & Kohne, 1968).

Apart from a few exceptions, single-copy DNA comprises most of the protein-coding sequence, and includes small multi-gene families and their related pseudogenes. Highly repetitive DNA consists of clustered, tandemly repeated sequences with a relatively simple monomeric unit of 100-400bp (Sutton & McCallum, 1971). Enriched preparations of this DNA can be obtained by collecting fractions containing satellite bands located either above or below the main DNA band in a CsCl density gradient of total cell DNA. This "satellite" DNA varies tremendously in amount among different species, but is generally found at centromeric and telomeric positions, and is thus thought to be a structural component of the chromosomes, perhaps maintaining the species barrier by its involvement in chromosome pairing (Peacock et al., 1978).

Moderately repeated sequences, reiterated about 10^4 times or more, are further classified into short interspersed repeats (SINEs) typically <500bp in length, and long interspersed repeats (LINEs) which are >5kbp long (Singer, 1982). SINEs and LINEs are interspersed with unique sequences to form two characteristic patterns. Sea urchins, Xenopus and humans possess the "short period" pattern, with most DNA organized into single copy regions of 1000-2000bp separated by SINEs. The "long period" interspersion pattern, the prototype example of which is Drosophila DNA, contains unique sequence of average length 35kbp interspersed with the >5kbp long LINEs (see Jelinek & Schmid, 1982, for a review).

SINEs and LINEs form large families of elements which are similar in length and sequence, but not identical. Some families, such as Alu, Bam and Kpn I are characterized by the restriction sites which are conserved in the majority of their members. The major SINE family in mouse DNA is the B1 family, which is closely related to the human Alu family (Jelinek & Schmid, 1982). RNA copies of these elements are found in cells, and

in vitro, cloned examples are transcribed by RNA polymerase III. Interest in these SINEs was aroused when it became clear that their large numbers were probably due to their amplification by duplicative transposition (Jagadeeswaran et al., 1981). This belief was based on the observation that (a) one end of each Alu repeat bears a stretch of dAMP-rich sequence, suggesting an RNA intermediate step, and (b) each unit is flanked by short (7-20bp) direct repeats, indicative of transposition.

Many pseudogenes may be the result of a similar mechanism, as their lack of introns suggests that a processed mRNA has been used as a template by a cellular reverse transcriptase (see Sharp, 1983 for a review). Recent evidence suggests that the majority of Alu-like sequences are themselves pseudogenes, and that the genuine article, exemplified by an element found in Drosophila DNA, is represented only a few times in a genome (Brown, 1984). It is proposed that the progenitor of this particular class of SINEs encodes 7SL RNA, which is an essential part of the signal recognition particle used to transport proteins across membranes (Walter & Blobel, 1982; Ullu & Tschudi, 1984).

The major LINE family of mouse DNA (30,000 copies) has been described variously as L1, Bam H1 or MIF-I, and it is partly homologous to the major primate Kpn I LINE family (Singer et al., 1983). MIF-I elements vary in size from 5-7kbp, and some appear to be at least partly made up of smaller, SINE-like repetitive elements. Thus one MIF-I unit described by Martin et al. (1984) contains a composite element derived by the interaction of a 475bp-long member of the Alu-like R family (Gebhard et al., 1982) with a 500bp BAM5 element (Fanning, 1982). In addition, a highly conserved 1350bp Eco RI fragment of MIF-I DNA (contiguous with the BAM5/R composite) includes sequences sharing close homology with an IAP LTR (Brown & Huang, 1982; Brown, 1983). It has been suggested that this interaction of repetitive DNA elements in the

mammalian genome is not a rare event (Wilson & Storb, 1983).

The discovery of a solo IAP LTR in a MIF-I member is particularly interesting as it is a further indication of the types of interaction which may occur in the formation of retrovirus-like elements from smaller genetic building-blocks, as proposed by Temin (1980). Retrovirus-like sequences may be considered as a slightly different category of moderately repeated element than the majority of SINEs and LINEs, as although both categories show evidence of mobility within the genome, B1/Alu-like or MIF-I/Kpn-like elements lack the long symmetrical ends which are characteristic of both retrovirus-like elements and transposable elements (Shimotohno *et al.*, 1980).

I-6.2 Transposable elements and retrovirus evolution

Transposable elements are DNA segments of defined length which can insert themselves into new areas of a DNA genome. They are found in prokaryotes and eukaryotes, and are associated with a wide variety of genotypic and phenotypic changes in their host organisms. Those of bacteria range from relatively simple, 1200-1400bp-long IS elements to transposons (eg Tn5, Tn10) which are several kbp long and which also encode different gene products.

Some transposons may have arisen as a result of a structural gene becoming flanked by two IS sequences, thereby causing it to be mobilized. For example Tn9, coding for chloramphenical resistance, carries a copy of IS1 at each of its termini (Alton & Vapnek, 1979). Most transposons have long inverted, rather than direct, terminal repeats (Finnegan *et al.*, 1982).

As suggested above, elements exhibiting close structural and functional parallels with bacterial transposons also exist in eukaryotes. These include the *copia*, *copia*-like and P elements of *Drosophila*, and the Ty-1 elements of *Saccharomyces cerevisiae*. Both

copia and Ty-1 are about 5kbp long and possess 300-500bp terminal direct repeats. Ty-1 occurs about 35 times in yeast DNA, and its repeats are termed delta (δ) sequences. These are reiterated approximately 100 times in the yeast genome, and are often found as "solo" elements (Cameron et al., 1979). Although solo δ elements are sometimes left behind when a Ty-1 element is excised by recombination (Roeder & Fink, 1980) there is no evidence that the δ unit itself can transpose (Calos & Miller, 1980).

Copia and copia-like elements of Drosophila comprise about 30 families of transposable elements, each with 20-40 members distributed throughout the genome. Together these families of sequences make up half of the moderately repeated DNA of Drosophila, or 5-10% of the whole genome (Finnegan et al., 1982).

The exact mechanism of DNA transposition is unknown, although several models have been proposed (eg Shapiro, 1979; reviewed by Bukhari, 1981). The only sequence structure common to all bacterial, Drosophila and yeast transposable elements seems to be their short inverted terminal repeats (Finnegan et al., 1982). All of these mobile elements generate small regions of duplicated host DNA sequence at the site of insertion, suggesting a common mode of integration.

As mentioned above, structural similarities suggest an evolutionary link between transposable elements and pro-retroviruses (reviewed by Temin, 1980; Flavell, 1981; Finnegan, 1981). The most popular element proposed to be the "missing link" is copia.

Circular copia DNA has been detected in Drosophila tissue culture cells, significantly as two size classes relecting the presence of one or two terminal repeats in tandem (Flavell & Ish Horowicz, 1981). Copia shares with retroviruses those sequences thought to be important for reverse transcription and transcription (Kulgushkin et al., 1981; Will et al., 1981). Perhaps most importantly, retrovirus-like particles

(VLPs) from sonicated Drosophila cell nuclei have been shown to contain 5kbp RNA molecules homologous to copia DNA (Shiba & Saigo, 1983). This VLP RNA stimulates the in vitro synthesis of VLP structural proteins, and purified VLPs possess reverse transcriptase activity (Shiba & Saigo, 1983).

Similarities between these "copia retroviruses" and especially the IAP class of mouse endogenous retroviruses have been observed (Finnegan, 1983). Both are encoded by families of genes which comprise a significant proportion of their respective host genomes (Potter et al., 1979; Lueders & Kuff, 1977) and neither particle is infectious (Shiba & Saigo, 1983; Kuff et al., 1972). Furthermore both DNA elements are transposable and can cause insertional mutations (Bingham & Judd, 1981; Rubin et al., 1982; Hawley et al., 1982, 1984). Mouse VL30 elements have also been proposed as links between transposable elements and retroviruses (Flavell, 1981) but as yet their protein-coding potential is unknown, and their only association with reverse transcriptase activity is that supplied by a competent retrovirus when VL30 RNA is packaged into a pseudotype particle.

Just as Tn9 may be the result of capture of the chloramphenicol resistance gene by two IS elements, and Ty-1 the result of a similar action by solo δ elements, so Temin (1980) proposes that retroviruses emerged in a series of similar events. The initial move would be the capture of a DNA polymerase gene by two eukaryotic IS-like elements. Recently Wirth et al. (1983) described a family of 500 LTR-like sequences (the LTR-IS family) in the mouse genome which resemble the insertion-like elements of Temin's protovirus hypothesis. More recent evidence suggests that LTR-IS are also mobile (Wirth et al., 1984). Interestingly, solo VL30 LTRs have also been isolated from libraries of mouse genomic DNA (Rotman et al., 1984). LTR-IS elements share short regions of homology with both these and the LTRs of the NVL-3 VL30 cDNA

clone (this thesis; Norton et al., 1984b) but are apparently unrelated to IAP gene LTRs (Wirth et al., 1983).

I-6.3 Organization and expression of endogenous retroviruses

Endogenous proviruses detected in the genomes of most vertebrates show evidence of great evolutionary age (see section D-5). However most examples still resemble closely the proviruses of exogenous retroviruses recently acquired by infection, and this has led to the wide belief that they arose from infection of the host animal in its germ line cells (Steffen & Robinson, 1982). These relatively simple packets of genetic information have been useful model systems in the study of both the evolution (Benveniste et al., 1977) and the expression of eukaryotic gene families (Hoffman et al., 1982). Some of the most recently described endogenous proviruses include those isolated from libraries of human genomic DNA (O'Connell et al., 1984) but the best studied examples are the MuLV- and ALV-related proviruses of the mouse and chicken genomes (reviewed by Coffin, 1982; Rovigatti & Astrin, 1983; Risser et al., 1983).

The major endogenous MuLV family is made up of xenotropic MuLV (X-MuLV) members, whilst most strains of chicken contain several copies of closely-related ALV-type proviruses which encode subgroup E viruses (Steffen & Robinson, 1982). Interestingly both families of virus are prevented from reinfecting their host species by a specific blockage at the penetration stage. Thus most mice lack cell-surface receptors for X-MuLV particles, and most chickens lack subgroup E receptors. Furthermore strains of chicken that do possess these receptors often express high levels of viral envelope glycoproteins, encoded by otherwise defective proviruses, which serves to render the receptors inaccessible to endogenous virus particles (Robinson et al., 1981). It has been argued that these mechanisms are the result of evolutionary

pressure to minimize the deleterious effects of these viruses on their hosts (Steffen & Robinson, 1982). Examples of retrovirus-induced changes which would probably not be evolutionarily favoured include the activation of cellular oncogenes by downstream promotion (Hayward et al., 1981; Westaway et al., 1984) or by the provirus integrating within the oncogene itself (Rechavi et al., 1982; Kuff et al., 1983), and the similar disruption of developmentally regulated structural genes (Breindl et al., 1984; Harbers et al., 1984).

How are these endogenous retroviruses organized and expressed? The ALV-related proviruses of inbred White Leghorn chickens have been studied by restriction enzyme mapping, and can be identified as at least 16 distinct proviruses. Ten chromosomal loci (ev loci) have been identified, and unexpectedly 6 of these are all located on chromosome 1. Three of these proviruses contain 5' terminal deletions, and it has been proposed that their asymmetric distribution is the result of duplication of part of chromosome 1 (Tereba, 1983). Importantly, this type of distribution may be peculiar to this strain, since other chickens such as White Plymouth Rock possess randomly scattered ev loci (Tereba, 1983).

In some cases the structure of endogenous ALVs correlates well with their transcriptional activity. As expected, two of the 5'-deleted chromosome 1 proviruses (ev-4 and ev-5) are not transcribed due to loss of the LTR-encoded promoter sequences. However ev-6 appears to be under the control of a fortuitously-located cellular DNA promoter, so that low levels of a truncated, genome-length RNA are produced (Rovigatti & Astrin, 1983).

The most commonly occurring endogenous ALV provirus is the ev-1 locus (Tereba & Astrin, 1980). This is also located on chromosome 1 and may be the progenitor of the other ev proviruses of this chromosome (Tereba, 1983). Although this provirus contains extremely short (273bp)

LTRs (Hishinuma et al., 1981) it is otherwise free from detectable sequence rearrangements. Nevertheless the normal 35S and 21S mRNAs which it encodes are produced at extremely low abundance in the cell (Rovigatti & Astrin, 1983). Since its efficient expression can be induced by treatment of chick embryos with 5-azacytidine, it is thought that in this instance the chromosomal location of ev-1 plays a central role in the control of its expression (Groudine et al., 1981).

Not all proviral deletions result in poor expression. The ev-3 locus (on a chicken microchromosome) contains a provirus suffering a deletion in its gag and pol genic regions (Tereba, 1981; Rovigatti & Astrin, 1983). Nevertheless a truncated (31S) mRNA and a normal 21S mRNA are expressed at moderately high levels (50-150 copies per cell, and 30-80 copies per cell, respectively) (Hayward et al., 1980; Baker et al., 1981). Unlike ev-1 loci, ev-3 provirus sequences are undermethylated and the LTRs each contain nuclease-hypersensitive sites (Groudine et al., 1981; see below).

As stated above, most of the endogenous MuLV of mice are of the X-MuLV class (Coffin, 1982). These can be distinguished from ecotropic MuLV proviruses by using as a DNA hybridization probe a 500bp-long restriction fragment derived from cloned X-MuLV env gene sequence (Hoggan et al., 1982). Restriction enzyme digestion and Southern transfer analysis of mouse genomic DNA have enabled these workers to arrange X-MuLV into at least 7 closely related but detectably different families (Hoggan et al., 1983). Interestingly the relative abundance of a particular family of X-MuLV proviruses did not correlate with the recovery of its corresponding infectious virus (Hoggan et al., 1982). This observation is reminiscent of the chicken system, where only a few ALV proviruses may be efficiently expressed.

As lymphocytes are common targets for retrovirus-induced malignancies these cells have been particularly well studied with regard

to the expression of endogenous retroviruses [for several reviews, see Current Topics in Microbiology and Immunology, volume 98 (1982)]. Only a few strains of mice contain lymphocytes with efficiently expressed X-MuLV proviruses (Morse & Hartley, 1982). Furthermore these mice do not present a consistent pattern of lymphocyte X-MuLV expression. For example in NZB mice the highest X-MuLV expression occurs in bone marrow lymphocytes, but in F/St mice the highest X-MuLV particle producers are the thymocytes (Morse & Hartley, 1982). It has been suggested that these rare examples of spontaneous X-MuLV expression are the result of assaying immunologically stimulated mice, and that usually, endogenous X-MuLV genes in lymphocytes are silent until specifically activated (Wecker & Horak, 1982). Surveys using a range of mitogens and allo-antigens have indicated that whole particle induction only occurs in B-cells, whether resulting in infectious virus or in defective particles (Wecker & Horak, 1982). Nevertheless, T cell-specific mitogens such as concanavalin A, whilst unable to induce the expression of infectious virus particles, can enhance the expression of X-MuLV envelope gp70 in the plasma membrane of these cells (Wecker & Horak, 1982). Thus gp70 expression may be used as a general marker for activated mouse lymphocytes of B- or T-cell lineage. The implications of X-MuLV <u>env</u> gene expression for leukaemogenesis are discussed in section D-7.

Since proviral LTRs themselves contain all the information necessary for retrovirus gene expression, it is likely that the regulation of endogenous virus expression is achieved by modification of the proviral DNA by the host, and/or influence of the DNA sequences which flank each provirus.

A large body of experimental data indicates that eukaryotic organisms control their gene activity using specific patterns of DNA methylation (see Razin & Szyf, 1984, for a review). As described in section R-6, this occurs at CpG dinucleotides, moreover these modified

bases must be at specific locations within (or upstream of) the gene to ensure transcriptional silence. The formation of tissue-specific patterns of expression is believed to occur during development by sequential changes in the DNA methylation pattern. Sperm DNA, for instance, is extensively methylated (Razin & Riggs, 1980). Inhibitors of the DNA methylase enzyme, such as 5-azacytidine or DMSO, can induce cellular differentiation (Creusot et al., 1982). This observation supports the hypothesis that differentiation of the zygote into its various tissues may involve the hypomethylation of specific sites in the DNA. It is unknown whether this is achieved by a demethylase enzyme or by post-replicative inhibition of methylation.

Host control of retrovirus provirus expression has been studied in Rudolf Jaenisch's laboratory, where MuLV infection of early mouse embryos has resulted in the development of mouse strains each of which carries new endogenous proviruses in its germline tissue (Jaenisch 1976, 1979, 1980). Although some proviruses were found to be structurally defective, other, apparently normal, proviruses were either expressed at different times during development, or they remained silent throughout the life of the mouse (Jaenisch, 1980; Jaenisch et al., 1981; see section D-6). This again implies that different integration sites may confer different opportunities for transcriptional activity. In some cases, extra proviruses became integrated and expressed in the normal target tissues of the MuLV. A positive correlation was observed between the tissue-specific expression of these proviruses and the undermethylation of their CpG dinucleotides (Harbers et al., 1981; Stuhlman et al., 1981).

Undifferentiated embryonal carcinoma cells (EC) have also been used as a model system for studying early gene expression. As also observed by Jaenisch, retroviral genomes are not expressed in the undifferentiated cells of early embryos (Peries et al., 1977). Most

newly integrated proviruses in EC cells become methylated (Stewart et al., 1982) but only several days after integration. This suggests that methylation may be a result, rather than a cause of transcriptional inactivity (Gautsch & Wilson, 1983). Again the site of integration may be important, as approximately 1 in 5000-10,000 of these newly acquired proviruses remain transcriptionally active, suggesting that integration into an actively expressed part of the chromosome has allowed the provirus to escape methylation (Sorge et al., 1984).

Once differentiation is complete, de novo methylation is a very rare event (Razin & Szyf, 1984). Therefore infection of differentiated cells invariably results in newly integrated proviruses that are undermethylated and expressed. As an example, the expression of endogenous MMTV proviruses is suppressed in both normal and transformed tissues of most inbred strains of mice, whereas exogenously acquired proviruses are expressed (McGrath et al., 1978). Thus even in tissues in which closely related or identical MMTV proviruses are acquired as a result of both genetic and milk-borne infection (eg MMTV (GR) tumours in GR mice) the endogenous proviruses are methylated and the acquired ones are hypomethylated (Cohen, 1980).

I-6.4 Mouse VL30 elements

The introduction of these will be deliberately brief, as further details are given at the beginning of each section of Results, and some aspects are expanded upon in the Discussion.

When MuLV is propagated on mouse or rat cells the progeny virus particles often selectively package an abundant cellular 30S RNA which shares many structural properties with 35-38S retroviral genomic RNA (Howk et al., 1978; Besmer et al., 1979).

In the mouse, these virus-like 30S (VL30) RNA species are encoded by a family of 100-200 endogenous retrovirus-like elements of unknown

origin (Keshet & Itin, 1982; Keshet & Shaul, 1981; Courtney et al., 1982a,b). Although their biological role is unknown, VL30 species are of particular interest for many reasons. These include: (a) on two separate occasions, members of the closely related rat VL30 family have participated in the genesis of the strongly oncogenic Ki-MuSV and Ha-MuSV genomes (Ellis et al., 1980) and (b) VL30 RNA may be used as a template for reverse transcription into provirus-like VL30 DNA. It can thus be transmitted from cell to cell by a retrovirus vector (Scolnick et al., 1979) and may therefore be capable of mutagenizing the genome of its new host, as known to occur with replication competent retroviruses.

The work presented in this thesis describes the cDNA cloning and characterization of retrovirus-transmissible mouse VL30 RNA species. These clones have been used to study the organization, evolution, and transmission of this mouse retrovirus-like family.

MATERIALS

MATERIALS

ML-1 Chemicals were obtained from the following sources:

Amersham International plc, Amersham, Bucks. UK

Adenosine 5'-[γ -^{32}P] triphosphate (5'-γ ^{32}P) ATP, 3000Ci mmol^{-1}

Deoxycytidine 5'-[α-^{32}P] triphosphate (^{32}P-dCTP), 3000Ci mmol^{-1}

[methyl-^{3}H] Thymidine 5'-triphosphate (^{3}H-TTP), 50Ci mmol^{-1}.

BDH Chemicals Ltd, Atherstone, Warks., UK

Amberlite monobed resin, MB-3; 4-amino salicylic acid, sodium salt; bromophenol blue; dimethyl sulphoxide; N, N'-methylenebisacrylamide (specially purified for electrophoresis); 2-mercaptoethanol; Nonidet P-40; sodium dodecyl sulphate; urea.

Bio-Rad Laboratories Ltd, Watford, Herts., UK

TEMED (N,N,N',N'-tetramethylenediamine).

Boehringer-Mannheim (BCL) The Boehringer Corporation Ltd, London, UK

Proteinase K (EC 3.4.21.14).

BRL (Bethesda Research Laboratories) Ltd, Cambridge, UK

Low melting point agarose, electrophoresis grade; Bam HI synthetic molecular linkers.

Difco Laboratories, Detroit, Michigan, USA

Bacto-Agar; Bacto-Tryptone, Nutrient Broth.

Fisons Scientific Apparatus, Loughborough, Leics., UK

Acrylamide, EDTA.

FMC Corporation, Marine Colloids Div., Rockland, ME, USA

"Sea Plaque" low gelling agarose.

Hopkin and Williams Ltd, Chadwell Heath, Essex, UK

HEPES; "Repelcote" water repellant for silicone treatment.

Oxoid Ltd, Basingstoke, Hants., UK

Yeast extract (powder).

Pharmacia Fine Chemicals (Great Britain) Ltd, Hounslow, Middx., UK

Dextran sulphate (sodium salt) from dextran with M.wt. approx. 500,000; Ficoll 400 (M.wt. 400,000); Sephadex G-50, G-100, G-200 (Superfine).

Sigma Chemical Company, Poole, Dorset, UK

Agarose type II, medium EEO; ampicillin, anhydrous; antifoam A emulsion; bovine serum albumin, fraction V; calcium chloride (dihydrate) grade I; chloramphenicol; C-TAB (hexadecyltrimethylammonium bromide); dithiothreitol; DEAE-dextran; 2'-deoxyadenosine 5'-triphosphate (disodium salt); 2'-deoxycytidine 5'-triphosphate (sodium salt); 2'-deoxyguanosine 5'-triphosphate (sodium salt); deoxyribonucleic acid, type I: sodium salt "highly polymerized" from calf thymus; type III: sodium salt from salmon testes; ethidium bromide; lysozyme, grade I from egg white (EC 3.2.1.17); polyvinyl pyrrolidone (PVP-40) average M.wt. 40,000, pharmaceutical grade; "Sarkosyl" (N-lauroylsarcosine) sodium salt; tetracycline HCl; thiamine HCl; thymidine (Sigma grade); thymidine 5'-triphosphate (sodium salt); Trizma base; triton X-100.

ML-2 Enzymes used in molecular cloning were obtained from the following sources:

Boehringer-Mannheim (BCL) Ltd

Calf intestinal alkaline phosphatase (EC 3.1.3.1); DNA polymerase I (*E. coli*; EC 2.7.7.7); nuclease S1 (*Aspergillus oryzae*, EC 3.1.30.1).

BRL Ltd

T4 DNA ligase (T4-infected *E. coli*, EC 6.5.1.1).

Uniscience Ltd, Cambridge, UK

T4 Polynucleotide kinase (T4-infected *E. coli*, EC 2.7.1.78).

Sigma Chemical Company

Ribonuclease A (type IA from bovine pancreas, EC 3.1.27.5).

Worthington Biochemical Corporation, New Jersey, USA

Deoxyribonuclease I, electrophoretically purified (bovine pancreas, EC 3.1.21.1).

RNA-dependent DNA polymerase (reverse transcriptase) from avian myeloblastosis virus was a gift from Dr. J.W. Beard, Life Sciences Inc., St. Petersberg, Florida, USA.

ML-3 Restriction endonucleases were obtained from:

Amersham International plc, Boehringer-Mannheim Ltd, BRL Ltd, and New England BioLabs, Beverley, MA, USA. These were routinely tested on wild-type bacteriophage lambda DNA, obtained from Miles Laboratories Ltd, Slough, UK.

ML-4 Materials for tissue culture

All cell lines were grown using Dulbecco's modified Eagle's medium (DMEM), made up by the virus group media laboratory. This was supplemented with 10% (v/v) newborn calf serum (Flow Laboratories Inc., Irvine, Scotland, UK) and 100 units ml^{-1} of cristamycin [600 mg (1 million units) of penicillin to 1g of streptomycin, Glaxo Laboratories Ltd, Greenford, Essex, UK]. Batches of DMEM (stored at $4^{o}C$) which were over 1 month old were further supplemented with 20 ml L-glutamine (200 mM). For pH control, DMEM contained 5% (w/v) sodium bicarbonate solution to pH 7.2. Cells were grown in a humidified atmosphere containing 5% CO_2.

All media were tested for sterility by inoculation into capped test tubes of Nutrient Broth (Difco Laboratories) followed by incubation at $37^{o}C$ for 5 days.

ML-5 Buffers and solutions

(i) General

These were sterilized by autoclaving at 15lb psi for 15 minutes. Solutions containing sucrose were autoclaved at 10lb psi for 20 minutes.

10 x DNA extraction buffer: 2M NaCl, 200mM Tris HCl, 10mM EDTA (pH 7.5).

Luria Broth: 10g bactotryptone, 5g yeast extract, 5g NaCl in $1dm^3$ distilled water. pH was adjusted to 7.2 with 1 M NaOH solution. Usually supplemented with 10 ml 20% glucose, plus 1 ml of a filter-sterilized solution of thiamine HCl and thymidine ($10mg\ ml^{-1}$ each), per dm^3.

PBS: 140mM NaCl, 2.7mM KCl, 8.1mM Na_2HPO_4, 1.5mM KH_2PO_4, pH 7.4.

Sephadex Column Buffer: 20mM Tris HCl, 50mM NaCl, 1.0mM EDTA, 0.05% (w/v) Sarkosyl.

SSC: 150mM NaCl, 15mM tri-sodium citrate.

STE: 100mM NaCl, 20mM Tris HCl, 1mM EDTA, pH 7.5.

TAE: 40mM Tris, 20mM sodium acetate, 2mM EDTA, adjusted to pH 7.7 with glacial acetic acid.

10 x TBE: 108g Trizma base, 55g boric acid, 9.5g EDTA in $1dm^3$ distilled water.

TCM: 10mM Tris HCl, 10mM $CaCl_2$, 10mM $MgCl_2$, pH 7.6.

TE: 10mM Tris HCl, 0.5mM EDTA, pH 7.5.

TNE: 100mM NaCl, 1mM EDTA, 50mM Tris HCl, pH 7.5.

"Trypsin" for cell passage: 8g NaCl, 0.38g KCl, 0.1g Na_2HPO_4, 1g D-glucose, 3g Trizma base, 15ml 1% (w/v) phenol red, 10^5 units of penicillin, 0.1g streptomycin and 2.5g trypsin (EC 3.4.21.4) in $1dm^3$ distilled water, pH 7.7.

"Versene" for cell passage: 8g NaCl, 0.2g KCl, 1.15g Na_2HPO_4, 0.2g KH_2PO_4, 0.2g EDTA, 1% (w/v) phenol red, in $1dm^3$ distilled water.

VSB (Viral Suspension Buffer): 200mM NaCl, 10mM Tris HCl, 1mM EDTA, pH 7.5.

(ii) Buffers for enzyme catalyzed reactions

Restriction endonuclease buffers were made up as 10X concentrates and stored at room temperature as 20ml aliquots in glass Universal bottles. Immediately before use, 20mg BSA and 100µl 2-mercaptoethanol were added (final [1X] concentrations, 100µgml^{-1} BSA, 7mM 2-mercaptoethanol).

Hind III, Pst I

10mM Tris HCl, 50mM NaCl, 10mM $MgCl_2$, pH 7.4.

Eco RI

100mM Tris HCl, 50mM NaCl, 10mM $MgCl_2$, pH 7.4.

Kpn I

6mM Tris HCl, 6mM NaCl, 6mM $MgCl_2$, pH 7.5.

Bam HI, Xba I, Xho I

6mM Tris HCl, 150mM NaCl, 6mM $MgCl_2$, pH 7.9.

Sal I

6mM Tris HCl, 150mM NaCl, 6mM $MgCl_2$, pH 7.4.

Hinf I

6mM Tris HCl, 20mM NaCl, 6mM $MgCl_2$, pH 7.4.

Sma I

6mM Tris HCl, 20mM KCl, 6mM $MgCl_2$, pH 8.0.

Bgl II

10mM Tris HCl, 60mM NaCl, 10mM $MgCl_2$, pH 7.4.

Pvu II

6mM Tris HCl, 60 mM NaCl, 6mM $MgCl_2$, pH 7.5.

Msp I, Hpa II

10mM Tris HCl, 10mM KCl, 10mM $MgCl_2$, pH 7.4.

Sac I, Sst I

14mM Tris HCl, 90mM NaCl, 6mM $MgCl_2$, pH 7.5.

(iii) Other DNA-modifying enzyme buffers

Kinase buffer (made up as 10X concentrate, stored at room temperature)

6.6mM Tris HCl, 10mM $MgCl_2$, pH 7.6 (7mM 2-mercaptoethanol).

Ligase buffer (made up as 5X concentrate)

66 mM Tris HCl, 6.6mM $MgCl_2$, 1mM ATP, 10mM DTT, pH 7.6. (Stored at -20°C in 0.5ml aliquots).

DNAase I buffer (made up as 1X)

10mM Tris HCl, 5mM $MgCl_2$, (1mg ml^{-1} BSA), pH 7.5. (Stored at room temperature in 10 ml aliquots. BSA added immediately before use).

Nick-translation buffer (made up as 10X concentrate)

50mM Tris HCl, 50mM NaCl, 10mM $MgCl_2$, 7mM 2-mercaptoethanol, 50 $\mu g ml^{-1}$ BSA, pH 7.5. (Stored at -20°C in 0.5 ml aliquots).

SI Nuclease buffer (made up as 5X concentrate)

100mM sodium acetate, 200mM NaCl, 5mM $ZnCl_2$, pH 4.5.

METHODS

METHODS

M-1 Virus and cell lines

Ki-MuSV/MuLV was supplied as a culture of infected rat kidney cells (KNRK) producing 8 x 10^4 pfu and 1 x 10^6 ffu per ml of tissue culture fluid (TCF). This cell line was a gift to Dr. Roger Avery from Dr. S. Aaronson (NIH, Bethesda, Maryland, USA).

NIH-3T3 cells were originally obtained from Dr. J. Levy (University of California, San Francisco, California, USA) and sub-cloned to give the cell line NIH 4 (Morris *et al.*, 1980).

M-1.1 Establishment of a MuLV-infected mouse cell line (A1)

NIH 4 cells were infected with KNRK TCF virus to generate the K-NIH 4 line. This produced 1.3 x 10^2 pfu and 1 x 10^5 ffu per ml of TCF.

K-NIH 4 TCF virus was passed at low multiplicity of infection (< 1 pfu per cell) five times in succession in NIH 4 cells. The resulting line was designated K-NIH SP. Virus produced by these cells had titres of 3 x 10^5 pfu and 1 x 10^4 ffu per ml of TCF. Thus, a cell line was selected which produced 30 times more MuLV than MuSV.

NIH 4 cells were sub-cloned to give the line NIH (-3T3) 4E. These cells were passaged 16 times, following which they were infected with the virus stock from KNIH SP TCF. 24h after infection, the cells were trypsinized, seeded onto micro-test plates at a density of 100 cells per well and incubated at 37°C. Culture fluids were tested for MuLV by the XC assay using uninfected NIH-3T3 cells. XC positive wells were trypsinized and grown up. Cell lines with high reverse transcriptase activity were again subjected to both XC and focus assays to identify lines producing MuLV alone. The cell line, A1, was one such line.

These cells had a flat, non-transformed morphology and grew as a monolayer to low density in a manner similar to the parental NIH 4E cells. They also exhibited a low plating efficiency in soft agar and

were not tumourigenic at an inoculum of 10^6 cells per mouse (Morris *et al.*, 1980). The A1 line produced a titre of 6 x 10^6 pfu per ml of TCF by the XC assay, but no sarcoma virus was detected in focus assays or in rescue experiments. It was assumed that the MuLV obtained from the A1 line was the same as the input Ki-MuLV, since NIH-3T3 cells do not release endogenous retroviruses (Avery & Levy, 1978).

M-1.2 Rat cells infected with MuLV

NRK (normal rat kidney) cells from Osborne-Mendall rats were obtained from Dr. J. Levy and cloned to give the cell line NRK 4.

NRK 4 cells were infected with the TCF from A1 cells (producing Ki-MuLV), passaged four times as a mass-infected culture and stored in liquid nitrogen. This cell line, designated KLNRK-1, was used on two separate occasions to make high molecular weight DNA preparations (at 15 and 19 passages post-infection, respectively).

M-2 Cell culture

M-2.1 Recovery of cells stored in liquid N_2

Vials removed from liquid N_2 were thawed quickly by warming in a 37°C water bath. Cells were resuspended then diluted with a further 2 ml of warmed DMEM. Following a low-speed centrifugation in a bench centrifuge the pelleted cells were resuspended in fresh DMEM and transferred to a sterile 25cm^2 or 75cm^2 plastic tissue culture flask.

M-2.2 Growth of tissue culture cells

Mouse and rat fibroblast cell lines were grown at 37°C in plastic tissue culture flasks and in 2.5dm^3 glass roller bottles in 5% CO_2 and 95% air. Cell cultures were incubated until the monolayers were confluent; if confluency had not occurred by 3 to 4 days, fresh medium was added. Continuous culture of cell lines was achieved by passaging cells as soon as confluency was reached. The cell culture fluid was removed by suction and the monolayer carefully rinsed with warmed

versene. Cells were detached from their substrate by addition of a 1:4 (v/v) mixture of trypsin:versene followed by a short period (1 to 2 minutes) of incubation at 37°C. An equal volume of fresh DMEM and 10% (v/v) new born calf serum was added; this served both to maintain the correct pH and to inactivate the trypsin. Cells were gently dispersed by pipetting, and diluted for further growth. Fast-growing cell lines were passaged 1:10, whilst slower growing lines were passaged 1:3 to 1:5.

Normally, cell lines were maintained in 75cm^2 plastic tissue culture flasks, however for larger quantitites of cells 150cm^2 flasks or roller bottles were used.

M-2.3 Storage of tissue culture cells

A 75cm^2 flask containing a recently confluent monolayer of cells was trypsinized and the cells suspended in 10ml of fresh DMEM. Cells were pelleted by 5 minutes of low speed centrifugation and the medium removed. The pellet was resuspended in 2ml freezing medium: DMEM, 25% (v/v) new born calf serum, 8% (v/v) DMSO, and 1ml placed in each of two sterile freezing vials. These were frozen at -70°C overnight in polystyrene boxes to ensure the slow rate of cooling which is necessary to minimize damage to the cells by the formation of large ice crystals. Freezing vials were then transferred to liquid nitrogen. Cells stored in this way for over 7 years have been successfully recovered, with minimum loss (<25%) of viability.

M-3 Infection of cells by retroviruses

For all preparations, cells were grown to approximately 60% confluency as estimated by visual examination. TCF was removed and the cells treated with DEAE-dextran (25μg ml^{-1}) in DMEM. After 30 minutes the dextran was removed and the cells washed with warmed DMEM. This was removed and the virus inoculum added. Filtered TCF from productively

infected cell lines was used without further concentration. Virus was allowed to adsorb for 30 to 60 minutes at 37°C, then a further volume of fresh medium was added. Infected cells were cultured as described for uninfected cells (section M-2.2).

M-4 Assay of retroviruses

Although these assays were not used directly for the experiments described in this thesis, they are included to more fully explain the terms pfu and ffu as applied to retroviruses.

M-4.1 MuLV plaque assay (XC assay)

A modification of the techniques described by Klement *et al*. (1969) and Rowe *et al*. (1970) were used.

The XC cell line is a rat cell line which was transformed by infection with the Prague strain of Rous sarcoma virus (RSV) (Svoboda *et al*., 1963). When this is placed in contact with cells previously infected with murine leukaemia virus, syncytium formation occurs. Plaques in the XC assay are seen as holes in the cell monolayer containing multiple giant cells, or as focal masses of giant cells. Plaque formation in permissive cells (eg NIH-3T3 cells) follows one-hit kinetics (Rowe *et al*., 1970).

5cm plastic Petri dishes were seeded at a density of 3.5×10^5 cells per plate with NIH-3T3 cells in DMEM containing 10% (v/v) NBCS, and these were incubated at 37°C in a humidified 5% CO_2, 95% air atmosphere for 24 hours.

The TCF was removed and the cells were treated with DEAE-dextran as described in section M-3, using 2ml of warmed solution per plate.

After 30 minutes the dextran solution was removed and the monolayer washed once with warmed DMEM. Virus samples or control (DMEM) were added (0.4ml per plate) in duplicate.

Virus was allowed to adsorb for 30 minutes at 37°C, followed by

addition of 4ml DMEM containing 10% (v/v) NBCS, and the plates incubated at 37°C. Culture medium was changed on the 2nd and 4th days after infection.

On the 6th day post infection, TCF was removed and the cell monolayers irradiated for 1 to 2 minutes at a distance of 65cm from the tissue culture hood UV lamp (60erg/mm^2/sec). A suspension of 10^6 XC cells in 4ml DMEM, 10% (v/v) NBCS was then added to each plate.

Culture medium was changed very carefully on the first or second day after irradiation; the cell monolayers were very fragile at this stage.

On the 3rd or 4th day after irradiation (when the XC layer was confluent), TCF was carefully removed and the cells washed with PBS. Monolayers were fixed and stained by addition of 0.11% (w/v) basic fuschin, 0.33% (w/v) methylene blue in methanol for 30 minutes. The stain was removed and "plaques" were counted.

Dishes, each containing 100 A1 cells in DMEM, 20% (v/v) NBCS, incubated until the test plates were ready for XC overlay, were used as positive controls.

M-4.2 MuSV Focus Assay

This was adapted from a quantitative *in vitro* assay developed by Hartley & Rowe (1966). The focus assay relies on the property of MuSV to transform fibroblasts and induce the formation of cell clumping, or foci, on the monolayer.

5cm plastic Petri dishes were seeded with 4ml DMEM, 10% (v/v) NBCS containing 3 x 10^5 NRK cells. These were then incubated at 37°C overnight, by which time they were approximately 50% confluent.

Monolayers were checked for absence of clumping, and the TCF removed. Each monolayer was incubated with 2ml DEAE-dextran (25μg ml^{-1} in DMEM) at 37°C for 30 minutes. Medium was removed and the plates washed carefully with warmed DMEM. Virus samples or DMEM (0.4ml each)

were added to the appropriate dishes, followed by incubation at 37°C for 30 minutes to allow virus adsorption to take place. Plates were inoculated in duplicate. After incubation, 4ml DMEM, 5% (v/v) heat-inactivated NBCS was added per plate. (Heat treatment was incubation at 55°C for 1 hour).

Plates were incubated at 37°C and the TCF replaced with fresh medium every 3 days, with an extra change on the day before the plates were read. This was usually about 11 days after infection. Dishes were drained of TCF before the foci were counted, as the foci became easier to see as the monolayers dried out.

M-5 Concentration of virus from tissue culture fluid

Cells producing virus were either grown in 2.5dm^3 roller bottles and harvested manually, or were grown in 4dm^3 Bellco bottles and harvested automatically by a Bellco Smith-Kozoman Auto-Harvester, supplied by Arnold Horwell Ltd, London, UK.

Virus intended for the preparation of intact viral RNA was harvested from the TCF every 2 hours; longer periods of contact with the cell monolayers resulted in more degraded RNA, possibly due to ribonuclease activity within the viral particles themselves (J. Norton, personal communication).

Virus to be used for infecting cells or for endogenous reactions was harvested every 4 hours from the TCF, with the former processed no further than filtration through a 1.2μM Millipore filter. TCF containing virus for endogenous reactions was stored at 4°C until 6 harvests had been collected and pooled, when it was processed as follows.

All manipulations were performed at 4°C. The chilled TCF was filtered through two layers of Whatman No. I paper discs (as a prefilter), then through a 1.2 μM Millipore disc filter. The filters

were retained in a Sartorius filter unit and the process was accelerated by maintaining a pressure of 5lb psi N_2 gas to the top half of the unit. One Millipore filter could be used for up to 250ml A1 TCF.

Virus was concentrated from the TCF by centrifugation in a 6 x 300 ml rotor at 20,000 rpm for 2.5 hours in an MSE "Superspeed" 65 centrifuge. The supernatant was removed and the tubes drained. The virus pellets were gently washed with ice-cold 50mM Tris HCl, pH 7.5 and resuspended in 2 ml of the same buffer. This crude preparation was further purified by pelleting through a 20% (w/v) sucrose cushion in 3 x 6ml centrifuge tubes, using a 3 x 6.5 ml rotor at 54,000 rpm for 1.5 hours.

Pellets were washed as before and resuspended in 300 to 500µl of 50mM Tris HCl, pH 7.5. The virus preparation was a fairly homogeneous, opalescent suspension. Virus yield was determined spectrophotometrically, using the relationship: 8 OD_{260} = 1mg ml^{-1} virus.

A typical virus yield from a 24 hour harvest using the Bellco Autoharvester varied between 4 to 8mg. Preparations of virus were either used immediately, or quickly frozen and stored at -70°C.

M-6 Preparation of virus RNA

Virus was purified as above, except that TCF was harvested every 2 hours from the A1 cells. For these smaller volumes, virus was pelleted by centrifugation at 40,000 rpm for 30 minutes at 4°C in an MSE 8 x 50 ml angle rotor. Since speed was the important factor in the recovery of intact RNA, the sucrose cushion step was omitted.

Virus pellets were resuspended in 10ml VSB and the particles were lysed by the addition of SDS to 1% (w/v) and 2-mercaptoethanol to 1% (v/v). An equal volume of phenol:chloroform (1:1, v/v) was added and the mixture was vigorously extracted at 4°C. After separation of the phases by centrifugation, the phenol phase was re-extracted with VSB and

the combined aqueous phases were once more extracted with phenol: chloroform. RNA was recovered from the aqueous phase by addition of 2.5 volumes of ice-cold ethanol and overnight precipitation at -20°C.

M-7 High molecular weight DNA

M-7.1 Non-Mus rodent DNA sources

Sciurus griseus (grey squirrel) liver was dissected from the fresh corpse of a female specimen caught in the wild near the University of Warwick campus.

High molecular weight DNA from *Clethrionomys glareolus* (bank vole) and *Apodemus sylvaticus* (woodmouse) were a kind gift from Ms Candice O'Connor, Dept of Biological Sciences, University of Reading, Berkshire. These two DNA samples were extracted from cell lines which had been established from whole embryos of locally trapped feral specimens.

M-7.2 Isolation of DNA from squirrel liver

Fresh squirrel liver was washed in PBS and cut into approximately $0.5cm^3$ portions. These were frozen rapidly by immersion in liquid N_2 and homogenized in the presence of solid CO_2 pellets with a sterile mortar and pestle. DNA was extracted as in section M-7.4.

M-7.3 Human DNA

This was prepared by Dr. John Norton from a mixed population of human lymphocytes. These were supplied as a fresh buffy coat suspension by the Birmingham Blood Transfusion Centre.

M-7.4 High molecular weight DNA from unfractionated cells

This was routinely extracted from 4 confluent roller bottles of cells. TCF was removed, and cell monolayers were carefully washed once with DNA extraction buffer (DEB; 200mM NaCl, 20mM Tris HCl, 10mM EDTA, pH 7.5).

To each roller bottle was added 50ml DEB, 0.5% (w/v) SDS. Cells were lysed by allowing the bottles to roll for 15 minutes at 37°C.

The cell lysate was decanted off and each roller bottle rinsed out with 10-20ml (DEB + SDS).

The pooled solutions were made 4% (w/v) with solid 4-amino salicylate, which serves as a protein denaturant. Once this was fully dissolved the solution was extracted for 10 minutes on ice with an equal volume of phenol:chloroform (1:1, v/v).

Following separation of phases by low-speed centrifugation the aqueous phase was removed. If this was too viscous to allow separation it was further diluted with DEB, the mixture re-shaken and again separated by centrifugation. Phenol extractions were repeated until the interface upon phase separation was clear.

The de-proteinized aqueous phase was poured into a glass beaker and 2 volumes of 100% ethanol carefully layered on top. High molecular weight DNA was recovered by spooling off from the aqueous/ethanol interface with a clean glass rod.

Spooled DNA was washed in 70% (v/v) aqueous ethanol to remove salts, immersed once more in 100% ethanol, scraped into a sterile Universal bottle and dried under a stream of N_2 gas.

The dried pellet of DNA was dissolved by incubation in 20ml TE at 4^oC over several days. The DNA solution was adjusted to 10mM with EDTA and incubated with RNAase A (50μg ml^{-1} final concentration) at 37^oC for 1 hour. RNAase A stock solution was first boiled for 10 minutes to inactivate any contaminating DNAases.

At the end of the RNAase digestion period, the solution was diluted to a reasonable viscosity and adjusted to 1 x DEB using a 10X stock solution. DNA was purified from enzyme and ribonucleotides by a further round of phenol:chloroform extraction and spooling from ethanol.

The final solid was dissolved in 1 to 5ml 0.1X TE at 4^oC and adjusted to 1mg ml^{-1}. DNA concentrations were determined spectrophotometrically, where 1 OD_{260} = 50μg DNA ml^{-1}.

M-8 Isolation of low molecular weight DNA from cells

This DNA fractionation procedure was adapted from that of Hirt (1967).

TCF was removed from the cells and the monolayers rinsed with PBS. To each roller bottle was added 10ml of 10mM Tris HCl, 10mM EDTA, 0.6% (w/v) SDS, pH 7.5, and the cells allowed to completely lyse by leaving to roll at 37°C for 15 minutes. The lysate was removed and 10-20ml lysing solution was used to rinse out the roller bottles. Both solutions were combined. The lysate was adjusted to 1M with NaCl, mixed gently by inversion and left at 4°C overnight.

Next day the tube was centrifuged at 17,000 x g for 30 minutes in a pre-chilled 8 x 50 ml angle rotor. High molecular weight DNA and SDS were pelleted by this procedure, while RNA and unintegrated virus DNA remained in the supernatant. Both fractions were recovered as follows:

(i) Supernatant

4-amino-salicylate was added to 4% (w/v) and the mixture was extracted twice with phenol:chloroform (1:1, v/v) then precipitated with 2 volumes of ethanol overnight at -20°C.

Next day the DNA was pelleted by low speed centrifugation and dissolved in TE. RNA was removed by adding LiCl to 2M and incubating on ice for 2 hours. The precipitated RNA was removed by centrifugation at 2,000 x g for 15 minutes. DNA in the supernatant was recovered by ethanol precipitation.

(ii) Pellet

This was dissolved in TE at room temperature, a process which could take 2-3 days. 4-amino-salicylate was added to 4% (w/v) and NaCl was added to a final concentration of 150mM. Following two rounds of phenol:chloroform extraction the DNA was recovered from solution by spooling under ethanol.

M-9 Preparation of unintegrated virus DNA from fractionated cells

Cells were fractionated into nuclei and cytoplasm, as originally described by Bachelor & Fan (1979).

Four roller bottles of sub-confluent cells were infected with A1 TCF as normal and incubated overnight at 37°C.

18 to 20 hours post infection, cells were removed by trypsinization and pelleted by low speed centrifugation. After a brief wash with ice-cold Tris-buffered saline (tissue culture saline containing 10mM Tris HCl, pH 7.5) the cells were again pelleted and resuspended in 19.5ml of lysing buffer (150mM NaCl, 10mM $MgCl_2$, 10mM Tris HCl, pH 7.5).

The cell suspension was lysed by adding NP-40 to a final concentration of 0.5% (v/v), followed by vortexing vigorously for 30 seconds.

Nuclei were pelleted by centrifugation for 3 minutes at 1,250 x g in an MSE "Chilspin" swing-out rotor. The supernatant (cytoplasmic fraction) was phenol extracted and ethanol precipitated as normal.

Nuclei were resuspended in 10mM Tris HCl, 10mM EDTA, pH 7.5 and lysed by addition of SDS to a final concentration of 0.6% (w/v). Low molecular weight DNA was recovered by Hirt fractionation (section M-8).

M-10 Enrichment for unintegrated virus DNA by velocity or density gradient fractionation

M-10.1 Isokinetic sucrose gradients

These were used for their resolution, which is superior to that of linear sucrose gradients (Noll, 1967). Isokinetic sucrose gradients are designed to provide constant velocity sedimentation. The linear increase in the driving force acting upon the particle with increasing distance from the centre of the rotor is exactly compensated by an equivalent increase in the opposing forces of viscous drag and buoyancy.

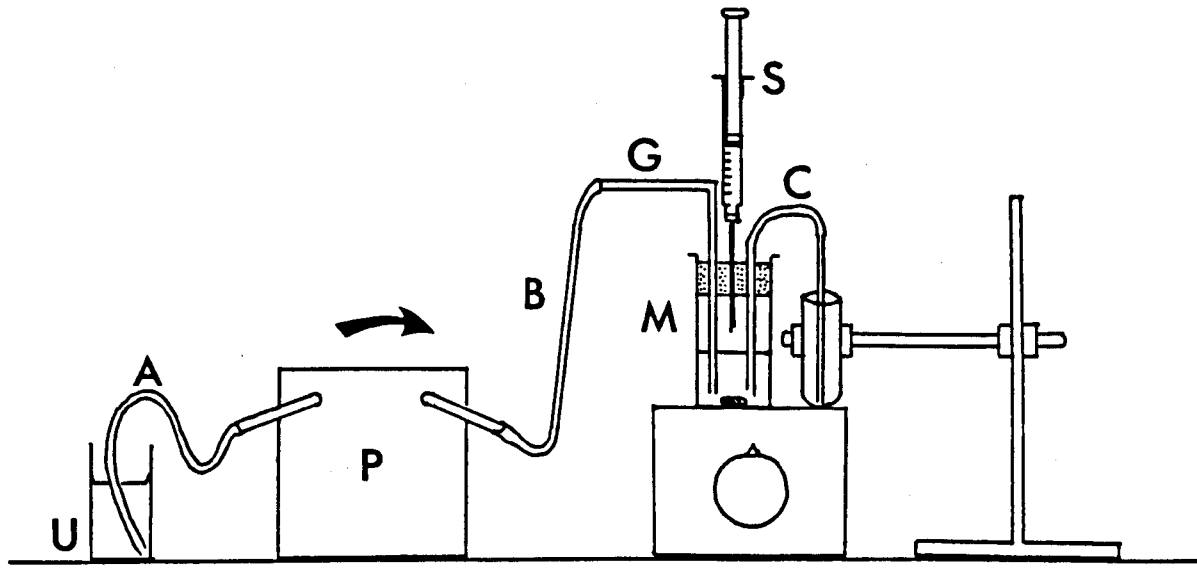


Figure M-1 Apparatus for constructing isokinetic sucrose gradients. See text for explanation. A, B, C; silicone rubber tubing; U, glass universal bottle (reservoir for heavy sucrose solution); P, peristaltic pump; M, mixing chamber (contains light sucrose solution and a magnetic follower), placed on a magnetic stirrer. The top assembly of M consists of a rubber bung interrupted by the glass inlet tube G, the syringe S, and the outlet tube C.

Specifications for each type of gradient have to be calculated using a computer approximation.

The modification of Noll's (1967) original method and specifications for the gradient using the MSE "Superspeed" 65 centrifuge and the 3 x 25ml swing-out rotor were kindly supplied by Dr. John Norton.

Figure M-1 accompanies the details concerning the establishment of a 10.4 - 21.8% (w/v) convex exponential sucrose gradient; the mixing chamber was made by John Norton. The method was as follows:

(i) Tubes A and B were flushed with 34.5% (w/v) sucrose solution.

(ii) 17.5ml of 34.5% (w/v) sucrose (heavy solution) was added to the glass Universal U, and 24.9 ml of 10.4% (w/v) sucrose (light solution) to the mixing chamber M.

(iii) The 25ml centrifuge tube was clamped gently to avoid its distortion, its bottom standing on a magnetic stirrer (as shown) to keep it level with M.

(iv) The top assembly of M was introduced (syringe plunger removed) pushing in the bung to achieve an air-tight seal.

(v) The delivery tube (A,B) was filled with heavy solution; the pump was stopped when this was 0.5 cm from the bottom of the glass inlet tube, G, in M.

N.B. This heavy solution represented "dead space" in the apparatus and was not taken from the reservoir in U.

(vi) Tube A was placed in the heavy solution in U.

(vii) Introducing the plunger to the syringe, slight pressure was applied to cause the outlet tube C to fill completely with light solution (and no further). The syringe setting was not altered, subsequently acting as a constant pressure device.

(viii) The magnetic stirrer was switched on, and the speed was adjusted to cause a slight distortion in the meniscus of the solution in M.

(ix) The pump was switched on and its speed adjusted to produce the gradient over a period of about 15 minutes.

(x) As U became empty, it was tilted to ensure all the sucrose was used. The pump was stopped immediately all the sucrose had entered tube A.

(xi) Tube C was pinched and drawn slowly and steadily out of the gradient.

(xii) The DNA sample in 1-2ml of light sucrose solution was layered on top of the gradient. This was centrifuged at 26,000-28,000 rpm at 5°C for 18 hours.

Gradients were unloaded from the top with a Densi-Flow probe (Buchler Instruments, UK) connected to a peristaltic pump. 1ml fractions were collected and stored at -20°C. For Southern blot analysis, 25µl aliquots were removed from each fraction, electrophoresed in agarose gels and transferred to nitrocellulose. Fractions containing the appropriate species of DNA, as detected by the cDNA probe (section M-18.1) were ethanol precipitated to recover this DNA.

M-10.2 CsCl-Ethidium bromide density gradients

Closed circular, supercoiled virus DNA was purified from low molecular weight nuclear DNA by equilibrium centrifugation in a 20ml CsCl-ethidium bromide density gradient (Giani *et al.*, 1975) using the modification of Anet & Strayer (1969) to enhance the separation of linear and supercoiled molecules. Supercoiled plasmid Col E1 DNA was added to the CsCl before centrifugation to act as a marker, since the picogramme quantities of virus DNA were not visible.

Low molecular weight DNA from infected cell nuclei, plus 20µg Col E1 DNA were made up to 2ml with TE and sheared twice by forcing through a 25 gauge hypodermic needle. This reduced the viscosity of contaminating chromosomal DNA, which could affect the resolution of chromosomal and circular DNA.

The sheared DNA was made up to 20ml TE containing 16g CsCl and ethidium bromide (final concentration, 300μg ml^{-1}). The refractive index before centrifugation was 1.3910. Gradients were formed by centrifugation in an 8 x 25ml angle titanium rotor on the MSE "Superspeed" 65 ultracentrifuge. A steep gradient was formed at first by centrifugation at high speed overnight (54,000 rpm). After 24 hours the gradient was "relaxed" by turning the speed down to 40,000 rpm for 64 hours. This had the effect of broadening the central part of the gradient, so improving the resolution of chromosomal and supercoiled circular DNA species (Anet & Strayer, 1969).

The resulting gradient was viewed under UV light, and 1ml fractions collected using an MSE tube piercer. As the gradient collected from the bottom of the tube, the fractions which contained plasmid marker DNA were noted.

