https://theses.gla.ac.uk/

Theses Digitisation:

https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

CHARACTERISATION OF ENDOGENOUS RETROVIRAL ELEMENTS IN CANINE DNA.

Richard Adrian Squires BVSc., DVR., Dip.ACVIM., MRCVS.

A thesis submitted for the degree of Doctor of Philosophy
in the University of Glasgow.

Departments of Veterinary Surgery and Veterinary Pathology,
University of Glasgow,
November 1990

© Richard A. Squires 1990

ProQuest Number: 10984131

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10984131

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

TABLE OF CONTENTS

page number

TABLE OF CONTENTS		i
ACKNOWLEDGEMENTS		ii
DECLARATION		iii
ABBREVIATIONS		iv
LIST OF FIGURES		vii
LIST OF TABLES		xii
SUMMARY		xiii
CHAPTER 1 :	INTRODUCTION	1
CHAPTER 2 :	MATERIALS AND METHODS	40
CHAPTER 3 :	SUB-CLONING AND SEQUENCING OF TWO DEFECTIVE CANINE RETROVIRAL ELEMENTS	81
CHAPTER 4 :	DISTRIBUTION OF ENDOGENOUS RETROVIRAL ELEMENTS IN CANINE GENOMIC DNA	126
CHAPTER 5 :	RNA EXPRESSION OF CANINE ENDOGENOUS RETROVIRAL ELEMENTS	163
CHAPTER 6 :	CHARACTERISATION OF A DNA SEQUENCE PRESENT BOTH IN CANINE GENOMIC DNA AND A BABOON ENDOGENOUS VIRUS PROBE	177
CHAPTER 7 :	GENERAL DISCUSSION	200
REFERENCES		212

ACKNOWLEDGEMENTS

I wish to thank my supervisors, Professors Neil Gorman and David Onions, for their support, guidance and enthusiasm during the past three years. Neil Gorman inspired me with confidence in the project and tutored me skillfully through the initial Wellcome Trust interview. More recently he has been unstinting in his efforts to help me finish writing the thesis on time. Throughout the past three years he has been most generous with his time and has always provided me with friendly advice and reassurance. He fully understands and affirms my career objectives. David Onions has inspired me with his enthusiasm and energy. He created the environment which allowed me to make the formidable transition from clinician to molecular biologist. He has been equally generous with his time and has been very tolerant of my frequent "corridor consultations". I consider myself very lucky to have had such understanding and competent supervisors.

Simon Hettle and Carolyn Johnston had an enormous impact on my scientific development, since they effectively trained me in the basic techniques of molecular biology during my first few months at Glasgow. I shall never forget their patience, generosity and friendship. Marcello Riggio and Lesley Nicolson trained me in DNA sequencing and bravely tolerated my over-enthusiasm. Gillian Lees and Gerry Mosson taught me the rudiments of cell culture and helped with the reverse transcriptase assays. Rose Ann Padua started the ball rolling by providing the seven recombinant bacteriophage lambda clones. Ruth Jarrett, Norman Spibey, Oswald Jarrett and Heather Cavanagh gave unstintingly of their time, to provide much helpful advice. Other members of the Department of Veterinary Pathology, including Alice Gallacher, Sarah Gledhill, Duncan Clark and Matthew Golder were a source of joy, friendship, assistance and helpful advice. Jim Anderson and Janice Lloyd were of enormous assitance in the final preparation of the thesis; staying up until the wee small hours of the morning on several occasions.

I am most grateful to the Wellcome Trust for providing financial support in the form of a Research Training Scholarship. I thank Professor W.F.H. Jarrett, head of the Department of Veterinary Pathology, for allowing me to carry out this work in his department. Finally, I wish to thank Alan May and John Fuller for their professional assistance in reproducing (respectively) the photographs and diagrams which form part of this thesis.

DECLARATION

The work described in this thesis was carried out by me, with the following exceptions. The FeLV ELISA and subgroup determination was carried out by Mr M. Golder in the laboratory of Professor O. Jarrett. The RNA-directed DNA polymerase assays were carried out by Mr G. Mosson in the laboratory of Dr G. Lees.

Richard A. Squires, November 1990

ABBREVIATIONS

A	adenine or adenosine
AIDS	acquired immunodeficiency syndrome
ALL	acute lymphoblastic leukaemia
ALV	avian leukosis virus
5-AZA	5-azacytidine
BaEV	baboon endogenous virus
bp.	base pair(s)
C	cytosine or cytidine
CA	major structural capsid protein
CERV	canine endogenous retrovirus
CLL	chronic lymphocytic leukaemia
cm	centimetre
c-*onc*	cellular proto-oncogene
d	(as a prefix) deoxy-
dNTP	one or more of the 4 deoxynucleoside triphosphates
D	Dalton
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
ds	double-stranded
EM	electron microscopy
env	envelope protein(s)-encoding gene
ER	endoplasmic reticulum
ERV	endogenous retrovirus
FeLV	feline leukaemia virus
FMuLV	Friend murine leukaemia virus
g.	gramme(s)
G	guanine or guanosine the gravitational field strength of the earth
gag	group-specific antigen(s)-encoding gene

HERV	human endogenous retrovirus
HIV	human immunodeficiency virus
HTLV-1	human T-lymphotropic virus type 1
IAP	intracisternal A-particle
IdU	5-iododeoxyuridine
IN	endonuclease/integrase
kb.	kilobase(s)
kbp.	kilobase pair(s)
kD	kilo-Dalton
LINE	long interspersed nuclear element (long, interspersed, repeated DNA elements)
LMP	low melting point
LTR	long terminal repeat
m^7Gppp	7-methyl guanosine triphosphate
MA	envelope-associated matrix protein
MCF	mink cell focus-forming
MDCK	Madin-Darby canine kidney (cells)
Mg^{++}	ionised magnesium
ml.	millilitre(s)
um	micrometre(s)
uM	micromolar
MMTV	mouse mammary tumour virus
Mn^{++}	ionised manganese
mRNA	messenger ribonucleic acid
MuLV	murine leukaemia virus
NC	RNA-binding protein
nm	nanometre(s)
-OH	hydroxyl
^{32}P	phosphorus-32

PBS	primer binding site
pol	polymerase-encoding gene
poly-A	poly-adenosine
PR	protease
R	terminal redundant retroviral sequences
RAV	Rous-associated virus
RE	restriction endonuclease
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal ribonucleic acid
RT	reverse transcriptase (RNA-dependent DNA polymerase)
S	sedimentation co-efficient
^{35}S	sulphur-35
SFFV	spleen focus forming virus
SINE	short interspersed nuclear element (short, interspersed, repeated DNA elements)
spp.	species
ss	single-stranded
SU	surface envelope glycoprotein
T	thymine or thymidine
T-cell	thymus-derived lymphocyte
TM	transmembrane envelope protein
tRNA	transfer ribonucleic acid
TVT	transmissible venereal tumour
U	uracil or uridine
U_3	unique 3' retroviral sequence
U_5	unique 5' retroviral sequence
UV	ultraviolet
v-*onc*	viral oncogene

LIST OF FIGURES

page number

CHAPTER 1

Fig. 1.1a:	Phylogeny of vertebrate retroviruses based on a comparison of reverse transcriptase sequences.	6
Fig.1.1b:	Phylogeny of vertebrate retroviruses based on a comparison of endonuclease sequences.	7
Fig. 1.2:	Genomic organization of a typical oncoviral provirus (AKV).	10
Fig. 1.3:	The retroviral lifecycle.	11
Fig. 1.4:	Reverse transcription of the retroviral ss RNA genome to produce ds DNA.	13-14

CHAPTER 2

Fig. 2.1:	pBR322	44
Fig. 2.2:	pBluescript KS (+)	45
Fig. 2.3:	Bacteriophage Lambda 2001	47
Fig. 2.4:	Bacteriophage Lambda GT10	48

CHAPTER 3

Fig. 3.1:	Sequencing autoradiographs: typical results achieved with acid phenol miniprep. DNA templates.	84
Fig. 3.2:	Restriction analysis of the cloned FMuLV insert provided by Dr Padua: a photograph of the ethidium bromide-stained gel and a diagram of the deduced restriction map.	86
Fig. 3.3:	Two photographs of restriction digestions of L1 to L7 with a variety of single REs, to show that L4, L6 and L7 are identical.	90

page number

Fig. 3.4:	A photograph of an ethidium bromide-stained gel including *Xba*I and *Bam*HI digestions of L1 to L5. Below, the corresponding autoradiograph showing strong hybridisation of the FMuLV probe to sub-fragments of L3 and L4.	91
Fig. 3.5:	Double RE digestions of L4 alongside the corresponding autoradiograph, to show hybridising sub-fragments. Below, the deduced restriction map of the L4 insert, with retrovirus-like sequences denoted.	93 - 94
Fig. 3.6:	Double RE digestions of L3 alongside the corresponding autoradiograph, to show hybridising sub-fragments. Below, the deduced restriction map of the L3 insert, with retrovirus-like sequences denoted.	95 - 96
Fig. 3.7:	Sub-cloning and sequencing strategy for L4.	98
Fig. 3.8:	Sub-cloning and sequencing strategy for L3.	99
Fig. 3.9a:	L4 sequence data, showing important landmarks in the sequence.	101
Fig. 3.9b:	Overview diagram of the defective L4 proviral element.	102
Fig. 3.10:	A dotplot of L4 versus the AKV complete genome.	103
Fig. 3.11:	Defectiveness of L4: a diagram to show that protein homology of L4 to AKV is shared among the 3 plus strand L4 reading frames. Frequent stop codons are present in the 3 plus strand L4 reading frames.	104
Fig. 3.12a:	L3 sequence data, showing important landmarks in the sequence.	106 - 107
Fig. 3.12b:	Overview diagram of the defective L3 proviral element.	108

		page number
Fig. 3.13:	Aligned L3 LTR sequences.	110
Fig. 3.14:	Alignment of the nine, highly-conserved L3 tandem repeats.	111
Fig. 3.15a:	A dotplot of L3 versus AKV to show that a sequence closely related to the L3 tandem repeat consensus sequence (minisatellite) is present at single copy in AKV *gag*.	113
Fig. 3.15b:	A dotplot of L3 versus L3 to show the two sets of tandem repeat elements.	114
Fig. 3.15c:	A dotplot of L3 versus the AKV complete genome.	115
Fig. 3.16:	Defectiveness of L3: a diagram to show that protein homology of L3 to AKV is shared among the 3 plus strand reading frames. Frequent stop codons are present in all 3 plus strand L3 reading frames.	117
Fig. 3.17a:	A dotplot of L3 versus L4 to show regions of high homology and deletions, one relative to the other.	118
Fig. 3.17b:	Summary diagram to show the structures of L4 and L3 compared with AKV.	119
Fig. 3.18:	Comparison of retroposon sequences at a locus in L3 and L4.	120

CHAPTER 4

Fig. 4.1:	Diagram of L4, to show the derivation of the *Hinf*I-*Pst*I and *Spe*I-*Eco*RV probes.	131
Fig. 4.2:	Hybridisation of canine genomic DNA to the L4 2.5 kbp. *Sst*I fragment.	133
Fig. 4.3:	18 canine genomic DNA samples, digested with *Pst*I and probed with the *Pst*I-*Hinf*I fragment.	135
Fig. 4.4:	18 canine genomic DNA samples, digested with *Eco*RV and probed with the *Spe*I-*Eco* RV fragment.	137

page number

Fig. 4.5:	Double digestion of canine genomic DNAs with *Pst*I/*Hinf*I and *Spe*I/*Eco*RV. probed with the corresponding L4 probes.	139
Fig. 4.6:	Hybridisation of FMuLV to canine genomic DNA samples.	140
Fig. 4.7:	Hybridisation of HTLV-1 to canine genomic DNA samples.	141
Fig. 4.8:	Hybridisation of BaEV to canine genomic DNA samples.	143 - 144
Fig. 4.A1.1	Digestion of genomic DNAs from several wild canid species with *Pst*I. Probed with the L4 *Pst*I-*Hinf*I probe.	151
Fig. 4.A1.2	Digestion of genomic DNAs from several wild canid species with *Eco*RV. Probed with the *Spe*I-*Eco*RV probe.	152
Fig. 4.A4.1	Hybridisation of a *c-myc* probe to *Eco*RI-digested canine genomic DNAs.	160
CHAPTER 5		
Fig. 5.1:	Diagram showing the new *Pvu*II/*Sst*I *env*-LTR probe derived from L4.	169
Fig. 5.2:	Autoradiograph showing multiple bands representing short L4 *pol*-related RNA molecules or DNA contamination.	171
Fig. 5.3	Repetition and extension of the work described in Fig. 5.2.	172
CHAPTER 6		
Fig. 6.1:	A photograph of an ethidium bromide-stained gel of double RE digestions of the Lambda GT10 construct which contained 6.7 kbp. of canine genomic DNA derived from normal kidney. Below, the corresponding autoradiograph, showing a BaEV-hybridising *Eco*RI/*Bam*HI 2.8 kbp. fragment.	181

		page number
Fig. 6.2:	A restriction map of the 6.7 kbp. canine genomic DNA fragment with the BaEV-hybridising 2.8 kbp. *Eco*RI/*Bam*HI and 0.7 kbp. *Xba*I sub-fragments denoted.	182
Fig. 6.3:	Sub-cloning and sequencing strategy for the 2.8 kbp. *Eco*RI/*Bam*HI sub-fragment of the Lambda GT10 insert.	184
Fig. 6.4:	NK986 sequence data.	185
Fig. 6.5:	Double RE digestions of the new BaEV construct, cloned in pBluescript KS (+).	187 - 189
Fig. 6.6:	Restriction map of our BaEV M7 clone compared with that reported by Tamura (1983).	190
Fig. 6.7:	Back probing the BaEV clone with NK702: an ethidium bromide-stained gel and the corresponding autoradiograph.	192 - 193
Fig. 6.8:	Sub-cloning and sequencing strategy for the 0.9 kbp. *Xba*I-*Bam*HI fragment in our BaEV clone. Summary diagram to show the insertion of canine DNA and reversed *gag* in the BaEV LTR.	194
Fig. 6.9:	Sequence data from our BaEV M7 clone.	195

LIST OF TABLES

		page number
Table 3.1:	shows phage titres (pfu/ul) at each stage of plaque purification and the final yields of phage DNA, as calculated from optical densitometry.	88
Table 4.1:	gives summary data of all genomic DNA samples used to carry out the work described in Chapter 4.	129
Table 4.A1.1:	Details of the geographic origins of the wild canid DNAs studied in Appendix 1 of Chapter 4.	150
Table 5.1:	summarises details of all RNA samples used to carry out the work described in Chapter 5.	167

SUMMARY

Dogs suffer from a variety of neoplastic, immunological, degenerative and proliferative disorders for which a retroviral aetiology is suspected. Despite decades of effort, the significance of retroviruses in canine disease remains obscure. This murky subject has been clouded further by a dearth of information concerning canine endogenous retroviral complement and expression. Previous reports have described retroviral expression in both normal and neoplastic canine tissues; but have not defined the origin of the expression. It was considered of interest to determine whether such expression reflected the presence of endogenous retroviral elements in canine genomic DNA. If such elements were found, their aetiological significance in canine diseases would be investigated.

At the start of this project, cloned canine genomic DNA from the lymphoma cell line 3132, which has been reported to express a retrovirus, was probed with a murine C-type retroviral sequence. Hybridising canine DNA fragments were sub-cloned and sequenced. Two related, but distinct, highly defective retroviral proviruses were identified and characterised by DNA sequence data analysis. Highest homology was found to the murine endogenous retrovirus, AKV.

Having identified retroviral elements in one canine cell line, it was considered of interest to determine whether the presence of such elements was a universal feature of canine DNA. Canine genomic DNAs from a variety of normal and neoplastic tissues were probed for the presence of retroviral sequences. Well-established murine and primate retroviral probes were used; as well as two *pol* probes derived from one of the newly-sequenced canine proviral elements. Numerous retrovirus-like elements were found in all samples of canine DNA examined. The arrangement of proviral elements was identical in all dogs, irrespective of whether the tissues were normal or neoplastic. Related canid species had an arrangement of proviral elements which could not be distinguished from that of the domestic dog, indicating that these are ancient sequences. There was evidence that some of these proviral sequences have undergone a DNA amplification process.

A preliminary investigation into the RNA expression of canine endogenous retroviral elements was carried out. Total RNA was extracted from a variety of

normal and neoplastic tissues and probed for the presence of retroviral *pol* sequences using probes derived from a canine proviral element. Canine RNA samples were also probed with BaEV, FMuLV, HTLV-1 and MMTV *gag-pol*. No hybridisation was found to the murine or primate retroviral probes. The canine *pol* probes detected a consistent pattern of short (<1 kb.) hybridising molecules in a small minority of the RNA samples. Both normal and neoplastic tissues were represented in the group of samples which hybridised to the canine *pol* probes. Attempts to induce detectable endogenous retroviral expression in 3132, A72-F and MDCK canine cells by growth in the presence of 5-iododeoxyuridine and 5-azacytidine were unsuccessful.

Canine DNA was found to contain an element which hybridised strongly to a BaEV M7 probe. It was suspected that this element might represent a relatively intact endogenous provirus unrelated to those previously characterised. Consequently, a genomic DNA library was prepared from normal canine kidney and this element was selected from it. When the BaEV-hybridising canine DNA sub-fragment was sequenced, it was found to be non-retroviral. Subsequent back-probing of the BaEV M7 clone from which the probe had been prepared, identified an identical non-viral sequence upstream of the PBS. It is proposed that the BaEV clone under study acquired this segment of canine DNA at some time in the past, during *in vitro* culture in canine cells.

The significance of all of these experimental results is discussed in Chapter 7. Sequences which could explain the production of an infectious retrovirus by cell line 3132 were not found. This certainly does not preclude the existence of such sequences among the many proviral elements which are present in canine genomic DNA. It appears that many of the endogenous retroviral sequences studied in this project have undergone an intriguing amplification process which may involve an amplified unit larger than the provirus itself. However, there is no evidence that this amplification process is continuing apace.

Chapter 1

INTRODUCTION

1.1 RETROVIRUSES

1.1.1 OVERVIEW

1.1.2 TAXONOMY

1.1.3 GENOME ORGANIZATION

1.1.4 LIFE CYCLE

1.1.5 PATHOGENESIS OF RETROVIRAL DISEASES

1.2 ENDOGENOUS RETROVIRUSES AND RETROVIRUS-LIKE ELEMENTS

1.2.1 THE VIRAL SUPERFAMILY OF RETROPOSONS

1.2.2 POSTULATED ORIGINS OF RETROVIRUSES

1.2.3 AVIAN, MURINE & FELINE ENDOGENOUS RETROVIRUSES

1.2.4 HUMAN ENDOGENOUS RETROVIRUSES

1.2.5 EXPRESSION OF ENDOGENOUS RETROVIRUSES

1.2.6 RELEVANCE IN HEALTH AND DISEASE OF THE HOST

1.3 EVIDENCE FOR THE EXISTENCE OF A CANINE RETROVIRUS

1.3.1 CANINE DISEASES WITH A SUSPECTED RETROVIRAL AETIOLOGY

1.3.2 EXPERIMENTAL EVIDENCE FOR THE EXISTENCE OF A CANINE RETROVIRUS

1.4 AIMS OF THIS PROJECT

1.1 RETROVIRUSES

1.1.1 OVERVIEW

Retroviruses are single-stranded (ss) RNA viruses which, upon infection of a cell, transcribe their RNA genome into double-stranded (ds) DNA. This ds DNA can then become stably-integrated into cellular chromosomal DNA as a provirus. This *retro*grade flow of genetic information gives the family its name.

Each spherical retroviral particle is 75-150nm in diameter and consists of a lipoprotein envelope surrounding an icosahedral protein core. The envelope is derived from cellular unit membrane and contains virally-encoded glycoproteins which may appear as protruding spikes when viewed by electron microscopy. The viral core is composed of several low molecular weight structural proteins. Within the core are two identical copies of the 4 to 10 kb. ss RNA genome. The two copies are non-covalently linked to form a dimer. A single tRNA molecule is attached by base-pairing close to the 5' end of each genomic RNA molecule. Also present within the viral core is the viral enzyme RNA-directed DNA polymerase or reverse transcriptase (RT) which catalyses transcription of viral RNA into DNA. Progeny viral genomes and viral mRNA are formed by transcription of proviral DNA by cellular RNA polymerase II. Cellular mechanisms are also used to translate viral mRNAs into protein. New viral particles are released by budding at cell membrane surfaces, rather than by cell lysis. In most cases, the host cell is not destroyed by retroviral infection.

Retroviruses are very widespread in nature. They infect an enormous variety of animal species by diverse routes. Transmission may be horizontal (by aerosol, contact, biting or scratching) or alternatively, virus particles may be transmitted vertically from parent to offspring, usually in milk or trans-placentally. Because retroviruses are enveloped, fragile viruses, they are susceptible to adverse environmental influences. Transmission therefore requires intimate contact between susceptible individuals. Uniquely among animal viruses, retroviruses may infect oocytes or early embryos and be transmitted from generation to generation in the germ line by classical Mendelian inheritance of chromosomally-integrated endogenous DNA proviruses. These endogenous proviruses are present in normal,

healthy members of the host species. Most endogenous proviruses are thought to be transcriptionally-inactive and non-pathogenic.

Retroviruses have been shown to cause a profusion of different diseases, including neoplastic, degenerative, proliferative and immunological disorders. Almost all families of animal DNA virus can cause neoplasia, but retroviruses are alone among the RNA viruses in manifesting oncogenicity. Perhaps this is because retroviruses are the only RNA viruses which replicate their genomes through a DNA intermediate. Intimate, persistent association between chromosomally-integrated viral DNA and cellular sequences may be necessary to effect neoplastic transformation. Certainly, interaction of retroviral proviruses with certain regulatory cellular genes (proto-oncogenes) has been shown to induce neoplasia (*q.v.*).

1.1.2 TAXONOMY

Retroviral taxonomy has been reviewed (Teich, 1982, 1985; Varmus and Brown, 1989; Doolittle *et al.*, 1989, 1990). Retroviruses are defined by their replication strategy: the family Retroviridae includes all viruses which replicate a ss RNA genome through a ds DNA intermediate by use of the viral enzyme RT. Viruses from diverse host species which use this system of replication have many other features in common, to be discussed later. Over the years, various systems of retroviral classification have been devised, dictated by the available technologies of the day. Originally retroviruses were subdivided according to morphological characteristics observed by electron microscopy (for reviews, see Bernhard, 1958, 1960). In this system, virus particles are categorised into one of 4 types, designated A to D:

Type A particles are relatively small (60-90 nm diameter), non-infectious, intracellular forms. They may be intracisternal or intracytoplasmic. The significance of intracisternal A particles is unknown. Intracytoplasmic A particles are the cores of type B and type D retroviruses. Type A particles have an electron-lucent core surrounded by a double shell.

Type B particles consist of an intracytoplasmic A particle surrounded by a prominently-spiked envelope (Bernhard and Bauer, 1955). The overall diameter is 125-130 nm. Mature type B particles contain an electron-dense, eccentrically-

located core. Mouse mammary tumour virus (MMTV) is the prototypic member of this group.

Type C particles constitute the great majority of retroviruses characterised to date. Most oncogenic retroviruses are type C particles. These particles are formed entirely at the cell membrane with no cytoplasmic precursor forms. Immediately inside the cell membrane, an electron-dense crescent forms, which subsequently becomes the viral core. Short spikes may be evident protruding from the overlying cell membrane. As budding proceeds, the crescent becomes a sphere, with an electron-lucent core. The cell membane surrounds and eventually pinches off the nascent particle. Mature type C particles are 80-110nm in diameter with a central, electron-dense core.

Type D particles are 100-120 nm in diameter. Like type B particles, they are formed by budding of type A intracytoplasmic particles. Their characteristic feature is a distinctly bar-shaped core. Their envelopes are smooth. Mason-Pfizer monkey virus is the best-characterised type D virus.

More recently, retroviruses have been classified according to their biological properties into 3 subfamilies: Oncovirinae, Lentivirinae and Spumavirinae (Matthews, 1979). The largest and most studied subfamily, Oncovirinae, includes all oncogenic and closely-related non-oncogenic viruses. As well as neoplasia, viruses in this group can cause degenerative, immunological and non-neoplastic proliferative diseases. Lentiviruses (or slow viruses), as their name would suggest, cause slowly-progressive diseases. These are usually degenerative and ultimately fatal diseases, often with long latent periods. The human immunodeficiency viruses (HIV-1 and HIV-2) which cause acquired immunodeficiency syndrome (AIDS) are important and increasingly well-characterised members of this group. Spumaviruses are non-pathogenic retroviruses, which establish persistent infections in many host species. The name derives from their tendency to produce a foamy cytopathic effect in cells grown in tissue culture. This is caused by marked vacuolation of host cell cytoplasm.

Lentiviruses and Spumaviruses were described after the original morphological studies and their subsequent inclusion in that classification system would be rather artificial. HIV-1, the best characterised lentivirus is 100-150 nm in diameter, with a bar or cone-shaped core. While the viral particle is forming, the

envelope has short spikes; but these are no longer apparent after the completion of budding. Spumaviruses are 100-140 nm in diameter. A 35-50 nm diameter core is present, with an electron-lucent centre. The envelope has long, prominent spikes.

The most recent system of retroviral classification is based upon computer-assisted comparisons of the amino acid sequences of relatively conserved retroviral enzymes, such as RT (McClure *et al.*, 1988; Doolittle *et al.*, 1989, 1990). Using this system to deduce phylogeny, it is possible to categorise the most studied retroviruses into at least four groups. These are: (i) a group of lentiviruses, (ii) a group of T-cell leukaemia viruses, (iii) a morphologically varied group, including endogenous and exogenous viruses as well as intracisternal A particles and (iv) another heterogeneous group which includes various leukaemia viruses, spumaviruses and endogenous retroviruses (Figs. 1.1a and 1.1b). This last group includes feline leukaemia virus, baboon endogenous virus, avian reticuloendotheliosis virus and Moloney murine leukaemia virus.

1.1.3 GENOME ORGANIZATION

The structure of retroviral genomes has been reviewed (Coffin; 1982a, 1985). All replication-competent retroviruses have a broadly similar genomic organization. Each of the two dimeric components of the positive-sense RNA genome is post-transcriptionally modified, so that it closely resembles a eukaryotic mRNA molecule; with a 5' methyl cap (m^7Gppp), a 3' poly-adenosine (poly-A) tail and several interspersed methylated A residues. Immediately internal to these 5' and 3' termini are redundant sequences (R), 20-80 bases in length. These are involved in reverse transcription (see next section). Internal to the R sequences, the unique region of the RNA genome (from 5' to 3') consists of U_5, primer binding site (PBS), leader sequence/packaging signal, *gag*, *pol*, *env*, polypurine tract and U_3. Some lentiviruses, spumaviruses and type C oncoviruses have several additional regulatory genes between *pol* and the 3' end of the genome that are expressed using complex, spliced mRNAs (for reviews, see Varmus, 1988; Varmus and Brown, 1989). Approximately eighteen bases at the 3' terminus of a specific tRNA molecule are attached to the PBS by precise base-pairing. This unit functions as a primer for the initiation of DNA synthesis. Different retroviral sub-groups utilize different tRNA molecules as primers.

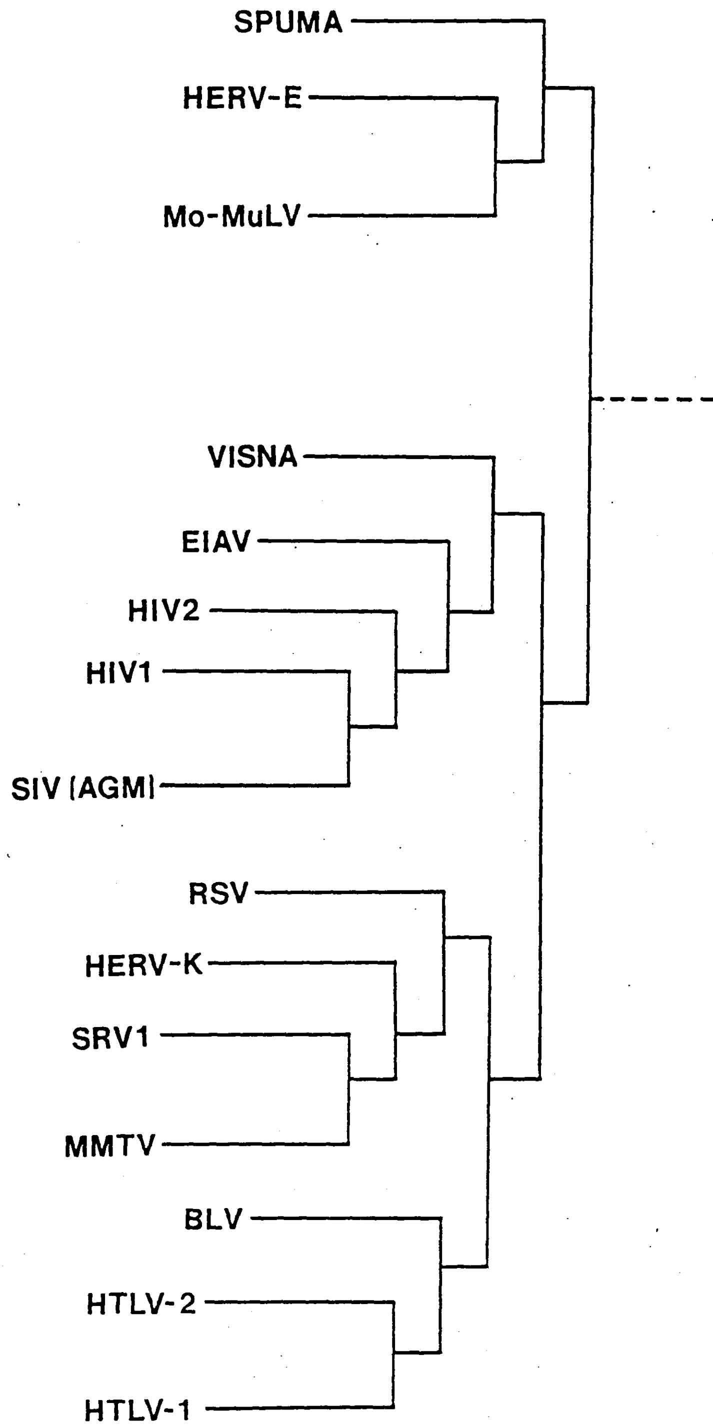


Fig. 1.1a: Phylogeny of vertebrate retroviruses based on a comparison of reverse transcriptase sequences (Adapted from Doolittle *et al.*, 1990).

Abbreviations: SPUMA, human spumavirus; HERV-E, human endogenous retrovirus E; Mo-MuLV, Moloney murine leukaemia virus; VISNA, visna lentivirus; EIAV, equine infectious anaemia virus; HIV 1 & 2, human immunodeficiency viruses types 1 & 2; SIV (AGM), simian immunodeficiency virus, African green monkey; RSV, Rous sarcoma virus; HERV-K, human endogenous retrovirus K; SRV1, simian type D acquired immunodeficiency syndrome retrovirus; MMTV, mouse mammary tumour virus; BLV, bovine leukaemia virus; HTLV-1 & -2, human T-cell leukaemia viruses types 1 & 2.

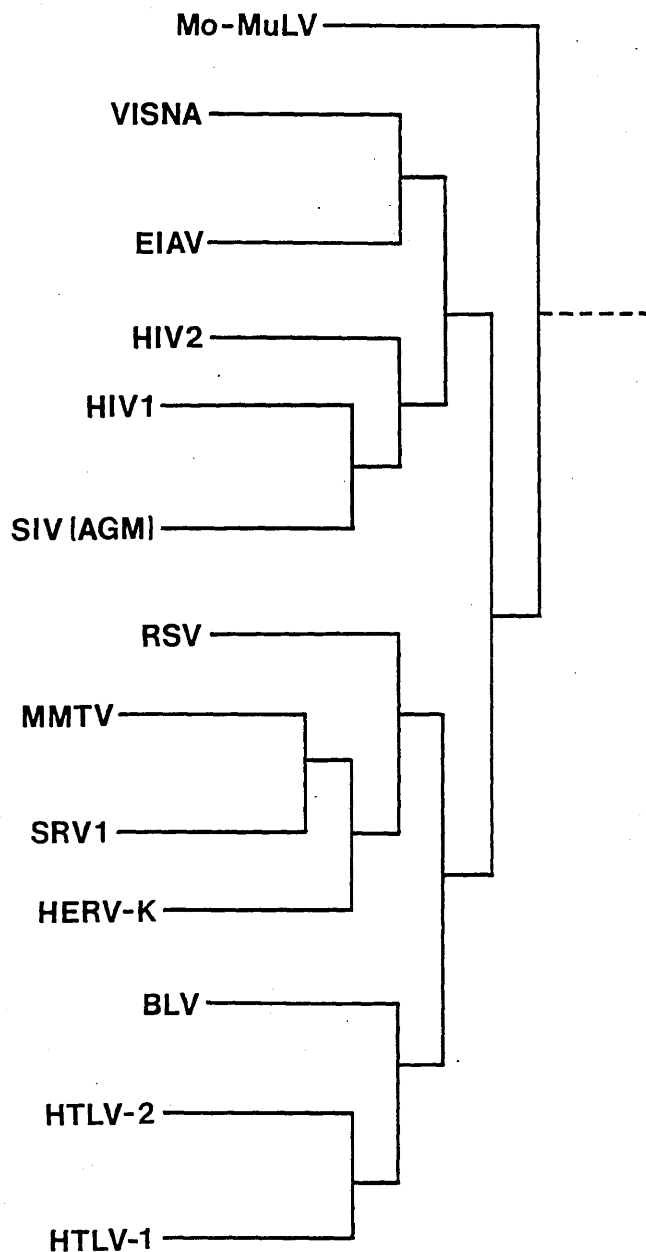


Fig.1.1b: Phylogeny of vertebrate retroviruses based on a comparison of endonuclease sequences (Adapted from Doolittle *et al.*, 1990).

Abbreviations: Mo-MuLV, Moloney murine leukaemia virus; VISNA, visna lentivirus; EIAV, equine infectious anaemia virus; HIV 1 & 2, human immunodeficiency viruses types 1 & 2; SIV (AGM), simian immunodeficiency virus, African green monkey; RSV, Rous sarcoma virus; MMTV, mouse mammary tumour virus; SRV1, simian type D acquired immunodeficiency syndrome retrovirus; HERV-K, human endogenous retrovirus K; BLV, bovine leukaemia virus; HTLV-1 & -2, human T-cell leukaemia viruses types 1 & 2.

The three major protein-coding genes are *gag*, *pol* and *env*. Their products will be described using the two-letter system of nomenclature proposed by Leis *et al.* (1988). The *gag* gene (about 2kb) codes for 3 to 5 internal structural proteins. The initial translation product of this gene is a precursor polyprotein. This is subsequently cleaved by viral protease to produce the mature proteins, including the major structural capsid protein (CA), an RNA-binding protein (NC) and an envelope-associated matrix protein (MA). Generally, *pol* (about 3kb) codes for protease (PR), reverse transcriptase (RT) and endonuclease/integrase (IN); although PR is part of *gag* in some retroviruses. RT can function as RNA-directed DNA polymerase, DNA-directed DNA polymerase and RNase H; the last of which degrades RNA in a DNA-RNA hybrid (Mölling *et al.*, 1971; Temin and Baltimore, 1972). The separate polymerase and RNase H functional domains of the RT molecule are joined by a non-enzymatic region, termed a tether (Johnson *et al.*, 1986). The *env* gene codes for a polyprotein precursor which is subsequently cleaved to produce a large surface envelope protein (SU) and a smaller transmembrane protein (TM). SU (and in some viruses, TM) is glycosylated after translation. TM is either disulphide-linked or non-covalently bonded to SU.

Unintegrated linear, and integrated proviral DNA differ from retroviral genomic RNA, in that U_5 and U_3 are duplicated so that both ends of the provirus have the structure U_3-R-U_5. These identical termini are termed Long Terminal Repeats (LTRs) and they encode the promoters of viral gene expression. TATA and CCAAT boxes are present in U_3, approximately 25 and 80 bp upstream (respectively) of the 5' cap site at the junction of U_3 and R. Enhancer elements are present further upstream, often present as direct repeats 70-100 bp long. These enhancers and promoters of mRNA transcription are more active in the 5' LTR than in its identical 3' counterpart (Herman and Coffin, 1986). It has been suggested that promoter activity of the 3' LTR is suppressed by the arrival of the transcription apparatus from upstream ("promoter occlusion"; Cullen *et al.*, 1984). In the 3' LTR, sequences in R direct mRNA cleavage and polyadenylation. *gag* and *pol* proteins are translated from a full genome-length viral mRNA, *pol* by occasional read-through of the termination signal at the 3' end of *gag*. If necessary, frame-shifting between *gag* and *pol* can occur by ribosomal slippage (for a review, see Varmus and Brown, 1989).

env proteins are translated from a spliced, sub-genomic mRNA. The splice donor site is 5' to, or just within the *gag* gene; the acceptor site is a variable distance upstream of the initiation codon of *env*, usually within *pol*. Thus, the *pol* and *env* genes generally overlap, although they are translated in different reading frames. Certain lentiviruses, including the immunodeficiency viruses, are exceptions to this rule. They have regulatory genes which separate *pol* and *env*. Figure 1.2 shows the genomic organization of a typical type-C retroviral provirus, AKV.

1.1.4 LIFE CYCLE

Replication of retroviruses has been comprehensively reviewed (Varmus and Swanstrom, 1982, 1985; Swanstrom and Vogt, 1990) and is represented in Fig. 1.3. Extracellular retroviral particles become attached to target cells by interaction between viral SU and a specific host cell membrane receptor. Cells which do not express the particular receptor, or in which the receptors are occupied, are much less readily infected by virus. Thus HIV, which binds the CD4 antigen present on human T-helper lymphocytes, will readily infect these cells (Dalgleish *et al.*, 1984). Other human cells, normally resistant to infection, may be rendered susceptible to HIV infection by transfection of CD4-encoding DNA and expression of CD4 antigen (Maddon *et al.*, 1986). Failure of retrovirus to infect cells with occupied receptors is the basis of the phenomenon of viral interference: cells infected by a retrovirus cannot usually be superinfected with another virus of the same subgroup. This is because receptors on the surface of an infected cell and within its cytoplasm are already occupied by retroviral products. After the specific interaction between SU and the cellular receptor, fusion of cell and viral unit membranes occurs, allowing the viral core particle to enter the cell. This may occur at the surface of the plasma membrane (as with HIV; Stein *et al.* 1987) or, following endocytosis, within a lysosome (Marsh, 1984). Previously, it was thought that the acidic environment of the lysosome was essential for membrane fusion. Recent work suggests that most mammalian retroviruses infect cells in a pH-independent manner (McClure *et al.*, 1990). Membrane fusion is thought to be mediated by a hydrophobic component of TM. In other enveloped RNA virus families, TM analogues have this function (White, *et al.*, 1983).

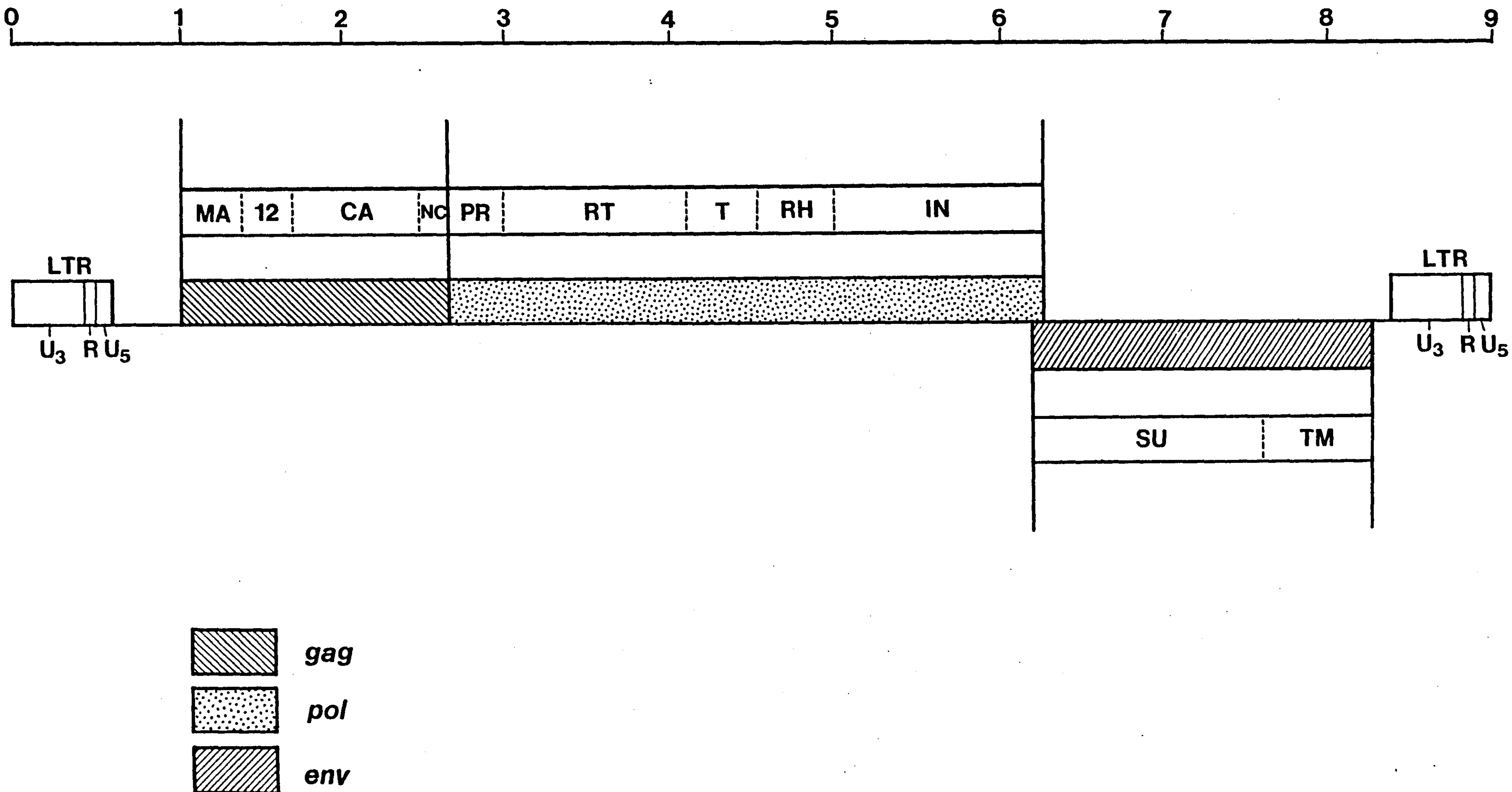


Fig. 1.2: Genomic organization of a typical oncoviral provirus (AKV).
Abbreviations: LTR, long terminal repeat; MA, envelope-associated matrix protein; 12, *gag* protein of uncertain function; CA, major structural capsid protein; NC, nucleocapsid RNA-binding protein; PR, protease; RT, reverse transcriptase; T, tether; RH, ribonuclease H; IN, integrase; SU, major surface glycoprotein; TM, transmembrane envelope protein.

Retrovirus

RNA

Viral RNA

Reverse transcriptase

Plasma membrane

Viral DNA

Provirus

Nucleus

Integration

Cell DNA

Genomic RNA

mRNA

Viral protein

Viral glycoprotein

Virus released by budding

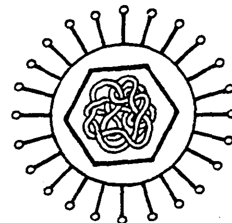


Fig. 1.3: The retroviral lifecycle.

Virus enters the cell by specific interaction of its surface glycoprotein with a cell membrane receptor. Once within the cytoplasm, the virus uncoats and its RNA genome is reverse transcribed into linear ds DNA. This ds DNA is integrated into the host cell genome, via a two-LTR circular intermediate (as shown); or, alternatively, via a linear intermediate with its ends approximated. Once integrated, the provirus can be transcribed to produce viral mRNAs and new viral genomes. Cellular machinery is used for synthesis of viral proteins. C-type retroviral particles are assembled at the cell membrane and released by budding.

Once inside the cell, the core particle uncoats and the ss RNA genome is ready for reverse transcription into ds DNA. The process by which this occurs is intricate (Fig. 1.4; Gilboa *et al.*, 1979). DNA synthesis begins at the PBS, using the bound tRNA molecule as a primer, and proceeds the short distance to the 5' end of the RNA genome. Once RT reaches the 5' end of the RNA and runs out of template, it switches to its RNase H function and removes the RNA in the dna-rna hybrid that it has just formed. Thus the 5' end of the plus strand RNA genome is removed to the level of the PBS, leaving minus strand ss DNA. At the free (3') end of this ss DNA is the R sequence, which is complementary to the R sequence at the 3' end of the RNA genome. Base-pairing between the two complementary R sequences occurs and since a template/primer is thus once again formed, RT continues to synthesise minus strand DNA, copying the whole of the remaining RNA genome up to and including the PBS. At this point RT once again runs out of template; genomic RNA 5' to the PBS having been removed by RNase H. The next step is to synthesise a second, plus sense DNA strand to complete the formation of ds DNA. RT makes a specific nick in the genomic RNA immediately 5' to U_3, in the polypurine tract, and using its RNase H function degrades genomic U_3-R. Using the nascent minus strand DNA as template and the 3'-OH terminal of genomic RNA at the nick as primer, plus strand DNA synthesis begins and continues up to and including part of the tRNA primer. All remaining RNA (transfer and genomic) is removed by RNase H. The plus strand DNA copy of the primer can then base-pair with the complementary minus strand copy of the PBS and DNA synthesis can continue on both strands, 5' to 3', to completion. The final product is the linear double-stranded structure U_3-R-U_5.......U_3-R-U_5. The linear ds DNA is transported to the cell nucleus where some molecules have their ends joined to form covalently closed circles. The mechanism of viral DNA integration is unclear. All integrated proviruses are precisely bounded by LTRs, but the site of chromosomal integration appears to be random (Panganiban, 1985). This does not exclude the possibility that certain regions of the host cell genome may be more or less accessible for integration. Two nucleotides are always removed from each end of the provirus upon integration. Since there are 4-6 bp. direct repeats in flanking cellular sequences at either end of the provirus, it is likely that integration involves a staggered cut in the cellular DNA. Once integrated, the retroviral provirus

Fig. 1.4: Reverse transcription of the single-stranded RNA retroviral genome to produce the proviral double-stranded DNA form. All of the steps are carried out by RT. (Adapted from Watson *et al.*, 1987).

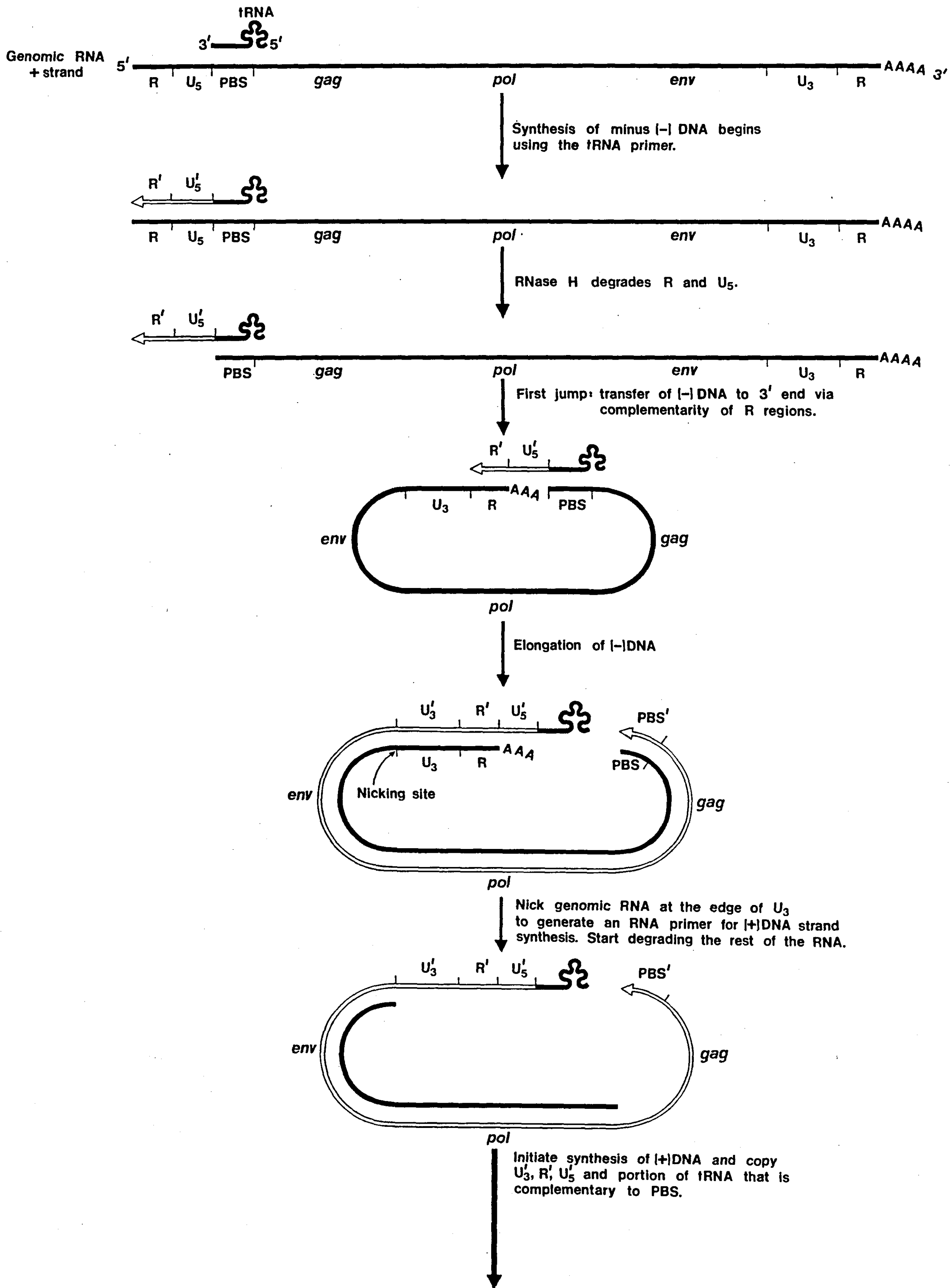
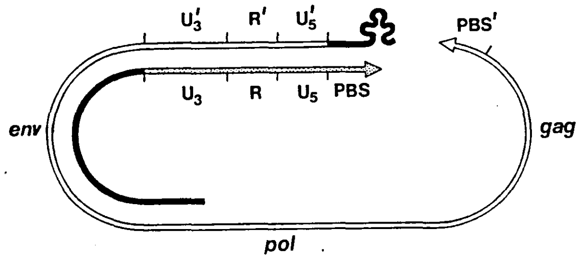
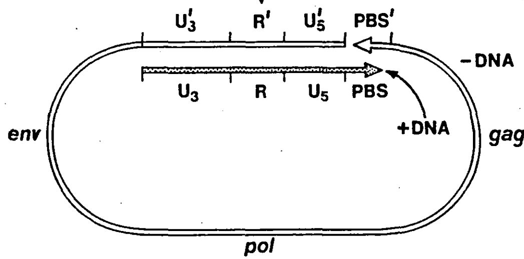


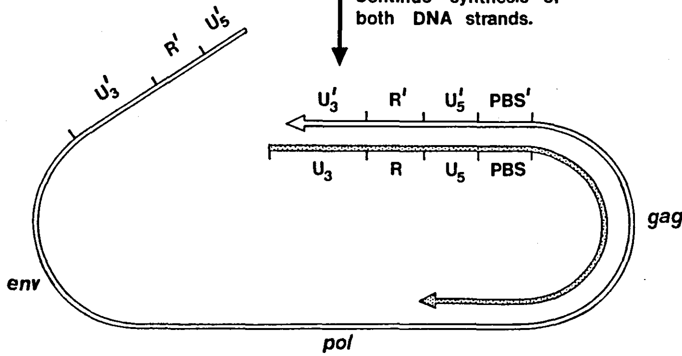
Fig. 1.4: (continued)



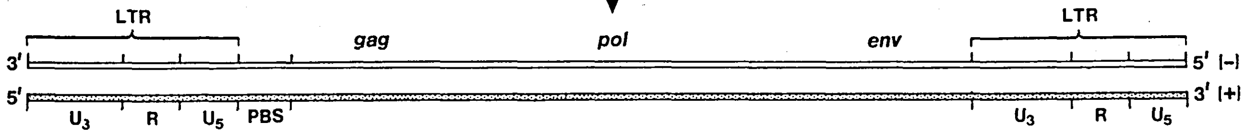
Degrade tRNA and remainder of genome with RNase H. Use complementarity between PBS and PBS' to tranfer (+)DNA to (-)DNA template (second jump).



Continue synthesis of both DNA strands.



Complete synthesis of both DNA strands



becomes a stable part of the genetic complement of the host cell and is inherited by all progeny cells.

Transcription of proviral DNA by cellular RNA polymerase II is regulated according to the metabolic and genetic status of the host cell. This sensitivity to host cell status is likely to be mediated by interactions between labile cellular factors and the enhancer elements in the 5' LTR. A striking example of this form of regulation involves MMTV. Expression of this virus is stimulated by glucocorticoids; the effect being the result of direct interaction of the steroid-receptor complex with the MMTV LTRs (Govidan *et al.*, 1982). More recent dissection of this interaction has demonstrated that it involves a specific enhancer element within the MMTV LTRs with the core sequence 5'-tgttct-3' (Scheidereit and Beato, 1984). RNA transcripts of the integrated provirus have several possible fates. Some are transported to the cytoplasm or cell membrane to serve as new genomes for budding progeny viruses. Other full-length RNAs serve as mRNA for *gag* and *pol* genes. Still others are spliced to serve as mRNAs for *env* or the regulatory genes present in some retroviruses. Retroviral protein synthesis and assembly has been reviewed by Dickson *et al.*, 1982, 1985. The *env* mRNA product has a leader peptide signal enabling its co-translational translocation into endoplasmic reticulum (ER). The covalent bond between TM and SU is cleaved by a cellular protease, but the two molecules remain linked, either non-covalently or by disulphide bridges. Glycosylation of SU and, in lentiviruses, of TM, is accomplished during translocation from the ER to the plasma membrane. The mechanisms which allow co-ordinated association of core proteins, RT, envelope proteins and genomic RNA in the cytoplasm or at the cell membrane to form a budding retroviral particle have not been fully elucidated.

1.1.5 PATHOGENESIS OF RETROVIRAL DISEASES

This topic is the subject of numerous reviews (for examples see Teich *et al.*, 1982, 1985; Ho *et al.*, 1987; Peters, 1989). As previously stated retroviruses cause a bewildering variety of diseases in several categories, including neoplastic, immunologic, degenerative, proliferative and hereditary disorders. An illustrative, rather than encyclopaedic approach must therefore be adopted in this section. At first glance, it may seem surprising that such accomplished parasites as

retroviruses, with their generally limited cytopathicity, should have such devastating effects. As will become apparent, it is the ability of retroviruses to modulate cellular function without killing the host cell which is responsible for many of their pathological effects.

PATHOGENESIS OF RETROVIRUS-INDUCED NEOPLASIA

Integration of a retroviral DNA provirus into a cellular chromosome represents, by definition, a mutational event. Many such insertional mutations will cause no deleterious effects for the cell and produce no phenotypic change. Integration at other sites may produce a lethal effect. If integration confers a selective advantage upon the cell, then its progeny cells will increase in number. An extreme example of this last case is retroviral oncogenesis, in which infected cells grow and replicate autonomously.

Certain cellular genes, termed proto-oncogenes, can be activated by adjacent retroviral provirus integration to increase the probability of neoplastic transformation of the infected cell (for a review see Peters, 1989). Tumours caused in this way often develop only after a long latent period. This is, at least in part, because proviral insertion adjacent to a proto-oncogene is a stochastic event which takes time to occur. Because of the long latent period, viruses employing this oncogenic mechanism are sometimes termed "weakly oncogenic". Less frequently, proto-oncogenes are transduced or captured by retroviruses, which can then, with a helper virus, infect new cells and rapidly cause neoplastic transformation. Such viruses are termed "strongly oncogenic". Oncogenes were first discovered in such "strongly oncogenic", recombinant viruses. It was only later realised that retroviral oncogenes (v-*oncs*) are derived from cellular progenitors (c-*oncs*). Novel proto-oncogenes continue to be identified by the study of clonal proviral integration sites in tumour DNA. Some mechanisms of retroviral oncogenesis are further considered below:-

Transduction of proto-oncogenes

With the exception of Rous sarcoma virus, transduction of a proto-oncogene involves loss of part of the retroviral genome. Such defective recombinant genomes can be packaged, but require a non-defective helper virus to provide structural proteins and/or enzymes. Transduced v-*oncs* are expressed at high levels, irrespective of integration site. When the v-*onc* is sufficiently potent, a

high percentage of infected cells undergo independent transformation to produce a polyclonal tumour. Clonality of a viral oncogene-induced tumour implies that the v-*onc* alone is insufficient for oncogenesis. Such a finding is commonplace and in keeping with multi-step theories of oncogenesis (Fulton, *et al.*, 1987).

Oncogenic mosaic env sequences

The Friend virus complex, which consists of a replication-competent type C murine leukaemia virus (FMuLV) and a replication-defective spleen focus forming virus (SFFV), causes rapidly progressive erythroleukaemia in susceptible strains of mice (Ruscetti and Wolff, 1984). The oncogenic component, SFFV is rapidly or "strongly" oncogenic, yet it does not carry a v-*onc*. The pathogenic effect of SFFV is mediated partly by its *env* gene glycoprotein product, which is a truncated, mosaic molecule derived by recombination of endogenous ecotropic and non-ecotropic proviral sequences. The mechanism by which this recombinant envelope glycoprotein induces erythroleukaemia in infected animals has been the subject of intensive study (for a thorough review, see Kabat, 1989). In brief, the recombinant *env* is an erythroblast mitogen, causing constitutive polyclonal expansion of infected cells, without immortalization. In the absence of helper FMuLV, polyclonal expansion is usually followed by cell differentiation and recovery. If FMuLV is present, continuing infection of erythroblasts can occur. SFFV integration will eventually occur stochastically at a site which causes immortalization of the infected cell. Clonal expansion of the immortalised cell follows.

Insertional mutagenesis and cis-*activation*

Most retrovirally-induced tumours are thought to arise by this mechanism. Integration of a retrovirus either adjacent to or within a proto-oncogene alters its expression, either quantitatively or qualitatively, so as to increase the probability of neoplastic transformation of the cell. The provirus may serve to distance the oncogene from regulatory sequences which would normally limit its expression. Alternatively promoter and enhancer sequences in the retroviral LTRs may increase the expression of the adjacent proto-oncogene. Promoters must be inserted in the same transcriptional orientation and upstream (5') of the target sequence to increase transcription. Promoters are present in both LTRs, but it is more usual for hybrid virus-c-*onc* RNAs to be initiated in the 3' LTR. Viral

enhancer elements upstream or downstream and in either orientation relative to the target gene will effectively increase expression from the oncogene promoter.

An example of retrovirally-induced, predominantly quantitative change in proto-oncogene expression involves avian leukosis virus (ALV)-induced lymphomas and c-*myc* in the domestic fowl. c-*myc* is a nuclear proto-oncogene, the protein-encoding domain of which is in exons II and III. Most characterised proviral insertions in chicken c-*myc* occur in exon I or in the intron between exons I and II. Thus the protein-coding domain is not disturbed. Moreover, the vast majority of proviruses are in the same transcriptional orientation as the c-*myc* gene. This is a classic example of promoter insertion which routinely causes a 20- to 100-fold increase in the level of *myc* RNA which codes for normal *myc* protein (Linial and Groudine, 1985).

An example of predominantly qualitative change in oncogene expression in the domestic fowl involves ALV-induced erythroblastosis and c-*erb*B (for a review, see Martin, 1986). This gene encodes the receptor for epidermal growth factor. Proviral insertion occurs in the same coding-region of c-*erb*B in almost every tumour. Proviruses are in the same transcriptional orientation as the gene. Virus-c-*erb*B hybrid RNAs code for a truncated form of the receptor protein which lacks the ligand-binding amino terminus. Tumour induction is thought to arise as a result of constitutive receptor activity in the absence of the growth factor (Raines *et al.*, 1985).

PATHOGENESIS OF NON-NEOPLASTIC DISORDERS

Compared with neoplastic diseases, the mechanisms of non-neoplastic retroviral disorders have, until recently, been relatively poorly characterised (Teich *et al.*, 1985). Recent efforts to understand the pathogenesis of AIDS have redressed this imbalance somewhat. HIV, the causative agent of AIDS, differs from many retroviruses in that it kills the $CD4^+$ cells which it infects. Loss of $CD4^+$ helper/inducer lymphocytes is responsible for much of the immunodeficiency caused by this virus. How HIV kills these cells has been the subject of much work. In early studies, interest focused on the novel, regulatory genes in the HIV genome (Varmus, 1988). Could one or more of these genes be responsible for the cytopathic effect? Studies involving induced mutagenesis and

transfection of these genes tended to refute this hypothesis (for examples see Rosen *et al.*, 1986; Terwilliger *et al.*, 1986). Another hypothesis concerned the finding that after infection, much of the DNA of HIV remained unintegrated in the cytoplasm (Ho *et al.*, 1987). Since unintegrated DNA had been associated with cytopathicity in cells infected with ALV (Keshet and Temin, 1979), this was suggested as a possible cause of HIV-induced cell death. However, transfection of HIV into $CD4^-$ cells led to accumulation of unintegrated viral DNA without cell death, arguing against a causal relationship (De Rossi, *et al.*, 1986). It now seems most likely that interaction between viral envelope glycoprotein (SU) and CD4 receptors ultimately causes cell death. In cell culture, HIV causes the cytopathic effect of syncytium formation because HIV-infected, SU-expressing cells become fused to other $CD4^+$ cells. Syncytia rapidly develop ballooning cytoplasm and die. Uninfected as well as infected cells are killed in this way. *In vivo*, cell death may be caused by similar fusion between cells or, perhaps, when separate parts of the plasma membrane of a single infected cell become attached to one another (Ho *et al.*, 1987).

A quite different form of immunodeficiency is caused by TM, the other *env* gene product. TM from FeLV (Mathes *et al.*, 1979) and MuLV (Schmidt *et al.*, 1987) decreases the usual blastogenic response of T-lymphocytes to alloantigens. Cianciolo *et al.* (1985) showed that a synthetic peptide identical to a hydrophilic, external domain of TM conserved amongst oncoviruses could produce the same immunosuppressive effect in murine and human cells *in vitro*. The equivalent peptide from HIV has a similar immunosuppressive effect (Ruegg *et al.*, 1989).

1.2 ENDOGENOUS RETROVIRUSES AND RETROVIRUS-LIKE ELEMENTS

1.2.1 THE VIRAL SUPERFAMILY OF RETROPOSONS

Retroposons have been defined as dispersed, repeated DNA elements, formed by the reverse flow of genetic information from RNA to DNA (Rogers, 1983). Retroposons constitute about 10% of the human genome and a similar figure may be true for other mammals (Temin, 1985). Retroposons can be divided into viral and non-viral superfamilies (for a review of non-viral retroposons, see Wiener *et al.*, 1986). The non-viral superfamily is much the larger, and includes

elements derived from all classes of cellular RNA except, intriguingly, the four ribosomal RNA species (5S, 5.8S, 18S and 28S). Non-viral retroposon intermediates include RNA polymerase II and III transcripts. SINEs (short interspersed nuclear elements or short interspersed repeated DNA elements) are typical non-viral retroposons, which are present in high copy number in mammalian genomic DNA (Deininger, 1989). SINEs are transcribed by RNA polymerase III. Retroposition of SINEs and of most other non-viral retroposons is thought to be a passive process, employing cellular mechanisms. However, certain non-viral retroposons, such as LINEs (long interspersed nuclear elements or long interspersed repeated DNA elements), encode an RT-related protein and may mediate their own retroposition (Hutchison *et al.*, 1989). Transposition of non-viral retroposons usually generates a 7-21 bp. direct repeat at the integration site. Many non-viral retroposons have a 3' terminal poly-A tract because they are derived from mRNA. In contrast, viral retroposons are all actively transposed using their encoded reverse transcriptase and integrase functions. Defective elements, lacking these functions, may be transposed using enzymes supplied *in trans* by intact elements. Viral retroposons have no poly-A tail, but are instead bounded by LTRs. A shorter (4-6 bp.) target sequence repeat is generated upon integration. Viral retroposons include exogenous and endogenous retroviruses, either intact or defective. A larger sub-category of the viral retroposon group consists of virus-like elements with no obligatory extracellular phase in their lifecycle. These have been termed, rather confusingly, retrotransposons (Boeke *et al.*, 1985). The boundaries between these different categories are not hard and fast.

Viral retroposons are found in eukaryotes as diverse as yeast (*Saccharomyces cerevisiae*), *Drosophila spp.* and man and are thought to constitute about 0.1-0.5% of the mammalian genome. The Ty elements of yeast have LTRs and have been shown to transpose through an RNA intermediate (Boeke *et al.*, 1985) justifying their inclusion in this category. A representative of the copia family of elements *of Drosophila melanogaster* has been sequenced (Mount and Rubin, 1985) and shown to have weak homology with vertebrate retroviral RT, but good homology with IN (integrase). Copia elements more closely resemble yeast Ty than vertebrate retroviruses. In humans, elements named THE-1 (transposonlike human element 1) are present in about 10,000 copies per genome.

Each element is 2.3 kbp. long, bounded by 350 bp. LTRs, and encodes a 2 kb. poly-adenylated RNA. The DNA sequence of THE-1 differs from that of known retroviruses (Paulson *et al.*, 1985).

As mutagenic agents, retroposons have the potential to produce profound deleterious effects (for an extensive review, see Lambert *et al.*, 1988). Despite this potential, there are few unequivocal examples of naturally-occurring retroposon-induced pathology. Most of these concern copia-like elements in *Drosophila spp.* In canine transmissible venereal tumour (TVT), a 1.8 kbp. non-viral retroposon is inserted 5' to the first exon of c-*myc* (Katzir *et al.*, 1985). Quantitation of recombinant c-*myc* expression was hampered by the lack of an appropriate control tissue, but these authors conjectured that retroposon-induced changes in c-*myc* expression might have caused neoplastic transformation. Another authority has suggested that since retroposons constitute as much as 10% of the mammalian genome, integration at this site is just as likely to be irrelevant and co-incidental (Peters, 1989).

1.2.2 POSTULATED ORIGINS OF RETROVIRUSES

Twenty years ago Temin (1970) postulated that retroviruses evolved from cellular "protoviruses". Since then, cellular retrotransposable elements (*q.v.*) have been characterised and Temin viewed this as strong support for his hypothesis (Temin, 1980; Shimotohno *et al.*, 1980). Recent dramatic increases in the availability of retrovirus and retroposon sequence data and the accessibility of computing power has allowed a thorough analysis of retroviral origins and evolutionary relationships (Doolittle *et al.*, 1989, 1990). Great care must be taken when attempting to establish phylogenetic relationships among retroviruses using sequence data. This is partly because recombination of sequences among distantly-related members of the family is commonplace. Another problem relates to the inability of computer software to compare highly dissimilar sequences. Certain exogenous retroviruses (for example lentiviruses) are evolving very rapidly and have highly divergent sequences. For example, the protease (PR) amino acid sequences of HIV and visna virus are as different from each other as are the acid proteases of mammals and fungi (Doolittle *et al.*, 1989). Nevertheless, using the relatively conserved *pol* product sequences of a wide variety of viral and non-viral

retroposons, interesting results have emerged. Eukaryotic reverse transcriptase-bearing elements can be divided into four groups. Vertebrate retroviruses constitute one group and various viral and non-viral retroposons, the other three. The weight of evidence suggests that the basic catalytic machinery of vertebrate retroviruses has evolved from very ancient sequences, probably pre-dating the prokaryote-eukaryote divergence. This view is supported by the recent discovery that certain bacteria can synthesise strange branched RNA-DNA molecules using their own encoded RT (Furuichi *et al.*, 1987; Lampson *et al.*, 1989; Inouye *et al.*, 1989). *gag*-related sequences are present among members of all four retroposon groups and are probably the next most ancient sequences. Doolittle *et al.* (1989) conjecture that *gag* sequences may have been obtained from other viruses. Only *env*-related sequences are unique to "true" retroviruses. The ability to form a spliced *env* product is essential for the extracellular stage of the retroviral life cycle. This ability seems to have evolved relatively recently, consistent with the fact that known infectious retroviruses are limited to certain vertebrates. Analysis of sequence data from vertebrate retroviruses indicates that horizontal transmission between birds and mammals has occurred. A fascinating and highly controversial question which has been raised in the past, is whether recently-evolved vertebrate retroviruses might be spreading through the biological world by horizontal infection to produce transposable elements (Doolittle *et al.*, 1990; Finnegan, 1983; Shiba and Saigo, 1983). This scenario, although intriguing, is not consistent with the available data from sequence analysis. Thus, it seems most likely that retroviruses evolved from transposable elements, rather than the other way around.

One further aspect of retroviral evolution should be discussed. The evolutionary rates of endogenous and exogenous retroviruses differ considerably. This is in part because the two groups are exposed to quite different selection pressures. Another consideration is the fact that exogenous retroviruses may undergo many more replication cycles per unit time than their quiescent endogenous counterparts. This is significant, because exogenous retroviral replication involves transcription by the error-prone enzyme RT, which lacks proof-reading capability (Doolittle *et al.*, 1989). Present-day evolutionary rates for infectious lentiviruses are astonishingly rapid. Indeed, Smith *et al.* (1988) have

proposed that HIV-1 and HIV-2, whose RTs are only 80% homologous, diverged as recently as 40 years ago. Yokoyama and Gojoboro (1987) have estimated that HIV and visna virus diverged only 300 to 500 years ago. In stark contrast, several copies of endogenous retroviruses are present at the same integration sites in both human and chimpanzee genomic DNA. Presumably, these proviruses were introduced into the germline before chimpanzees and humans diverged more than 8 million years ago and have remained stably-integrated ever since (Steele *et al.*, 1986).

Doolittle *et al.* (1989) have proposed a hypothesis to explain why there are presently no known retroviruses *very* distantly related to all the others, as one would predict might be the fate of the descendants of fast-evolving viruses such as HIV. They suggest that new exogenous retroviruses are continually arising from endogenous sequences but, in evolutionary terms, are short-lived. Their life spans might be measured in decades, or perhaps centuries. Extinction of a particular exogenous virus might be caused by the development of immunity within the host population. Alternatively, a particularly pathogenic virus might decimate the population of its host species. In any case, a proportion of exogenous viruses will become endogenised during their life span to continue the cycle (Doolittle *et al.*, 1989).

1.2.3 AVIAN, MURINE & FELINE ENDOGENOUS RETROVIRUSES

For reviews, see Coffin (1982b) and Stoye and Coffin (1985).

Historical Perspectives

In 1939, Andrewes speculated that ubiquitous, "indiginous", latent viruses might be activated in neoplastic tissues. Darlington (1948) suggested that such viruses might be part of the genetic complement of the cell and coined the term "provirus". Lieberman and Kaplan (1959) reported that radiation-induced murine lymphoma tissues expressed a filterable agent which could cause lymphoma in unirradiated mice. The possibility that a horizontally-transmitted virus was confounding these experiments could not be ruled out. Convincing evidence of the inheritance of viral genes came in 1968, when Payne and Chubb showed in chickens that expression of a retroviral antigen, in the absence of horizontally transmitted virus, was inherited in a typical Mendelian fashion as an autosomal,

dominant trait. Parallel experiments with foster-suckled mice showed that, as well as being milk-borne, MMTV can be transmitted genetically (Bentvelzen *et al.*, 1970).

With the advent of restriction enzymes and more sophisticated DNA hybridisation methodologies (Southern, 1975), it was possible to show that normal mammalian DNA contains numerous copies of retrovirus proviral genomes (for example see Steffen and Weinberg, 1978).

Avian Endogenous Viruses Endogenous retroviruses were first described in chicken cells (Dougherty and DiStefano, 1966) and it has turned out that chickens have a relatively simple arrangement of endogenous viruses compared with mammals. All characterised endogenous viruses of the chicken are very closely related to each other and to exogenous avian oncoviruses. About 14 distinct proviral loci (*ev*s) have been identified, but any one chicken is unlikely to have more than 6 of these (Coffin, 1982b). Some of these loci encode infectious virus (*e.g.*, *ev*-10, *ev*-11), while others are defective. Intriguingly, *in situ* hybridisation studies show that several of the defective loci (*ev*-1, -4, -5, -8 and -13) are clustered on chromosome 1, prompting the suggestion that their replication may involve an amplification process distinct from reverse transcription and random integration (Tereba, 1983). Because endogenous proviruses are inherited as stable Mendelian genes, interest focused on the possibility that they could be used to establish phylogenetic relationships among related avian species. This was tried and found to be unreliable by Frisby *et al.* (1979). These workers found that presence of DNA sequences related to Rous-associated virus (RAV-0) amongst members of the genera *Gallus* and *Phasianus* did not parallel phylogenetic relationships. Specifically, RAV-0 is present in both chickens and red jungle fowl, but not other closely-related galliform birds. Yet a few other distantly-related species also have RAV-0-related sequences in their germlines. Presumably germline infection of the ancestors of some species, but not others, has occurred.

Murine Endogenous Viruses

The endogenous viruses of laboratory mice represent one of the most complex of all virological systems (Coffin, 1982b). The murine genome contains multiple copies of at least seven types of viral retroposon; types A, B, C, VL30, VL30-GLN, ETn and MuRRS (Varmus and Brown, 1989). Of these, the type C

proviruses, which include the murine leukaemia viruses (MuLVs), are best characterised. Because of their extensive inbreeding, laboratory mice are probably not a good model to study the relevance of endogenous proviruses in relatively outbred mammals such as dogs or humans. Nevertheless endogenous viruses have been more intensively studied in mice than in any other species. Some of the concepts which have emerged from these studies are presented here.