Fractions were diluted by addition of 1ml TE and 10μg yeast RNA carrier was added to each. Ethidium bromide was removed from the DNA by three extractions with butanol which had been previously saturated with TE. Fractions were diluted to 8ml with TE (ie approximately 0.5M CsCl) and precipitated with 2.5 volumes of ethanol.

Pellets were dissolved in 100μl TE and 10μl aliquots were run in agarose gels then transferred to nitrocellulose. These filters were hybridized with ^{32}P-labelled cDNA prepared using virion RNA as template (see figure R-3).

Once it had been established that supercoiled plasmid DNA banded at the same position as supercoiled Ki-MuLV and VL30 DNA species, this region of the gradient was recovered by insertion of a hypodermic needle through the side of the centrifuge tube.

M-11 Gel electrophoresis

M-11.1 Agarose-formamide gels for RNA

RNA gel running buffer was 40mM Tris, 20mM sodium acetate, 2mM EDTA, 0.05% DEPC (v/v) adjusted to pH 7.0 with glacial acetic acid, then made 50% (v/v) with formamide.

Before use, formamide was deionized to pH 7.0 by stirring with 12g dm^{-3} of Amberlite MB-3 resin for 5 to 10 minutes. Once the pH had dropped to 7.0, the formamide was filtered through Whatman No. 1 paper using a Buchner funnel, to remove the resin.

1.5g agarose was added to 150ml of running buffer, and dissolved by boiling in a water bath. Horizontal slab gels were cast by pouring agarose, cooled to 40°C, into a tray formed by applying autoclave tape to the sides of the gel plate. A perspex well former was adjusted to stand vertically, approximately 4cm from one end of the tray. Gels were cast at 4°C and required at least 45 minutes to set. Using a 21cm by 14.5cm gel plate, the gels were approximately 6mm thick.

The well former was removed carefully to leave well bottoms intact and the gel placed in a horizontal gel tank. Gels were submerged with running buffer to a depth of at least 0.5cm.

RNA was dissolved in TE containing 0.5% (w/v) SDS. 10µl RNA was added to 25µl loading buffer and 15µl dye mix (total 50µl), denatured by heating to 65°C for 1 minute followed by cooling on ice, then loaded into the gel slots. Loading buffer was deionised formamide, pH 7.0, 0.1% (w/v) SDS, 0.1% (v/v) DEPC. Dye mix was 50% (v/v) loading buffer, 49.6% (v/v) glycerol, 0.2% (v/v) agarose beads, 0.1% (w/v) SDS, 0.1% DEPC, 0.02% (w/v) bromophenol blue.

Gels were electrophoresed for 12-16 houurs using a DC power pack set at 30mV constant voltage. After electrophoresis, the gel was stained by gentle shaking in TE containing ethidium bromide (1µg ml^{-1}) for 20-30 minutes. RNA was visualized by UV fluorescence.

M-11.2 Large scale agarose gels for DNA

These were cast and electrophoresed in TAE buffer (see Materials). The percentage agarose used for a gel was varied according to the sizes of the DNA species to be resolved, and ranged from 0.8% (w/v) for 1 to 5 kbp fragments, to 2.0% (w/v) for 0.1 to 0.5 kbp fragments. Two sizes of gel were routinely cast; a 21 x 14 cm (150ml) gel, and a 20 x 21 cm (225ml) gel. The larger gels were used for more precise Southern blotting experiments where it was important to avoid any "edge effect" in the electrophoresis. Unlike agarose:formamide gels, DNA gels could be cast and set at room temperature.

Restriction digests were stopped by addition of SDS to 0.5% (w/v). DNA was loaded in a total volume of 50µl. A typical loading mix was:- DNA 30µl; agarose beads (see M-11.2.1) 10-15µl; glycerol (80% v/v plus bromophenol blue) 5-10µl.

DNA gels were electrophoresed while submerged under 0.5cm of buffer at 30mV constant voltage for 12-16 hours. Staining was as for RNA gels.

M-11.2.1 Agarose beads

These were used when loading DNA or RNA samples into agarose gels, as they produced tighter bands upon electrophoresis.

0.02g agarose was boiled until molten in 10ml of 10% (v/v) glycerol, 10mM EDTA and 10mM Tris HCl pH 7.5, then stained dark blue by the addition of solid bromophenol blue. The mixture was poured into a plastic 10ml syringe and allowed to set at 4°C. A slurry of agarose beads was produced by forcing the gel through a 25 gauge needle several times. Beads were stored at 4°C.

M-11.3 Mini-agarose gels for DNA

These were found to be convenient for the rapid analysis of small scale cleared lysates, restriction enzyme digests and new DNA preparations. The mini gel electrophoresis system was supplied by Cambridge Biotechnology Laboratories, Cambridge, UK. Mini gels were 25

to 50 ml in volume and were electrophoresed in TBE (see Materials), according to the suppliers' recommendations.

M-11.4 Polyacrylamide gel electrophoresis for testing molecular linkers

Vertical gels containing 20% (w/v) acrylamide, 42% (w/v) urea were prepared as follows. The gel sandwich (20 x 20cm) was prepared using teflon side spacers (1.5mm deep) and silicone tubing. Gel polymerization was initiated by adding 24µl TEMED and 360µl ammonium persulphate (100mgml^{-1}) to 60ml of acrylamide/urea in running buffer, and mixing well. The gel was poured in a vertical position with a 10ml pipette and allowed to set overnight at room temperature. Wells were formed by a teflon gel comb.

The silicone tubing and gel comb were removed and the gel placed in a vertical electrophoresis tank. Running buffer (100mM Tris, 2.5mM EDTA, adjusted to pH 8.3 with boric acid) was added to the top and bottom tanks and the samples were loaded. Electrophoresis was performed for approximately 4 hours at 20mA constant current, at which time the bromophenol blue marker had migrated 12 to 14cm.

Gels were rinsed with water to remove crystallized urea and autoradiographed without drying.

Samples were loaded as a 20µl volume containing 10µl sample together with 10µl dye mix (6M urea, 50% (w/v) sucrose, 0.2% xylene cyanol FF, 0.2% bromophenol blue, 15mg ml^{-1} bacterial tRNA).

M-12 Endogenous reactions

The retrovirus endogenous reaction relies on the virus reverse transcriptase for synthesis of double-stranded DNA copies of encapsidated RNA. The reaction takes place within the virion which is partially lysed to allow the entry of deoxyribonucleoside triphosphates. Using optimum conditions large quantities of full-length linear proviral DNA can be obtained (see Benz & Dina, 1979; Gilboa *et al.*, 1979 and

Norton et al., 1982).

For Ki-MuLV the optimized conditions were as follows:

Virus was purified as described in section M-5 and resuspended in a small volume (approx. 400µl) of 50mM Tris HCl, pH 7.5. The endogenous reaction mix comprised: virus (4mg ml^{-1}); 50mM Tris HCl, pH 7.5; 2mM $MgCl_2$; 20mM DTT; 1mM each of the four dNTPs, and either TX-100 (0.02% v/v) or melittin (60µg ml^{-1}). The reaction was allowed to proceed for 12 hours at 37°C. Virus was lysed by the addition of SDS to 0.5% (w/v) and extracted with an equal volume of phenol:chloroform (1:1, v/v) saturated in TNE containing 0.1% (w/v) SDS. Nucleic acid recovered from the aqueous phase by ethanol precipitation was incubated with 75 units of SI nuclease per mg of virus in a total reaction volume of 500µl, containing 100mM sodium acetate, pH 4.5, 200mM NaCl and 5mM $ZnCl_2$. Digestion proceeded for 30 minutes at 37°C, and was stopped by the addition of 5 volumes of TNE. The products were phenol extracted and precipitated with ethanol prior to further purification by preparative gel electrophoresis (section M-13). This removed residual single-stranded nucleic acid intermediates.

M-13 Preparative gel electrophoresis

DNA was recovered from low gelling temperature agarose gels using the method described by Langridge et al. (1980).

(a) Preparation of reagents

150ml butan-1-ol and 150ml distilled water were equilibrated by shaking and allowed to settle out. 1g of hexadecyltrimethylammonium bromide (C-TAB) was dissolved in 100ml of the butanol fraction. 50µl antifoam A was added to 100ml of the equilibrated aqueous fraction, which was then shaken with the butanol:C-TAB. The solution was left overnight for the phases to separate, then each was separately stored at 37°C.

(b) Method

DNA was electrophoresed through a 1.0% (w/v) low gelling temperature agarose gel in TAE buffer. Buffer salts were removed during staining by gently shaking the gel for 1 hour in 1 litre TE containing ethidium bromide (1μg ml^{-1}). Low salt conditions were essential for the recovery procedure, which relies on the displacement by C-TAB of sodium ions bound to the DNA, thereby causing the DNA to partition into the organic phase (thus removing it from agarose which remains in the aqueous phase).

Using a long-wave UV light source (to reduce damage to the DNA) the fluorescent DNA bands were removed with a flamed scalpel and placed in a sterile plastic tube. Agarose was melted by incubating the tube at 65°C for 5 to 15 minutes. Once the agarose was completely molten, all procedures were performed at 37°C. Where necessary, carrier nucleic acid such as lambda DNA or tRNA was added at this stage.

The volume of molten agarose was estimated, and diluted with an equal volume of the C-TAB aqueous phase. A similar volume of C-TAB: butanol was added and the mixture was extracted by inverting gently for 2-5 minutes. Phases were separated by centrifugation in an Eppendorf microfuge and the butanol (top) phase was transferred to a fresh tube. The aqueous layer was extracted twice more with C-TAB:butanol, and all three butanol phases were pooled.

The butanol fraction was extracted with an equal volume of C-TAB: water and the organic phase recovered. The following stages were performed at room temperature.

DNA was recovered from the C-TAB:butanol by extracting twice with a one quarter-volume of 200mM NaCl. Residual C-TAB was precipitated from solution by extracting the two combined high salt fractions with chloroform for 10-30 minutes at 4°C.

Small quantities of DNA were recovered more efficiently in the

subsequent ethanol precipitation step if 100μg of pure glycogen were added as carrier. Recoveries of DNA were >50%, and the product was suitable for restriction enzyme digestion and ligation reactions.

M-14 Restriction endonuclease digestion of DNA

High molecular weight DNA was digested at a final concentration of 0.2mg ml^{-1} with a 4 to 5-fold excess of restriction enzyme, under the conditions recommended by the suppliers (see Materials for buffers). When DNA was cleaved with combinations of enzymes, digestions were performed sequentially with adjustment of reaction conditions appropriate for the second enzyme. 10μg quantities of DNA for Southern blotting experiments were digested for 4 to 6 hours with the addition of fresh enzyme every 1.5 to 2 hours. Preparative scale digestions were performed for longer periods with larger additions of enzyme. Reactions were stopped by adding SDS to a final concentration of 0.5% (w/v) and the digested DNA was electrophoresed through a horizontal agarose gel in either TAE or TBE buffer.

M-14.1 Preparation of molecular weight markers

Wild type bacteriophage lambda DNA was digested to completion with *Eco* RI or *Hind* III. The reactions were stopped with SDS as described above, and pooled. The concentration of DNA was adjusted to 67μg ml^{-1}, then undigested lambda DNA was added to a final concentration of 6.7μg ml^{-1}. The marker preparation was stored at $4^{o}C$. Fragment sizes of the *Eco* RI and *Hind* III digestion products were taken from Thomas & Davis (1975) and Wellauer *et al.* (1974), respectively. These were (in kbp):

DNA	47.00	
1	22.10	(H)
2	20.80	(E)
3	8.85	(H)
4	7.18	(E)
5	6.14	(H)
6	5.65	(E)
7	5.27	(E)
8	4.58	(E)
9	4.05	(H)
10	3.23	(E)
11	2.12	(H)
12	1.83	(H)
13	0.52	(H)

where H = Hind III fragment, E = Eco RI fragment.

The sizes of smaller restriction fragments were estimated by comparison of their mobility with those of fragments generated by Hinf I digestion of pBR322. Fragment sizes were derived from the complete nucleotide sequence of this plasmid, obtained by Sutcliffe (1979):

Fragment	Size (bp)
1	1631
2	517
3	506
4	396
5	344
6	298
7	221
8	220
9	154
10	75

M-15 Restriction mapping

Low resolution restriction maps were constructed by analyzing the fragments produced by simultaneous or sequential digestion of the DNA with pairs of restriction enzymes which were found to cleave the DNA infrequently (ie, enzymes that recognized hexanucleotide sequences). From the sizes of the fragments it was possible to deduce the relative locations of at least some of the cleavage sites. As the number of pairwise combinations of enzymes was increased, the number of defined sites on the map increased until, eventually, no ambiguities remained. Since the resolution of these maps depended on the accuracy of size estimation of fragments relative to those of the markers, it should be noted that most sites were not accurate to less than 50-100 bp. Accuracy was improved by taking averages based on several different determinations, from data obtained using a range of gels with different agarose concentrations.

M-16 Transfer of nucleic acid from agarose gels to nitrocellulose filters

M-16.1 Transfer of DNA to nitrocellulose filters (Southern blotting)

The method was based on that of Southern (1975).

Reagents used:-

Denaturing solution:	1.5M NaCl, 0.5M NaOH
Neutralizing solution:	3.0M NaCl, 0.5M Tris HCl, pH 7.0
20 x SSC:	3.0M NaCl, 0.3M tri-sodium citrate.

After electrophoresis, gels were stained as described above. The positions of molecular weight marker fragments were marked with indian ink using a 25 gauge hypodermic needle. Excess agarose was removed, care being taken to leave the bottoms of the wells intact.

DNA fragments were denatured *in situ* by gently shaking the gel in $1.5dm^3$ of denaturing solution for 1.5-2 hours, at room temperature. The

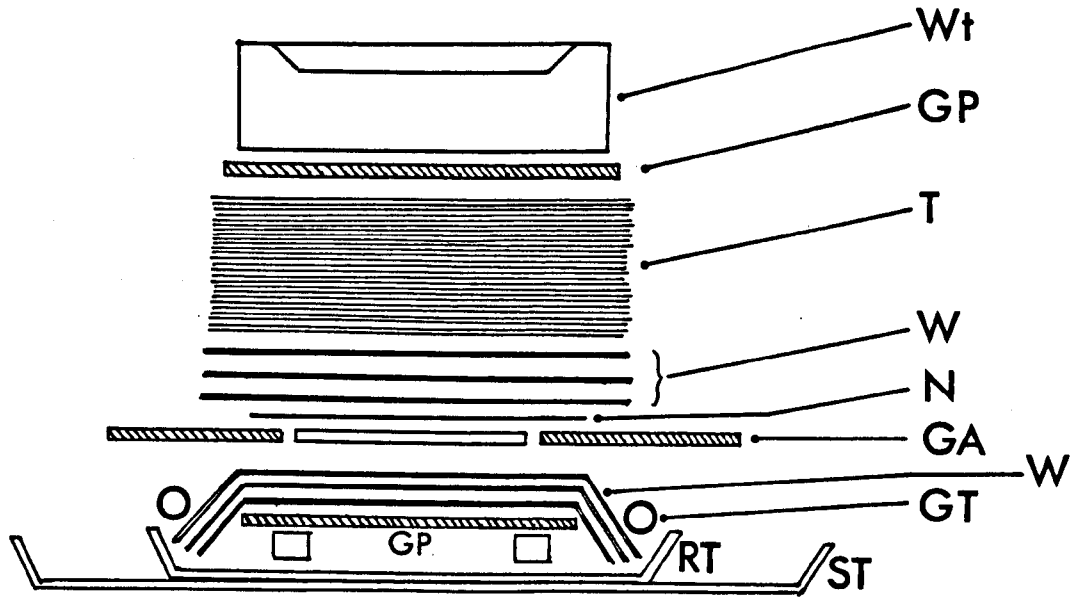


Figure M-2 Southern blotting apparatus. See text for explanation. ST, 40 x 30cm spill tray; RT, 25 x 30cm reservoir tray (filled with 20 x SSC); GT, glass tubes; W, Whatman 3MM paper; GA, agarose gel framed by glass plates; N, nitrocellulose; T, tissues; GP, glass plate; Wt, weight (eg house brick).

gel was then shaken in 1.5dm^3 of neutralizing solution for 1.5-2 hours.

Transfer to nitrocellulose was accomplished using a blotting system designed by John Norton, shown in figure M-2. Care was taken to avoid bubbles between the sheets of 3MM paper forming the wick, and between the gel and nitrocellulose filter. Glass plates were placed around each side of the gel, leaving a 0.5cm gap to avoid capillary suction, and the filter was placed on top of the gel with its edges resting on the plates. The filter was soaked in 2 x SSC prior to use, as were three sheets of 3MM paper which was carefully placed on top of the filter. Dry Kleenex tissues were used to draw 20 x SSC from the reservoir up through the gel and filter. Transfer of high molecular weight DNA, or DNA from high concentration (>2% w/v agarose) gels required up to 2 days blotting with replenishment of tissues and 20 x SSC after 24 hours to ensure complete transfer. Otherwise DNA transfer was performed overnight.

After transfer the nitrocellulose filter was carefully peeled off the blotting platform, leaving the gel still attached to the filter. The positions of wells and marker fragments (localized by transfer of the ink to the filter) were marked with a soft lead pencil. Excess filter was removed, and the gel discarded. Traces of salt were removed by shaking the filter in 500ml of 5 x SSC for 5 minutes, after which it was blotted dry with 3MM paper.

Having removed excess moisture the filter was placed between two sheets of 3MM paper and baked in a vacuum oven at 80°C for 2 hours.

M-16.2 Transfer of RNA to nitrocellulose filters ("northern blotting")

The migration of ribosomal RNA markers in agarose:formamide gels was recorded by taking a Polaroid photograph of the gel alongside a ruler.

After being shaken gently in 10% (v/v) formaldehyde in TE for 60 minutes, the gel was equilibrated in 20 x SSC for 90 minutes. Transfer

of RNA to nitrocellulose involved the same protocol as that for Southern blotting, except that after wetting in 2 x SSC the filter was soaked in 20 x SSC before placing on the gel. The three sheets of 3MM were also dipped in 20 x SSC, and the filter after RNA transfer had taken place was rinsed quickly in 10 x SSC. RNA was fixed on the nitrocellulose filter by baking for 2 hours at 80°C *in vacuo*.

M-17 Dot-blotting experiment (see R-5.3)

This procedure was modified from that of Itin *et al*. (1983).

High molecular weight DNA samples were diluted to a final concentration of 0.6, 2.0, 4.0 or 6.0μg ml^{-1} in TE. 2ml of each solution were denatured by mixing with an equal volume of 1M NaOH then allowing the solution to stand on ice for 10 minutes. DNA samples were prepared for binding by the addition of 2ml 1M Tris HCl, pH 8.0, 2ml 1M HCl and 4ml 3M NaCl. Four replicates of each sample were prepared by passing 3ml aliquots through a 13mm diameter nitrocellulose filter which had been pre-wetted in 2 x SSC. Each sample was passed through the filter three times.

After receiving the DNA the filters were dried at room temperature, then baked and prehybridized as for a Southern blot. Duplicates of each sample were hybridized with either nick-translated VL30 clone NVL-3 probe (specific activity 3.3 x 10^7cpm μg^{-1}) or with cloned Ki-MuLV DNA probe (specific activity 4.9 x 10^7cpm μg^{-1}) using 2 x 10^6cpm ml^{-1} of hybridization solution.

Following hybridization filters were washed for 3 hours in 2 x SSC at 60°C, then placed in 2 x SSC at room temperature for Cerenkov counting. This procedure was repeated using the same filters but at 0.5, 0.1, 0.03 and 0.01 x SSC at 60°C, each wash being followed by Cerenkov counting.

Results were calculated using corrections both for background

signal (two blank filters for each probe) and radioactive decay of the ^{32}P-labelled probes during the course of the experiment. Graphs were plotted of cpm probe hybridized versus amount of DNA loaded, using a linear regression formula;

$$m = \frac{N\Sigma x_1 y_1 - \Sigma x_1 \Sigma y_1}{N\Sigma x_1^2 - (\Sigma x_1)^2}$$

which was a pre-programmed function of a "Commodore" calculator. Regressions were calculated using 0,0 as the first value. Mouse DNA samples hybridized with VL30 probe gave a curve which reached a maximum between 2 and 3μg of DNA loaded; in this instance the regression was calculated for the first part of the curve.

Data obtained from these plots were used to construct a melting curve for each DNA sample. The 100% hybridization value was taken as the slope of each curve at the 2 x SSC (least stringent) wash. Gradients of curves at subsequent wash stringencies were expressed as a percentage of this maximum value, for each sample.

M-18 ^{32}P-labelled probes

M-18.1 cDNA probe

The method used was based on that of Shank *et al.* (1978), using exogenous reverse transcriptase with random primers.

Reagents

Random oligodeoxynucleotide primers were prepared by digestion of commercial calf thymus DNA with pancreatic DNAase I for 30 minutes at 37°C. The randomized DNA fragments were phenol extracted and denatured by boiling for 15 minutes. These were recovered by ethanol precipitation, dissolved and the solution adjusted to a concentration of approximately 1mg ml^{-1} (Taylor *et al.*, 1976).

Reaction mix

50mM Tris HCl (pH 8.3);

40mM KCl;

8mM $MgCl_2$;

0.5mM DTT;

0.2mM each of dGTP, dATP and TTP;

100µg ml^{-1} oligomers of calf thymus DNA;

5-20µg ml^{-1} RNA template;

500 units ml^{-1} AMV reverse transcriptase;

50-250µCi α ^{32}P dCTP (3000Ci $mmol^{-1}$).

Reaction volumes were made up to 50µl with sterile distilled water, and the mix was incubated at 37°C for 60 minutes, followed by the addition of: 50µl ss E. coli carrier DNA (350µg ml^{-1}), 250µl 2 x extraction buffer (1X = 400mM NaCl, 100mM Tris HCl, pH 7.5, 1mM EDTA, 0.2% (w/v) SDS), 150µl distilled water and 500µl phenol:chloroform (1:1, v/v). After vigorous extraction at 4°C the phases were separated by centrifugation and the aqueous phase was ethanol precipitated with 200µg of glycogen carrier, at -20°C.

Alkaline hydrolysis

The nucleic acid pellet was re-dissolved in 450µl H_2O, and the RNA template was hydrolysed by addition of NaOH to 0.3M, followed by incubation at 60°C for 50 minutes. Hydrolysis was stopped by adding HCl to neutralize the solution and the cDNA was ethanol-precipitated in the presence of 0.25M sodium acetate and glycogen carrier.

The cDNA was recovered by centrifugation, dissolved in 200µl TE and applied to a 12ml G-100 column in Sephadex column buffer (see Materials). 1ml fractions were collected, and those containing the first peak were pooled, NaCl added to 150mM, and ethanol-precipitated with glycogen carrier.

The final cDNA pellet, which had a specific activity of

5.5×10^9cpm μg^{-1}, was dissolved in 400µl TE and stored at -20°C until use.

M-18.2 Nick-translated probe

This method was modified from those of Maniatis *et al.* (1975) and Rigby *et al.* (1977).

Reaction mix

5µl 10 X nick translation buffer;

20µM each of the 3 unlabelled dNTPs (5µl);

0.5-1.0µg DNA;

20ng ml^{-1} DNAase I;

10 units *E. coli* DNA polymerase I;

125-250µCi $\alpha^{32}P$ dCTP (3000Ci $mmol^{-1}$);

H_2O to 50µl.

The mix was incubated at 15°C for 2 hours. ^{32}P-labelled DNA products were extracted with phenol:chloroform (1:1, v/v) and unincorporated label was removed by G-100 chromatography, as described for the cDNA probe preparation. DNA recovered by ethanol precipitation was dissolved in 200µl TE and a 2µl aliquot was counted in scintillant. Specific activity ranged from 3×10^7 to 2×10^8cpm μg^{-1}.

M-19 Measurement of DNA-associated radioactivity

This was achieved by adsorption of labelled DNA to Whatman DE-81 paper. Duplicate 5-30µl aliquots were spotted onto a 3 x 2cm rectangle of DE-81 paper. While still damp, one filter was washed five times, 5 minutes per wash in 0.5M Na_2HPO_4; twice in water (1 minute per wash) and twice in 95% ethanol (30 seconds per wash).

Both filters were dried under a heat lamp on aluminium foil. Filters were placed in Beckman Ready-solv EP aqueous scintillation fluid and their associated radioactivity determined by scintillation counting. The unwashed filter measured the total radioactivity in the sample,

while the washed filter measured only radioactivity which was incorporated into the nucleic acid.

M-20 Filter hybridization

Prehybridization solution was:

50% (v/v) formamide;

3 x SSC;

10μg ml^{-1} sheared salmon sperm DNA;

500μg ml^{-1} yeast RNA;

50mM HEPES pH 7.0;

5 x Denhardt's solution: 0.1% (w/v) BSA; 0.1% (w/v) Ficoll 400; 0.1% (w/v) Polyvinylpyrrolidone-40.

Hybridization solution was the same as prehybridization solution, except that 1 x Denhardt's (1966) solution was used, and dextran sulphate was added to a final concentration of 10% (w/v) (Wahl *et al.*, 1979). Both solutions were stored at 4°C and warmed for 30 minutes at 42°C before use.

Each filter was wetted in 3 x SSC, then immersed in 10ml prehybridization solution and sealed into a polythene bag. The bag was "massaged" to ensure even distribution of the solution, then incubated in a 42°C water bath between two glass plates, for 12-16 hours.

^{32}P-labelled probe was denatured by boiling for 3 minutes followed by rapid transfer to a dry ice:ethanol bath. Probe was used to give 1 x 10^6 to 3 x 10^6cpm per filter, and was mixed with hybridization solution before this was added to the filter in a new polythene bag. Filters were hybridized in 1-6ml, according to their dimensions. Hybridization was for 70 hours at 42°C.

M-20.1 Post-hybridization washing

This was adjusted to give the required stringency. For moderately

high stringency washes (as were normally performed) the filter was rinsed four times in 2 x SSC at room temperature with shaking, 1-2 minutes each rinse. The filter was subjected to another four washes in 2 x SSC at room temperature, for 15 minutes per wash, followed by two low salt washes of 15 minutes each in 0.1 x SSC at 55°C-60°C. Finally, after two brief rinses in 0.1 x SSC at room temperature, the filter was blotted dry on Whatman 3MM paper. Before preparing for autoradiography, the filter was dried thoroughly at 37°C for 1 hour.

Low stringency washes were essentially similar to the high stringency wash, except that all washes were performed at room temperature. Least stringent washes were performed using 2 x SSC at room temperature.

M-21 Autoradiography

Sensitivity of X-ray film was increased by the use of intensifying screens (Swanstrom & Shank, 1978). The most sensitive combination was found to be a Dupont Cronex Lightning-Plus calcium-tungstate-phosphor screen, with Fuji RX X-ray film (Fuji Photo Film Co. Ltd, Japan).

The hybridized nitrocellulose filter was fastened to one side of an X-ray film cassette. One sheet of X-ray film was placed next to the filter, and an intensifying screen was placed on top. The cassette was placed at -70°C for 1-14 days.

Events recorded by the X-ray film when exposed to ^{32}P β particles are long-wavelength photons which result from the fluorescence that occurs when the β-particle strikes the screen (Laskey & Mills, 1977). Since the response of film to low intensities of light is non-linear, exposure was carried out at -70°C in order to prolong the period of fluorescence. Pre-exposure of the film to red light (flashing) was not carried out as sensitivity of detection of ^{32}P is not increased by this procedure when using calcium-tungstate screens at -70°C (Laskey & Mills,

1977).

Films were developed for 1-5 minutes in Kodak DX-80 diluted 1 + 3 with water, using a red safety lamp. After a brief rinse in tap-water, films were placed in Kodak FX-40 fixer (diluted 1 + 3 with water) for 5-10 minutes. Films were washed in running tap-water for 20-30 minutes before being dried at 37°C.

M-22 VL30 copy number in mouse genomic DNA

A 10μg quantity of high molecular weight mouse DNA was digested to completion with a restriction enzyme and run on an agarose gel. In a parallel track was loaded 50-300pg of cloned VL30 DNA restriction fragments, the particular digest being varied according to the size of the genomic DNA fragment under study. The DNA was transferred to nitrocellulose and the filter hybridized with ^{32}P-labelled NVL-3 DNA. Copy number of a genomic DNA VL30 restriction fragment was calculated using a comparison of its autoradiographic signal intensity with that of a similar-sized cloned marker fragment. Knowing the mass of VL30 DNA represented by the signal intensity on the autoradiograph, the mass of the mouse DNA fragment could be calculated and used in the following formula:

$$\text{Copies/haploid genome} = \frac{\text{mass of DNA in band (pg)}}{\text{mass DNA digested (pg)}} \times \frac{\text{mouse genome size (bp)}}{\text{fragment size (bp)}}$$

The size of the mouse haploid genome was taken to be 3×10^9bp. Estimations of signal intensity were carried out using a Joyce-Loebl microdensitometer by calculating the area under the peak corresponding to the band under study.

When high molecular weight DNA is digested with a restriction enzyme and electrophoresed through an agarose gel, the viscosity of the larger DNA fragments can retard the migration of smaller fragments.

Thus cloned DNA marker fragments of relatively small size which are electrophoresed in parallel with these digests will not migrate at the same rate as their genomic DNA counterparts. To compensate for this, low molecular weight DNA was always added to an appropriate amount of Eco RI-digested high molecular weight calf thymus DNA. The latter bore no detectable sequence homology to the cloned VL30 probe.

Calculations of VL30 copy number were based on the mean values obtained from several different experiments, using different restriction enzyme digests and different strengths of agarose gels. Even so, due to the variability of such Southern blotting data for a number of technical reasons, copy numbers of VL30 sub-sets which are mentioned in the text are tentative, and are given to emphasize important trends in VL30 organization, rather than to be taken as absolute values.

M-23 Molecular cloning

M-23.1 Phosphorylation of linkers

Commercially available flush-ended synthetic molecular linkers lack 5' terminal phosphate groups which are necessary for ligation to cDNA. The method used to phosphorylate linkers was adapted from Maniatis et al. (1978).

2 OD_{260} units of Bam HI linkers were dissolved in 100µl 0.1 x TE to give a concentration of 1µg μl^{-1}.

To 50µCi $\alpha\,^{32}P$ ATP (3000Ci $mmol^{-1}$; 16.7pmoles) which had been dried down, were added:

500ng Bam HI linkers	0.5µl;
10 x Kinase buffer (+ 2-mercaptoethanol)	1.0µl;
H_2O	7.5µl
Polynucleotide kinase (5 units)	1.0µl,

and the mixture was incubated at 37°C for 30 minutes, followed by the addition of:

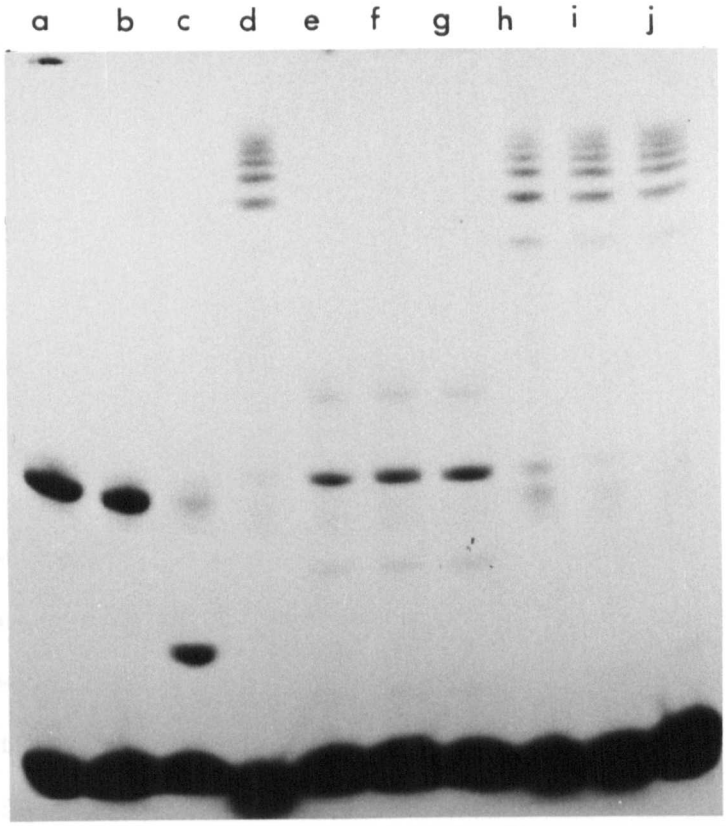


Figure M-3 Polyacrylamide gel electrophoresis of ^{32}P-labelled molecular linkers. 20µl samples of ^{32}P-labelled Bam HI linkers, specific activity 3.4 x 10^5cpmµg^{-1}, were loaded on to a vertical 20 x 20cm, 20% (w/v) polyacrylamide, 42% (w/v) urea gel and electrophoresed for 4h at 20mA constant current. The gel was rinsed with water and autoradiographed without drying. Each lane contained 10ng of linkers; a, unincubated; b, after overnight (19h) incubation at room temperature (approx. 20°C); c, incubated with 5 units Bam HI, 2.5h, 37°C; d, incubated with 0.5 units T4 DNA ligase, 19h, 20°C; e-g as d, followed by incubation with 5, 10 and 20 units, respectively, Bam HI, 2.5h, 37°C; h-j, incubated with 0.5, 2.0 and 4.0 units, respectively, T4 DNA ligase, 19h, 20°C.

10mM unlabelled ATP	1µl;
10 x Kinase buffer	1µl;
H_2O	7µl;
Polynucleotide kinase (5 units)	1µl.

The 20µl mixture was incubated for 30 minutes at 37°C, when the reaction was terminated by the addition of 80µl 10mM Tris HCl, 20mM NaCl, 1mM EDTA, pH 8.0. Specific activity of the linker preparation was calculated as described above (section M-19).

The reaction mixture was extracted with phenol:chloroform, and after phase separation by centrifugation the organic phase was re-extracted with 25µl of the stopping buffer. The pooled aqueous phases were adjusted to 200mM with NaCl, and ethanol precipitated in the presence of 30µg tRNA carrier, at -20°C overnight.

Linkers were pelleted by centrifugation in an Eppendorf centrifuge for 30 minutes at 4°C. The pellet was washed briefly in 70% aqueous ethanol, centrifuged once more and drained carefully. After drying *in vacuo* the preparation was dissolved in 50µl 0.1 x TE, and 2 x 1µl aliquots were DE-81 counted to assess recovery.

M-23.1.1 Testing kinased linker preparations

New batches of ^{32}P-labelled molecular linkers were routinely tested to assess their integrity and ability to be efficiently ligated and be digested with the appropriate enzyme. The products of these reactions were checked by autoradiography, following polyacrylamide gel electrophoresis (section M-11.4). An example of such a routine analysis is shown in figure M-3.

M-23.2 Ligating ^{32}P-labelled linkers to VL30 cDNA

The Km for the activity of T4 DNA ligase on blunt-ended DNA is nearly 100 times higher than its Km on DNA with cohesive ends (Maniatis *et al.*, 1982). Therefore ligation of blunt-ended linker to blunt-ended VL30 DNA requires a high concentration of enzyme and a high

concentration of DNA ends. Because the synthetic linkers are very small (10bp), the ligation reaction is driven by a high concentration of the linkers (in terms of DNA ends).

Since the *in vitro* DNA products of the endogenous reactions were purified by digestion with S1 nuclease, it was assumed that VL30 DNA recovered from agarose gels was flush-ended. The yield of VL30 DNA was estimated by fluorescence under UV light in a stained agarose gel; recovery from the gel was assumed to be 50%.

Ligation was carried out at room temperature for 20 hours in a total volume of 5µl:

50ng VL30 in 0.1 x TE	2.5µl;
5 x ligase buffer	1.0µl;
Bam HI linker (25ng)	1.0µl;
T4 DNA ligase (2 units μl^{-1})	0.5µl.

The reaction mix was then heated to 65°C for 3 minutes to inactivate the ligase.

M-23.3 Cleaving linkers to generate cohesive ends

Since the molar amount of the linkers is so large (25ng = 3.8pmoles) a large amount of restriction enzyme is required to achieve a complete digest. This was added at 1 hour intervals to compensate for any premature loss of enzymic activity that might have occurred. The following reactants were mixed:

Overnight ligation reaction mix	5µl;
H_2O	37µl;
10 x *Bam* HI endonuclease buffer	5µl;
5 + 5 + 5 units *Bam* HI	1 + 1 + 1µl,

and incubated at 37°C for 3 hours. 20µg tRNA carrier was added and the mix was extracted with phenol:chloroform (as in section M-23.1). Pooled aqueous phases were precipitated with ethanol in the presence of 50µg glycogen carrier, at -20°C overnight. Nucleic acid was pelleted by

centrifugation at 10,000 x g for 1 hour at -10°C, dried under vacuum and dissolved in 20µl 0.1 x TE.

M-23.4 Sephadex G-200 (Superfine) column chromatography

This was used to remove linker fragments, generated by digestion of cDNA with Bam HI (section M-23.3), from cDNA bearing cohesive Bam HI termini.

Sephadex G-200 SF in TE was swollen by incubation at 90°C for 5 hours, as recommended by the suppliers. 150ml of slurry was de-gassed in vacuo for 15 minutes and used to pack a 12ml column. The column was washed with 20ml TE and eventually a flow rate of 1ml per 9 minutes was achieved.

The products of the Bam HI digestion reaction (section M-23.3) were applied to the moist column surface in 40µl of 0.1 x TE. 20 fractions (0.5ml) were taken and each counted for 10 minutes to assess Cerenkov radioactivity. Figure M-4 shows a typical G-200 SF column profile.

Fractions corresponding to the first peak, as indicated, were pooled and NaCl was added to a final concentration of 200 mM. 30µg tRNA and 50µg glycogen carrier were added, together with 150ng of Bam HI-linearized, CIAP-treated pAT153 plasmid vector (see section M-23.5). The DNA was recovered by ethanol precipitation at -20°C overnight, followed by centrifugation. The pellet was washed in 70% (v/v) aqueous ethanol, re-centrifuged and the pellet drained and dried under vacuum. Recovery of DNA was assessed by Cerenkov counting.

M-23.5 Phosphatase treatment of plasmid vector DNA

DNA ligase is capable of joining two adjacent nucleotides only if one contains a 5'-phosphate group and the other a 3'-hydroxyl group. Re-circularization of plasmid vector DNA was therefore prevented by removal of the 5'-phosphate groups from each end of the linear DNA. During ligation of insert DNA with phosphatased vector DNA, only recombinant circles should result.

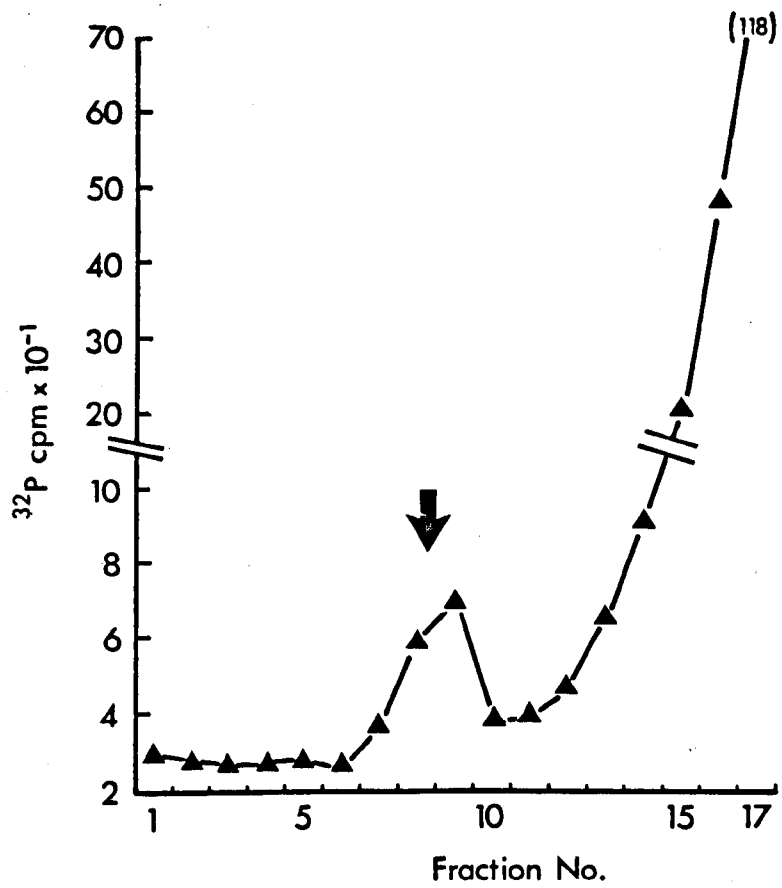


Figure M-4 Sephadex column chromatography of molecular linker DNA. Approx. 50ng VL30 cDNA was incubated with 25ng Bam HI molecular linkers (specific activity 3.4 x 10^5 cpmμg^{-1}) and 1 unit T4 DNA ligase, 20h, 20°C. "Linkered" VL30 DNA and free linkers were digested with 15 units Bam HI, 3h, 37°C, deproteinized by extraction with phenol:chloroform (1:1, v/v), precipitated by ethanol, re-dissolved in 20μl 0.1 x TE, and loaded onto a 12ml Sephadex G-200 superfine column. The figure indicates the position of VL30 DNA (arrowed) within the column profile, detected by measurement of Cerenkov radioactivity in the 0.5ml column fractions. The second peak indicates the position of linker fragments and unincorporated γ^{32}P-ATP.

1µg Bam HI-linearized pAT153 was dissolved in 50µl 10mM Tris HCl, pH 8.0, and incubated for 60 minutes with 0.5 units of calf intestinal alkaline phosphatase. The mix was adjusted to 200mM with NaCl, and extracted twice with phenol:chloroform. DNA recovered by ethanol precipitation was added to the Sephadex G-200 SF column fractions of VL30 cDNA prior to ethanol precipitation. This reduced the final volume of the subsequent ligation reaction.

M-23.6 Ligation of VL30 cDNA into plasmid vector

VL30 DNA was co-precipitated with CIAP-treated vector DNA as described in section M-23.4, and its recovery assessed by Cerenkov counting. The pellet was dissolved in 4.5µl 1 x ligase buffer and incubated for 24 hours at room temperature with 1 unit T4 DNA ligase.

M-23.7 Transformation of E. coli cells with recombinant DNA

This procedure is based on the observation of Mandel & Higa (1970) that uptake of bacteriophage lambda DNA is enhanced by treatment of bacterial cells with calcium chloride. The method used was slightly different for E. coli HB101 or MC1060 cells.

M-23.7.1 Using HB101 cells

This "weakened" strain is restriction deficient and recombination deficient. It is an E. coli K-12 x E. coli B hybrid, and is commonly used in transformation and large-scale growth of plasmids (Boyer & Roulland-Dussoix, 1969).

HB101 cells were sub-cultured three times before use. On the day of transformation, 2ml of overnight culture was inoculated into 50ml of L-broth and shaken vigorously at 37°C. Growth was followed by measuring absorbance using a spectrophotometer. When cells reached OD_{590} 0.6 (2×10^8 cells ml^{-1}), two 10ml aliquots were centrifuged at 4°C in an MSE "Minor" bench centrifuge. Cells were resuspended in two 5ml volumes of ice-cold 50mM $CaCl_2$ and incubated on ice for 30 minutes.

The suspension was pelleted by centrifugation as before, and the

cells resuspended in a final volume of 2ml 50mM $CaCl_2$.

Meanwhile, the recombinant plasmid DNA was dissolved in 50µl TCM buffer (see Materials) and 20µg tRNA carrier added. It was found that this addition increased the transformation efficiency (John Norton, unpublished observation).

100µl of competent cell suspension was mixed with each TCM solution, and the tubes incubated on ice for 60 minutes with occasional agitation. Each tube was then heat-pulsed for 3 minutes at 42°C, and 850µl L-broth added. Tubes were incubated at 37°C for 1 hour, then 50µl-300µl aliquots were spread onto well-dried plates of L agar containing ampicillin (100µg ml^{-1}). Plates were incubated in an inverted position overnight at 37°C.

Controls

0.1ng, 1.0ng or 10.0ng of closed circular, supercoiled pAT153 DNA were dissolved in 50µl TCM, and 20µg tRNA carrier added. Each solution was used to transform the recipient cells as described above. The next day, the number of transformants per µg of plasmid DNA indicated the transformation efficiency. With HB101 this varied from 10^5-10^6 colonies μg^{-1}.

Tubes were also set up containing (a) 50µl TCM, 20µg tRNA but no plasmid DNA, and (b) 50µl TCM, 20µg tRNA, 100µl 50mM $CaCl_2$. Tube (a) contents were used to transform bacteria as described above; this control was designed to indicate that the bacteria had not already been transformed with a plasmid. Tube (b) received no bacteria, but otherwise its contents were treated identically to those of the other tubes mentioned above. This was a sterility check for the solutions used.

M-23.7.2 Using MC1060 cells

These are *E. coli* cells which are *recA*$^+$ and which transform at a much higher efficiency than HB101 (eg 5 x 10^5 - 5 x 10^6 colonies μg^{-1} of

supercoiled circular plasmid DNA).

MC1060 cells were grown for 3 successive overnight sub-cultures. On the day of transformation, 0.5ml of overnight culture was inoculated into 50ml L-broth in a 500ml flask. This was shaken vigorously at 37°C. The culture reached an OD_{590} 0.6 after 2-2.5 hours, when two 10ml aliquots were pelleted by low speed centrifugation at 4°C. Pellets were resuspended in two 5ml volumes of ice-cold, 100mM $MgCl_2$ and once more centrifuged. Pellets were resuspended and pooled in 1ml of ice-cold 100mM $CaCl_2$, then were incubated on ice for 30 minutes. 100µl of cell suspension was then added to each 50µl TCM mix, as with the method for HB101 cells. After a further incubation on ice for 40 minutes, each tube was heat-pulsed at 42°C for 2 minutes, then placed on ice for a further 20 minutes. L-broth was added to a final volume of 1ml and the tubes were incubated at 37°C for 60 minutes. Samples were plated (with appropriate controls) as described for the HB101 method.

M-23.8 Screening transformants

Ampicillin-resistant transformant colonies can contain (a) plasmid which escaped phosphatase treatment and has re-circularized during the ligation reaction, (b) plasmid molecules containing an insert of linker fragment, another piece of plasmid DNA or a portion of bacterial genomic DNA, or (c) plasmid molecules which have received the insert of interest to the investigator.

Three methods were used to detect VL30 molecular clones. These were (1) antibiotic screening, (2) Grunstein-Hogness colony hybridization and (3) agarose gel electrophoresis of small-scale preparations of plasmid DNA ("cleared lysates").

M-23.8.1 Antibiotic screening

VL30 DNA was inserted into pAT153 at the unique Bam HI site. This restriction site is situated within the gene coding for tetracycline

resistance. Consequently, an insert at the Bam HI site will remove tetracycline resistance but leave the transformants resistant to ampicillin, the other antibiotic resistance gene carried on pAT153.

Transformants were therefore plated first on ampicillin-containing agar plates as described above. Colonies obtained after overnight incubation at 37°C were picked from these plates using sterile wooden toothpicks, and stabbed onto a master ampicillin-containing plate in a grid pattern. Each toothpick was then stabbed into the appropriate position in an identical grid pattern on a tetracycline-containing plate (20μg ml^{-1}). Both sets of plates were incubated at 37°C overnight and scored for amp^R, tet^S transformants. Large (>100) numbers of tet^S colonies were further examined by colony hybridization, whilst smaller numbers were screened using a mini-cleared lysate procedure.

M-23.8.2 In situ hybridization of bacterial colonies

This method of Grunstein & Hogness (1975) was used as described by Maniatis et al. (1982). Bacteria were transferred with a wooden toothpick from a master ampicillin plate to a nitrocellulose filter. The resulting colonies were lysed and the liberated DNA was fixed to the filter by baking. A colony whose DNA hybridized with the ^{32}P-labelled probe could then be recovered from the master plate.

8cm diameter nitrocellulose filters (Millipore HAWP, 0.45μM) were sterilized by irradiation with UV light for 2 minutes per side. Using gloved hands these were placed onto agar plates containing ampicillin (100μg ml^{-1}).

Using sterile wooden toothpicks, transformant colonies were transferred to the filter, then onto a master plate containing ampicillin but no filter. Small dots were arranged in a grid pattern, 100 colonies per filter. A colony containing pAT153 alone was included on each filter as a negative hybridization control. Plates were inverted and incubated at 37°C overnight, when colonies reached 3-4mm in

diameter.

The master plates were sealed with parafilm and stored at 4°C in an inverted position until the results of hybridization were available. Each nitrocellulose filter was carefully peeled from its plate and placed, colony side up, on Whatman 3MM paper impregnated with 10% (w/v) SDS, for 3 minutes. This treatment gave a sharper hybridization signal, probably by limiting the diffusion of plasmid DNA in the following stages.

Filters were transferred to a second sheet of 3MM paper that had been saturated with denaturing solution (0.5M NaOH, 1.5M NaCl) and left for 5 minutes. Next they were placed on a third sheet of 3MM paper that was saturated with neutralizing solution (1.5M NaCl, 0.5M Tris HCl, pH 8.0) and left for 5 minutes. Filters were then laid, colony side up, on dry 3MM paper and allowed to dry at room temperature for 60 minutes. Dry filters were sandwiched between two sheets of 3MM paper, and baked for 2 hours at 80°C *in vacuo*.

Baked filters were hybridized with a ^{32}P-labelled probe, as described in section M-20. Any signal on the subsequent autoradiograph which was significantly higher than that attributable to the pAT153 control, was further investigated by a cleared lysate analysis of the appropriate master plate colony.

M-23.8.3 Preparation of plasmid DNA

Two methods were used. The first was a rapid procedure for screening several bacterial colonies at the same time, and was performed on a small scale to produce enough DNA for 1-2 restriction enzyme digestions. The second was a gentler method for the preparation for larger quantities of recombinant plasmid DNA.

(a) Rapid method for mini-cleared lysates

Based on the alkali lysis method of Birnboim & Doly (1979), this procedure relied on the fact that covalently closed circular plasmid DNA

strands cannot be separated when their hydrogen bonds are broken by mild alkali. When returned to neutral pH they regain their native configuration, whilst E. coli DNA remains denatured.

A single colony was inoculated into 5ml of L-broth containing ampicillin (100μg ml^{-1}) and incubated at 37°C overnight. (Amplification by chloramphenicol was not necessary for this procedure). 1.5ml of the overnight culture was centrifuged for 1 minute in an Eppendorf centrifuge, the remainder of the culture being stored at 4°C. After draining the bacterial pellet, this was vortexed in 100μl of 50mM glucose, 10mM EDTA and 25mM Tris HCl (pH 8.0) to which solid lysozyme had been added to a final concentration of 4mg ml^{-1}.

The tube containing the mixture was incubated at room temperature for 5 minutes, then 200μl of freshly prepared, ice-cold 0.2M NaOH, 1.0% (w/v) SDS was added and mixed in by inverting the tube 3 times. Following a 5 minute incubation on ice, 150μl of ice-cold potassium acetate (pH 4.8) was added. (This was prepared by adding 11.5ml of glacial acetic acid to 60ml of 5M potassium acetate, and making up to 100ml with distilled water. The resulting solution was 3M with respect to potassium and 5M with respect to acetate). The tube's contents were vortexed for 10 seconds, then placed on ice for 5 minutes.

Bacterial genomic DNA and cell debris were pelleted by centrifugation in an Eppendorf centrifuge for 5 minutes at 4°C. The supernatant was transferred to a fresh tube and extracted once with phenol:chloroform, then once with chloroform:isoamylalcohol (24:1, v/v). The aqueous phase was transferred to a new tube and the DNA precipitated with two volumes of ethanol at room temperature for 2 minutes. DNA was recovered by centrifugation in an Eppendorf centrifuge for 5 minutes at room temperature. After a wash in 70% aqueous ethanol, the pellet was dessicated in vacuo and dissolved in 50μl TE. 10μl aliquots were suitable for restriction enzyme analysis as described in section M-14.

1µl of a 10mg ml^{-1} (boiled) solution of RNAase A was added to each digest, the products of which were analysed using a mini-gel system (section M-11.3).

(b) Large scale cleared lysates

10ml of an overnight culture was inoculated into 1 litre of pre-warmed L-broth containing ampicillin (100µg ml^{-1}). The culture was split into two 500ml volumes, each of which were grown with vigorous shaking at 37°C in $2dm^3$ flasks. Chloramphenicol was added to a final concentration of 200µg ml^{-1} when the cells had reached a density corresponding to an OD_{590} of 0.6-0.7. The cultures were incubated as before for a further 14-16 hours.

Next day, cells were pelleted by centrifugation for 10 minutes at 10,000 rpm, 4°C in an MSE "18" 6 x 300ml angle rotor. Cells were resuspended in ice-cold TE (pH 8.0) and pelleted once more. All pellets were resuspended and pooled in 8ml of 25% (w/v) sucrose, 10mM Tris HCl, pH 8.0 in a glass conical flask. Following the addition of 48mg lysozyme dissolved in 3.2ml 0.25M Tris HCl, pH 8.0, the suspension was incubated on ice for 5 minutes. 10ml of 0.25M EDTA was added prior to a further incubation on ice, for 10 minutes.

Cells were lysed by the addition of 24ml 0.5% (v/v) TX-100; the detergent was carefully mixed into the solution by gentle swirling of the flask. After a 20 minute incubation on ice, the lysate was heat-pulsed at 40°C for 2-5 minutes, then centrifuged at 34,000 x g for 45 minutes in an 8 x 50ml angle rotor, at 4°C. The supernatant was adjusted with NaCl to a final concentration of 200mM, phenol:chloroform extracted twice, and the DNA recovered by precipitation under 2 volumes of ethanol at -20°C, overnight.

After low speed centrifugation the nucleic acid pellet was washed with 70% aqueous ethanol, dried *in vacuo* and dissolved in TE.

Much of the RNA was removed by addition of LiCl to a final

concentration of 2M, and incubating on ice for 2 hours. The RNA precipitate was pelleted by centrifugation at 2000 x g for 30 minutes at $4^{o}C$. The DNA in the supernatant was de-salted by G-100 column chromatography, concentrated by ethanol precipitation and purified by preparative agarose gel electrophoresis. Alternatively, closed circular plasmid DNA was purified by centrifugation to equilibrium in a caesium chloride - ethidium bromide density gradient, as described in section M-10.2.

Yields of up to 1mg of supercoiled plasmid DNA per litre of culture were routinely obtained.

RESULTS

RESULTS

R-1 Unintegrated VL30 DNA in virus-infected cells

R-1.1 Introduction

As described in the General Introduction, retroviral DNA synthesis is initiated in the cytoplasm of the newly infected cell, and results in the formation of a linear double-stranded DNA copy of the viral RNA template. Redundant sequences present at each end of the viral DNA, the LTRs, are a consequence of the two transfers of template which occur during reverse transcription.

The linear viral DNA migrates from the cytoplasm to the nucleus (Shank & Varmus, 1978) where it is the precursor for two forms of circular viral DNA. The smaller of the two circles contains a single copy of the LTR and the larger, two copies of the LTR (Shank *et al.*, 1978a,b; Hsu *et al.*, 1978). The fact that circles are found predominantly in the cell nucleus has implicated these structures in the process of integration (Shank & Varmus, 1978; Norton *et al.*, 1982; also see Panganiban & Temin, 1984). The experiments described in this section were designed to investigate (a) whether full-length copies of VL30 DNA could also be detected in retrovirus-infected cells, (b) the structure of this reverse-transcribed DNA, and (c) the existence of useful restriction enzyme sites in VL30 DNA for molecular cloning experiments.

R-1.2 Identification of Ki-MuLV and VL30 DNA in infected cells

Sub-confluent NIH-3T3 cells were infected with A1 cell tissue culture fluid (TCF) containing Ki-MuLV (VL30) particles at approximately 5×10^6 pfu per ml. The cells were incubated at 37 C for 18-20 hours, then harvested and fractionated into nucleus and cytoplasm. Low molecular weight DNA was extracted from each fraction as described in Methods, sections M-8 and M-9.

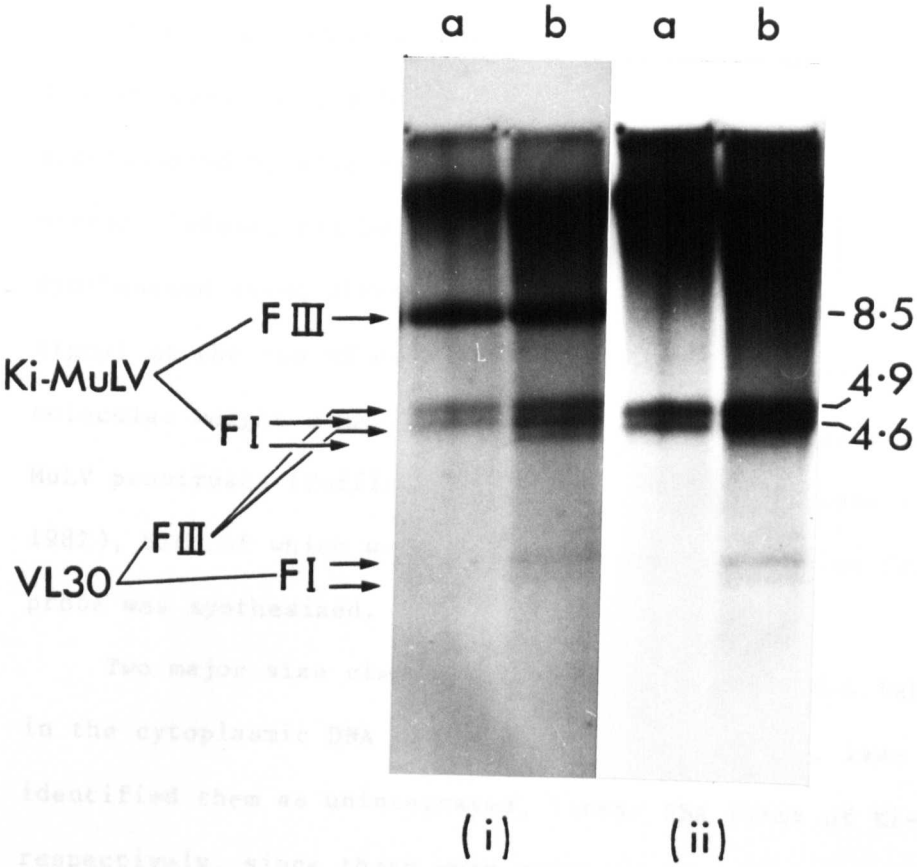


Figure R-1 Unintegrated forms of Ki-MuLV and VL30 DNA. Low mol.wt. DNA was extracted from the cytoplasm or nucleus of NIH-3T3 mouse fibroblasts after recent infection with Ki-MuLV. Following electrophoresis through a 1.0% (w/v) agarose gel, DNA species were immobilized on nitrocellulose and hybridized with (i) Ki-MuLV/VL30 cDNA probe, and later (ii) "nick-translated" VL30 clone NVL-3 DNA probe. Lanes a and b cytoplasmic and nuclear DNA, respectively. Abbreviations used in this figure: FI, form I DNA (supercoiled closed circular DNA); FIII, form III DNA (linear DNA). Sizes of DNA species are indicated in kilobase pairs.

Figure R-1(i) shows the result of a Southern blot analysis of the low molecular weight DNA preparations. The samples have been fractionated by electrophoresis through an agarose gel, transferred to nitrocellulose, and hybridized with a cDNA ^{32}P-labelled probe synthesized using virion (ie Ki-MuLV and VL30) RNA as template. The signal at the top of each track was due to contamination by high molecular weight mouse DNA. This contains multiple copies of endogenous MuLV proviruses (Coffin, 1982) and VL30 genetic elements (Keshet & Itin, 1982), both of which were related to the RNA species from which the probe was synthesized.

Two major size classes of DNA (8.5kbp and 4.6-4.9kbp) were detected in the cytoplasmic DNA preparation (figure R-1(i), lane a). Their sizes identified them as unintegrated, linear DNA forms of Ki-MuLV and VL30, respectively, since these were commensurate with the 35-38S, and 30S RNA species previously detected in Ki-MuLV-encapsidated RNA preparations (see figure R-14). The appearance of the VL30 DNA as an apparent doublet in this gel system reflected the size heterogeneity observed for the 30S virion RNA (Clewley & Avery, 1982; J.D. Norton, unpublished).

Low molecular weight DNA isolated from the cell nuclei gave a more complex gel profile (lane b). This was due partly to the presence of cytoplasmic viral DNA contaminating the nuclear DNA preparation and also to the fact that closed circular Ki-MuLV DNA co-migrated with the various linear VL30 DNA species.

The inferred identities of the different unintegrated DNA species are depicted as in the figure. For clarity, the same filter is shown following hybridization with a ^{32}P-labelled, cloned (see later) VL30 DNA probe [figure R-1(ii)]. The nuclear fraction (lane b) can be seen to contain linear, cytoplasmic-derived VL30 DNA species, also a faint doublet towards the bottom of the track. This migrated in a position expected for closed circular VL30 DNA. Its identity as such was

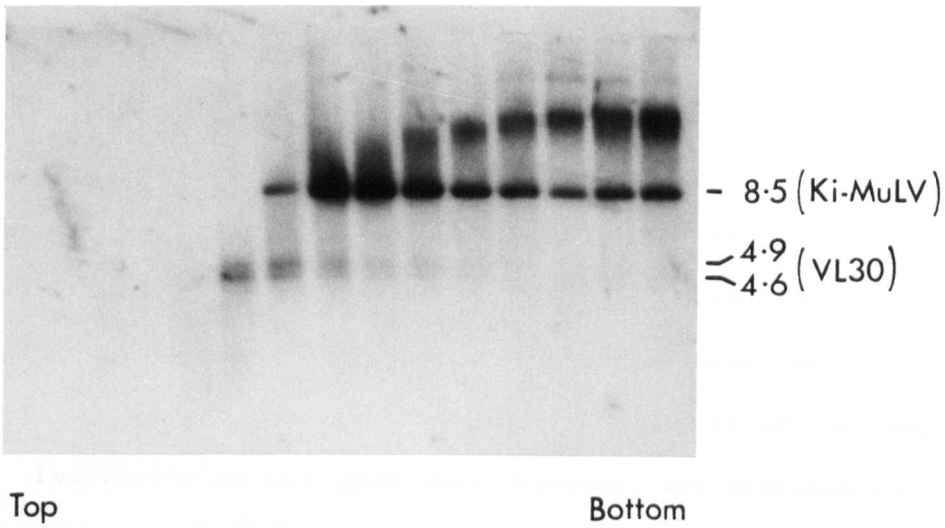


Figure R-2 Sucrose gradient fractionation of unintegrated Ki-MuLV and VL30 DNA. Low mol.wt. DNA was extracted from the cytoplasm of NIH-3T3 mouse fibroblasts after recent infection with Ki-MuLV, and layered onto a 10.4 to 21.8% (w/v) convex exponential sucrose gradient. Following centrifugation at 28,000rpm for 18 hours at 5°C in an MSE 3 x 25ml swing-out rotor, the gradient was unloaded from the top and 1ml fractions collected. 25µl samples from each fraction were electrophoresed through a 0.8% (w/v) agarose gel, the DNA immobilized on nitrocellulose and hybridized with ^{32}P-labelled Ki-MuLV/VL30 cDNA probe. Lane V contains VL30 DNA but no Ki-MuLV DNA. Sizes of DNA species are indicated in kilobase pairs.

confirmed after it was banded on a caesium chloride-ethidium bromide density gradient (see below). Thus, the unintegrated VL30 DNA species detected soon after retrovirus infection appeared to be exactly analogous to those found with retroviruses.

R-1.3 Further enrichment for VL30 retrovirus-like DNA

To obtain an efficient separation between the linear forms of VL30 and Ki-MuLV DNA, 25ml isokinetic sucrose gradients were employed. In normal linear gradients resolution between two components does not improve with sedimentation beyond one third to one half of the length of the tube. Isokinetic sucrose gradients, however, are designed to provide a constant velocity sedimentation, allowing separation to proceed for the entire depth of the gradient (Noll, 1967).

Low molecular weight DNA extracted from the cytoplasmic fraction of Ki-MuLV-infected cells was subjected to isokinetic sucrose gradient fractionation and analysis by Southern transfer. A typical result is shown in figure R-2. Hybridization of the filter with the cDNA probe showed that good separation of VL30 DNA from the linear Ki-MuLV DNA had been achieved in one fraction (labelled V in the figure). However this fraction was still contaminated by mouse chromosomal DNA. The presence of this high molecular weight DNA has caused "trailing" of the unintegrated Ki-MuLV DNA down to the bottom of the gradient. Although not visible in this preparation, some cytoplasmic DNA preparations were contaminated by the supercoiled forms of circular Ki-MuLV DNA. Upon sucrose gradient fractionation these were invariably found at the bottom of the gradient.

When low molecular weight DNA extracted from infected cell nuclei was centrifuged to equilibrium in a CsCl-ethidium bromide density gradient, several viral DNA species were found which co-banded with closed circular plasmid Col El DNA [figure R-3(i)]. The two most

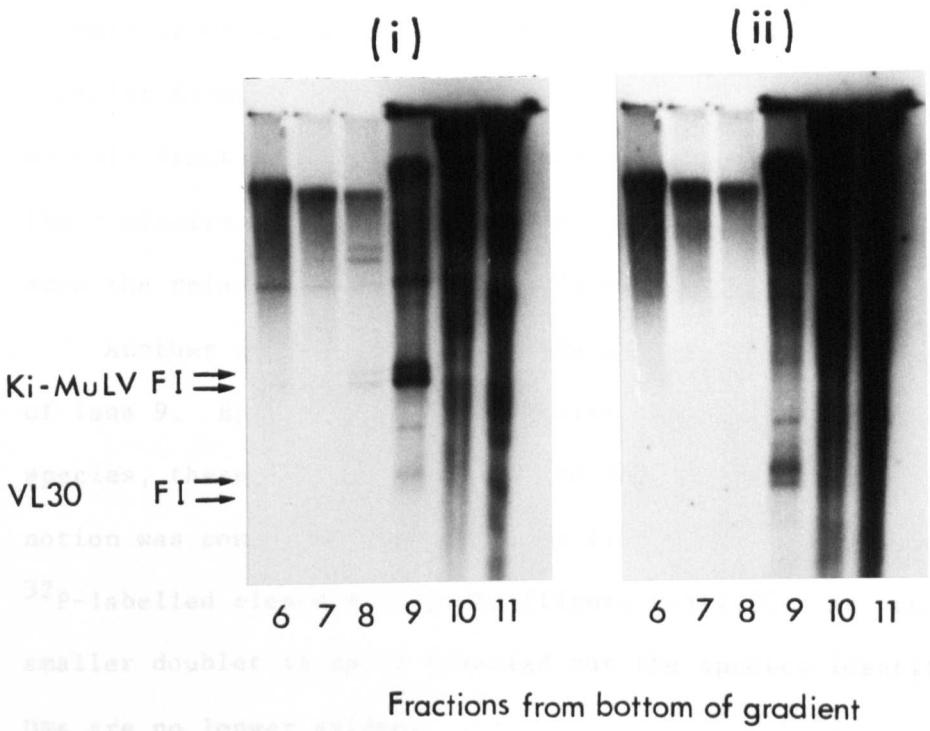


Figure R-3 Circular forms of unintegrated Ki-MuLV and VL30 DNA. Low mol.wt. DNA was extracted from the nucleus of NIH-3T3 mouse fibroblasts 20 hours post infection with Ki-MuLV, and added to 20ml caesium chloride solution (0.8g ml^{-1}) containing ethidium bromide (300µg ml^{-1}). A steep gradient was formed by centrifugation at 54,000rpm for 24 hours at 18°C in an MSE 8 x 25ml angle rotor, then "relaxed" by turning the speed down to 40,000rpm for 64 hours. 1ml fractions were collected from the bottom of the tube and extracted with butanol to remove ethidium bromide. One tenth of each fraction was electrophoresed through a 0.8% (w/v) agarose gel, the DNA transferred to nitrocellulose and hybridized with (i) Ki-MuLV/VL30 cDNA probe and later (ii) "nick translated" VL30 clone NVL-3 DNA probe. Abbreviation: FI, form I DNA (supercoiled closed circular DNA).

intense bands in lane 9 were the large and small forms of closed circular Ki-MuLV DNA. These are better resolved by the gel in lane 8, as this fraction contained less contaminating chromosomal DNA to retard their electrophoretic mobility. At the top of this track can also be seen the relaxed (open circular) forms of these two species.

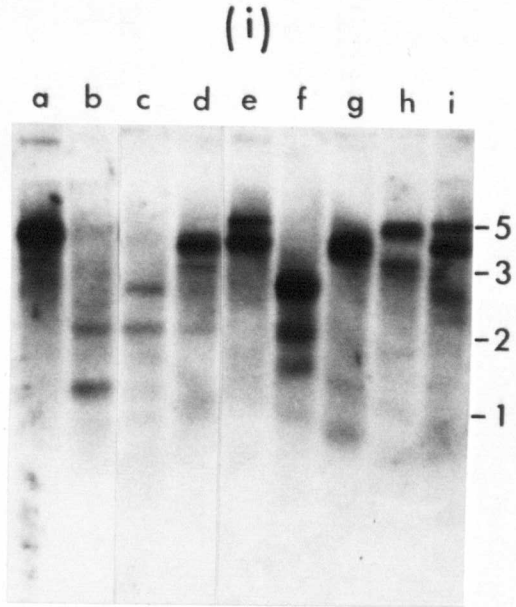
Another doublet of smaller DNA species migrated towards the bottom of lane 9. By their relative mobility in comparison to the Ki-MuLV species, these were suspected to be supercoiled VL30 circles. This notion was confirmed when the same filter was later hybridized with a ^{32}P-labelled cloned VL30 probe [figure R-3(ii)]. As can be seen, the smaller doublet is again revealed but the species identified as Ki-MuLV DNA are no longer evident.

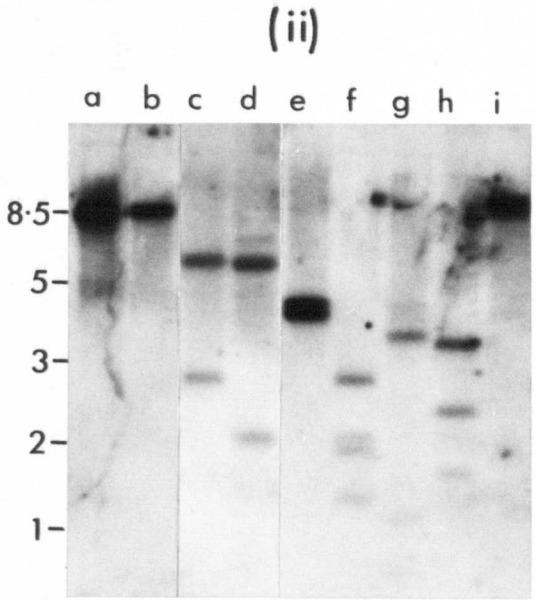
More preparations of unintegrated viral DNA are shown in figure R-10. These were fractionated more successfully than that shown in figure R-1, as less contamination of the nuclear DNA fraction by cytoplasmic DNA species has occurred. Band intensities obtained using the cDNA probe appeared at first to indicate that more Ki-MuLV DNA was synthesized *in vivo* than was VL30 DNA. However it is possible that the conditions of cDNA probe synthesis favour one class of RNA template over another. This situation may not necessarily be reflected in the infected cell.

R-1.4 Restriction enzyme analysis of in vivo-synthesized DNA

Unintegrated linear forms of both the Ki-MuLV and VL30 DNA species were studied, (a) to investigate the possibility of any sequence relationship between them, and (b) to identify enzymes which might be useful for molecular cloning manipulations.

Southern blotting experiments with these DNAs invariably suffered as a result of the high background signal generated by the chromosomal DNA which contaminated the preparations. Since less VL30 DNA-specific


(i)
a b c d e f g h i
-5
-3
-2
-1


(ii)
a b c d e f g h i
8·5-
5-
3-
2-
1-

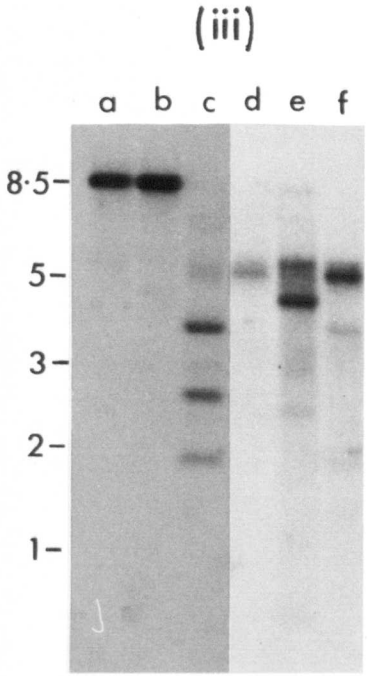

(iii)
a b c d e f
8·5-
5-
3-
2-
1-

Figure R-4 Restriction enzyme analysis of *in vivo*-synthesized VL30 and Ki-MuLV DNA. Samples of sucrose gradient-enriched linear *in vivo* VL30 and Ki-MuLV DNA were incubated with an excess of the appropriate restriction enzyme then electrophoresed through a 1.0% (w/v) [(i) and (ii)] or a 0.8% (w/v) [(iii)] agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled Ki-MuLV/VL30 cDNA. (i) VL30 DNA, (ii) Ki-MuLV DNA; lanes a, undigested; lanes b to i, digested with *Hind* III, *Sac* I, *Xba* I, *Xho* I, *Pvu* II, *Bgl* II, *Bam* HI and *Eco* RI, respectively. (iii) a, undigested Ki-MuLV DNA; b, *Eco* RI-digested Ki-MuLV DNA; c, *Bam* HI-digested Ki-MuLV DNA; d, undigested VL30 DNA; e, *Eco* RI-digested VL30 DNA; f, *Bam* HI-digested VL30 DNA.

signal was generated in filters hybridized with the cDNA probe, the problem of a high background signal was particularly noticeable with this DNA. This is exemplified in figure R-4, where the two autoradiographs (i) and (ii) compare restriction enzyme digestion profiles of VL30 and Ki-MuLV DNA, respectively. It can be seen that the background smear of hybridization is far worse in section (i).

This preliminary study showed that some restriction fragments appeared to be shared in the two classes of DNA. For example, a 2.7kbp SacI band (lanes c), a 2.0kbp XbaI band (lanes d) and at least three bands generated by digestion with PvuII (lanes f). Once enzymes which failed to cleave Ki-MuLV DNA had been identified, these apparently common restriction fragments were found to be the result of contamination of the VL30 DNA size fraction with truncated fragments of Ki-MuLV DNA. This is best illustrated in figure R-4(iii) where a BamHI digest of Ki-MuLV DNA (lane c) generated two bands which were also seen (more faintly) in the VL30 DNA profile (lane f). This autoradiograph also demonstrated that Ki-MuLV DNA lacked EcoRI sites (compare lanes a and b) and that the major part of the VL30 DNA preparation lacked sites for BamHI.

The results presented in this section have indicated that: (a) species of Ki-MuLV and VL30 DNA are synthesized in mouse cells which have been infected with virus produced by the Al cell line. Both genetic elements were typically retrovirus-like as linear forms predominated in the cytoplasm and circular forms were detected in the nucleus. (b) Restriction enzyme analysis of sucrose gradient-enriched preparations showed that Ki-MuLV and VL30 DNA could be distinguished by digestion with BamHI, which cleaves Ki-MuLV but not VL30, and with EcoRI, which cleaves VL30 species but not Ki-MuLV DNA. (c) VL30 DNA was apparently synthesized less efficiently in newly infected mouse cells than was Ki-MuLV (though see earlier comments). The disappointing

yields of VL30 DNA signified that attempts to molecularly clone these unintegrated species would be hampered by the need to screen a large number of transformants with a cDNA probe that was also related to numerous endogenous provirus sequences.

On the strength of these findings it was decided to attempt the synthesis of full-length VL30 DNA *in vitro* using the retrovirus endogenous reaction (see below). If successful, this would ensure a degree of purity which would be much more favourable for the molecular cloning of VL30 DNA.

R-2 Synthesis and characterization of in vitro VL30 DNA

R-2.1 Introduction

Purified preparations of retrovirus particles contain all the machinery necessary to carry out viral DNA synthesis; virions treated with mild detergents which disrupt the virus membranes and permit the entry of deoxynucleoside triphosphates carry out extensive DNA synthesis under the appropriate reaction conditions. In some cases a complete DNA copy of the viral RNA can be synthesized (Rothenberg *et al.*, 1977; Gilboa *et al.*, 1979b; Benz & Dina, 1979) which retains biological activity (Rothenberg *et al.*, 1977; Gilboa *et al.*, 1979b; Boone & Skalka, 1980). In many cases these reactions, which use the endogenous reverse transcriptase and are thus sometimes referred to as "endogenous reactions", parallel the events in the infected cell, and have been valuable in elucidating the mechanisms of reverse transcription. The advantages of this *in vitro* DNA synthesis are that a large number of virions can be used in a single reaction, yielding substantial amounts of viral DNA, and that the viral DNA which is synthesized is essentially free from host cell DNA. Having established that Ki-MuLV (VL30) particles synthesize both Ki-MuLV and VL30 DNA *in vivo*, it seemed possible that an endogenous reaction using these same virions *in vitro* would yield VL30 DNA of sufficient purity for molecular cloning.

Ki-MuSV DNA has been successfully synthesized in an endogenous reaction (Norton *et al.*, 1982). Since Ki-MuSV particles consist of the defective Ki-MuSV genome which is pseudotyped by Ki-MuLV particles, optima for the Ki-MuLV (VL30) reaction were expected to be similar. However it was considered necessary to perform a series of experiments to determine the need for any modification to these reaction conditions. Several parameters were considered: (1) The final concentration of virus in the reaction mix that could be routinely achieved using Ki-MuLV particles from the A1 cell line TCF. (2) The optimum detergent

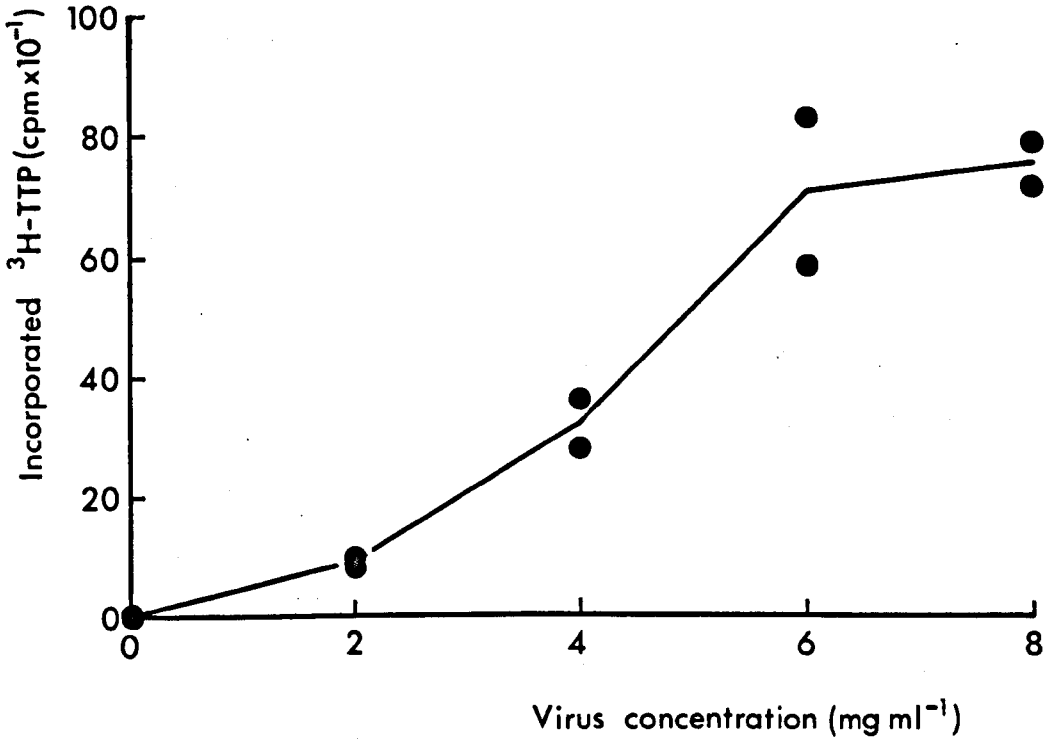


Figure R-5 Effect of virus concentration on endogenous reactions. Ki-MuLV was concentrated from A1 cell TCF and used to set up 5 duplicate 100 µl endogenous reactions at the final concentrations indicated. Each tube received 25µCi of ^{3}H-TTP, and the reactions were initiated by adding TX-100 to 0.03% (v/v). After incubation at 37°C for 16 hours, SDS was added to 1.0% (w/v), and duplicate 45µl samples were used to measure DE-81-bindable radioactivity.

concentration. (3) Whether storage at -70°C affected the endogenous enzyme activity of concentrated virus particle preparations. (4) The optimum time of incubation for the reaction.

R-2.2 Effect of varying the concentration of Ki-MuLV particles

In this experiment the final concentration of virus in an endogenous reaction was varied over a range of values that could be achieved by routine purification procedures. In each reaction, the concentration of detergent (TX-100) was maintained at 0.03% (v/v). All other reactants were present as described in Methods (section M-12) except that each tube received 25µCi of ^{3}H-TTP that had been lyophilized and re-dissolved in water.

Virus was concentrated from one litre of A1 cell line TCF, as detailed in Methods (section M-5). A total of 4mg of virus concentrate was used to set up 5 duplicate 100µl endogenous reactions. These were incubated at 37°C for 16 hours and stopped by addition of SDS to 1.0% (w/v). Aliquots were spotted directly onto DE-81 paper for measurement of DNA synthesis by incorporation of ^{3}H-TTP.

The results (figure R-5) showed an increase in DNA synthetic activity up to a concentration of 6mg virus ml^{-1}, after which the curve formed a plateau. This may have been caused by sub-optimal lysis of the virions by the detergent, or by one of the other reactants reaching a limiting concentration.

Since the useful range of final virus concentration was observed to extend between 2-6mg ml^{-1} (ie the exponential portion of the curve in figure R-5) a mid-range value of 4mg ml^{-1} was chosen to be used in all subsequent endogenous reactions.

R-2.3 TX-100 concentration

The concentration of detergent used to activate virion-associated

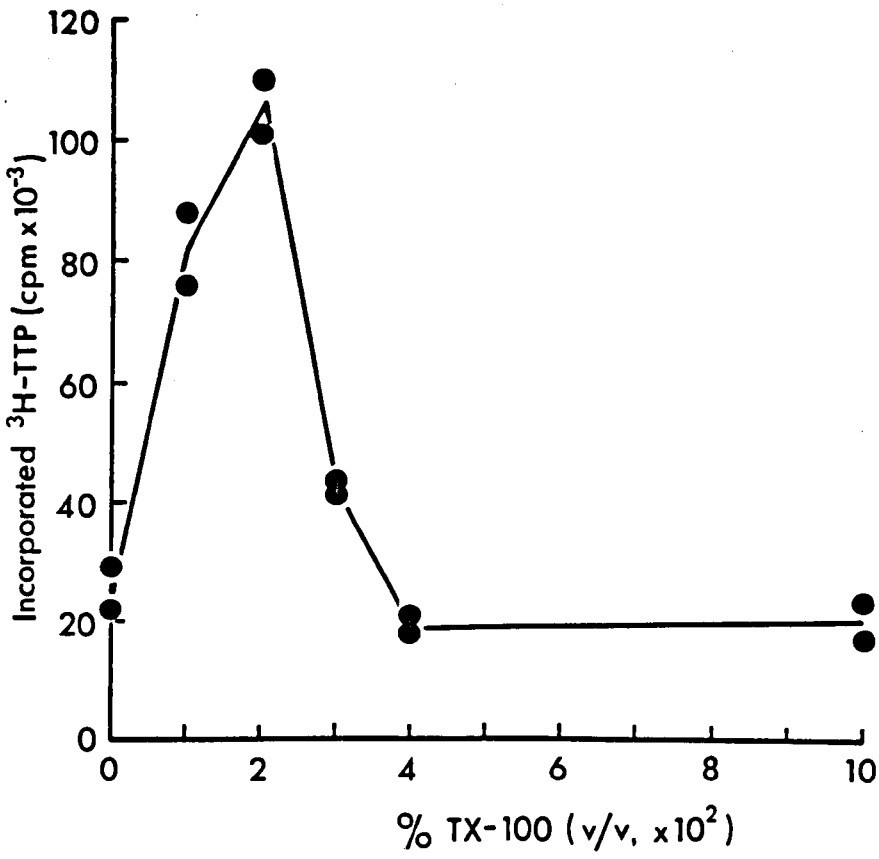


Figure R-6 Effect of detergent concentration on endogenous reactions. Purified Ki-MuLV particles were adjusted to 4.2mg ml^{-1} for 6 duplicate 100µl endogenous reactions. Each tube received 20µCi of ^{3}H-TTP, and the reactions were initiated by adding TX-100 to one of the final concentrations indicated. After incubation at 37°C for 16 hours, SDS was added to 1.0% (w/v), and duplicate 45µl samples were used to measure DE-81-bindable radioactivity.

reverse transcriptase is a critical factor in retrovirus endogenous reactions. A balance must be reached to compromise between, on the one hand, virion lysis, which effectively destroys reverse transcriptase activity, and on the other, under-lysis resulting in the inefficient access of triphosphates and Mg^{2+} ions to the virion core. In this experiment, DNA synthesis was measured as a function of TX-100 concentration. A detergent concentration of 0.03% (v/v) was used with success by Junghans et al. (1975) for RSV and also by Norton et al. (1982) for Ki-MuLV (Ki-MuSV). It was reasonable to assume that a similar concentration would be applicable to the Ki-MuLV (VL30) reaction.

As illustrated in figure R-6, there was a sharp peak of DNA synthesis at 0.02% (v/v) TX-100. Increasing the concentration of TX-100 to 0.04% reduced the level of DNA synthesis to that of the background obtained with no added detergent. The former concentration was used in all ensuing experiments.

R-2.4 Time course of Ki-MuLV endogenous reaction

Results of the previous two experiments had given no indication of (a) degree of integrity of the DNA species produced in the endogenous reaction, (b) relative quantitites of Ki-MuLV and VL30 DNA synthesized, or (c) the optimum time of incubation. To investigate these parameters a time course experiment was performed in which ^{32}P-labelled DNA products were examined by agarose gel electrophoresis (figure R-7). All reaction conditions were optimized as indicated in Methods, except that the virus preparation had been stored frozen at -70°C for 14 days prior to its use.

Discrete products were seen to appear after 3h and to accumulate up to a maximum of 12h. Thereafter degradation occurred. The sizes of the two major reaction products (8.5kbp and 4.6-4.9kbp) were identical to

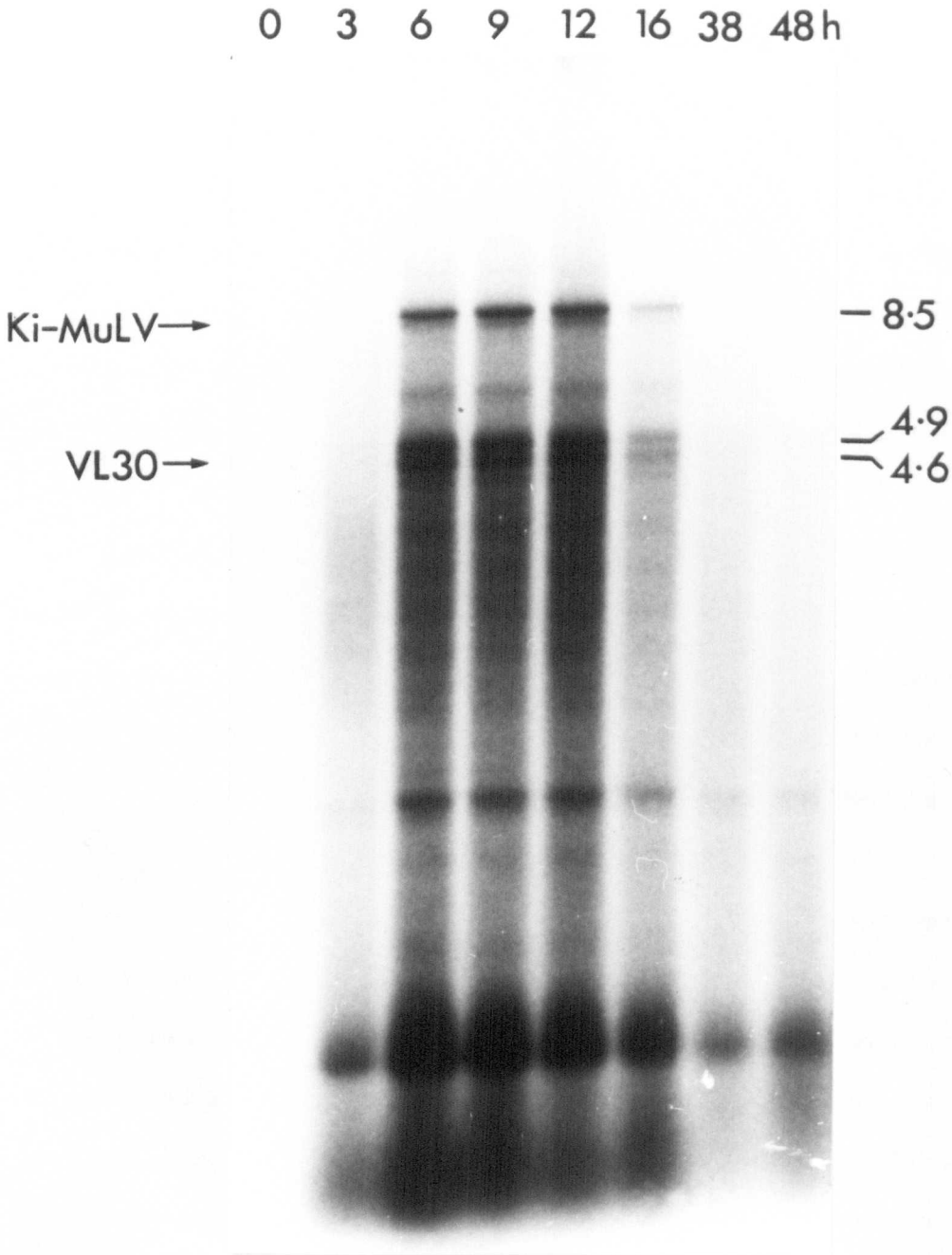

0 3 6 9 12 16 38 48 h
Ki-MuLV
VL30
8·5
4·9
4·6

Figure R-7 Agarose gel electrophoresis of ^{32}P-labelled *in vitro* DNA. A Ki-MuLV endogenous reaction was performed using ^{32}P-labelled dCTP under standard conditions. Samples were removed at the indicated times and purified as described in Methods, section M-12. Double-stranded DNA products were electrophoresed through a neutral 0.8% (w/v) agarose gel at 30V for 16.5h. The gel was dried on Whatman DE-81 paper for subsequent detection of DNA species by autoradiography. Sizes of major reaction products are given in kilobase pairs.

the *in vivo* synthesized Ki-MuLV and VL30 DNA species (see figure R-10 where gel-purified *in vitro* DNA is seen to co-migrate with the corresponding *in vivo*-synthesized species). Once more the *in vitro* VL30 DNA appeared as a doublet. Judging by relative band intensity, approximately equal amounts of Ki-MuLV and VL30 DNA had been synthesized.

In addition to the major reaction products shown in figure R-7, other DNA species can be seen. The sizes and relative amounts of these minor products were similar to those observed by workers with other retroviruses (Benz & Dina, 1979; Gilboa *et al.*, 1979b). Such species arise as intermediates in the formation of the final endogenous reaction products (Dina & Benz, 1980; Gilboa *et al.*, 1979a) (see General Introduction) but were not further characterized.

Not shown are the results of an identical time course experiment which used freshly prepared virus. The yield of *in vitro*-synthesized DNA products, and their degree of integrity were indistinguishable in each case, showing that storage at -70°C was adequate for retention of faithful reverse transcriptase activity.

Confluent Al cells on the Bellco Autoharvester machine produced virus continuously for 7-10 days before the monolayers began to strip off from the inside of the roller bottles. Thus freshly prepared batches of virus concentrate were stored at -70°C during the production run, to be processed for *in vitro* DNA production at a more convenient time.

R-2.5 Endogenous reactions using melittin

Melittin is the major component in bee (*Apis mellifera*) venom. It is a cationic peptide of 26 amino acids with an unusual arrangement of largely hydrophobic (1-20) and hydrophilic (positions 21-26) residues which makes it highly surface active. It has been characterised by its

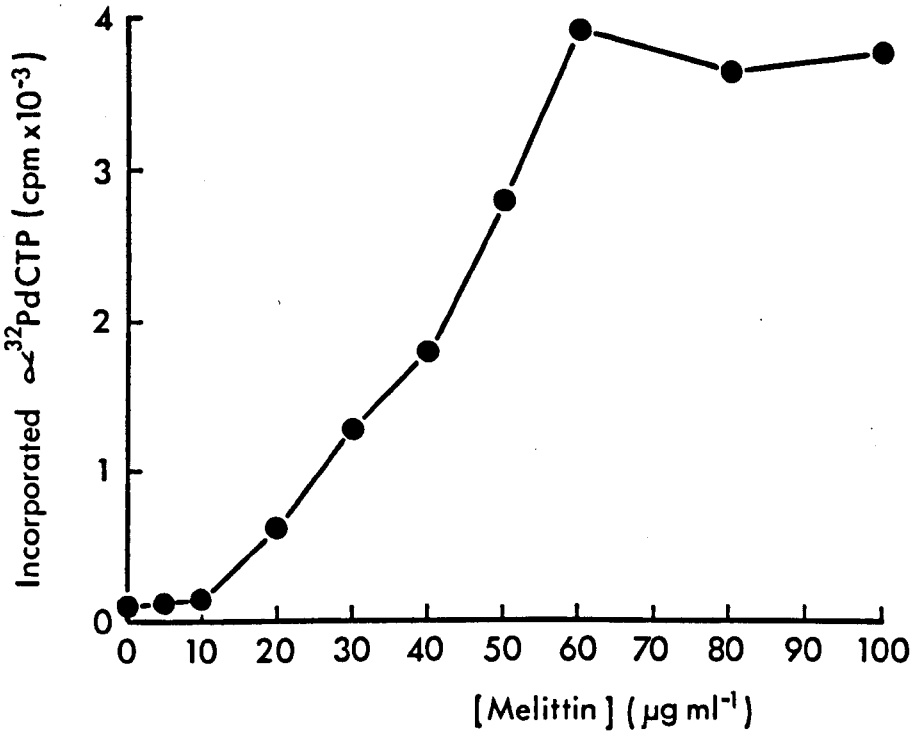


Figure R-8 Effect of melittin concentration on endogenous reactions. Ten 50μl Ki-MuLV/Ki-MuSV endogenous reactions were performed using ^{32}P-labelled dCTP. Reactions were initiated by adding melittin to the final concentration indicated. After incubation at 37°C for 1 hour, SDS was added to 1.0% (w/v), and duplicate 20μl samples were used to measure DE-81-bindable radioactivity.

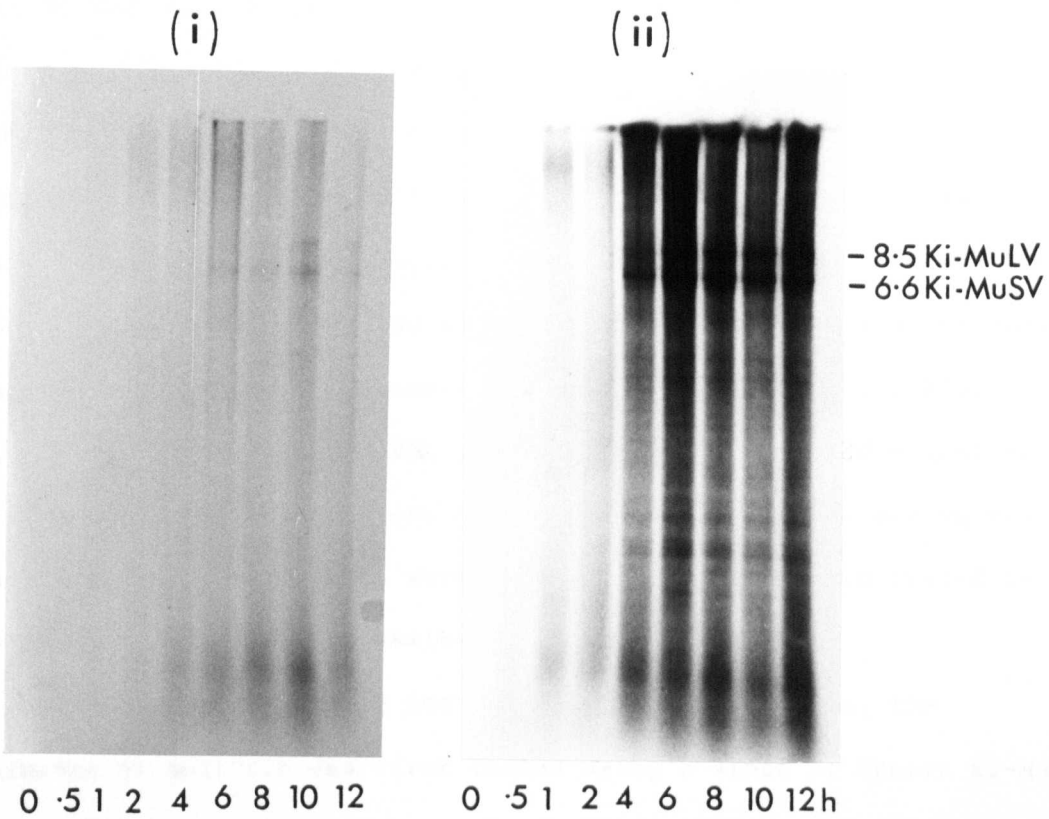


Figure R-9 In vitro DNA synthesized using TX-100 or melittin. 250µl Ki-MuLV/Ki-MuSV endogenous reactions were performed using ^{32}P-labelled dCTP under standard conditions. Reactions were initiated by adding (i) TX-100 to a final concentration of 0.02% (v/v) or (ii) melittin to a final concentration of 60µg ml^{-1}. 25µl samples were removed at the indicated times and purified as described in Methods, section M-12. Double-stranded DNA products were electrophoresed through a neutral 1.0% (w/v) agarose gel at 30V for 16h. The gels were dried on Whatman DE-81 paper for subsequent detection of DNA species by autoradiography. Sizes of major reaction products are given in kilobase pairs.

action on both natural and artificial membranes (Sessa et al., 1969; Williams & Bell, 1972; Dawson et al., 1978) and due to its high affinity for phospholipids does not lyse membranes by solubilizing their components as a detergent would but rather disrupts them by producing a wedge-like effect (Dawson et al., 1978). Membrane lesions are smaller than those caused by TX-100 and so melittin has also been investigated for use in retrovirus endogenous reactions (Boone & Skalka, 1980). Studies with RSV showed that DNA yields were up to 3.5-fold higher with melittin than with the detergent NP-40, yielding 250ng DNA per mg virus. Furthermore, the DNA products were biologically active when tested by transfection assay (Boone & Skalka, 1980).

In the hope of obtaining increased yields of VL30 DNA, the performance of melittin was first tested using a stock of frozen Ki-MuLV (Ki-MuSV) (kindly donated by Dr. J.D. Norton). Figure R-8 shows the result of incubating this virus (at 4mg ml^{-1}) with increasing concentrations of melittin. It can be seen that a peak of incorporated radioactivity occurred at 60μg ml^{-1} but that this peak was much broader than the one obtained using TX-100 (see figure R-6), a phenomenon also observed by Boone & Skalka (1980) using RSV. The use of melittin, therefore, allows for a greater margin of error when setting up an endogenous reaction.

Figure R-9 compares the yields of Ki-MuSV DNA obtained using either melittin or TX-100 in a time course experiment similar to the one described above. It can be seen that although both reactions produced full-length DNA molecules (Ki-MuLV is also seen since all MuSV preparations contain a significant proportion of helper virus [MuLV]), the melittin reaction gave a far higher yield.

DNA products synthesized using either reaction method appeared to be identical in every sense except for that of final yield. Since melittin activation was more efficient, it was used in most of the later

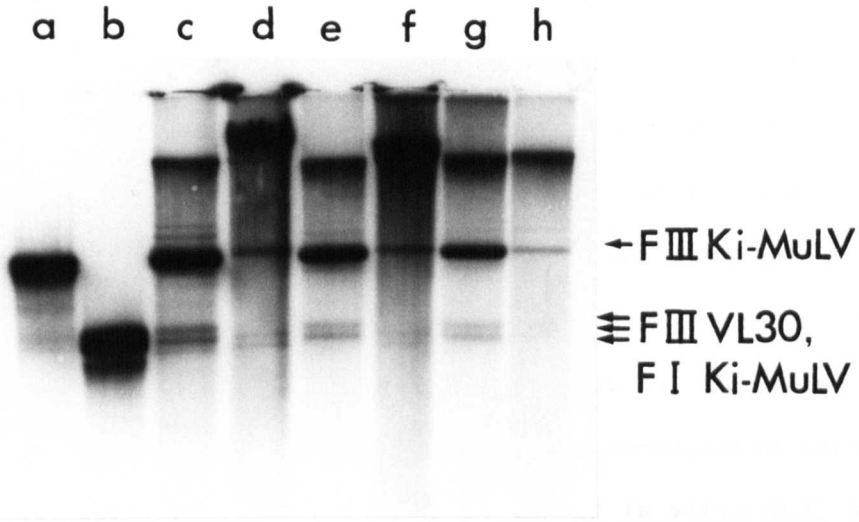


Figure R-10 Electrophoretic mobility of *in vitro* and *in vivo* DNA preparations. DNA species were fractionated on a 0.8% (w/v) agarose gel, immobilized on nitrocellulose and hybridized with ^{32}P-labelled Ki-MuLV/VL30 cDNA probe. Lanes a, b: *in vitro*-synthesized Ki-MuLV and VL30 DNA, respectively. Lanes c-h: low mol.wt. DNA extracted from NIH-3T3 cells after recent infection with Ki-MuLV. c, e, g: cytoplasmic fractions; d, f, h: nuclear fractions. Compare with figure R-1.

experiments with *in vitro*-synthesized DNA.

R-2.6 Characterization of in vitro-synthesized DNA

As shown by results of the time course experiment (figure R-7) the Ki-MuLV endogenous reaction yielded DNA species whose migration in agarose gels suggested they were full-length versions of Ki-MuLV and VL30 proviral DNA. Figure R-10 is a Southern blot comparing gel-purified, *in vitro*-synthesized DNA species with those detected in low molecular weight DNA from recently infected mouse cells. The major classes of this *in vivo* linear Ki-MuLV and VL30 DNA co-migrated with their *in vitro* counterparts. In this preparation of *in vitro* VL30 DNA, the various species electrophoresed as a triplet of bands, the smallest band representing either a minor VL30 species or perhaps a degradation product of full-length Ki-MuLV or VL30 DNA.

The Southern blots depicted in figure R-11 show a restriction enzyme analysis of *in vitro*-synthesized Ki-MuLV DNA (i) and VL30 DNA (ii). The cDNA probe hybridized with both classes of DNA, and as with *in vivo* DNA some digestion profiles appeared to be similar for both Ki-MuLV and VL30 (for instance the *Pvu*II digests in lanes (b) and the *Sal*I digests in lanes (f) of each gel). Despite the confusion caused by bands which were apparently of Ki-MuLV origin, it was clear that many restriction fragments derived solely from the VL30 DNA preparation.

Judging by the small degree of complexity attributable to VL30 restriction digest profiles, the *in vitro*-synthesized VL30 DNA was composed of a relatively homogeneous population of molecules. This finding added weight to an earlier observation, based on the oligonucleotide mapping of VL30 RNA rescued by Ki-MuLV particles, which similarly showed it to possess a limited sequence heterogeneity (Clewley & Avery, 1982).

A restriction enzyme analysis the results of which are presented in

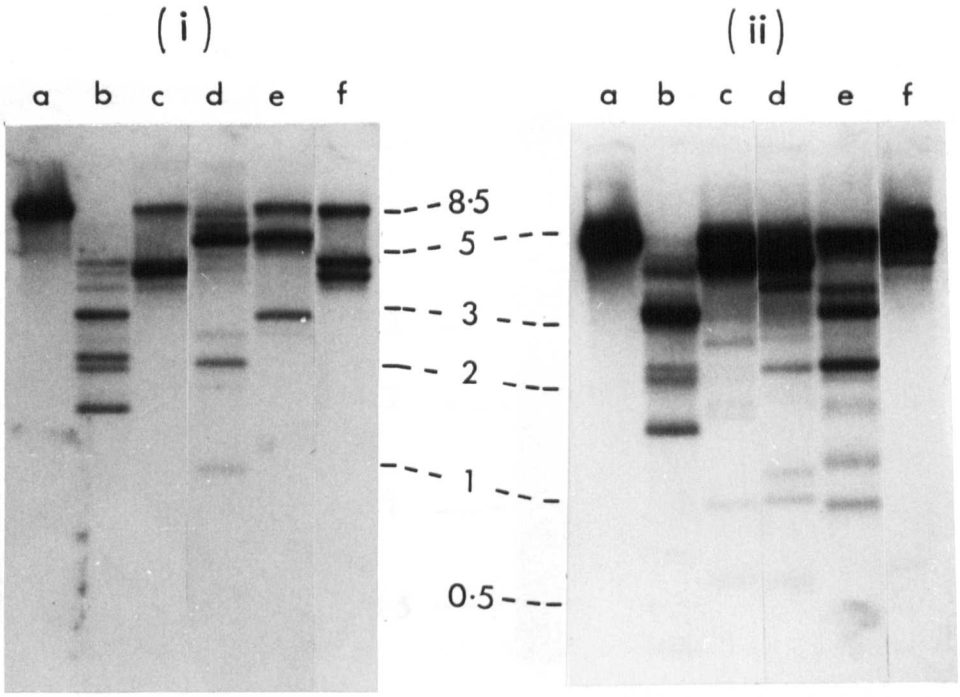


Figure R-11 Restriction enzyme analysis of *in vitro* Ki-MuLV and VL30 DNA. Samples of agarose gel-purified Ki-MuLV DNA (i) or VL30 DNA (ii) were incubated with an excess of the appropriate restriction enzyme, electrophoresed through a 1.5% (w/v) agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled Ki-MuLV/VL30 cDNA. Lanes a, undigested DNA; b-f, digested with *Pvu* II, *Xho* I, *Xba* I, *Sac* I and *Sal* I, respectively. Scale is given in kilobase pairs of DNA.

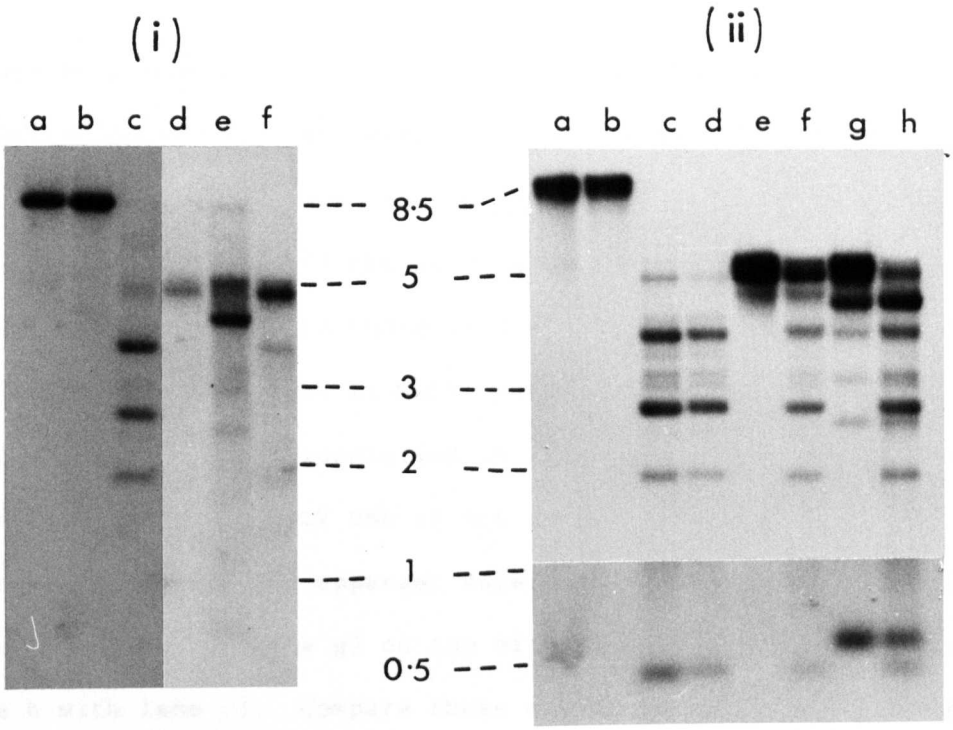


Figure R-12 Comparison of restricted *in vivo* DNA with *in vitro* DNA. Samples of sucrose gradient-enriched linear *in vivo* DNA (i) or agarose gel-purified *in vitro* DNA (ii) were incubated with the appropriate restriction enzyme, electrophoresed through a 0.8% (w/v) agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled Ki-MuLV/VL30 cDNA. Lanes a-c, Ki-MuLV DNA, undigested, digested with *Eco* RI, or with *Bam* HI, respectively. Lanes (i)d and (ii)e, undigested VL30 DNA; (i)e and f and (ii)g and f, VL30 DNA digested with *Eco* RI and *Bam* HI, respectively. Lanes (ii)d and h, Ki-MuLV and VL30 DNA, respectively, digested with *Eco* RI and *Bam* HI. The bottom section of gel (ii) is a photograph of a longer autoradiographic exposure than that used for the top section. Scale is given in kilobase pairs of DNA.

figure R-12 showed that *in vitro*-synthesized DNA species shared identical restriction fragment gel profiles with the *in vivo*-synthesized DNAs. Section (ii) shows that *in vitro* VL30 DNA was not cleaved by *Bam*HI, since in lane (f) the bands generated upon digestion by this enzyme were identical to those produced by the *Bam*HI-digested Ki-MuLV DNA (lane c). Truncated Ki-MuLV contaminants of the VL30 DNA preparation were also identified in a (*Bam*HI + *Eco*RI) double digestion (lane h). Since Ki-MuLV DNA is not cleaved by *Eco*RI (lane b), this digest resulted in the apparent superimposition of a VL30 *Eco*RI digestion profile (lane g) on top of a Ki-MuLV *Bam*HI profile (compare lane h with lane c). Compare these digestion profiles with the *in vivo* profiles of section (i) (previously described in figure R-4).