Murine C-type viruses may be classified according to host-range properties as ecotropic, xenotropic, polytropic or amphotropic. Host range is primarily a function of the viral *env* gene product, SU. *Ecotropic viruses* are restricted to growth in murine cells and cells of closely-related rodents. The endogenous virus AKV is a member of this class. In a screen of 54 strains of laboratory mice, Jenkins *et al.* (1982) found that 30 strains contained one copy of an endogenous ecotropic provirus and 12 strains contained multiple copies (from one to six). *Xenotropic viruses* will not grow well in murine cells, but grow in cells from a wide variety of other species, including birds. They are widely distributed in the germlines of laboratory mice, which contain 5 to 16 copies per genome (Stoye and Coffin, 1988). They are also present in Asian and East European wild mice (Kozak and O'Neill, 1987). *Polytropic viruses* will infect both murine and foreign cells. They produce characteristic foci when grown in mink lung cells, leading to their alternative designation as mink cell focus-forming (MCF) viruses. MuLV-like elements with MCF-related *env* genes are present in high copy number (20 to >30) in genomic DNA from laboratory mice (Stoye and Coffin, 1988). DNA from some wild mice also contains MCF *env*-related sequences (Kozak and O'Neill, 1987). *Amphotropic viruses* will replicate in murine and certain foreign cells, but can be distinguished from polytropic viruses by interference testing and the fact that they do not form mink cell foci. They have only been isolated from Californian wild mice (Hartley and Rowe, 1976). Hybridisation studies have not demonstrated amphotropic virus *env*-related sequences in the murine germline (O'Neill *et al.*, 1987).

Many strains of laboratory mice have inducible ecotropic, xenotropic and polytropic proviruses in their germlines. Despite their different host range properties, these endogenous viruses are closely-related, as shown by restriction enzyme analyses and sequence data (for example, see Chattopadhyay *et al.*, 1981).

As one would predict, the *env* genes of these groups show the greatest sequence divergence. Hybridisation experiments, using probes derived from divergent parts of the *env* genes to probe murine genomic DNA, show that non-ecotropic sequences outnumber ecotropic proviruses by 10- to 20-fold (Jenkins *et al.*, 1982; Hoggan *et al.*, 1983; Stoye and Coffin, 1988). Despite their numerical superiority, very few of the non-ecotropic proviruses encode infectious virus, whereas many of the characterised ecotropic loci do (Stoye and Coffin, 1985). These differences in expression probably reflect dissimilar levels of proviral DNA methylation and mutation (*q.v.*).

Among the different MuLVs, pathogenicity is very variable. Polytropic MuLVs are strongly implicated in induction of T-cell lymphoma in certain strains of mice. These leukaemogenic viruses arise by somatic recombination between an infectious, ecotropic virus and one or more non-ecotropic endogenous viruses (Khan, 1984). The polytropic host-range is conferred by the non-ecotropic parent(s). Leukaemogenic, polytropic viruses derived from the ecotropic endogenous virus AKV, have a polytropic proviral substitution in SU and xenotropic sequences at the 3' end of TM and 5' U_3. The TM-U_3 substitution is important in determining tissue tropism and leukaemogenic potential. The SU substitution may also play a leukaemogenic role by acting as a mitogen, in a similar manner to the chimaeric SU of SFFV. (Holland *et al.*, 1985; Davis *et al.*, 1987).

In addition to their oncogenic potential, C-type murine endogenous retroviruses have been shown to act as insertional mutagens in non-neoplastic disorders. One of the dilute coat colour mutations (d^V) in certain inbred mouse strains is caused by insertion of an ecotropic endogenous virus (*Emv*-3) in a non-coding region at or near the *d* locus. Affected animals are homozygous for the mutation. A causal relationship between presence of *Emv*-3 and dilute coat colour was proved when Jenkins *et al.* (1981) showed that revertant, wild-type colour animals of *d* stock had lost *Emv*-3. Subsequent work with revertants showed that a single LTR remained at the former integration site, excision of *Emv*-3 presumably having been a consequence of homologous recombination between LTRs (Copeland *et al.*, 1983).

MMTV is the prototypic B-type retrovirus. Most strains of laboratory mice carry several endogenous MMTV proviruses, although they are absent from the

DNA of many wild mice. Certain endogenous MMTV proviruses have been shown to cause late-onset mammary tumours. MMTV has also been implicated as a cause of thymic lymphoma. Intracisternal A-type particle (IAP) proviruses are present in about 1000 copies in murine germ line DNA and have been shown to act as insertional mutagens, influencing both positively and negatively the expression of adjacent cellular genes (Kuff *et al.*, 1983). Another group of endogenous retroelements, termed VL30 sequences encode virus-like 30S RNAs. These genomes can be co-packaged into virions released by C-type virus-expressing rodent cells. VL30 elements are present in multiple copies in mouse DNA and most likely represent defective retroviral proviruses. Other murine endogenous retrovirus-like elements have been described which are recombinant DNAs containing components of VL30, MuLV and other endogenous retroviral elements (GLN family, Itin and Keshet, 1986; B-26, Obata and Khan, 1988). One such element (B-26) is present in similar integration sites in woodchucks, guinea pigs and mice. It is therefore thought to be ancient (*i.e.* it has been in the rodent germ line for at least 50 million years).

Feline Endogenous Viruses

In common with some other species, cats have several endogenous proviruses closely related to their most important exogenous retroviral pathogen. In the case of cats, that pathogen is feline leukaemia virus (FeLV), which can be classified into 3 subgroups (A, B and C) on the basis of interference testing (Sarma and Log, 1973). Subgroup A viruses, which are strictly ecotropic, are found in every field isolate of exogenous virus. Subgroup B and C viruses are polytropic and, in the wild, are always isolated with a subgroup A virus. Subgroup B viruses are thought to be generated by recombination between an exogenous subgroup A virus and endogenous FeLV sequences (Overbaugh *et al.*, 1988). Certainly, feline genomic DNA contains between 15 and 20 endogenous proviruses with homology to FeLV-B (Stewart *et al.*, 1986; Kumar *et al.*, 1989). Subgroup C viruses are isolated more rarely and are probably generated by a similar recombinational mechanism, although the nature of the endogenous sequences involved in such a recombination has not yet been fully defined. The generation of FeLV-B and -C has close parallels with the situation described above for mice, in which an ecotropic infectious virus combines with non-ecotropic endogenous sequences to

produce a polytropic MCF virus. A further parallel is the fact that MCF viruses are more proximal leukaemogenic agents than their ecotropic, infectious parents. The same may be true of FeLV: the appearance of lymphoproliferative disease is strongly (though not exclusively) associated with A/B mixed infection (Jarrett *et al.*, 1978). Unlike the situation in mice, none of the endogenous FeLV elements are known to produce infectious virus, although sub-genomic transcripts are expressed in a tissue-specific manner (Niman *et al.*, 1980). Berry *et al.* (1988) have shown that endogenous FeLV LTRs retain their ability to promote and enhance transcription *in vitro*, despite having considerable sequence divergence from exogenous FeLV in the U_3 region. Opposing these functional promoter and enhancer elements, some endogenous proviruses have *cis*-acting regulatory elements in 5' flanking cellular DNA which exert a negative influence on expression. Transcription is also limited by numerous nonsense and frameshift mutations which have been detected in the *gag* coding region. This defectiveness correlated with a lack of infectivity of the proviral DNA.

Feline genomic DNA contains another family of retroviral proviruses unrelated to FeLV, termed RD 114 or feline endogenous virus (Reeves and O'Brien, 1984). One or more of these elements produce an infectious xenotropic virus. RD 114 has sequence homology with baboon endogenous virus (BaEV). It has been found only in four species of felid closely related to the domestic cat. This finding indicates that RD 114 became incorporated into the germline of a common ancestor of these four species, but not of other felids.

1.2.4 HUMAN ENDOGENOUS RETROVIRUSES

This subject has been reviewed recently (Callahan, 1988; Larrson *et al.*, 1989). Unlike the species discussed in the previous sections, humans do not have endogenous sequences closely related to their important, known exogenous retroviral pathogens. They do, however, have hundreds and perhaps thousands of endogenous retroviruses (ERVs) or retrovirus-like elements detected by their homology to retroviruses of other vertebrates. Several human ERVs (HERVs) have been cloned and sequenced and have had their mRNA products characterised. To date, all known HERVs are defective and, in contrast to some other species, it is unlikely that there are many infectious ERVs in human DNA.

HERVs can be divided into two diverse groups (designated class I and II) on the basis of their sequence homology to known retroviruses. Class I elements have high homology to mammalian type C viruses. Class II elements share homology with types A, B, D and avian type C viruses. A further system of nomenclature (cited in Larrson *et al.*, 1989), which categorises HERVs according to the tRNA species complementary to their primer binding site (PBS) has been devised. Thus, HERV-K (UUU) denotes a family of endogenous elements using the lysine tRNA with the UUU anticodon as primer (K is the one-letter code for the amino acid lysine). This system, although useful, poses problems when categorising proviruses for which the PBS sequence has not been established, or in which the PBS has undergone mutation.

CLASS I HERVS

Single copy elements

ERV 1 (PBS sequence not established), located on chromosome 18q22-23, lacks a 5' LTR and has not been found to be expressed as mRNA. An almost identical provirus with identical 3' flanking cellular sequence is present in chimpanzee DNA. It seems highly likely that the progenitor of ERV 1 was integrated at this chromosomal locus in the common ancestor of chimpanzees and humans (Bonner *et al.*, 1982). HERV-R (ERV 3) is a full-length, but defective provirus located on chromosome 7 (O'Connell *et al.*, 1984). An equivalent proviral element has been identified in other primates. HERV-R *env* contains a 1940 bp. open reading frame and is highly transcribed in normal human placental villi. Interestingly this element is not transcribed in choriocarcinoma cells (Kato *et al.*, 1988). S71, an incomplete provirus on chromosome 18q21, contains *gag*, *pol* and 3' LTR sequences with sequence homology to simian sarcoma virus and its helper, simian sarcoma-associated virus (Werner *et al.*, 1990).

Multiple copy elements

The HERV-E family consists of 50-100 type-C proviral elements, some of which are complete (for example, clone 4-1; Repaske *et al.*, 1985) and others of which lack *env* sequences. Solitary HERV-E LTRs are also present. One truncated form of HERV-E is fascinating, in that it is bounded by tandem arrays of 8 to 13 or more imperfect repeats, each repeat being 70 to 80 bp. in length (Steele *et al.*, 1984). The arrangement of the terminal repeat elements indicates that this

form of HERV-E has been amplified subsequent to the process which produced the tandem repeats. A variety of truncated HERV-E mRNAs are expressed in normal placenta, but those characterised contain in-frame premature termination signals in the coding regions.

Another group of sequences, HERV-H , is present in very high copy number (800-1000 proviral copies, plus about 1000 solitary LTRs; Mager and Henthorn, 1984). *In situ* studies show that these copies are dispersed among the chromosomes with clusters on 1p and 7q (Fraser *et al.*, 1988). cDNA clones have been isolated which contain HERV-H LTRs, some of which have provided the polyadenylation signal for non-retroviral transcription units (Mager, unpublished; cited in Larrson *et al.*, 1989). If HERV-H LTRs are, indeed, providing 3' processing signals for the transcription of cellular genes, that would be an exciting finding.

CLASS II HERVS

HERV-K (CUU) proviral elements were independently isolated from human genomic DNA using probes derived from Syrian hamster IAP (HERV-K10, Ono, 1986) and MMTV (HLM-2, Callahan *et al.*, 1982). There are approximately 50 HERV-K (CUU) proviruses per haploid human genome. HERV-K (CUU) elements are mosaic sequences, with LTRs related to type D viruses, *gag* and *pol* sequences related to types A, B, D and avian type C viruses and a putative *env* gene (Callahan *et al.*, 1985). A member of this group (HERV-K10) has been sequenced (Ono, 1986). Its 968 bp. LTRs differ from each other by only 2 bp., prompting the suggestion that this element is, in evolutionary terms, relatively young. In contrast, another member of the group, HLM-2, appears to be integrated at the same chromosomal site in humans, chimpanzees, gorillas and orangutans, but not in ape lar gibbons (Mariani-Constantini *et al.*, 1989). These authors suggest that HLM-2 entered the ancestral hominoid germline between 40 and 17 million years ago.

A distinct group of elements, despite their similar name, is HERV-K (UUU). A member of this group was cloned from breast tumour DNA using a MMTV *gag-pol* probe (May and Westley, 1986). Human DNA contains 10-50 of these elements. Yet another group of elements, HERV-P was discovered using the proline tRNA PBS sequence as a probe (Harada *et al.*, 1987). This novel approach

was adopted because many type-C retroviruses use proline tRNA as their primer. Interestingly, one HERV-P provirus has an *Alu* repeat sequence interrupting its 5' LTR (Harada *et al.*, 1987). There are 20-40 copies of HERV-P per haploid genome. Lastly, a provirus with a PBS complementary to isoleucine tRNA (HERV-I) was found serendipitously during characterisation of the haptoglobin-related locus (Maeda, 1985). This 9 kbp., full-length provirus also contains an *Alu* repetitive sequence, this one interrupting the TM-encoding segment of *env*. There are 15-30 copies of HERV-I per haploid genome.

1.2.5 EXPRESSION OF ENDOGENOUS RETROVIRUSES

This subject has been reviewed (Coffin, 1982b; Lambert *et al.*, 1988) and some aspects have been mentioned in the preceding sections. Factors which influence expression of endogenous retroviral sequences are categorised and briefly discussed below:

Proviral Defectiveness

The majority of proviral elements for which sequence data is available have multiple mutations (mis-sense, nonsense, deletions, non-viral insertions) in their coding-regions which preclude production of infectious virus. However, many elements have an open reading frame in *env* and could thus code for truncated or possibly functional SU if *env* were transcribed. Defects in U_3 of the 5' LTR may partly account for transcriptional inactivity of some groups of elements.

DNA methylation

Eukaryotic DNA is extensively methylated at ^{5}C in the dinucleotide CG (for a review, see Razin and Riggs, 1980). The extent of methylation of a particular gene is easily assessed by contrasting restriction patterns produced by pairs of restriction enzymes with the same basic recognition sequence, only one of which will cut methylated DNA. In any particular tissue, actively-transcribed genes are demethylated and inactive genes are methylated. Methylation, then, is a mechanism by which genes not in use can be silenced. Moreover, DNA methylation, and hence the silencing of inactive genes, is inherited: because CG is palindromic, daughter DNA strands are also CG. DNA methylase recognizes ^{m5}CG on one strand of newly synthesised ds DNA and methylates its complement.

Methylation is involved in the control of retroviral provirus expression. In chickens, actively-expressed proviral loci (*e.g.*, *ev*-3) are demethylated compared with transcriptionally-silent loci such as *ev*-1 (Groudine *et al.*, 1981). Brief treatment of growing chicken cells with the demethylating agent 5-azacytidine causes high and permanent induction of expression of *ev*-1 (Groudine *et al.*, 1981).

cis-acting repressors

The chicken endogenous provirus *ev*-2 contains all sequences necessary for production of infectious virus, but is expressed at a very low rate. The regulation of virus production is *cis* to the *ev*-2 provirus and is not inherited by its progeny. Transfection of *ev*-2-containing DNA into chicken cells did not efficiently induce a productive infection. Shearing of *ev*-2-containing DNA prior to transfection enhanced infectivity, prompting the suggestion that removal of repressor sequences close to the provirus could increase expression (Cooper and Silverman, 1978). Another example of *cis*-acting repression concerns endogenous FeLV. As previously mentioned, endogenous FeLV proviruses differ in their U_3 sequence from their exogenous counterparts. Nevertheless, the endogenous LTRs have been shown to have competent enhancer and promoter function *in vitro* (Berry *et al.* 1988). A *cis*-acting repressor element in the cellular DNA immediately 5' to one of the proviruses has a strong regulatory effect upon transcription of this element.

trans-activators

In certain strains of inbred mice, the *Gv-1* locus regulates the expression of multiple endogenous retroviral elements *in trans* (Wilson *et al.*, 1988). Only a subset of the total complement of murine proviruses are responsive to *Gv-1*. Some of the responsive elements have distinctive *env* deletions, presumably generated by homologous recombination between 7 bp. direct repeats identified in non-deleted, but otherwise identical elements. Using oligonucleotide probes spanning the deletion breakpoint, *Gv-1*-responsive elements have been cloned and subsequently characterised. These elements have unusual 192 bp. insertions in U_3, which may account for their *Gv-1* responsiveness (Wilson *et al.*, 1988).

Post-transcriptional regulation

McDonald *et al.* (1988) have shown that, in *Drosophila* fruit flies, the abundance of copia transcripts does not correlate with the presence of retrovirus-

like particles. Copia transcripts were found to be abundant in a variety of tissues of young adult flies, yet retrovirus-like particles have never been observed in flies of this age. Transcripts were also found in tissues which have not been reported to produce retrovirus-like particles. These authors suggested that a significant part of host-mediated, *trans*-regulation of expression of retrovirus-like elements occurs at the post-transcriptional level. Further studies to prove that the transcripts present in the tissues of young flies are capable of being translated and that there is truly a host-mediated translational block are needed.

1.2.6 RELEVANCE IN HEALTH AND DISEASE OF THE HOST

The fact that multiple endogenous proviruses have been part of the vertebrate genome for millions of years has led many investigators to speculate that some of these elements might have adopted an essential physiological role; rather than being benign, but selfish genes. There is presently no concrete evidence to support this attractive hypothesis. For example, strains of chickens vary enormously in their complement of proviruses, some strains carrying none of the characterised elements. Yet these apparently virus-free chickens remain perfectly healthy (Astrin *et al.*, 1979). This finding does not rule out the possibility that ancient, poorly-characterised, perhaps heavily-defective elements might have an essential function. Many defective proviruses express partial or complete SU glycoprotein. Such SU could bind its specific receptors in and on the surface of the host cell, rendering the receptors unavailable to potential pathogens. Another frequently-encountered postulate is that post-embryonic proviral expression could benefit the host by sensitising its immune system to antigenically-similar, potentially pathogenic exogenous retroviruses. Conversely, embryonic expression, as is known to occur for many proviruses, would produce immunological tolerance with presumably deleterious consequences. Larrson *et al.* (1989) conjecture that HERV-R expression might have an anti-oncogenic effect because HERV-R mRNA is present in normal cells, but absent from choriocarcinoma and other trophoblastic neoplasms. Another interpretation of this data is that the neoplastic cells do not produce regulatory signals required for HERV-R expression. Clearly, more work must be done to investigate these interesting findings.

The deleterious effects of endogenous viruses are better understood than their potential beneficial ones. Endogenous elements related to exogenous retroviral pathogens can recombine with infectious viruses and, in some cases, enhance their pathogenicity. Exogenous viruses with lethal lesions in essential genes can be repaired by homologous recombination with endogenous retroviral sequences (Schwartzberg *et al.*, 1985). In the absence of closely related exogenous virus (as seems to be the case in man and other primates) endogenous proviruses play a less conspicuous role. They may occasionally retrotranspose and act as insertional mutagens as has been shown to occur at the *dilute* locus in inbred mice (Jenkins *et al.*, 1981). Insertion of a retrotransposed provirus with a functional promoter or enhancer adjacent to a proto-oncogene could induce neoplasia by *cis*-activation, but this has yet to be definitively demonstrated.

1.3 EVIDENCE FOR THE EXISTENCE OF A CANINE RETROVIRUS

There are many reasons to suspect that an infectious canine retrovirus might exist. Despite decades of effort, no such virus has been discovered. Neither have endogenous canine retroviral elements been characterised. This section describes the canine diseases for which a retroviral aetiology might be suspected and gives details of previous experimental work carried out in the search for a canine retrovirus.

1.3.1 CANINE DISEASES WITH A SUSPECTED RETROVIRAL AETIOLOGY

Dogs suffer from a variety of neoplastic, immunological, degenerative and proliferative disorders for which a retroviral causation is suspected. This suspicion arises mainly because retroviruses have been shown to cause similar diseases in other species. Greatest suspicion has been focused upon the haematopoeitic neoplasms; and in particular, upon canine lymphoma and lymphoid leukaemias (for reviews, see Madewell and Theilen, 1987; MacEwen and Young, 1989). Lymphoma (or malignant lymphoma, lymphosarcoma) is one of the more common canine neoplasms. The annual, age-adjusted incidence for lymphoma and lymphoid leukaemias has been reported as 26.4 per 100,000 dogs at risk (Schneider, 1983). Lymphoma occurs most commonly in middle-aged animals, although very young and old dogs are affected. There is no sex predilection.

Airedales, basset hounds, boxers, bulldogs, St. Bernards and Scottish terriers are at increased risk. In one report, there was a very high incidence of lymphoma in the Bull mastiff breed, with a familial distribution (Onions, 1984). The most frequent presenting sign of canine lymphoma is non-painful, generalised lymphadenopathy, although the disease can present in many other guises. Diagnostic investigations, including biopsy, are conducted to establish the definitive diagnosis and the clinical stage of the disease (a measure of its severity). Treatment, which is palliative, usually involves combination chemotherapy.

Paraneoplastic hypercalcaemia is relatively common in canine lymphoma and is often associated with the cranial mediastinal form of the disease. In a recent study, 11 of 176 lymphoma patients were hypercalcaemic (Greenlee *et al.*, 1990). Hypercalcaemia was strongly correlated with the "T-cell" phenotype (*i.e.*, pan-T$(LQ_1)^+$ SIg^-; Brundage-Anguish and Quimby, in press) and *vice versa*: Of 12 pan-T$(LQ_1)^+$ SIg^- tumour-bearing patients, 9 were hypercalcaemic. Only two of the hypercalcaemic patients did not have tumours of this phenotype. The fact that paraneoplastic hypercalcaemia occurs so frequently in canine lymphoma is interesting, because it is also a frequent finding in the retrovirally-induced human neoplasm, adult T-cell leukaemia (Kiyokawa *et al.*, 1987). The mechanism by which hypercalcaemia is induced seems to be identical in the two species, involving secretion of a parathyroid hormone-related protein (PTHrP) by tumour cells (Motokura *et al.*, 1989; Weir, 1990; Weir *et al.*, 1988). Despite these parallels, efforts to isolate a lymphotropic retrovirus from canine T-cell lymphomas have so far been unsuccessful (Quimby, F.W. unpublished; cited in Greenlee *et al.*, 1990).

Lymphoid leukaemia in the dog may be acute lymphoblastic (ALL) or chronic lymphocytic (CLL). In both disorders neoplastic lymphoid cells infiltrate the bone marrow, disrupting normal haematopoiesis. Neoplastic cells frequently spill over into the peripheral blood. ALL is much more common in females than males. There is no known breed predilection for either disorder. Clinical signs are very non-specific. Definitive diagnosis relies upon haematological and bone marrow examination.

Other canine diseases for which a retroviral aetiology has at one time or another been mooted include systemic lupus erythematosus, autoimmune haemolytic anaemia, cyclic neutropenia of grey collies and mammary carcinoma.

1.3.2 EXPERIMENTAL EVIDENCE FOR THE EXISTENCE OF A CANINE RETROVIRUS

In the 1960s and 70s, electron microscopy (EM) of naturally-occurring tumour tissues and transmission studies (cellular and cell-free) were used in attempts to demonstrate the existence and potential aetiological significance of a canine retrovirus. French, German and American workers independently identified retroviral particles in a proportion of canine lymphoma tissues using EM (Chapman *et al.*, 1967 ; Seman *et al.*, 1967; Rudolph, 1971). More recently, retroviral particles have been identified in neoplastic cells from a case of granulocytic leukaemia which was complicated by hypercalcaemia (Sykes *et al.*, 1985). Turning to transmission studies, in 1966 Moldovanu *et al.* reported experimental cellular transmission of canine malignant lymphoma to X-irradiated canine neonates. Two years later Kakuk *et al.* carried out a similar transmission study using healthy, unirradiated beagle neonates. They identified retroviral particles in the recipients using EM (Kakuk *et al.*, 1968). Cell-free transmission of canine lymphoma has not been reported. However, mast cell leukaemia, a very rare disorder, has been transmitted using cell-free filtrates on two separate occasions (Lombard *et al.*, 1963; Post *et al.*, 1970).

Bonner and Todaro (1979) used a DNA hybridisation technique to demonstrate endogenous retroviral sequences in carnivore DNA. They probed genomic DNA from several species of carnivore, including dogs, with the primate endogenous type-C virus, MAC-1. They showed that felid DNAs, irrespective of the presence or absence of RD114 and endogenous FeLV, contained sequences that hybridised to MAC-1. Canine DNA also hybridised, although to a lesser extent.

In the same year (1979), Strandstrom and Rimaila-Parnanen reported tantalizing findings from Finland. They described a case of atypical malignant lymphoma in a 3 year-old Belgian shepherd dog with some histopathological similarities to human Hodgkin's disease. In the last paragraph of their paper they mentioned briefly that a cell line (3132) established from ascitic fluid from the dog

released retrovirus-like particles. A subsequent publication by Strandstrom and Bowen (1982) described this retrovirus in more detail. The retroviral particles had a buoyant density of 1.16 to 1.17 g/cm^3 in sucrose gradients and contained 60-70S RNA and Mn^{++}-dependent RT. Particles contained a major protein of 30 kilo-Daltons (kD) and a structural glycoprotein of 70 kD. The 30 kD protein was antigenically related to the major proteins of gibbon ape leukaemia virus, simian sarcoma virus and baboon endogenous virus. It was not related to rodent or feline type-C viruses. The virus infected normal canine and human cells in tissue culture and induced productive infection in neonatal puppies. Unfortunately, attempts to reproduce these exciting results have been unsuccessful (D.E. Onions, personal communication). The possibilities that the 3132 virus was a canine leukaemia virus, or perhaps a transiently-expressed canine endogenous retrovirus remain unproven. Likewise, the possibility of inadvertent contamination of the canine cell line cannot be ruled out.

During the 1980s a few studies demonstrated RT activity and particles of retroviral density in a variety of neoplastic and non-neoplastic canine tissues. Onions (1980) recognized RT activity of borderline significance in supernatants from 3 of 14 short-term lymphoma cell cultures and in 2 of 11 crude preparations of tumour tissue. In a larger study, Tomley *et al.* (1983) studied tissues from 43 dogs with lymphoma and 40 normal dogs. 76% of short-term culture supernatants from dogs with lymphoma contained particles of retroviral density, or demonstrated RT activity, or both; compared with 24% of culture supernatants from normal canine lymphoid cells. In this study, many of the lymphoma patients had received chemo- or radiation therapy prior to tissue sampling. This may partly explain the higher percentage of retrovirus-positive samples as compared with the previous work by Onions (1980). In a further study from the Cambridge veterinary school (Littlewood *et al.*, 1984) RT activity was demonstrated in milk from 7 of 80 healthy beagles. Repeat milk samples were taken from four of the seven RT-positive bitches: two remained positive on the second sampling. Further work is needed to establish the aetiological significance, if any, of these findings.

1.4 AIMS OF THIS PROJECT

The primary objective of this project was to identify and characterise canine endogenous retroviral elements by application of molecular biological techniques. The role of such elements as insertional mutagens could then begin to be investigated. No canine retroviral sequence data was available at the start of this project; although there were strong reasons to believe that the dog, like many other vertebrates, would harbour endogenous retroviral elements in its genome. Elucidation of the canine endogenous retroviral complement would allow a better understanding of the experimental findings described in the preceding section. An appreciation of background endogenous retroviral expression in normal canine tissues would help in the search for a horizontally-transmissible, exogenous canine retrovirus. Had an exogenous, aetiologically-significant, canine retrovirus been found during the characterisation of endogenous sequences, that virus would obviously have been studied with vigour.

Specific intentions of the project were to:-

1. Clone and sequence elements from canine genomic DNA that hybridised with an established retroviral probe and determine whether such elements represented canine endogenous virus.

2. Using probes derived from the cloned canine endogenous retroviral elements, establish the arrangement and approximate copy number of such elements in canine genomic DNA from normal and lymphomatous tissues.

3. By probing Northern transfer membranes, determine the pattern of RNA expression of canine endogenous retroviral elements.

4. Attempt to induce expression of endogenous retrovirus from canine cell lines using established techniques.

5. Determine whether the retrovirus produced by the canine cell line A72-E represented an endogenous or exogenous canine retrovirus.

6. Begin to investigate a possible role for canine endogenous proviruses as insertional mutagens by probing canine lymphoma tissue for proto-oncogene rearrangements associated with proviral insertion.

Secondary objectives of the project were to investigate whether there are several different families of endogenous virus in the canine genome by probing canine genomic DNAs with a variety of different established retroviral probes. It

was considered of interest to determine whether different breeds of dog had dissimilar proviral arrangements and whether this could be correlated with breed-specific, hereditary diseases.

Chapter 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 CHEMICALS

2.1.2 RADIOCHEMICALS

2.1.3 EQUIPMENT

2.1.4 BACTERIAL STRAINS

2.1.5 CLONING VECTORS

2.1.6 RESTRICTION ENDONUCLEASES

2.1.7 OTHER ENZYMES

2.1.8 BUFFERS, SOLUTIONS & GROWTH MEDIA

2.1.9 PROVIDED RECOMBINANT DNA CLONES

2.1.10 CELL LINES

2.1.11 CANINE TISSUE BIOPSIES AND POST MORTEM MATERIAL

2.2 GROWTH AND MANIPULATION OF BACTERIOPHAGES AND CELLS

2.2.1 GROWTH AND MANIPULATION OF BACTERIOPHAGES

2.2.2 GROWTH AND MANIPULATION OF BACTERIA

2.2.3 GROWTH AND MANIPULATION OF MAMMALIAN CELLS

2.3 EXTRACTION, PURIFICATION AND SEPERATION OF NUCLEIC ACIDS

2.3.1 PREPARATION OF BACTERIOPHAGE LAMBDA DNA

2.3.2 PREPARATION OF PLASMID DNA

2.3.3 EXTRACTION OF NUCLEIC ACIDS FROM MAMMALIAN TISSUES

2.3.4 AGAROSE GEL ELECTROPHORESIS

2.3.5 EXTRACTION OF DNA FROM LOW MELTING POINT (LMP) AGAROSE GELS

2.3.6 POLYACRYLAMIDE GEL ELECTROPHORESIS

2.3.7 PREPARATION AND PURIFICATION OF SYNTHETIC OLIGONUCLEOTIDES

2.4 ENZYMATIC MANIPULATION OF NUCLEIC ACIDS

2.4.1 RESTRICTION ENDONUCLEASE DIGESTIONS

2.4.2 DNA LIGATION

2.4.3 ^{32}P RADIO-LABELLING OF ds DNA

2.4.4 SEQUENCING OF ds DNA & SEQUENCE DATA ANALYSIS

2.4.5 RNA-DIRECTED DNA POLYMERASE (RDDP) ASSAYS

2.5 HYBRIDISATION OF NUCLEIC ACIDS

2.5.1 SOUTHERN TRANSFER AND HYBRIDISATION

2.5.2 NORTHERN TRANSFER AND RNA-DNA HYBRIDISATION

2.5.3 BACTERIOPHAGE PLAQUE HYBRIDISATION

2.5.4 AUTORADIOGRAPHY

2.1 MATERIALS

The following section lists only those materials used routinely. Less-frequently used materials are detailed in the appropriate Methods section.

2.1.1 CHEMICALS

All chemicals were of ultrapure, analaR or equivalent purity and were obtained from BDH chemicals Ltd., Boeringer Mannheim Corp., Gibco BRL, Pharmacia LKB, or The Sigma Chemical Company, except:

Ampicillin (Penbritin™)	Beecham Ltd.
Bacto Agar, tryptone & yeast extract	Oxoid Ltd.
Ecoscint	International Diagnostics Inc.
Ethanol	James Burroughs Ltd.
Formamide	Fluka Chemie AG.
RNAzol™	Biogenesis Ltd.

2.1.2 RADIOCHEMICALS

[Alpha ^{32}P]dCTP for labelling of DNA probes and [^{35}S]dATP alpha S for DNA sequencing were obtained from Amersham International plc.

2.1.3 EQUIPMENT

Microcentrifuge tubes, pipette tips and other disposable plasticware were obtained from Scotlab Ltd., Gibco BRL. and Elkay Lab Products Inc.

Tissue culture flasks and other plasticware - Nunclon

Collodion Dialysis tubes - Sartorius

Hybridisation Membranes - "Hybond N" Amersham International plc.

Autoradiography film - "Hyperfilm MP" Amersham International plc.

Sequencing Apparatus - Pharmacia LKB macrophor system.

Filter paper - Whatman International Ltd.

Disposable filter assemblies (0.2um & 0.45um) - "Acrodisc" Gelman Sciences Inc.

2.1.4 BACTERIAL STRAINS

Escherichia coli (*E. coli*) DS941: This derivative of *E. coli* K-12 strain AB1157 (Bachmann, 1972) was used in all plasmid experiments. It is AB1157 *recF lacI^q* *lacZ* delta.M15 (Summers & Sherratt, 1988). The strain is resistant to streptomycin and because of complementation *by lacZ* delta.M15, blue/white selection for plasmids or phagemids carrying the appropriate *LacZ* amino-terminal fragment is possible.

E. coli Q358: This strain was used as the host for growth of the seven bacteriophage lambda 2001 recombinant clones provided at the start of the project by Dr Padua. It is *supE, hsdR*; thus, it is permissive for vectors carrying amber mutations and will modify but not restrict *Eco*K sites (Karn, *et al.*, 1980).

E. coli C600 and C600 Hfl$^-$: These strains were purchased from Stratagene and used as hosts for the cloning vector Lambda GT10 (Huynh *et al.*, 1985). Strain C600 Hfl$^-$ contains the high frequency lysogeny mutation *HflA150.* In *hfl* mutants, the bacteriophage lambda *c*II gene product is much more stable than usual, and it accumulates to high levels (Hoyt *et al.*, 1982). *c*II is a positive regulator of the *c*I gene, which promotes lysogeny. Non-recombinant Lambda GT10 clones are *c*I$^+$ and their lytic growth is repressed in C600 Hfl$^-$, since they undergo highly efficient lysogenisation. However, recombinant Lambda GT10 clones are *c*I$^-$, because foreign DNA insertion inactivates the *c*I gene. Recombinant clones form normal, clear, lytic plaques on C600 Hfl$^-$. Both recombinant and non-recombinant phages grow on C600, allowing an assessment of packaging efficiency.

2.1.5 CLONING VECTORS

pBR 322 (Fig. 2.1)

This prototypic plasmid vector (Bolivar *et al.*, 1977) which confers ampicillin and tetracycline resistance to successfully transformed cells was used in a few early experiments. The FMuLV probe provided by Dr Padua was cloned in this vector.

pBluescript (Fig. 2.2)

This versatile 2964 bp. plasmid (phagemid) vector, derived from pUC 19 (Yanisch-Perron *et al.*, 1985), was used extensively in sequencing and cloning experiments. The particular version used in all experiments was

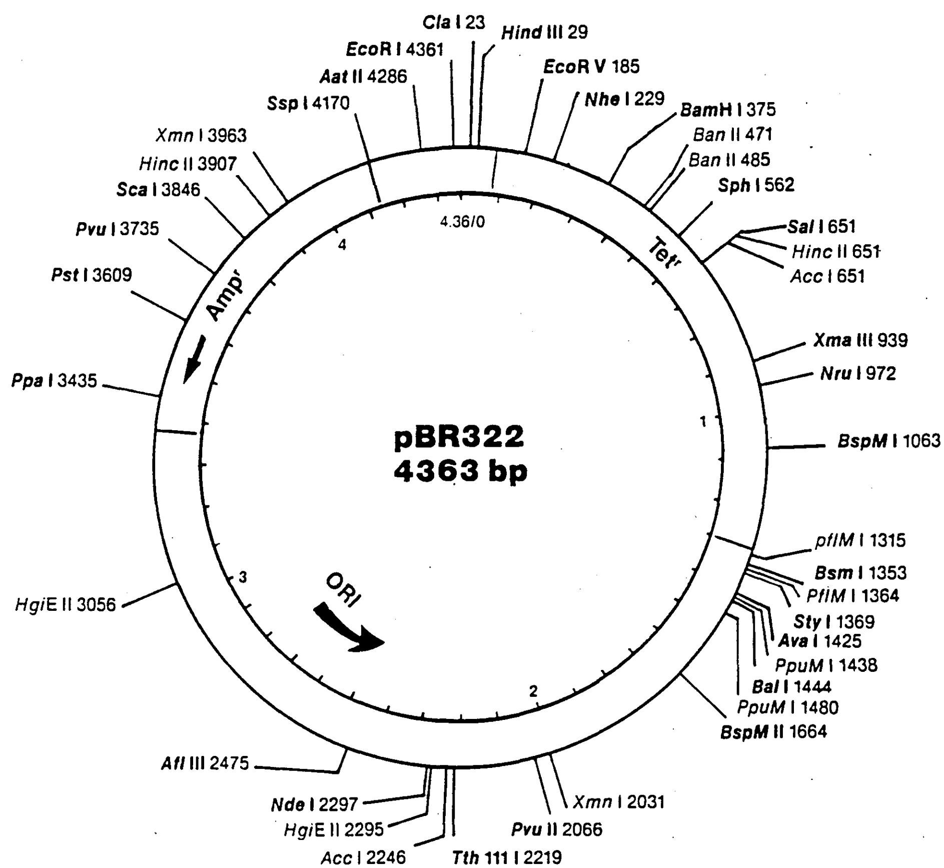
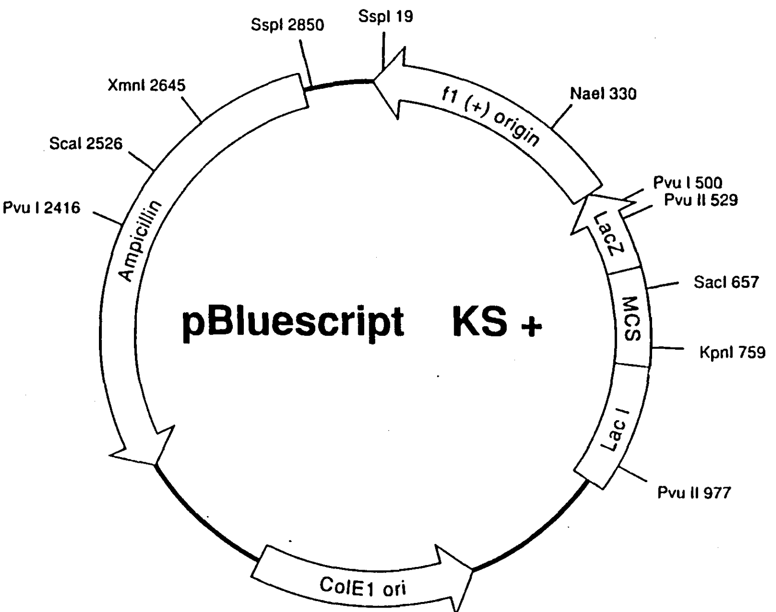


Fig. 2.1: pBR322. An *E. coli* cloning vector with two antibiotic resistance genes. The numbering system begins with the unique *Eco*RI site and continues clockwise.