R-3 Molecular cloning of in vitro synthesized VL30 DNA

R-3.1 Introduction

Although the yields of DNA synthesized in TX-100 or melittin-activated endogenous reactions represented an approximately one thousand-fold increase upon that obtained from infected cells, considerable effort was necessary to produce material that could still only be examined successfully by the Southern blot technique. Furthermore the VL30 DNA species obtained by this method not only displayed a moderate degree of heterogeneity with respect to size and restriction enzyme sites, but they were also contaminated by fragments of sub-genomic Ki-MuLV DNA. Molecular cloning in an *Escherichia coli* host offered the ideal approach for obtaining biochemically pure DNA of those VL30 species which were packaged by Ki-MuLV particles.

Of the various strategies available for molecular cloning, it was decided to insert linear VL30 DNA molecules into the unique restriction enzyme site of a plasmid vector using synthetic linker molecules. This method had been recently used with success to clone *in vitro*-synthesized Ki-MuSV DNA (Norton & Avery, 1982) and offered the advantage that cloned inserts which were released from vector DNA would be faithful copies of a complete VL30 genome in a non-permuted form.

R-3.2 Cloning procedure

The various manipulations performed during the molecular cloning of VL30 cDNA have been described in detail in Methods (section M-23). Briefly, several milligrammes of Ki-MuLV (VL30) particles were used in two separate endogenous reactions. DNA products were purified by digestion with S1 nuclease and preparative agarose gel electrophoresis. Phosphorylated molecular linkers containing the *Bam*HI recognition site were tested in polyacrylamide gels, and if acceptable were blunt-end ligated onto the flush-ended VL30 cDNA. Excess concatemers of linkers

VL30 yield after purification (ng)	*E. coli* host	Transformation efficiency (colonies μg^{-1})	Transformants (amp^R)	amp^R, tet^S transformants	VL30 inserts	Ki-MuLV inserts
50	HB 101	2×10^5	646	42	6	11
100	MC 1060	4×10^6	> 12,000	146/6000	6	22

Table R-1 Results of two attempts to molecularly clone VL30 cDNA from endogenous Ki-MuLV reactions.

were removed by digestion with BamHI, which left each VL30 DNA molecule with cohesive BamHI termini. To prevent their competition with VL30 DNA for BamHI termini in the vector DNA, the excess linker fragments were removed by Sephadex G-200 SF column chromatography.

Meanwhile, vector DNA was prepared by (1) cleavage with BamHI to generate a linear molecule and (2) removal of 5'-phosphate groups by calf intestinal alkaline phosphatase (CIAP).

Finally, VL30 DNA with its cohesive BamHI termini was ligated into the linearized pAT153 vector, and the resulting recombinant molecules used to transform *E. coli* HB101 or MC1060 cells to ampicillin resistance. Transformants with DNA inserts were selected by their lack of growth on agar plates containing the antibiotic tetracycline (since the unique BamHI site of pAT153 lies within the gene encoding resistance to this drug).

Amp^R, tet^S transformants were further screened by Grunstein-Hogness colony hybridization and small-scale cleared lysate preparation of plasmid DNA.

R-3.3 Transformants

Table R-1 summarizes the results of the two VL30 cloning exercises. 646 amp^R transformants were obtained from the first experiment, in which HB101 cells were used as recipients. Of 42 tet^S clones, only 6 contained plasmids with VL30 DNA inserts (see later). The higher efficiency of transformation obtained with MC1060 cells in the second experiment led to the production of >12,000 amp^R transformants. 6000 were stabbed onto tetracycline plates to identify 146 tet^S clones. Again, however, only 6 VL30 inserts were present.

Over 60% of inserts were found to be cloned BamHI fragments of *in vitro* Ki-MuLV DNA. These could be identified as such since the BamHI digestion profile of Ki-MuLV DNA had been previously characterized in

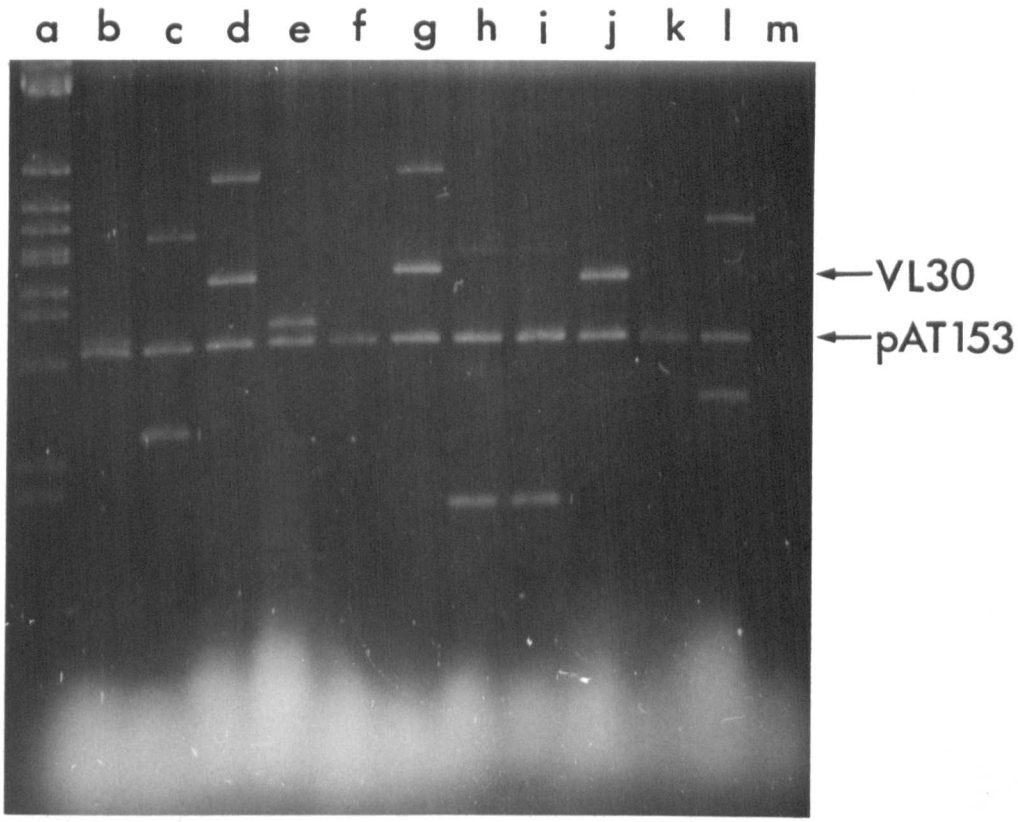


Figure R-13 Screening VL30 clones by the cleared lysate method. Plasmid DNA from single amp^R, tet^S colonies was extracted as detailed in Methods, section M-23.8.3(a), dissolved in TE and incubated with 5 units Bam HI for 2 hours at 37°C. Restricted DNA was electrophoresed through a 0.8% (w/v) agarose gel at 30V for 16h, and visualized using UV light after staining with ethidium bromide (1μg ml^{-1}) for 30 minutes. Molecular sizes of DNA species were estimated by reference to λ.Eco RI plus λ.Hind III restriction fragments (lane a).

Southern blotting experiments.

Figure R-13 gives an example of the mini-cleared lysate analysis of some of the tet^S transformants which were generated during the first cloning experiment. Apart from lambda DNA marker fragments in the first lane, each track represents a BamHI digestion profile of plasmid DNA extracted from a different tet^S clone. Linearized pAT153 vector DNA was easily recognized as it formed a continuous line of bands across the ethidium bromide-stained gel. Some tracks appeared to possess this band alone. In these instances the tet^S phenotype must have been caused during the ligation reaction by linker fragments successfully competing with VL30 DNA for the cohesive termini on the vector molecule. Alternatively, contaminating exonuclease activity may have removed one or two nucleotides from one end of the vector DNA. Upon re-ligation, this would result in an open circular molecule which was capable of transforming its host to an amp^R, tet^S phenotype, but which contained no insert.

Although some of the digests were not complete, three 4.6-4.9kbp-sized inserts could be identified from this particular gel. Of the four inserts that can be seen below the line of vector DNA bands, three were fragments of Ki-MuLV DNA, while the fourth was probably the result of a plasmid:plasmid interaction. This conclusion was reached after the gel had been blotted and hybridized with ^{32}P-labelled virion cDNA. All tracks containing VL30 or Ki-MuLV inserts hybridized with the probe (data not shown), but the fourth insert (lane 1) did not.

As stated above, each cloning experiment generated 6 clones of VL30-sized DNA fragments that were released from plasmid DNA by digestion with BamHI. All 12 clones hybridized with the cDNA probe (not shown).

R-3.4 Identification of 4.6-4.9kbp inserts as VL30 clones

VL30-sized inserts may have been VL30 clones, or the truncated forms of Ki-MuLV DNA which had been found to co-migrate with the *in vivo* and *in vitro* VL30 DNA doublets. In fact the latter possibility was less likely since restriction mapping studies of *in vivo* Ki-MuLV proviral DNA had shown that the largest *BamHI* fragment of Ki-MuLV was about 3.5kbp (see figure R-12, also a restriction map of Ki-MuLV in figure R-17). A formal proof of VL30 identity was sought, however, as the possibility remained that the 4.6-4.9kbp clones were variants of Ki-MuLV which had lost *BamHI* sites by mutation.

Figure R-14 is a northern blot of total Ki-MuLV RNA which had been electrophoresed through an agarose:formamide gel (thanks here to Peter Roberts [Montana State University, Montana, USA] for running the gel in order to demonstrate the technique). In lane (b) a ^{32}P-labelled VL30 clone hybridized with the VL30-sized RNA (30S RNA) but not with the Ki-MuLV genomic RNA. The Ki-MuLV/VL30 cDNA probe hybridized with both bands (lane c). A similar result was obtained with the other putative VL30 clones. Thus the cloned VL30 species did not represent truncated Ki-MuLV molecules. This conclusion was also supported by a comparison of the VL30 restriction maps with that of Ki-MuLV DNA (figure R-17).

VL30 clones also hybridized specifically with *in vivo*-synthesized unintegrated VL30 DNA. Figure R-1 showed a cytoplasmic and nuclear fractionation of low molecular weight DNA from Ki-MuLV (VL30) infected mouse cells. Whilst both Ki-MuLV and VL30 species were revealed by hybridization of the Ki-MuLV (VL30) cDNA probe to the blot, a nick-translated VL30 probe was specific for the bands which had been tentatively designated as linear and circular species of VL30 DNA. A similar result was also shown in figure R-3 where CsCl-ethidium bromide gradient fractions of nuclear unintegrated viral DNA were hybridized with either the cDNA or cloned VL30 probes.

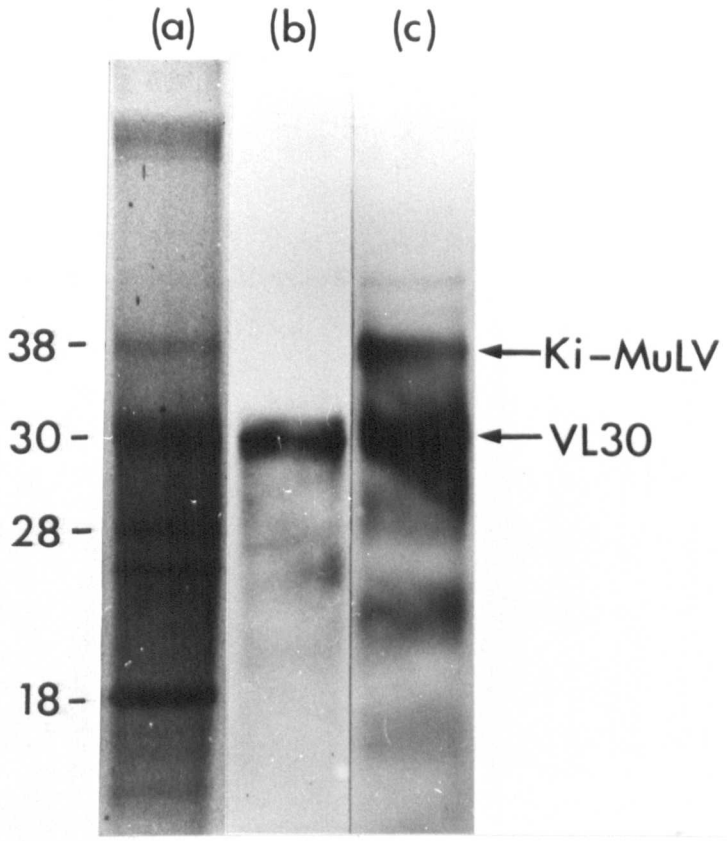


Figure R-14 Filter hybridization of virion RNA with cloned VL30 DNA probe. Purified Ki-MuLV particles were lysed by addition of SDS to 0.5%. Nucleic acid was deproteinized by phenol:chloroform (1:1, v/v) extraction and precipitated by ethanol. Following denaturation by heating to 65°C, RNA species were resolved by electrophoresis through a 1.0% (w/v) agarose/formamide gel. Size estimation was by comparison with mouse ribosomal RNA markers. Lane (a), ethidium bromide-stained gel photographed using UV light; (b), northern blot of (a) annealed to ^{32}P-labelled "nick translated" VL30 clone NVL-3 DNA; (c), the same filter reannealed to ^{32}P-labelled cDNA synthesized from total poly(A)-containing Ki-MuLV particle RNA. In lane (a) the bands at positions corresponding to 28S and 18S markers represent virion-associated ribosomal RNA.

R-3.5 Characterization of VL30 clones

The 4.6-4.9kbp length of the 12 clones, and their hybridization characteristics all fitted the known properties of mouse VL30 RNA species that had been reported by others (Howk *et al.*, 1978; Besmer *et al.*, 1979). Clones were further characterized by restriction enzyme mapping. Several hundred microgrammes of each recombinant plasmid were isolated from large scale cleared lysates, and the *Bam*HI-released inserts obtained in a pure form by preparative agarose gel electrophoresis.

Initial characterization consisted of a comparison of VL30 insert sizes, also a preliminary restriction enzyme analysis with one or two enzymes. This was to gain an idea of the level of heterogeneity displayed by the 12 clones.

It soon became obvious that despite using two different batches of virus as a source of *in vitro*-synthesized DNA, only 4 different classes of VL30 clone could be distinguished. The range of insert size is shown in figure R-15. Section (a) is an ethidium bromide-stained agarose gel. Section (b) is a Southern blot of a similar gel that has been loaded with much less VL30 DNA (amounts of DNA visible after staining would grossly overload a Southern blot). Here the 4 clones were run in parallel with uncloned *in vitro* VL30 DNA. Figure R-15 part (c) demonstrates that the VL30 clones were representative of this uncloned *in vitro*-synthesized VL30 DNA, since common *Sac*I restriction fragments were detected by this Southern blot analysis. A similar digestion profile of uncloned VL30 DNA cleaved by *Sac*I may be seen in figure R-11 (ii) lane (e). These data show that despite the rather limited range of heterogeneity displayed by the 12 VL30 clones, this was reflected in cDNA synthesized in the virion. In other words selection for a sub-set of VL30 cDNA had not occurred during the manipulations required for molecular cloning.

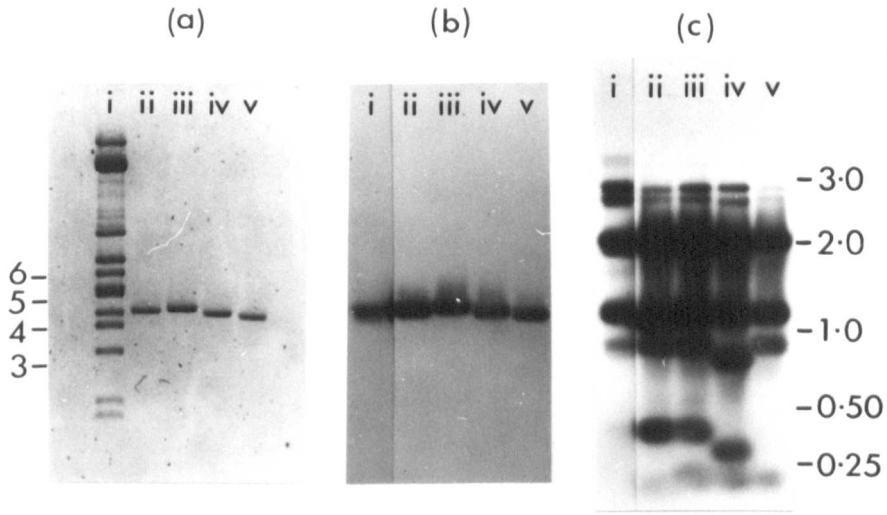
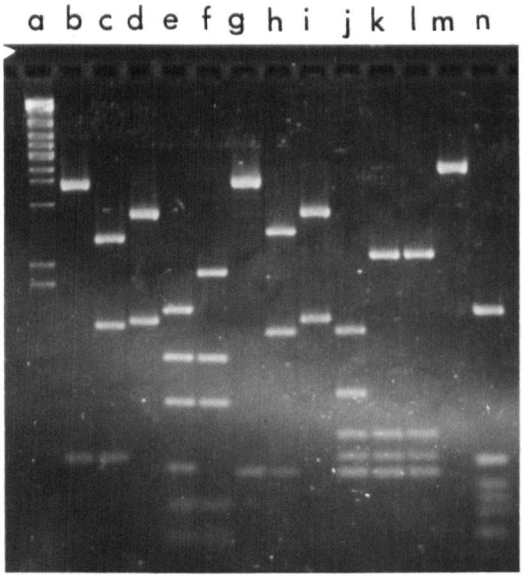
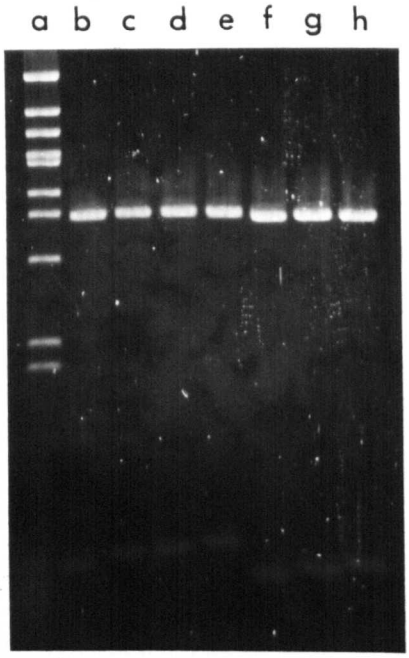


Figure R-15 Characterization of VL30 cDNA clones. (a) Ethidium bromide-stained 0.6% agarose gel visualized using UV light. Lane (i), λ.Eco RI plus λ.Hind III restriction fragments. Lanes (ii) to (v) undigested VL30 clones NVL-1 to NVL-4, respectively. (b) Southern blot of a similar gel; lane (i) in this case contains uncloned VL30 cDNA for comparison. (c) Southern blot of Sac I restriction fragments of the samples shown in (b), following electrophoresis through a 1.5% agarose gel. Both (b) and (c) were hybridized with ^{32}P-labelled VL30 clone NVL-3 DNA probe. Sizes of DNA species are given in kilobase pairs.

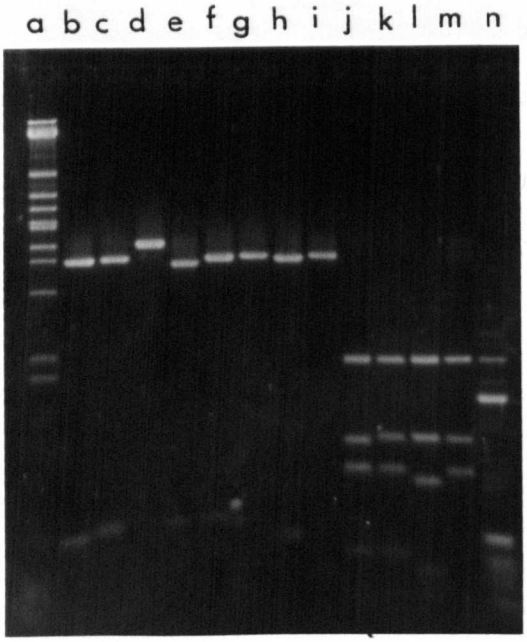
A



B



C



D

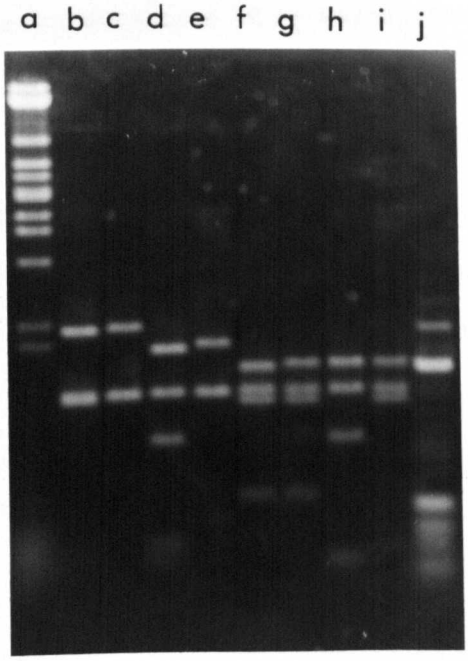


Figure R-16 Restriction mapping of VL30 clones. 1μg quantities of gel-purified VL30 DNA inserts were digested to completion with the appropriate restriction endonuclease, fractionated on a 1.5% (A), 0.8% (B) or 1.0% (C,D) (w/v) agarose gel, and visualized using UV light after staining with ethidium bromide (1μg ml^{-1}). Gel A, NVL-3 DNA digested with b, Eco RI; c, Pst I + Eco RI; d, Pst I; e, Pst I + Sac I; f, Sac I; g, Sma I; h, Sma I + Pst I; i, Pst I; j, Pst I + Msp I; k, Msp I; l, Msp I + Sal I; m, Sal I. Gel B, NVL-2 DNA digested with b, Sma I; c, Xho I; d, Xba I; e, Eco RI; and NVL-3 DNA digested with f, Sma I; g, Xho I and h, Eco RI; note small fragments. Gel C, NVL-1 to NVL-4 DNA, respectively, digested with b-e, Xba I; f-i, Xho I, j-m, Sac I. Gel D, NVL-1 to NVL-4 DNA, respectively, digested with b-e, Hind III, and f-i, Hind III + Xho I. Size estimation was by reference to DNA marker fragments; Lanes a, all gels, λ.Eco RI plus λ.Hind III restriction fragments; Lane A(n), C(n) and D(j), pBR322.Hinf I restriction fragments.

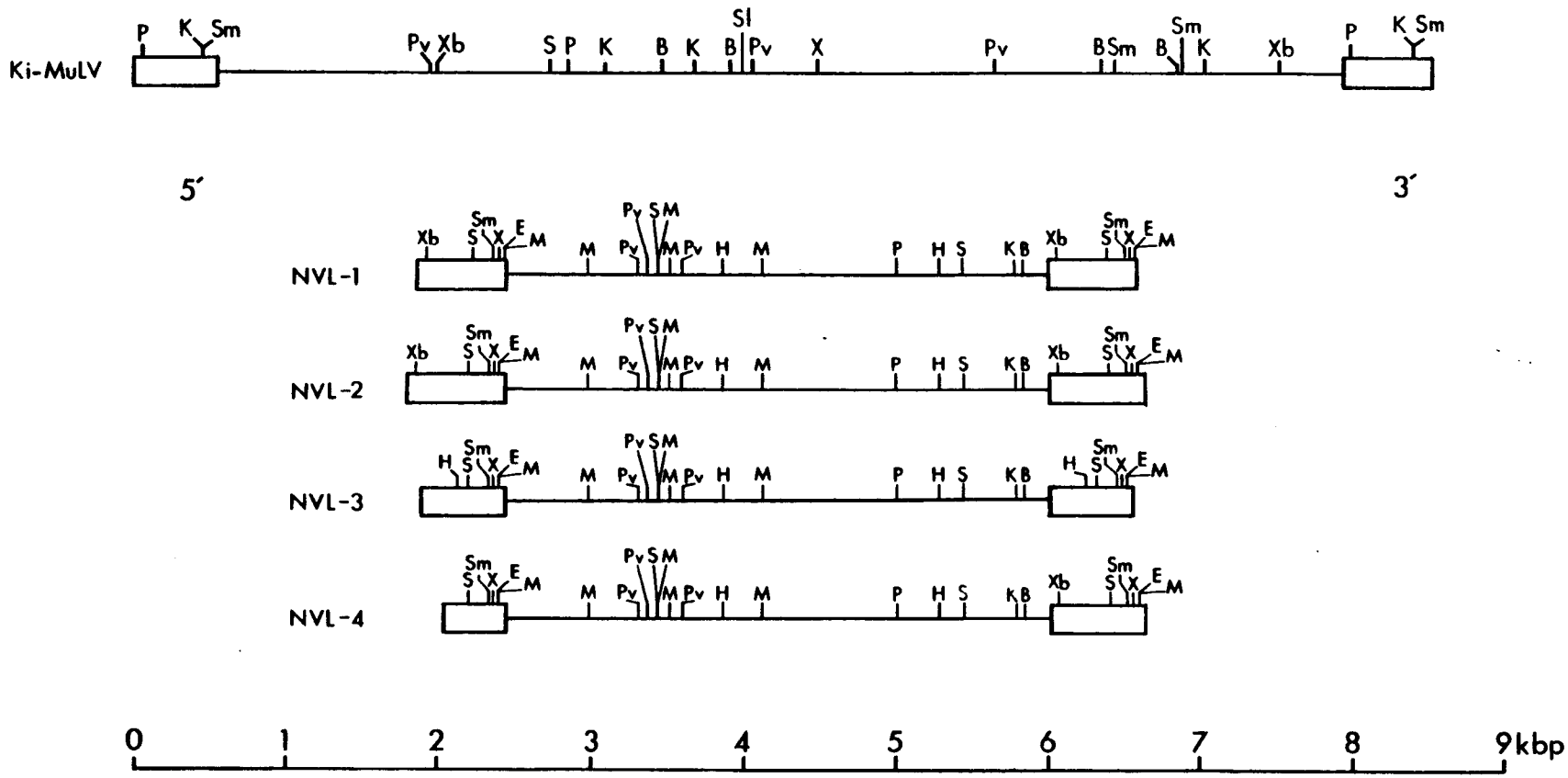

Ki-MuLV
P K Sm
Pv Xb
S P K
B K B Pv
Sl
X
Pv
B Sm
B Sm K
Xb
P K Sm
5′
3′
NVL-1
NVL-2
NVL-3
NVL-4
Xb S Sm X E M
M Pv Pv M Pv H M
P H S K B
0 1 2 3 4 5 6 7 8 9kbp

Figure R-17 Comparison of Ki-MuLV DNA and NVL DNA restriction maps. Restriction enzyme abbreviations used: Xb, Xba I; S, Sac I (Sst I); Sm, Sma I; X, Xho I; E, Eco RI; M, Msp I; Pv, Pvu II; H, Hind III; P, Pst I; K, Kpn I; B, Bgl II, Sl, Sal I. Boxed areas indicate LTR sequences. Enzymes Bam HI and Sal I do not cleave NVL clone DNA. Sites for Msp I and Bgl II, not determined for Ki-MuLV DNA; enzymes Hind III and Eco RI do not cleave Ki-MuLV DNA.

R-3.6 Restriction mapping

All mapping experiments were performed with unlabelled DNA visualized by ethidium bromide-staining and UV fluorescence. VL30 inserts were released from plasmid DNA and purified by preparative gel electrophoresis. Each DNA preparation was first cleaved by a number of different enzymes (one per digest) and the number of fragments produced was noted. The positions of cleavage sites were determined by double digestion experiments as described in Methods (section 15). Using a variety of different enzyme combinations an unambiguous physical map of each clone was constructed. The gels shown in figure R-16 are a representative selection of those used in mapping experiments - see the legends for details.

As indicated by initial studies of VL30 insert size and primary restriction digest profiles, only four classes of VL30 clones could be identified. The maps of these classes are presented in figure R-17. The 5' and 3' orientation of the clones was determined from the nucleotide sequence of the NVL-3 LTRs, as discussed in section D-3 (Norton et al., 1984b). Clones were designated "NVL" after NIH VL30. None of the NVL clones bore any resemblance to the physical map of Ki-MuLV or indeed to any retrovirus restriction map that had been published. Of the 12 clones isolated, 8 were of the type designated NVL-3, 2 were of the NVL-2 type and NVL-1 and NVL-4 were each represented once.

From the relative position of restriction sites at each end of the clones it was apparent that all possessed long terminal repeats (LTRs). These have been indicated in figure R-17 by open boxes. At this level of analysis it was also probable that the unique region of the clones was identical in each case. The only detectable differences between the four classes of VL30, therefore, were found in their LTRs.

VL30 LTRs varied in length from approximately 550bp to 630bp. As shown in figure R-17, NVL-4 possessed an incomplete LTR at one end of the molecule; this probably arose as a consequence of aberrant reverse transcription. Such molecules can be predicted to occur from a consideration of models of retrovirus reverse transcription, and have been more fully discussed in the General Introduction.

R-3.7 Summary

The data obtained so far indicated that:-

(1) VL30 RNA could direct the reverse transcription *in vivo* of dsDNA, which by its linear and circular forms and its distribution in newly infected mouse cells was exactly analogous to the unintegrated DNA of retroviruses.

(2) As with retroviruses, VL30 RNA which had been packaged into a replication-competent virus particle could also be reverse-transcribed into full-length DNA in an endogenous reaction.

(3) Cloned examples of this *in vitro*-synthesized VL30 DNA were again found to be retrovirus-like by their possession of LTRs. These LTRs were of a similar length to those of retrovirus proviruses, and presumably were formed by a similar mechanism of reverse transcription.

(4) From 12 independently isolated VL30 clones only four types of retrovirus-transmissible VL30 could be identified. Areas of sequence heterogeneity were confined to the LTRs, whereas the unique region which is flanked by these structures was common to all clones.

(5) Clewley & Avery (1982) had characterized the VL30 RNA of A1 tissue culture fluid virus particles using oligonucleotide fingerprinting. The limited sequence heterogeneity displayed by the NVL clones confirmed the findings of these authors. Together, these results

suggested that Ki-MuLV-packaged VL30 RNA represents a closely-related family of molecules. The experiments described in the following sections of this thesis examine the organization of the VL30 genes encoding these RNA species.

R-4 Organization of VL30 elements in mouse genomic DNA

R-4.1 Introduction

Proviruses which are transmitted vertically as a result of their integration into germ line DNA are known as endogenous retroviruses. These have been studied most intensively in mice and chickens (see Steffen & Robinson, 1982 for a review). Whereas those of chickens constitute a closely related family of avian leukosis virus (ALV)-type proviruses (Hughes *et al.*, 1981b) mouse DNA contains at least four unrelated classes of retrovirus-like genetic elements. Type A particle sequences are present at up to 1000 copies per haploid genome (Leuders & Kuff, 1977) type B sequences, represented by MMTV, have less than 10 copies (Varmus *et al.*, 1972) and MuLV-related proviruses may number up to 50 copies per haploid genome (Coffin, 1982).

This section describes the results of a detailed study of the fourth class of provirus-like information - the VL30 family. Initial reports by other workers show that all species of the genus *Mus* so far examined possess some VL30 elements in their genomes (Courtney *et al.*, 1982b; Itin *et al.*, 1983) and in *Mus musculus* DNA these are reported to number 100-200 copies (Keshet & Itin, 1982). In the present work NVL-3 derived probes were used to study VL30 elements in NIH-3T3 chromosomal DNA. Could the VL30 family be ordered into sub-classes on the basis of restriction enzyme sites, as done for ecotropic and xenotropic MuLVs (Chan *et al.*, 1980; Hoggan *et al.*, 1982)? Also of particular interest were those VL30 elements which encode NVL-type RNA. If preliminary evidence which suggested that these were a small subset of the VL30 family was correct, this difference might be detected in a structural analysis of genomic DNA VL30 elements.

R-4.2 VL30-related restriction fragments of mouse DNA

As detailed in the restriction maps presented in figure R-17, at

this level of analysis there were few structural differences between the NVL clones. Since NVL-3 was the most frequently isolated example of VL30 cDNA (8 out of 12) this was taken as the most representative VL30 species, and was used as a nick-translated probe for Southern blotting studies.

In each experiment, 10μg quantities of high molecular weight mouse DNA were digested to completion with an excess of restriction enzyme, electrophoresed through agarose gels and transferred to nitrocellulose by Southern blotting. Three results are shown in figure R-18. In each case the nitrocellulose filter was hybridized with a ^{32}P-labelled NVL-3 DNA probe and used to produce an autoradiogram. Apart from generating preliminary restriction fragment data from NIH-3T3 DNA, these experiments also compared restriction digest profiles of this uninfected mouse DNA with that of the A1 cell line. As stated previously, this cell line was derived from an NIH-3T3 sub-clone which had been productively infected by Ki-MuLV. It was of interest to determine whether establishment of the A1 cell line had altered the genetic organization of the VL30 gene family.

In the examples given in sections (i) and (ii), picogramme quantities of clone NVL-3 DNA were run in parallel with the mouse DNA digests. This technique aided the interpretation of Southern blotting data greatly. Several observations could be made from these preliminary results:-

(1) The hexanucleotide-recognizing enzymes used here generated 20 to 40 restriction fragments, each of different intensity, which ranged in size from >30kbp to approximately 0.3kbp. This size range represented the limits of detection afforded by the Southern blotting technique.

(2) In several comparisons of NIH-3T3 and A1 DNA digests no differences in restriction digest profile could be detected. This suggested that no gross changes in VL30 gene organization had occurred during or after

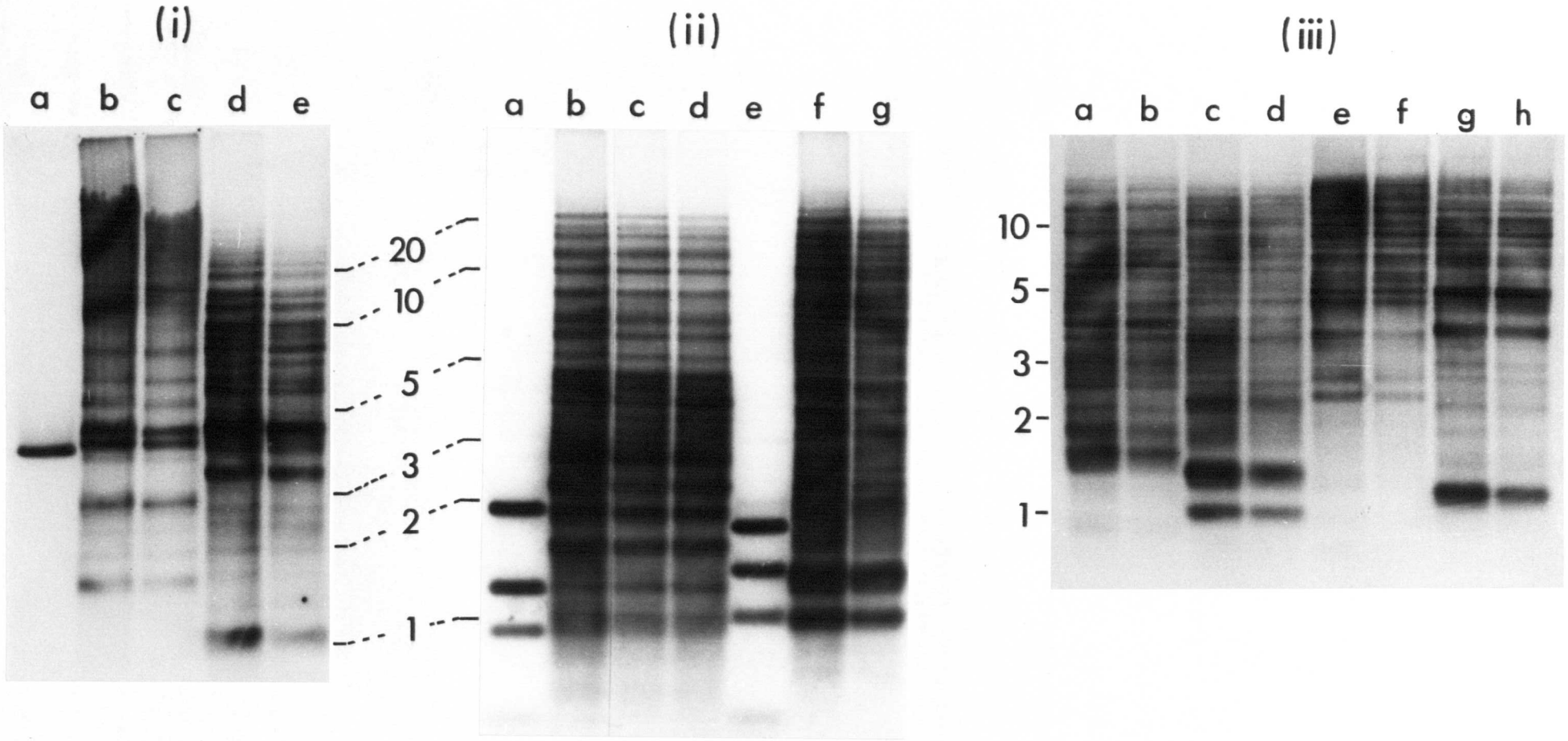

(i)
a b c d e
20
10
5
3
2
1
(ii)
a b c d e f g
(iii)
a b c d e f g h
10
5
3
2
1

Figure R-18 Primary digests of NIH-3T3 and A1 cell high molecular weight DNA. 10μg samples of DNA were digested to completion with the appropriate rectriction enzyme, electrophoresed through a 0.6% (i), 0.8% (ii) or 1.0% (iii) (w/v) agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled NVL-3 DNA. Gel (i) NIH-3T3 and A1 DNA, respectively, digested with b, c, Xho I and d, e, Xba I. Gel (ii) NIH-3T3 DNA, lanes b and f and A1 DNA, lanes c, d and g, digested with Sst I (b,c) Sac I (d) or Hind III (f,g). Gel (iii) NIH-3T3 and A1 DNA, respectively, digested with a, b, Pst I; c, d, Hind III; e, f, Eco RI and g, h, Xba I. Size estimation (in kbp) was by comparison with λ.Eco RI plus λ.Hind III restriction fragments (not shown) and with 275pg quantities of NVL-3 DNA digested with Xho I [(i)a], Sst I [(ii)a] or Hind III [(ii)e], each added to 10μg of Eco RI-digested calf thymus DNA before loading.

productive infection by Ki-MuLV.

(3) Rough estimates of VL30 copy number in the mouse haploid genome were attempted. This was done by comparing band intensity of similar-sized genomic DNA fragments and NVL-3 DNA marker fragments (see Methods section 22). Intense fragments larger than about 10kbp were assigned one complete copy of a VL30 unit, and fragments smaller than 4-5kbp were considered to derive from VL30 units that bore internal recognition sites for the enzyme. Assuming all VL30-related bands were 100% homologous to the probe, and that all fragments had blotted with equal efficiency, then copy number estimates from various digestion patterns ranged from 80 to 150 complete VL30 units per haploid genome. This agreed closely with the results of Keshet & Itin (1982) who detected VL30 clones in a mouse genomic DNA library at a frequency which suggested that VL30 elements are reiterated 100-200 times in the mouse genome. These investigators also used a Cot analysis to compare IAPs, VL30 genes and MuLV proviruses in mouse DNA, arriving at 17 copies for MuLVs, 150 copies of VL30, and 1200 copies of IAP sequences.

(4) Most restriction enzyme digestion profiles included a few high copy number bands (ie with more intense autoradiographic signal). These more intense bands were detected in the <4.5kbp size range, and their presence suggested the possibility of highly conserved restriction site positions within VL30 elements. However at this stage it was not known whether some of these high copy fragments represented VL30 termini which were contiguous with equally conserved cellular DNA flanking sequence.

(5) The enzyme BamHI produced a digestion profile where the majority of fragments were greater than 5kbp in length (see figure R-25A, lane R, where a similar result is shown with woodmouse DNA). Since those VL30 units which have been cloned from mouse DNA (Keshet & Itin, 1982; Hodgson *et al.*, 1983), in addition to the NVL cDNA clones are all in the size range 4.6-5.2kbp, this result confirmed that most mouse VL30

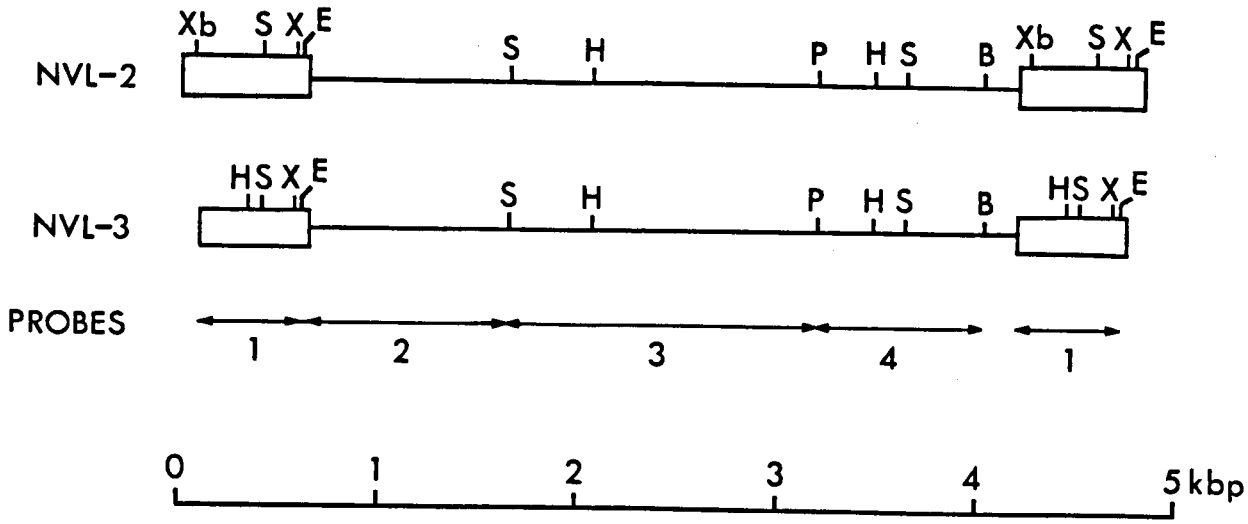


Figure R-19 Sub-genomic VL30 DNA probes. Areas of clone NVL-3 used for sub-genomic probes 1 to 4 are delimited by arrowheads under the restriction map. Restriction enzyme abbreviations used: Xb, Xba I; S, Sac I; X, Xho, I; E, Eco RI; H, Hind III; P, Pst I; B, Bgl II. Boxed areas indicate LTR sequences.

elements lacked internal BamHI sites. This fact should prove useful in the isolation of genomic DNA VL30 clones. A similar result was obtained using EcoRI (figure R-18 part iii). In this case the reduction in signal below about 5kbp was more gradual, indicating the existence of some VL30 members which possessed internal recognition sites for this enzyme.

(6) SstI- and HindIII-digested NVL-3 marker fragments appeared to co-migrate with multi-copy bands in the corresponding genomic DNA tracks (figure R-18 part ii). Relative intensity was different in each case, however, suggesting that if restriction sites in NVL clones were shared with some genomic DNA VL30 units, they were not shared equally amongst these species.

R-4.3 Sub-genomic VL30 probes

Four sub-genomic VL30 probes were used in Southern blotting analyses of mouse genomic DNA. These were obtained by nick-translation of NVL-3 DNA restriction fragments, rigorously purified by successive rounds of preparative agarose gel electrophoresis. Probes 1-4 covered the entire genome of NVL-3 apart from 160bp at the 3' end of the unique sequence (figure R-19). This "missing sequence" did not significantly alter the results obtained from the Southern analyses, since comparison of hybridization patterns using probe 4 with those of a fifth probe, homologous to DNA extending between the PstI site and the 3' XbaI site of NVL-2, revealed no detectable differences. Results obtained with probe 5 are therefore not presented.

Each experiment was constructed as follows. Identical restriction digests of NIH-3T3 DNA were run in parallel tracks of an agarose gel and Southern blotted. Each track was hybridized separately to one of probes 1-4; later one of the filter strips was re-probed with total ^{32}P-labelled NVL-3 DNA as a control. Autoradiographs were aligned using

this control together with NVL-3 DNA marker fragments which had been run in alternate tracks of the original gel (see figure R-21A, where these markers have been retained).

The results of three mapping experiments are shown in figure R-20. Data from these (and other experiments not shown) strongly suggested that, (a) the majority of NIH-3T3 mouse VL30 units were organized into provirus-like units possessing unique sequence bounded by LTRs, and (b) the position of many restriction enzyme sites had been conserved within these genetic elements. The more intense bands produced by these sites, therefore, were initially the most useful source of data for interpretation.

The enzyme HindIII (part A) generated two such bands, one of 1.2-1.4kbp (henceforth referred to as 1.3kbp) and one of 1.0kbp. PstI (part B) produced a major band of 1.4-1.5kbp, whilst SacI generated several high copy restriction fragments (part C). By comparing the relative intensity of signal with each probe, such fragments could be roughly localized to various areas of a generalized VL30 physical map. For instance an examination of figure R-20A lanes 3 and 4 indicated that in well over 50% of mouse VL30 units, the 1.3kbp HindIII fragment mapped to the central part of the VL30 unique sequence. This fragment was similar, therefore, to the central 1.4kbp HindIII fragment of the NVL clones (figure R-17). The less intense 1.0kbp HindIII band mapped chiefly to an area 3' of this fragment, suggesting the presence of a conserved recognition site in the 3' LTR (a characteristic of clone NVL-3). Signal with probes 2 and 3 also revealed a sub-set of VL30 units whose physical map would include a 1kbp fragment to the 5' of the central 1.3kbp one. Interestingly, a high copy band indicating the presence of a conserved 5' LTR site could not be detected. The significance of this will be discussed later.

Similar interpretation of the autoradiographs shown in figure R-20

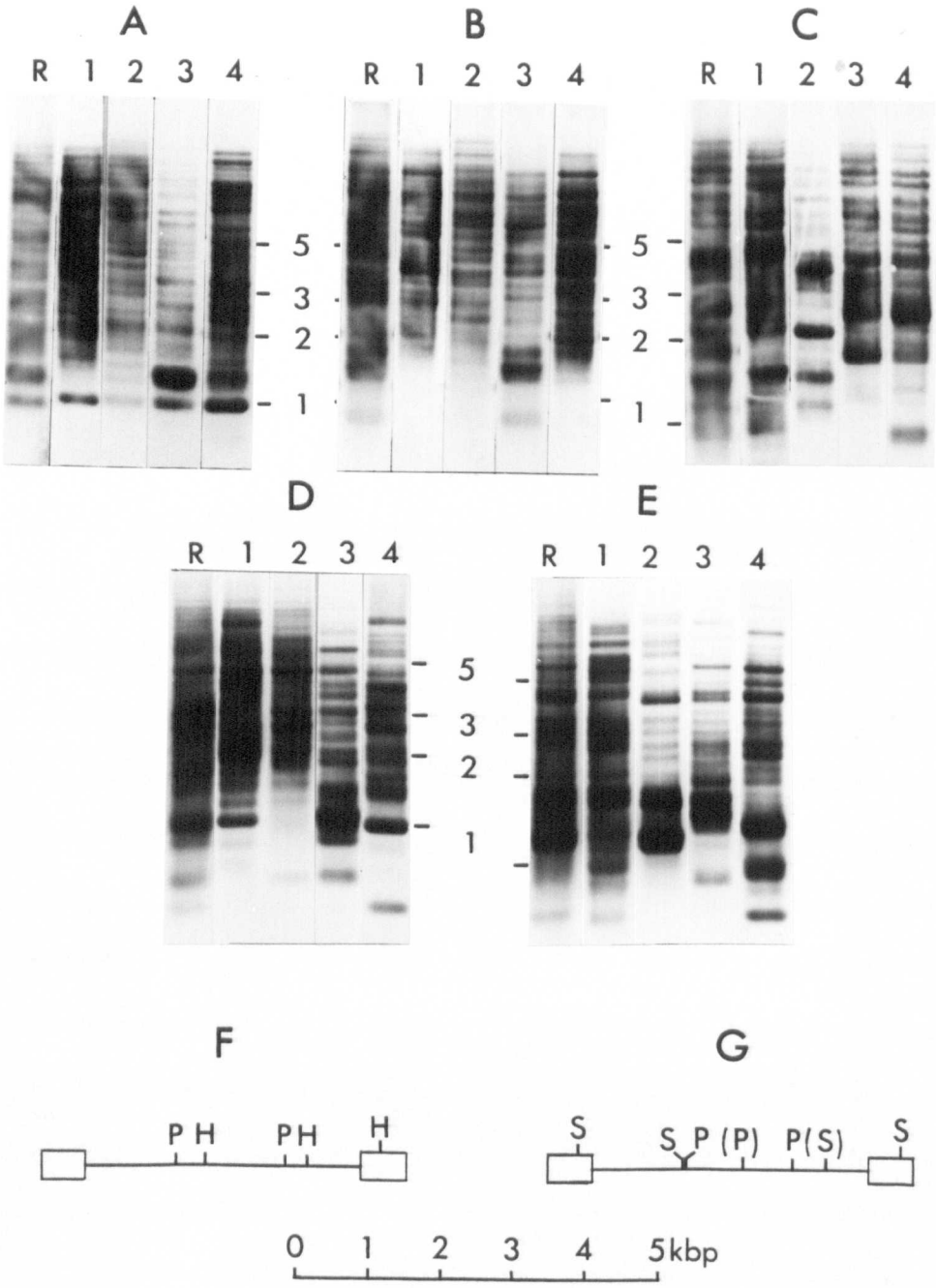

A
R 1 2 3 4
5
3
2
1
B
R 1 2 3 4
5
3
2
1
C
R 1 2 3 4
D
R 1 2 3 4
5
3
2
1
E
R 1 2 3 4
F
P H
P H
H
G
S
S P (P)
P (S)
S
0 1 2 3 4 5kbp

Figure R-20 Sub-genomic probe analysis of VL30 clone NVL-3-related sequences in NIH-3T3 DNA. 10μg DNA samples were digested to completion with a single (A-C) or two sequentially added (D,E) restriction enzymes and fractionated on 1.0% (w/v) agarose gels. Following transfer of restriction fragments to nitrocellulose, each track was separately hybridized with the ^{32}P-labelled representative NVL-3 probe (R) or one of the sub-genomic probes (1 to 4). A, Hind III; B, Pst I; C, Sac I; D, Hind III + Pst I; E, Sac I + Pst I. F, G, an interpretation of Hind III + Pst I and Sac I + Pst I mapping data, respectively. Sizes of DNA species are indicated in kilobase pairs.

parts B and C enabled conserved PstI and SacI restriction fragments to be approximately mapped. Not all of these high copy bands, however, could be accounted for. For example a 2.5kbp SacI band (see part C, lane 2) hybridized exclusively with probes 1 and 2. Significantly, these probes cover a contiguous stretch of the NVL-3 map which spans only 1.5kbp. Normally a band of this size which hybridized to probes 1 and 2 would also be detected by probe 3. This irregularity in the SacI data is discussed later.

Relative positions of the conserved HindIII, PstI and SacI restriction enzyme sites were mapped using double digestion experiments. The simplest interpretation of the (HindIII + PstI) and the (SacI + PstI) digestions (parts D and E) are diagrammatically interpreted in parts F and G. The reasoning behind the HindIII + PstI map was as follows.

As described above, a 1.3kbp HindIII fragment (part A) mapped with probes 3 and 4, thus demonstrating its central position. PstI digestion reduced this to a fragment of about 1.1kbp which now hybridized to probe 3 only (part D lane 3). Thus one of the conserved PstI sites mapped near the more 3' HindIII site. An approximately 300bp fragment (part D lane 4) was the expected smaller product of this cleavage. This more finely mapped the position of the PstI site with respect to the HindIII site. Since the single PstI site of cloned NVL elements delineates the border between probes 3 and 4 (figure R-19), these data signified that the same PstI site plus its surrounding sequences are conserved in the majority of mouse VL30 units. Furthermore, the large genomic DNA VL30 PstI fragment was cleaved at the more 5' HindIII site to produce an approximately 400bp digestion product (lanes 2 and 3). Apart from the fact that the NVL map lacked the second PstI site, this suggested that much of the central part of the VL30 unique region showed the same gross sequence organization that was exemplified in the NVL clones.

Interpretation of PstI + SacI digestion profiles was complicated by the ambiguous SacI primary digestion data. However the majority of VL30 units possessed at least some of the sites indicated in figure R-20 part G. SacI sites seemed to be (a) more numerous among VL30 units and (b) more variable in their positioning (at least in the VL30 unique sequence). For example lane 4 (part E) shows three major bands at approximately 1.45, 0.95 and 0.45kbp. These bands revealed at least two types of VL30 3' end; those with a 1.45kbp fragment extending from the 3' unique sequence PstI site through to a 3' LTR SacI site, and those units possessing an extra SacI site (bracketed in figure R-20G) which generated the two smaller restriction fragments.

A doublet of high copy 1.8 and 1.9kbp bands was similarly interpreted to represent two different types of VL30 5' end. As shown in part E lanes 1-3, the 1.8kbp band was the more conventional by its hybridization with probes 1, 2 and 3. The larger fragment, however, mapped only with probes 1 and 2. It seemed likely that this fragment could be a PstI digestion product of the anomalous 2.5kbp SacI band described above. The structure of a VL30 sub-set which would display these hybridization characteristics will be more fully discussed later (section D-4).

R-4.4 Analysis of isolated high copy VL30 restriction fragments

Several bands identified as high copy VL30 restriction fragments from the mapping experiments described above were selected for further structural analysis. Specific size fractions of preparative scale digests of NIH-3T3 DNA were purified from agarose and used in Southern blotting experiments. The most informative data were obtained from a study of XbaI digestion products.

As shown previously in figure R-18 parts (i) and (iii), digestion of NIH-3T3 DNA with this enzyme produced a restriction fragment profile

which was dominated by three intense bands. These can be seen once more in figure R-21 part A lane R, where they are indicated by their size in kbp. From the mapping data (lanes 1 to 4) two major VL30 sub-groups could be identified. Group A members possessed XbaI sites only in the LTR, and were characterized by the 4.7kbp band. Group B members shared these two sites but in addition were cleaved once more in the 5' part of the unique sequence. This extra site produced the two smaller bands seen in part A, a 5' 1.15kbp and a 3' 3.45kbp fragment. By relative band intensity estimations the two Xba groups A and B comprised over two thirds of the total mouse VL30 complement.

In figure R-21 parts B and C are presented the results of a restriction enzyme analysis of gel-purified 4.7kbp, 3.45kbp and 1.15kbp high copy VL30 XbaI fragments. Each Southern blot included undigested DNA run in parallel with a sample of XbaI-digested mouse DNA. This was to ensure no degradation of the samples had occurred during gel purification, and that the correct size fraction had been excised from the preparative scale restriction digest.

The 4.7kbp fragment preparation in part B can be seen to include some smaller DNA species. These were not excluded from the experiment as they also shared the mapping characteristics of the larger fragment, and were therefore regarded as examples of a minority VL30 group whose members were slightly smaller than the standard 5.2 to 5.3kbp size. The latter figure could be calculated by first assuming that, as with VL30 clones isolated from genomic DNA libraries (Hodgson et al., 1983) in addition to the NVL cDNA examples, members of VL30 groups A and B each possessed only one XbaI site per LTR. Thus members of group A would be 4.7kbp long, plus the equivalent of one LTR of sequence (about 600bp) = 5.3kbp. Group B members would be (3.45 + 1.15 + 0.6) = 5.20kbp long. The smaller size class, which included NVL-sized units (note the co-migration of the largest marker fragment in figure R-21 part B, lane d)

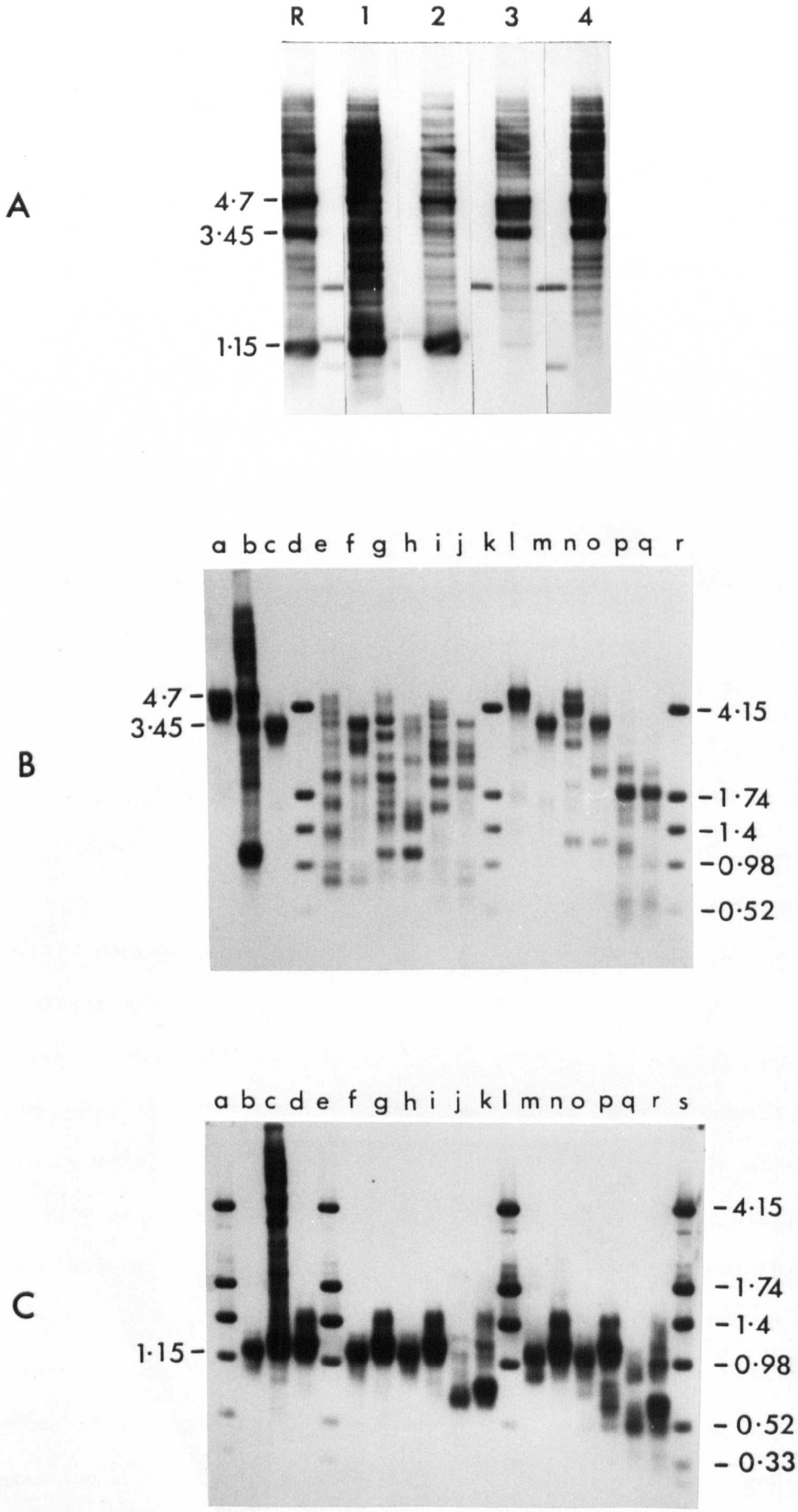
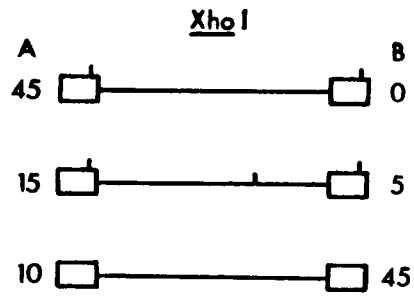
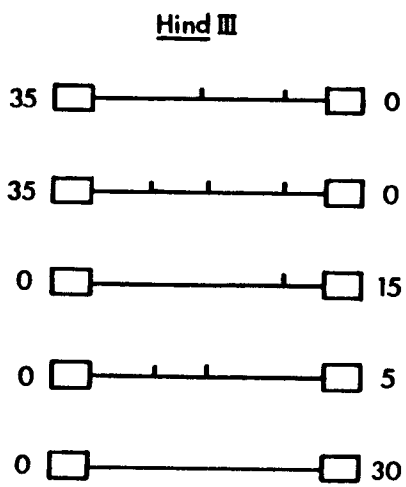
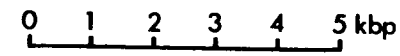
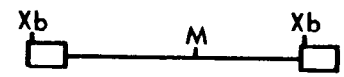

A
R 1 2 3 4
4·7
3·45
1·15
B
a b c d e f g h i j k l m n o p q r
4·7
3·45
4·15
1·74
1·4
0·98
0·52
C
a b c d e f g h i j k l m n o p q r s
1·15
4·15
1·74
1·4
0·98
0·52
0·33

Figure R-21 Analysis of VL30 Xba I sub-groups in NIH-3T3 DNA. A, characterization by sub-genomic probes of Xba group A (4.70kbp band) and group B (3.45kbp and 1.15kbp bands). See legend of fig. R-20 for explanation. Marker fragments in intervening lanes are 275pg quantities of NVL-3 DNA digested with Sac I. B, 4.70kbp and 3.45kbp NIH-3T3 DNA Xba I fragments were gel-purified, digested with the appropriate restriction enzyme, electrophoresed through a 1.0% (w/v) agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled NVL-3 DNA. a, undigested 4.70kbp fragment; b, 5µg Xba I-digested NIH-3T3 DNA; c, undigested 3.45kbp fragment; d, k and r 250pg each of NVL-3 DNA. Eco RI and Hind III markers. The following pairs of tracks are digests of the 4.70kbp and 3.45kbp fragments, respectively; e,f, Hind III; g, h, Pst I; i, j, Sac I; l, m, Eco RI; n, o, Xho I; p, q, Msp I. All samples except b were loaded with 5µg Eco RI-digested calf thymus DNA. C, similar analysis of two size fractions of 1.15kbp NIH-3T3 DNA Xba I fragments (undigested in lanes b and d). Enzyme-digested DNAs were loaded in the same order as in gel B. Lanes a, e, l and s, Eco RI and Hind III NVL-3 DNA markers. Sizes of DNA species are given in kilobase pairs.

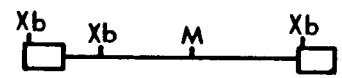




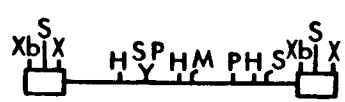




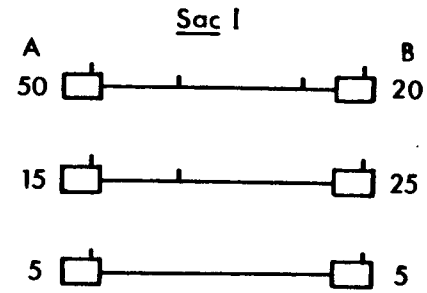
GROUP A : 70



GROUP B : 50



Conserved sites



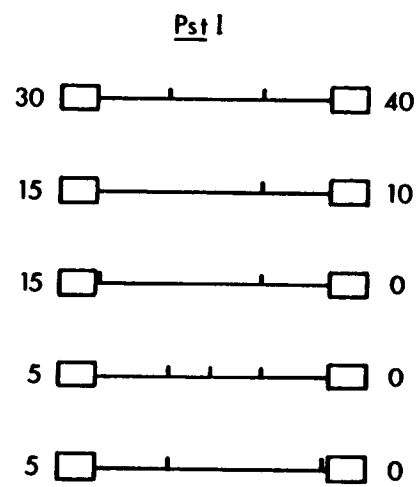


Figure R-22 Restriction maps of VL30 Xba I subgroups in NIH-3T3 DNA. Interpretation of the restriction mapping data presented in fig. R-21 suggests that 70 copies per haploid genome of Xba group A members and 50 copies of Xba group B members (centre) fall into further subdivisions as indicated under each restriction enzyme heading. Members of group A are indicated at the left hand side of each map, members of group B are indicated at the right hand side of each map. The most conserved restriction enzyme sites for both Xba groups are shown in the restriction map at bottom centre. The scale indicates length of DNA in kilobase pairs.

ranged from about 4.7kbp to 5.0kbp.

The data shown in part B indicated that some restriction fragments were unique to either group A or B, while others were common to both. The Southern analysis depicted in figure R-21C was designed to reveal conserved restriction enzyme sites within the 5'-located 1.15kbp XbaI fragments of the group B VL30 units. Two different DNA preparations were compared. One contained the upper half of the 1.15kbp band (which actually extends from about 1.0-1.2kbp) the other represented the smaller size fraction (compare tracks b and d). Digestion with SacI reduced each preparation by 0.35-0.45kbp, ie to about 0.75kbp and 0.65kbp, respectively. This result was consistent with those of the mapping studies described in the previous section, which revealed a highly conserved VL30 LTR SacI site. It also exemplified the size variation (ie plus or minus 100bp) displayed by virtually all high copy VL30 DNA restriction fragments.

The restriction maps of figure R-22 represent the simplest interpretation of the XbaI data. Arbitrary copy number estimates are intended merely to indicate the general relative abundance of each VL30 sub-set. Although full-sized members numbered about 50 for each group, group A included an extra 20 copies (approximately) of the smaller, NVL-sized units. Several observations could be made from the maps.

(1) Unique region variation HindIII data showed that all members of group A shared the two most conserved sites which generated the well-characterized 1.3kbp fragment. However group B members lacked either one or the other of these sites (see figure R-21B lanes e and f). In contrast the two most frequently observed PstI restriction site patterns, plus a central MspI site, were equally represented in both groups. Thus restriction site differences were not confined to a single part of the unique VL30 sequence. Rather, they were scattered throughout the unique region.

(2) LTR region variation Three enzymes were found to possess conserved recognition sites in many genomic DNA VL30 LTRs. XhoI maps to the NVL U_5 region (Norton et al., 1984b) and similarly cleaved most group A VL30 units. In comparison relatively few group B (5') LTRs were digested by this enzyme (see figure R-21C, lanes o and p). SacI, which in NVL clones and the BALB/c clone BVL-1 (Hodgson et al., 1983) maps to the R region of the LTR, cleaved all XbaI-defined VL30 group members at a similar location.

Although it is difficult to make even a semi-quantitative assessment of Southern blotting data, these results tend to suggest that mouse VL30 LTR restriction sites are rather more conserved that those located in the unique region. Within the latter, however, there was no evidence for an obvious mutational hot-spot; restriction site polymorphism, in other words, was not localized. This implied that the restriction fragment heterogeneity observed in total mouse VL30 DNA Southern analyses is probably the result of relatively small sequence changes which, apart perhaps from the LTR regions, are randomly scattered throughout a family of otherwise closely related members.

(3) NVL-type VL30 units The Southern blotting experiments described above failed to detect a major VL30 sub-set which was exactly like the NVL cDNA clones in both size and restriction enzyme sites. Apart from their significantly smaller size than the majority of mouse VL30 units, NVL clones differed mostly by virtue of their LTR restriction sites. NVL-3 was conspicuously different by its lack of the highly conserved LTR XbaI site. Other NVL clones have retained these sites but possess an EcoRI site in the U_5 region of the LTR (Norton et al., 1984b). As revealed by sub-genomic probing of EcoRI-digested mouse DNA, (data not shown) and as seen in figure R-21B and C, VL30 units with these LTR sites must only account for a small fraction of the total VL30 membership (an estimated <5% of approximately 150 VL30 units).

Returning to the NVL-3 LTR, this possesses a HindIII site in the U_3 region. Although this enzyme has been shown to cleave a significant number of VL30 3' LTRs (figure R-20) it did not seem to cut many 5' LTRs. The results of the 1.15kbp XbaI fragment analysis (figure R-21C) also showed no significant digestion by HindIII of group B 5' LTRs. In this case no evidence exists for HindIII sites in 3' LTRs of the Xba group members, as the XbaI site lies to the 5' of the HindIII site in NVL-3 (and presumably in most genomic VL30 units). These results implied (a) that several VL30 units in mouse DNA have different LTRs at each end of their unique sequence, and (b) that very few VL30 members (perhaps less than 5 copies) could possess an NVL-3 type LTR. The significance of the atypical NVL-type LTRs is discussed later.

R-5 VL30-related sequences in different rodent genomes

R-5.1 Introduction

Based on studies of their distribution in feral subspecies of *Mus musculus*, the two families of endogenous xenotropic MuLV and VL30 proviruses appeared to be of similar evolutionary age (Steffen *et al.*, 1980; Dolberg *et al.*, 1981; Courtney *et al.*, 1982b; Itin *et al.*, 1983).

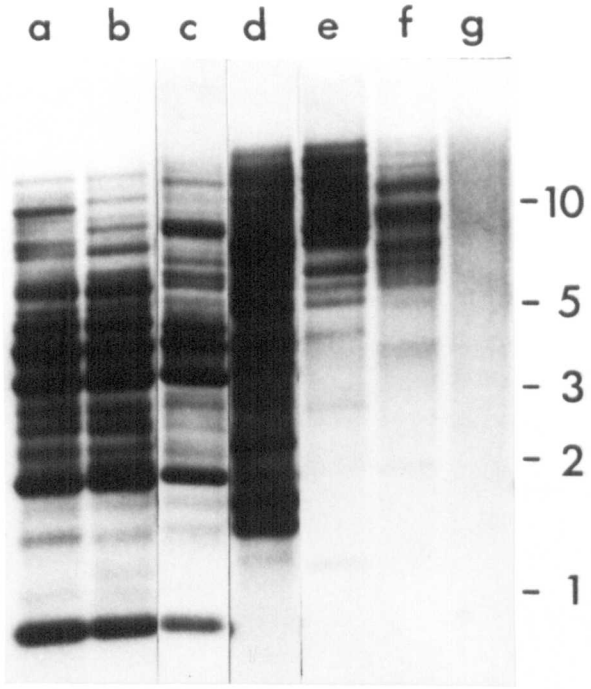
Using relaxed hybridization conditions Giri *et al.* (1983) detected a limited homology between a cloned VL30 element and cloned examples of MuLVs. It was suggested that this relationship could be explained by a putative recombination event (or series of events) between VL30 and MuLV progenitor sequences which must have occurred approximately 1-2 million years ago.

The experiments described in this section used several rodent genomic DNA samples in a survey for sequences related to cloned MuLV and NVL probes. In order to investigate further any relationship which may exist between the two classes of genetic element, these sequences were examined for differences in their degree of conservation and in their relative abundance.

R-5.2 MuLV and VL30-related sequences detected by Southern blotting

High molecular weight DNA from examples of different rodent genera was digested by *Sac*I, fractionated by electrophoresis through agarose gels and transferred to nitrocellulose. Blots were hybridized with ^{32}P-labelled NVL-3 and Ki-MuLV DNA probe, then washed at low stringency. The autoradiographs in figure R-23 show the range of rodent species whose genomic DNA contained Ki-MuLV or VL30-related sequences. Differences in distribution of the two provirus families were most evident in grey squirrel DNA (lanes f). Here the NVL probe generated a slight background smear of hybridization, whilst the MuLV probe revealed several discrete bands. Similar low level signal was obtained for human

A



B

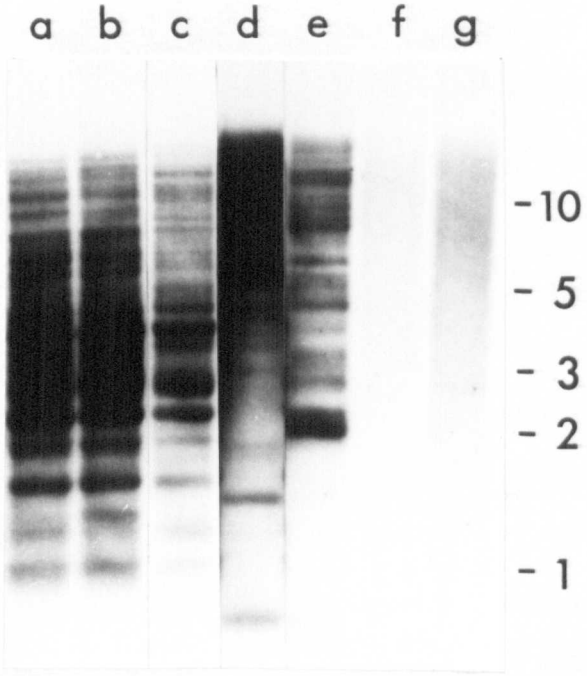


Figure R-23 Survey of rodent genomes for Ki-MuLV and VL30-related sequences. The autoradiographs are of Sac I-digested rodent DNA samples after fractionation on a 1.0% (w/v) agarose gel, transfer to nitrocellulose and hybridization to a ^{32}P-labelled Ki-MuLV DNA probe (A) or NVL-3 DNA probe (B). Following a room-temperature post-hybridization wash at 2 x SSC then 0.1 x SSC, lanes a-c (gel A) were exposed overnight, and lanes d-g of A, and all those of gel B were exposed for 10 days before development of the X-ray film. Lanes a, 2μg NIH-3T3; b, 2μg C3H 10T½; c, 2μg Apodemus sylvaticus (woodmouse); d, 10μg NRK; e, 10μg Clethrionomys glareolus (bank vole); f, 10μg Sciurus griseus (grey squirrel); g, 10μg human mixed lymphocyte DNA. Sizes of DNA species are given in kilobase pairs.

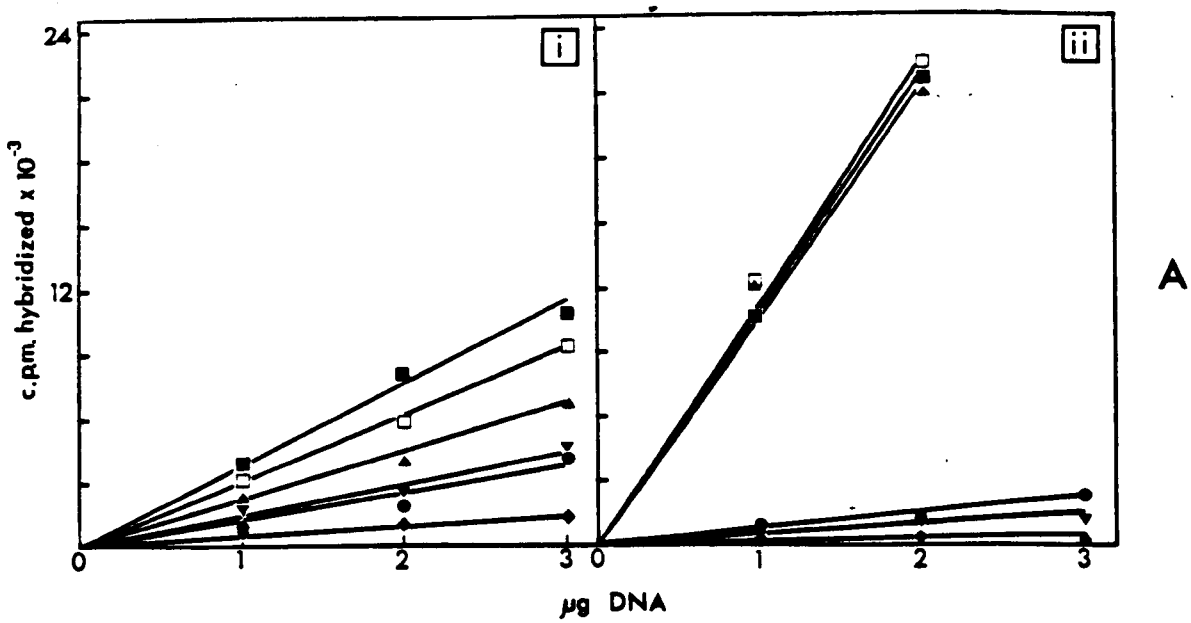
DNA with both probes (lanes g); this, together with the non-specific VL30 squirrel signal could be removed almost entirely by a 60°C wash, indicating a fairly distant homology in each case (data not shown).

For both probes, a comparison of lanes (a), (b) and (c) showed that two strains of Mus musculus DNA (NIH- and C3H-derived cell lines) and feral Apodemus sylvaticus (woodmouse) DNA exhibited slight differences in band profiles which were of the same order irrespective of which two lanes were compared. This suggested a close evolutionary relationship between these two mouse species. Signal with rat (NRK cells, Rattus norvegicus) and bank vole (feral Clethrionomys glareolus embryo cell line) DNAs was indicative of less abundant sequences with a lower homology to either probe, but the degree of such differences was difficult to assess. As pointed out in the figure legend, loadings of cellular DNA and autoradiograph exposure times have been adjusted in order to compensate for the wide range of signal intensity that was produced.

R-5.3 Comparison of MuLV- and VL30-related sequences by dot-blotting

This approach was adopted in order to provide a semi-quantitative analysis of the MuLV and VL30-related sequences detected by Southern blotting. The method was modified from Itin et al. (1983), and is detailed in Methods (M-17).

High molecular weight rodent DNA was denatured with alkali, neutralized and immobilized on 13mm diameter nitrocellulose filters. Duplicates of each DNA loading were hybridized with either the cloned Ki-MuLV probe or the NVL-3 probe. The moderately stringent conditions of hybridization originally used for the Southern blots (figure R-23) were used once more, as detailed in section M-17. Following hybridization all filters were subjected to a series of increasingly stringent washes at 60°C. Between each wash they were placed in 2 x SSC for determination of their radioactivity by scintillation counting.


24
12
c.p.m. hybridized x 10^-3
i
ii
0
1
2
3
0
1
2
3
µg DNA
A

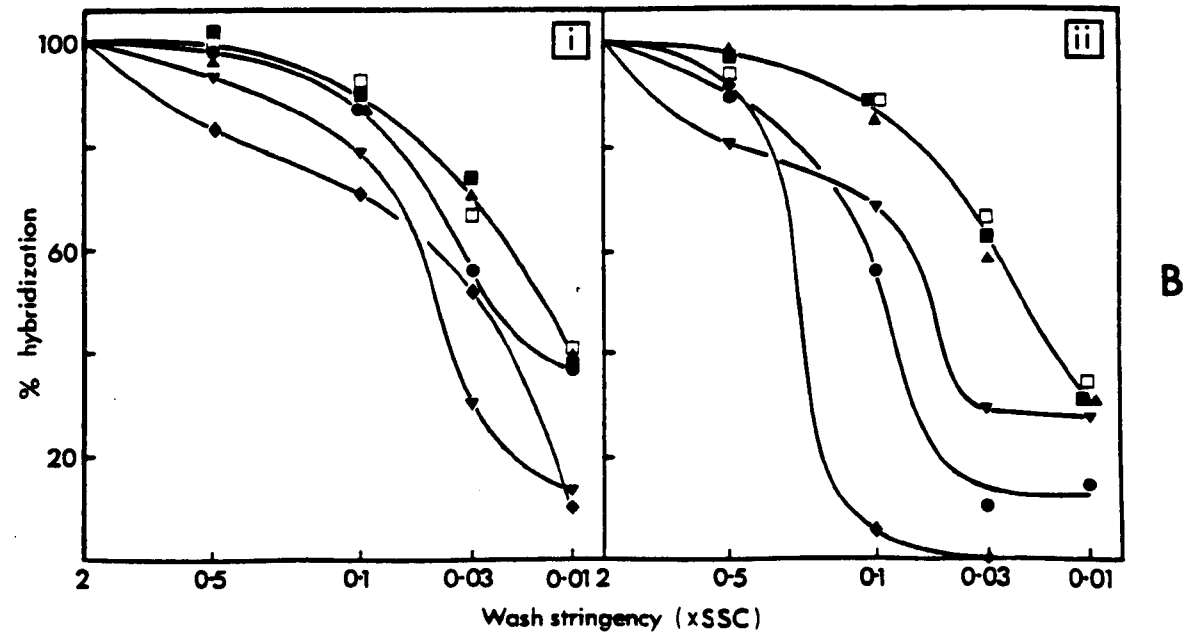

100
60
20
% hybridization
i
ii
2
0·5
0·1
0·03
0·01
2
0·5
0·1
0·03
0·01
Wash stringency (xSSC)
B

Figure R-24 Dot-blot hybridization analysis of rodent DNA samples. A, The indicated amounts of chromosomal DNAs were immobilized on nitrocellulose disc filters in duplicate and hybridized with a ^{32}P-labelled Ki-MuLV DNA probe (i) or NVL-3 DNA probe (ii) as detailed in Methods, section M-17. Following hybridization, filters were washed in 2 x SSC at 60°C before being placed in room temperature 2 x SSC for Cerenkov counting. Curves were plotted using linear regression calculations. B, The same filters were subjected to a series of 60°C washes at the indicated SSC concentrations. Filters were counted after each wash as before. Curves similar to those shown in A were plotted for each wash stringency. The melting curves shown were obtained by expressing the slope of a curve at each wash as a percentage of the slope of the curve at the 2 x SSC (least stringent) wash. DNA samples were as for fig. R-23 lanes a-f; NIH-3T3 (■); C3H10T½ (□), woodmouse (▲), rat (▼), bankvole (●) and grey squirrel (◆).

Results of the least stringent (2 x SSC) wash are presented in figure R-24A. The slope of each curve on the graphs is a function of two parameters; the relative abundance of MuLV- or VL30-related sequences, and the respective homology of these sequences to the cloned DNA probes. As shown in section A part (i), results with the MuLV probe indicated the existence of an inverse relationship between the amount of hybridization and increasing phylogenetic distance from NIH-3T3 DNA. In contrast, signal generated by hybridization with the VL30 probe was of a similar intensity with all three mouse DNA samples, but was dramatically reduced for the non-mouse DNAs [figure R-24A part (ii)].

Similar graphs were drawn from the results of each post-hybridization wash. The slope of each curve therein was calculated by linear regression and expressed as a percentage of the slope obtained with the appropriate DNA following the first (least stringent) wash (ie representing 100% hybridization). In this way melting curves for each rodent DNA:probe hybridization reaction were constructed [figure R-24B]. Since each point plotted on these curves represented a ratio of two hybridization values, parameters affecting the shape of the curves were independent of differences in copy number between the DNA samples.

Schildkraut & Lifson (1965) showed that the Tm of a given species of DNA decreased with decreasing salt (monovalent cation) concentration. They postulated that the negatively charged phosphate groups of DNA in solution tended to repel one another, and that monovalent cations were able to screen these mutual electrostatic repulsions. At lower salt concentrations where there was less cationic shielding, therefore, less thermal energy was needed to separate the DNA strands. They combined their equation relating Tm to log [salt] with that derived by Marmur & Doty (1962) which dealt with the effect on Tm of the G+C content of DNA, to give:

$$Tm = 16.6 \log M + 0.41 (\%G+C) + 81.5,$$

where M represents the concentration of monovalent cation (molar).

As demonstrated by the melting curves of figure R-24B, this equation predicted that Tm can be achieved by lowering the [salt] while maintaining a constant temperature. Here, the concentration of SSC in the post-hybridization wash (1 x SSC = 0.195M Na^+) was decreased in stages while maintaining a washing temperature of 60°C (± 3°C).

Tm values for each DNA:probe hybridization were calculated by inserting the value of M at which 50% hybridization occurred into the Schildkraut equation. A DNA sample which contained a population of sequences with 100% homology to the probe should achieve this equilibrium state at an SSC concentration which predicts a Tm of 60°C (ie the washing temperature). MuLV or VL30-related sequences with less than 100% homology to the probe would reach equilibrium under less stringent (higher [salt]) conditions. Thus a calculated Tm value greater than the washing temperature of 60°C indicated less than 100% homology.

Since the complete sequence of Ki-MuLV or NVL-3 is not yet known, G+C was estimated. NVL-3 was assigned a G+C content of 45% [oligonucleotide fingerprint analysis of Al virion VL30 RNAs showed them to possess a high G content (Clewley & Avery, 1982)]. Considerable cross-hybridization can be detected between exogenous and endogenous MuLVs, for instance between Mo-MuLV and AKR mouse endogenous proviruses (Chien et al., 1980), so the G+C content of these MuLVs is likely to be similar. The entire sequence of AKV-MuLV is known (Etzerodt et al., 1984) and its 52.6% G+C value has been used for the calculation described here.

At a temperature of 60°C, a DNA species with 100% homology to the Ki-MuLV probe should achieve equilibrium at 0.013 x SSC. This was calculated as follows:

	MuLV Probe				VL30 Probe			
DNA	ΔTm (°C)	% homology	% NIH signal	Copies per haploid genome	ΔTm (°C)	% homology	% mouse signal	Copies per haploid genome
NIH-3T3	0	100	100	50	0	100	100	150
C3H-10T½	0	100	82.3	41	0	100	100	150
Woodmouse	0	100	59.3	30	0	100	100	150
Rat	8.9	93.6	38.1	20	7.3	94.8	6.8	11
Bank vole	4.4	96.9	32.7	17	11	92.2	10.3	17
Grey squirrel	5.3	96.2	11.5	6	17.5	87.5	1.6	2-3

Table R-2 Relative homologies of retrovirus-like sequences in different rodent DNAs.

$$Tm = 16.6 \log M + 0.41 (\% G+C) + 81.5$$

$$\log M = Tm - [0.41 (\% G+C) + 81.5] / 16.6$$

when $Tm = 60^{o}C$, and $\% G+C = 52.6$,

$$\log M = 60 - 103.07 / 16.6$$

$$= -2.5943$$

$$M = 2.55 \times 10^{-3} M \quad (0.013 \times SSC)$$

Similarly DNA sharing 100% homology with the NVL-3 probe should achieve equilibrium at 0.021 x SSC and $60^{o}C$. As can be seen from the melting curves, the mouse DNA samples approached these two values very closely.

The ΔTm values from $60^{o}C$ are summarized in table R-2, together with the percentage homology to probe DNA which was commensurate with this value. These were calculated using the approximation that every 1% base mismatch between two DNA species is equivalent to a reduction in Tm of $1.4^{o}C$ (Hyman _et al._, 1973).

Having obtained an idea of the relative homologies to the probes of each rodent DNA provirus-like family, a superficial estimate of their relative copy number was attempted. Although presented as absolute values in table R-2, these copy number estimates are intended merely to display a trend, as their calculation was based on several assumptions. These included (1) that amount of probe annealed increased linearly with percentage homology, (2) that all rodent genomes tested here were of identical size (3×10^{9}bp) and (3) that NIH-3T3 mouse DNA contained 150 copies of VL30 elements and 50 copies of MuLV proviruses. As an example: From the graph shown in figure R-24A(ii), bankvole DNA can be seen to generate 10.3% of the maximum (mouse DNA) signal at 2 x SSC with the VL30 probe. As shown in table R-2, this signal was due to a 92.2% homology to the probe. Thus bankvole DNA contained:

$$150 \times \frac{10.3}{100} = 15.45, + (100 - 92.2)\% = 16.7 \text{ copies/haploid genome of VL30-related elements}$$

The data presented in table R-2 provided a more quantitative confirmation of the preliminary Southern blotting results described previously. MuLV-related sequences showed a greater degree of conservation across the various rodent species than did those related to the VL30 probe. In rodent DNA other than mouse, the latter have rapidly diverged and have also been maintained in dramatically fewer numbers. The most distantly related rodent to mouse, the grey squirrel, contained sequences which were only 88% homologous to mouse VL30 DNA. These were only just detectable by dot-blotting, as the moderately stringent conditions of hybridization (50% formamide, 3 x SSC, 42°C) only allowed the formation of hybrids sharing 85% homology or greater (Howley et al., 1979). Interestingly rat, which is generally accepted to be more closely related to mouse than the vole (Chaline, 1977), had slightly fewer copies of VL30-related sequence than the latter. However the fact that the rat DNA data produced a biphasic melting curve with VL30 probe [figure R-24B(ii)] indicated the existence of more than one class of these sequences in the rat genome.

R-5.4 Organization of VL30-related sequences in woodmouse, rat and vole

The dot-blot analysis described above indicated that VL30 sequences were present in non-murine rodent DNAs, but that compared to MuLV-related sequences, they were less conserved. The mechanism by which these VL30-like sequences had been retained was examined in a Southern blot analysis of Apodemus, Rattus and Clethrionomys DNA.

As shown in figure R-25A, a sub-genomic probe analysis of Apodemus DNA produced a similar result to those seen earlier with NIH-3T3 DNA. All probes revealed approximately the same number and intensity of bands, many of which undoubtedly represented complete VL30 provirus-like elements. In contrast, rat and vole DNA demonstrated markedly different

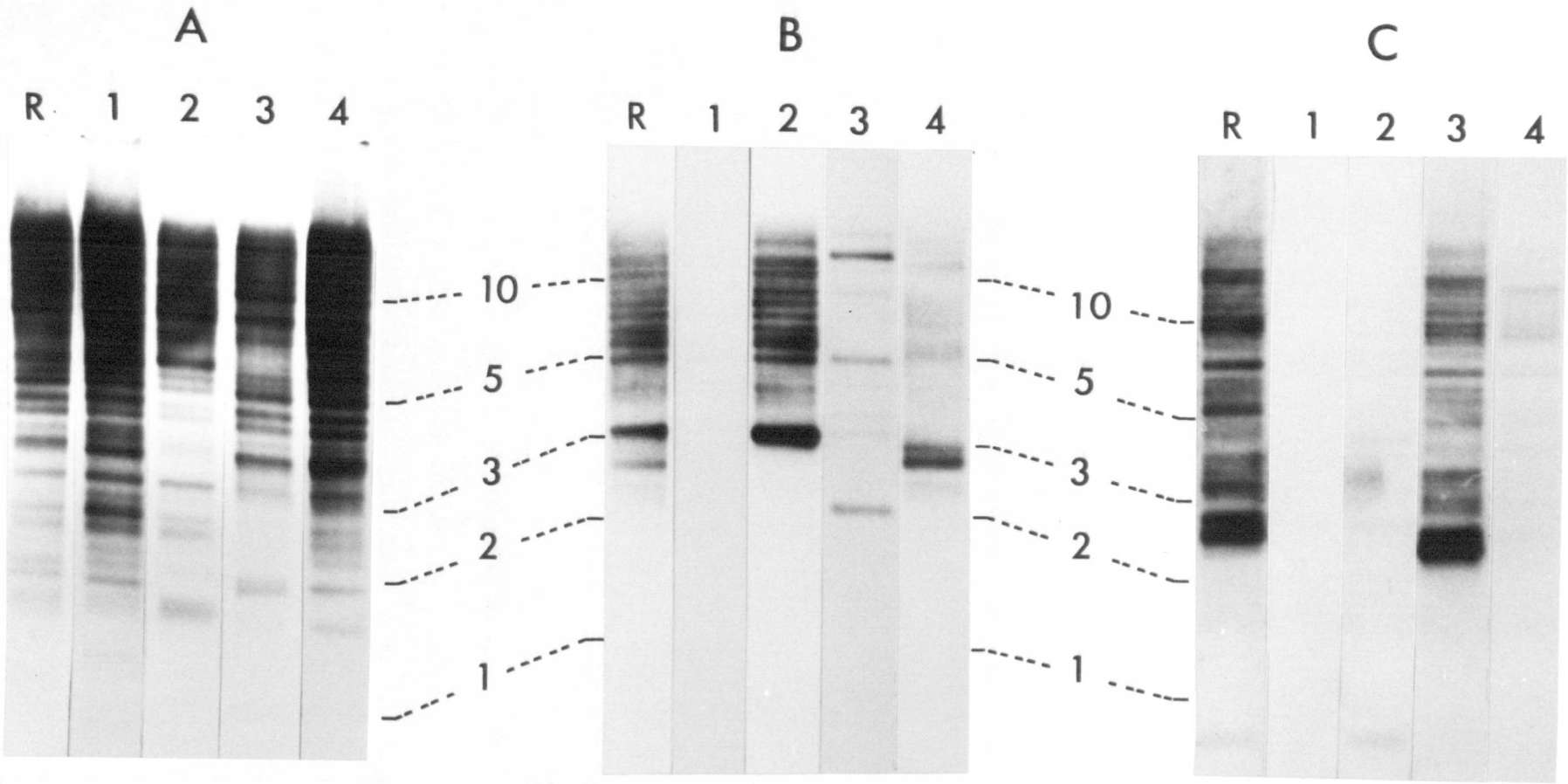

A
R
1
2
3
4
B
R
1
2
3
4
C
R
1
2
3
4
10
5
3
2
1
10
5
3
2
1

Figure R-25 Organization of NVL-3-related sequences in three rodent DNAs. 10μg samples of chromosomal DNA were digested to completion with Bam HI (A), Eco RI (B) or Sac I (C) and fractionated in five lanes of a 1.0% (B) or 0.8% (A,C) agarose gel. Following transfer to nitrocellulose each lane was separately hybridized with a ^{32}P-labelled NVL-3 DNA probe (R) or with one of the subgenomic NVL-3 DNA probes 1-4 (see fig. R-19). Filters were washed in 0.1 x SSC at 60°C, dried and reassembled for autoradiography. A, woodmouse DNA (2 day exposure); B, rat DNA (14 day exposure); C, bank vole DNA, (14 day exposure). Sizes of DNA species are given in kilobase pairs. During printing, the bottom half of A was exposed a few seconds longer than the top half to enable visualization of low molecular weight bands.

patterns of VL30 sequence conservation (figure R-25B, C). Although both genomes lacked VL30 LTR information, rat DNA revealed multiple copies of a discrete set of sequences mapping predominantly to the NVL 5' unique sequence, while vole DNA possessed a slightly more abundant set of fragments which mapped to the central part of the NVL unique region. Rat DNA also possessed restriction fragments which were revealed by probe 4, which recognizes sequences at the 3' end of the NVL unique region. These bands, though, were obviously a minor population compared to the 5' end-related set. The two different sub-sets of mouse VL30 information in rat DNA revealed by the sub-genomic probes may account for the biphasic melting curve described in the previous experiment. The significance of the fact that different blocks of VL30-related sequence are conserved in different rodent species will be discussed later.

R-6 Expression and transmissibility of VL30 RNA

R-6.1 Introduction

Detailed restriction enzyme analyses of VL30 elements present in NIH-3T3 genomic DNA (section R-4) provided evidence which, together with the cDNA cloning results (section R-3), strongly suggested that only a few members of this gene family were transcriptionally active. In the first part of this section results of an investigation into the methylation status of integrated NIH-3T3 VL30 units adds further support to this hypothesis.

As discussed in section D-5, many present day mammalian and avian species contain evidence of ancient, inter-specific retrovirus infection events in their germ-line DNA. Knowledge that VL30 RNA can be packaged and transmitted by retroviruses (section R-1) introduces the interesting possibility that VL30 elements may also be passed from species to species.

As a first step towards demonstrating this possibility, it must be established that the *in vivo*-synthesized mouse VL30 DNA detected in newly infected cells can integrate into the genomic DNA of a heterologous cell. The final experiments described in this section show that the NVL-like VL30 species, packaged into A1-produced Ki-MuLV particles, can be transmitted to normal rat cells. Evidence is presented which suggests that these VL30 species have been used as templates for the synthesis of full-length VL30 DNA. This has become stably integrated into the rat genomic DNA and is apparently transcriptionally active.

R-6.2 Methylation status of NIH-3T3 mouse VL30 elements

Evidence from oligonucleotide fingerprinting of virion-encapsidated VL30 RNA (Besmer *et al.*, 1979; Clewley & Avery, 1982) and the restriction maps of the NVL clones has suggested that only a limited

number of the approximately 150 mouse VL30 genes are transcriptionally active. As discussed in the General Introduction, it has been shown that the level of expression of some endogenous proviruses correlates with their degree of DNA methylation (Cohen, 1980; Groudine et al., 1981). It was decided, therefore, to examine the methylation status of NIH-3T3/A1 mouse VL30 elements in order to obtain evidence which might support this notion.

5-Methylcytosine (m^5C) comprises 2 to 7 percent of the cytosine present in the genome of higher eukaryotes (Vanyushin et al., 1970) and almost all of these m^5C bases are found in the dinucleotide sequence CpG (Razin & Riggs, 1980). One of the accepted methods of assaying methylation at CpG residues is by the use of the enzyme MspI which recognizes the sequence CCGG and which will cut here irrespective of a m^5C at the internal C residue (Waalwijk & Flavell, 1978). This enzyme is used in comparison with HpaII which also recognizes CCGG but which will not cut if the internal C residue is methylated (Bird & Southern, 1978). Thus a restriction fragment which migrates at the same rate following either MspI or HpaII digestion may be regarded as evidence for undermethylation at CC*GG sites.

MspI/HpaII analyses of NIH-3T3 DNA, or A1 cell DNA produced identical results, so only those of A1 DNA are shown in figure R-26. The largest NVL-3 DNA MspI marker fragment (lane c of part B) co-migrated with a major 2.3kbp band in the MspI-digested DNA track (lane a), also with a fainter band in the HpaII-digested DNA track (lane b). This fragment maps to the 3' half of clone NVL-3 (see restriction map, part A). Bands corresponding to the three smaller NVL-3/MspI fragments were just detectable in the MspI track, but could not be seen in the HpaII track.

Relative intensity of the 2.3kbp MspI genomic DNA band (lane a) to that of the corresponding marker fragment (lane c) suggested an

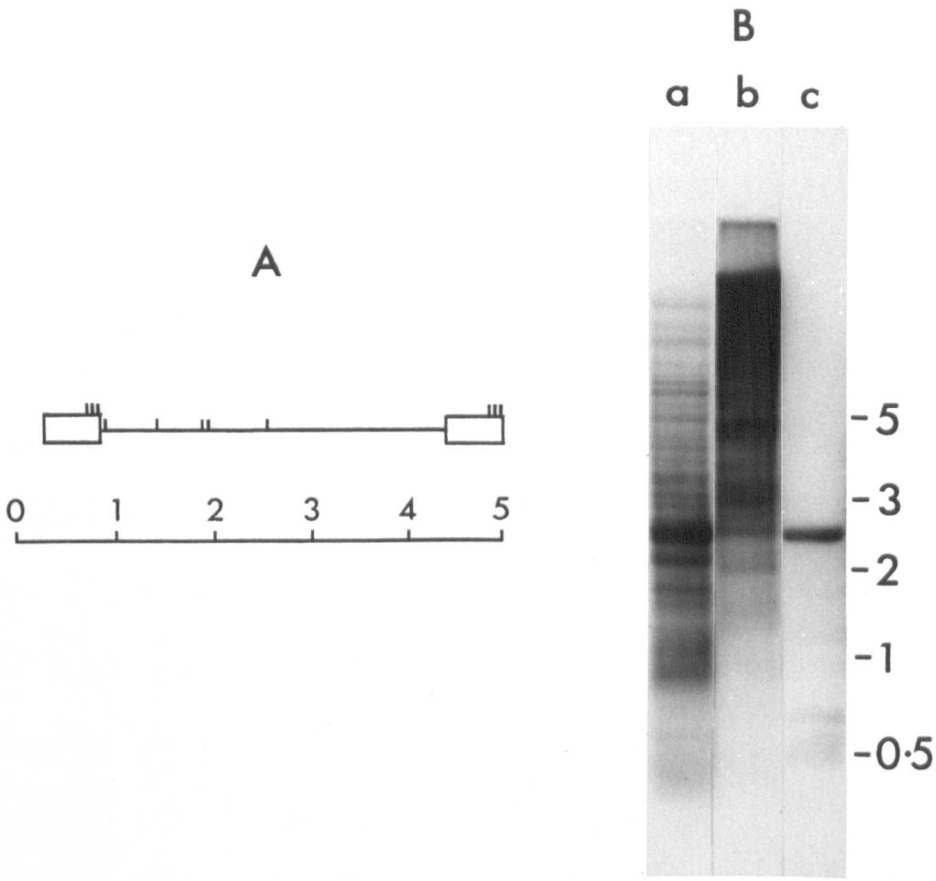


Figure R-26 Methylation status of VL30 sequences in NIH-3T3 DNA. A, physical map of clone NVL-3 indicating Msp I/Hpa II cleavage sites. B, DNA samples were digested to completion with Msp I (a) or Hpa II (b,c), electrophoresed through a 1.0% (w/v) agarose gel, immobilized on nitrocellulose and hybridized with ^{32}P-labelled VL30 clone NVL-3 DNA. Lanes a, b, 10μg NIH-3T3 DNA; c, approx. 275pg NVL-3 DNA. Sizes of DNA species are given in kilobase pairs.

approximate copy number of 30-50 fragments per haploid genome. Its counterpart in the HpaII track (lane b) was about 10 fold less intense, ie 3-5 copies per haploid genome.

Other bands also appeared to co-migrate (compare lanes a and b), but from this preliminary experiment it was concluded that only 3-5 copies of NVL-like VL30 elements were undermethylated, and could therefore be transcriptionally active. This figure represents about 2-3% of the total mouse VL30 complement, and was consistent with the restriction mapping studies described in section R-4.

If methylation at HpaII sites does correlate with reduced or non-expression of VL30 units, then this finding could explain the similarity between the NVL cDNA clones discussed in section R-3. However these data do not preclude a sub-set of VL30 units which (a) express RNA, (b) are packaged (as RNA) by retroviruses, but (c) are less efficient at, or are incapable of being reverse-transcribed by the virion-packaged enzyme. It might therefore be interesting to examine the total VL30 RNA population in these cells, either by cDNA cloning, or by attempting to determine the percentage of these molecules which were closely related to NVL clones (ie by liquid hybridization studies). The latter experiment might be difficult to perform, however, as restriction mapping studies (section R-4) suggest a high degree of sequence homogeneity between members of the mouse VL30 gene family.

The results discussed here were consistent with those of Keshet & Itin (1982) who performed a similar analysis of BALB/c mouse DNA to conclude that only a small fraction of mouse VL30 genes were transcriptionally active.

R-6.3 Newly integrated proviruses in the rat cell line KLNRK-1

This cell line was derived by acute infection of NRK cells (Duc-Nguyen *et al.*, 1966) by Ki-MuLV (VL30) particles from A1 tissue culture

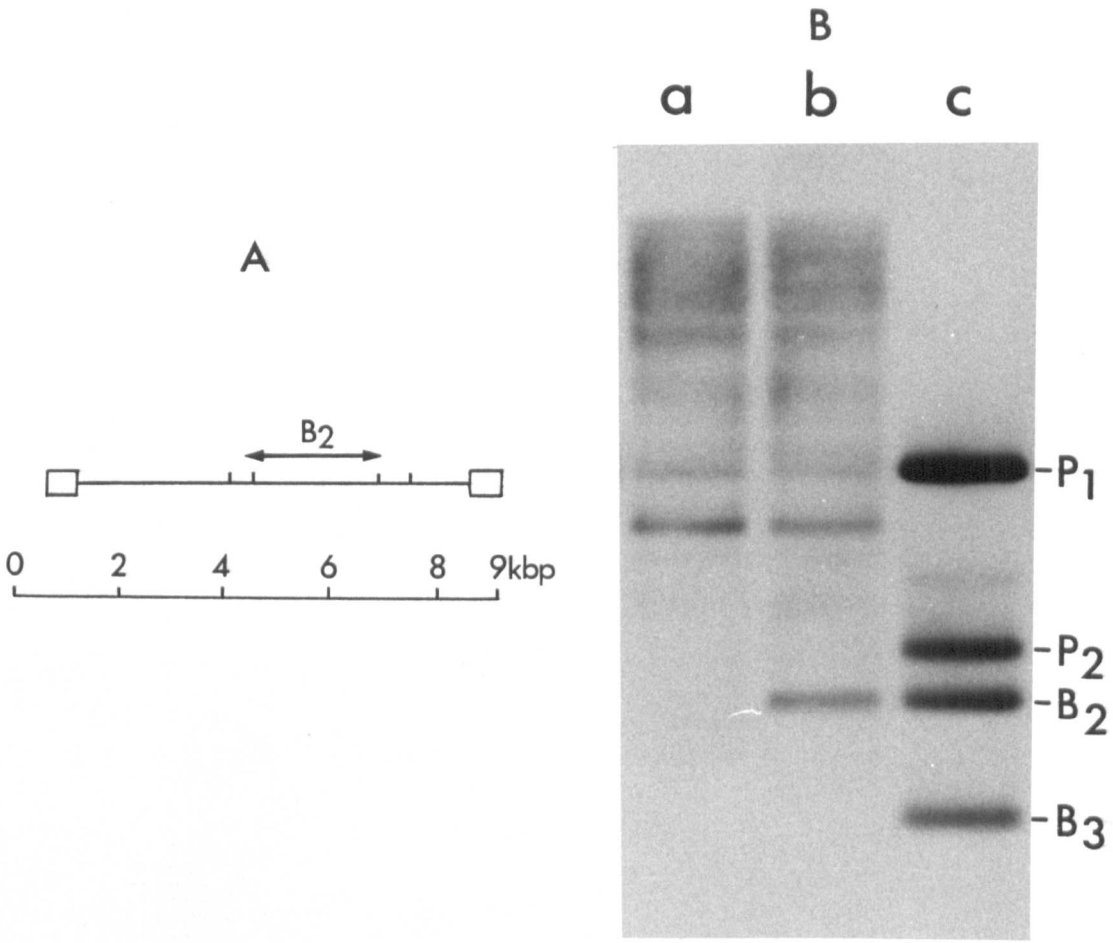


Figure R-27 Detection of Ki-MuLV proviruses in KLNRK-1 cells. A, physical map of Ki-MuLV DNA indicating Bam HI cleavage sites. B, 10μg samples of uninfected NRK DNA (a) and KLNRK-1 DNA (b) were digested to completion with Bam HI, electrophoresed through a 1.0% (w/v) agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled Ki-MuLV DNA. Lane c was loaded with approx. 100pg each of the following cloned Ki-MuLV DNA Pst I and Bam HI restriction fragments; P1 (5.12kbp), P2 (2.83kbp), B2 (2.42kbp) and B3 (1.6kbp).

fluid (section M-1.2). High molecular weight DNA was prepared from these cells at the fifteenth passage post infection, and was examined by the Southern transfer technique for evidence of stably integrated proviruses that had been introduced since the infection event. Ki-MuLV particles are released by these cells (Clewley & Avery, 1982) so newly integrated Ki-MuLV proviruses were expected and probed for as a positive control. The question of interest concerned the ability of mouse VL30 elements to successfully enter the genome of a heterologous cell.

Figure R-27B shows the results of a Southern analysis of BamHI-digested NRK DNA (lane a) and KLNRK-1 DNA (lane b) electrophoresed in parallel with Ki-MuLV DNA. Restriction sites for BamHI are situated in the Ki-MuLV genome as depicted in part A of this figure. The smaller two marker fragments of lane c correspond to a central 2.4kbp fragment (arrowed in the map) and a smaller, 3' terminal BamHI fragment. The internal 2.4kbp fragment co-migrated with a major band in lane b but not in lane a. This was taken as evidence for newly integrated Ki-MuLV proviruses. Two smaller BamHI fragments which flank this central one in Ki-MuLV DNA are not seen as they were not retained by this gel. Also seen were bands which co-migrated in both lanes a and b. These indicated the presence of several endogenous MuLV-related species that were established before infection with Ki-MuLV.

The absence of 5' or 3' terminal Ki-MuLV BamHI fragments in KLNRK-1 DNA, or their covalently joined product demonstrated that the Ki-MuLV proviruses were integrated, and were not merely present in the DNA preparation as unintegrated linear and circular forms, respectively. Furthermore the absence of any other Ki-MuLV-related bands in KLNRK-1 DNA, but not in NRK DNA, suggested that these proviruses were inserted at random into the rat cell DNA.

Comparison of autoradiographic signal generated by the integrated 2.4kbp Ki-MuLV BamHI fragment, and by internal PstI fragments from a

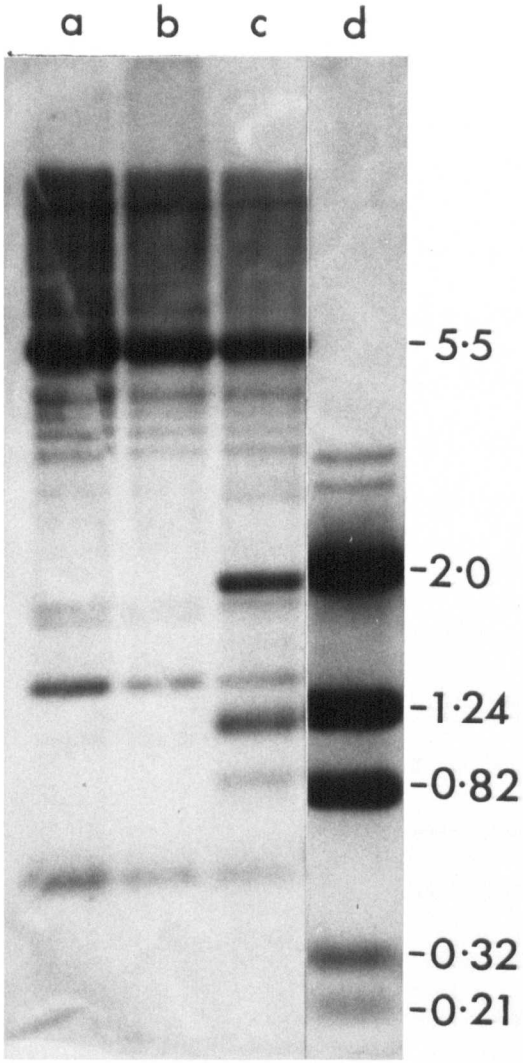

a
b
c
d
-5·5
-2·0
-1·24
-0·82
-0·32
-0·21

Figure R-28 Detection of mouse VL30 sequences in KLNRK-1 cells. Samples were digested to completion with Sac I, electrophoresed through a 1.0% (w/v) agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled NVL-3 DNA. a, 10μg NRK DNA; b, 10μg KNRK DNA; c, 10μg KLNRK-1 DNA; d, approx. 275pg NVL-3 DNA. Sizes of DNA species are given in kilobase pairs.

similar analysis (data not shown) with that of their corresponding marker fragments suggested that 40-50pg of each Ki-MuLV-specific fragment was present per 10μg of KLNRK-1 DNA. Assuming the rat genome to comprise 3 x 10^9bp, this indicated the presence of about 10 copies of the provirus per cell. Such a figure was consistent with the findings of Steffen & Weinberg (1978) who infected NRK cells with Mo-MuLV, then sub-cloned a number of cells to demonstrate the presence of several randomly integrated proviruses in the DNA of each cell line.