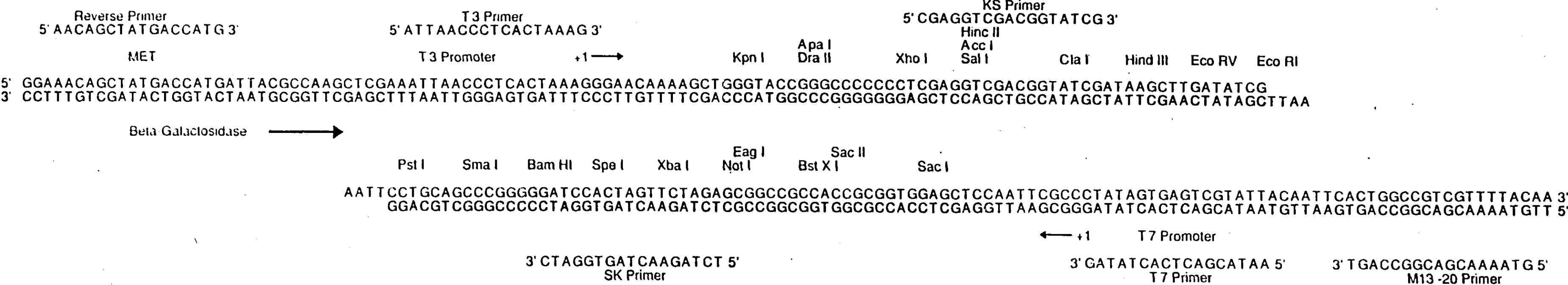


Fig. 2.2: pBluescript KS (+). A 2,958 bp. phagemid derived from pUC19. This vector confers ampicillin resistance and possesses a multiple cloning site (MCS) containing 21 unique restriction sites. This plasmid contains a *lacZ* promoter which allows blue/white colour selection of recombinants when the plasmid is grown in a suitable *E. coli* strain.

pBluescript KS (+). The KS designation indicates that the polylinker is orientated so that Beta-galactosidase transcription proceeds through the *Kpn*I site first and the *Sst*I site last. The (+) designation indicates that the f1 phage origin is directed towards the ampicillin resistance gene. pBluescript phagemids replicate autonomously as plasmids, thus bacterial colonies rather than phage plaques are obtained after transformation. Presence of pBluescript confers ampicillin resistance. The vector incorporates an extensive polylinker with 21 unique restriction endonuclease recognition sites. Since the polylinker is within the *lacZ* gene fragment and DNA insertion disrupts this gene, blue/white selection for recombinant plasmids is possible when pBluescript is grown in an appropriate *LacZ* delta M15 bacterial strain. Sequencing is facilitated using this vector, since the polylinker region incorporates six sites to which commercially-available sequencing primers (M13, reverse M13, T3, T7, SK and KS) will anneal.

pIC20H

This 2.7 kbp. vector is a derivative of pBR 322 (Marsh *et al.*, 1984). It confers ampicillin resistance and has a polylinker in the *lac* gene fragment, allowing blue/white selection of recombinant transformed clones.

Bacteriophage Lambda 2001 (Fig. 2.3)

This lambda replacement vector has an insert capacity of 9 to 23 kbp. (Karn *et al.*, 1984). It is a development of Lambda 1059 (Karn *et al.*, 1980) with multiple cloning sites at both ends of the *red+/gam+* stuffer fragment. The stuffer fragment is replaced by foreign DNA during cloning experiments. At the start of this project Dr Padua kindly provided seven recombinant bacteriophage lambda 2001 clones containing canine 3132 DNA. These clones had been selected from a genomic DNA library on the basis of hybridisation to a FMuLV probe.

Bacteriphage Lambda GT10 (Fig. 2.4)

This cloning vector was purchased from Stratagene. It will accept DNA fragment inserts up to 7.6 kbp. in length (Huynh *et al.*, 1985). It contains a unique *Eco*RI site in the *c*I imm^{434} gene. This site is used for cloning and allows selection between *c*I^{+}imm^{434} and *c*I^{-}imm^{434} phages when they are grown in an *E. coli* strain carrying the high frequency lysogeny mutation *HflA150* (see discussion above). After ligation, recombinant phage genomes were packaged using "Gigapack II Gold" packaging extract from Stratagene.

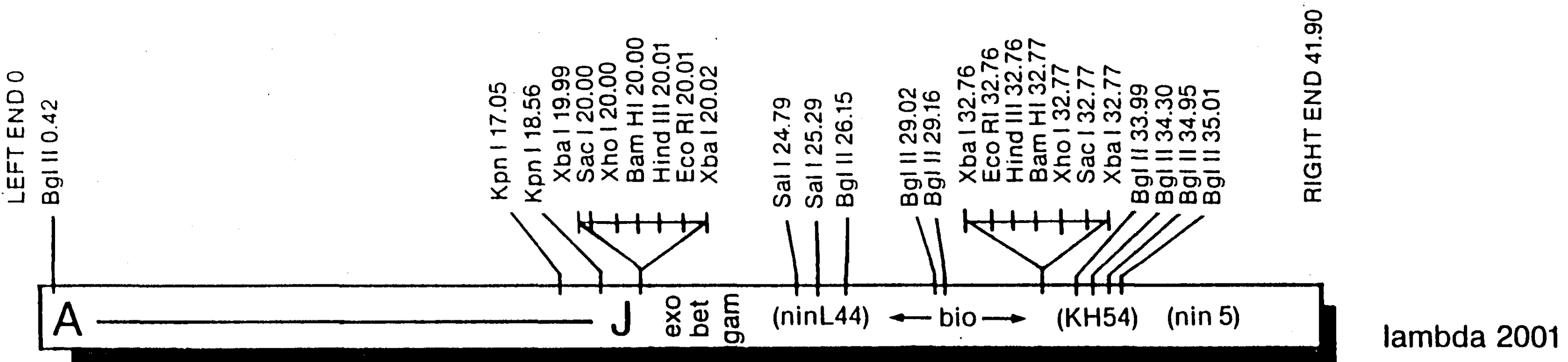


Fig. 2.3: Bacteriophage lambda 2001 genome (Karn *et al.*, 1984). A lambda replacement vector with an insert capacity of 9 to 23 kbp. Multiple cloning sites are present at both ends of the *red*+/*gam*+ stuffer fragment. Figures shown adjacent to restriction sites represent kbp. from the left end of the genome.

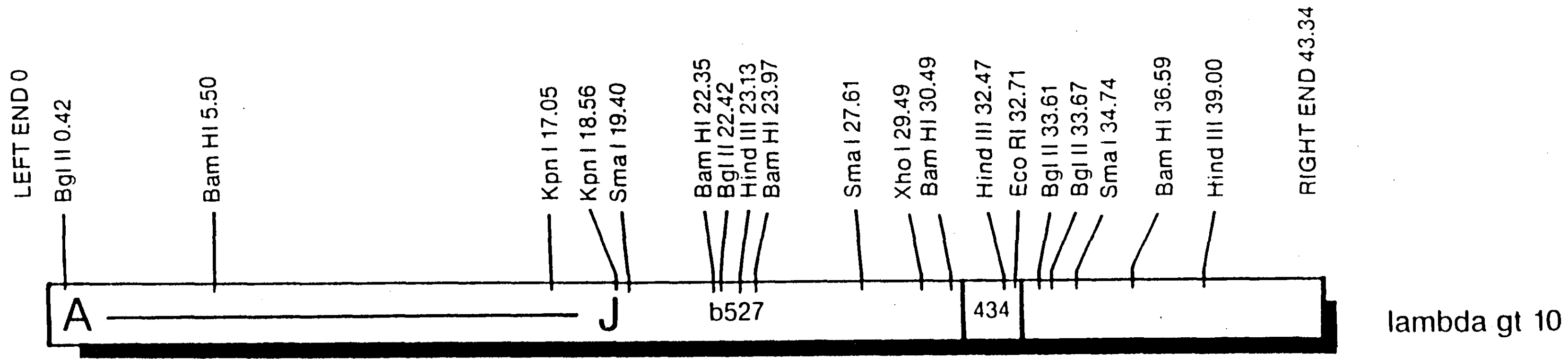


Fig. 2.4: Bacteriophage lambda GT10 genome (Huynh *et al.*, 1984). An insertion vector which will accept DNA fragments up to 7.6 kbp. in length. Figures shown adjacent to restriction sites represent kbp. from the left end of the genome

2.1.6 RESTRICTION ENDONUCLEASES

All were from Gibco BRL except for *Spe*I from Northumbria Biologicals Ltd. All were supplied and used with appropriate reaction buffers.

2.1.7 OTHER ENZYMES

All were from Gibco BRL., The Boeringer Mannheim Corp., or The Sigma Chemical Co., except:-

Klenow fragment- Amersham International plc. (in their Multiprime DNA labelling system).

T7 DNA polymerase- Pharmacia Ltd. (in their T7 DNA sequencing kit).

T4 DNA Ligase- Stratagene (in their DNA ligation kit).

2.1.8 BUFFERS, SOLUTIONS & GROWTH MEDIA

Where mentioned, autoclaving was carried out at 121°C for 20 minutes.

ANTIBIOTICS

Ampicillin: sodium ampicillin at 100mg/ml in distilled water sterilised by filtration and stored in 0.5ml aliquots at -20°C. Used at a final concentration of 50ug/ml.

Gentamycin sulphate: purchased as a sterile 10mg/ml aqueous solution from Gibco BRL. Stored at +4°C.

Penicillin G: purchased as a 10,000iu/ml sterile solution from Gibco BRL. Stored at +4°C.

Streptomycin: streptomycin sulphate prepared as a 50mg/ml solution in distilled water, sterilised by filtration and stored in 0.5ml aliquots at -20°C. Used at a final concentration of 50 ug/ml.

Tetracycline: tetracycline hydrochloride as a 12.5mg/ml solution in ethanol/water (50% v/v) sterilised by filtration and stored in 0.5ml. aliquots at -20°C away from light. Used at a final concentration of 15ug/ml.

BUFFERS & SOLUTIONS

Acid citrate dextrose solution B (ACD): 0.48g citric acid, 1.32g sodium citrate, 1.47g glucose, distilled water to 100ml. Sterilised by filtering. Used 1 part ACD to 6 parts fresh blood as an anticoagulant.

DNA Gel Loading Buffer (6x): 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in water. Stored at 4°C.

Dulbecco's Phosphate-buffered saline (PBS): Calcium and magnesium-free. Purchased as 10x liquid from GibcoBRL.

EDTA (0.5M, pH 8.0): 186.1g. disodium ethylene diamine tetraacetate.$2H_2O$ added to 800ml. of distilled water. NaOH pellets (approximately 20g.) added to make pH 8.0. EDTA will then dissolve. Volume adjusted to 1 litre with distilled water. Sterilised by autoclaving.
Ethidium bromide (10mg/ml): 1g. ethidium bromide made up to 1 litre with distilled water in a fume cupboard while wearing gloves. Stored away from light at $4^{o}C$. Ethidium bromide was also prepared as a 3mg/ml solution.
Foetal bovine serum: Virus screened, mycoplasma screened, bacteriophage screened. Purchased from Gibco BRL.
Formaldehyde gel loading buffer: 50% glycerol, 1mM EDTA (pH 8.0), 0.25% bromophenol blue, 0.25% xylene cyanol FF [for RNA electrophoresis].
Formamide gel loading buffer: 10ml formamide, 200ul 0.5M EDTA (pH 8.0), 10mg xylene cyanol FF, 10mg bromophenol blue [used for sequencing gels].
Hanks' Balanced Salt Solution (HBSS): purchased as 1x liquid from GibcoBRL.
L-Glutamine: purchased as a 200mM (100x) aqueous solution from Gibco BRL.
"Lytic Mix": 1% triton X-100, 50mM tris HCl, 60mM $Na_2EDTA.2H_2O$, pH adjusted to 8.0 with HCl.
Miniprep Solution 1: 50mM glucose, 10mM EDTA, 25mM tris HCl (pH 8.0). 4mg/ml lysozyme added just before use.
Miniprep Solution 2: 0.2M NaOH, 1% SDS (freshly prepared).
Miniprep Solution 3: 60ml of 5M potassium acetate, 11.5ml glacial acetic acid, 28.5ml distilled water. [This solution is 3M for potassium, 5M for acetate, pH 4.8]
MOPS buffer (10x): 200mM MOPS (3-[*N*-morpholino]propanesulphonic acid), 10mM EDTA, 50mM sodium acetate in distilled water, made pH 7.0 with 10M NaOH. Sterilise by filtration. Goes straw-coloured over time. [Dark brown solutions were not used].
NaCl (5M): 292.2g. NaCl made up to 1 litre with distilled water and sterilised by autoclaving.
NaOH (10M): 400g. NaOH pellets made up to 1 litre with distilled water.
Phenol (Equilibrated, neutral): 250g. redistilled phenol was melted at $68^{o}C$ and repeatedly extracted with an equal volume of 1M tris HCl (pH 8.0) and then 0.1M tris HCl (pH 8.0) until the pH of the aqueous phase was > 7.6. Stored in aliquots at $-20^{o}C$ away from light.
Phenol (Equilibrated, acid): same as equilibrated (neutral) phenol except 50mM sodium acetate (pH 4.0) was used in place of tris HCl (pH 8.0).
RPMI 1640: purchased as 1x sterile liquid from GibcoBRL.
SDS (20%): 200g. sodium dodecyl sulphate made up to 1 litre with distilled water and sterilised by filtering.

SM: 5.8g. NaCl, 2g. $MgSO_4.7H_2O$, 50ml. 1M Tris.Cl (pH 7.5), 5ml. 2% gelatin made up to 1 litre with distilled water. Sterilised by autoclaving. [Used for phage storage and dilution.]
SSC (20x): 175.3g. NaCl and 88.2g. sodium citrate dissolved in 800ml. distilled water. pH adjusted to 7.0 with a little 10M NaOH. Volume adjusted to 1 litre with distilled water and sterilised by autoclaving. [1xSSC is 150mM NaCl, 15mM sodium citrate]
STE: 10mM tris HCl, 1mM EDTA, 100mM NaCl, pH 8.0.
TAE (50x): 242g. tris base, 57.1ml. glacial acetic acid, 100ml. 0.5M EDTA (pH 8.0) made up to 1 litre with distilled water. Used at a final concentration of 1x. [1xTAE pH 8.15 is 40mM tris-acetate, 2mM EDTA]
TBE (10x): 108g. tris base, 55g. boric acid, 40ml. 0.5M EDTA (pH 8.0). Made up to 1 litre with distilled water and pH adjusted to 8.3. Sterilised by autoclaving. Used at a final concentration of 1x. [1xTBE pH 8.3 is 89mM tris-borate, 89mM boric acid]
TE: 10mM tris HCl, 1mM EDTA, pH adjusted as appropriate [made up as a 100X solution and sterilised by autoclaving].
Tris (1M): 121.1g. tris base dissolved in 800ml of distilled water. pH adjusted as required with concentrated HCl. Made up to 1 litre with distilled water and sterilised by autoclaving.
Trypsin: purchased as 1x liquid (0.25%, 1:250) from Gibco BRL.

BACTERIAL GROWTH MEDIA

L agar: L broth containing 1.5% agar. Sterilised by autoclaving.
LB medium (L broth): 10g. tryptone, 5g. yeast extract and 10g. NaCl made up to 1 litre with distilled water. pH adjusted to 7.5 with NaOH. Sterilised by autoclaving.
Maltose (20% w/v): 20g. maltose made up to 100ml. with distilled water. Sterilised by filtering. [0.2% maltose was added to medium during growth of bacteria to be used for plating bacteriophage lambda. This was because maltose induces expression of the maltose operon, which includes the gene (*lamB*) coding for the bacteriophage lambda receptor]
NZCYM Medium: 10g. NZ amine (casein hydrolysate type A), 5g. NaCl, 5g. yeast extract, 2g. $MgSO_4.7H_2O$, 1g. casamino acids made up to 1 litre with distilled water. pH adjusted to 7.5 with NaOH. Sterilised by autoclaving.
NZY Medium: the same as NZCYM medium without the casamino acids.
Soft Agarose (Top Agarose): L broth or another growth medium containing 0.7% agarose. Sterilised by autoclaving.

HYBRIDISATION REAGENTS

Denhardt's Solution (100x): 2% bovine serum albumin (fraction V), 2% ficoll, 2% polyvinylpyrrolidone in distilled water. Sterilised by filtering and stored in 25ml. aliquots at -20°C.

Formamide: purchased from Fluka Chemie AG and de-ionised by mixing 50ml of formamide with 5g. of Amberlite MB1 ion-exchange resin. Filtered through Whatman No. 1 filter paper to remove filter beads. Stored in 1ml aliquots at -20°C.

Salmon testis DNA (10mg/ml): 1g of desiccated salmon testis DNA (Sigma type III sodium salt) was dissolved in 100ml of distilled water by stirring for 2-4 hours at room temperature with a magnetic stirrer. The DNA was sheared by passing several times through a 19 guage needle and was then boiled for 10 minutes. It was stored in 1ml aliquots at 20°C until needed. Immediately before use in hybridisation experiments the DNA was boiled for 5 minutes before being quickly chilled on ice.

Southern Hybridisation buffer:

FINAL CONCENTRATION	AMOUNT FOR 1 LITRE
25% (or 50%) formamide	250ml (or 500ml)
5x Denhardt's Solution	50mlof 100x
0.1% SDS	5ml of 20%
50mM tris pH 7.4	50ml of 1M
10mM EDTA	20ml of 0.5M
3xSSC	150ml of 20x
10% dextran sulphate	200ml of 50%

2.1.9 PROVIDED RECOMBINANT DNA CLONES

Lambda 2001; 1-7:

Seven recombinant lambda 2001 bacteriophage clones, designated L1 to L7, containing canine DNA inserts were kindly provided by Dr R. Padua. The inserts were of 12 to 16 kbp. in length and were derived from a *Sau*3A partial digest of genomic DNA from canine lymphoma cell line 3132. The inserts were cloned into *Bam*HI-digested lambda 2001 arms.

FMuLV.2(pBR322):

Dr Padua kindly provided a clone which consisted of an unusual arrangement of two pBR322 vectors joined in tandem array at one *Eco*RI site; with a single, permuted, full length FMuLV provirus between them as an *Eco*RI insert at their other, free ends. The FMuLV insert was orientated so that its right hand end (as shown in Fig. 3.2) was adjacent to position 0 of one of the vector molecules and its left end adjacent to position 4,361 of the other.

The FMuLV insert was re-cloned as a 1:1 construct in pBR322 and was later cloned in pIC20H and pBluescript.

MMTV.pAT153 (5' probe):

Dr Clive Dickson provided a clone containing a MMTV *gag-pol* domain. The insert was a 3.2 kbp. *Pst*I-*Eco*RI fragment. The recombinant plasmid conferred only tetracycline resistance to host bacteria, since utilisation of the *Pst*I site insertionally inactivates the ampicillin resistance gene of pAT vectors.

pMT2:

This clone was provided by Dr Ruth Jarrett. It consisted of a pUC vector containing a 9 kbp. *Sst*I DNA fragment of HTLV-1 provirus from clone Lambda 23-3. *Sst*I cleaves the LTRs of this (and other) HTLV-1 clones freeing an almost full-length, unpermuted provirus.

BaEV.pBR322:

This clone was obtained indirectly. The vector contained a 9 kbp. *Eco*RI insert which consisted of a full-length, permuted M7 baboon endogenous retroviral provirus, but with only one LTR. Subsequent work described in Chapter 6 showed that it also contains 286 bp. of canine genomic DNA, presumably transduced during passage through a canine cell line.

2.1.10 CELL LINES

3132:

This cell line was derived from ascitic fluid of a 3 year old male Belgian shepherd dog with multicentric lymphoma (Strandstrom and Rimaila-Parnanen, 1979). It has been reported to produce a retrovirus (Strandstrom and Bowen, 1982). The cells were maintained in RPMI 1640 (Gibco BRL)/20% Foetal Bovine Serum (FBS)/4mM glutamine.

A-72F:

These cells originated from a tumour of the thigh of an 8 year old female golden retriever (Binn, 1980). A histological diagnosis was not obtained, but the dog had a previous history of mammary adenocarcinoma. The cells, however, have a fibroblastic morphology and develop cytopathic effects or demonstrate haemadsorption after infection with canine coronaviruses, canine adenovirus types

I & II, canine herpesviruses, canine parainfluenza virus and canine parvovirus. Cells were maintained in RPMI 1640 (GibcoBRL)/20% FBS/4mM glutamine.

A-72E:

Details of the history of this cell line are very poor. It arose during repeated passage of A-72F. The cells demonstrate epithelioid morphology and produce a retrovirus. Cells were maintained in RPMI 1640 (GibcoBRL)/20% FBS/4mM glutamine. Work described in Chapter 4 shows that A72E is not, in fact, a canine cell line and is expressing FeLV-B. Presumably it arose by contamination of A72F with FeLV-infected cells of another species.

MDCK (Madin-Darby canine kidney cells):

Derived in 1958 from a kidney of an apparently normal adult female cocker spaniel. The initial cells appeared fibroblastic, but since the 6th passage, the cells have maintained an epithelioid morphology. Maintained in Eagle's MEM with Earle's BSS/10% FBS.

MRC-5:

Normal diploid human fibroblasts. Maintained in Eagle's basal medium in Hanks' buffered salt solution with 10% FBS or in RPMI 1640/20%FBS/4mM glutamine.

2.1.11 CANINE TISSUE BIOPSIES AND POST MORTEM MATERIAL

Biopsy samples of canine tissue were obtained from the Department of Veterinary Surgery, University of Glasgow. The majority of these samples were lymph nodes from dogs with a tentative (and subsequently histologically confirmed) diagnosis of lymphoma. Biopsy material was taken into liquid nitrogen immediately after surgery and stored in liquid nitrogen, or at -70°C. Post mortem material was from the Department of Veterinary Pathology, University of Glasgow. Most dogs had been dead for less than one hour when examined. Samples of normal kidney, normal brain and lymphomatous tissues (lymph nodes, spleen, thymus, skin) were taken into liquid nitrogen and stored at -70°C.

2.2 GROWTH AND MANIPULATION OF BACTERIOPHAGES & CELLS

Unless otherwise stated, centrifugation was carried out at room temperature.

2.2.1 GROWTH AND MANIPULATION OF BACTERIOPHAGES

A). GROWTH OF LAMBDA 2001 CLONES

Preparation of Plating Bacteria

Q358 plating cells were streaked out from a glycerol stock onto an L-agar plate and grown up overnight at 37°C. A single colony was picked and inoculated into 100ml of NZCYM medium supplemented with 0.2% maltose in a 500ml flask. The cells were grown overnight at 37°C and harvested the next morning by centrifugation at 4000g for 10 minutes. The bacterial pellet was resuspended in a sufficient volume (usually about 20ml) of 10mM $MgSO_4$ to give an optical density ($O.D._{600}$) of 2.0. The cell suspension was stored at 4°C for up to 48 hours.

Plating, Quantitating and Plaque Purifying Phage Stocks

The seven phage stocks provided by Dr Padua (L1 to L7) were each serially diluted 10-fold in SM to provide 10^{-1}, 10^{-2} and 10^{-3} fold dilutions. 0.1ml of each dilution was added to a 1.5ml microcentrifuge tube containing 0.1ml of the plating cell suspension. Cells and phage were mixed by gentle vortexing and incubated at 37°C for 20 minutes. 3ml aliquots of melted NZCYM top agarose supplemented with 0.2% maltose were added to 15ml test tubes and held at 47°C in a waterbath. After the 20 minute incubation, each mixture of plating cells and adsorbed phage was added to an aliquot of top agarose, shaken (avoiding bubbles) and poured onto an L agar plate. The top agarose was swirled to encourage even spread. After 10 minutes at room temperature to allow hardening of the top agarose, the plates were inverted and incubated overnight at 37°C. Phage plaques were counted after 12-16 hours, and the number of plaque forming units (pfu)/ml of the original phage stocks was calculated. Single plaques were removed from each plate along with a plug of agar using the open end of a glass pasteur pipette. Each plaque "plug" was added to 1 ml of SM in a microcentrifuge tube and held for an hour at 4°C to allow the bacteriophages to resuspend. Serial 10-fold dilution of the new phage stock, plating and quantitation was then repeated. This process of single plaque purification was carried out twice.

Large scale Preparation of Bacteriophage Lambda

The following procedure was carried out for each of the seven plaque-purified Lambda 2001 clones in turn. It is essentially the technique described by Sambrook *et al.* (1989a). Plating cells were prepared as described above and four aliquots of 10^{10} cells were made (assuming 1 $O.D._{600}$ = $8x10^8$ cells/ml). The cells were centrifuged at 4000g for 10 minutes and resuspended in 3ml SM. $5x10^7$ pfu of phage was added, the mixture was vortexed and incubated at 37^oC for 20 minutes. Each aliquot of infected plating cells was then added to 500ml of pre-warmed (37^oC) NZYCM broth in a 2 litre, ribbed conical flask. The culture was shaken vigorously overnight. The following morning, the cultures were inspected for bacterial lysis. If lysis was not apparent, 10ml of chloroform was added to the culture, and incubation was continued for a further 30 minutes. This was found to be necessary for three of the seven cultures (L1, L2 and L5).

[PURIFICATION OF BACTERIOPHAGE LAMBDA FROM THE LYSED CULTURE AND EXTRACTION OF PHAGE DNA WILL BE DESCRIBED BELOW IN SECTION 2.3.1.]

B). LAMBDA GT10 LIBRARY CONSTRUCTION

Materials

Bacteriophage Lambda GT10 *Eco*RI digested arms (non-phosphorylated), T4 DNA ligase and packaging extract (Gigapack II Gold) were purchased from Stratagene along with appropriate host bacterial strains (C600 and $C600Hfl^-$) and reaction buffers. High molecular weight genomic DNA was extracted from the kidney of a normal adult male greyhound. 100ug of this canine DNA was digested with *Eco*RI and subjected to electrophoresis through LMP agarose. DNA fragments in the size range 6-7kbp. were excised from the gel and extracted (see section 2.3.5 for method).

Ligation

After resuspension of the size-selected canine DNA fragments in TE (pH 8.0) and adjustment to 0.1ug/ul, 0.15ug was added to 1ug (1ul) of Lambda GT10 arms. This approximates an equimolar ratio of arms to insert. 0.5ul 10x ligation buffer (0.5M tris HCl, pH 8.0; 70mM $MgCl_2$; 10mM dithiothreitol), 0.5ul 10mM rATP (pH 7.5) and 0.5ul T4 DNA ligase (2-3 Weiss units) was added and the volume adjusted to 5ul. The ligation reaction was incubated at 4^oC overnight.

Packaging

A set of packaging extracts (sonic and freeze-thaw) was removed from the -70°C freezer. The freeze-thaw extract was held on dry ice, while the sonic extract was allowed to thaw. The freeze-thaw extract was partially thawed between a finger and thumb and 1ul of the ligation reaction was *immediately* added. 15ul of sonic extract was quickly added and mixed in by gentle stirring, avoiding bubbles. The reaction was incubated at room temperature for 2 hours, after which 0.5ml SM and 20ul chloroform were added and gently mixed in. The tube was centrifuged briefly (10,000g, 10 seconds) to sediment debris.

Plating Out

Single colonies of host strains C600 and C600Hfl- were inoculated separately, each into 50ml L broth supplemented with 0.2% maltose and 10mM $MgSO_4$ and grown overnight at 30°C, rather than 37°C, to avoid overgrowth. Plating cells were prepared from the two host strains as described for Lambda 2001 except they were diluted to $O.D._{600}$ = 0.5. 200ul plating cells were infected with 1ul of the packaging reaction product. Infected cells were incubated and plated out as described above, but onto NZY plates. Non-recombinant phage grew as cloudy plaques on C600 and did not grow on C600Hfl-. Recombinant phages grew as clear plaques on both C600 and C600Hfl-.

Library Amplification

C600Hfl- plating cells were prepared as described immediately above. 2.5ml of plating cells were infected with 150,000 recombinant phage pfus and plated out in 25ml of top agarose onto a 22.5cm x 22.5cm NZY culture plate. After 6 hours at 37°C to allow phage growth, the top agarose was overlaid with 30ml SM and rocked gently at 4°C overnight to allow diffusion of phage into the SM. The bacteriophage suspension was harvested and chloroform was added to 5% (v/v). After 15 minutes at room temperature cell debris was removed by centrifugation (2000g, 5 minutes), the supernatant was harvested and chloroform was added to 0.3% (v/v). The titre (number of pfu/ml) of the amplified library was checked by serial dilutions and plating out. The library was stored in aliquots at 4°C.

2.2.2 GROWTH AND MANIPULATION OF BACTERIA

Production of Competent Cells

0.5 ml of an overnight culture of *E. coli* strain DS941 was inoculated into 50ml of L broth plus 50ug/ml of streptomycin sulphate. The culture was shaken vigorously at 37°C until the O.D.$_{600}$ was approximately 0.4. This usually took 1.5 to 2 hours. The bacterial cells were harvested by centrifugation (4000g, 5 minutes) and resuspended in one half the original volume (25ml) of ice cold 100mM $MgSO_4$. The cells were immediately recentrifuged and resuspended in 25ml ice cold 50mM $CaCl_2$. After further centrifugation the cells were resuspended in one tenth the original volume (5ml) of ice cold 50mM $CaCl_2$. 0.1ml aliquots of competent cells were dispensed into 1.5ml microcentrifuge tubes and held on ice for immediate use in transformation experiments.

Transformation of Cells

Up to 40ng of DNA in 1-50ul of TE or ligation buffer was added to 0.1ml of competent cells (see above) and the mixture was held on ice for 15 minutes. The cells were then "heat shocked" at 37°C for 5 minutes, before being returned to ice for 1 hour. 1ml of L broth was added to each tube and the tubes were transferred to a 37°C waterbath for 1 hour to allow expression of plasmid-encoded antibiotic resistance. The cells were centrifuged (12,000g, 1 minute) and washed twice with 1ml L broth before final resuspension in 0.1ml L broth. This volume was plated out on appropriate selective culture plates.

Screening and Selection techniques

Transformed cells expressing antibiotic resistance were selected from non-transformed cells by growth on appropriate selective (antibiotic-containing) L agar plates. When a host/vector combination allowing blue/white selection of recombinant from non-recombinant plasmids on the basis of intact B-galactosidase activity was used, then IPTG (isopropyl B-D thiogalactopyranoside) and X-Gal (5-bromo, 4-chloro, 3-indolyl B-D galactoside) were incorporated in the L agar. IPTG is an inducer of the *lac* promoter and X-gal is a substrate of B-galactosidase which produces a blue colour upon hydrolysis. Cells transformed by non-recombinant plasmids have intact B-galactosidase function and grow as blue colonies. In cells transformed by recombinant plasmids, the *LacZ* gene fragment

in the plasmid is disrupted by DNA insertion and B-galactosidase activity is absent. These cells grow as white colonies and can be selected on this basis.

The most frequently used host/vector combination was *E. coli* DS941 cells transformed with pBluescript KS (+) vector. This combination was grown on L agar plates containing 50ug/ml streptomycin (to which the DS941 strain is innately resistant), 50ug/ml ampicillin (to which pBluescript KS (+) confers resistance), 50ug/ml X-gal and 20mM IPTG.

2.2.3 GROWTH AND MANIPULATION OF MAMMALIAN CELLS

Maintenance

Mammalian cells were maintained in disposable plasticware at 37°C in 95% air, 5% CO_2. Suspension cells, such as cell line 3132, were grown up to a density of $5x10^6$ cells/ml before splitting 1:6 to 1:10. This was done approximately weekly. Adherent cells were grown in 0.2ml medium/cm^2 surface area until they reached sub-confluence. Cells were then washed in HBSS and trypsinised with 0.25% trypsin EDTA before replating in fresh medium at the correct density. A72F and A72E cells were passaged twice weekly at 1:3 to 1:6. MRC-5 cells were split 1:2 to 1:3 twice weekly. MDCK cells grew more slowly and were split 1:3 approximately weekly.

Long term storage

After trypsinisation and centrifugation, cells were washed in growth medium. After repeat centrifugation, cells were resuspended at approximately 2-$5x10^6$/ml, in RPMI 1640/20% FBS/10% dimethyl sulphoxide (v/v). The cells were cooled in the vapour phase of a liquid nitrogen cannister for a minimum of two hours, before immersion in the liquid phase of liquid nitrogen. After long term storage, frozen cells were rapidly thawed, by immersion of their plastic container in sterile water at 37°C.

Harvesting the A72E virus

Virus harvests were taken 2 or 3 days after splitting A72E cells at 1:3 when the culture was just sub-confluent. 20ml of medium was removed from the culture and filtered through a 0.45um filter. The filtered medium was then centrifuged at 10,000g for 15 minutes and the cleared supernatant was either used immediately to infect MRC-5 cells or stored at -70°C for RDDP assay.

Infection of MRC-5 cells with the A72E virus

Day 0: MRC-5 cells were seeded at $5x10^5$ in 5ml of medium in a $25cm^2$ bottle.

Day 1: The medium was removed and replaced with 2ml of filtered A72E supernatant (RDDP assay positive). Polybrene was added to 4ug/ml. The culture was incubated at $37^{O}C$ for 2 hours. Fresh medium was added to 10ml.

Day 3: The cells were washed in HBSS, trypsinised and transferred into a $75cm^2$ bottle.

Days 4-21: The cells were fed and split as required while the infection spread through the culture.

Characterisation of the A72E virus

An ELISA test for detection of FeLV antigen was carried out in the laboratory of Professor O. Jarrett. Subgroup determination was carried out by M. Golder in that laboratory by a reverse interference test. FEA cells (non-established primary feline embryo fibroblasts; Jarrett *et al.*,1973) were infected with the A72E virus and challenged with three types of viral particle consisting of the genome of Moloney murine sarcoma virus pseudotyped with FeLV envelopes of sub-groups A, B or C (Russell and Jarrett, 1976).

Strategies designed to induce RNA expression of endogenous retroviruses

The pyrimidine analogues 5-iododeoxyuridine (IdU) and 5-azacytidine (5-AZA) have been reported to induce expression of endogenous retroviral sequences (IdU: Lowy *et al.*, 1971; Teich *et al.*, 1973; 5-AZA: Groudine *et al.*, 1981). 5-AZA works by a mechanism which is thought to involve inhibition of DNA methylation (Jones and Taylor, 1980). IdU was used at a concentration of 20ug/ml and 5-AZA at 3uM in all experiments.

Canine cell lines 3132, MDCK and A72F were split into $75cm^2$ bottles, allowed to grow for 24 hours and then exposed to IdU or 5-AZA for 48 hours (in the case of IdU, in darkness) before washing in HBSS and providing fresh medium. After a further 48 hours, or when the cells were just sub-confluent, they were washed in PBS and total RNA was collected by the guanidinium thiocyanate-phenol-chloroform extraction method (*q.v.*).