Figure R-28 shows a similar analysis of rat DNA, this time using the enzyme SacI and an NVL-3 DNA probe. Newly integrated VL30 proviruses in KLNRK-1 DNA were identified by (a) their specificity for this DNA, and (b) by the fact that their SacI restriction fragments co-migrated with marker fragments electrophoresed in parallel with the genomic DNA digests. Digestion of NVL DNA with SacI produces 5 restriction fragments, the two smallest ones generated by sites in the two LTRs (see map in figure R-19). As with the previous analysis, these terminal fragments were not seen in SacI-digested KLNRK-1 DNA (lane c), indicating that VL30 integration had taken place. As might have been predicted from analysis of the NVL (virion-derived) cDNA clones, these newly integrated VL30 species were NVL types. Thus three internally derived NVL SacI fragments (lane d) co-migrated with their counterparts in lane c. The smallest of the latter was fainter than expected. On closer inspection this was a result of its slight size heterogeneity. In fact two bands were present, one (the fainter) corresponding in size to an NVL-3 type of provirus, the other (about 2-3 times more intense) indicating an NVL-1 or NVL-2 type.

Endogenous rat VL30 sequences were readily identified by their presence in all three rat DNA digestion profiles. Neither NRK DNA (lane a) nor KNRK DNA (lane b) contained the NVL-specific bands. This finding was not unexpected for the latter, as the KNRK cell line was

established by infecting NRK cells with a Ki-MuLV/KiMuSV stock that had been propagated on rat cells, thus avoiding the possibility of transmitting mouse VL30 elements.

As with the newly introduced Ki-MuLV proviruses, the absence of detectable cell DNA/VL30 termini fragments demonstrated that the mouse VL30 proviruses were integrated at random in the rat genome. The copy number of these NVL-like proviruses was estimated at 1-2 per cell, which suggests that they are less infectious than the MuLV sequences.

R-6.4 Mouse VL30 RNA in mouse and rat cells

The data discussed in this section are the unpublished findings of Yvonne Gibson (Department of Microbiology, University of Leicester) and Dr. J.D. Norton (Department of Haematology, Royal Free Hospital School of Medicine, London), but since they are directly relevant to the work described in this results section, they are included at this point rather than in a discussion section.

Total RNA from NIH-3T3 cells was fractionated by electrophoresis in an agarose:formamide gel and transferred to nitrocellulose. Following hybridization with the NVL-3 DNA probe, at least 3 bands were detected (figure R-29A). In the absence of a definitive RNA marker system for these gels, it was difficult to determine which species represented 30S RNA. The most satisfactory interpretation identified the middle species as 30S, the upper species as an aggregation artefact previously described by Fan & Mueller-Lantzsch (1976), and the lower species as a putative sub-genomic VL30 RNA species. It was not clear whether the latter represented a processed form of 30S RNA, a discrete degradation product, or a hybrid RNA composed partly of VL30 sequence and partly of flanking cellular DNA sequence.

This situation was clarified by an analysis of RNA extracted from KLNRK-1 cells. As described above, these cells each contained

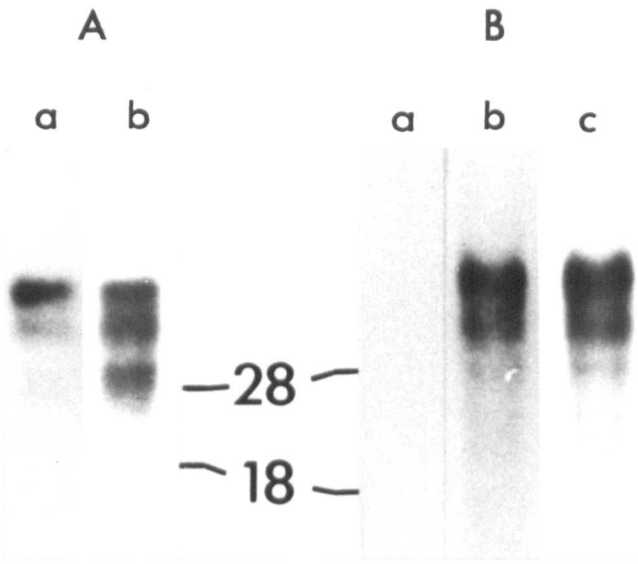


Figure R-29 Mouse VL30 RNA detected in NIH-3T3 and KLNRK-1 cells. Total RNA was fractionated by electrophoresis through a 1.0% (w/v) agarose/formamide gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled NVL-3 DNA. Gel A, approx. 5µg (a) and 10µg (b) NIH-3T3 RNA. Gel B, equal loadings (approx. 10µg) of NRK RNA (a) and KLNRK-1 RNA (b, 24 hour autoradiographic exposure, c, 10 day exposure). Size estimation (S values) was by comparison with mouse ribosomal RNA markers (not shown). Gel A is reproduced from unpublished work of Yvonne Gibson and John Norton; gel B was repeated for clarity.

1-2 NVL-like proviruses as a result of infection by Ki-MuLV (VL30) particles. It was of interest, therefore, to determine whether these VL30 elements were capable of RNA expression in a heterologous cell.

Figure R-29B shows a northern analysis of total RNA from NRK cells (lane a) and KLNRK-1 cells (lanes b and c). The rat genome contains a family of rat VL30 elements which are structurally and genetically related to the mouse VL30 family (Scolnick et al., 1976a, 1979; Itin et al., 1983). These are known to express 30S RNA in normal rat cells (Tsuchida et al., 1974). However these species were not detectable in the NRK RNA (lane a) by this method, even after prolonged autoradiographic exposure.

This result implied that the prominent signal obtained with KLNRK-1 RNA (lanes b and c) was directly attributable to the newly integrated mouse VL30 proviruses. Although the three major RNA species detected by the NVL-3 probe were present in slightly different proportions to those of the NIH-3T3 preparation, their sizes were similar if not identical to the latter species.

R-6.5 Methylation status of VL30 proviruses in KLNRK-1 cells

High molecular weight DNA from these cells was examined using the MspI/HpaII system, as described above. Uninfected NRK cell DNA contained a major mouse VL30-related fragment of about 400bp (figure R-30, lane a). Smaller fragments related to the NVL-3 DNA probe were either not retained by this gel, or did not blot efficiently. Upon HpaII digestion of the same DNA, all of the mouse VL30-related sequences were contained in larger restriction fragments, indicating the presence of methylated CpG positions.

MspI or HpaII digestion of cloned NVL-3 DNA produces a distinct restriction profile composed of 4 major internally-derived fragments (lane e). Their proviral DNA counterparts were easily recognizable

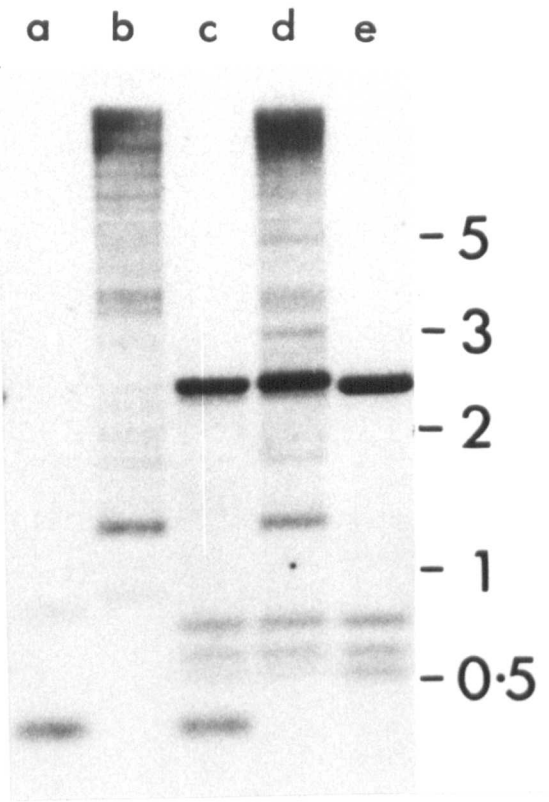


Figure R-30 Methylation status of newly integrated NVL-like elements in KLNRK-1 cells. DNA samples were digested to completion with the appropriate enzyme, fractionated by electrophoresis in a 1.2% (w/v) agarose gel, transferred to nitrocellulose and hybridized with ^{32}P-labelled NVL-3 DNA. Lanes a, b, 10μg NRK DNA; c, d, 10μg KLNRK-1 DNA, digested with Msp I (a,c) or Hpa II (b,d). Lane e, approx. 35pg Hpa II-digested NVL-3 DNA added to 10μg Eco RI-digested calf thymus DNA. Sizes of DNA species given in kilobase pairs.

against the background pattern of endogenous rat VL30 sequences, following digestion of KLNRK-1 DNA with either MspI (lane c) or HpaII (lane d). This result provided further evidence that the newly integrated VL30 proviruses were very closely related to NVL-type species. A slight doublet effect of the largest NVL-related band in lanes c and d once more suggested that NVL-1 or NVL-2 elements were more strongly represented in this DNA than were the NVL-3 (smaller) types. Most importantly, these data also showed that all of the newly introduced mouse VL30 species were undermethylated.

R-6.6 Summary

The results of this section included (1) an MspI/HpaII analysis of NIH-3T3 mouse DNA which showed that the majority of VL30 elements were hypermethylated, a physical state which, for endogenous proviruses, often correlates with transcriptional silence.

(2) This finding was consistent with the discovery that the NVL cDNA clones were closely related, which suggested that they were encoded by a relatively small sub-set of NIH-3T3 mouse VL30 elements.

(3) DNA from KLNRK-1 cells was found to contain examples of Ki-MuLV and mouse VL30 proviruses which were not present in uninfected NRK DNA. Interestingly, the MuLV proviruses were present in greater numbers than were the VL30 proviruses. On the basis of restriction sites, the latter appeared to be closely related to NVL-type units.

(4) The work of Yvonne Gibson and John Norton showed that NIH-3T3 RNA contained 3 major VL30-related species, at least one of which may have been the result of RNA processing. Analysis of RNA from KLNRK-1 cells revealed a similar set of mouse VL30-related species. Since endogenous rat VL30 RNA was not detectable in these cells, these data strongly suggested that the novel RNAs were encoded by the newly introduced NVL-like proviruses. Furthermore they implied that the

putative sub-genomic RNA species were encoded entirely from within these VL30 units. By inference, these findings were further evidence that the majority of NIH-3T3 mouse VL30 RNA species were also encoded by NVL-type units.

Finally (5) the methylation status of NVL-like proviruses in KLNRK-1 DNA showed that all of these were undermethylated and were therefore capable of expressing the RNA species detected by northern analysis.

DISCUSSION

DISCUSSION

D-1 In vivo-synthesized VL30 DNA

Three major forms of retrovirus DNA have been defined in infected cells: linear duplexes, closed circular duplexes, and DNA covalently integrated into the host genome (Varmus et al., 1975; Weinberg, 1977). Studies on the kinetics of appearance of these viral DNA species have shown that linear duplexes are synthesized in the cytoplasm during the first 3-4 hours post infection. This is followed by the appearance of circular forms in the nucleus as early as 5 hours (ASV in cultured quail cells) or 7 hours (MuLV in cultured mouse cells) post infection (Shank & Varmus, 1978; Yang et al., 1980). Evidence that the circular DNA with two tandem LTRs is a precursor to the provirus was recently provided by Panganiban & Temin (1984) who showed that a retrovirus vector could efficiently integrate at the site of a 49bp DNA fragment insertion which contained the junction formed by blunt-end ligation of two LTRs. However these experiments did not formally exclude the possibility that the other two forms of unintegrated DNA could integrate.

The results described in section R-1 demonstrated that Ki-MuLV behaved as a typical retrovirus when inoculated onto a culture of rapidly dividing mouse cells. The cDNA probe detected an 8.5kbp linear DNA species, which in the most successfully fractionated cells (figure R-10) was the predominant Ki-MuLV DNA form in the cytoplasm. Two DNA species isolated from cell nuclei were judged by their electrophoretic mobility to possess a closed circular conformation, and were identified as such when they co-banded with supercoiled plasmid DNA in a caesium chloride:ethidium bromide density gradient. A restriction enzyme analysis of circular Ki-MuLV DNA provided data which suggested that the larger of the two species contained two copies of the LTR, whilst the smaller form possessed only one copy of this region (Norton

et al., 1984a). Thus in every respect, the retrovirus used in these studies had demonstrated its ability to synthesize replicative intermediates which were exactly analogous to those described for other MuLVs (Yoshimura & Weinberg, 1978) in addition to the those of other retroviruses as distantly related as ASV (Shank et al., 1978b) and MMTV (Shank et al., 1978a).

The cDNA probe synthesized using virion RNA as template also detected a major 4.6-4.9kbp DNA species in recently infected cells. This was of a size commensurate with that of the virion 30S RNA (figure R-14) and its appearance in agarose gels as a doublet or triplet reflected the size heterogeneity of the VL30 RNA. Fractionation on an isokinetic sucrose gradient provided evidence which supported the theory that these species were linear DNA molecules; as with Ki-MuLV they were more abundant in the cytoplasm than in the nucleus. Low molecular weight DNA extracted from the nuclei was found to contain a DNA species which, when electrophoresed through agarose gels, migrated as a doublet in a position expected for closed circular VL30 DNA. Compared to other VL30 DNA, and the Ki-MuLV DNA species, this was extremely difficult to detect in the bulk DNA preparations. However following caesium chloride:ethidium bromide density gradient centrifugation, which also provided proof of its circular structure, this species was enriched to a degree whereby its existence as a discrete entity was no longer in doubt. The identity of these circular DNAs was confirmed when a cloned VL30 probe became available for hybridization to the original nitrocellulose filters on which these species had first been immobilized.

Although liquid hybridization studies have detected VL30-specific sequences in the cytoplasm of infected cells or in virions released from them (Scolnick et al., 1979), there have been no reports of a Southern blot analysis of unintegrated VL30 DNA. The data described in this

thesis, therefore (and published in Carter et al., 1983) remain the only available information on this in vivo-synthesized VL30 DNA. That linear VL30 DNA is found in the cytoplasm, and circular VL30 DNA is found in the nucleus of Ki-MuLV (VL30)-infected cells indicates that the unintegrated DNA intermediates of VL30 appear to be exactly analogous to those found with retroviruses. Since the amounts of VL30 DNA circles recovered from the nucleus were very small, no further characterization of these species was undertaken. It would be of interest, nevertheless, to determine whether these molecules completed the analogy with retrovirus DNA by their adoption of two size classes, corresponding to the one- or two-LTR version.

Restriction enzyme digestion of sucrose gradient-fractionated preparations of linear Ki-MuLV and VL30 DNA species produced band patterns in agarose gels which tended to confirm the initial observation that Ki-MuLV was a discrete genetic element, in contrast to VL30 species which displayed a limited heterogeneity. An annoying background hybridization signal in Southern blot experiments, caused by contaminating mouse chromosomal DNA, hindered any detailed observations. However, judging by the complexity and relative band intensity of the VL30 DNA digestion profiles it appeared that different species were present in different relative proportions. Further conclusions were prevented when it was shown that some of the "VL30" restriction fragments were derived from Ki-MuLV information. It could not be determined whether these contaminants represented randomly degraded Ki-MuLV DNA, or a more discrete population of sub-genomic species, either of which by chance had co-purified with the VL30 DNA size fraction.

In conclusion, it was shown that apparently full-length VL30 DNA was recoverable from recently infected cells, but that this was present in minute amounts compared to the bulk of contaminating mouse

chromosomal DNA. Since the cDNA probe shared sequence homology with two different families of endogenous proviruses in the mouse genome, any molecular cloning attempt using the *in vivo* source of VL30 DNA would be technically very difficult. Instead, as discussed in the following section, the possibility was explored of using a retrovirus endogenous reaction as a means of obtaining relatively pure VL30 DNA.

D-2 In vitro-synthesized VL30 DNA

Each retrovirus core contains 40-70 copies of reverse transcriptase, at least one molecule of which is closely associated with each subunit of the virion RNA complex (Panet *et al.*, 1975; Krakower *et al.*, 1977; Stromberg *et al.*, 1974). Faithful double-stranded DNA copies of the virus genome may be obtained using this endogenous reverse transcriptase activity if conditions are arranged so that they mimic the environment of partially uncoated virus particles within the newly infected cell. In an attempt to obtain VL30 DNA suitable for molecular cloning, conditions for the endogenous reaction of Ki-MuLV (VL30) particles were optimized (section R-2).

Initial experiments used the detergent Triton X-100 to partially lyse the virions and allow Mg^{2+} ions and dNTPs access to the transcription complex. Having identified the optimum concentrations of virus and detergent, a time course experiment was conducted using ^{32}P-labelled DNA reaction products (figure R-7). Newly synthesized linear DNA molecules of Ki-MuLV and VL30 began to be detected between 3-6 hours, indicating an approximate rate of synthesis of 24bp/minute. This compared well with that observed for other retrovirus systems, both *in vivo* and in the endogenous reaction (Rothenberg & Baltimore, 1977; Varmus *et al.*, 1978). The significance of this rate of polymerization, which is much slower than that observed for other cellular and viral reactions, is not known (Varmus & Swanstrom, 1982). Full-length DNA products (their size estimated from electrophoretic mobility in agarose gels) continued to accumulate up to 12 hours, after which time they started to degrade. The precise cause of this degradation was unknown, but may have been a result of cellular nuclease activity packaged during virion assembly. Approximately 100-200ng of viral DNA could be recovered from an endogenous reaction with an initial input of 4mg virus (data not shown). This was comparable to yields obtained using other

retroviruses (Benz & Dina, 1979; Gilboa et al., 1979b; Norton et al., 1982) and represented a one thousand fold greater yield than that achieved in vivo.

The identical sizes of the in vitro DNA products and their in vivo linear DNA counterparts was confirmed by their co-migration in an agarose gel (figure R-10). However it was considered necessary to examine the fidelity of the in vitro DNA in more detail, since it was possible that the intracellular process had not been faithfully recapitulated in some way. This degree of caution was fuelled by reports that AMV polymerase can exhibit a high error rate under certain conditions, especially when transcribing homopolymeric RNA and DNA templates or a single-stranded phage DNA template (Battula & Loeb, 1974; Mizutani & Temin, 1976; Gopinathan et al., 1979). A high virus mutation rate observed during passage of a clonal stock of MuLV may also have been due to misincorporation by the reverse transcriptase (Shields et al., 1978) but this phenomenon is a recognized feature of RNA viruses in general (Weissman et al., 1973; Tsipis & Bratt, 1976). A nucleotide sequence comparison between two cloned examples of Mo-MuLV, one representing an integrated provirus (Van Beveran et al., 1980) the other from DNA synthesized in vitro (Sutcliffe et al., 1980) revealed a difference of three nucleotide changes in the U_3 region (3 changes in 448 nucleotides = 0.67%). This may be an overestimate of error incurred by in vitro synthesis, since the U_3 region of retroviral LTRs is known to be less conserved than, for instance, the U_5 domain (Van Beveran et al., 1982; Temin, 1982). Indeed, several examples of retrovirus DNAs synthesized in vitro have possessed the appropriate biological activity in the transfection assay (Rothenberg et al., 1977; Clayman et al., 1979; Gilboa et al., 1979b; Lai & Verma, 1980). This indicates that reverse transcriptase mistakes in optimized endogenous reactions are relatively rare and do not involve large regions of the genome. The

results of a comparative restriction enzyme analysis of *in vivo* and *in vitro* VL30 DNA provided data which tended to support this hypothesis (figure R-12). As with DNA recovered from infected mouse cells, VL30 DNA was not cleaved by *Bam*HI, and Ki-MuLV DNA bore no recognition sites for *Eco*RI. Enzymes which cleaved the *in vivo* DNA generated identical restriction fragment profiles with the *in vitro*-synthesized DNA. Interestingly, the 5kbp VL30 size fraction was again shown to be contaminated by Ki-MuLV DNA species. These truncated molecules appeared to represent the entire Ki-MuLV genome, since *Bam*HI digestion of VL30 DNA generated a digestion profile which was indistinguishable from that of full-length Ki-MuLV DNA. Some preparations of VL30 DNA possessed substantial quantities of the sub-genomic Ki-MuLV DNA (eg see figure R-12). Indeed, over 60% of the molecular clones obtained after using this source of VL30 DNA were found to be of Ki-MuLV origin (section R-3). One might expect to obtain more *Bam*HI fragments of Ki-MuLV DNA than full-length VL30 clones, however, simply because each Ki-MuLV fragment would possess at least one natural *Bam*HI terminus. Therefore these fragments would be cloned at a higher efficiency than VL30 elements which bore synthetic linkers at each end.

This Ki-MuLV DNA may have arisen (a) by random degradation of the full-length (8.5kbp) Ki-MuLV transcript or (b) as the result of a more ordered process. It is known that retroviruses often package RNA molecules which do not necessarily play a role in virus replication (for review see Bishop, 1978). Several species of small RNA derived from the cytoplasm of the host cell (including 4S, 5S and 7S species) have been identified in mammalian retrovirus particles, also small amounts of cellular mRNA [Friend leukaemia virus, propagated in erythroid cells, packages about one copy of globin mRNA per 1000 copies of viral RNA (Ikawa *et al*., 1974)]. Of most significance in this context is the observation that sub-genomic viral RNA species are also encapsidated

into retroviruses. Thus, the 21S env mRNA of ALV has been found to be associated with high molecular weight virion RNA (Stacey, 1979). RAV-2 particles may package RNA of which 5-10% is 21S mRNA and 90-95% is the 35S genomic RNA (Linial & Blair, 1982). Since this env mRNA has information for both 5' and 3' ends of the genome, it was proposed that if successfully used as template by the virion reverse transcriptase, the sub-genomic DNA thus formed could integrate into chromosomal DNA and even be capable of expression (Stacey, 1979). This was put forward as a possible mechanism to explain the long-term expression of 21S mRNA in RSV(-) cells following its micro-injection into the cytoplasm. Stacey's (1979) hypothesis was supported by the belief that a truncated version of an integrated MMTV provirus described by Majors & Varmus (1981) could be a reverse transcription product which had used a packaged MMTV env mRNA as template, since the deletion endpoints corresponded to conventional sites for splicing eukaryotic RNA (Linial & Blair, 1982).

Mo-MuLV particles have also been reported to contain low molecular weight viral RNA species which are homologous to all parts of the complete viral genome (Ball et al., 1979). These species possessed an average molecular weight of 1.6×10^6 daltons (4.6kb) but were only detected in virus grown in lymphoid-derived cell lines. Mo-MuLV propagated in fibroblast cell-lines did not contain this RNA (Ball et al., 1979).

The above examples suggest that it is entirely possible that Ki-MuLV particles may also package sub-genomic viral RNA species which are responsible for the synthesis of truncated DNA duplexes. If this is indeed the case, then cells infected at high multiplicity with this virus may occasionally receive an integrated version of an incomplete provirus.

This digression apart, the results of the analysis of in vitro VL30 DNA indicated that it was a suitable source of material for molecular

cloning. In order to obtain full length clones, it was decided to insert the VL30 DNA into its plasmid vector using molecular linkers containing the BamHI recognition site. Subsequent clones of VL30 DNA could then be easily distinguished from Ki-MuLV inserts during mini-cleared lysate analysis of plasmid DNA, since digestion by BamHI would release approximately 5kbp VL30 DNA inserts from the vector, whilst the largest Ki-MuLV derived fragment could not be larger than 3.5kbp (see the Ki-MuLV restriction map in figure R-17).

The following section discusses the characterization of the NVL clones which were obtained using *in vitro*-synthesized DNA.

D-3 Retrovirus-like features of VL30 NVL clones

Plasmid DNA extracted from amp^R, tet^S bacterial transformants was digested with BamHI. The DNA inserts released by this treatment were characterized by agarose gel electrophoresis and by hybridization with the Ki-MuLV/VL30 cDNA probe. Inserts were of several sizes. VL30 clones were initially identified by their approximately 5kbp length. Most of the smaller inserts were found to be BamHI fragments of Ki-MuLV DNA; the characterization of one of these in vitro-synthesized Ki-MuLV clones has been described elsewhere (Norton et al., 1984a).

Twelve VL30 cDNA clones (NVL clones) obtained using DNA synthesized from two different preparations of Ki-MuLV were successfully tested for their ability to hybridize with 30S virion RNA and in vivo-synthesized VL30 DNA. Further evidence that these clones were not merely examples of truncated Ki-MuLV molecules was provided by their complete lack of hybridization to the Ki-MuLV-specific components of these nucleic acid preparations.

On the basis of restriction enzyme sites all 12 NVL clones were shown to have LTRs, a feature which is also characteristic of retrovirus proviruses and transposable genetic elements (Shimotohno et al., 1980). In this respect the NVL clones resembled examples of VL30 elements cloned from BALB/c mouse genomic DNA (Keshet & Shaul, 1981; Foster et al., 1982). The NVL clones, however, would be expected to differ from randomly isolated genomic VL30 elements in two ways; (a) they were complementary to transcriptionally active members of the VL30 gene family, and (b) they represented VL30 RNAs which were both selectively packaged by murine retroviruses and were templates for reverse transcriptase. NVL-type VL30 elements should therefore be capable of transmission to a new host cell. Direct evidence for this was later obtained by the detection of newly integrated NVL DNA in the chromosomal DNA of Ki-MuLV- infected rat cells (section R-6).

Only four types of NVL clone could be identified by restriction enzyme mapping. Areas of sequence heterogeneity were confined to the LTRs, whereas the unique region which is flanked by these structures was common to all clones. This evidence for the close relatedness of retrovirus-transmissible VL30 elements was concordant with that of Clewley & Avery (1982) who detected a small degree of heterogeneity in a ribonuclease T1 oligonucleotide fingerprint analysis of virion VL30 RNA, and was also in keeping with the results of an analysis of *in vivo* and *in vitro*-synthesized VL30 DNA (sections R-1 and R-2). Keshet & Itin (1982) used a BALB/c mouse genomic VL30 clone as a probe to detect 100-200 VL30 elements in mouse DNA. These displayed a considerable restriction enzyme site diversity. Taken together, these data suggested that only a limited subset of the mouse VL30 family of genetic elements was transcriptionally active. Evidence to support this hypothesis was presented in section R-6, where NIH-3T3 mouse VL30 DNA was shown to be extensively methylated.

NVL-3 was the most abundant type of VL30 that was cloned (8 out of 12) and was therefore considered to be most representative of the NVL-VL30 family. A nucleotide sequence analysis of its LTRs and their adjacent regions identified several features that were structurally analogous to those of retroviruses (Norton *et al.*, 1984b). NVL-3 LTRs were 572bp long and each was bounded by a slightly imperfect inverted repeat about 9bp long. Using S1 nuclease mapping and the primer extension technique, transcriptional boundaries were identified which marked the borders of the R region with the U_3 and U_5 domains. As with retroviruses, sequences corresponding to the minus- and plus-strand primer binding sites were discovered in the unique region sequence at the boundary of the U_5 and U_3 domains respectively, and enabled 5' and 3' ends of the NVL clones to be identified.

One of the most significant discoveries concerned the sequence of

the minus strand primer binding site. This was complementary to the 3' terminal nucleotides of $tRNA^{gly}$. All known murine type C retroviruses, including Ki-MuLV, use $tRNA^{pro}$ as the virion RNA-associated primer for initiation of reverse transcription. This might provide a further explanation for the variable yield of VL30 DNA observed in preparations from Ki-MuLV-infected cells (compare figure R-1 with R-10), since it is known that the tRNA population packaged in retrovirus particles is not representative of that found in the host cell cytoplasm. In RSV, for instance, the priming species are generally dominant among those included in the particles (Sawyer & Dahlberg, 1973). Although not proven for murine retroviruses it seems that with avian retroviruses it is their binding affinity for the reverse transcriptase which is at least partly responsible for determining which tRNA species will be packaged (Haseltine et al., 1977b). Likewise, if Ki-MuLV reverse transcriptase causes the preferential encapsidation of $tRNA^{pro}$, on some occasions VL30 RNA species which are otherwise suitable templates may not be reverse transcribed through lack of the appropriate primer molecule.

LTR sequences of two BALB/c-derived integrated VL30 clones have recently been published. Although clone BVL-1 shares with NVL-3 a primer binding sequence which is complementary for $tRNA^{gly}$ (Hodgson et al., 1983), that of "clone 3" carries a $tRNA^{pro}$ binding site (Itin & Keshet, 1983b). Obviously more VL30 clones must be sequenced before the true significance of these findings can be assessed.

Norton et al. (1984b) were able to align the NVL-3 LTR sequence with that of the clone 3 VL30 element derived from BALB/c mouse DNA. Most sequence homology occurred within the U_3 region of the two LTRs. In figure D-1 is shown a diagrammatical representation of these two LTRs, in addition to that of another BALB/c derived clone, BVL-1 (Hodgson et al., 1983). As described above, the functional domains of

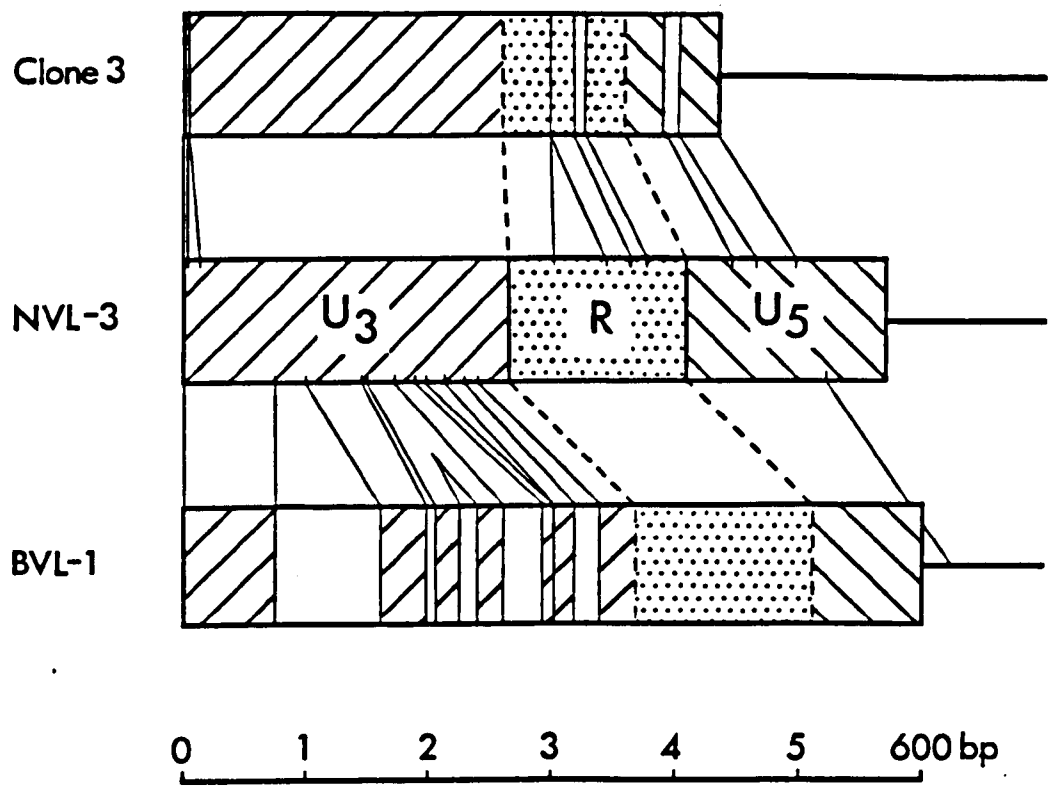


Figure D-1 Nucleotide sequence comparison of three VL30 LTRs. Sequences of BALB/c genomic DNA VL30 clones 3 (Itin & Keshet, 1983b) and BVL-1 (Hodgson et al., 1983) were compared visually with that of clone NVL-3 (Norton et al., 1984b). Borders between the R and the U_3 or U_5 domains have been determined for NVL-3 by S1 mapping and primer extension (Norton et al., 1984b) and are tentatively identified by sequence homology in the other two cloned DNAs (vertical dashed lines). Thin lines joining the boxed LTR structures indicate areas of sequence homology. Unshaded or unstippled areas of clone 3 or clone BVL-1 indicate stretches of sequence not present in clone NVL-3. Thick horizontal lines indicate 5' unique sequence. The scale indicates length of DNA in base pairs.

the NVL-3 LTR have been identified (Norton et al., 1984b) and are labelled according to standard retrovirus nomenclature. The boundaries of the U_3, R and U_5 domains in clone 3 and BVL-1 have not been defined experimentally, but can be recognized by comparison with the NVL-3 sequence. Inspection of figure D-1 shows that the NVL-3 LTR contains large, uninterrupted stretches of sequence homology with the 5' half of the clone 3 LTR, and with the 3' half of the BVL-1 LTR. The abrupt change from one type of LTR sequence to that of another immediately suggests that a recombination event has occurred in the generation of NVL-3 type VL30 units. It is possible that this event has taken place either at the DNA or at the RNA stage of VL30 duplicative transposition.

The former possibility could occur if two closely situated VL30 proviruses were to homologously recombine at their LTRs, such that part of the 3' LTR of one unit was replaced by LTR sequences from the other. For example, the recombinant 3' LTR might now possess the U_3 and R region of the second VL30 unit, whilst retaining its original U_5 region. Restriction mapping studies of genomic DNA VL30 units (section R-4) suggested that VL30 units with different LTRs may not be uncommon. Transcription of an actively expressed, integrated provirus commences at the 5' boundary of the R region in the 5' LTR. Thus the RNA transcript of the hybrid VL30 unit would contain U_3 sequences homologous only to the newly acquired region. Upon reverse transcription, the resulting DNA duplex would now possess the hybrid LTR at each end of the provirus.

The second mechanism by which the NVL-3 type of LTR may have occurred involves recombination during reverse transcription. It is conceivable that a retrovirus has packaged a mixture of VL30 RNA templates which included examples of both BVL-1 and clone 3-type elements. Coffin (1979) has proposed that recombinant retrovirus genomes may be generated by switching of the reverse transcriptase from one RNA component of a heterodimer to another. Since the homology of

the NVL-3 LTR with those of the other two VL30 elements seems to be suddenly disrupted within the R region, or at its border with the U_3 or U_5 regions, a somewhat more simple form of recombination may have taken place. The proposed mechanism by which this could occur is as follows: (1) reverse transcription of a BVL-1 type unit is initiated as normal at the tRNA primer site, and the minus-sense "strong-stop" DNA is synthesized. (2) Instead of the first transfer of templates occurring between the 5' end of a BVL-1 type RNA to the 3' end of another (or the same) BVL-1 RNA molecule, this strong-stop DNA becomes attached to the 3' end of a clone 3-type RNA template. For this to occur there must be at least partial homology between the two types of template within their 'R' regions. (3) Reverse transcription now proceeds normally, and will result in a linear, double-stranded DNA duplex with the BVL-1 U_5 region included in each LTR. This occurs because the U_5 domain of the viral RNA template is only used once during reverse transcription, ie to generate the minus-sense strong-stop DNA. There are at least two major postulated mechanisms for reverse transcription (eg Dina & Benz, 1980; Gilboa et al., 1979a) but in each case the second U_5 domain is synthesized either directly from the minus-sense strong-stop template, or indirectly, by using the plus-sense strong-stop DNA as template (ie by displacement of the latter from its own, minus-sense template; see General Introduction for a schematic of these two models).

That the NVL-3-encoding provirus has become genetically fixed by its incorporation into germ line DNA implies that such events may be relatively common in somatic cells. Moreover, since VL30 LTRs contain sequences necessary for transcriptional regulation (Hodgson et al., 1983; Norton et al., 1984b) homologous recombination at the RNA or DNA stage may be a mechanism whereby new VL30 LTRs with different degrees of transcriptional activity can be generated. The LTRs of endogenous avian retroviruses, eg RAV-0, ev-1 and ev-2 are generally shorter than those

of exogenous retroviruses, such as RAV-2, and also contain promoter sequences which are much less efficient that those of RAV-2 LTRs (Cullen *et al.*, 1983). A similar range of activity could account for the variation in frequency of isolation of the four types of NVL clone. Alternatively, efficiency of virion packaging and reverse transcription may provide the necessary explanation for their apparent relative abundance.

D-4 VL30 proviruses in NIH-3T3 DNA

No detailed analysis of integrated mouse VL30 elements has yet been described. To date, conclusions regarding the organization of this family in mouse DNA have been drawn from the results of brief surveys using one or two restriction enzymes, Southern blotting and a representative or a single sub-genomic VL30 probe (Courtney et al., 1982b; Keshet & Itin, 1982; Itin et al., 1983). DNAs from several different species of Mus have been shown to contain VL30-related sequences, but the only evidence that these may be organized into provirus-like structures was derived from the restriction enzyme analysis of a few randomly cloned examples (Courtney et al., 1982a; Keshet et al., 1980). Although the results described provided useful data, it was felt that a more detailed analysis involving the majority of integrated VL30 elements was still necessary in order to form a basis for understanding their role (if any) in the cell and the mechanisms of molecular evolution by which they have been generated. The work described in section R-4 and discussed here used modifications of the Southern blotting technique to investigate in detail those VL30 elements which are found in high molecular weight NIH-3T3 DNA.

Hybridization profiles of DNA which had been digested with a single restriction enzyme were of a complexity which suggested a VL30 copy number of about 150 per haploid genome (assuming each unit was approximately 5kbp long). This result was in agreement with the data of Keshet & Itin (1982) who detected 100-200 VL30 elements in BALB/c mouse DNA, and thus identified VL30 as the second most abundant family of endogenous retrovirus-like elements in Mus musculus DNA [the A-type particle genes number approximately 1000 per haploid genome (Lueders & Kuff, 1977)].

Initial studies used enzymes which recognized hexanucleotide sequences of DNA and which should therefore cut a random sequence once

in every 4^6 (4096) base pairs. The observed VL30-specific digestion profiles, however, were markedly different with each enzyme used. In some patterns most bands were clustered in the region of the gel corresponding to a size of 5kbp or greater, and others consisted of a background signal which was occasionally punctuated by low molecular weight bands of an intensity which suggested they comprised multiple copies (ie 10-50) of fragments. These findings agreed closely with those of others (Courtney et al., 1982b; Keshet & Itin, 1982) and implied that (a) most VL30 units were indeed of 5kbp or longer, and that (b) many restriction enzymes consistently cut without these elements while others recognised internally located sites.

Some of the VL30-specific hybridization patterns described by other workers contain bands of very similar or identical electrophoretic mobility to the NIH-3T3-derived ones. Results agreeing closely with those reported here include those of Courtney et al. (1982b) who used EcoRI to digest the DNA of several species of Mus to show that most of the VL30-specific signal was associated with digestion fragments >4kbp in length. Keshet & Itin (1982) detected two major XbaI bands in the DNA of several strains of Mus musculus which appeared to be identical to the 4.7 and 3.45 kbp fragments described in detail in section R-4. Furthermore, HindIII digestion of the same DNA samples generated a major 1.3kbp band which was reminiscent of a similarly prominent band in NIH-3T3 DNA digestion profiles. In contrast to the M. musculus results, this band was absent from the profiles of DNA from mice such as M. pahari, M. platathryx, M. shortridgei and M. cervicolor which have been genetically separate from M. musculus for 3-4 million years (Itin et al., 1983; Benveniste et al., 1977).

Sub-genomic VL30 probes were used to analyze NIH-3T3 DNA in more detail. These derived from four contiguous sections of the NVL-3 genome, one homologous to the LTR and three to regions of the unique

sequence. High-copy genomic DNA VL30 restriction fragments could now be mapped to the appropriate position on a generalized VL30 structure, and for most enzymes the simplest interpretation of the results confirmed that the majority of mouse VL30 sequences were organized into complete provirus-like structures.

Complex hybridization patterns produced with the LTR-specific probe suggested that sequences which flanked these proviruses were not related to each other, in other words the VL30 elements were scattered apparently at random throughout the mouse genome. Keshet & Itin (1982) reached a similar conclusion after using mouse-hamster hybrid cell lines to show that VL30 copies were distributed among different chromosomes, although these workers could not dismiss the possibility that this distribution might be a non-random one. Other mouse endogenous retrovirus-like families are also spread throughout chromosomal DNA, with no evidence of any tandem arrangements of proviruses (Coffin, 1982).

Deleted forms of proviruses have been characterized in studies of the endogenous retrovirus families of both mouse and chicken (Tereba, 1983; Rovigatti & Astrin, 1983) but the mapping characteristics of a few VL30-specific restriction fragments indicated that an insertion of foreign DNA into a VL30 provirus may have taken place, followed by amplification of this hybrid element. A 2.5kbp SacI band hybridized with NVL probes 1 and 2, which derived from a contiguous portion of NVL-3 sequence spanning only 1.5kbp (figure R-20C). Absence of signal with probe 3, which is homologous to that portion of unique sequence immediately adjacent to probe 2, suggested that at least 1kbp of DNA had been inserted into the 5' end of a VL30 unit. Figure D-2 depicts a hybrid VL30 element which could explain this result.

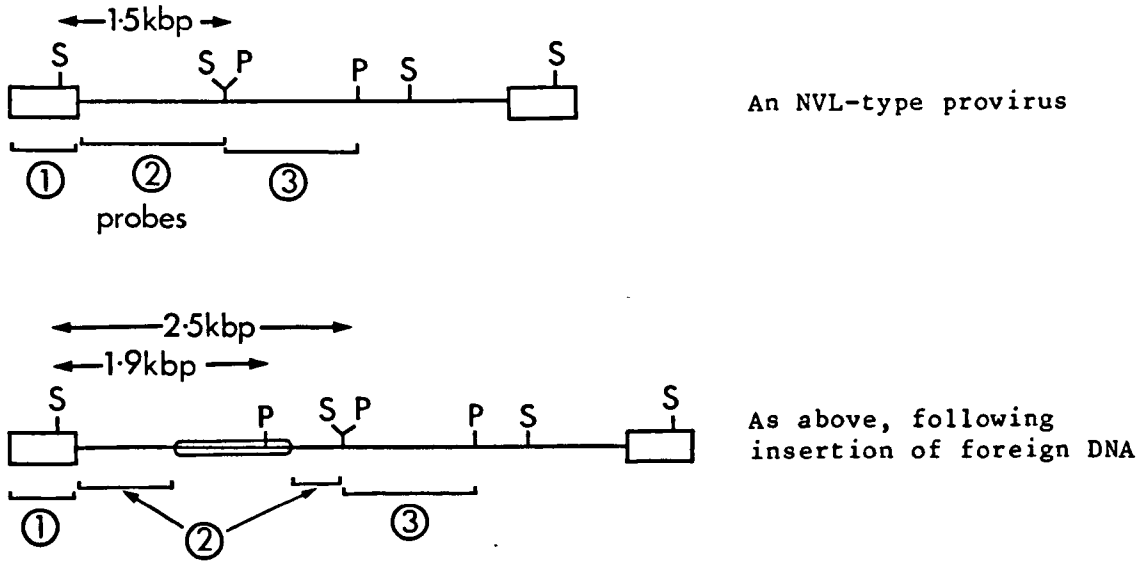


Figure D-2 An interpretation of SacI digestion data

Double-digestion of NIH-3T3 DNA with PstI and SacI also produced a fragment which hybridized with probes 1 and 2 but not with 3 (the 1.9kbp band of figure R-20E). This might derive from the same hybrid element, especially if the foreign sequence possessed a PstI site (as indicated in figure D-2). This is just one of several possible explanations. It is conceivable that a VL30 unit has been amplified together with its 5' flanking sequence, such that the 2.5kbp SacI fragment extends from the VL30 DNA to a SacI site in this amplified flanking sequence. Such a hypothesis must also assume the deletion of a highly conserved SacI site in the LTR.

VL30 units containing an insertion of foreign information have been molecularly cloned from BALB/c mouse DNA (Itin & Keshet, 1983a). These elements included VL30 termini which bordered two sets of sequence that possessed extensive homology to the MuLV gag and pol genes (figure D-3).

That the MuLV inserts were ordered in the same 5' to 3' orientation as the VL30 element prompted the authors to suggest that recombination had occurred within a retrovirus particle during reverse transcription of a heterodimer molecule (Coffin, 1979). Hybrids of this type represent the

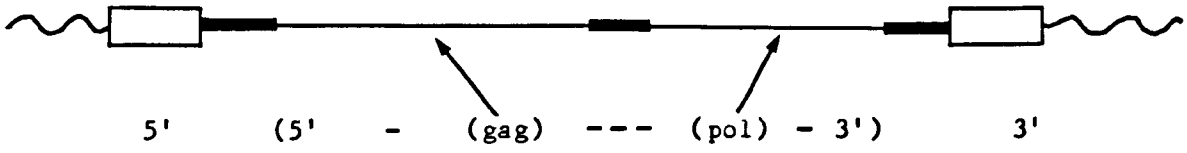
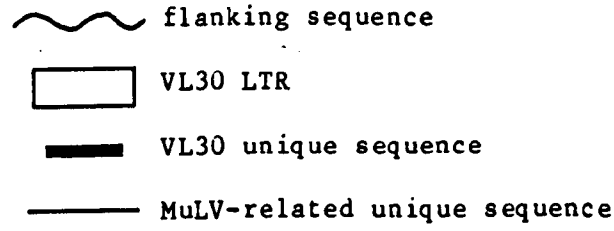


Figure D-3 Sequence organization of the VM-1 clone (Itin & Keshet, 1983a)



reciprocal of those recombination events which generated the genomes of Ha-MuSV and Ki-MuSV, as these possess MuLV-derived terminal sequences which border parts of a rat VL30 element containing the ras oncogene (Shih et al., 1978; Chien et al., 1979). It is conceivable that the NIH-3T3 DNA restriction fragments discussed above derive from the same or a similar family of elements as exemplified by the VM-1 clone in figure D-3. Certainly the autoradiographic signal intensity of these bands suggest that they exist as multiple copies in the mouse genome

(figure R-20).

Whatever their origin, most of the data described in section R-4 indicated that these VL30 elements must be a minority, since the members of the VL30 family appeared to be very closely related. This belief was strengthened by an analysis of high copy VL30 restriction fragments which had been removed from their normal context by preparative agarose gel electrophoresis. Isolated thus from the bulk of more variant VL30 elements, these DNA preparations could be subjected to a restriction enzyme analysis in which specific questions concerning VL30 evolution could be answered. Two VL30 sub-sets, defined by certain XbaI sites, and together constituting 60-70% of the total mouse VL30 complement were compared for their possession of several previously mapped restriction sites. The results (summarized in figure R-22) indicated that no particular region of the VL30 provirus could be identified as a mutational hot-spot. It was concluded that during the evolution of these elements in *Mus musculus* small sequence changes in individual proviruses must have occurred randomly throughout the VL30 genome. In this way restriction digest profiles of VL30-specific DNA appear to be complex but actually represent rather modest evolutionary change over a period of tens of millions of years (see section D-5). In their study of the evolution of type C viral genes, Benveniste & Todaro (1976) also concluded that endogenous "virogenes" evolve at rates approximating that of unique sequence cellular DNA. They contrasted endogenous proviruses to the genomes of horizontally transmitted viruses which unlike the former are under a greater selection pressure for their ability to replicate. These viruses might therefore be expected to acquire mutations or recombine with other infectious retroviruses at a much greater frequency than endogenous viral genes (Benveniste *et al.*, 1977). Thus, as seems also to be the case with VL30 genes, the close relatedness of members of these endogenous provirus families might more

closely reflect the evolutionary conservation of genes required for non-replicative, cellular functions (whichever the latter turn out to be).

Analysis of the gel-purified VL30 restriction fragments failed to provide evidence of a strongly represented NVL-like group within the mouse VL30 family. It was shown that unlike the vast majority of family members, which were 5.2-5.3kbp long, NVL-encoding units ranging from 4.6-4.9kbp must be in a very small minority. This was estimated to be no more than 5 copies per haploid genome (approximately 4% of the total). NVL units were also different by virtue of their LTR restriction enzyme sites. These were shown to be unusual in comparison to those shared by most VL30 LTRs. A restriction site may be generated or removed by the alteration of a single nucleotide, but if this occurred in a sequence concerned with transcriptional regulation the consequences could be profound. As discussed in section D-3, nucleotide sequencing studies of VL30 LTRs have shown that such functions are a part of VL30 LTR information; perhaps the rare NVL-type LTR restriction site differences are associated with the ability to express 30S RNA. Certainly the methylation status of most VL30 units suggests that few can be transcriptionally active (section R-6). As discussed in the General Introduction, such a phenomenon is not uncommon with endogenous proviruses, most of which are subject to cellular controls that render them transcriptionally silent (see Coffin, 1982). This is not to say that all non-NVL VL30 elements are forever incapable of expression as a consequence of their sequence divergence. NIH-3T3 DNA and C3H-10T½ DNA generate virtually identical VL30-specific hybridization profiles when digested with restriction enzymes (see figure R-23). However, Sherwin *et al*. (1978) failed to detect any VL30 RNA expression in C3H-10T½ cells, and Courtney *et al*. (1982a) could only detect a very limited expression (the equivalent of 14 VL30 RNA molecules per cell). This reinforces the idea that the functional capacity to encode 30S RNA may

be present in many units, but that in most cases, expression is strictly controlled. Whatever mechanisms the cell uses to repress the production of VL30 RNA in C3H-10T½ cells, this can be partly overcome by chemical transformation. Thus, methylcholanthrene-treated cells were found to contain 15-fold higher levels of VL30-related RNA than did their non-transformed counterparts (Courtney et al., 1982a). It is still not clear, however, whether this treatment merely enhanced the expression of VL30 elements which were already (slightly) transcriptionally active, or whether it induced the expression of previously silent VL30 genes. Aspects of VL30 RNA expression will be discussed more fully in section D-6.

D-5 Evolution of the rodent VL30 family

A Southern blotting survey of high molecular weight DNA samples revealed that sequences related to MuLV and VL30 were distributed amongst several evolutionary distant rodent genomes. MuLV- and VL30-specific bands were detected in the hybridization patterns of rat, vole and mouse DNA, indicating the possibility that these genetic elements were present in the genome of a rodent ancestor to the families Cricetidae and Muridae, which diverged 20 million years ago (Benveniste *et al.*, 1977; Chaline, 1977). MuLV-related restriction fragments were also detected in grey squirrel DNA; since the last common ancestor of the squirrel, mouse and rat existed at least 60 to 65 million years ago, it can be tentatively suggested that these retrovirus genes are of a similar age (Chaline, 1977; see figure D-4).

Thermal stabilities of MuLV or VL30 probe:cellular DNA duplexes were determined using dot-blotted DNA in order to assess the extent to which the two classes of genetic element have been conserved in each rodent genome. The results confirmed those of the Southern analysis, and further suggested that each family has evolved at a different rate. Although MuLV sequences are less highly reiterated in the genome of *Mus musculus* (up to 50 copies) than are VL30 units (150 copies), in other rodent genomes their copy number and nucleotide sequence homology to the MuLV probe indicated that they enjoyed the greater degree of evolutionary conservation. The relative extent of this conservation seemed to reflect the evolutionary trends of the order Rodentia, whereas that of VL30-related sequences indicated a dramatic amplification and homogenization of this family during the last 4 to 5 million years of mouse speciation (Benveniste *et al.*, 1977; Chaline, 1977). Interestingly the dot-blot analysis also revealed the existence of sequences more distantly related (88% homology) to mouse VL30 in the squirrel genome. The smear of background-type hybridization signal

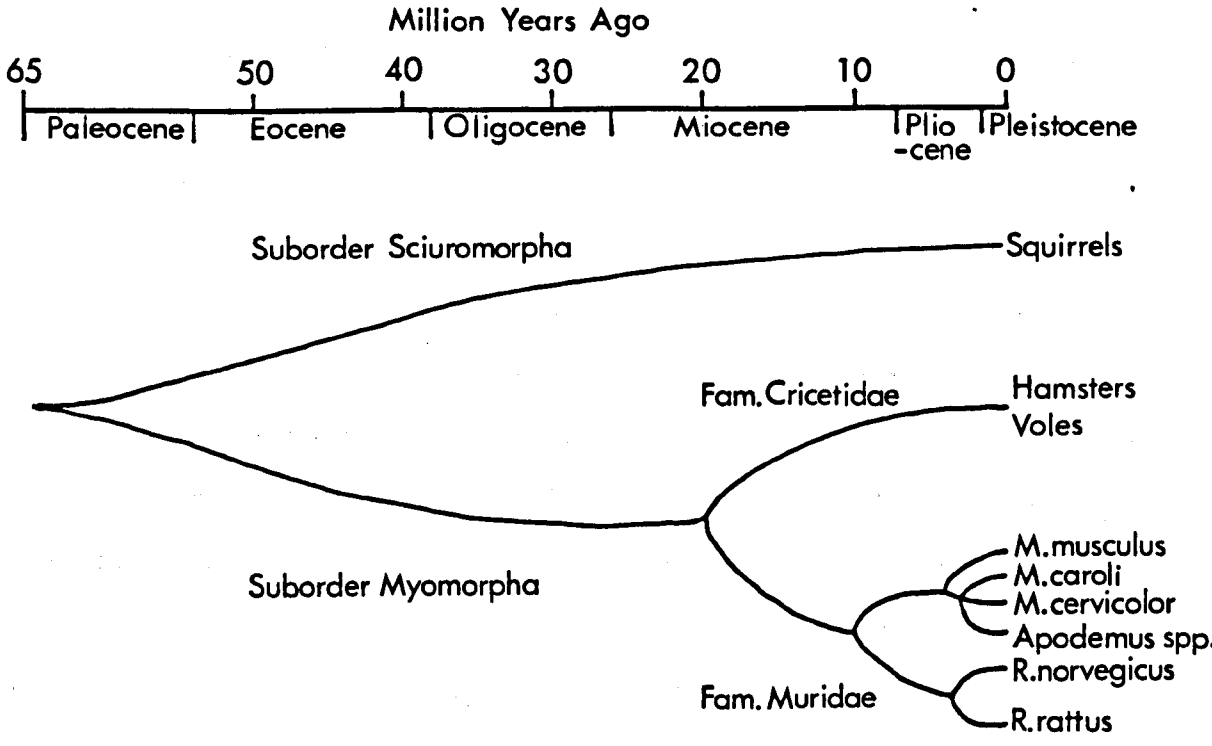


Figure D-4 Simplified phyletic tree of the Order Rodentia. After Benveniste *et al*. (1977) and Chaline (1977).

obtained from the Southern blot experiment with this DNA must therefore have been at least partly due to these weakly hybridizing components. In retrospect, since the conditions of hybridization were such that only duplexes sharing greater than 85% homology could form, it is perhaps not surprising that there were no discrete bands in the squirrel DNA digestion profile. Nevertheless, the identification of these VL30-related sequences placed the apparent age of this family on a par with that of MuLV.

Evidence exists which could be interpreted as suggesting an even greater antiquity for these two classes of endogenous proviruses. MuLV-related sequences have been detected in gibbon apes, a woolly monkey (Benveniste et al., 1977) and African green monkey DNA (Martin et al., 1981) whilst weak homology has been reported to exist between a genomic DNA mouse VL30 probe and dot-blots of rat, hamster, monkey and human DNA (Itin et al., 1983). In an attempt to repeat Itin's (1983) finding, human DNA was also included in the dot-blot experiment reported here, but the signal obtained was indistinguishable from the background level (data not presented).

Interpretations of surveys for retrovirus-like sequences in distantly-related DNAs must, however, be treated with caution. Many endogenous proviruses are or were at one time closely related to infectious viruses, some of which may have displayed a very wide host range. The present members of the xenotropic MuLV family, for instance, can infect rat, mink, human and even quail cells (Levy, 1975; 1978). Several examples have been described which demonstrate that in the past retroviruses have not always respected the species barrier. These inter-specific infections therefore complicate any determination of the evolutionary age of a retrovirus-like family that is based on the detection of proviruses in the genomes of various taxonomic classes.

Studies by Benveniste and his colleagues (Benveniste & Todaro,

1973; Benveniste et al., 1974) have shown that the SSAV/GALV group of infectious primate type C viruses are not in fact endogenous to the primate genome, but probably originated by trans-species infection of gibbons with an endogenous, xenotropic type C virus from the Asian mouse Mus caroli or one of its close relatives (Lieber et al., 1975).

Another primate retrovirus, the baboon endogenous virus (BaEV) was unexpectedly found to be serologically and genetically related to the endogenous feline type C virus, RD114 (Hellman et al., 1974; Sherr et al., 1974). BaEV sequences can be detected in the genomes of all Old World monkeys that have been tested, in addition to the DNAs of the domestic cat and European wildcat, but not in those of apes or of New World monkeys (see Teich, 1982).

Similarly in birds the RAV-0 group of endogenous chicken retroviruses also shows a distribution which does not coincide with established phylogenetic relationships. Thus the DNA of three junglefowl species which are closely related to Gallus gallus (the domestic chicken) were shown to lack RAV-0 proviruses, whilst the DNA of two more distantly related Phasianus species (pheasant and quail) shared a significant homology with the RAV-0 probe (Frisby et al., 1979).

In addition to the dangers involved in drawing conclusions about retrovirus evolution from the phylogenetic relations of their various host species, endogenous retroviruses also exhibit a remarkable degree of fluidity within the genomes of individuals of a particular host species. The number and distribution of endogenous proviruses can vary greatly from individual to individual. For example 16 distinct avian retrovirus loci (ev loci) have been characterized from a variety of inbred and outbred White Leghorn chicken flocks (Rovigatti & Astrin, 1983), and although ev 1 has been reported to be present in 506 out of 508 White Leghorn embryos examined (Tereba & Astrin, 1980) the other loci are much more variable, such that individuals usually contain

between 0 and 6 proviruses (Coffin, 1982). Studies conducted on ev loci in breeds of chicken other than the White Leghorn have indicated an even greater variability in the number and location of proviruses per chicken (Hughes et al., 1981b).

However even sublines or individuals of some highly inbred lines of mice can vary in their content of endogenous proviruses, apparently as a result of new germ-line integration during the breeding programme. This phenomenon is probably peculiar to certain viraemic strains of mice, for instance those derived from Japanese breeds which contain ecotropic viral loci. Unlike the majority of xenotropic MuLV proviruses, which are more common, these are frequently expressed and can re-infect their host. Thus Rowe & Kozak (1980) showed that the number of AKR MuLV genomes were gradually increasing in AKV-1 congenic mice, presumably by infection of embryos or gonadal tissues.

Mere variation in provirus number and their chromosomal positioning, nevertheless, could not fully account for the results obtained when sub-genomic NVL probes were used in a Southern analysis of rat and vole DNA (section R-5). These clearly demonstrated that different portions of the mouse VL30 genome had been conserved in each rodent genome, moreover as multiple copies of each sub-set. Both DNAs lacked NVL LTR-related information, but rat DNA possessed two discrete sets of sequence which were homologous to LTR-adjacent areas, while vole DNA:VL30 signal was due to hybridization with the central part of the NVL unique region. The fact that specific portions of VL30 information have been conserved in each rodent suggests that these experimental observations are valid for the respective species as a whole and are not due to the type of individual variation described above. If the latter were true, one might expect to see a greater diversity of conserved VL30 genome fragments.

The results of this VL30 analysis are remarkably similar to those

obtained by Lueders & Kuff (1983) in a study of the sequence organization of IAP genes in the mouse, rat and hamster genomes. Sequence relationships between the three families were determined by measuring the thermal stabilities of heteroduplexes formed between cloned examples of IAP genes, also between these clones and their respective genomic DNA. Like the mouse IAP family (Lueders & Kuff, 1980) that of the Syrian hamster consisted of relatively homogeneous, well-conserved units. In contrast the rat IAPs appeared to be heterogeneous units composed of several subsets of sequences, only some of which were related to mouse IAP genes. As with the VL30 example described here, no mouse IAP-related LTR sequences were detected in the rat and hamster genomes. Of interest was the discovery that although R. norvegicus and R. rattus IAPs were heterogeneous in organization, these two species of rat possessed a similar IAP sequence copy number. However IAP gene copy number in the Chinese and Syrian hamster genomes was quite different, IAP sequences being barely detectable in the former. As with the endogenous RAV-0 proviruses in bird genomes (Frisby et al., 1979) these authors concluded that amplification or homogenization of IAP genes must have occurred independently and at different periods of time during their evolution (Lueders & Kuff, 1983). Incidentally, no sequence homology exists between members of the IAP and VL30 proviral families (Lueders & Kuff, 1979; Keshet et al., 1980).

The current organization of mouse VL30-related sequences in rat and vole DNA may be a result of the several molecular mechanisms of gene turnover which are postulated to be the cause of evolutionary rapid changes to members of multi-gene families (Brown & Dover, 1981). Equally possible, the apparently radical VL30 sequence alterations may be entirely due to the considerable ability of retrovirus-like genetic elements to form recombinant genomes. For example, the B-type M432 retrovirus of M. cervicolor and the closely related M. cookii was found

to contain a major 3.7kbp region of homology to a M. musculus type of IAP genome (Kuff et al., 1978; Callahan et al., 1981). Since the 5' end sequences of the M432 genome can only be detected in the cellular DNA of its two host mouse species, this virus probably arose as a result of recombination between an IAP unit and an unidentified class of retrovirus-like element (Callahan et al., 1981).

The interactions of rat VL30 sequences with those of MuLV to form the defective Ha-MuSV and Ki-MuSV genomes have been the subject of much study (Young et al., 1980; Ellis et al., 1980, 1981; Norton & Avery, 1982; Norton et al., 1984c). Since the mouse and rat VL30 families are structurally and genetically related, it seemed possible that mouse VL30 sequences could have been involved in the generation of mouse-derived transforming viruses such as Abelson-MuLV, Mo-MuSV and BALB-MuSV. However no sequence homology between VL30 and these viruses could be detected (Howk et al., 1978; Besmer et al., 1979).

It is probable that the sequences detected in rat DNA by the NVL probes form part of the genomes of a rat VL30 subset, since Itin et al. (1983) showed that cross-hybridization between a BALB/c mouse VL30 clone and rat genomic DNA was fully accounted for by homology between mouse and rat VL30 elements. This region of homology was confined within a 1kbp stretch of their VL30 clone unique sequence adjacent to the 3' LTR. Courtney et al. (1982b) and Giri et al. (1983) on the other hand identified two regions of their BVL-1 VL30 clone that hybridized with rat VL30 DNA. These mapped to unique sequence regions that are similar in location to the rat DNA-related sequences in NVL-3. The extent of the relationships between Ha-MuSV, a cloned rat VL30 element and the three mouse VL30 clones is depicted in figure D-5. It can be seen that the region of rat VL30 DNA which hybridizes with Keshet's VL30 clone 3 is different from the two regions which are homologous to portions of the BVL-1 VL30 genome. At present the exact relationship between NVL-3

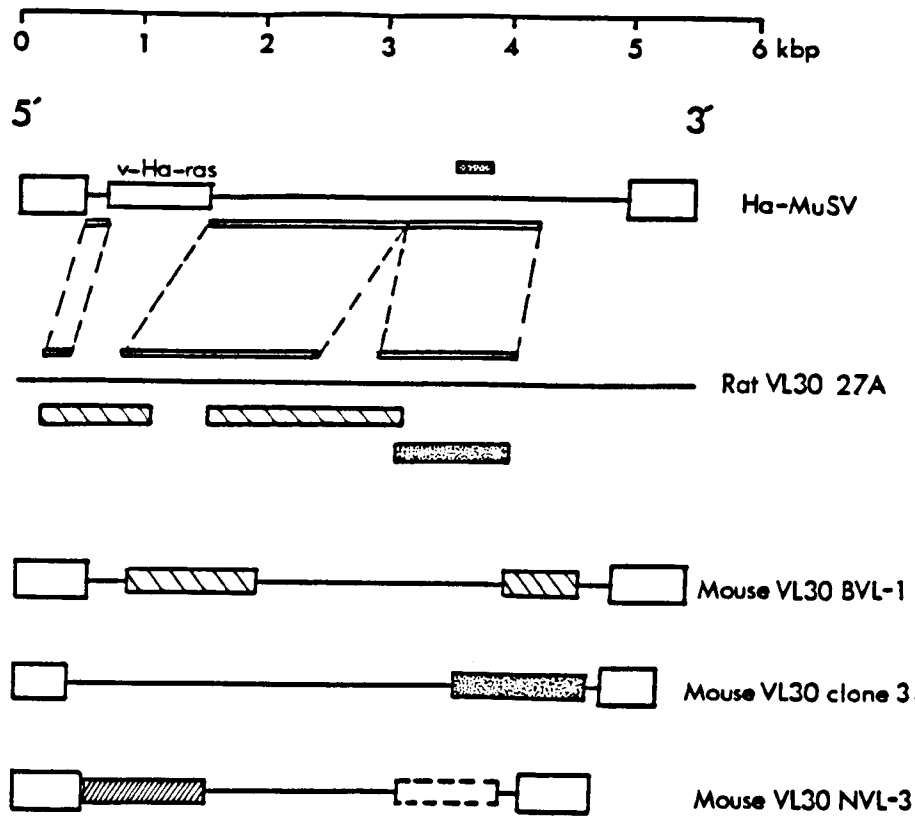


Figure D-5 Sequence relationships of Ha-MuSV, rat VL30 and mouse VL30 elements. Approximate locations of sequence homology between a cloned example of Ha-MuSV DNA and rat VL30 clone 27A were determined by heteroduplex and hybridization analysis (Ellis et al., 1980; Young et al., 1980)). Areas of homology between clone 27A and mouse VL30 clone BVL-1 (Courtney et al., 1982b; Giri et al., 1983), between clone 27A and mouse VL30 clone 3, and between Ha-MuSV DNA and clone 3 (Itin et al., 1983) were determined by Southern transfer hybridization analysis, and are indicated by appropriately positioned boxes sharing a similar shading pattern. Boxed areas of NVL-3 unique sequence represent regions that hybridize with rat (NRK) genomic DNA (see fig. R-25). Large open boxes indicate LTR structures; these have not yet been identified for clone 27A.

and rat VL30 DNA has not been fully characterized, but studies described in section R-6 failed to reveal homology between an NVL-3 probe and uninfected NRK cellular RNA. Although this implies that the NVL-3-related sequences present in NRK genomic DNA do not form part of transcriptionally active rat VL30 units, it is conceivable that the conditions used in this experiment were not sensitive enough to detect this expression.

Recently a nucleotide sequence comparison of Ki-MuLV and Ki-MuSV DNA has identified the most likely points of recombination between Ki-MuLV and a rat VL30 element which resulted in the generation of Ki-MuSV (Norton et al., 1984c). At both 5' and 3' ends of the rat VL30 insertion there were short regions of interrupted homology and non-homology before complete divergence from the Ki-MuLV progenitor sequence. As stated by the authors, these could be the remnants of more extensive regions of homology between the two types of genetic element which allowed a homologous recombination event to take place. As mentioned in section R-5, Giri et al. (1983) detected weak regions of homology between cloned examples of rat and mouse VL30 elements and discrete parts of the MuLV genome. These mapped in the same 5' and 3' areas of the MuLV genome that are implicated in the genesis of Ki-MuSV (Norton et al., 1984c).

Although no known mouse-derived infectious virus genomes carry regions of mouse VL30 sequence (Howk et al., 1978; Besmer et al., 1979), Itin & Keshet (1983a) have described an endogenous retrovirus-like element whose organization appears to represent the reciprocal arrangement to that of Ha-MuSV or Ki-MuSV (see figure D-3). Here, VL30 LTRs and terminally located unique sequences border an insertion of MuLV-related unique sequences. It is not known whether such an element is capable of expression, but the discovery of several hybrids of this type in a mouse gene library suggests that they occur at a high

frequency over evolutionary time periods (Itin & Keshet, 1983a).

It is possible that the discrete subsets of NVL-related sequence discovered in rat and vole genomic DNA are the result of a similar recombination event. Conceivably, a single hybrid virus was formed by the recombination of a mouse VL30 unit and an unknown genetic element such that only VL30 unique sequence was incorporated into the new genome. Following integration the novel virus in each species was able to amplify itself by successive rounds of replication and germ-line re-integration. Alternatively these putative hybrid viruses may be the amplified remnants of a retrovirus-like genome formed by "copy choice" of a reverse transcriptase using a heterodimer of VL30 and the unknown genetic element RNA as alternate templates (Coffin, 1979).

Another possibility need not invoke the mechanisms of recombination available to a retrovirus-like element. VL30 units may have evolved at different rates in different rodent species via gene conversion and/or transposition events. Dover (1982) suggests that these processes of gene family turnover can be either random or directional in nature. His "molecular drive" hypothesis explains how, in the directional mode, one particular variant (which could be a complete unit or just a fragment of one) may become fixed in germ-line DNA at the expense of the original family members. Furthermore, where the rate of gene family homogenization within individuals is slow compared to that at which the pool of chromosomes "shared" by a population of sexually reproducing individuals is being effectively randomized, then molecular drive over thousands of generations could lead to a simultaneous increase of that variant in all the individuals of that population (Dover, 1982). Clearly, this particular aspect of VL30 evolution requires the detailed characterization of several VL30-related molecular clones from different rodent genomes before the relevant hypothesis can be identified.

D-6 VL30 RNA expression in mouse and Ki-MuLV-infected rat cells

Two experiments described in section R-6 used the MspI/HpaII system to assess the methylation status of integrated VL30 proviruses in mouse and rat DNA, respectively. Although the expression of some viral genes appears to be methylation insensitive (Graessmann et al., 1983) as a general rule retrovirus proviruses that are actively transcribed are undermethylated, whereas those that are not transcribed are highly methylated at their CpG dinucleotides (Cohen, 1980; Groudine et al., 1981; Harbers et al., 1981). In NIH-3T3 mouse cells most of the VL30-related sequences were methylated at their HpaII restriction sites. However, preliminary evidence suggested that at least some restriction fragments were derived from undermethylated DNA sequences. Using NVL-3 DNA fragments as comparative markers, it was concluded that 3-5 copies of NVL-like elements were undermethylated, and might therefore be capable of expression.

This observation was consistent with the finding that VL30 cDNA clones were structurally very similar to one another, and suggested that the NVL type of RNA may be the major class of VL30 RNA expressed in NIH-3T3 cells.

High molecular weight DNA from the Ki-MuLV (VL30)-infected rat cell line KLNRK-1 was found to contain recently introduced Ki-MuLV and VL30 proviruses. The latter were all (not unexpectedly) of the NVL class, and were present at 1-2 copies per rat cell.

When RNA from KLNRK-1 cells was examined, it was found to include mouse VL30-related species which were undetectable in uninfected NRK cells. The agarose gel profile of these VL30 RNAs was remarkably similar to that of VL30 RNA extracted from NIH-3T3 cells, a further indication that NIH-3T3 cells express predominantly NVL-type RNA. Consistent with the presence of NVL-type RNA in KLNRK-1 cells, the newly introduced NVL proviruses were unmethylated at their HpaII sites, unlike

the endogenous mouse VL30-related sequences.

The VL30 family shares with mouse IAP genes similar patterns of sequence organization and distribution amongst distantly related rodent species (section D-5). This similarity may be extended to certain characteristic features of RNA expression which are shared by the two classes of genetic element. In the following discussion it will be shown that, like VL30 elements, IAP genes may be ordered into discrete sub-sets of proviruses each with their own expression characteristics. Furthermore both types of gene family seem to be implicated, by their patterns of RNA expression, with the transformed phenotype.

IAPs are retrovirus-like particles encoded in Mus musculus DNA by about 1000 genes (Lueders & Kuff, 1977). Large amounts of particles are present only in early (two-cell stage onwards) mouse embryos and a variety of mouse tumour cells. This has suggested a possible link between IAP expression and embryonic development and differentiation (Biczysko et al., 1973; Calarco et al., 1973; Chase & Piko, 1973). In neuroblastoma cells most of the IAP RNA is expressed by the major type of provirus, which is 7kbp long and is highly conserved (Lueders & Kuff, 1980; Kuff et al., 1981). However a smaller subset of 90 genes has been identified (Ono et al., 1980). A still smaller portion of this subset, representing 2% of the total IAP sequences in mouse DNA, expresses the majority of IAP RNA in mouse myeloma cells (Shen-Ong & Cole, 1982, 1984). Rather reminiscent of the VL30 NVL sub-set, these minority IAP genes are significantly smaller than average, their 4.8kbp size apparently due to a major deletion in their 5' unique sequence region (Shen-Ong & Cole, 1982). IAP RNA is expressed as two major size classes, 29-30S and 35-36S. Hojman-Montes de Oca et al. (1983) showed that different levels of each RNA species could be produced in stem embryonic carcinoma cells caused to differentiate by different methods. Again this suggested the presence of different populations of IAP genes

within the family, each induced to express in different cellular environments.

Mouse VL30 RNA expression has also been correlated with the transformed phenotype. In tissue culture cells this state is associated with a number of spontaneous phenotypic changes which include "immortality", ability to grow in the presence of low levels of calf serum, ability to grow in soft agar (anchorage independence) and the attainment of a high saturation density in monolayer culture. Different NIH-3T3-derived clonal cell-lines showed a positive correlation between their ability to grow in soft agar and cytoplasmic VL30 RNA expression (Howk et al., 1978). In another study, cell lines from several strains of M. musculus were tested for VL30 RNA expression. It was of interest that the only completely negative cell line for VL30 RNA was C3H-10T½, which is a cloned line that displays contact inhibition during culture, and is non-tumourigenic when injected into animals (Reznikoff et al., 1973; Sherwin et al., 1978).

VL30 gene expression is not enhanced, however, upon Ki-MuSV-induced transformation of BALB-3T3 cells (Howk et al., 1978). These results suggested to members of M.J. Getz's research group a possible role for VL30 expression in spontaneous transformation events (Courtney et al., 1982a). Mouse embryo cells were transformed by chemicals or by growth in mitogens and levels of VL30 RNA expression were examined. The polycyclic aromatic hydrocarbon, 5-methylcholanthrene (5-MCA) was found to enhance VL30 expression by approximately 10-fold in two different mouse strains. Interestingly, whereas AKR cells induced by 5-MCA produced higher levels of both VL30 and endogenous MuLV RNA, similar treatment of C3H10T½ cells induced only VL30 expression (Courtney et al., 1982a).

It was also shown that a differential response by specific classes of DNA sequences could be elicited by exposure of the cell to epidermal

growth factor (EGF). The binding of this peptide growth factor to specific cell membrane receptors initiates a complex series of events which stimulate cells in the quiescent G_0 state to renewed proliferation (Hershko et al., 1971). These cells display several growth characteristics of transformed cells, but in a completely reversible (ie EGF-dependent) fashion. Mitogenic stimulation by EGF was shown to enhance levels of VL30 RNA expression in AKR mouse cells, but not that of other poly(A)$^+$ RNA species such as α and β globin mRNA (Courtney et al., 1982a). Foster et al. (1982) showed that EGF treatment of these cells induced VL30 expression but not AKR-type MuLV expression to a great extent; furthermore three from twelve cDNA clones prepared from EGF-induced RNA were found to contain VL30 sequences. These findings led Foster et al. (1982) to suggest that enhanced VL30 RNA expression in transformed cells may be related to the secretion of and possible autostimulation by growth factor-like compounds which are known to be produced by certain transformed cells and tumours (Roberts et al., 1980; Todaro et al., 1980; Moses et al., 1981). This is not to suggest, however, that VL30 gene expression may have any more than a coincidental role in the events which lead up to the spontaneous transformation of tissue culture cells.

An apparent difference between the regulation of VL30 gene expression and that of MuLV, as highlighted in this example by EGF treatment, has also been noted in another experimental system. Howk et al. (1978) showed that cycloheximide or IdU treatment of Ki-MuSV-transformed BALB-3T3 cells resulted in the elevated expression of xenotropic MuLV RNA. Assay of VL30 RNA in these cells detected a dramatic increase in the IdU-treated cells, but only a two-fold increase in VL30 RNA levels of the cycloheximide treated cells.

Induction of endogenous retroviruses and retrovirus-like proviruses by nucleoside analogues and inhibitors of protein synthesis is

consistent with the existence of a short-lived repressor. In the case of the former treatment, one can imagine such a repressor would fail to bind to DNA which contained the modified bases. However the results of Lowy (1978) imply an indirect, cell-mediated effect rather than direct repression, since pretreatment of recipient mouse cells with IdU enhanced the infectivity of DNA from non-producer AKR cells (ie of endogenous proviruses in this DNA), whereas IdU-substituted AKR DNA did not have enhanced infectivity. It is more likely that the effect of these inducing agents is similar to that of 5-azacytidine. This analogue of cytidine becomes converted to the deoxy form in the cell, and is incorporated instead of cytidine. Since this base cannot be methylated during DNA synthesis (as is postulated to be the case normally) 5-azacytidine treatment of cells induces extensive and permanent demethylation of DNA, resulting in permanent expression of many endogenous proviruses (eg see Groudine et al., 1981).

Studies of artificially introduced MuLVs have added significantly to our knowledge of the mechanisms governing their expression. Jaenisch and his colleagues have attempted to "create" endogenous proviruses by infecting mice at early stages of their development with the ecotropic Mo-MuLV (Jaenisch, 1976). These experiments showed that several conditions had to be fulfilled before expression of these proviruses could occur.

Firstly, the stage of development at which infection took place was critical; mice infected after the 16 cell stage would develop into adults which were rarely capable of vertically transmitting Mo-MuLV to their offspring. In contrast, adult mice which had been infected at the 4-16 cell stage contained a copy of the provirus in the cells of every tissue, including that of the germ line (Jaenisch, 1979). By inbreeding, strains of mice were established with one or more proviruses integrated at a different position in the genome. These exhibited an

interesting phenotypic variation in the appearance of infectious virus. The majority of mice carrying Mo-MuLV in their germ line did not develop viraemia, indicating either that integration of a defective provirus had occurred, as has been shown for the Mov-4 and Mov-6 strains, or, as is apparently the case for six other Mov strains, that integration of a normal provirus had occurred but at a site within the chromatin which remained inactive throughout life (Jaenisch et al., 1981). The remaining strains differed in the time at which provirus activation occurred. Mov-1 mice (ie with a provirus at the locus designated mov-1) became viraemic only after birth, thus virus spread was restricted to the natural target tissues of Mo-MuLV (thymus and spleen) (Jaenisch, 1980). Mov-13 mice have been shown to contain a provirus inserted into the first intron of the 1(I) collagen gene, which is strongly expressed from days 11 and 12 of gestation (Harbers et al., 1984; Schnieke et al., 1983). Embryos homozygous for viral insertion are arrested in development at this time (Jaenisch et al., 1983) but heterozygous mice become viraemic as early as day 16 of gestation (Jaenisch et al., 1981). In contrast to the cells of an adult mouse, those of the midgestation embryo are susceptible to virus integration and expression in all organs (Jaenisch, 1980). Thus in heterozygous Mov-13 mice most if not all organs of the adult contain extra copies of Mo-MuLV, acquired by superinfection during this early stage of embryogenesis (Jaenisch et al., 1981).

Mice possessing activated proviruses only in target tissues were assessed for methylation status and DNAase I sensitivity of these proviruses. In each case the endogenous provirus (acquired by germ line infection) remained hypermethylated and transcriptionally inactive, whereas exogenously introduced proviruses (ie acquired after birth) were undermethylated and expressed (Harbers et al., 1981; Stuhlman et al., 1981). Thus in vivo expression of virus in target organs is not due to

activation of a previously silent endogenous virus, but the tissue-specific replication of a provirus acquired by infection.

It should be stressed that only ecotropic or dualtropic (MCF) viruses could exhibit these phenomena; the majority of endogenous MuLVs are of the xenotropic variety and consequently can only infect heterologous hosts. IAP genes face a similar problem since they never leave the cell. It is possible to imagine, however, that a certain amount of IAP superinfection may occur during embryogenesis simply by transfer of infected cytoplasm from parent to daughter cells.

Jaenisch's work may have provided some answers to questions which arise from this study of VL30 organization and expression. It is possible that the minority subsets of transcriptionally active proviruses identified in NIH-3T3 and other mouse cell lines are also the result of superinfection by VL30 elements acquired during early development. These exogenously acquired VL30 proviruses may have been transmitted from cell to cell within the respective mouse embryos by endogenous MuLVs activated during these early stages. Differences in VL30 expression observed in different mouse cell lines may indeed correlate with the transformed phenotype, where embryological differentiation may be mimicked. Alternatively these could be the result of random integration of superinfecting VL30 elements into areas of the mouse genome enjoying different degrees of transcriptional activity.

Also by analogy with retroviruses, VL30 expression may be tissue specific, so that cell lines derived from one part of the mouse may contain VL30 RNA and others may not. Evidence has accumulated which suggests that the VL30 LTR regions are the most likely location for these tissue specific control sequences.

Returning once more to analogy with retroviruses, in addition to other sequences concerned with transcriptional regulation the LTRs

contain enhancer elements. These are defined as cis-acting control elements which increase the level of expression of an adjacent gene in a fashion which is relatively independent of position and orientation [see Khoury & Gruss (1983) for a review]. The "prototype" enhancer element is contained within a 72bp tandem repeat of SV40, and is responsible for early viral gene activity. Without this repeat, such activity is 10^2- to 10^3-fold less (Benoist & Chambon, 1981). Although the tandem repeats often found in retroviral LTRs have only minor sequence homology with those found in SV40, Levinson et al. (1982) have shown that the 72bp tandem repeat present in the Mo-MuSV LTR can functionally replace the SV40 tandem repeat. Considerable evidence has accumulated for the tissue-specific role of such enhancer elements in the determination of the complex phenotypes of MuLV-induced diseases in mice. Friend MuLV and Mo-MuLV have related but not identical LTR sequences, each containing slightly different enhancer-like elements. After injection of these viruses into new-born NFS mice, Fr-MuLV induces erythroleukaemias but Mo-MuLV induces T-cell lymphomas. Chatis et al. (1983) found that when a 0.62kbp fragment encompassing the U_3 LTR region (containing the enhancer element) is replaced with a corresponding fragment of Mo-MuLV, the recombinant DNA now induced T-cell lymphomas. A reciprocal experiment showed that a 380bp fragment (again containing the enhancer) from the Mo-MuLV U_3 region converted Fr-MuLV to a lymphoma-inducing virus (Chatis et al., 1984). Although they enhance transcriptional activity, these elements may not actually induce leukaemogenesis by MuLV. Postulated mechanisms for induction of leukaemia include the activation of cellular onc genes by promoter insertion, and the mitogenic action of virus-encoded envelope glycoproteins (see General Introduction and section D-7).

Enhancer-like elements have been identified in the nucleotide sequences of both the BVL-1 and NVL-3 U_3 regions (Hodgson et al., 1983;

Norton et al., 1984b). By analogy with MuLV proviruses, therefore, these elements may be responsible for the specific induction of VL30 expression in the examples described above. It remains to be seen whether the differential expression of IAP subsets in different mouse tumours is also connected with tissue-specific control elements in their LTRs. Several of the latter have been sequenced (Kuff et al., 1983; Ono & Ohishi, 1983; Ono et al., 1984) and although none contain an exact copy of the consensus "core" sequence of an enhancer element [(G)TGG$^{AAA}_{TTT}$ (Weiher et al., 1983)] at least two (Canaani et al., 1983; Leuders et al., 1984) contain a similar sequence (GTGGTAAA). These two mouse-derived IAP LTRs also contain stretches of alternating purines and pyrimidines which could form the Z-DNA conformation suggested to be a critical component of enhancer activity (Nordheim & Rich, 1983).

The ability of VL30 and MuLV RNA species to be efficiently co-packaged into MuLV particles is probably not a function of an LTR-located sequence. Study of an RSV-transformed quail cell line that is deficient in packaging viral RNA (Linial et al., 1978) has shown that the mutant virus has a 150bp deletion at the 5' end of its genome, upstream of the start of Pr76^{gag} (Shank & Linial, 1980). Furthermore a Mo-MuLV packaging mutant has been constructed in vitro by deletion of a 350bp sequence from an untranscribed region between the 5' donor splice site and the initiation codon for Pr65^{gag} (Mann et al., 1983). Cell lines containing stably integrated proviruses of this mutant produce apparently normal virions containing reverse transcriptase, but lacking genomic RNA. Of interest was that a minority of cells transfected with the mutated Mo-MuLV DNA produced non-defective virus after a lag of 4 to 6 days. The authors proposed that this was the result of recombination between a cellular sequence and the defective provirus. Likely candidates for providing these packaging site sequences were endogenous retrovirus-like elements such as VL30 (Mann et al., 1983). Certainly

MuLV and VL30 elements appear to possess sequence homology in this 5' unique sequence region (Giri et al., 1983; Norton et al., 1984b). This may explain why VL30 RNA is selectively packaged by C-type retroviruses but not B-types, which do not possess sequence relatedness to MuLV (Howk et al., 1978; Sherwin et al., 1978).

The experiments described in section R-6 showed that mouse VL30 RNA, packaged by Ki-MuLV can be successfully transmitted to a heterologous cell line. Detection of NVL-like proviruses in the DNA of these cells indicated that the retrovirus-transmitted RNA had been used as a template for the synthesis of provirus-like molecules that had become stably integrated into genomic DNA. Other workers have also attempted to demonstrate the transmissibility of mouse VL30 RNA. Sherwin et al. (1978) rescued VL30 RNA from feral mouse SC-1 cells using the baboon M7 retrovirus. However upon infection of human cells with this pseudotype, only the M7 viral genome was successfully transmitted. More success was enjoyed by Scolnick et al. (1979) using mouse VL30 RNA packaged by Mo-MuLV. Both genomes were transmitted to normal (FRE) rat cells or rat tumour cells. VL30 RNA was detected in both cytoplasmic RNA or RNA isolated from released viral particles. Nevertheless, VL30 was only poorly infectious for these cells, as virions released from the tumour cells (the most successful example) contained MuLV:VL30 RNA at a ratio of 100:1 (Scolnick et al., 1979).

Until the work described here, no retrovirus-transmitted mouse VL30 provirus has been shown by Southern transfer analysis to be integrated in heterologous cell DNA. However that of a rat VL30 element, transmitted to bat lung cells by woolly monkey virus [WO-LV] (Scolnick et al., 1979) has been successfully identified in a restriction enzyme digest of bat chromosomal DNA (Young et al., 1980) thus strengthening the analogy between rat and mouse VL30 elements.

As discussed in section D-5, these experimental results imply that

VL30 evolution may not only have involved genetic recombination with the progenitors of MuLV (Giri et al., 1983) but may also have been accelerated by intra- or inter-specific retrovirus-mediated infection events. In this way, new variants of VL30 species which have arisen by spontaneous mutation or as a result of gene turnover mechanisms may have taken the opportunity to spread, not only within the genomes of their host species, but also in the genomes of heterologous species. However it is questionable whether VL30 elements have used the retrovirus-mediated route during their evolution. Inter-specific retrovirus infections may be rare in the natural environment, and the vast majority of integrated proviruses which result will be lost with the death of their new host, having failed to enter the germline. Moreover, it has been noted that the patterns of IAP gene organization in rodent genomes closely parallel those of VL30 elements. The former particles, nevertheless, occur as intracellular forms only, and are non-infectious, even when isolated in large amounts from the cells containing them (Kuff et al., 1968, 1972; Minna et al., 1974).

Both the origin of VL30 genetic elements, and their biological significance remain a mystery. Some of the possible advantages to the cell in containing these endogenous retrovirus-like sequences will be discussed in the final section.

D-7 Concluding remarks, speculations and future prospects

Eukaryotic genomes contain many different families of middle repetitive DNA sequences whose structure suggests that they are, or have once been, mobile genetic elements (Cameron *et al.*, 1979; Jagadeeswaran *et al.*, 1981; Jelinek & Schmid, 1982). The discovery that transposable elements of this kind can promote genetic rearrangements in chromosomal DNA led to the view that eukaryotic genomes are more fluid than was once supposed (Calos & Miller, 1980; Roeder & Fink, 1980). During the last few years a search has been in progress to find an evolutionary link between transposable elements, able to rearrange sequences within genomes, and retroviruses, which can rearrange sequences between genomes (Temin, 1980). Obvious candidates for intermediates in the pathway from transposable element to retrovirus (or vice versa) are the retrovirus-like IAP and VL30 elements present in rodent genomic DNA and the copia-like elements of *Drosophila melanogaster* cells (Temin, 1980; Flavell, 1981, 1984; Finnegan, 1983) (see General Introduction). The work presented in this thesis has described the cloning and characterization of several mouse VL30 cDNA species, and their use as probes to study the genetic organization of the mouse VL30 family, in addition to that of related sequences in other rodent genomes.

Results of these analyses showed that many aspects of VL30 gene organization and expression were similar to those of endogenous retroviruses. Most of the mouse VL30 "proviruses" were hypermethylated and transcriptionally inactive, a characteristic of endogenous proviruses derived from germ line DNA. The genes which probably encoded the NVL RNA species, moreover, were found to represent a small minority of the total number of NIH-3T3 mouse VL30 elements. Taken together with the fact that NVL-like genes were apparently undermethylated, these data suggested that they were the result of relatively rare exogenous infections, which had occurred either at an early stage in the

establishment of the NIH-3T3 cell line, or during NIH mouse embryogenesis. The newly acquired NVL-like elements in KLNRK-1 cell DNA demonstrated how this might occur (sections R-6, D-6). If the latter hypothesis were correct, then successive generations of mice might be expected to contain a gradually increasing number of VL30 proviruses. The capacity for VL30 germ-line amplification may be large, since certain mouse strains tolerate up to 1000 IAP genes in their chromosomal DNA, in addition to numerous other proviruses. However VL30 amplification may be dependent upon the presence of infectious C-type viruses which are capable of re-infecting the tissues of their own host. In nature this seems to be strictly controlled by the phenomenon of xenotropism.

It is possible, nevertheless, that VL30 elements may be capable of duplicative transposition within their host cell genome in the absence of infectious retroviruses. Evidence for the action of cellular reverse transcriptase is frequently discovered in mammalian genomes, for instance the intronless pseudogenes, or "retrogenes" as they have been called (eg Hollis et al., 1982; Gwo-Shu Lee et al., 1983; Karin & Richards, 1982; for review see Sharp, 1983). What is the source of this enzyme activity? Free and particle-bound (possibly membrane-bound) reverse transcriptase has been described in normal Japanese quail embryos, both in the nucleus and cytoplasm (Mondal & Hofschneider, 1983). Although free reverse transcriptase has not been discovered in human cells, retrovirus-like particles have been observed in normal human embryonic tissue (Chandra et al., 1970) and in placenta (Dirksen & Levy, 1977). These particles were not infectious, but contained reverse transcriptase activity (Nelson et al., 1978). In fact Nelson et al. (1981) have also described a specific inhibitor for the human placental reverse transcriptase, and from this propose that the enzyme plays a role in placental development. Correlations have also been made between

reverse transcriptase activity and human embryonic differentiation (Mondal & Hofschneider, 1982).

It is known that the reverse transcriptase of retroviruses is not enzymically active until it has been proteolytically cleaved from an inactive precursor during (or after) packaging into virions (Dickson et al., 1982). If free reverse transcriptase does indeed play a role in embryogenesis, one could speculate that it may be encoded by endogenous retroviruses. Perhaps the enzyme must be activated in cytoplasmic retrovirus-like particles before being transported to its site of action in the nucleus. One could imagine that "budding-out" of these particles would deplete the cell of its source of activated reverse transcriptase; the mouse IAPs and the human non-infectious retrovirus-like particles may therefore be the result of specific selection for this characteristic.

The role of reverse transcriptase during embryonic development remains a mystery. The re-insertion of processed pseudogenes may be a chance by-product of a more useful cellular process. "Retrogenes" of mRNA, snRNA or rRNA sequences would possess no transcriptional activity, since these are transcribed by RNA polymerases I and II. They will therefore have lost the sequences which specify the initiation of transcription, which lie upstream of the mature RNA coding region (Grummt, 1981; Shenk, 1981). Only RNA polymerase III-transcribed units could be templates for a DNA form retaining its transcriptional competence. Such transcripts include the Alu-like transposable elements (Jagadeeswaran et al., 1981; Brown, 1984) which appear to play an important role in protein signal recognition by encoding 7SL RNA (Walter & Blobel, 1982; Ullu et al., 1982; Ullu & Tschudi, 1984).

If one goes so far as to speculate that it is indeed the synthesis of reverse transcribed copies of certain cellular RNA species that plays some role in embryonic development, then one must also suppose that

Drosophila flies have adopted a different mechanism. The cells of these invertebrates contain an abundant source of reverse transcriptase, used to generate extrachromosomal copies of *copia* DNA (Flavell, 1984) but their genomic DNA bears no evidence of processed pseudogenes (Saluz *et al.*, 1983). Whatever the developmental role of a cellular reverse transcriptase, it could explain the successful spread of apparently defective retrovirus-like genomes, such as VL30, in mouse germ line DNA.

Are VL30 elements examples of "selfish" DNA, or do they also serve a selectable biological function? As originally discussed by Doolittle & Sapienza (1980) and Orgel & Crick (1980), any element which produces a faithful copy of itself elsewhere in the genome, while preserving the original copy at its original site need not justify its existence by benefiting its host. Replication by this method alone ensures the survival of a piece of DNA, since it would be impossible to eliminate all the replicas by simultaneous deletion. The observation that endogenous retroviruses and virus-like elements vary in structure and location between individuals of the same species, and vary widely in number between closely related species is indeed consistent with the view that they are benign parasites, existing solely as the result of a tenuous balance between random deletion and occasional germ line integration.

Is it possible that this apparently random duplication of endogenous retrovirus-like sequences confers a long term evolutionary advantage to their hosts? As emphasized by Ohno (1970) and Jeffreys & Harris (1982) the genetic repertoire of present day eukaryotic organisms may have been gradually increased by the process of gene duplication, since this appears to be the method of choice by which mutation and gene conversion can create new protein products by their action upon coding sequences which are no longer vital to the cell. Over long periods of evolutionary time, different organisms may have entrained their own

particular classes of endogenous proviruses to useful functions. These would not necessarily be universal attributes of all provirus-like elements.

Most of the ways in which a provirus could be useful to an organism have been postulated for endogenous C-type gene products, probably because these have been the most extensively studied due to their carcinogenic activities. Structural genes of endogenous retroviruses are frequently expressed under natural conditions (Weiss & Biggs, 1972). The expression of retrovirus _env_ genes in the mouse has been particularly well studied, and a few of the important discoveries will be briefly discussed here.

The _env_ gene products in question are probably the result of xenotropic (X)-MuLV expression. These sequences are present in virtually all mouse genomes, and whereas many strains of mouse are virus particle-negative, most still express X-MuLV _env_ proteins. These are located on the surface of cell membranes and their distribution appears to be tissue specific. For example high levels of gp70 may be detected in mouse seminal vesicles, thus implicating this protein in the potentiation of sperm in the seminal fluid (Lerner _et al._, 1976).

Moroni & Schumann (1977) argued that expression of endogenous MuLV genes may be physiologically required for B cells to participate in the immune response. They observed that B lymphocytes can be induced by mitogens to release virus, whereas T cells are refractory to induction by either mitogens or BrdU. Moreover gp70 antiserum was immunosuppressive _in vivo_ and _in vitro_ (Moroni & Schumann, 1977; Theofilopoulos _et al._, 1981).

Six X-MuLV loci have recently been mapped to as many different chromosomes by using recombinant inbred mouse strains (Wejman _et al._, 1984). It was of interest to find that three of these loci were closely linked to genes encoding mouse lymphocyte cell surface antigens. The

authors speculated that this close association indicated that the proviruses may have some function in normal lymphocyte development, or may even encode these particular antigens (Wejman et al., 1984).

The attempts to implicate env-encoded gene products in leukaemogenesis have also generated theories which suggest they may indirectly affect rodent evolution. McGrath & Weissmann (1979) maintained that a specific subset of thymus-derived T lymphocytes can bind MCF (dualtropic) virus. MCF viruses arise by recombination of ecotropic virus sequences with xenotropic sequences in the region of the env gene (Elder et al., 1977). Stimulation of T lymphocytes by the specific binding of this recombinant p15(E) product may be one of the stages leading towards induction of thymic leukaemia (McGrath & Weissmann, 1979). Ihle & Lee (1982), on the other hand, do not accept that any T cell receptor binds to a soluble antigen, but always reacts by an "associated recognition" of the foreign antigen (they suggest gp70) with a "self" MHC component. These workers admitted, however, that it may be possible for endogenous retroviruses to contain determinants which actually mimic this combination of antigens.

Coutinho (1982) took this observation a stage further by suggesting that this ability of retroviruses could indirectly regulate the evolution of MHC antigen polymorphism in a species. This would be achieved by allowing selection for a germ line-encoded T cell receptor repertoire whose products could recognize a new range of MHC variants. In the absence of a system which mimics part of the MHC function, an MHC mutant would not be viable since it could not develop a T cell immune system (Coutinho, 1982).

VL30 genes have probably been members of the rodent genome for as long as have these endogenous MuLV sequences. In order to assess the potential of any VL30-encoded contributions, therefore, it is imperative to gain a deeper understanding of this gene family. As achieved for

several C-type retroviruses, the complete nucleotide sequence of several genomic and cDNA VL30 examples should be determined in order to define open reading frames for protein coding regions. Sequence analysis of the NVL-3 terminal regions has revealed a 5'-located region containing a sequence which is very similar to the donor splice site of a retrovirus (Norton et al., 1984b). Further sequencing might also establish the existence of acceptor splice sites, which in turn should aid the interpretation of the VL30 RNA gel profiles described in this work.

The VL30 proteins predicted by sequence analysis might be most easily detected in KLNRK-1 cells, which contain 1-2 copies per cell of transcriptionally active NVL-like VL30 units. Mouse VL30-specific protein could be purified from these cells, or synthesized by the engineered expression of NVL clones in vitro, and used to raise antibodies. Anti-NVL protein serum could then be used to examine mouse tumour tissues for the abnormal expression of VL30 protein.

It is quite possible that no VL30 encoded protein will be found. However these elements are still theoretically capable of inducing leukaemias by the mechanism of promoter insertion (see General Introduction). A non-producer bat cell line has been established which contains exogenously introduced rat VL30 sequences, but no helper MuLV (Scolnick et al., 1979; Young et al., 1980). It is conceivable, therefore, that exogenously introduced VL30 elements can occasionally induce virus-negative tumours. Some mouse tumour cells contain the amplified genomes of retroviruses in their cellular DNA, such as AKV (Berns & Jaenisch, 1976; Steffen et al., 1979) and IAPs (Shen-Ong & Cole, 1982). The latter were probably the result of duplication by an RNA intermediate. However several other examples have been described in which tumour cells contain amplified karyotypic abnormalities that include abundantly expressed cellular oncogenes (Collins & Groudine, 1982; Dalla-Favera et al., 1982; Alitalo et al., 1983; Schwab et al.,

1983). It may also be of interest to search for VL30 elements in regions of amplified DNA from mouse tumours.

Hopefully, the preliminary studies of VL30 gene organization, expression and transmission described in this thesis will provide a useful basis for further investigation into the biological significance of this retrovirus-like family.

REFERENCES

REFERENCES

AARONSON, S.A. & ROWE, W.P. (1970) Virology 42, 9-19.

AARONSON, S.A. & WEAVER, C. (1971) J. Gen. Virol. 13, 245-252.

AARONSON, S.A., HARTLEY, J.W. & TODARO, G.J. (1969) Proc. Natl. Acad. Sci. U.S.A. 64, 87-94.

AARONSON, S.A., JAINCHILL, J.L. & TODARO, G.J. (1970) Proc. Natl. Acad. Sci. U.S.A., 66, 1236-1243.

ALITALO, K., SCHWAB, M., LIN, C.C., VARMUS, H.E. & BISHOP, J.M. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1707-1711.

ALTON, N.K. & VAPNEK, D. (1979) Nature (London) 282, 864-869.

ANDERSON, G.R. & ROBBINS, K.C. (1976) J. Virol. 17, 335-351.

ANET, R. & STRAYER, D.R. (1969) Biochem. Biophys. Res. Commun. 34, 328-334.

ASTRIN, S.M., BUSS, E.G. & HAYWARD, W.S. (1979) Nature (London) 281, 339-341.

AVERY, R.J. & LEVY, J.A. (1978) J. Gen. Virol. 39, 427-435.

AXELRAD, A.A. & STEEVES, R.A. (1964) Virology 24, 513-518.

BACHELER, L.T. & FAN, H. (1979) J. Virol. 30, 657-667.

BALL, J.K., DEKABAN, G.A., LOOSMORE, S.M., CHAN, S.K. & McCARTER, J.A. (1979) Nuc. Acids Res. 7, 1091-1108.

BALTIMORE, D. (1970) Nature (London) 226, 1209-1211.

BALTIMORE, D. (1975) Cold Spring Harbor Symp. Quant. Biol. 39, 1187-1200.

BALUDA, M. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 576-580.

BAKER, B., ROBINSON, H., VARMUS, H.E. & BISHOP. J.M. (1981) Virology 114, 8-22.

BASSIN, R.H., TUTTLE, N. & FISCHINGER, P.J. (1970) Int. J. Cancer 6, 95-107.

BASSIN, R.H., TUTTLE, N. & FISCHINGER, P.J. (1971a) Nature (London) 229, 564-566.

BASSIN, R.H., PHILLIPS, L.A., KRAMER, M.J., HAAPALA, D.K., PEEBLES, P.T., NOMURA, S. & FISCHINGER, P.J. (1971b) Proc. Natl. Acad. Sci. U.S.A. 68, 1520-1524.

BATTULA, N. & LOEB, L.A. (1974) J. Biol. Chem. 249, 4086-4093.

BEEMON, K.L. (1978) Curr. Top. Microbiol. Immunol. 79, 73-110.

BEEMON, K.L. & KEITH, J.M. (1976) In: Animal Virology (ed. D. Baltimore et al.) pp. 97-105. Academic Press, New York.

BEIJERINCK, M.W. (1899) Zentr. Bakteriol. Parasitenkd. Infektionskr. 5, 27-35.

BENDER, W. & DAVIDSON, N. (1976) Cell 7, 595-607.

BENOIST, C. & CHAMBON, P. (1981) Nature (London) 290, 304-310.

BENOIST, C., O'HARE, K., BREATHNACH, R. & CHAMBON, P. (1980) Nuc. Acids Res. 8, 127-142.

BENTVELZEN, P. (1968) PhD Thesis, Hollandia, Amsterdam.

BENTVELZEN, P., DAAMS, J.H., HAGEMAN, P. & CALAFAT, J. (1970) Proc. Natl. Acad. Sci. U.S.A. 67, 377-384.

BENVENISTE, R.E. & TODARO, G.J. (1973) Proc. Natl. Acad. Sci. U.S.A. 70, 3316-3320.

BENVENISTE, R.E. & TODARO, G.J. (1976) Nature (London) 261, 101-108.

BENVENISTE, R.E., CALLAHAN, R., SHERR, C.J., CHAPMAN, V. & TODARO, G.J. (1977) J. Virol. 21, 849-862.

BENVENISTE, R.E., HEINEMANN, R., WILSON, G.L., CALLAHAN, R. & TODARO, G.J. (1974) J. Virol. 14, 56-67.

BENZ, E.W. & DINA, D. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3294-3298.

BENZ, E.W., WYDRO, R.W., NADAL-GINARD, B. & DINA, D. (1980) Nature (London) 288, 665-669.

BERNHARD, W. (1958) Cancer Res. 18, 491-509.

BERNHARD, W., BONAR, R.A., BEARD, D. & BEARD, J.W. (1958) Proc. Soc. Exp. Biol. Med. 97, 48-52.

BERNS, A. & JAENISCH, R. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 2448-2452.

BESMER, P., OLSHEVSKY, U., BALTIMORE, D., DOLBERG, D. & FAN, H. (1979) J. Virol. 29, 1168-1176.

BESMER, P., SNYDER, H.W., MURPHY, J.E., HARDY, W.D. & PARODI, A. (1983) J. Virol. 46, 606-613.

BICZYSKO, W., PIENKOWSKI, M., SOLTER, D. & KOPROWSKY, H. (1973) J. Natl. Cancer Inst. 51, 1041-1050.

BINGHAM, P.M. & JUDD, B.H. (1981) Cell 25, 705-711.

BIRD, A.P. & SOUTHERN, E.M. (1978) J. Mol. Biol. 118, 27-47.

BIRNBOIM, H.C. & DOLY, J. (1979) Nuc. Acids Res. 7, 1513-1523.

BISHOP, J.M. (1978) Annu. Rev. Biochem. 47, 35-88.

BISHOP, J.M. (1983a) Cell 32, 1018-1020.

BISHOP, J.M. (1983b) Annu. Rev. Biochem. 52, 301-354.

BISHOP, J.M. & VARMUS, H. (1982) In: RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds.) pp. 999-1108. Cold Spring Harbor Laboratory, New York.

BITTNER, J.J. (1936) Science 84, 162.

BITTNER, J.J. (1942) Science 95, 462-463.

BLAIR, D.G., MASON, W.S., HUNTER, E. & VOGT, P.K. (1976) Virology 75, 48-59.

BOLOGNESI, D.P., MONTELARO, R.C., FRANK, H. & SCHAFER, W. (1978) Science 199, 183-186.

BOONE, L.R. & SKALKA, A. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 847-851.

BOYER, H.W. & ROULLAND-DUSSOIX, D. (1969) J. Mol. Biol. 41, 459-472.

BRAHIC, M., STOWRING, L., VENTURA, P. & HAASE, A.T. (1981) Nature (London) 292, 240-242.

BREATHNACH, R. & CHAMBON, P. (1981) Annu. Rev. Biochem. 50, 349-383.

BREINDL, M., HARBERS, K. & JAENISCH, R. (1984) Cell 38, 9-16.

BRITTEN, R.J. & KOHNE, D.E. (1968) Science 161, 529-540.

BROWN, A. & HUANG, R.C.C. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6123-6127.

BROWN, A.L. (1984) Nature (London) 312, 106.

BROWN, S.D.M. (1983) Gene 23, 95-97.

BROWN, S.D.M. & DOVER, G.A. (1981) J. Mol. Biol. 150, 441-466.

BUETTI, E. & DIGGELMANN, H. (1980) Virology 102, 251-261.

BUKHARI, A.I. (1981) TIBS 6, 56-60.

BURT, D.W., REITH, A.D. & BRAMMAR, W.J. (1984) Nuc. Acids Res. 12, 8579-8594.

CAIRNS, J. (1981) Nature (London) 289, 353-357.

CAIRNS, J. & LOGAN, J. (1983) Nature (London) 304, 582-583.

CALAFAT, J. & HAGEMAN, P. (1969) Virology 38, 364-368.

CALAFAT, J. & HAGEMAN, P.C. (1973) Nature (London) 242, 260-262.

CALARCO, P.G. & SZOLLOSI, D. (1973) Nature (London) New Biol. 243, 91-93.

CALLAHAN, R., KUFF, E.L., LUEDERS, K.K. & BIRKENMEIER, E. (1981) J. Virol. 40, 901-911.

CALOS, M.P. & MILLER, J.H. (1980) Cell 20, 579-595.

CAMERON, J.R., LOH, E.Y. & DAVIS, R.W. (1979) Cell 16, 739-751.

CANAANI, E. & DUESBERG, P. (1972) J. Virol. 10, 23-31.

CANAANI, E., HELM, K.V.D. & DUESBERG, P. (1973) Proc. Natl. Acad. Sci. U.S.A. 70, 401-405.

CANAANI, E., DREAZEN, O., KLAR, A., RECHAVI, G., RAM, D., COHEN, J.B. & GIVOL, D. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 7118-7122.

CARDIFF, R.D., PUENTES, M.J., TERAMOTO, Y.A. & LUND, J.K. (1974) J. Virol. 14, 1293-1303.

CARTER, A.T., NORTON, J.D. & AVERY, R.J. (1983) Nuc. Acids Res. 11, 6243-6254.

CELANDER, D. & HASELTINE, W.A. (1984) Nature (London) 312, 159-162.

CHALINE, J. (1977) Endeavour 1, 44-57.

CHAN, H.W., BRYAN, T., MOORE, J.L., STAAL, S.P., ROWE, W.P. & MARTIN, M.A. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5779-5783.

CHANDRA, S., LISZCZAK, T., KOROL, W. & JENSEN, E.M. (1970) Int. J. Cancer 6, 40-45.

CHASE, D.G. & PIKO, L. (1973) J. Natl. Cancer Inst. 51, 1971-1973.

CHATIS, P.A., HOLLAND, C.A., HARTLEY, J.W., ROWE, W.P. & HOPKINS, N. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 4408-4411.

CHATIS, P.A., HOLLAND, C.A., SILVER, J.E., FREDERICKSON, T.N., HOPKINS, N. & HARTLEY, J.W. (1984) J. Virol. 52, 248-254.

CHIEN, Y.-H., JUNGMANS, R.P. & DAVIDSON, N. (1980) In: Molecular biology of RNA Tumor Viruses (J.R. Stephenson ed.) pp 395-446. Academic Press, New York.

CHIEN, Y.-H., LAI, M., SHIH, T.Y., VERMA, I.M., SCOLNICK, E.M., ROY-BURMAN, P. & DAVIDSON, N. (1979) J. Virol. 31, 752-760.

CHISWELL, D.J. & PRINGLE, C.R. (1977) J. Gen. Virol. 36, 551-555.

CHISWELL, D.J. & PRINGLE, C.R. (1979a) J. Gen. Virol. 43, 429-434.

CHISWELL, D.J. & PRINGLE, C.R. (1979b) J. Gen. Virol. 44, 145-152.

CHOPRA, H.C. & MASON, M.M. (1970) Cancer Res. 30, 2081-2086.

CLARK, D.P. & DOUGHERTY, R.M. (1980) J. Gen. Virol. 47, 283-291.

CLAYMAN, C.H. & MOSHARRAFA, E. & FARAS, A.J. (1979) J. Virol. 29, 242-249.