2.3 EXTRACTION, PURIFICATION AND SEPARATION OF NUCLEIC ACIDS

2.3.1 Preparation of Bacteriophage Lambda DNA

Purification of Bacteriophage particles from the lysed culture

The method described here is a modification of that of Yamamoto *et al.*, (1970). The fully-lysed culture was allowed to cool to room temperature and crude pancreatic DNase and RNase were added, each to 1ug/ml. After incubation at room temperature for 30 minutes, NaCl was added to 1M and dissolved by swirling. The mixture was held on ice for 1 hour and then centrifuged at 11,000g for 10 minutes to remove debris. To the pooled supernatant in a clean flask, 10% w/v polyethylene glycol (PEG 6000) was added and dissolved by gentle magnetic stirring. Once the PEG had dissolved, the mixture was held on ice for 2 hours. Precipitated bacteriophage particles were then harvested by centrifugation at 11,000g, $4^{O}C$ for 10 minutes. The supernatants were discarded and the pellets gently resuspended in SM (16ml per litre of original lysed culture). The resuspended bacteriophages were mixed with an equal volume of chloroform, vortexed and centrifuged at 1600g for 15 minutes. The aqueous phase was retained and its volume measured. 0.75g/ml of cesium chloride was added and dissolved. The bacteriophage suspension was added to cellulose nitrate tubes which were sealed and centrifuged in a Beckman 50Ti rotor at 38,000 rpm for 24 hours at $4^{O}C$.

After equilibrium centrifugation, bacteriophage particles were visible as a smokey blue horizontal band in the middle of the tube. The band was removed by side puncture of the tube using a needle and syringe.

Extraction of Bacteriophage Lambda DNA

The purified suspension of bacteriophage was dialysed against a 1000x volume of 10mM NaCl, 10mM $MgCl_2$, 50mM tris HCl (pH 8.0) for 1 hour. This was repeated with fresh buffer. EDTA (pH 8.0) was added to 20mM, proteinase K was added to 50ug/ml and SDS was added to 0.5% using stock solutions. The mixture was incubated at $65^{O}C$ for 1 hour. After incubation, an equal volume of equilibrated phenol was added and mixed in by repeated inversion rather than vortexing. The mixture was centrifuged at 1600g for 5 minutes and the aqueous (upper) phase was collected. The aqueous phase was extracted with an equal

volume of 50:50 (v/v) equilibrated phenol and chloroform and once again with chloroform. The aqueous phase was transferred to dialysis tubing and dialysed overnight against a 1000x volume of TE. The dialysis was repeated for 3 successive nights, then phage DNA was harvested from the dialysis tubing and stored at $4^{o}C$.

2.3.2 Preparation of Plasmid DNA

Mini-preparation

This is a modification of the alkaline lysis method of Birnboim and Doly (1979). A single bacterial colony carrying the plasmid of interest was inoculated into 5ml of L broth containing appropriate antibiotic(s) and shaken vigorously overnight at $37^{o}C$. The bacterial pellet was harvested after centrifugation (2500rpm, benchtop centrifuge; 10 minutes) and any remaining drops of supernatant were aspirated with a pipette. The pellet was resuspended in 100ul of miniprep solution 1 (50mM glucose, 10mM EDTA, 25mM tris HCl (pH 8.0); 4mg/ml lysozyme added just before use), transferred to a 1.5ml microcentrifuge tube and held at room temperature for 5 minutes. 200ul of ice cold, freshly-prepared miniprep solution 2 (0.2M NaOH, 1% SDS) was then added, the tube vortexed and held on ice for another 5 minutes. 150ul of miniprep solution 3 (60% 5M potassium acetate, 11.5% glacial acetic acid, 28.5% distilled water; v/v) was added and the tube held on ice for 5 more minutes.

The mixture was centrifuged (13,000rpm, eppendorf centrifuge, 5 minutes) to sediment cell debris and precipitated SDS. The supernatant was transferred to a fresh tube. An equal volume of 1:1 v/v equilibrated phenol/chloroform was added, the mixture vortexed and centrifuged as before. The upper aqueous phase was collected and transferred to a fresh tube. When miniprep DNA was intended for use in sequencing experiments, acid phenol followed by separate chloroform extraction was used in place of 1:1 (v/v) phenol/chloroform (Weickert and Chambliss, 1989). 2 volumes of 100% ethanol was added and mixed in. After 2 minutes at room temperature, the tube was centrifuged (13,000rpm, eppendorf centrifuge, 1 minute). The supernatant was discarded and the pellet was washed with 1ml of 70% ethanol and recentrifuged. Again the supernatant was discarded and the pellet was dried at $37^{o}C$ for 15 minutes. The pellet was resuspended in

50ul of TE containing 50ug/ml DNase-free pancreatic RNase. Typically, this procedure yielded several ug of plasmid DNA.

Large-scale preparation

Large scale plasmid preparation was carried out using a modified version (Hettle, 1985) of the SDS lysis method described by Sambrook *et al.* (1989b). Large scale plasmid preparations were carried out frequently at the beginning of this project; but later, when miniprep sequencing became reliable, most sub-cloning and sequencing experiments were carried out using acid phenol mini-preparations only.

A single bacterial colony carrying the plasmid of interest was inoculated into 200ml of L broth containing appropriate antibiotic(s) in a 500ml flask and shaken vigorously at 37°C overnight. The bacterial pellet was harvested by centrifugation of the culture in a 250ml polycarbonate flask (Beckman JS-7.5 rotor, 10,400g/7.5 krpm, 15 minutes, 4°C) and the supernatant was discarded. 2ml of 25% sucrose, 10mM EDTA (pH 8.0) was added, the pellet was dispersed and the mixture was transferred to a 30ml polypropylene tube. After 5 minutes on ice, 0.3ml of freshly-prepared 20mg/ml lysozyme was added and the mixture incubated on ice for 15 minutes. 2ml of 0.25M EDTA (pH 8.0) was added and the mixture was returned to ice for a further 5 minutes. 4ml of "lytic mix" (*q.v.*) was added and immediately mixed in gently by pipetting to avoid shearing of DNA. The tube was held on ice for a further 30 minutes and then centrifuged (Beckman JA-20 rotor, 46,000g/19.5krpm, 1 hour, 4°C) to sediment cell debris including chromosomal DNA. To the supernatant, 50ul of proteinase K (10mg/ml) was added and the mixture incubated at 55°C for 20 minutes. The volume of the solution was measured. For each 1ml, 1.035g CsCl and 68.32ul ethidium bromide (3mg/ml) was added. The CsCl was dissolved at 30°C and the solution was loaded into Beckman Quick-Seal tubes. Equilibrium centrifugation to produce a continuous CsCl-ethidium bromide gradient was carried out (Beckman VTi65 rotor, 49krpm, 16 hours, 15°C).

After centrifugation, 2 red bands were present close to the middle of the gradient. The lower band, consisting of closed circular plasmid DNA was more intense. This band was removed by careful side-puncture of the tube and

aspiration with a 21G needle into a syringe. Care was taken not to contaminate the preparation with any of the upper band, which is composed of linear (chromosomal) DNA and nicked circular (plasmid) DNA. The volume of the plasmid solution was measured and 3 volumes of TE were added. Ethidium bromide was removed by repeated extractions with an equal volume of 2-butanol until the pink colour had completely disappeared from both the lower aqueous and upper organic phases. The lower aqueous, plasmid-containing phase was retained and dialysed for several hours against repeatedly changed 1000x volumes of TE (pH 8.0).

The DNA was precipitated by addition of 2 volumes of 100% ethanol, 0.1 volume 5M NaCl and held at -20°C for 1 hour. After centrifugation (10,000g, 15 minutes) the DNA pellet was washed in 70% ethanol, dried in a vacuum drier and resuspended in 0.5ml TE (pH 7.6) overnight. The $O.D._{260}$ of the final solution was measured and the DNA concentration calculated*. Purity was assessed by determination of the $O.D._{260}/O.D._{280}$ ratio. This was usually found to be approximately 1.8. If the ratio was less than 1.7, the DNA was reprecipitated, washed and resuspended.

*($O.D._{260}$ x Dilution Factor x 50 = DNA concentration in ug/ml).

2.3.3 EXTRACTION OF NUCLEIC ACIDS FROM MAMMALIAN TISSUES

EXTRACTION OF HIGH MOLECULAR WEIGHT GENOMIC DNA

Technique for tissue biopsies and suspension cells

0.1g of tissue or 10^7 cells were disrupted in 5ml PBS in a Stomacher. The suspension was centrifuged (benchtop centrifuge, 2.5krpm, 10 minutes, 4°C) and the pellet washed in 25ml PBS. After repeat centrifugation the pellet was resuspended in 25ml STE. NaCl was added to 0.4M, SDS to 0.5% and proteinase K to 50ug/ml. The mixture was incubated at 50°C for 1 hour then extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 v/v). To avoid shearing of the DNA, the two phases were mixed by 10 minutes of manual inversion of the tube rather than by vortexing. After centrifugation (benchtop centrifuge, 2.5krpm, 10 minutes) the upper aqueous phase was collected with a wide-bore pipette, transferred to a clean flask and extracted with an equal volume

of chloroform. The resulting aqueous extract was added to 2 volumes of 100% ethanol at room temperature and DNA was spooled out with a heat-sealed pasteur pipette. The DNA was washed with 70% ethanol, recentrifuged and the pellet was air dried at room temperature. 0.5-1ml TE (pH 8.0) was added and the DNA allowed to resuspend overnight at $4^{o}C$. The DNA was quantitated spectrophotometrically (For high molecular weight DNA, an $O.D._{260}$ of 1 is approximately equal to 50ug/ml). Purity was assessed by determination of the $O.D._{260}/O.D._{280}$ ratio, which was usually found to be approximately 1.8. If the ratio was less than 1.7, the DNA was reprecipitated, washed and resuspended.

Technique for cells grown in monolayer

Culture medium was discarded and the cells rinsed with an excess of PBS. $0.1ml/cm^2$ of a buffer consisting of 10mM tris HCl (pH 8.0), 10mM EDTA (pH 8.0), 100mM NaCl, 0.5% SDS and 250ug/ml proteinase K was added. The cells were incubated with gentle shaking for 2 hours at $37^{o}C$. After incubation the buffer became viscous because of dissolved cell components, including DNA. The aqueous phase was extracted with equilibrated phenol twice, phenol:chloroform twice and chloroform:isoamyl alcohol (24:1, v/v) once. DNA was precipitated with 2 volumes of ethanol and 0.1 volume of 5M NaCl and collected by centrifugation (5000g, 5 minutes). The pellet was washed twice with 70% ethanol, air dried and resuspended in TE containing 50ug/ml DNase-free pancreatic RNase.

Technique for blood lymphocytes or lymphoblasts

This technique was used to collect DNA from neoplastic lymphoblasts in two dogs with acute lymphoblastic leukaemia. Blood was collected into ACD and centrifuged at 1300g for 15 minutes. The buffy coat was collected with a wide bore pipette, resuspended in 10 volumes of HBSS and carefully layered onto 20ml Ficoll-Paque in a 50ml Falcon tube. After centrifugation (benchtop centrifuge, 2000 rpm, 20 minutes) mononuclear cells formed a layer at the HBSS-Ficoll interface. The cells were removed, diluted in one volume of HBSS and pelleted by centrifugation (benchtop centrifuge, 1000 rpm, 10 minutes). After resuspension in PBS, the cells were counted in a haemocytometer. 10^7 cells were pelleted and DNA was extracted as described above for tissue biopsies and suspension cells.

EXTRACTION AND PURIFICATION OF TOTAL RNA

When handling RNA, it is important to avoid degradation by RNases. Disposable plastic gloves were worn at all times. RNA manipulations were carried out in disposable, sterile plasticware, which is essentially free of RNases, or in glassware which had been treated with diethylpyrocarbonate (0.1% DEPC, 12 hours, $37^{o}C$) and then autoclaved. Solutions intended for use with RNA were made up using specially-designated chemicals which were handled only with baked spatulas. When possible, solutions were DEPC-treated and autoclaved. Since DEPC reacts with amines, tris-containing solutions could not be treated in this way.

RNA was isolated by the guanidinium thiocyanate-phenol-chloroform extraction method described by Chomczynski and Sacchi (1987). This method was conveniently carried out using the Biogenesis Ltd. product "RNAzol". Briefly, 0.1g of tissue or 10^7 suspension cells were macerated in 2ml "RNAzol". Cells grown in monolayer were lysed by direct addition of $0.15ml/cm^2$ "RNAzol" to the culture flask, before transfer to an appropriate-sized tube. 0.1 volume of chloroform was added to the solution which was shaken vigorously in a sealed tube for 15 seconds, then held on ice for 15 minutes. After centrifugation (12,000g, $4^{o}C$, 15 minutes) the upper aqueous phase was carefully transferred to a clean tube, mixed with an equal volume of isopropanol and stored at $-20^{o}C$ for 45 minutes. Precipitated RNA was harvested by centrifugation (12,000g, 15 minutes, $4^{o}C$) and washed twice with 1ml of 75% ethanol. After recentrifugation the pellet was air dried and resuspended in 50-100ul 1mM EDTA (pH 8.0). RNA was quantitated spectrophotometrically (For RNA, an $O.D._{260}$ of 1 is approximately equal to 40ug/ml). Purity was assessed by determination of the $O.D._{260}/O.D._{280}$ ratio, which was found to be approximately 1.9.

2.3.4 AGAROSE GEL ELECTROPHORESIS

DNA

DNA molecules of different sizes were separated by electrophoresis through 0.8% agarose in 1xTBE. A variety of perspex gel sets of different sizes with varying numbers of sample wells were used. The cast agarose gel was immersed in TBE buffer in the gel apparatus. Each DNA sample was mixed 5:1

(v/v) with DNA gel loading buffer (*q.v.*) and carefully loaded into one of the wells. This was repeated for each of the samples. When restriction endonuclease-digested genomic DNA was the sample, the DNA solution was heated in a waterbath (65°C, 10 minutes) prior to mixing with gel-loading buffer. This was intended to disrupt hydrogen bonds between molecules with compatible cohesive termini.

Electrophoresis was carried out at approximately 5V/cm, usually overnight. After electrophoresis, the separated DNA fragments were stained with ethidium bromide (0.5ug/ml in distilled water, 45 minutes), destained (distilled water, 45 minutes) and identified by direct examination of the gel using an ultraviolet transilluminator (Vilber Lourmat, 312 nm wavelength output). DNA fragments were photographed using Polaroid type 57 high-speed film, or occasionally with type 55 positive-negative film. Fragment sizes were determined by comparison with DNA size markers. The size markers most frequently used were *Hin*dIII or *Hin*dIII/*Eco*RI-digested bacteriophage lambda genomic DNA. When an accurate size estimation was required, a semi-log scale graph was produced with size marker fragment lengths on the log axis and migration distances on the linear axis.

When a rapid result was required, DNA samples were loaded onto a minigel apparatus, and electrophoresis was carried out at 20V/cm. Ethidium bromide was included in the gel mixture and buffer, so that gels could be photographed immediately after electrophoresis. Linear double stranded DNA runs approximately 15% slower in the presence of ethidium bromide, and the bands often appeared slightly ragged using this method.

RNA

All manipulations of RNA and solutions intended for use with RNA were carried out using disposable, sterile plasticware or DEPC-treated glassware, as described above. Electrophoresis tanks were cleaned with detergent, rinsed with water, dried with 100% ethanol and filled with 3% hydrogen peroxide. After 15 minutes at room temperature, the hydrogen peroxide was discarded and the tank thoroughly rinsed with DEPC-treated water.

A 350ml 1.2% agarose/formaldehyde gel was prepared by boiling 4.2g agarose in 304.5ml DEPC-treated water. The solution was cooled to 60°C before

adding 35ml 10x MOPS buffer and 10.5ml of 37% formaldehyde in a fume cupboard. The gel was cast in the usual way.

For each RNA sample the following premix was prepared in a microcentrifuge tube:

10x MOPS buffer	5.00ul
formaldehyde	8.75ul
formamide	25.00ul

RNA in 1mM EDTA, pH 8.0 (upto 10ug) was mixed in and the volume was adjusted to 50ul. The samples were incubated at 65°C for 5 minutes, chilled on ice and briefly centrifuged. 10ul of formaldehyde gel loading buffer was added and the samples were loaded onto the gel. 10ug of an RNA ladder (Gibco BRL) was treated in an identical way to the test samples and loaded into each of the outside lanes of the gel to serve as size markers. Electrophoresis was carried out in 1xMOPS buffer at 5V/cm for 3 hours, or until the bromophenol blue marker had migrated halfway through the gel. It was essential to run the gel quickly when using this method, because the concentration of the denaturant, formaldehyde, was lower than with many other methods. Overnight electrophoresis requires upto 5 times more formaldehyde in the gel mixture.

After electrophoresis, the section of gel to be transferred to nitrocellulose or nylon membrane was cut out and processed as described below in Section 2.5.2. On most occasions, samples were loaded onto the gel in duplicate, one half for Northern transfer and the duplicate half for ethidium bromide staining. To stain the duplicate sample lanes, that section of gel was rinsed in several changes of water and left to stand overnight in a large excess of water. The next day, the gel was stained with ethidium bromide (0.5ug/ml, 45 minutes) and photographed.

2.3.5 EXTRACTION OF DNA FROM LOW MELTING POINT (LMP) AGAROSE GELS

Preparation of LMP agarose gels and electrophoresis of DNA samples was carried out exactly as described above for normal agarose gels. DNA bands of interest were identified using minimum exposure to UV light, cut out of the gel using a scalpel blade and transferred to a clean tube. The segment of gel containing the DNA of interest was melted in a waterbath (65°C, 10 minutes) and

2 volumes of TBE (65°C) was added. The diluted sample was allowed to cool to 37°C in a waterbath for 10 minutes. An equal volume of equilibrated phenol was added and mixed in by thorough vortexing. After centrifugation (13,000rpm, eppendorf centrifuge, 5 minutes), the upper aqueous phase was collected. White flocculent material at the interface was carefully avoided. The sample was extracted once more with phenol and twice with chloroform before precipitation in 2 volumes of 100% ethanol, 0.1 volume 5M NaCl. After one hour at -20°C, the DNA was collected by centrifugation (13,000rpm, eppendorf centrifuge, 20 minutes), washed in 70% ethanol, dried and resuspended in an appropriate volume of TE (pH 7.6).

2.3.6 POLYACRYLAMIDE GEL ELECTROPHORESIS

Polyacrylamide gel electrophoresis was used to separate and accurately measure the size of DNA molecules less than 1 kbp in length. Acrylamide was used at concentrations between 3.5% and 20% (w/v) in TBE, depending upon the size of the DNA of interest. An acrylamide:*N,N'*-methylenebisacrylamide ratio of 20:1 was used. Polymerisation was achieved by addition of ammonium persulphate (0.08%, w/v) and TEMED (*N,N,N',N'*-tetramethylethylenediamine, 0.08%, v/v). 15 x 17cm gels were cast at a thickness of 1.5mm between sealed, vertical, siliconised glass plates. 20 well teflon combs were used. After removal of the comb and before the samples were loaded, the wells were thoroughly flushed with TBE to remove incompletely-polymerised acrylamide. DNA samples were mixed 5:1 (v/v) with DNA gel loading buffer. A total volume of 20ul was loaded. Gels were run in TBE buffer at approximately 5V/cm until the marker dyes had run the desired distance. DNA bands were detected and photographed after staining with ethidium bromide as described for agarose gel electrophoresis.

Denaturing polyacrylamide gel electrophoresis using 6% polyacrylamide gels incorporating 7M (42%, w/v) urea was used for DNA sequencing. This method is described below.

2.3.7 PREPARATION AND PURIFICATION OF SYNTHETIC OLIGONUCLEOTIDES

Oligonucleotides (17 to 21mers) were synthesised using an Applied Biosystems 381A DNA synthesiser. The 5' methoxytrityl protecting group was removed after incorporation of the last (5') nucleotide. Oligonucleotides were eluted from the synthesis column with 2ml concentrated (30%) ammonia solution. Elution was carried out over a 2 hour period. 0.25 ml increments of the ammonia solution were passed across the column from one 2 ml syringe to another. Each increment was left in contact with the column for 15 minutes. After elution, the oligonucleotide solution was transferred to an airtight tube and side-chain protecting groups were removed by overnight incubation at 55°C. The oligonucleotide was precipitated in 2 volumes of 100% ethanol and 0.1 volume of 5M NaCl, collected by centrifugation (13,000rpm, eppendorf centrifuge, 20 minutes), washed in 70% ethanol, dried and resuspended in distilled water to a concentration of 5ng/ul. Typical yields measured by spectrophotometry (calculated using a multiplication factor of 20ug/ml*) were approximately 1 mg.

*$O.D._{260}$ x Dilution Factor x 20 = DNA Concentration (ug/ml)

2.4 ENZYMATIC MANIPULATION OF NUCLEIC ACIDS

2.4.1 RESTRICTION ENDONUCLEASE DIGESTIONS

A variety of restriction endonucleases were used in the preparation and analysis of cloned and genomic DNA. Typically, 1-2ug of plasmid or bacteriophage DNA was digested in a reaction volume of 20ul. When working with genomic DNA, 10ug was digested in 50ul. Restriction enzymes were used at a concentration of 3U/ug of substrate DNA in the buffer provided by the manufacturer. Digestions were carried out at 37°C for at least 2 hours (cloned DNA) or 12 hours (genomic DNA). Genomic DNA and occasional samples of cloned DNA were digested in the presence of 3mM spermidine, which is reported to prevent endonuclease inhibition by polyanions (Bouche, 1981). When samples were to be subjected to electrophoresis, the reaction was terminated by the addition of 0.2 volumes of DNA gel loading buffer. Otherwise reactions were

terminated by heating to 70°C for 10 minutes or by the addition of 0.5M EDTA (pH 8.0) to a final concentration of 10mM. If required, digested DNA fragments were purified by phenol:chloroform extraction and ethanol precipitation.

2.4.2 DNA LIGATION

All ligations involved DNA molecules with either compatible cohesive or blunt termini. Purified, precipitated insert DNA was resuspended in an appropriate small volume of TE (pH 8.0) and mixed with vector DNA in TE. For plasmid vectors, the molar ratio of insert:vector was usually 10:1 and a reaction volume of 50ul was used. When ligating into lambda arms, an equimolar ratio of insert:vector was used in a total reaction volume of 5ul. 0.1 volume of 10x ligation buffer (0.5M tris HCl, pH 8.0; 70mM $MgCl_2$; 10mM dithiothreitol), 0.5ul 10mM rATP (pH 7.5) and 0.5ul T4 DNA ligase (2-3 Weiss units) was added and the volume adjusted with water to 5ul or 50ul as appropriate. The ligation reaction was incubated at 4°C overnight.

2.4.3 ^{32}P RADIO-LABELLING OF ds DNA

This was carried out using the Multiprime DNA labelling system from Amersham International plc. The method was introduced by Feinberg and Vogelstein (1983). It uses random sequence hexanucleotides to prime DNA synthesis from a denatured DNA template. DNA from any source can be used as the template.

Synthesis of DNA was carried out using [alpha-^{32}P]dCTP and unlabelled dATP, dGTP and dTTP. The Klenow fragment of *E. coli* DNA polymerase 1 was used to catalyse the reaction. This enzyme lacks 5' to 3' exonuclease activity, so that the radioactive product, once synthesised, was not degraded by exonucleolysis. Briefly, 25ul (25ng) of template DNA in distilled water (usually a restriction fragment purified from LMP agarose) was boiled in a waterbath for 2 minutes, then placed on ice. 10ul of multiprime buffer (dATP, dGTP and dTTP in a concentrated buffer solution containing tris HCl (pH 7.8), $MgCl_2$ and 2-mercaptoethanol) and 5ul of primer (random hexanucleotides in an aqueous solution containing nuclease-free bovine serum albumin) were added. 8ul of [alpha-^{32}P]dCTP (800Ci/mmol, 10uCi/ul) was added taking the usual precautions

observed when working with radioactive materials, followed by 2ul (2 units) of Klenow fragment (in 50mM potassium phosphate (pH 6.5), 10mM 2-mercaptoethanol and 50% glycerol). The reaction was incubated at room temperature for 5 to 16 hours. Unincorporated label was separated from incorporated by chromatography using a 20cm column of Sephadex G-50 (medium grade). Typically, this procedure yielded 500ul of probe solution containing 1.5 to 5.0 x 10^5 cpm/ul. The specific activity and probe length were not determined for each probe. Random priming typically yields probes with specific activities in the range 5 x 10^8 to 4 x 10^9 cpm/ug. Probe lengths vary enormously about a mean of approximately 250 bases. Probes were routinely heat-denatured in a 100°C waterbath for 2 minutes before use in hybridisation experiments.

2.4.4 SEQUENCING OF ds DNA & SEQUENCE DATA ANALYSIS

DNA was sequenced using the base-specific, dideoxy chain termination method (Sanger *et al.*, 1977) on alkali-denatured double stranded DNA templates. pBluescript KS (+) was used exclusively as the sequencing vector and [alpha-^{35}S]dATP alpha S as the label. During the course of the project, a variety of sequencing template preparation methods, DNA polymerases and apparatuses were used. Best results were obtained using acid phenol mini-preparation DNA templates (*q.v.*), bacteriophage T7 DNA polymerase (Pharmacia LKB) and the Pharmacia LKB Macrophor sequencing apparatus. Sequencing was primed using either purchased "universal primers" (M13, reverse M13, KS, SK, T7 & T3) or custom-synthesised 17 to 20mers. "Universal primers" were used when sequencing sub-cloned DNA sequences directly adjacent to the vector. Custom-synthesised primers were used when sequencing longer stretches of DNA.

At the start of the project, mixtures of dNTPs and dideoxy (dd) NTPs for sequencing were manually prepared using 0.5mM dNTP and 10mM ddNTP stock solutions, as described by Sambrook *et al.* (1989c). The optimum dNTP:ddNTP molar ratio depended upon the particular nucleotide termination reaction (A,C,G or T) and the sequencing enzyme in use. By the end of the project, commercially-available sequencing kits with pre-mixed buffers and solutions containing optimal (but unstated) dNTP:ddNTP ratios were available. The sequencing protocol

outlined below was slightly modified from that described in the Pharmacia T7 Sequencing handbook.

Denaturation of the double-stranded template

The concentration of template was adjusted to 0.25mg/ml. 8ul of template was mixed with 2ul of 2M NaOH and held at room temperature for 10 minutes. 3ul of 3M sodium acetate (pH 4.8) and 7ul of distilled water were mixed in. 60ul of 100% ethanol was added and the mixture placed on dry ice/ethanol for 15 minutes. Precipitated DNA was collected by centrifugation (13,000rpm, eppendorf centrifuge, 10 minutes), washed with 70% ethanol, recentrifuged and dried in a vacuum desiccator. The pellet was resuspended in 8ul distilled water.

Annealing Reaction

To the 8ul of denatured template DNA, 4ul of primer (5ng/ul) and 2ul of annealing buffer (Pharmacia: a buffered solution containing $MgCl_2$ and dithiothreitol) was added. The mixture was incubated for 20 minutes to allow annealing of template to primer. The tube was placed at room temperature for 10 to 30 minutes before proceeding to the labelling reaction.

Labelling Reaction

T7 DNA polymerase was diluted to 1.5 units/ul in cold enzyme dilution buffer (Pharmacia: a buffered solution containing glycerol, bovine serum albumin and dithiothreitol) and held on ice. 2.5ul of "A mix", "C mix", "G mix" and "T mix" were pipetted separately into each of 4 labelled microcentrifuge tubes and placed at $37^{O}C$ (Pharmacia: "A mix" contains an optimal mixture of ddATP with all 4 dNTPs, "C mix" an optimal mixture of ddCTP with all 4 dNTPs, and so on).

To the tube containing the annealed template and primer was added 3ul labelling mix (Pharmacia: dCTP, dGTP and dTTP in aqueous solution), 1ul [alpha-^{35}S]dATP alpha S (>1200Ci/mmol, 10uCi/ul) and 2ul T7 DNA polymerase (1.5 u/ul). The components were mixed by gentle pipetting and incubated at room temperature for 5 minutes.

Termination Reaction

After 5 minutes, 4.5ul of the labelling reaction was transferred and mixed into each of the four pre-warmed 2.5ul aliquots of "N mix". The reaction was incubated at $37^{O}C$ for 5 minutes. The reaction was terminated by addition of 5ul

of formamide gel loading buffer (*q.v.*). The reaction products were stored at -70°C or were heated to 100°C for 3 minutes and loaded immediately onto the sequencing gel. Each sequencing reaction produced enough product for 3 sequencing gels when using the system described below.

Denaturing Polyacrylamide Gel Electrophoresis

The radio-labelled oligonucleotide products of sequencing reactions were separated and resolved by electrophoresis through 6% polyacrylamide gels containing 7M (42%, w/v) urea. Gels were run at a constant temperature of 60°C, at 2000V (approximately 30V/cm) for 2 hours in TBE buffer. Under these denaturing conditions, using 55cm long wedge gels in the high resolution LKB Macrophor system, it was possible to obtain excellent oligonucleotide separations in the size range 40-375 bases in a single run.

In the Macrophor system, the acrylamide gel is cast horizontally between a glass plate, to which it is bonded and the siliconised glass face of a water-filled thermostatic circulator. The two plates were held apart by 0.2mm thick plastic strips at either side. By introducing extra 0.4mm spacers at the bottom of the gel, a wedge-shaped gel was made. This wedge shape produced an electric field strength gradient during electrophoresis, retarding the normally rapid migration of the smallest oligonucleotides. This allowed much more information to be obtained from a single gel.

Polymerisation of the gel took approximately 45 minutes. Once this had occurred, the gel and thermostatic circulator were installed into the electrophoresis apparatus and the buffer reservoirs were filled with TBE. The thermostatic circulator was connected to a water heater set at 60°C. The sample comb was removed and the wells were thoroughly flushed with TBE to remove any incompletely-polymerised acrylamide. The gel was pre-run for 30 minutes at 30V/cm to remove any charged contaminants in the gel. During this time the thermostatic plate came up to temperature. The samples in formamide gel loading buffer were heated to 100°C for 3 minutes prior to loading onto the gel using drawn-glass pipettes.

Autoradiography

After electrophoresis, the glass plate with the gel bonded to it, was prised from the face of the thermostatic circulator. Urea was removed from the gel by immersion of the glass plate with gel attached in 10% acetic acid for 30 minutes. The gel was thoroughly dried with a hair dryer and taken to the darkroom. An appropriately cut piece of autoradiography film was sandwiched between the dried gel on its plate and another, similarly-sized plain glass plate. The plates were held firmly together using bulldog clamps and put up for autoradiography overnight at -70°C in a light-proof box. Autoradiographs were developed using a Kodak automatic processor. After autoradiography, the dried gel was softened and removed from its glass plate by overnight immersion in dilute NaOH solution.

Analysis of Sequence Data

DNA sequence data was initially compiled and analysed using the Microgenie^{TM} software package (Queen and Korn, 1984). The MAKESEARCH command was used extensively, since it allowed comparison of newly-acquired sequence data against a large database of eukaryotic and prokaryotic DNA sequences. A major advantage of MAKESEARCH was that it allowed rapid comparison of sequences up to 2,200 bp. in length against much longer sequences in the database. The ANALYSE command allowed generation of restriction maps which were useful in the development of sub-cloning strategies, or when identifying suitable fragments for use as probes. The TRANSLATE command contributed in the assessment of the degree of defectiveness of the proviruses.

More extensive data analysis was carried out using the University of Wisconsin genetics computer group (G.C.G.) sequence analysis software programmes (Devereux *et al.*, 1984). These programmes, written in Fortran 77, were run on a Digital Equipment Corp. VAX computer using the VMS operating system. BESTFIT, which employs the "local homology" algorithm of Smith and Waterman (1981), was used to find the best segments of homology between pairs of DNA sequences. The gap weight was 5.0 and the length weight 0.3 in all BESTFIT analyses. The maximum permissible "surface of comparison" was 10^6 units (*i.e.* 1,000 bp. compared with 1,000 bp.; or 100 bp. compared with 10,000 bp. *etc.*). Although considerably more laborious, BESTFIT did detect a few areas of DNA homology not detected by Microgenie MAKESEARCH. The G.C.G.

COMPARE and DOTPLOT programmes were used to generate dot-matrix analyses. The COMPARE programme compares two sequences in every register, searching for places where a given number of matches (stringency) occur within a given range or window size (Maizel and Lenk, 1981). Window size and stringency were varied as necessary.

2.4.5 RNA-DIRECTED DNA POLYMERASE (RDDP) ASSAYS

These were carried out by Mr G. Mosson in the laboratory of Dr G.Lees during the characterisation of the A72E virus. The purpose was to determine whether or not MRC-5 cells had been successfully infected prior to Northern and Southern analysis. The method used was a modification of that described by Onions (1980). Briefly, virus particles were pelleted by centrifugation from clarified cell culture supernatants and disrupted with detergent. RDDP activity was assayed using a poly rA-oligo dT template and ^{3}H-dTTP as substrates in an appropriate buffer. The reaction was duplicated in the presence of both Mg^{2+} and Mn^{2+} ions. Poly rA-^{3}H-dT was precipitated in yeast carrier RNA with 5% trichloroacetic acid and collected on Whatman GF/C glass-fibre filters. After thorough rinsing to remove unincorporated ^{3}H-dTTP, the filters were dried in scintillation vials, immersed in scintillation fluid (Ecoscint, International Diagnostics Inc.) and counted for 10 minutes using the tritium channel of a Beckman liquid scintillation counter. Background count levels and activity due to cellular polymerases were controlled by replacing poly rA-oligo dT in the reaction with buffer or poly dA-oligo dT, respectively.

2.5 HYBRIDISATION OF NUCLEIC ACIDS

2.5.1 SOUTHERN TRANSFER AND HYBRIDISATION

Transfer

The method used was modified from that of Southern (1975). Agarose gels intended for Southern transfer underwent electrophoresis and were stained and destained in the usual way. A control lane containing a quantity of probe DNA equivalent to single copy in the haploid genome was included in each gel. A transparent ruler was placed alongside the gel during photography, so that subsequent findings in the autoradiograph could be correlated directly with the

photograph of the ethidium bromide-stained gel. After photography, the gel was immersed for 45 minutes in 1.5M NaCl, 0.5M NaOH to denature the DNA. The gel was then neutralised by immersion in 1M tris HCl (pH 7.4), 1.5M NaCl for 45 minutes. During the neutralisation step, the apparatus required for capillary transfer of DNA fragments was set up. A 40 x 30 x 5 cm plastic tray was half-filled with 20xSSC. This served as the reservoir of transfer buffer. A glass plate, longer but slightly narrower than the tray was laid lengthwise on top of the tray. A piece of Whatman 3MM paper 30 x 40 cm was folded over the glass plate with its long axis perpendicular to that of the plate, so that both of its ends were hanging in the buffer. The Whatman paper was thoroughly wetted with 20xSSC and all bubbles between it and the glass plate were removed by rolling a 10ml glass pipette over the paper. The neutralised agarose gel was inverted and laid down in the centre of the 3MM paper on top of the glass plate. The wells of the gel were trimmed away with a scalpel and one small corner was cut away to allow subsequent orientation. Saran wrap or Cling film was used to surround, but not cover the gel. This was to prevent buffer from flowing directly from the reservoir to the tissue stack on top of the gel (*i.e.* to prevent "short-circuiting"). A pre-cut piece of nylon membrane (Hybond N) the same size as the gel was labelled with pencil. One corner of the membrane was cut off to enable subsequent orientation. The membrane was laid on top of the gel. Pre-wetting of the membrane with 2xSSC was not necessary. Sometimes, however, by reducing friction it made accurate positioning of the membrane on the gel easier. Two pieces of pre-wetted Whatman 3MM paper the same size as the gel were laid on top of the nylon membrane and all bubbles between layers were removed using a 10ml pipette as before. An 8cm stack of tissues was laid on top, followed by a 0.5kg weight. As the tissues became saturated with buffer, they were discarded and replaced. Transfer was allowed to continue for 20-24 hours, after which the membrane was peeled off the dehydrated gel and baked at $80^{O}C$ for two hours to bond the DNA to the membrane. Occasionally, to check the efficiency of transfer, the dehydrated gel was stained with ethidium bromide (0.5ug/ml) and examined under UV light. Although large DNA fragments were often still present in the gel, enough was transferred to produce a strong signal. Depurination prior to alkaline denaturation was not found to be necessary.

Hybridisation

The baked nylon membrane was laid flat in a heat sealable bag. Southern hybridisation fluid (62.5ul/cm^2 of membrane) and boiled salmon testis DNA (10mg/ml, 1.5ul/cm^2 of membrane) were added. All bubbles were excluded from the bag which was then heat sealed. The membrane in its bag was placed in a water-filled sandwich box and was pre-hybridised for 4-16 hours (usually overnight) in a rocking waterbath. Previously-prepared probe (*q.v.*) was heated to 100°C for 2 minutes in a waterbath and rapidly chilled on ice. Using a 21G needle and 1ml syringe, probe ($1x10^6$ cpm/ml of prehybridisation fluid) was added by injection into a corner of the bag. The bag was re-sealed and incubated in the waterbath for a further 16-20 hours to allow hybridisation of probe to target sequences. After hybridisation, the membrane was removed from the bag and washed before being wrapped in cling film and put up for autoradiography.

The stringency of hybridisation was varied by altering the concentration of formamide in the hybridisation fluid (25% or 50%) and the temperature at which pre-hybridisation and hybridisation took place. The stringency of washing was varied by altering the washing temperature and the concentration of SSC in the washing solution. All membranes were first washed in 2xSSC for 10 minutes at room temperature to remove excess probe. They were then washed in n x SSC, 0.1% SDS for 90 minutes. This 90 minute wash was repeated once more. The value of n and the temperature of these washes were dictated by the desired stringency. Membranes were given a final 10 minute rinse in 2xSSC at room temperature before being wrapped in cling film for autoradiography.

The decision as to what stringency parameters to use depended on the particular probe-target combination. Most membranes were hybridised under one or other of two sets of stringency conditions: HIGH STRINGENCY consisted of pre-hybridisation and hybridisation at 45°C in the presence of 50% formamide. Washing was for 2 x 90 minutes in 0.5xSSC, 0.1% SDS at 65°C; with a 10 minute, room temperature, 2xSSC rinse at the beginning and end. LOW STRINGENCY consisted of pre-hybridisation and hybridisation at 37°C in the presence of 25% formamide. Washing was for 2 x 90 minutes in 1xSSC, 0.1% SDS at 55°C; with a 10 minute, room temperature, 2xSSC rinse at the beginning and end. The

empirical formula derived from work by Bolton and McCarthy, 1962 and cited by Meinkoth and Wahl, 1984:

$$T_m = 81.5^oC + 16.6 \log M + 0.41(\%G+C) - 500/L - 0.61(\%formamide)$$

in which T_m is the melting temperature, M the molarity of the hybridisation solution and L the probe length, was applied occasionally, but was not considered to be particularly useful in the specific context of this project. This was because:

1. The degree of homology between probe and target was completely unknown in most experiments. Under the conditions used in this project T_m is reported to decrease by 1^oC for every 1% of mismatched base pairs (Bonner *et al.*, 1973).

2. The random priming method produces probes of very variable and unpredictable length, thus melting will tend to occur over a wide range of temperatures, rather than at one particular temperature.

3. The G+C content of target and probe were often unknown.

2.5.2 NORTHERN TRANSFER AND RNA-DNA HYBRIDISATION

Gels for Northern transfer underwent agarose/formaldehyde gel electrophoresis as usual. After electrophoresis, a duplicate part of the gel was excised and retained for ethidium bromide staining (*q.v.*). The part of the gel intended for transfer was not stained. It was rinsed several times in distilled water to remove formaldehyde, immersed in 10xSSC and shaken gently for 45 minutes. RNA was transferred from the gel to a nylon membrane (Hybond N) exactly as described above for Southern transfer. The membrane was then baked as usual. Pre-hybridisation and hybridisation were carried out essentially as for Southern blots using 25% or 50% formamide hybridisation buffer at 37^oC or 45^oC, respectively. To detect potentially low-abundance mRNAs, up to 5 x 10^6 cpm of probe/ml of hybridisation buffer was used. The membrane was washed for 10 minutes in 2xSSC at room temperature, followed by 3 washes of 20 minutes each at 65^oC in 0.5xSSC, 0.1% SDS. Finally, a 10 minute rinse in 2xSSC at room temperature was carried out. The membrane was dried with tissue paper, wrapped in cling film and put up for autoradiography (*q.v.*).

2.5.3 BACTERIOPHAGE PLAQUE HYBRIDISATION

Bacteriophages intended for hybridisation screening were grown on 20 x 20cm NZY culture plates. Once visible plaques had formed, the plate was removed from the 37^{0}C incubator and stored at 4^{0}C for 30 minutes. This allowed the top agarose to harden somewhat and prevented it from tearing when the hybridisation membrane was lifted off. A pre-cut piece of Hybond-N membrane was carefully laid on top of the agarose avoiding creases or bubbles. This was easily achieved by starting with one edge of the membrane and slowly lowering it onto the gel. When the membrane was in position on the gel, holes were stabbed through into the agar in an assymetric pattern with a 21G needle. This left complementary holes in the membrane and gel, which allowed subsequent accurate orientation of the autoradiograph relative to the gel for plaque selection. After one minute in contact with the gel the membrane was removed and placed, phage side up, on a 0.5cm pad of Whatman 3MM paper soaked in denaturation solution (1.5M NaCl, 0.5M NaOH). The solution was allowed to diffuse through the membrane, but not to wash over it, for 7 minutes. The membrane was transferred to a similar pad of paper soaked in neutralisation solution (1M tris HCl (pH 7.4), 1.5M NaCl) and left for 3 minutes. Neutralisation was repeated once more with a fresh soaked pad and then the membrane was washed briefly in 2xSSC. The membrane was baked, pre-hybridised, hybridised and washed exactly as for a high stringency Southern blot and put up for autoradiography.

2.5.4 AUTORADIOGRAPHY

After washing, excess fluid was blotted off hybridisation membranes with tissue paper. They were wrapped in cling film and sandwiched in between fast tungstate intensifying screens in a film cassette. Hyperfilm MP (Amersham) autoradiography film was exposed to the membrane at -70^{0}C for 16 hours and was then developed. Depending upon the results, the membrane was put up for further autoradiography, with exposure times varying from days to weeks.

Chapter 3

SUB-CLONING AND SEQUENCING OF TWO DEFECTIVE CANINE RETROVIRAL ELEMENTS

3.1 INTRODUCTION

3.2 DNA SEQUENCING

3.2.1 TEMPLATE DNAs

3.2.2 KLENOW FRAGMENT OF *E. COLI* DNA POLYMERASE

3.2.3 BACTERIOPHAGE T7 DNA POLYMERASE

3.3 PRELIMINARY EXPERIMENTS

3.3.1 PREPARATION AND CHARACTERISATION OF THE FMuLV PROBE

3.3.2 PREPARATION AND PRELIMINARY RESTRICTION ANALYSIS OF THE BACTERIOPHAGE LAMBDA DNAs

3.3.3 PRELIMINARY HYBRIDISATION EXPERIMENTS

3.4 CHARACTERISATION OF LAMBDA INSERT DNA FRAGMENTS WHICH HYBRIDISED WITH FMuLV

3.4.1 RESTRICTION MAPPING AND PROBING OF LAMBDA 4

3.4.2 RESTRICTION MAPPING AND PROBING OF LAMBDA 3

3.5 SUB-CLONING AND SEQUENCING OF HYBRIDISING FRAGMENTS

3.5.1 SUB-CLONING AND SEQUENCING STRATEGY FOR LAMBDA 4

3.5.2 SUB-CLONING AND SEQUENCING STRATEGY FOR LAMBDA 3

3.6 SEQUENCE DATA ANALYSIS

3.6.1 SEQUENCE DATA ANALYSIS OF LAMBDA 4

3.6.2 SEQUENCE DATA ANALYSIS OF LAMBDA 3

3.6.3 COMPARISON OF L3 AND L4

3.7 DISCUSSION

3.1 INTRODUCTION

This chapter describes how cloned canine genomic DNA from the lymphoma cell line 3132 was probed with a murine C-type retroviral sequence (permuted, full length proviral FMuLV). Hybridising canine DNA fragments were sub-cloned into the phagemid vector pBluescript KS (+) and sequenced by the Sanger dideoxy chain termination method. Two similar, but distinct, highly defective retroviral proviruses were identified and characterised.

The strategies described in this chapter to identify, clone and sequence canine retroviral elements have been used to good effect in other species. In a rather convoluted, but successful attempt to identify human endogenous retroviral elements, Martin *et al.* (1981) first cloned an element from African green monkey DNA which hybridised with a cloned AKV ecotropic provirus. They probed human DNA with the African green monkey clone and identified the first human endogenous retroviral elements. Bonner *et al.* (1982) probed human genomic DNA with a chimpanzee retroviral *pol* probe and were the first to determine partial nucleotide sequence of a human endogenous provirus. More recently, low stringency hybridisation of human DNA with the C-type simian sarcoma-associated virus allowed identification of the defective human endogenous retroviral element S71, which was found to contain *gag*- and *pol*-related regions, interspersed non-viral DNA and a 3' LTR (Leib-Mösch *et al.*, 1986; Werner *et al.*, 1990).

The cloning and sequencing techniques used to carry out the work described in this chapter are well-established and have been detailed in Chapter 2. Certain interesting aspects of the sequencing methodology are briefly discussed below.

3.2 DNA SEQUENCING

3.2.1 TEMPLATE DNAs

All sequencing was carried out using alkali-denatured, double-stranded DNA templates (method adapted from Chen and Seeberg, 1985). Although it is more difficult to achieve uniformly good results with double-stranded, as against single-stranded templates (Sambrook *et al.*, 1989c), satisfactory results are attainable, especially with the newer, more processive DNA polymerases.

A critical factor when sequencing double-stranded templates is the quality of input DNA. During the first two years of the project, plasmid DNA purified by equilibrium centrifugation in CsCl-ethidium bromide gradients was used. In the last year, excellent results were achieved using acid-phenol miniprep DNA templates and T7 DNA polymerase (for example, see Fig. 3.1). Conventional miniprep DNA produced unsatisfactory results with the same polymerase. Interestingly, DNA prepared by the conventional miniprep method could not be improved as a sequencing template by subsequent acid phenol extraction (personal observations and Weickert and Chambliss, 1989). Acid phenol has been reported to remove nicked and linear DNA from supercoiled DNA preparations (Zasloff *et al.*, 1978). This is probably relevant to its beneficial effect in miniprep DNA sequencing.

3.2.2 KLENOW FRAGMENT OF *E. COLI* DNA POLYMERASE

This enzyme was used in early experiments and produced inconsistent results. A recurrent problem was the presence of bands at the same level in all four tracks. These so-called "stops" occurred in several regions of the sequencing autoradiograph, particularly among the larger fragments. This problem was probably caused in part by the inherent low processivity of Klenow fragment and partly by the presence of small amounts of contaminating bacterial chromosomal DNA in the plasmid preparations. Another problem was that the sequences under analysis contained long homopolymeric tracts which are known to be copied poorly by Klenow fragment. The problems with Klenow fragment were alleviated by increasing the temperature of the polymerization reaction from $37^{o}C$ to $50^{o}C$.

3.2.3 BACTERIOPHAGE T7 DNA POLYMERASE

Compared with the Klenow fragment of *E. coli* DNA polymerase I; bacteriophage T7 DNA polymerase is highly processive, rapid in action and more tolerant of nucleotide analogues, such as dITP and 7-deaza-dGTP, which are occasionally used to resolve compressions in sequencing gels (Tabor and Richardson, 1987). Commercially-available products are chemically-modified (*e.g.* T7, Pharmacia LKB.) or genetically-engineered (*e.g.* Sequenase 2.0, U.S. Biochemical Corp.) to remove the 3' to 5' exonuclease activity of the native enzyme. The Pharmacia LKB product was used in this project.

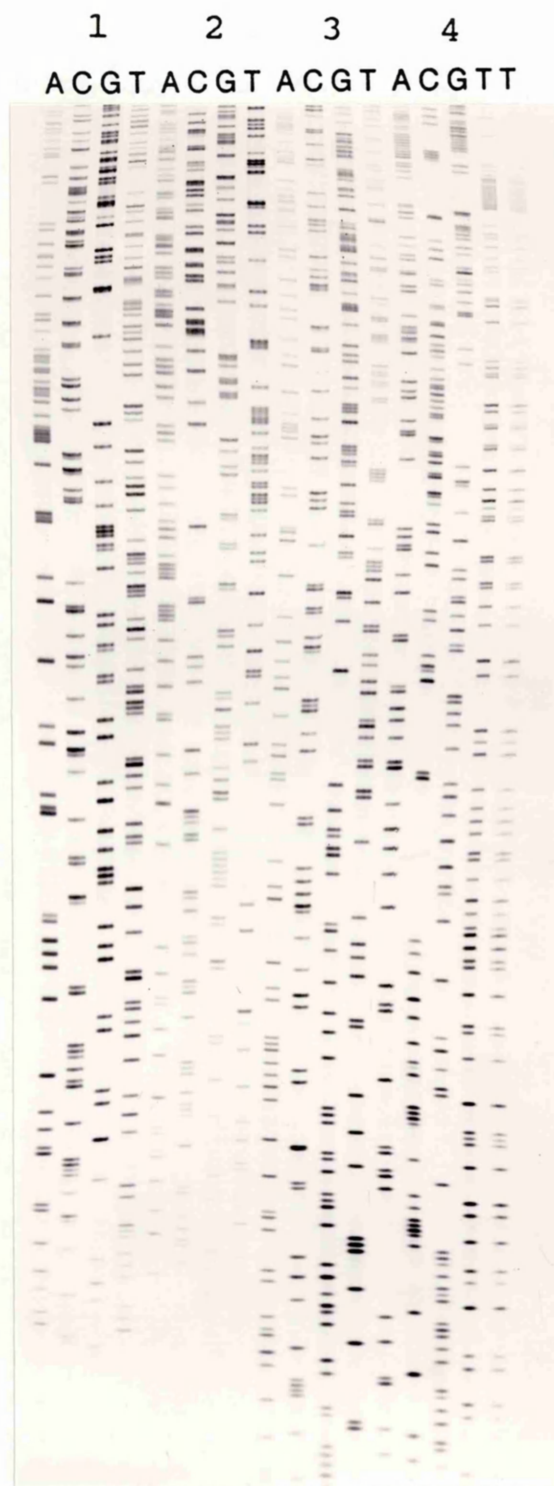


Fig. 3.1: A sequencing autoradiograph showing typical results achieved with acid phenol minipreparation DNA templates. The autoradiograph shows the results of four sets of sequencing reactions: (1). Lambda 3 *Sst*I/*Xho*I 362 bp. fragment in pBluescript KS (+), M13 primer; (2). Same template as (1), T3 primer; (3). Lambda 3 *Sst*I 1197 bp. fragment in pBluescript KS (+), custom-synthesised primer No. 4; (4). Same template as (3), custom-synthesised primer No. 5.

To take advantage of the high processivity of T7 DNA polymerase, the sequencing reaction was divided into two stages. In the first stage, low concentrations of dNTPs (including a single radiolabelled dNTP) and a low incubation temperature ensured limited DNA synthesis but efficient incorporation of radio-labelled dNTP. In the second stage, the mixture was divided into the standard 4 reactions, each containing high concentrations of dNTPs and a single ddNTP. The second stage was carried out at a higher temperature ($37^{O}C$).

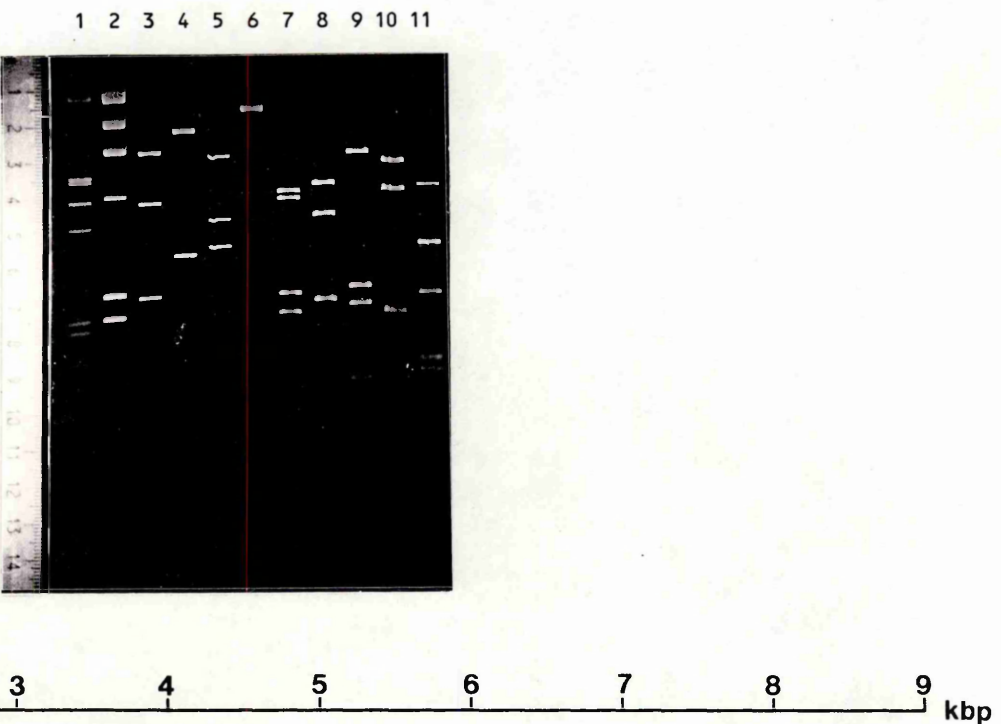
3.3 PRELIMINARY EXPERIMENTS

3.3.1 PREPARATION AND CHARACTERISATION OF THE FMuLV PROBE

At the beginning of the project Dr R. Padua kindly provided 50 ul of a 0.9 ug/ul DNA solution which was thought to contain permuted, full length FMuLV provirus as an *Eco*RI insert in an unstated plasmid vector. 1 ug of this DNA was used to transform competent DS 941 cells. Transformed clones were selected on the basis of ampicillin resistance, patched out on ampicillin/streptomycin plates and DNA minipreparations were carried out. The vast majority of clones had been transformed by a contaminating, non-recombinant plasmid with size and restriction pattern characteristics of pBR 322. *Eco*RI digestion of DNA from one clone (R14) produced two electrophoretic bands, 8.5 kbp. and 4.3 kbp. in length. These were thought to represent the 8.5 kbp. permuted FMuLV proviral DNA insert in pBR 322 vector. Interestingly, the intensity of the 4.3 kbp. band on repeated ethidium bromide-stained gels was equal to that of the 8.5 kbp. band. Large scale plasmid DNA preparation of R14 was carried out and aliquots of the product were digested with a variety of restriction enzymes. Restriction mapping demonstrated that the R14 construct was, in fact, a single permuted FMuLV provirus in a double pBR 322 vector . The restriction pattern of the permuted FMuLV insert was close to that of the 8.5 kbp. clone with a single LTR described by Oliff *et al,*(1980; Fig. 3.2).

The FMuLV insert was freed by *Eco*RI digestion, purified from LMP agarose and cloned into the *Eco*RI site of pIC20H. Large scale plasmid preparation of the new construct was carried out so that milligram quantities of insert DNA would be available for probe preparation.

a).



b).

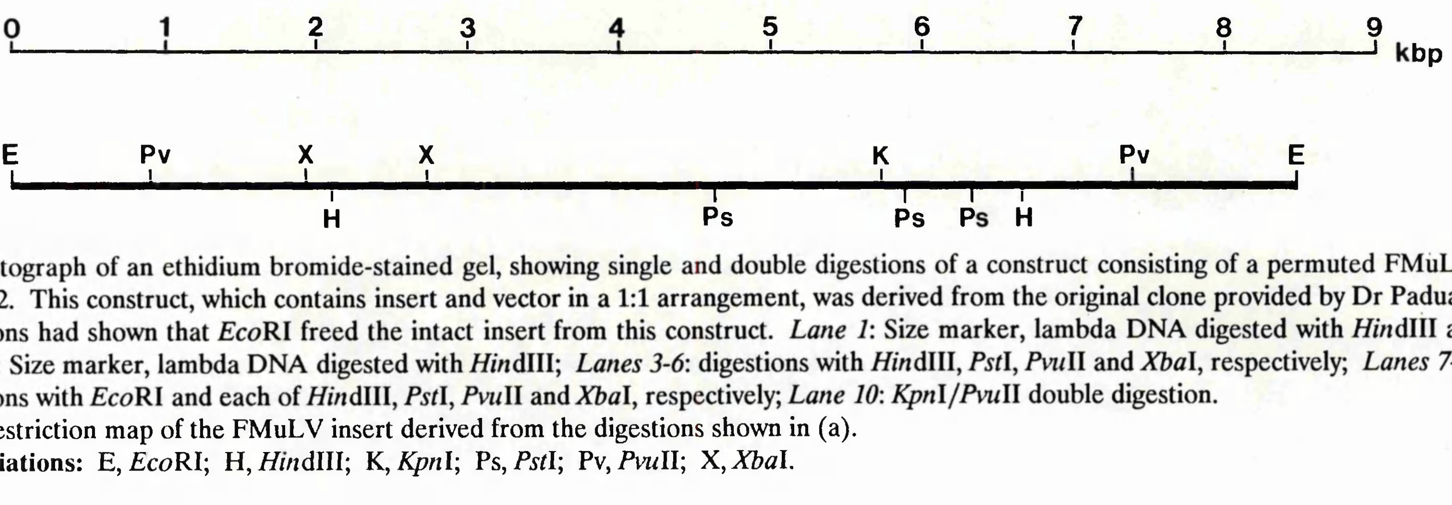


Fig. 3.2: a). Photograph of an ethidium bromide-stained gel, showing single and double digestions of a construct consisting of a permuted FMuLV insert in pBR322. This construct, which contains insert and vector in a 1:1 arrangement, was derived from the original clone provided by Dr Padua. Previous digestions had shown that *Eco*RI freed the intact insert from this construct. *Lane 1*: Size marker, lambda DNA digested with *Hin*dIII and *Eco*RI; *Lane 2*: Size marker, lambda DNA digested with *Hin*dIII; *Lanes 3-6*: digestions with *Hin*dIII, *Pst*I, *Pvu*II and *Xba*I, respectively; *Lanes 7-10*: double digestions with *Eco*RI and each of *Hin*dIII, *Pst*I, *Pvu*II and *Xba*I, respectively; *Lane 10*: *Kpn*I/*Pvu*II double digestion.
b). A restriction map of the FMuLV insert derived from the digestions shown in (a).
Abbreviations: E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; Ps, *Pst*I; Pv, *Pvu*II; X, *Xba*I.

3.3.2 PREPARATION AND PRELIMINARY RESTRICTION ANALYSIS OF THE BACTERIOPHAGE LAMBDA DNAs

Seven FMuLV-hybridising recombinant lambda 2001 clones containing 3132 lymphoma DNA were kindly provided by Dr Padua. These clones, designated L1 to L7, were each separately double plaque-purified, quantitated and prepared on a large scale. Phage DNA was extracted, purified and quantitated. Table 1 shows phage titres (pfu/ul) at each stage of plaque purification (PP) and the final yields of phage DNA, as calculated from optical densitometry.

DNAs from the seven lambda 2001 clones were digested with *Bam*HI, *Hin*dIII, *Kpn*I, *Pst*I, *Sst*I, *Xba*I and *Xho*I and fragments were separated by electrophoresis. These enzymes, with the exception of *Pst*I, were chosen because their few restriction sites within lambda 2001 are known (Karn *et al.*, 1984). Restriction mapping of the constructs would thus be facilitated using these enzymes. L1 to L7 each consisted of *Sau*3A-digested insert 3132 DNA cloned into the *Bam*HI sites of the left and right lambda 2001 polylinkers. In such constructs *Sst*I, *Xba*I and *Xho*I cut vector sequence in the polylinkers only, freeing a variable number of insert fragments from the intact 20 kbp. left and 9 kbp. right lambda arms. Examination of the digestion products in ethidium bromide stained gels revealed the following information:-

1). *Sst*I, *Xba*I and *Xho*I digestions generated 20 kbp. and 9 kbp. lambda arm fragments in addition to a variable number of insert fragments in all seven lambda clones. This confirmed that the basic structure of the recombinant bacteriophage genomes (Left Arm--Insert--Right Arm) was intact.

2). L4, L6 and L7 had completely identical restriction patterns and were thus clonal. Since these clones were originally selected as separate FMuLV-hybridising plaques from a relatively small screening of a lambda library, it seems likely that artifactual amplification of this particular clone must have occurred prior to selection. This could have occurred if the soft agarose on the culture plates was too wet, allowing diffusion of replicating bacteria and bacteriophage to produce multiple plaques from a single initial clone. Once L4, L6 and L7 were determined to be identical, L4 was used to represent all three in subsequent experiments.

Clone No.	L1	L2	L3	L4	L5	L6	L7
Initial pfu/ul (as provided)	36	29	19	4	7	10	40
1st PP cycle	216	120	220	280	1,100	1,000	286
2nd PP cycle	10,000	7,000	2,300	4,000	600	250	6,500
Final DNA yield (ug/ml)*	225	336	66	78	163	80	32

Table 3.1: Phage titres at each stage of plaque purification and the final yields of phage DNA.

Abbreviations: pfu, plaque forming unit(s); ul, microlitre; PP, plaque purification.

*As determined by spectrophotometry.

3). The intensity of staining of DNA fragments in the gels did not correlate with the DNA concentrations calculated from optical densitometry. This was because RNA or DNA oligonucleotides contaminated the bacteriophage DNA in L1, L2 and L5, artifactually elevating their O.D.$_{260}$ values and producing a brightly-stained "plume" near the bottom of each gel (see Fig. 3.3). Significantly, L1, L2 and L5 were the clones for which chloroform was used to bring bacterial lysis to completion during bacteriophage preparation (see Chapter 2, section 2.2.1 a). It is likely that the "plume" represents contaminating bacterial nucleic acids.

4). L1 to L5 had completely different arrangements of insert restriction fragments. Thus, it seemed unlikely that equivalent proviruses (which would be expected to have similar restriction patterns) were present in more than one of the clones.

5). *Pst*I produced a complex pattern of bands in all of the clones because it cut at multiple sites in the lambda arms. It was not considered useful in the analysis of these clones.

3.3.3 PRELIMINARY HYBRIDISATION EXPERIMENTS

DNAs from L1 to L7 were separately digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Sal*I, *Sst*I, *Xba*I and *Xho*I. Fragments were electrophoretically separated, transferred to nylon membranes and probed with FMuLV at low and high stringencies. The hybridisation patterns of L4, L6 and L7 were identical, consistent with their being identical clones. Therefore, in later hybridisation experiments, only L1 to L5 were probed.

Strongly hybridising bands were consistently present in digestions of L3 and L4. In particular, a 2.5 kbp. *Sst*I L4 fragment and a 3.5 kbp. *Bam*HI L3 fragment hybridised strongly to the FMuLV probe (Fig. 3.4). Weakly hybridising bands were present in L1 and L2. No hybridising bands were present in L5. In view of these findings, it was decided that the inserts of L3 and L4 should be fully mapped and their hybridising sub-fragments cloned and sequenced. L4 was characterised in this way before L3, hence the order of the descriptions in the sections below.

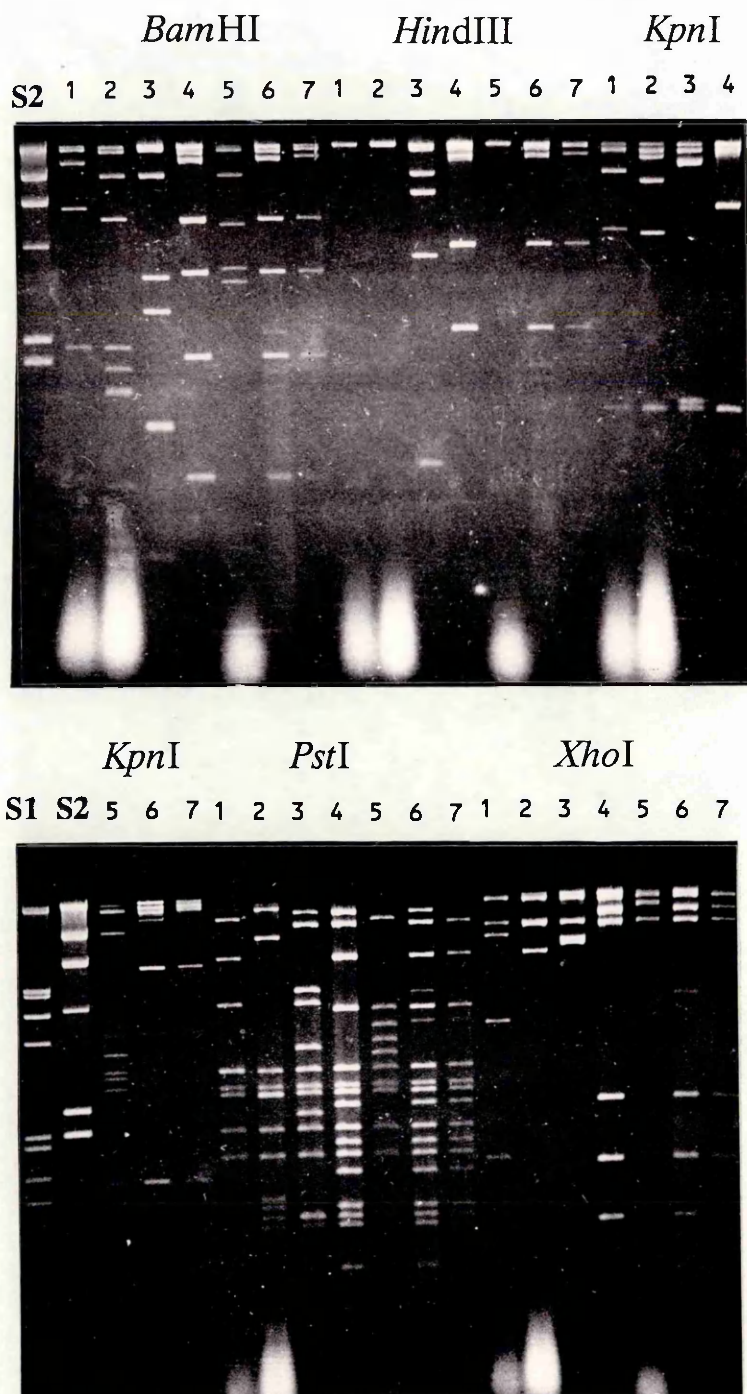
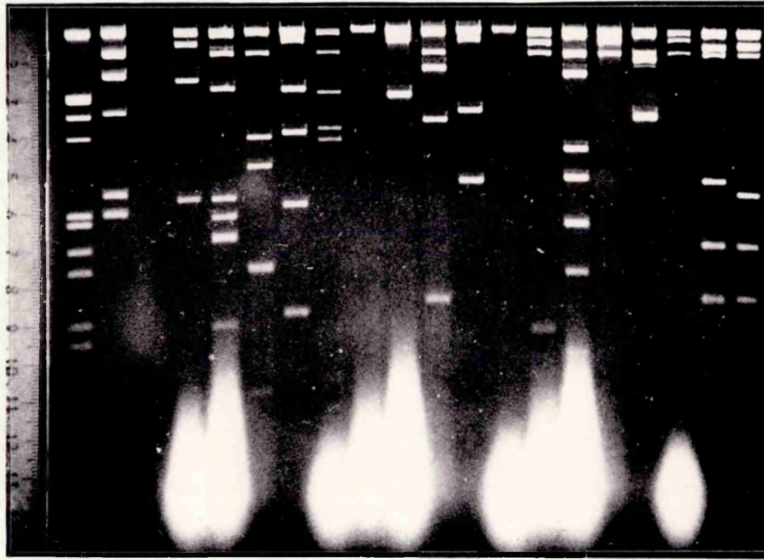


Fig. 3.3: Photographs of ethidium bromide-stained gels, showing digestions of L1 to L7 with a variety of solo restriction endonucleases. L4, L6 and L7 are shown to be identical clones. L1, L2 and L5 have prominent flares caused by the presence of oligonucleotides within the samples. These were the bacteriophage clones for which chloroform was used to encourage bacterial cell lysis.

Abbreviations: S1, DNA size marker: bacteriophage lambda DNA digested with *Hin*dIII and *Eco*RI; S2, DNA size marker, bacteriophage lambda DNA digested with *Hin*dIII.



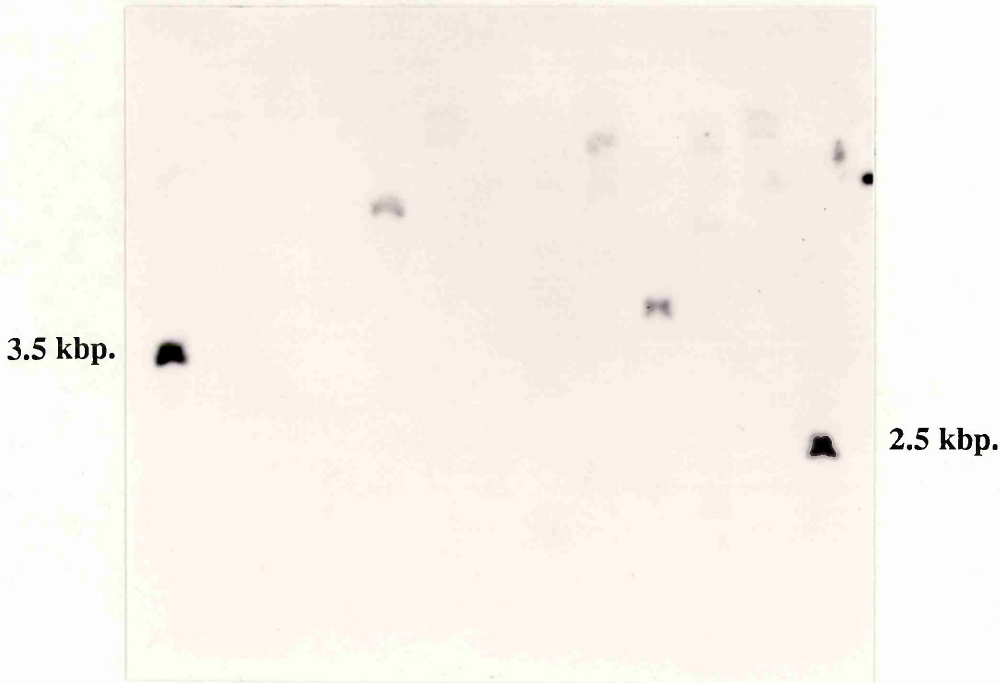


Fig. 3.4: a). A photograph of an ethidium bromide-stained gel, showing restriction digestions of L1 to L5. L1, L2 and L5 have prominent flares caused by the presence of oligonucleotides within these samples.
b). An enlarged part of an autoradiograph derived from a very similar gel (with L5 digestions not included). The Southern transfer membrane was probed with FMuLV. There is prominent hybridisation to L3 and L4 fragments; in particular, to a 3.5 kbp. L3 *Bam*HI fragment and to a 2.5 kbp. L4 *Sst*I fragment.

Abbreviations: S1, DNA size marker: bacteriophage lambda DNA digested with *Hin*dIII and *Eco*RI; S2, DNA size marker, bacteriophage lambda DNA digested with *Hin*dIII.

3.4 CHARACTERISATION OF LAMBDA INSERT DNA FRAGMENTS WHICH HYBRIDISED WITH FMuLV

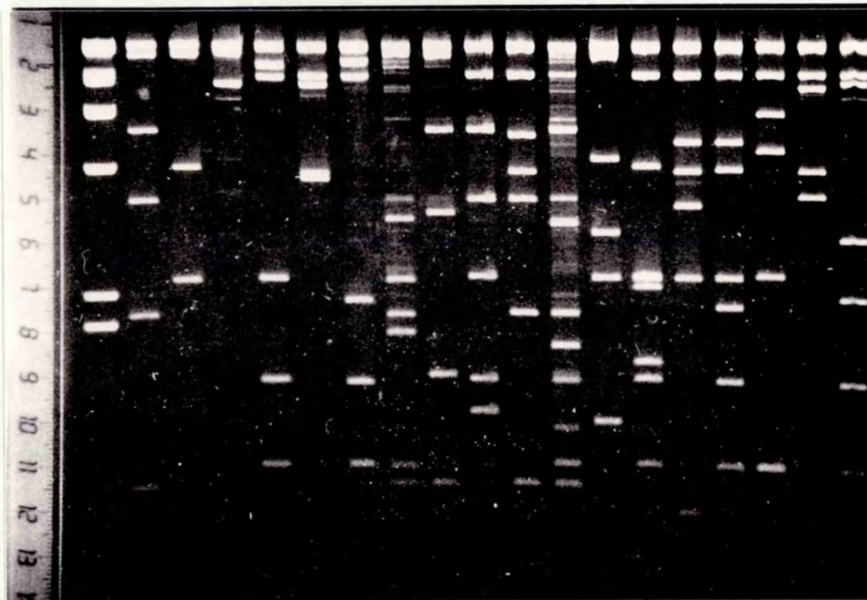
3.4.1 RESTRICTION MAPPING AND PROBING OF LAMBDA 4

L4 DNA was digested with *Bam*HI, *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I and all possible double digestion combinations of these enzymes to produce a restriction map of the 16.7 kbp. insert. DNA fragments in the gels used for this purpose were transferred to nylon membranes and probed with FMuLV. Fig. 3.5 shows an ethidium bromide-stained gel of single and double digestions of L4, and below, an autoradiograph of the same gel after Southern transfer and hybridisation. On p.94 is the derived restriction map of the L4 insert, showing the location of sub-fragments which hybridised to FMuLV. The legend details the logic by which the map was deduced. Fig. 3.5 shows that FMuLV-hybridising sequences in L4 were limited to 2.5 kbp. at the extreme right hand end of the insert. A 2.5 kbp. *Sst*I fragment and two *Xho*I fragments (1.5 and 1.0 kbp.) at the right hand end of the 16.7 kbp. insert were the only fragments to hybridise in their respective digestions. Adjacent sequences (a 0.6 kbp. *Sst*I fragment and a 2.2 kbp. *Xho*I fragment) did not hybridise, or hybridised very weakly. The 4.2 kbp. *Xba*I fragment which had previously been noticed to hybridise strongly included the right hand end of the insert. Why the smaller *Xho*I DNA fragments did not hybridise in the preliminary experiment is unclear. There may have been inadequate binding of the smaller DNA fragments to the nylon membranes. In a few early experiments, ultraviolet transillumination, rather than baking, was used to bind DNA to the nylon membranes. Although this method is recommended by the membrane manufacturer, it requires a particular wavelength and intensity of UV light which may not have been provided by the transilluminator used in early experiments. Baking of nylon membranes, which was used in all subsequent experiments, produced consistently good results.