CLEMENTS, J.E., NARAYAN, O., GRIFFIN, D.E. & JOHNSON, R.T. (1979) Virology 93, 377-386.

CLEWLEY, J.P. & AVERY, R.J. (1982) Arch. Virol. 72, 35-46.

COFFIN, J. (1982) In: RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H. & Coffin, J. eds.), 2nd edn, pp 1109-1203. Cold Spring Harbor Laboratory, New York.

COFFIN, J.M. (1979) J. Gen. Virol. 42, 1-26.

COFFIN, J.M. & HASELTINE, W.A. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 1908-1912.

COFFIN, J.M., HAGEMAN, T.C., MAXAM, A.M. & HASELTINE, W.A. (1978) Cell 13, 761-773.

COHEN, J.C. (1980) Cell 19, 653-662.

COHEN, M., REIN, A., STEPHENS, R., O'CONNELL, C., GILDEN, R., SHURE, M., NICOLSON, M., McALLISTER, R. & DAVIDSON, N. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 5207-5211.

COLLINS, S. & GROUDINE, M. (1982) Nature (London) 298, 679-681.

COOPER, G.M. & LANE, M.-A. (1984) Biochim. Biophys. Acta 738, 9-20.

COOPER, G.M. & NEIMAN, P.E. (1981) Nature (London) 292, 857-858.

COPELAND, N.G., JENKINS, N.A. & COOPER, G.M. (1981) Cell 23, 51-60.

COPELAND, N.G., ZELENETZ, A.D. & COOPER, G.M. (1980) Cell 19, 863-870.

CORDELL, B., STAVNEZER, E., FRIEDRICH, R., BISHOP, J.M. & GOODMAN, H.M. (1976) J. Virol. 19, 548-558.

CORDEN, J., WASYLYK, B., BUCHWALDER, A., SASSONE-CORSI, P., KEDINGER, C. & CHAMBON, P. (1980) Science 209, 1406-1414.

COURTNEY, M.G., SCHMIDT, L.J. & GETZ, M.J. (1982a) Cancer Research 42, 569-576.

COURTNEY, M.G., ELDER, P.K., STEFFEN, D.L. & GETZ, M.J. (1982b) J. Virol. 43, 511-518.

COUTINHO, A. (1982) Curr. Top. Microbiol. Immunol. 98, 113-126.

CREUSOT, F., ACS, G. & CHRISTMAN, J.K. (1982) J. Biol. Chem. 257, 2041-2048.

CRITTENDEN, L.B. & MOTTA, J.V. (1975) Virology 67, 327-334.

CULLEN, B.R., SKALKA, A.M. & JU, G. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 2946-2950.

DALLA FAVERA, R., WONG-STAAL, F. & GALLO, R. (1982) Nature (London) 299, 61-63.

DARLINGTON, C.D. (1948) Br. J. Cancer 2, 118-126.

DAWSON, C.R., DRAKE, A.F., HELLIWELL, J. & HIDER, R.C. (1978) Biochim. Biophys. Acta 570, 75-86.

DE FEO, D., GONDA, M.A., YOUNG, H.A., CHANG, E.H., LOWY, D.R., SCOLNICK, E.M. & ELLIS, R.W. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 3328-3332.

DENHARDT, D.J. (1966) Biochem. Biophys. Res. Commun. 23, 641-646.

DHAR, R., McCLEMENTS, W.L., ENQUIST, L.W. & VANDE WOUDE, G.F. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3937-3941.

DIAMOND, A., COOPER, G.M., RITZ, J. & LANE, M.-A. (1983) Nature (London) 305, 112-116.

DICKSON, C., SMITH, R., BROOKES, S. & PETERS, G. (1984) Cell 37, 529-536.

DICKSON, C., EISENMAN, R., FAN, H., HUNTER, E. & TEICH, N. (1982) In: RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H. & Coffin, J. eds.), 2nd edn., pp 513-648. Cold Spring Harbor Laboratory, New York.

DINA, D. & BENZ, E.W. (1980) J. Virol. 33, 377-389.

DINOWITZ, M. (1975) Virology 66, 1-9.

DIPPOLD, W.G., JAY, G., DELEO, A.B., KHOURY, G. & OLD, L.J. (1981) Proc. Natl. Acad. Sci. 78, 1695-1699.

DIRKSEN, E.R. & LEVY, J.A. (1977) J. Natl. Cancer Inst. 59, 1187-1192.

DOLBERG, D. & FAN, H. (1981) J. Gen. Virol. 54, 281-291.

DOLBERG, D.S., BACHELOR, L.T. & FAN, H. (1981) J. Virol. 40, 96-106.

DOOLITTLE, W.F. & SAPIENZA, C. (1980) Nature (London) 284, 601-603.

DOUGHERTY, R.M., DISTEFANO, H.S. & ROTH, F.K. (1967) Proc. Natl. Acad. Sci. U.S.A. 58, 808-817.

DOVER, G.A. (1982) Nature (London) 299, 111-117.

DUC-NGUYEN, H., ROSENBLUM, E.N. & ZEIGEL, R.F. (1966) J. Bacteriol. 92, 1133-1140.

DUDLEY, J.P. & VARMUS, H.E. (1981) J. Virol. 39, 207-218.

DUESBERG, P.H. (1968) Proc. Natl. Acad. Sci. U.S.A 60, 1511-1518.

DUESBERG, P.H. (1970) Curr. Top. Microbiol. Immunol. 51, 79-104.

DUESBERG, P.H. (1983) Nature (London) 304, 219-226.

DONOGHUE, D.J. & HUNTER, T. (1983) J. Virol. 45, 607-617.

DUFF, R.G. & VOGT, P.K. (1969) Virology 39, 18-30.

DUNSMUIR, P., BROREIN, W., SIMON, M. & RUBIN, G.M. (1980) Cell 21, 575-579.

EDWARDS, S.A. & FAN, H. (1981) Virology 113, 95-108.

EISENMAN, R., VOGT, V.M. & DIGGELMANN, H. (1975) Cold Spring Harbor Symp. Quant. Biol. 39, 1067-1075.

ELDER, J.F., JENSEN, F., LERNER, R., HARTLEY, J.W. & ROWE, W.P. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 4676-4680.

ELIYAHU, D., RAZ, A., GRUSS, P., GIVOL, D. & OREN, M. (1984) Nature (London) 312, 646-649.

ELLERMAN, V. & BANG, O. (1908) Zentralbl. Bakteriol. Parasitenkd. Infektionskr. 46, 595-609.

ELLIS, R.W., DEFEO, D., MARYAK, J.M., YOUNG, H.A., SHIH, T.Y., CHANG, E.H., LOWY, D.R. AND SCOLNICK, E.M. (1980) J. Virol. 36, 408-420.

ELLIS, R.W., DEFEO, D., SHIH, T.Y., GONDA, M.A., YOUNG, H.A., TSUCHIDA, N., LOWY, D.R. & SCOLNICK, E.M. (1981) Nature (London) 292, 506-511.

ETZERODT, M., MIKKELSEN, T., PEDERSEN, F.S., KJELDGAARD, N.O. & JORGENSEN, P. (1984) Virology 134, 196-207.

EVANS, W.H., MAGE, M., HSU, C.K., HIMMELHOCH, S.R. & SMITH, G.H. (1978) Cancer Res. 38, 130-136.

FAMULARI, N.G., KOEHNE, C.F. & O'DONNELL, P.V. (1982) Proc. Natl. Acad. Sci. 79, 3872-3876.

FAN, H. & BALTIMORE, D. (1973) J. Mol. Biol. 80, 93-117.

FAN, H. & MUELLER-LANTZSCH, N. (1976) J. Virol. 18, 401-410.

FANNING, T.G. (1982) Nuc. Acids Res. 10, 5003-5013.

FARABAUGH, P.J. & FINK, G.R. (1980) Nature (London) 286, 352-256.

FARAS, A.J., TAYLOR, J.M., LEVINSON, W.E., GOODMAN, H.M. & BISHOP, J.M. (1973) J. Mol. Biol. 79, 163-183.

FENNER, F. (1976) Intervirology 7, 8-102.

FINNEGAN, D.J. (1981) Nature (London) 292, 800-801.

FINNEGAN, D.J. (1983) Nature (London) 302, 105-106.

FINNEGAN, D.J., WILL, B.H., BAYEV, A.A., BOWCOCK, A.M. & BROWN, L. (1982) In: Genome Evolution (Dover, G.A. & Flavell, R.B., eds.). Academic Press, London.

FINSTON, W.I. & CHAMPOUX, J.J. (1984) J. Virol. 51, 26-33.

FLAVELL, A. (1981) Nature (London) 289, 10-11.

FLAVELL, A.J. (1984) Nature (London) 310, 514-516.

FLAVELL, A.J. & ISH-HOROWICZ, D. (1981) Nature (London) 292, 591-595.

FOSTER, D.N., SCHMIDT, L.J., HODGSON, C.P., MOSES, H. & GETZ, M.J. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 7317-7321.

FRISBY, D.P., WEISS, R.A., ROUSSEL, M. & STEHELIN, D. (1979) Cell 17, 623-634.

GALIBERT, F., DE DINECHIN, S.D., RIGHI, M. & STEHELIN, D. (1984) EMBO J. 3, 1333-1338.

GARRETT, J.S., COUGHLIN, S.R., NIMAN, H.L., TREMBLE, P.M., GIELS, G.M., & WILLIAMS, L.T. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 7466-7470.

GAUTSCH, J.W. & WILSON, M.C. (1983) Nature (London) 301, 32-37.

GEBHARD, W., MEITINGER, T., HOCHTL, J. & ZACHAU, H.G. (1982) J. Mol. Biol. 157, 453-471.

GERARD, G.F., ROTTMAN, F. & GREEN, M. (1974) Biochemistry 13, 1632-1641.

GIANI, A.M., SMOTKIN, D. & WEINBERG, R.A. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 447-451.

GILBOA, E., MITRA, S.W., GOFF, S. & BALTIMORE, D. (1979a) Cell 18, 93-100.

GILBOA, E., GOFF, S., SHIELDS, A., YOSHIMURA, F., MITRA, S. & BALTIMORE, D. (1979b) Cell 16, 863-874.

GIRI, C.P., HODGSON, C.P., ELDER, P.K., COURTNEY, M.G. & GETZ, M.J. (1983) Nuc. Acids Res. 11, 305-319.

GOPINATHAN, K.P., WEYMOUTH, L.A., KUNKEL, T.A. & LOEB, L.A. (1979) Nature (London) 278, 857-859.

GOUBIN, G. & HILL, M. (1979) J. Virol. 29, 799-804.

GOVINDAN, M.V., SPIESS, E. & MAJORS, J. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 5157-5161.

GRAESSMANN, M., GRAESSMANN, A., WAGNER, H., WERNER, E. & SIMON, D. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 6470-6474.

GROSS, L. (1951) Proc. Soc. Exp. Biol. Med. 76, 27-32.

GROSS, L. (1957) Proc. Soc. Exp. Biol. Med. 94, 767-771.

GROSS, L. (1958) Acta Haematol. 19, 353-361.

GROSS, L. (1970) Oncogenic Viruses, 2nd edn. Pergamon Press, Oxford.

GROUDINE, M., EISENMAN, R. & WEINTRAUB, H. (1981) Nature (London) 292, 311-317.

GRUMMT, I. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 727-731.

GRUNSTEIN, M. & HOGNESS, D.S. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 3961-3965.

GWO-SHU LEE, M., LEWIS, S.A., WILDE, C.D. & COWAN, N.J. (1983) Cell 33, 477-487.

HAASE, A.T., STOWRING, L., NARAYAN, O., GRIFFIN, D. & PRICE, D. (1977) Science 195, 175-177.

HACKETT, A.J. & MANNING, J.S. (1971) J. Am. Vet. Med. Assoc. 158, 948-954.

HALL, W.T., HARTLEY, J.W. & SANFORD, K.K. (1968) J. Virol. 2, 238-247.

HANAFUSA, H. (1965) Virology 25, 248-255.

HARBERS, K., JAHNER, D. & JAENISCH, R. (1981) Nature (London) 293, 540-542.

HARBERS, K., KUEHN, M., DELIUS, H. & JAENISCH, R. (1984) Proc. Natl. Acad. Sci. 81, 1504-1508.

HARRIS, J.D., BLUM, H., SCOTT, J., TRAYNOR, B., VENTURA, P. & HAASE, A. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 7212-7215.

HARTER, D.H., HSU, K.C. & ROSE, H.M. (1968) Proc. Soc. Exp. Biol. Med. 129, 295-300.

HARTLEY, J.W. & ROWE, W.P. (1966) Proc. Natl. Acad. Sci. U.S.A. 55, 780-786.

HARTLEY, J.W. & ROWE, W.P. (1976) J. Virol. 19, 19-25.

HARTLEY, J.W., ROWE, W.P., CAPPS, W.I. & HUEBNER, R.J. (1969) J. Virol. 3, 126-132.

HARTLEY, J.W., WOLFORD, N.K., OLD, L.J. & ROWE, W.P. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 789-792.

HARVEY, J.J. (1964) Nature (London) 204, 1104-1105.

HASELTINE, W.A., MAXAM, A.M. & GILBERT, W. (1977a) Proc. Natl. Acad. Sci. U.S.A. 74, 989-993.

HASELTINE, W.A., PANET, A., SMOLER, D., BALTIMORE, D., PETERS, G. HARADA, F. & DAHLBERG, J.E. (1977b) Biochemistry 16, 3625-2632.

HAWLEY, R.G., SHULMAN, M.J. & HOZUMI, N. (1984) Mol. Cell Biol. 4 2565-2572.

HAWLEY, R.G., SHULMAN, M.J., MURIALDO, H., GIBSON, D.M. & HOZUMI, N. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 7425-7429.

HAYWARD, W.S., BRAVERMAN, S.B. & ASTRIN, S.M. (1980) Cold Spring Harbor Symp. Quant. Biol. 44, 1111-1122.

HAYWARD, W.S., NEEL, B.G. & ASTRIN, S.M. (1981) Nature (London) 290, 475-480.

HEBERLING, R.L., BARKER, S.T., KALTER, S.S., SMITH, G.C. & HELMKE, R.J. (1977) Science 195, 289-292.

HEINE, C.W., KELLY, D.C. & AVERY, R.J. (1980) J. Gen. Virol. 49, 385-395.

HELLMAN, A., PEEBLES, P.T., STRICKLAND, J.E., FOWLER, A.K., KALTER, S.S., OROSZLAN, S. & GILDEN, R.V. (1974) J. Virol. 14, 133-138.

HERSHKO, A., MAMONT, P., SHIELDS, R. & TOMKINS, G.M. (1971) Nature (London) New Biol. 232, 206-211.

HILL, M. & HILLOVA, J. (1972a) Nature (London) New Biol. 237, 35-39.

HILL, M. & HILLOVA, J. (1972b) Virology 49, 309-313.

HIRT, B. (1967) Proc. Natl. Acad. Sci. U.S.A. 74, 989-993.

HISHINUMA, F., DE BONA, P.J., ASTRIN, S.M. & SKALKA, A.M. (1981) Cell 23, 155-164.

HODGSON, C.P., ELDER, P.K., ONO, T., FOSTER, D.N. & GETZ, M.J. (1983) Mol. Cell Biol. 3, 2221-2231.

HOFFMANN, J.W., STEFFEN, D., GUSELLA, J., TABIN, C., BIRD, S., COWING, D. & WEINBERG, R.A. (1982) J. Virol. 44, 144-157.

HOGGAN, M.D., BUCKLER, C.E., SEARS, J.F., CHAN, H.W., ROWE, W.P. & MARTIN, M.A. (1982) J. Virol. 43, 8-17.

HOGGAN, M.D., BUCKLER, C.E., SEARS, J.F., ROWE, W.P. & MARTIN, M.A. (1983) J. Virol. 45, 473-477.

HOJMAN-MONTES DE OCA, F., DIANOUX, L., PERIES, J. & EMANOIL-RAVICOVITCH, R. (1983) J. Virol. 46, 307-310.

HOLLIS, G.F., HIETER, P.A., McBRIDE, O.W., SWAN, D. & LEDER, P. (1982) Nature (London) 296, 321-325.

HOOKS, J.J. & DETRICK-HOOKS, B. (1979) J. Gen. Virol. 44, 383-390.

HOOKS, J.J., GIBBS, C.J., CUTCHINS, E.C., ROGERS, N.G., LAMPERT, P. & GAJDUSEK, D.C. (1972) Arch. Gesamte Virusforsch. 38, 38-55.

HOWK, R.S., TROXLER, D.H., LOWY, D., DUESBERG, P.H. & SCOLNICK, E.M. (1978) J. Virol. 25, 115-123.

HOWLEY, P.M., ISRAEL, M.A., LAW, M-F. & MARTIN, M.A. (1979) J. Biol. Chem. 254, 4876-4883.

HSU, T.W., SABRAN, J.L., MARK, G.E., GUNTAKA, R.V. & TAYLOR, J.M. (1978) J. Virol. 28, 810-819.

HUANG, A.L., OSTROWSKI, M.C., BERARD, D. & HAGER, G.L. (1981) Cell 27, 245-255.

HUEBNER, R.J. & TODARO, G.J. (1969) Proc. Natl. Acad. Sci. U.S.A. 64, 1087-1094.

HUGHES, S. (1982) In: RNA Tumor Viruses, (Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds.), 2nd edn., pp 1285-1386. Cold Spring Harbor Laboratory, New York.

HUGHES, S.H. (1983) Curr. Top. Microbiol. Immunol. 103, 23-49.

HUGHES, S.H., ROBINSON, H.L., BISHOP, J.M. & VARMUS, H.E. (1979) Virology 99, 437-442.

HUGHES, S.H., MUTSCHLER, A., BISHOP, J.M. & VARMUS, H.E. (1981a) Proc. Natl. Acad. Sci. U.S.A. 78, 4299-4303.

HUGHES, S.H., TOYOSHIMA, K., BISHOP, J.M. & VARMUS, H.E. (1981b) Virology 108, 189-207.

HUNTER, T. (1984) J. Natl. Cancer Inst. 73, 773-786.

HUNTER, T. & SEFTON, B.M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1311-1315.

HWANG, L-H.S., PARK, J. & GILBOA, E. (1984) Mol. Cell Biol. 4, 2289-2297.

HYMAN, R.W., BRUNOVSKIS, I. & SUMMERS, W.C. (1973) J. Mol. Biol. 77, 189-196.

IHLE, J.N. & LEE, J.C. (1982) Curr. Top. Microbiol. Immunol. 98, 85-101.

IKAWA, Y., ROSS, J. & LEDER, P. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1154-1158.

ITIN, A. & KESHET, E. (1983a) J. Virol. 47, 178-184.

ITIN, A. & KESHET, E. (1983b) J. Virol. 47, 656-659.

ITIN, A., ROTMAN, G. & KESHET, E. (1983) Virology 127, 374-384.

JAENISCH, R. (1976) Proc. Natl. Acad. Sci. 73, 1260-1264.

JAENISCH, R. (1979) Virology 93, 80-90.

JAENISCH, R. (1980) Cell 19, 181-188.

JAENISCH, R., JAHNER, D., NOBIS, P., SIMON, I., LOHLER, J., HARBERS, K. & GROTKOPP, D. (1981) Cell 24, 519-529.

JAENISCH, R., HARBERS, K., SCHNIEKE, A., LOHLER, J., CHUMAKOV, I., JAHNER, D., GROTKOPP, D. & HOFFMANN, E. (1983) Cell 32, 209-216.

JAGADEESWARAN, P., FORGET, B.G. & WEISSMAN, S.M. (1981) Cell 26, 141-142.

JAMJOOM, G.A., NASO, R.B. & ARLINGHAUS, R.B. (1977) Virology 78, 11-34.

JAMJOOM, G., KARSHIN, W.L., NASO, R.B., ARCEMENT, L.J. & ARLINGHAUS, R.B. (1975) Virology 68, 135-145.

JANSEN, H.W., PATSCHINSKY, T. & BISTER, K. (1983a) J. Virol. 48, 61-73.

JANSEN, H.W., RUCKERT, B., LURZ, R. & BISTER, K. (1983b) EMBO J. 2, 1969-1975.

JEFFREYS, A.J. & HARRIS, S. (1982) Nature (London) 296, 9-10.

JELINEK, W.R. & SCHMID, C.W. (1982) Annu. Rev. Biochem. 51, 813-844.

JENKINS, J.R., RUDGE, K. & CURRIE, G.A. (1984) Nature (London) 312, 651-654.

JOLICOEUR, P. & RASSART, E. (1980) J. Virol. 33, 183-195.

JUNGHANS, R.P., DUESBERG, P.H. & KNIGHT, C.A. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 4895-4899.

KARIN, M. & RICHARDS, R.I. (1982) Nature (London) 299, 797-802.

KELLY, K., COCHRAN, B.H., STILES, C.D. & LEDER, P. (1983) Cell 35, 603-610.

KEOGH, E.V. (1938) Br. J. Exp. Pathol. 19, 1-9.

KESHET, E. & ITIN, A. (1982) J. Virol. 43, 50-58.

KESHET, E. & SHAUL, Y. (1981) Nature (London) 289, 83-85.

KESHET, E., SHAUL, Y., KAMINCHIK, J. & AVIV, H. (1980) Cell 20, 431-439.

KHOURY, G. & GRUSS, P. (1983) Cell 33, 313-314.

KIRSTEN, W.H. & MAYER, L.A. (1967) J. Natl. Cancer Inst. 39, 311-335.

KIRSTEN, W.H. & MAYER, L.A. (1969) J. Natl. Cancer Inst. 43, 735-746.

KLEMENT, V., ROWE, W.P., HARTLEY, J.W. & PUGH, W.E. (1969) Proc. Natl. Acad. Sci. U.S.A. 63, 753-758.

KLEMPNAUER, K.-H., GONDA, T.J. & BISHOP, J.M. (1982) Cell 31, 453-463.

KLEMPNAUER, K.-H., SYMONDS, G., EVAN, G.I. & BISHOP, J.M. (1984) Cell 37, 537-547.

KLOETZER, W.S., MAXWELL, S.A. & ARLINGHAUS, R.B. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 412-416.

KRAKOWER, J.M., BARBACID, M. & AARONSON, S.A. (1977) J. Virol. 22, 331-339.

KUFF, E.L., LUEDERS, K.K. & SCOLNICK, E.M. (1978) J. Virol. 28, 66-74.

KUFF, E.L., SMITH, L.A. & LUEDERS, K.K. (1981) Mol. Cell Biol. 1, 216-227.

KUFF, E.L., WIVEL, N.A. & LUEDERS, K.K. (1968) Cancer Res. 28, 2137-2148.

KUFF, E.L., LUEDERS, K.K., OZER, H.L. & WIVEL, N.A. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 218-222.

KUFF, E.L., FEENSTRA, A., LUEDERS, K., SMITH, L., HAWLEY, R., HOZUMI, N. & SHULMAN, M. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1992-1996.

KULGUSKIN, V.V., ILYIN, Y.V., & GEORGIEV, G.P. (1981) Nuc. Acids Res. 9, 3451-3463.

KUNG, H.-J., SHANK, P.R., BISHOP, J.M. & VARMUS, H.E. (1980) Virology 103, 425-433.

LAI, M.-H.T. & VERMA, I.M. (1980) Virology 100, 194-198.

LAND, H., PARADA, L.F. & WEINBERG, R.A. (1983) Nature (London) 304, 596-602.

LANE, D.P. (1984) Nature (London) 312, 596-597.

LANGFORD, C.J. & GALLWITZ, D. (1983) Cell 33, 519-527.

LANGRIDGE, J., LANGRIDGE, P. & BERQUIST, P.I. (1980) Anal. Biochem. 103, 264-271.

LASKEY, R.A. & MILLS, A.D. (1977) FEBS Lett. 82, 314-316.

LATARJET, R., & DUPLAN, J.-F. (1962) Int. J. Radiat. Biol. Relat. Stud. Phys. Chem. Med. 5, 339-344.

LEDBETTER, J.A., NOWINSKI, R.C. & EISENMAN, R.N. (1978) Virology 91, 116-129.

LEDER, P., BATTEY, J., LENOIR, G., MOULDING, C., MURPHY, W., POTTER, H., STEWART, T. & TAUB, R. (1983) Science 222, 765-771.

LEE, J.S., VARMUS, H.E. & BISHOP, J.M. (1979) J. Biol. Chem. 245, 8015-8022.

LEE, W.-H., NUNN, M. & DUESBERG, P.H. (1981) J. Virol. 39, 756-758.

LEIS, J.P., McGINNIS, J. & GREEN, R.W. (1978) Virology 84, 87-98.

LEMONS, R.S., NASH, W.G., O'BRIEN, S.J., BENVENISTE, R.E. & SHERR, C.J. (1978) Cell 14, 995-1005.

LERNER, R.A., WILSON, C.B., DEL VILLANO, B.C., McCONAHEY, P.J. & DIXON, F.J. (1976) J. Exp. Med. 143, 151-166.

LEVIN, J.G. & ROSENAK, M.J. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 1154-1158.

LEVINSON, B., KHOURY, G., VANDE WOUDE, G. & GRUSS, P. (1982) Nature (London) 295, 568-572.

LEVINSON, W.E., VARMUS, H.E., GARAPIN, A.C. & BISHOP, J.M. (1972) Science 175, 76-78.

LEVY, J.A. (1973) Science 182, 1151-1153.

LEVY, J.A. (1975) Nature 253, 140-142.

LEVY, J.A. (1978) Curr. Top. Microbiol. Immunol. 79, 109-213.

LIEBER, M.M., SHERR, C.J., TODARO, G.J., BENVENISTE, R.E., CALLAHAN, R. & COON, H.G. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 2315-2319.

LIEBERMAN, M. & KAPLAN, H.S. (1959) Science 130, 387-388.

LILLY, F. & PINCUS, T. (1967) Adv. Cancer Res. 17, 231-277.

LINIAL, M. & BLAIR, D. (1982) In: RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds.), 2nd edn., pp 649-783. Cold Spring Harbor Laboratory, New York.

LINIAL, M., MEDEIROS, E. & HAYWARD, W.S. (1978) Cell 15, 1371-1381.

LITTLE, C.C. (1947) Biol. Rev. Cambridge Philos. Soc. 22, 315-343.

LIU, W.-T., NATORI, T., CHANG, K.S.S. & WU, A.M. (1977) Arch. Virol. 55, 187-200.

LOWY, D.R. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 5539-5543.

LUEDERS, K.K. & KUFF, E.L. (1977) Cell 12, 963-972.

LUEDERS, K.K. & KUFF, E.L. (1979) J. Virol. 30, 225-231.

LUEDERS, K.K. & KUFF, E.L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3571-3575.

LUEDERS, K.K. & KUFF, E.L. (1983) Nuc. Acids Res. 11, 4391-4408.

LWOFF, A. (1960) Cancer Res. 20, 820-829.

MAJORS, J.E. & VARMUS, H.E. (1981) Nature (London) 289, 253-258.

MALMQUIST, W.A., VAN DER MAATEN, M.J. & BOOTHE, A.D. (1969) Cancer Res. 29, 188-200.

MANDEL, M. & HIGA, A. (1970) J. Mol. Biol. 53, 159-162.

MANIATIS, T., FRITSCH, E.F. & SAMBROOK, J. (1982) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

MANIATIS, T., JEFFREY, A. & KLEID, D.G. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 1184-1188.

MANIATIS, T., HARDISON, R.C., LACY, E., LAUER, J., O'CONNELL, C., QUON, D., SIM, G.K. & EFSTRATIADIS, A. (1978) Cell 15, 687-701.

MANN, R., MULLIGAN, R.C. & BALTIMORE, D. (1983) Cell 33, 153-159.

MARMUR, J. & DOTY, P. (1962) J. Mol. Biol. 5, 109-118.

MARTIN, M.A., BRYAN, T., McCUTCHAN, T.F. & CHAN, H.W. (1981) J. Virol. 39, 835-844.

MARTIN, S.L., VOLIVA, C.F., BURTON, F.H., EDGELL, M.H. & HUTCHISON, C.A. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 2308-2312.

McCLINTOCK, P.R., IHLE, J.N. & JOSEPH, D.R. (1977) J. Exp. Med. 146, 422-434.

McGRATH, M.S. & WEISSMAN, I.L. (1979) Cell 17, 65-75.

McGRATH, C.M., MARINEAU, E.J. & VOYLES, B.A. (1978) Virology 87, 339-353.

MINNA, J.D., LUEDERS, K.K. & KUFF, E.L. (1974) J. Natl. Cancer Inst. 52, 1211-1217.

MITRA, S.W., GOFF, S., GILBOA, E. & BALTIMORE, D. (1979) Proc. Natl. Acad. Sci. 76, 4355-4359.

MIZUTANI, S. & TEMIN, H.M. (1976) Biochemistry 15, 1510-1516.

MOELLING, K., BOLOGNESI, D.P., BAUER, H., BUSEN, W., PLASSMAN, H.W., HAUSEN, P. (1971) Nature (London) New Biol. 234, 240-243.

MOLONEY, J.B. (1966) Natl. Cancer Inst. Monogr. 22, 139-142.

MONDAL, H. & HOFSCHNEIDER, P.H. (1982) Int. J. Cancer 30, 281-287.

MONDAL, H. & HOFSCHNEIDER, P.H. (1983) Biochem. Biophys. Res. Commun. 116, 303-311.

MORONI, C. & SCHUMANN, G. (1977) Nature (London) 269, 600-601.

MORRIS, A., CLEGG, C., JONES, J., RODGERS, B. & AVERY, R.J. (1980) J. Gen Virol. 49, 105-113.

MORSE, H.C. & HARTLEY, J.W. (1982) Curr. Top. Microbiol. Immunol. 98, 17-26.

MOSES, H.L., BRANUM, E.L., PROPER, J.A. & ROBINSON, R.A. (1981) Cancer Res. 41, 2842-2848.

MUHLBOCK, O. (1965) Eur. J. Cancer 1, 123-124.

MULLER, H.K., BALL, G., EPSTEIN, M.A., ACHONG, B.G., LENOIR, G. & LEVIN, A. (1980) J. Gen. Virol. 47, 399-406.

NAHARRO, G., ROBBINS, K.C. & REDDY, E.P. (1984) Science 223, 63-66.

NELSON, J., LEONG, J. & LEVY, J.A. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 6263-6267.

NELSON, J., LEVY, J.A. & LEONG, J.C. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 1670-1674.

NEMO, G.J., BROWN, P.W., GIBBS, C.J., GAJDUSEK, D.C. (1978) Infect. Immun. 20, 69-72.

NERMUT, M.V., HERMANN, F. & SCHAFER, W. (1972) Virology 49, 345-358.

NEWBOLD, R.F. & OVERELL, R.W. (1983) Nature (London) 304, 648-651.

NEXO, B.A. (1977) Virology 77, 849-852.

NOLL, H. (1967) Nature (London) 215, 360-363.

NORDHEIM, A. & RICH, A. (1983) Nature (London) 303, 674-678.

NORTON, J.D. & AVERY, R.J. (1982) Biochem. Biophys. Res. Commun. 108, 1631-1637.

NORTON, J.D. & AVERY, R.J. (1984) J. Gen. Virol. 65, 309-316.

NORTON, J.D., CARTER, A.T. & AVERY, R.J. (1982) J. Gen. Virol. 58, 95-106.

NORTON, J.D., CARTER, A.T. & AVERY, R.J. (1984a) Biochem. Biophys. Res. Commun. 119, 150-156.

NORTON, J.D., CONNOR, J. & AVERY, R.J. (1984b) Nuc. Acids Res. 12, 3445-3460.

NORTON, J.D., CONNOR, J. & AVERY, R.J. (1984c) Nuc. Acids Res. 12, 6839-6852.

NUSSE, R. & VARMUS, H.E. (1982) Cell 31, 99-109.

O'CONNELL, C., O'BRIEN, S., NASH, W.G. & COHEN, M. (1984) Virology 138, 225-235.

OHNO, S. (1970) Evolution by Gene Duplication. Springer, Heidelberg.

ONO, M. & OHISHI, H. (1983) Nuc. Acids Res. 11, 7169-7180.

ONO, M., COLE, M.D., WHITE, A.T. & HUANG, R.C.C. (1980) Cell 21, 465-473.

ONO, M., KITASATO, H., OHISHI, H. & MOTOBAYASHI-NAKAJIMA, Y. (1984) J. Virol. 50, 352-358.

ORGEL, L.E. & CRICK, F.H.C. (1980) Nature (London) 284, 604-607.

OSKARSSON, M., McCLEMENTS, W.L., BLAIR, D.G., MAIZEL, J.V. & VANDE WOUDE, G.F. (1980) Science 207, 1222-1224.

PANET, A., BALTIMORE, D. & HANAFUSA, H. (1975) J. Virol. 16, 146-152.

PANGANIBAN, A.T. & TEMIN, H.W. (1984) Cell 36, 673-679.

PARADA, L.F., LAND, H., WEINBERG, R.A., WOLF, D. & ROTTER, V. (1984) Nature 312, 649-651.

PARKER, R.C., VARMUS, H.E. & BISHOP, J.M. (1984) Cell 37, 131-139.

PAYNE, G.S., BISHOP, J.M. & VARMUS, H.E. (1982) Nature (London) 295, 209-213.

PAYNE, L.N. & CHUBB, R.C. (1968) J. Gen. Virol. 3, 379-391.

PEACOCK, W.J., LOHE, A.R., GERLACH, W.L., DUNSMUIR, P., DENNIS, E.S. & APPELS, R. (1978) Cold Spring Harbor Symp. Quant. Biol. 42, 1121-1135.

PERIES, J., ALVES-CARDOSO, E., CANIVET, M., DEBONS-GUILLEMEN, M.C. & LASNERET, J. (1977) J. Natl. Cancer Inst. 59, 463-465.

PETERS, G.G. & HU, J. (1980) J. Virol. 36, 692-700.

PETERS, G., HARADA, F., DAHLBERG, J.E., PANET, A., HASELTINE, W.A. & BALTIMORE, D. (1977) J. Virol. 21, 1031-1041.

POTTER, S.S., BROREIN, W.J., DUNSMUIR, P. & RUBIN, G.M. (1979) Cell 17, 415-427.

PROUDFOOT, N.J. & BROWNLEE, G.G. (1974) Nature (London) 252, 359-362.

QUADE, K. (1979) Virology 98, 461-465.

QUIGLEY, J.P., RIFKIN, D.B. & REICH, E. (1971) Virology 46, 106-116.

RAPP, U.R. & TODARO, G.J. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 624-628.

RAPP, U.R., REYNOLDS, F.H. & STEPHENSON, J.R. (1983) J. Virol. 45, 914-924.

RASHEED, S., GARDNER, M.B. & CHAN, E. (1976) J. Virol. 19, 13-18.

RAZIN, A. & RIGGS, A.D. (1980) Science 210, 604-610.

RAZIN, A. & SZYF, M. (1984) Biochim. Biophys. Acta 728, 331-342.

RECHAVI, G., GIVOL, D. & CANAANI, E. (1982) Nature (London) 300, 607-611.

REDDY, E.P., SMITH, M.J. & SRINIVASAN, A. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3623-3627.

RESNICK, R., OMER, C.A. & FARAS, A.J. (1984) J. Virol. 51, 813-821.

REZNIKOFF, C.A., BRANKOW, D.W. & HEIDELBERGER, C. (1973) Cancer Res. 33, 3231-3238.

RIGBY, P.W.J., DIECKMANN, M., RHODES, C. & BERG, P. (1977) J. Mol. Biol. 113, 237-251.

RINGOLD, G.M. (1979) Biochim. Biophys. Acta 560, 487-508.

RISSER, R., HOROWITZ, J.M. & McCUBREY, J. (1983) Annu. Rev. Genet. 17, 85-121.

ROBERTS, A.B., LAMB, L.C., NEWTON, D.L., SPORN, M.B., DELARCO, J.E. & TODARO, G.J. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3494-3498.

ROBINSON, H.L. (1967) Proc. Natl. Acad. Sci. U.S.A. 57, 1655-1662.

ROBINSON, H.L., ASTRIN, S.M., SENIOR, A.M. & SALAZAR, F.H. (1981) J. Virol. 40, 745-751.

ROEDER, G.S. & FINK, G.R. (1980) Cell 21, 239-249.

ROHDE, W., PAULI, G., HENNING, J. & FRIIS, R.R. (1978) Arch. Virol. 58, 55-59.

ROMMELAERE, J., FALLER, D.V. & HOPKINS, N. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 495-499.

ROTHENBERG, E. & BALTIMORE, D. (1977) J. Virol. 21, 168-178.

ROTHENBERG, E., DONOGHUE, D.J. & BALTIMORE, D. (1978) Cell 13, 435-451.

ROTHENBERG, E., SMOTKIN, D., BALTIMORE, D. & WEINBERG, R.A. (1977) Nature (London) 269, 122-126.

ROTMAN, G., ITIN, A. & KESHET, E. (1984) Nuc. Acids Res. 12, 2273-2282.

ROUS, P. (1910) J. Exp. Med. 12, 696-705.

ROUS, P. (1911) J. Exp. Med. 13, 397-411.

ROUSSEL, M.F., RETTENMIER, C.W., LOOK, A.T. & SHERR, C.J. (1984) Mol. Cell Biol. 4, 1999-2009.

ROVIGATTI, U.G. & ASTRIN, S.M. (1983) Curr. Top. Microbiol. Immunol. 103, 1-21.

ROWE, W.P., PUGH, W.E. & HARTLEY, J.W. (1970) Virology 42, 1136-1139.

ROWE, W.P. & KOZAK, C.A. (1980) Proc. Natl. Acad. Sci. 77, 4871-4874.

ROWLEY, J. (1983) Nature (London) 301, 290-291.

RUBIN, C.S. & ROSEN, O.M. (1975) Annu. Rev. Biochem. 44, 831-887.

RUBIN, G.M., KIDWELL, M.G. & BINGHAM, P.M. (1982) Cell 29, 987-994.

RUBIN, H. (1965) Virology 26, 270-276.

RULEY, H.E. (1983) Nature (London) 304, 602-606.

SALUZ, H.P., SCHMIDT, T., DUDLER, R., ALTWEGG, M., STUMM-ZOLLINGER, E., KUBLI, E. & CHEN, P.S. (1983) Nuc. Acids Res. 11, 77-90.

SAWYER, R.C. & DAHLBERG, J.E. (1973) J. Virol. 12, 1226-1237.

SAWYER, R.C. & HANAFUSA, H. (1979) J. Virol. 29, 863-871.

SCHEIDEREIT, C. & BEATO, M. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 3029-3033.

SCHEIDEREIT, C., GEISSE, S., WESTPHAL, H.M. & BEATO, M. (1983) Nature (London) 304, 749-752.

SCHILDKRAUT, C. & LIFSON, S. (1965) Biopolymers 3, 195-208.

SCHLOM, J., SPIEGELMAN, S. & MOORE, D.H. (1971) Nature (London) 231, 97-100.

SCHNIEKE, A., HARBERS, K. & JAENISCH, R. (1983) Nature (London) 304, 315-320.

SCHWAB, M., ALITALO, K., VARMUS, H.E., BISHOP, J.M. & GEORGE, D. (1983) Nature (London) 303, 497-501.

SCOLNICK, E.M., GOLDBERG, R.J. & WILLIAMS, D. (1976a) J. Virol. 18, 559-566.

SCOLNICK, E.M., YOUNG, H.A. & PARKS, W.P. (1976b) Virology 69, 148-156.

SCOLNICK, E.M., RANDS, E., WILLIAMS, D. & PARKS, W.P. (1973) J. Virol. 12, 458-463.

SCOLNICK, E.M., VASS, W.C., HOWK, R.S. & DUESBERG, P.H. (1979) J. Virol. 29, 964-972.

SEIF, I., KHOURY, G. & DHAR, R. (1979) Cell 18, 963-977.

SESSA, G., FREER, J.H., COLACICCO, G. & WEISSMAN, G. (1969) J. Biol. Chem. 244, 3575-3582.

SHANK, P. & LINIAL, M. (1980) J. Virol. 36, 450-456.

SHANK, P.R. & VARMUS, H.E. (1978) J. Virol. 25, 104-114.

SHANK, P.R., COHEN, J.C., VARMUS, H.E., YAMAMOTO, K.R. & RINGOLD, G.M. (1978a) Proc. Natl. Acad. Sci. U.S.A. 75, 2112-2116.

SHANK, P.R., HUGHES, S.H., KUNG, H.J., MAJORS, J.E., QUINTRELL, N., GUNTAKA, R.V., BISHOP, J.M. & VARMUS, H.E. (1978b) Cell 15, 1383-1395.

SHAPIRO, J. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1933-1937.

SHARP, P.A. (1981) Cell 23, 643-646.

SHARP, P.A. (1983) Nature (London) 301, 471-472.

SHENK, T. (1981) Curr. Top. Microbiol. Immunol. 93, 25-46.

SHEN-ONG, G.L.C. & COLE, M.D. (1982) J. Virol. 42, 411-421.

SHEN-ONG, G.L.C. & COLE, M.D. (1984) J. Virol. 49, 171-177.

SHEN-ONG, G.L.C., KEATH, E.J., PICCOLI, S.P. & COLE, M.D. (1982) Cell 31, 443-452.

SHERR, C.J., LIEBER, M.M., BENVENISTE, R.E. & TODARO, G.J. (1974) Virology 58, 492-503.

SHERWIN, S.A., RAPP, U.R., BENVENISTE, R.E., SEN, A. & TODARO, G.J. (1978) J. Virol. 26, 257-264.

SHIBA, R. & SAIGO, K. (1983) Nature (London) 302, 119-124.

SHIELDS, A., WITTE, O.N., ROTHENBERG, E. & BALTIMORE, D. (1978) Cell 14, 601-609.

SHIH, T.Y., WILLIAMS, D.R., WEEKS, M.O., MARYAK, J.M., VASS, W.C. & SCOLNICK, E.M. (1978) J. Virol. 27, 45-55.

SHILO, B.-Z. & WEINBERG, R.A. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 6789-6792.

SHIMOTOHNO, K. & TEMIN, H.M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 7357-7361.

SHIMOTOHNO, K., MIZUTANI, S. & TEMIN, H.M. (1980) Nature (London) 285, 550-554.

SHOPE, R.E. & HURST, E.W. (1933) J. Exp. Med. 58, 607-615.

SIGURDSSON, B. (1954) Br. Vet. J. 110, 255-270.

SIGURDSSON, B., THORMAR, H. & PALSSON, P.A. (1960) Arch. Gesamte Virusforsch. 10, 368-381.

SIMON, D., STUHLMANN, H., JAHNER, D., WAGNER, H., WERNER, E. & JAENISCH, R. (1983) Nature (London) 304, 275-277.

SINGER, M.F. (1982) Cell 28, 433-434.

SINGER, M.F., THAYER, R.E., GRIMALDI, G., LERMAN, M.I. & FANNING, T.G. (1983) Nuc. Acids Res. 11, 5739-5745.

SLAMON, D.J. & CLINE, M.J. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 7141-7145.

SMITH, G.C., HEBERLING, R.L., HELMKE, R.J., BARKER, S.T. & KALTER, S.S. (1977) J. Natl. Cancer Inst. 59, 975-979.

SOBIS, H. & VANDEPUTTE, M. (1978) J. Natl. Cancer Inst. 61, 891-893.

SORGE, J., CUTTING, A.E., ERDMAN, V.D. & GAUTSCH, J.W. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 6627-6631.

SOUTHERN, E.M. (1975) J. Mol. Biol. 98, 503-517.

SPECTOR, D.H., VARMUS, H.E. & BISHOP, J.M. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 4102-4106.

STACEY, D.W. (1979) J. Virol. 29, 949-956.

STEEVES, R.A., ECKNER, R.J., MIRAND, E.A. & PRIORE, R.L. (1971) J. Natl. Cancer Inst. 46, 1219-1228.

STEFFEN, D.L. & ROBINSON, H. (1982) Curr. Top. Microbiol. Immunol. 98, 1-10.

STEFFEN, D. & WEINBERG, R.A. (1978) Cell 15, 1003-1010.

STEFFEN, D.L., BIRD, S. & WEINBERG, R.A. (1980) J. Virol. 35, 824-835.

STEFFEN, D., BIRD, S., ROWE, W.P. & WEINBERG, R.A. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 4554-4558.

STEHELIN, D., VARMUS, H.E., BISHOP, J.M. & VOGT, P.K. (1976) Nature (London) 260, 170-173.

STEWART, C.L., STUHLMANN, H., JAHNER, D. & JAENISCH, R. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 4098-4102.

STILES, C.D. (1983) Cell 33, 653-655.

STOLL, E., BILLETER, M.A., PALMENBERG, A., & WEISSMANN, C. (1977) Cell 12, 57-72.

STROMBERG, K., HURLEY, N.E., DAVIS, N.L., RUECKERT, R.R. & FLEISSNER, E. (1974) J. Virol. 13, 513-528.

STRONG, L.C. (1935) Genetics 20, 586-591.

STUHLMAN, H., JAHNER, D. & JAENISCH, R. (1981) Cell 26, 221-232.

SUTCLIFFE, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90.

SUTCLIFFE, J.G., SHINNICK, T.M., VERMA, I.M. & LERNER, R.A. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3302-3306.

SUTTON, W.D. & McCALLUM, M. (1971) Nature (London) New Biol. 232, 83-85.

SVOBODA, J., CHYLE, P.P., SIMKOVIC, D. & HILGERT, I. (1963) Folia Biol. (Prague) 9, 77-81.

SWANSTROM, R. & SHANK, P.R. (1978) Anal. Biochem. 86, 184-192.

SWANSTROM, R., DE LORBE, W.J., BISHOP, J.M. & VARMUS, H.E. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 124-128.

SWANSTROM, R., PARKER, R.C., VARMUS, H.E. & BISHOP, J.M. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 2519-2523.

TAKEYA, T. & HANAFUSA, H. (1983) Cell 32, 881-890.

TANAKA, H. (1977) Virology 76, 835-850.

TAUB, R., MOULDING, C., BATTEY, J., MURPHY, W., VASICEK, T., LENOIR, G.M. & LEDER, P. (1984) Cell 36, 339-348.

TAYLOR, J.M. (1977) Biochim. Biophys. Acta 473, 57-71.

TAYLOR, J.M., ILLMENSEE, R. & SUMMERS, J. (1976) Biochim. Biophys. Acta 442, 324-330.

TEICH, N. (1982) In: RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds.), 2nd edn., pp. 25-207. Cold Spring Harbor Laboratory, New York.

TEICH, N., WYKE, J., MAK, T., BERNSTEIN, A. & HARDY, W. (1982) In: RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds.), 2nd edn., pp 785-998. Cold Spring Harbor Laboratory, New York.

TEMIN, H.M. (1961) Virology 13, 158-163.

TEMIN, H.M. (1962) Cold Spring Harbor Symp. Quant. Biol. 27, 407-414.

TEMIN, H.M. (1963) Virology 20, 577-582.

TEMIN, H.M. (1964) Natl. Cancer Inst. Monogr. 17, 557-570.

TEMIN, H.M. (1980) Cell 21, 599-600.

TEMIN, H.M. (1982) Cell 28, 3-5.

TEMIN, H.M. & BALTIMORE, D. (1972) In: Advances in Virus Research (Smith, K., Lauffer, M. & Band, F., eds.) 17, pp 129-185. Academic Press, New York.

TEMIN, H.M. & MIZUTANI, S. (1970) Nature (London) 226, 1211-1213.

TEMIN, H.M. & RUBIN, H. (1958) Virology 6, 669-688.

TEREBA, A. (1981) J. Virol. 40, 920-926.

TEREBA, A. (1983) Curr. Top. Microbiol. Immunol. 107, 29-50.

TEREBA, A. & ASTRIN, S.M. (1980) J. Virol. 35, 888-894.

THEILEN, G.H., GOULD, D., FOWLER, M. & DUNGWORTH, D.L. (1971) J. Natl. Cancer Inst. 47, 881-889.

THEOFILOPOULOS, A.N., SHAWLER, D.L., BALDERAS, R.S., ELDER, J.H., KATZ, D.H. & DIXON, F.J. (1981) J. Immunol. 126, 1154-1159.

THOMAS, M. & DAVIS, R.W. (1975) J. Mol. Biol. 91, 315-328.

TODARO, G.J. & HUEBNER, R.J. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1009-1015.

TODARO, G.J., FRYLING, C. & DE LARCO, J.E. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5258-5262.

TODARO, G.J., BENVENISTE, R.E., SHERR, C.J., SCHLOM, J., SCHIDLOVSKY, G. & STEPHENSON, J.R. (1978) Virology 84, 189-194.

TOOZE, J. (ed.) (1980) DNA Tumor Viruses, 2nd edn. Cold Spring Harbor Laboratory, New York.

TSICHLIS, P.N., STRAUSS, P.G. & HU, L.F. (1983) Nature (London) 302, 445-449.

TSIPIS, J.E. & BRATT, M.A. (1976) J. Virol. 18, 848-855.

TSUCHIDA, N.R., GILDEN, R.V. & HATANAKA, M. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 4503-4507.

ULLU, E. & TSCHUDI, C. (1984) Nature (London) 312, 171-172.

ULLU, E., MURPHY, S. & MELLI, M. (1982) Cell 29, 195-202.

URIELI-SHOVAL, S., GRUENBAUM, Y., SEDAT, J. & RAZIN, A. (1982) FEBS Lett. 146, 148-152.

VAN BEVEREN, C., ENAMI, S., CURRAN, T. & VERMA, I.M. (1984) Virology 135, 229-243.

VAN BEVEREN, C., GODDARD, J.G., BERNS, A. & VERMA, I.M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3307-3311.

VAN BEVEREN, C., RANDS, E., CHATTOPADHYAY, S.K., LOWY, D.R. & VERMA, I.M. (1982) J. Virol. 41, 542-556.

VANYUSHIN, B.F., TKACHEVA, S.G. & BELOZERSKY, A.N. (1970) Nature (London) 225, 948-949.

VARDIMON, L., KRESSMAN, A., CEDAR, H., MAECHLER, M. & DOERFLER, W. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 1073-1077.

VARMUS, H.E. (1982) Science 216, 812-820.

VARMUS, H. & SWANSTROM, R. (1982) In: RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds.) 2nd edn., pp 369-512. Cold Spring Harbor Laboratory, New York.

VARMUS, H.E., BISHOP, J.M., NOWINSKI, R.C. & SARKAR, N.H. (1972) Nature (London) New Biol. 238, 189-191.

VARMUS, H.E., GUNTAKA, R.V., DENG, C.T. & BISHOP, J.M. (1975) Cold Spring Harbor Symp. Quant. Biol. 39, 987-996.

VARMUS, H.E., HEASLEY, S., KUNG, H.-J., OPPERMANN, H., SMITH, V.C., BISHOP, J.M. & SHANK, P.R. (1978) J. Mol. Biol. 120, 55-82.

VARMUS, H.E., SHANK, P.R., HUGHES, S.E., KUNG, H.-J., HEASLEY, S., MAJORS, J., VOGT, P.K. & BISHOP. J.M. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 851-864.

VERMA, I.M. (1977) Biochim. Biophys. Acta 473, 1-38.

VOGT, V.M., EISENMAN, R. & DIGGELMANN, H. (1975) J. Mol. Biol. 96, 471-493.

VOGT, V.M., WIGHT, W. & EISENMAN, R. (1979) Virology 98, 154-167.

WAALWIJK, C. & FLAVELL, R.A. (1978) Nuc. Acids Res. 5, 3231-3236.

WAHL, G.M., STERN, M. & STARK, G.R. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3683-3687.

WALTER, P. & BLOBEL, G. (1982) Nature (London) 299, 691-698.

WANG, L.-H., DUESBERG, P., BEEMON, K. & VOGT, P.K. (1975) J. Virol. 16, 1051-1070.

WATERFIELD, M.D., SCRACE, G.T., & WHITTLE, N. (1983) Nature (London) 304, 35-39.

WECKER, E. & HORAK, I. (1982) Curr. Top. Microbiol. Immunol. 98, 27-36.

WEIHER, H., KONIG, M. & GRUSS, P. (1983) Science 219, 626-631.

WEINBERG, R.A. (1977) Biochim. Biophys. Acta 473, 39-55.

WEINBERG, R.A. (1980) Annu. Rev. Biochem. 49, 197-226.

WEINMASTER, G., ZOLLER, M.J., SMITH, M., HINZE, E. & PAWSON, T. (1984) Cell 37, 559-568.

WEISS, R. (1969) J. Gen Virol. 5, 511-528.

WEISS, R. (1982) In: RNA Tumor Viruses (Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds.) 2nd edn., pp 209-260. Cold Spring Harbor Laboratory, New York.

WEISS, R. (1984) Nature (London) 309, 12-13.

WEISS, R.A. & BIGGS, P.M. (1972) J. Natl. Cancer Inst. 49, 1713-1725.

WEISS, S.R., VARMUS, H.E. & BISHOP, J.M. (1977) Cell 12, 983-992.

WEISS, S.R., VARMUS, H.E. & BISHOP, J.M. (1981) J. Virol. 110, 476-478.

WEISSMANN, C., BILLETER, M.A., GOODMAN, H.M., HINDLEY, J. & WEBER, H. (1973) Annu. Rev. Biochem. 42, 303-328.

WEJMAN, J.C., TAYLOR, B.A., JENKINS, N.A. & COPELAND, N.G. (1984) J. Virol. 50, 237-247.

WELLAUER, P.K., REEDER, R.H., CARROLL, D., BROWN, D.D., DEUTCH, A., HIGASHINAKAGAWA, T. & DAWID, I.B. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 2823-2827.

WELLS, R.D., FLUGEL, R.M., LARSON, J.E., SCHENDEL, P.F. & SWEET, R.W. (1972) Biochemistry 11, 621-629.

WESTAWAY, D., PAYNE, G. & VARMUS, H.E. (1984) Proc. Natl. Acad. Sci. USA 81, 843-847.

WILHELMSEN, K.C., EGGLETON, K. & TEMIN, H.M. (1984) J. Virol. 52, 172-182.

WILL, B.M., BAYEV, A.A. & FINNEGAN, D.J. (1981) J. Mol. Biol. 153, 897-915.

WILLIAMS, J.C. & BELL, R.M. (1972) Biochim. Biophys. Acta 228, 255-262.

WILSON, R. & STORB, U. (1983) Nuc. Acids Res. 11, 1803-1817.

WILSON, S.H. & KUFF, E.L. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1531-1536.

WITTE, O.N. & BALTIMORE, D. (1978) J. Virol. 26, 750-761.

WYKE, J.A. (1973) Virology 52, 587-590.

WYKE, J.A. & LINIAL, M. (1973) Virology 53, 152-161.

YANG, W.K., KIGGANS, J.O., YANG, D.-M., OU, C.-Y., TENNANT, R.W., BROWN, A. & BASSIN, R.H. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 2994-2998.

YOSHIMURA, F.K. & WEINBERG, R.A. (1978) Cell 16, 323-332.

YOUNG, H.A., GONDA, M.A., DEFEO, D., ELLIS, R.W., NAGASHIMA, K. & SCOLNICK, E.M. (1980) Virology 107, 89-99.

ZAVADA, J. (1972) J. Gen. Virol. 15, 183-191.

ZAVADA, J. & MACPHERSON, I. (1970) Nature (London) 225, 24-26.