3.4.2 RESTRICTION MAPPING AND PROBING OF LAMBDA 3

A restriction map of the 14.5 kbp. L3 insert was deduced by the same double digestion method as was used for L4 (logic detailed in the legend of Fig. 3.6). Southern transfer and hybridisation showed that approximately 4 kbp. of FMuLV-like sequence was located near the middle of the insert (Fig. 3.6). The 3.5

a). S2 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18



b). 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

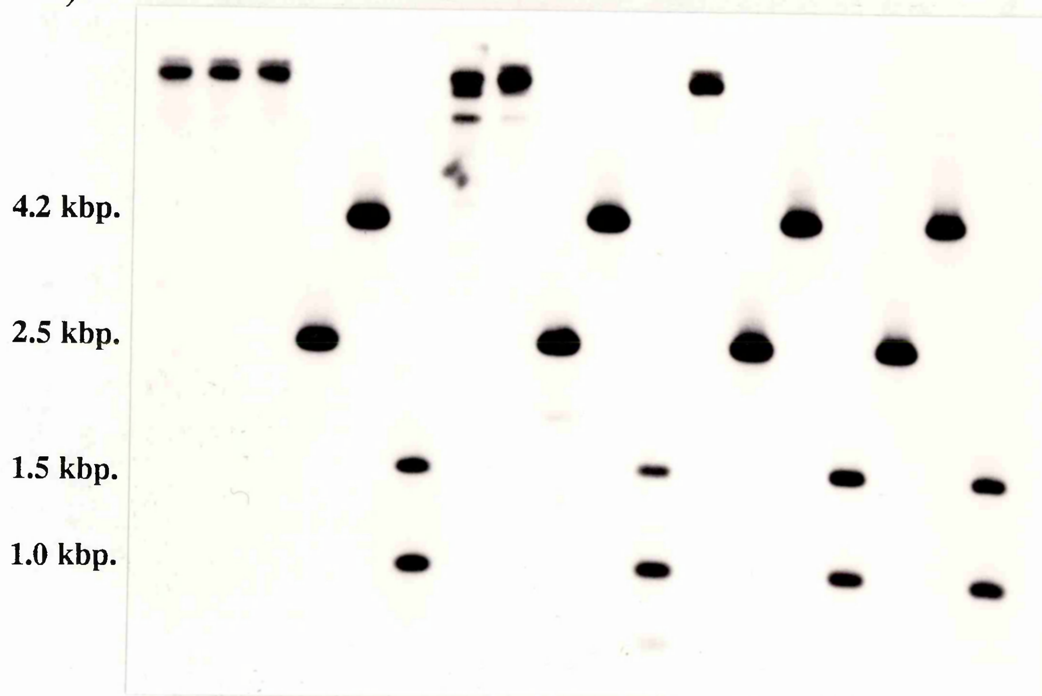
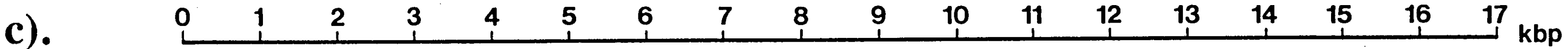


Fig. 3.5: **CONTINUED OVERLEAF. PLEASE TURN OVER FOR FULL ANNOTATION.**

Key: S2, DNA size marker: bacteriophage lambda DNA digested with *Hin*dIII; *Lanes 1-6*: single RE digestions with *Bam*HI, *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 7-11*: double digestions with *Bam*HI and *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 12-15*: double digestions with *Hin*dIII and *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 16-18*: double digestions with *Sal*I and *Sst*I, *Xba*I and *Xho*I, respectively.

c).



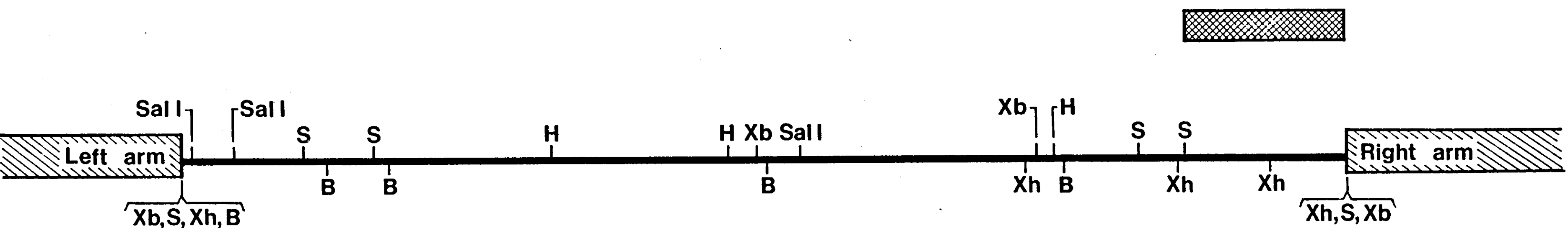


Fig. 3.5: a). A photograph of an ethidium bromide-stained gel, showing single and double restriction endonuclease digestions of L4. S2, DNA size marker: bacteriophage lambda DNA digested with *Hin*dIII; *Lanes 1-6*: single RE digestions with *Bam*HI, *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 7-11*: double digestions with *Bam*HI and *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 12-15*: double digestions with *Hin*dIII and *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 16-18*: double digestions with *Sal*I and *Sst*I, *Xba*I and *Xho*I, respectively.

b). An autoradiograph derived from the gel shown in (a). The Southern transfer membrane was probed at high stringency with FMuLV. Lane numbering is the same as in (a). A 4.2 kbp. *XbaI* fragment, a 2.5 kbp. *Sst*I fragment and two *Xho*I fragments (1.5 and 1.0 kbp.) account for almost all of the hybridisation in their respective digestions.

c). A restriction map of the 16.7 kbp. L4 bacteriophage lambda 2001 insert derived from the data presented in (a) and (b). The crosshatched box indicates the subfragments which hybridised to the FMuLV probe. As can be seen, they are restricted to the extreme right hand end of the insert.

Abbreviations: B, *Bam*HI; H, *Hin*dIII; S, *Sst*I; Xb, *Xba*I; Xh, *Xho*I

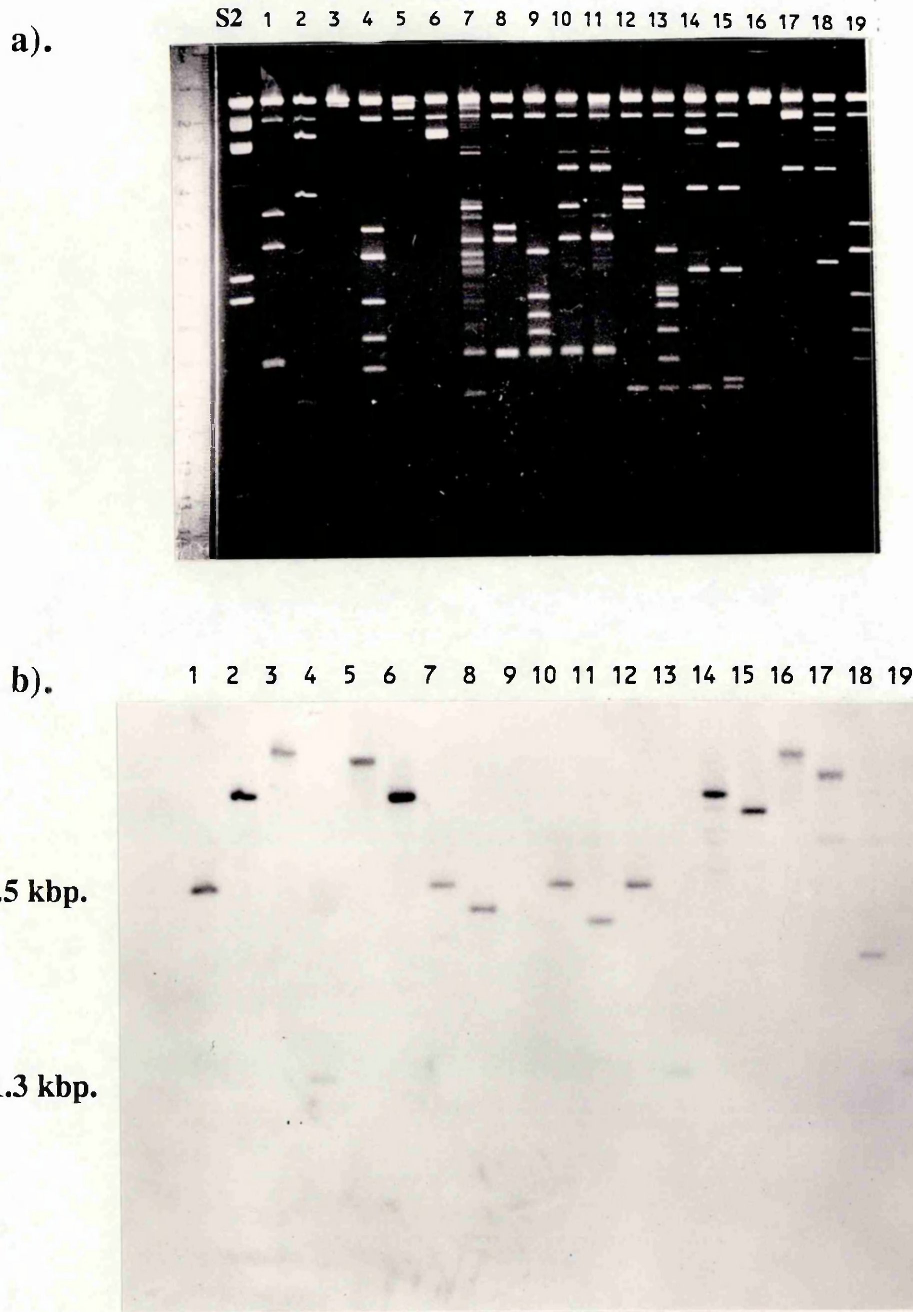


Fig. 3.6: **CONTINUED OVERLEAF. PLEASE TURN OVER FOR FULL ANNOTATION.**

Key: S2, DNA size marker: bacteriophage lambda DNA digested with *Hin*dIII; *Lanes 1-6*: single RE digestions with *Bam*HI, *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 7-11*: double digestions with *Bam*HI and *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 12-15*: double digestions with *Hin*dIII and *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 16-18*: double digestions with *Sal*I and *Sst*I, *Xba*I and *Xho*I, respectively; *Lane 19*: double digestion with *Sst*I and *Xba*I.

c).

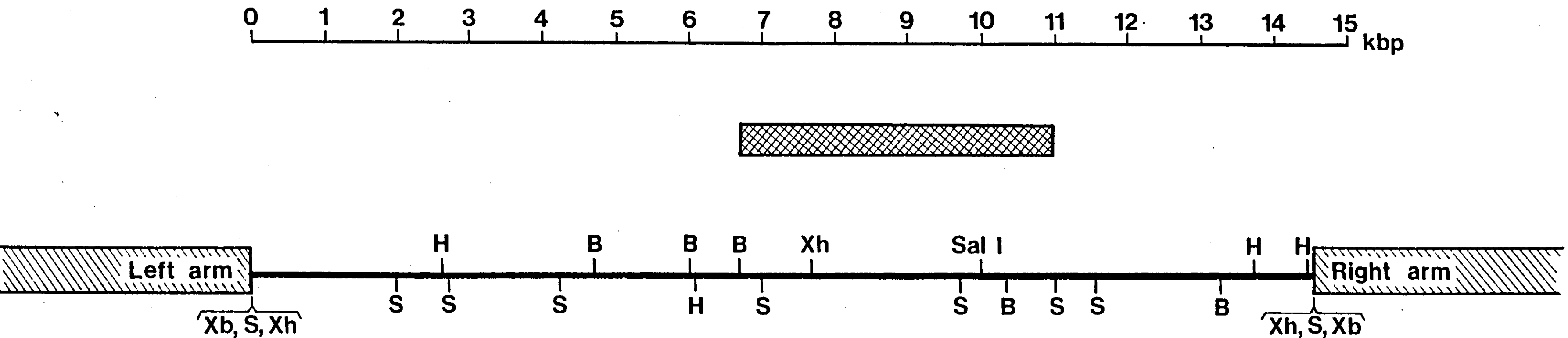


Fig. 3.6: a). A photograph of an ethidium bromide-stained gel, showing single and double restriction endonuclease digestions of L3. S2, DNA size marker: bacteriophage lambda DNA digested with *Hin*dIII; *Lanes 1-6*: single RE digestions with *Bam*HI, *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 7-11*: double digestions with *Bam*HI and *Hin*dIII, *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 12-15*: double digestions with *Hin*dIII and *Sal*I, *Sst*I, *Xba*I and *Xho*I, respectively; *Lanes 16-18*: double digestions with *Sal*I and *Sst*I, *Xba*I and *Xho*I, respectively; *Lane 19*: double digestion with *Sst*I and *Xba*I.

b). An autoradiograph derived from the gel shown in (a). The Southern transfer membrane was probed at high stringency with FMuLV. Lane numbering is the same as in (a). Several subfragments have hybridised to the FMuLV probe, including a 3.5 kbp. *Bam*HI fragment and a 1.3 kbp. *Sst*I fragment.

c). A restriction map of the 14.5 kbp. L3 bacteriophage lambda 2001 insert derived from the data presented in (a) and (b). The crosshatched box indicates the region of the insert which hybridised to the FMuLV probe.

Abbreviations: B, *Bam*HI; H, *Hin*dIII; S, *Sst*I; Xb, *Xba*I; Xh, *Xho*I

kbp. *Bam*HI fragment noticed in the preliminary experiment and an overlapping 1.3 kbp. *Sst*I fragment hybridised most strongly.

3.5 SUB-CLONING AND SEQUENCING OF HYBRIDISING FRAGMENTS

3.5.1 SUBCLONING AND SEQUENCING STRATEGY FOR LAMBDA 4

The hybridising fragments (2.5 kbp. *Sst*I, 1.5 kbp. *Xho*I and 1.0 kbp. *Xho*I) were purified from LMP gels and cloned separately into pBluescript KS (+). In addition the 0.6 kbp. *Sst*I fragment immediately to the left of the hybridising fragments was sub-cloned. Restriction analysis of the new constructs revealed a useful *Pst*I site in the middle of the 1.0 kbp. *Xho*I fragment. This site was used to generate two further constructs in pBluescript KS (+): one with a 2.0 kbp. *Sst*I/*Pst*I insert and the other with a 0.5 kbp. *Pst*I/*Sst*I insert (see Fig. 3.7).

Sequencing was carried out using appropriate universal primers and custom-synthesised 17 to 20mers. Fig. 3.7 shows the sequencing strategy adopted. Arrows represent the extent and direction of nucleotide sequences obtained from each sequencing reaction. In total, 3,089 bp. at the right hand end of the L4 insert was sequenced. Once sequence data became available, it was clear that the sequence should be numbered from right to left, to orientate it correctly with respect to the retroviral element it contained. This was because the 5' end of the positive sense proviral DNA strand was towards the right hand end of the lambda 2001 insert.

3.5.2 SUBCLONING AND SEQUENCING STRATEGY FOR LAMBDA3

Fig. 3.8 shows the sub-cloning and sequencing strategy used to determine the nucleotide sequence of 6,266 bp. of the L3 insert. Custom-synthesised oligonucleotides are rather expensive, hence the large number of sub-clones. With the advent of acid phenol miniprep DNA sequencing, with its considerable time-saving; construction of sub-clones was, in any case, quicker than waiting for oligonucleotide synthesis. As with L4, the L3 sequence had to be numbered from right to left, to orientate it correctly with respect to the retroviral element it contained.

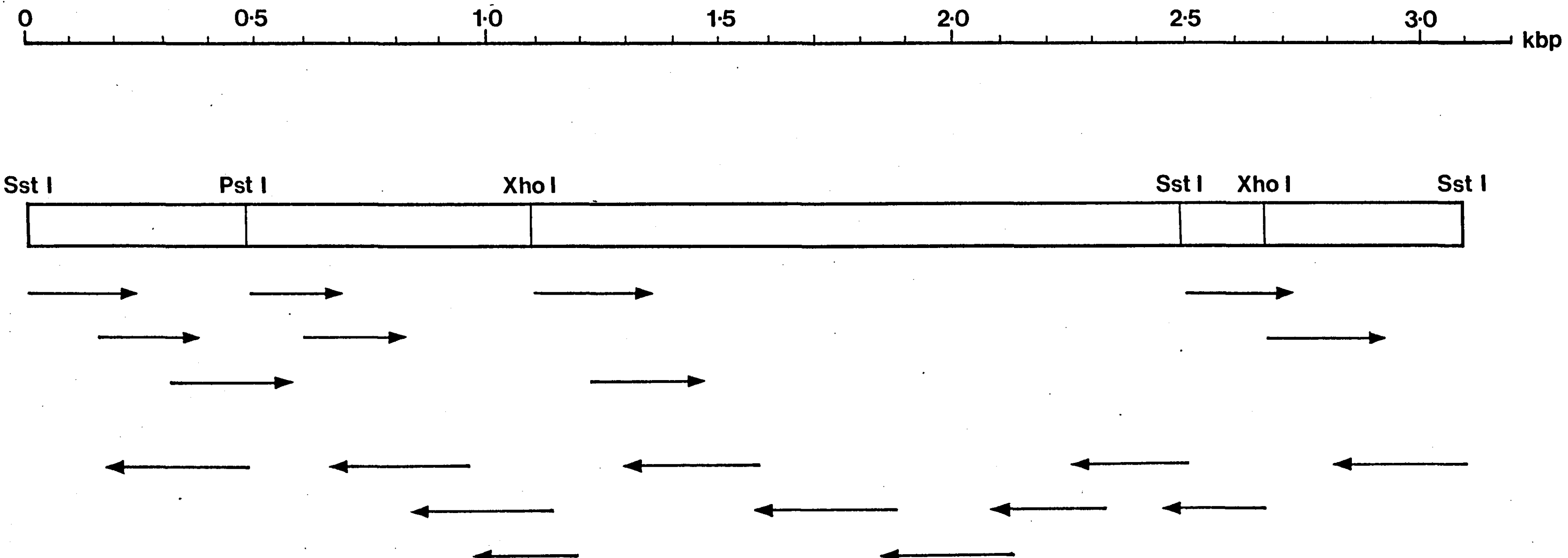


Fig. 3.7: Sub-cloning and sequencing strategy for Lambda 4.
The diagram represents 3.089 kbp. at the extreme right hand end of the L4 insert. The diagram is in opposite orientation to Fig. 3.5 (c), because retroviral sequences were found to be orientated from right to left in L4. Restriction sites used in the generation of sub-clones are shown. Specific fragments cloned into the pBluescript KS (+) sequencing vector were: 0.46 kbp. *Sst*I-*Pst*I, 0.62 kbp. *Sst*I, 1.07 kbp. *Sst*I-*Xho*I, 1.57 kbp. *Xho*I, 2.01 kbp. *Pst*I-*Sst*I and 2.47 kbp. *Sst*I fragments. Leftward-pointing arrows indicate the extent of individual DNA sequencing reactions on the upper DNA strand. Rightward-pointing arrows indicate the same for the lower DNA strand.

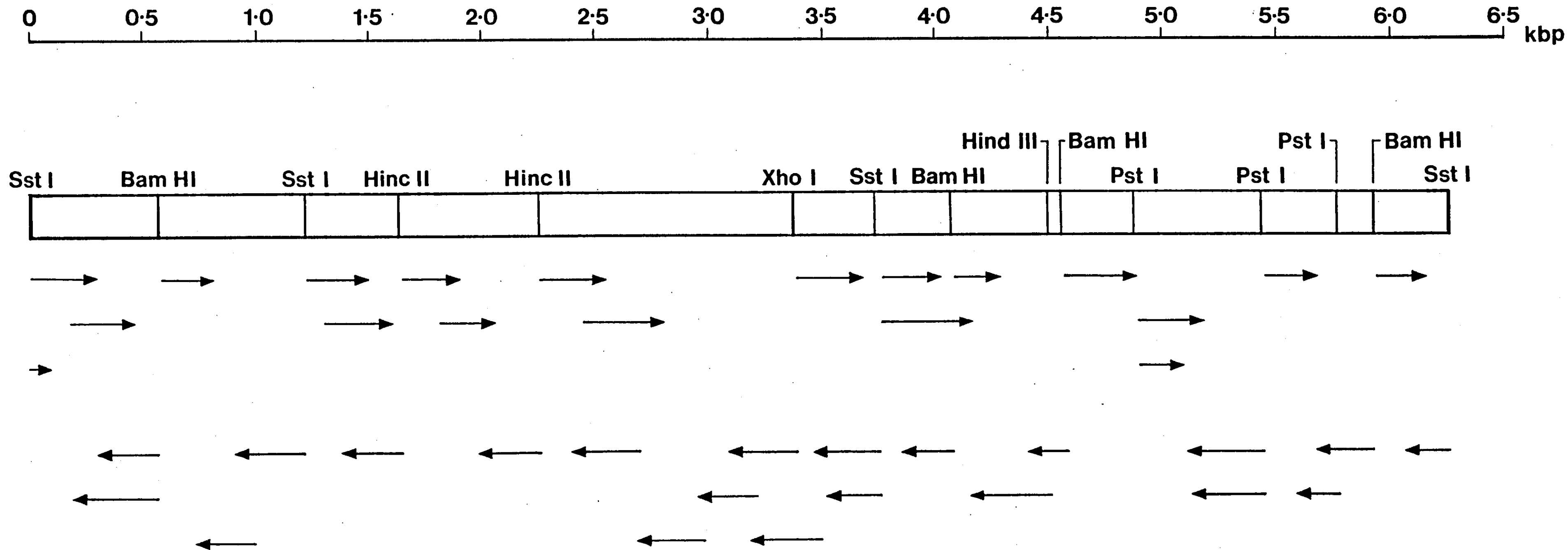


Fig. 3.8: Sub-cloning and sequencing strategy for Lambda 3.
The diagram represents 6.266 kbp. of the L3 insert, extending from the right hand end of the crosshatched area in Fig. 3.6 (c), leftward into sequences which did not hybridise to the FMuLV probe. The diagram is in opposite orientation to Fig. 3.6 (c), because retroviral sequences were found to be orientated from right to left in the L3 insert. Restriction sites used in the generation of sub-clones are shown. Leftward-pointing arrows indicate the extent of individual DNA sequencing reactions on the upper DNA strand. Rightward-pointing arrows indicate the same for the lower DNA strand.

3.6 SEQUENCE DATA ANALYSIS

3.6.1 SEQUENCE DATA ANALYSIS OF LAMBDA 4

L4 sequence data is presented in Fig. 3.9a. Fig. 3.9b shows an overview of its structure. The 3,089 bp. L4 element contains part of a highly defective retroviral provirus. Since retrovirus-like *pol* sequence begins within 30 bp of the right hand end of the lambda insert, it is presumed that the proviral element was cleaved during cloning. In any case, only part of the proviral genome has been cloned, the 5' LTR and *gag* sequences being absent. When the L4 sequence was compared with retroviral sequences in a large database (Microgenie^{TM} and GenEMBL), homology to *pol* and *env* sequences of several retroviruses was found. Highest and most extensive homology was to the *pol* gene of AKV murine leukaemia virus (sequence as determined by Herr, 1984). In the L4 element, sequences related to AKV *pol* extend from positions 28 to 586 (60.4% homology), 579 to 783 (67% homology) and 1161 to 1952 (62.9% homology). Shorter L4 sequences related to the TM-encoding part of AKV *env* extend from positions 2061 to 2169 (72.5% homology) and 2188 to 2290 (67.7% homology). Fig. 3.10 depicts as a dotplot the regions of nucleotide homology between AKV and L4. The large deletions and insertions in L4 compared with the intact AKV genome are evident as horizontal and vertical "steps", respectively. The AKV *pol* gene consists of components which encode PR, RT, tether, RNase H and IN (Etzerodt *et al.*, 1984; Johnson *et al.*, 1986; Werner *et al.*, 1990). L4 contains truncated parts of all of these components, with the possible exception of tether. Homology to regions encoding RNase H and IN was slightly higher than it was for PR and RT.

Translation of the L4 sequence demonstrated frequent stop codons in all three plus strand reading frames, consistent with its marked defectiveness. The longest open reading frame (ORF) was 403 bp in Frame 1 (positions 425 to 827). However, this ORF spanned a large deletion in L4 relative to AKV *pol* and is unlikely to be of biological significance. Predictably, in view of the many point deletions and insertions, homology to AKV at the protein level was shared among the L4 reading frames (see Fig. 3.11).

Fig. 3.9 (a): L4 sequence data [PTO FOR LEGEND]

```
                                      ┌──────── PR- and RT-encoding pol homology ────────►
1     GATCTTGGAA ATGGAGGAAC TGGACGTTAG GGGAGTCAGG ATTCGGCACC CTCCCAGAGC CCAGGGTAAC TCTTAGAGTG AAAGGGGCAA CCTCTTCATT

101   TCCTAGTGGA CAGTGGGGCA CAACATTAGG TTTTATTACA ACCCCAAGGG AAATTGGCAA ACAAGATTTC ACGGGTGCAA GGGGCCATGG GCACAAACAT

201   TTCTCATGGA CTACCTGAAG AACTGTGGGT CTCAGCATGG ACCAGGTATT CCATTCCTTC ATGGTCATCC CTGAGTGCCC CTACCTGTTG TTAGGTCGGG

301   ATTTGCTCAC CAAGAGGGGG CACAAATTCA CTTCCATCCC GAAGGACTAT TATCCTAACC AAAGAGGGGC ACCAATTCAG GTGCTTGTCC TGAGTTTAGA

401   AGACGAAATA TCGTCTCCAT TAAATGCCCT CAGCCTCAAT GACTGACATT GACCACTGGC TGCAGGAATT TCCCCAAGCG TGGGCAGAAA CTGGGGGAAT

                                                                                    ┌──────── pol (RH) ──►
501   TGGGCTGGCC TGGCACCGAC CAGCCATAGA CATAGAACTA AAGCTGGGGG CAGACCCTGT CAGGGTCCGC CAATACCCCA TGCCTCTTGT CTCCCTAATC

601   CCAGACTGGG ACCCCCCTCT GCATGACTGT CAAGAAATTT TGGCACAGGT GCACGGAATC AGAGCTGATC TCCAGGACCA GCCACTGGCC GAAGGCGGAT

                                                                                    ◄── pol ──┐
701   GCCACCTGGT ACACGGACGG CAGCAGTTTC GTTCAAGAAG GAATCAGATA CGCAGGGGTA GCCGTAACCA CGGAAATGGA GTTCCCATCC GGGGGCACGA

801   CCCTGCCAGA CACCCTCAAC TATACTGAGG CTGACTTACA CTGGATCAAA CGCTTGCCTA TGACCCAGTG CTTGCGTGGC TGGTGGAGGG CCACAGACTC

901   CAGCATTATC CTGCCAGAGG AACTAGGACG GTGAGTCCTA TCCAAGATGC ATCGGAGTAC TCATATGGGA ACAAGAAAGA TGGAAGACCT CATACGACAT

                                        RETROPOSON
1001  GCAAAGATCA CTATTAAAGA CTCTCGAAAA AAAATTTAAA AAAATAAATA AATAAATAAA TAAAGACTCT CGAGCAAAGA TCGAGCAGAT TGTGGCAAGC

                                                                ┌──────── IN-encoding pol ────────►
1101  TGTCACGCAT GCCAGTTAAC TAATGCCACC GCCCATGGAT CTAACCTGGG TGCCTGGCTC CGAGGGGACC GCCCAGGAGC CTACTGGGAA GTGGACTTCA

1201  CTGAGGTAAA ACCTGGAAAG TACGGATACA GGTATTTACT AGTGTTTGTA GATACTTTTT CAGGATGGAC AGAGGCATTT CCCACCAAAC ATGAAAACTG

1301  CACAGACTGT GACCAAGAAA CTGCTGGAAG ACATCTTACC CAAGGTATGG TTTTCCTGTT AAGATTGGAT CAGACAACGG CCCAGGATTC ATCTATAAGG

1401  TAACACGGGG AGTGGGCCTA GTACTTGGGG AAAATTGGAA ATTACATTGT GCATATAGGC CCCAAAGCTC AGGACAGGTA GAGAGGATGA ACAGAACATT

1501  AAAGGAGACC TTACTAAATT AGCCCTGGAG ACTGGCGGGG ACTGGGTGAC ACCCCTCCCC TACGCCCTAT ATAGGGTGAG GAACACCCCA TATAAGATGG

1601  GATTTACCTA CTATGAAATC ATGTTCGGTG TTCCTCCACC TATTATCCCC AATTTAAAAC ATGAAGTGAC TTGCTGAATT TGATGATCAC CAACTTTCTT

1701  TTCTTCCTCC TAAGTGTTAA AACGAACCCA AGAAGCTGTG GCCTAAGCTG AGGGCCCTCT ACGAGACTGG GCCACCCCCG GGCCCCCATC GATATCGGCC

1801  GGGTGACTGG GTGTATGTGT GGAGATACCA ACACCAGACA ATTCAACCTC GCTGGAAGGG ACCGTACATC GTGATCCTGA CCACTCCCAC CACTCTCAAG

          env begins ┌►   ◄── pol & 5' env homology ──┐
1901  GCCCGACGGG ATTACTCCCT GGGCCCACTA CACCCACGTC TGTCCAGCCA GCTGATCCAC ACGCTGTTCT CAAGGACTTT GTTCCAGAAT GGAAAAGCCA

                                                    ┌──────── TM-encoding env ────────►
2001  ACCAGACAAG GACAATCCCC TAAAGCTAAG ACTGCGCCGT TCTCACTTAT TTCCCACCTC CCTAGTCTGA GGTTGTCCTT CAAATAAAAG GGGATTAGAC

                                                                          ~54 bp. deletion
2101  TTAGTCTTCT TACAACCAGG GGGGGTGGTG GTTATGTGCT GCCTTAAACA GGATGTTGCT TCTTTGCTGG GATGTAAATG GGAGCGAGAA CAGCAACAAG

                                                                                                 P.P. TRACT
2201  ATTGGTTTCG TAATTCCTTG GTTTTATCAT TCCCCCTGGC ACTCTGAGAG CAACCTGGAC CCATGATTGG GTCCTTCCTT AGGTTCCTCT GAGAGGGGAA

      ┌──── 3' LTR ────►
2301  AATGTGGAGG CCAAGAAAAT TAAGGCCATT CCATTTTAAG TTCAGCGTTA CACAAGTACA GCCTACCCAG GCCCCCTGTG AATAAGAGCA GAATTTACCT
         IR

2401  TACTTCAGTT ACAGGGGAAA AAAACAGCTT TCAGCTTAAC GCCATAGAAA GCCCCCTATT AGAATGAGAA GAGAGCTCAA TGCCCTTGAA AGCCCCATAT

2501  CAGAATGTAA ACAGAACTTG AGAAATTCCT CCACCCCTTC TGGAGGTCCC CTAGACCAGC CCATAAAACT AAGCTGAAAC CCACCTTGGG GTCCAAGTCC

                                                     Poly(A) signal
2601  CTGCTCCGCT GTGTCGGGTA CACTTGGACC CAAGCTCGAG CTTGTTAATA AACCCTCGTG TGTTTGCATC GGTGTCGGCT TTTTGGTGGT TTCTCAGATA

        ◄── 3' LTR ──┐
2701  CGTAATCTTG CGCACAACAC TCTGACCTCT GGGGCCTCAC CTTCCCTGCA GGGACATGGT GAGCTGAGGT TGGCACACAG CCTCTGCAGC CAGCCCACCT
            IR                homology to L3

2801  AGGTTCAAAT ATCAACTCTC GTGCTCACTG ACTGTATTGT CTGAGGCAAG GGATGTCATC TTTCTCTGCT TCAGATTCTT TGTATCAGAG GAGAGGGCAC

2901  CCAGAGCAGC TCTGTGGGCA ACCAGAGATC AACTGGTAAC TGTTTGGCAA CCAGCTTGTT TGGGGGAGAC ACAGACCCTG CCTGATGGGC TGACCAATTT

3001  CCATGGCGTA CTCCCGCCAT GGCTGGTCCC AGCTGCCCAC GGCATAATCC TCTCTCACAA AGTCCCCAAA AGTTACCAGT CAAGAGCTC
```

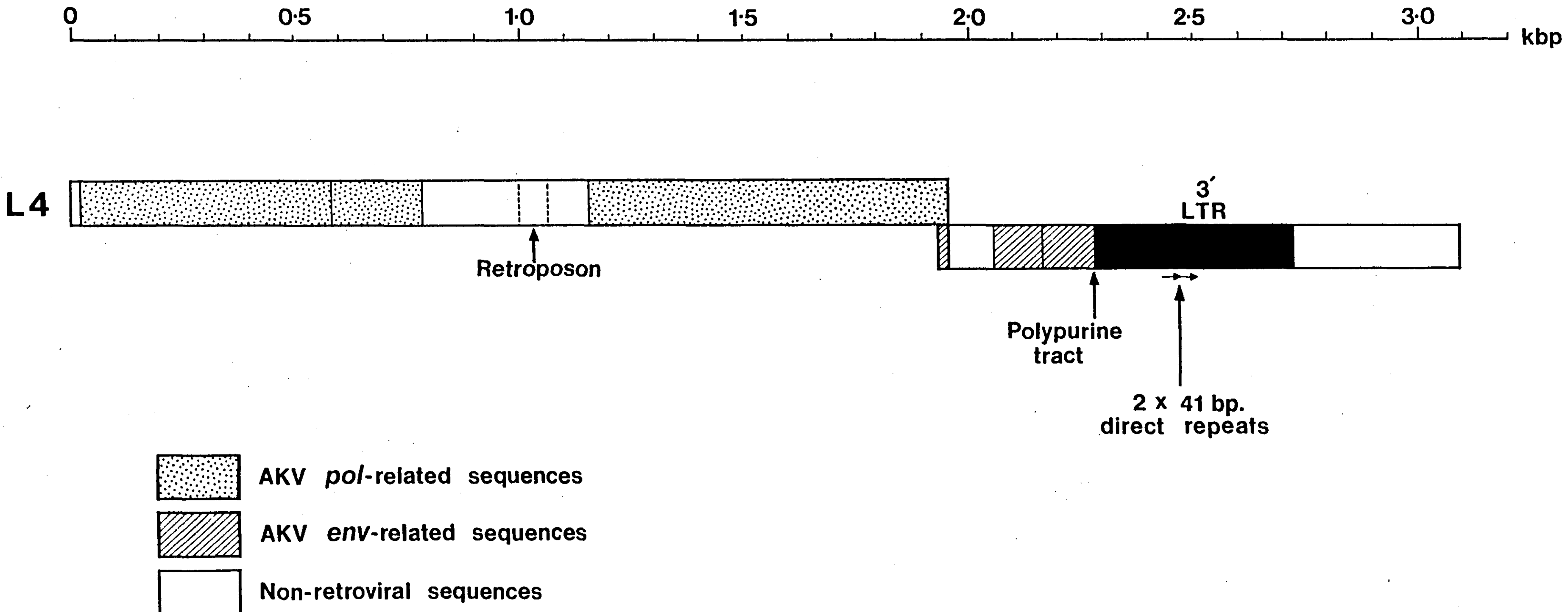


Fig. 3.9: a). L4 sequence data. [SEE PREVIOUS PAGE] Annotations show features of interest within the 3,089 bp. DNA sequence.
b). A diagram showing structural features of the defective retroviral element, L4.

Abbreviations: PR, protease; RT, reverse transcriptase; RH, RNase H; IN, integrase; TM, transmembrane protein; P.P., polypurine; IR, inverted repeat; poly(A), polyadenylation.

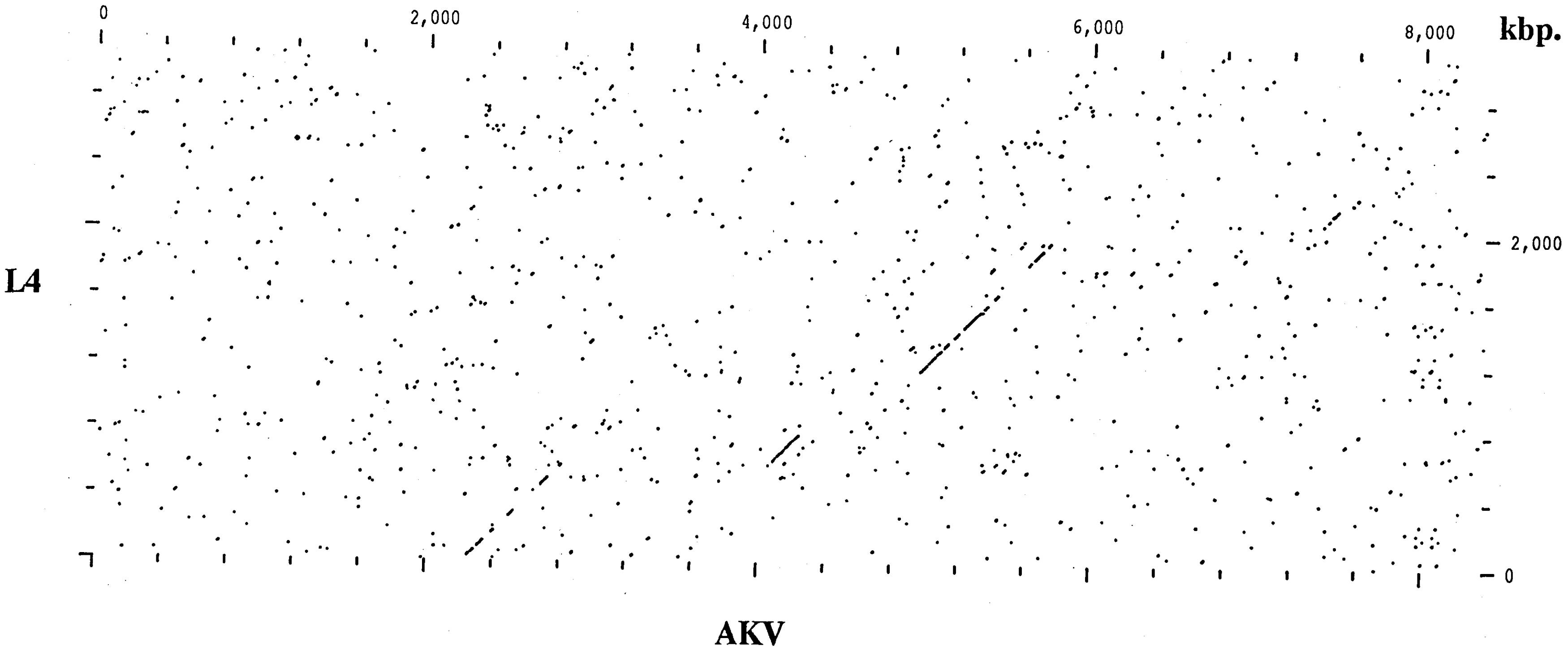


Fig. 3.10: A dotplot to show homology of the defective L4 proviral element to the murine endogenous retrovirus AKV at the nucleotide level. The 3,089 bp. L4 element is compared with the entire 8,374 bp. AKV genome (Herr, 1984). Regions of homology appear as broken diagonal lines. Deletions in L4 relative to AKV are evident as horizontal steps separating the diagonal lines. Generated using G.C.G. COMPARE, Window size: 21, Stringency: 14.

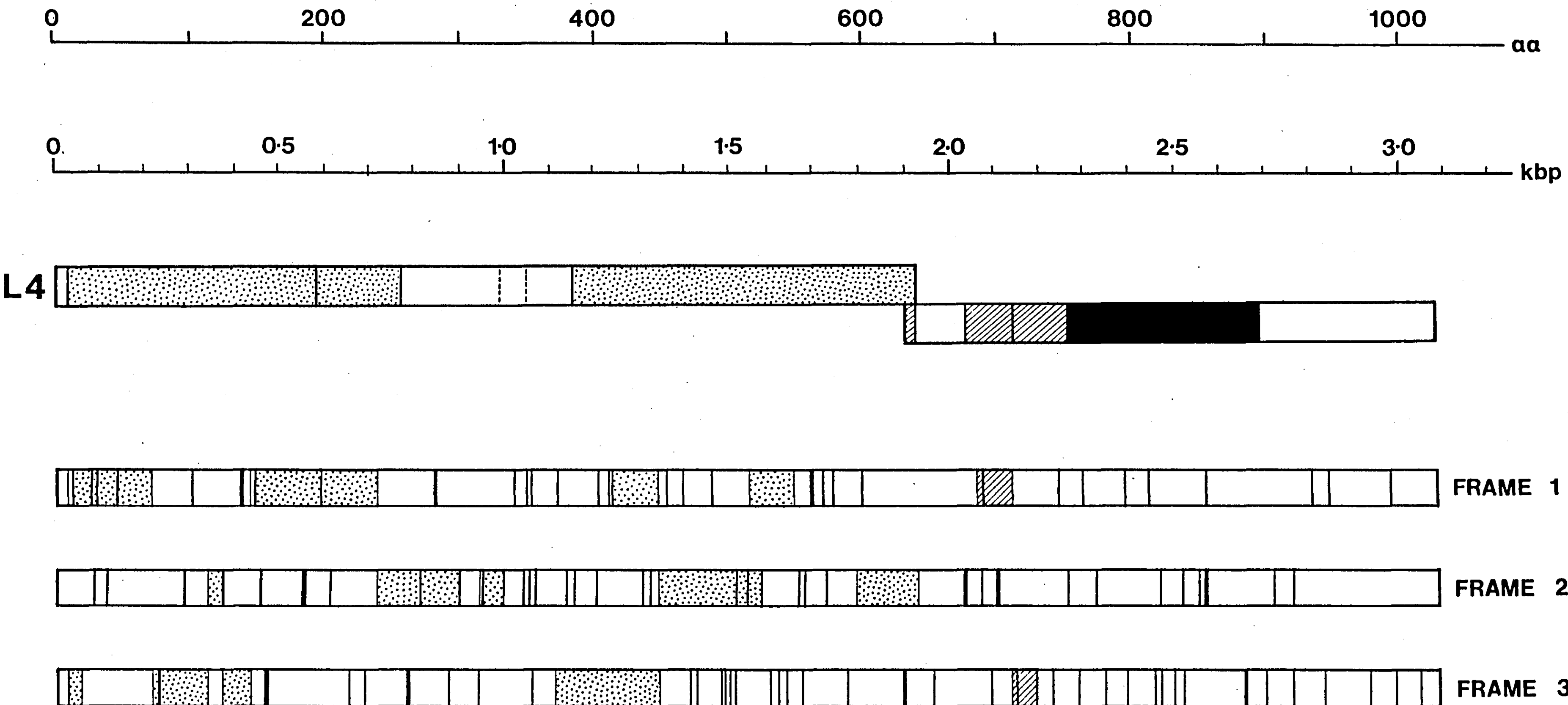


Fig. 3.11: Homology of the defective L4 proviral element to AKV at the protein level. The amino acid sequence of L4 was deduced from its nucleotide sequence for all three (+) strand reading frames. The resulting amino acid sequences were compared with that of AKV. Homology was found to be shared among the three reading frames. Stippled and hatched regions within the three reading frames represent regions of homology to AKV *pol* and *env*, respectively. The positions of stop codons in each frame are marked by vertical lines.

The L4 sequence contained other notable features. A small non-viral retroposon was found between positions 1015 and 1073, bounded by 13 bp. perfect direct repeats. Since this retroposon consisted mostly of a poly-A tail, it is likely that it originated in mRNA. This retroposon is discussed further in the section below comparing L3 with L4. Downstream of the truncated *env*-related sequences was a polypurine tract and 3' LTR. These structures were definitively identified only after comparison with the L3 element, which has both 5' and 3' LTRs. Beyond the 3' LTR of L4 there was 368 bp. of flanking cellular sequence, part of which was homologous to the human fibrinogen gamma and gamma-prime chain genes (positions 2776 to 2881, 66% homology). The first 43 bp. of this flanking sequence was homologous to the 3' flanking sequence of L3 (26/43 bp. correct, 60.4% homology). The L4 element is compared with L3 in Section 3.6.3.

3.6.2 SEQUENCE DATA ANALYSIS OF LAMBDA 3

The 6,266 bp. nucleotide sequence of L3 is shown in Fig. 3.12a. Fig. 3.12b shows an overview of its structure. L3 contains a truncated retroviral genome, bounded by non-identical LTRs. 5' cellular flanking sequence extends from position 1 to 473, with no significant homology to sequences in the Microgenie™ database. A small mRNA-derived retroposon occupies positions 477 to 599. The retroposon is bounded by imperfect 12 bp direct repeats (10/12 bp. correct) and its poly-A tail is orientated towards the start of the L3 sequence (*i.e.* the minus DNA strand corresponds to the retroposon transcript). Interestingly, this retroposon has been integrated just beyond the start of L3 retroviral sequence: its direct repeat sequence incorporates part of the inverse repeat (IR) which defines the start of the L3 proviral 5' U_3. A true case of "Big fleas have little fleas upon their backs to bite them......"!

The retroposon insertion within the 5' U_3 IR and point mutations in the adjacent sequences make it difficult to define precisely the 5' terminus of the L3 provirus. However, extrapolation from the intact 5' terminus of the 3' LTR suggests that the 5' LTR begins at position 474, extends to 476 and is there interrrupted by the retroposon. Retroviral sequence starts again at position 588, at the beginning of the 3' (plus strand) direct repeat of the retroposon.

Fig. 3.12 (a): L3 sequence data [PTO FOR LEGEND]

```
1     GAGCTCAGCT CTCAAATGGA AGCATGCCTC TAAGCCAGCT TTCTACTAGC TGCGTTCTTA TGGCCGTGAG AACGTAGCAG CATTGTAAGG CAGGAGGTTC

101   CATGAACACC ACTCTACTGC ACCTGGAGAC ACAACCATTG AAGCCTGCAG CTTGCCCCTT GCGTCCCATG AGCTTCCGGC AAACATCTCT ATTGAGGATG

201   TTCGAAACTG CATCTGCAGG CAGTGTGGAA CTGTGAGGGA CTGGGTGTGG GCTCCCCTCT CAGAATCAAT TTCCCCTGGA CGCTGCGTTC AACTAGCTTC

301   TCTTGTATGG TAGTTAGTAT GTAGTGCCAA TGTATGAGAG GAACTCCTCG GAAGTCGCTG GATTGAGCTT CAAGACACTT GTCTCCTGGC CTGCACCTTC

                                                                                           DR        5' LTR→
401   CAATTACCTC TCACACCACT TCTGTCACGC AACAAACCTA GGCCTGTTAG ACGCGAAGTG TACATGCAGT CAGTGTGGGT ACCAGGAATT TTTTTTTTTT
                                                                                         IR
                                           RETROPOSON                                                            DR
501   TTATTTATGA TAGTCACACA CAGAGAGAGA GAGGTAGAGG CAGAGACATA GGCAGAGGGA AAAGCAAGCT CCATGCACCG GGATCCCGGG ACCCAGGAAA
                                                                                                          IR (cont'd)

601   TTTAACAAAA ATCCCCTACC CCTGGACAAG CAGAGCAGGA CTGATTTCAT TTTGTGCTAC ACCTGCCACC TCCTGTATGA CCCCCACATA ACCTGCTTAT

701   TGCTTAAGGC GCTGCCCCAC CCTAGTCAAG CCGCTGGGCA TACCCTAATC GGAAATCGGC TCATAACAAC GTAACCCTGC TTTGTGCCCG CCAAAACTGT

801   GCGCCAATTC TGACCAAATA GGCCAGGTCA AGTGGATACT ATAGGGTAAG ATGTAATTCA ATCGGCCACC TGCGTGTGGA CCGACATGAC TGTGCAACTT

901   TCTGTGTATC CCATGCCACT GGCCCTTCTA AAGCTGTACG CCTCTTAGTC TCGGGGTCCA AGTCCCTGTC CGCTCTGTCA GGTGCTCTTG GACCCAAGCT

        Poly(A)-signal                                                            ←5' LTR    PBS
1001  CCAGCTTATA AATAAACCCT CGTGTGTTTG CATCTGTGTC GACTCCTTGG TGGTTTCTCG GATTTGCAAT CTTGGGCACA ACATTTGGGG GCTCGTCCGG
                                                                                    IR

1101  GATCTGAGAG ACCTCCAGGA CCCTATCCGG AGGGTTTCAC GCGTGGTGAG TGCACTCAAC TTTTTCCACC TTTTGAGTGC AAGTTCTCTG AGAGCTCTAA

1201  CCTGTAGGAA TTCCAATCTG TATTAGGTCG GCATTGGCTT AGGTGGACGC GCTGGTGAGC CATCGACCGG GGGTGTTGAG GAGACGTCCC TTGCCCCTGC

        TANDEM REPEATS (type 1)
1301  CTGGAGGACA GAGGTCCTCA TCTGTAAGGA GGACGGGGTC CTCATTTGTA AGGGGGATGG AGGTCTTCAT TTGTATGGAG GATGGGGGTC CTCACCTGTA

1401  TGGAGGACGG GGGTCCTCAC CTGTAAGGAG GATGGGGTCC TCATCTGTAT GGAGGACAGG GGTCCTCATC TGTATGGAGG ATGGGGGTCC TCATCTGTAA

1501  GGAGGGCTGG CTGCCAACCC TTCGGATTAC TTTCTGTCTT GTCTCCGCAG CCACACTAGA ACGTTTGTCT GTTACTGTGC TCTTCTTACC TCTTTGCGCG

        TANDEM REPEATS (type 2)
1601  CTGGTCTGTT ACCTGTTTGT GTGCTGGTCT GTGTTACTGT GTTCTTGTTA ACTGTGTGTT TGACTGTGTT ATTTGTGAAC TGTCATAACT ATCTGTGCCT

                  MA-encoding gag homology begins→
1701  TGGTGTGGAC GGGCGACATA ATGGGGCAGA TGCAAACCAT ACCCTTGTCT CTTATGCTCT CCTACTTCTC GGATGTTAGG GAAGCACTCT TATTCTGAGC

1801  ATAGACATAC AGAGAGAGAT TCCAAATTTA TTGCATATCA GAATGGCCAA CCTTTGATGT GGGATGGCCC CGAGAAGGAA ATATTCACCG ACCTATTATC

1901  TTACAGGTTA AGGCCGTGAA CTTCAGGGAC AAACCAGATG GCCACCCTGA CCAGATGCCC AACATCCTGG CCTGTCAGGA CATGATTGAG AATTCCTCTC

                                                    CA-encoding gag homology begins→
2001  CTTGGCTAAA ACCCGTTTTT ACTCCCCAAA GCACACAACT CGCCTGAAGG TAAGGAGAGA CTCCGGGTCC ACCGCCAGAC TCTAATGGCA GGTCTCCAGG

2101  CTGCCGCGTG CAAGACTACC AATCTGGCCA AGGTCTATGA TGTGAGACAG GGTAAGGATG AGAGTCCAGC AGCCTTCTTA GAGGGAGCAA TAGAGGCTTT

2201  TAGGCAGTAC ACCCCTATGA ACCCAGAAGC CCCGGAAACA AAGGCCGCAA TTATCATGGC TTTTGTTAAC AAGGCCGCTC TGAACATTAA AAAGAAATTG

2301  CAAAGAGTAG AGAGACTGGG AGAGAAGAGC TTGCAGGACT TAGTGATAAT AGCAGAACGA GTCTATAATA GAAAAAGTCC GGAAGAACTG CATGTTACCA

                          RETROPOSON                       → pol homology begins
2401  CCCCACACGC TATTAAAGGG GTACTCAAAA AAAAAAAAAT AAAGGGGTAC TCAAACAGCC CCCAGACAGC TGGATCAGCA ATGCCCATCT GACCCATTAT
                         DR                      DR

2501  CAGACCCTAC TGCTAAACCC CACCAGAATT TTATTTAAGC CACCGACAAC CCTGAATCTG GCAACGCTAC TCCCTAACCC AGACTGGGAA CCCCCTCTGC

2601  ACGAGTATCA AGAAATTTGG GCACAGGTGT ACGGAATCAG CGCTGATCTC CAGGACCAGC CACTGCCGAA CTCCGATGCC ACCTGGTACA CGGACGGCAG

2701  CAGTTTTGTC CGAGAAGGAG TCCGAATTGC GGGGGCAGCC ATACCATAAC CACGGAGACA GAGACCGTCT GGGCGGAGCC ATTGACAGCC TGAATGTTGG

2801  CTCAATGGGC AGAACTGATC CCATAGGCCA AGGCACTGAC CATGGAGGAA GGTAAACGAA TAAACATCTA CACCGACAGC AGGTACGCCT TTGCCACCGC

2901  TCCCATTCAC AGAGCCCTTT ACAGAGAGAG AGAGAGCTTC TGACAGCAGA GACAAAGACT GTTAAAAACA AAACGGAGAT TCTTGAACTC CTGAGGACCC

                                                                              pol interrupted
3001  TCTGGCTGCC CAAGGCCCTG GCCATCATCC ACTGTCCGGG GCACCAAAAG GCAGACACAC CAGTAGCCAG GGGAAACCGG CTAGCCGATT TAAAAGCAAA

3101  GAAGGTGTCC CTTTTGGTGA CCCAGGTCTT AGCAACCACG CTACCTGATC CGGGGGAACG GAACCTGCCG GACACCCCCA ACTATAGTGA TGCTGACTTA
```

Fig. 3.12 (a): L3 sequence data (continued) [PTO FOR LEGEND]

```
3201  CACTGGATCA AACACTTGCC TATGACCCAG TGATTGTGTG GCTGGTGGAG GGCCACAGAC TCCAGCATCA TCTTGCCAGA GGAACTGGGA CGGCGAGTCC

                                                                   DR         RETROPOSON          DR
3301  TATCCAAAAC GCATCGGAGT ACTCACACGG GAACAAGGAA GATGGGAGAC CTCATTCGAC ATGCAAAGAT CACTATTAAA GACTCTCGAG CAAAGATTGA

                                                                                  ------> pol resumes
3401  GCAGATCGTG GTGAGCTGCC ACTCATGCCA GTTAACTAAT GCCACTGCCC ATGGATCTAA CCCAGGCACC TGGCTCCGGG GGGACCGCCC AGGAGCGTAC

3501  TGGGAAGTGG ACTTCACTGA GGTAAAACCT GGAAAATATG GATACAGGTA TTTATTAGTG TTTGTAGATA CTTTTTCAGG ATGGACAGAG GCATTTCCCA

                                                   34 bp. deletion
3601  CCAAACATGA AATGGCACAG ACCGTGACCA AGAAACTGCT GGAAGACATC TTGACAACGG CCCAGGATTC ATCTCTAAGG TAACACAGGG AGTGGCACGA

3701  GTACTTGGGG CAGATTGGAA ATGACATACT GCATATAGGC CCCAGAGCTC AGGACAGGTA GAGAGGATGA ACAGAACATT AAAGGAGACC TTAACTAAAT

3801  TAGCCCTGGA GACTGGCGGG GACTGGGTGA CTCTCCTCCC CTTCGCCCTA TATAGGGTGA GGAACTCCCC ATATAATATG GGACTGACTT CCTACGAGAT

3901  CATGTTCGGT CTTCCTCCAT CTGTTATCCC CAATTTAAAA CCTGAGGGGC TTGCTGAATT TGATGATCAC CAACTTCTTT TCTCCCTCCA AATATTACAA

4001  CAAACCCATG AGCAGGTGTG GCCTAAGCTG AGGGCCCTCT ATGAGACTGG GCCACCCCTG GATCCCTATC AATATCGGCC AGGTGACTGG GTGTACGTGC

                                                                                               env begins -->
4101  AGAGATACCA ACACCAGACA CTTCAACCTC GCTGGAAGGG ACCTTACACC GTGATCCTGA CCACTCCCAC TGCTCTCAAG GTTGACAGGA TTACTCCCTG

      <--pol & 5' env homology--
4201  GGTCCACTAC ACCCACATCC GGCCAGCTGA CCCACACATC GTTCTCAAGA ACTTTGTTCC AGAATGGAAA AGCCACCCAG ACAAGGACAA TCCCCTAAAG

                                ------------TM-encoding env ------------>
4301  CTAAGACTGC GCCCTTCTCA TTTATTTCGC ACCTCCAAGA CTCCCTAGTC TGAGGTTGTC CTTCAAAATA GAAGGGCATT AGACTTAGTC TTCTTACAAC

                                                                                 ~54bp. deletion
4401  AAAGGGTGTG TGTGTGGTGT GTGTGTGGGG GGTTATGTGC TGCCTGAAAC GTAGAATGTT GCTTCTTCGC TGGAATATAA ATGGGAGCGA GAACAGCAAC

                                                                                                  P.P. TRACT
4501  AAGCTTGGTT AAAATCCTGG TTTAATCATT GGGGGTGGCT CACTCTAAGA GCAACCTGGA CCCTTGATTG GGTCCTGCAT AGGATCCTCT GAAAAAGGGG

        ---3' LTR --->
4601  GGAATGTGGG ACCCAGGAAA TTTAACAAAA ATCCCCAACC CCTGGACAAG CAGAGCAGGA CTGATTCCAT TTTGTGCTAC ACCTGCCACC TCCTGTATGA
            IR

4701  CCCCCACATA ACCTGCTTAT TGCTTAAGGC GCTGCCCCAC CCTAGTCGAG CCCCTGGGCA CACCCTAATC AGAAATCGGC TCATAACAAT GTAACCCTGC

4801  TTTGTGCCCG CCAAAACTGC GCGCCAATTC TGACCAAAGT AATAGGCCAG CTCAAGCCAC CTGCGTGTGG ACCGACATGA CTGCAGCTTT CTGCGTATCC

4901  CATGGGCCAC TGGCCCTTAA AGCTGCTACA CCTCTTAGTC TCGGGGTCCA AGTCCCTGCT CCGCTGTGTC GGGTATATTT GGACCCAAGC CCGAGCTTGT

      Poly [A] signal                                                   <-- 3' LTR --
5001  AAATAAACCC TCGTGTGTTT GCATCGGTGT CGGCTCCTCG GTGGTTTCTT GGATTCGCGA TCTTGGGCAC AACACAAGGG TCTGCTGCCC TGACCTCCCT
                                                                            IR       homology to L4

5101  TCCAGGGCCA CGCTGACTGC CCCACAGGCC CTCTGTGCCC GAGGCACACA GGCACGGGGA TGCGGGGAGG GGAGGCAGTG CCCCAGAACC CGGCCAGCCC

5201  AGGCCCCCAG GACAGAGGGA GCACAGACCT CCCAGGTCGC TTTGCACAGA CCCTGGTGAC CTGAGGTTCC CTGGCTGTCA CTCTGTCCCA GCCCCTCCTG

5301  GCCCGTGGGG CACAGGCTGA GTCACAGTAA AGTCTGTGGG GTCCTGACCC TGGCGTCTGG CCGCCCCTGG AAGGCACGTG GGGACGTCAG CTGTGTGTGT

5401  GCCTGGCGGC GCCCCTGCAG GGGCTGGTTG GGGCCTGGGT GGTGCCCCCT CCTCCTCTCT GGGGGTCACT CCACGCCCCG TGCACAGAGT TCCTGGAGAC

5501  ACAGAAAAGG GCTGGCTGTG CATTTTCACA CCACAACCCT TTATTGACTT TCTACAGCAA ACGGGCTCAG AGTGGAGGGT ATGCGCCTTC CCCTGGGGAG

5601  GGAGACTTTT CTGTCCTTTA GGGGAGGGCA AGGATCACAG GGGGTGGGGG GGCGGGGGCT GGAGTCCCCC ACCTGGGAGC ATGTGTGGAC TCCGTGTGGG

5701  GGGCGTGTGT CCTGAGGTTT CAGCCTGAGA GACGGTTCTG AGGACAGCAG AGCTGACCTG GACCTGCAGC TCATGGGCTC CTGCGCTGGG TCCCCGCAGG

5801  GCAGCTGGAC CCCGGCCTCG CCCCCTGGGG CCCCAGCACC TGGAGGAGCC TGCCTCCTGG TGTGGGTGCA GGTGGGGCCC ACCCGCCTAC CCCTGGTCTC

5901  CCCACCCTGG GGTGTCCCGG GATCCCCAGG GCTGCCCTAG GTCAGCGACC TGGATCTGTG GGGCACACCC AGGGCCCAGG AGGCCACATC CAGCTGTGAG

6001  AGGGCTCGCT CAGCCTGGGG CCAGAATCCA GGCCTGGCCA GGGCCATGGG GGGCTGGATC ACATTAGTCT TTGTGTTGGG CCTGGGGGGG GAGGCTCATA

6101  TCACCCCAGC CCCCACAGAG CCCGGGGCCA GGCCAGCACA GGTCCCTCTG GGTTAGGGGC TGGTGTCCCT GTGCAGGGTG TGCCGCCCTC CAGGAGGGAC

6201  AGTGGGAGGA GGCAGGGCAC AGTGCATCCT CTGCTCCACC CACCTGCACC CACCTGTCGG GAGCTC
```

b).

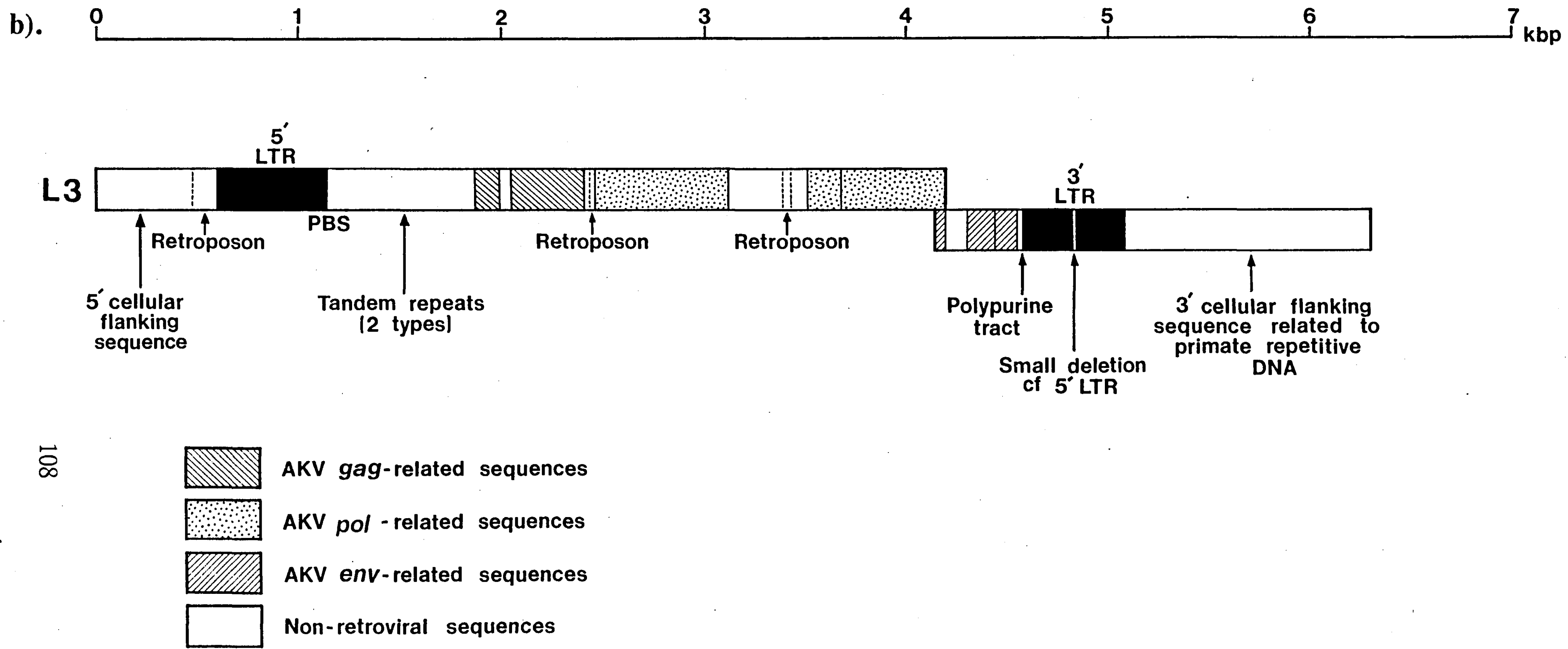


Fig. 3.12: a). L3 sequence data. [SEE PREVIOUS PAGES] Annotations show features of interest within the 6,266 bp. DNA sequence.
b). A diagram showing structural features of the defective retroviral element, L3.

Abbreviations: IR, inverted repeat; DR, direct repeat; LTR, long terminal repeat; poly(A), polyadenylation; PBS, primer binding site; MA, matrix *gag* protein; CA, major capsid protein; TM, transmembrane protein; P.P., polypurine.

The 5' and 3' LTRs of L3 share 92.8% homology and are 501 bp. and 471 bp. in length, respectively. There is a 33 bp. deletion in the 3' LTR compared with its 5' counterpart. There are six other small gaps (1 to 4 bp.) in the aligned sequences (Fig. 3.13). Comparison of the L3 LTRs with retroviral LTRs in the Microgenie^{TM} database revealed no significant homology. Rather, it was the similarity of the 5' and 3' ends of the L3 sequence and identification of the tRNA^{pro} PBS which allowed initial identification of the LTRs. Each LTR is bounded by 16 bp. imperfect IRs (11/16 bp. correct, the external 5 bp. being a perfect match). As is usual for oncoviruses, the first two base pairs of the 5' LTR (AA) and the last two of the 3' LTR (TT) have been removed, so that the integrated provirus begins with TG and ends with CA. There is no 4 bp. direct repeat in the immediately adjacent flanking cellular sequences, but the 5' terminus is flanked by the sequence 5'-cag-3' and the 3' terminus by 5'-CAAG-3'. Thus, a single point deletion to remove an A residue might have occurred in the 5' flanking sequence.

Analysis of the LTR sequences is hampered by the fact that the U_3-R and R-U_5 junctions are unknown. Assuming that R and U_5 are each approximately 70 bp. in length (as is the case for AKV and several other C-type retroviruses), there is neither a TATA nor a CCAAT box in the appropriate region of U_3. However, further downstream, 75 bp. from the 3' end of both LTRs, a TTATAAATAAA sequence is present. Although this sequence could include a TATA box; judging from its position, it is more likely that the AATAAA component represents the polyadenylation signal near the 3' end of R.

Immediately downstream of the 5' LTR is the PBS which is a 21/22 bp. match to the 3' end of murine tRNA^{pro}(2), the terminal 18 bp. being a perfect match. This tRNA species is used as a primer by many of the oncovirinae (including AKV). 3' to the PBS there is a 400 bp. region of short tandem repeats, or minisatellites. Interestingly, there are two distinct, adjacent groups of these elements. The 5' group consists of nine tandem repeats of a highly conserved 25 bp. motif. The last repeat element is truncated. Individual minisatellites within this group vary by only one or two base pairs (Fig. 14). A sequence closely related to this 25 bp. minisatellite is present as a single copy in AKV *gag*, near the junction of MA- and p12-encoding sequences (Fig. 3.15a). The second (3') group of

```
  580 GGGATCCCGGGACCCAGGAAATTTAACAAAAATCCCCTACCCCTGGACAAGCAGAGCAGGACTGATTTCATTTTGTGCTACACCTGCCACCTCCTGTATG
      ||||    ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
 4600 GGGAATGTGGGACCCAGGAAATTTAACAAAAATCCCCAACCCCTGGACAAGCAGAGCAGGACTGATTCCATTTTGTGCTACACCTGCCACCTCCTGTATG

  680 ACCCCCACATAACCTGCTTATTGCTTAAGGCGCTGCCCCACCCTAGTCAAGCCGCTGGGCATACCCTAATCGGAAATCGGCTCATAACAACGTAACCCTG
      |||||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||| ||||||||| |||||||||||||||||| |||||||||
 4700 ACCCCCACATAACCTGCTTATTGCTTAAGGCGCTGCCCCACCCTAGTCGAGCCCCTGGGCACACCCTAATCAGAAATCGGCTCATAACAATGTAACCCTG

  780 CTTTGTGCCCGCCAAAACTGTGCGCCAATTCTGACCA....AATAGGCCAGGTCAAGTGGATACTATAGGGTAAGATGTAATTCAATCGGCCACCTGCGT
      |||||||||||||||||||| ||||||||||||||||    |||||||||| ||||                                 |||||||||||
 4800 CTTTGTGCCCGCCAAAACTGCGCGCCAATTCTGACCAAAGTAATAGGCCAGCTCAA.................................GCCACCTGCGT

  876 GTGGACCGACATGACTGTGCAACTTTCTGTGTATCCCAT..GCCACTGGCCCTTCTAAAGCTG.TACGCCTCTTAGTCTCGGGGTCCAAGTCCCTG.TCC
      |||||||||||||||  |||| ||||||| |||||||||  ||||||||||||  |||||||| ||| |||||||||||||||||||||||||||| |||
 4867 GTGGACCGACATGAC..TGCAGCTTTCTGCGTATCCCATGGGCCACTGGCCCT..TAAAGCTGCTACACCTCTTAGTCTCGGGGTCCAAGTCCCTGCTCC

  972 GCTCTGTCAGGTGCTCTTGGACCCAAGCTCCAGCTTATAAATAAACCCTCGTGTGTTTGCATCTGTGTCGACTCCTTGGTGGTTTCTCGGATTTGCAATC
      ||| |||| |||    |||||||||||| | ||||| |||||||||||||||||||||||||| |||||| ||||| |||||||||| ||||| || |||
 4963 GCTGTGTCGGGTATATTTGGACCCAAGCCCGAGCTTGTAAATAAACCCTCGTGTGTTTGCATCGGTGTCGGCTCCTCGGTGGTTTCTTGGATTCGCGATC

 1072 TTGGGCACAACATTTGGG 1089
      ||||||||||||   |||
 5063 TTGGGCACAACACAAGGG 5080
```

Fig. 3.13: Alignment of the L3 LTRs. The 5' LTR nucleotide sequence is shown above that of the 3' LTR. The 33 bp. deletion in the 3' LTR versus its 5' counterpart is evident. Numbers indicate distance from the beginning of the L3 sequence.

```
1  GGAGGACAGAGGTCCTCATCTGTAA
2  GGAGGACGGGG-TCCTCATTTGTAA
3  GGGGGATGGAGGTCTTCATTTGTAT
4  GGAGGATGGGGGTCCTCACCTGTAT
5  GGAGGACGGGGGTCCTCACCTGTAA
6  GGAGGATGGGG-TCCTCATCTGTAT
7  GGAGGACAGGGGTCCTCATCTGTAT
8  GGAGGATGGGGGTCCTCATCTGTAA
9  GGAGGGCTGGCTGCC
   **-**---*----*-***--****-   (i)
   ******--*-*-******--****-   (ii)
```

Fig. 3.14: Alignment of the nine tandem "minisatellites" which begin at position 1,303 in L3. The last copy is truncated. Where necessary, gaps have been introduced to optimise the alignment. Stars in line (i) indicate positions where every "minisatellite" copy has the same nucleotide. Stars in line (ii) indicate positions where all but one of the copies have the same nucleotide.

elements has a completely different 24 bp. consensus sequence and is much less well conserved than the first. There are eight copies of the second type of element, but several are truncated and have non-repetitive sequences interspersed between them. Fig. 3.15b graphically illustrates these two sets of repetitive elements. Fig. 3.15c shows a dotplot of L3 versus the entire AKV genome.

Immediately downstream of the tandem repeats, at position 1715, homology to FeLV B *gag* begins. A little further downstream, there are two relatively short lengths of L3 related to different parts of AKV *gag*, with a large deletion between the two. L3 sequence which is 59.3% homologous to the MA-encoding component of AKV *gag* extends from positions 1839 to 1983. From positions 2068 to 2387 there is a section of L3 which is 62.7% homologous to CA-encoding AKV *gag*. BaEV CA-encoding *gag* homology begins a little earlier at position 2049. Retroviral sequence is once again interrupted by a retroposon between positions 2414 and 2455. This very short retroposon is bounded by 16 bp. perfect direct repeats and consists entirely of poly-A sequence. The downstream retroposon direct repeat overlaps the beginning of AKV *pol*-related sequences in L3, which extend from positions 2451 to 3088 (64.5% homology) and 3475 to 4227 (65.2% homology). Between these two *pol* segments, the 386 bp. gap contains yet another very short retroposon with 8 bp. perfect direct repeats. It is this particular L3 retroposon which is mentioned below, in the section which compares L3 with L4. The relatively long segments of *pol* homology in L3 are composed of sequences which encode tether, RNase H and integrase. The sequence colinear with AKV 3' *pol* also includes the first 30 bp. of *env*. There are substantial deletions in L3 *env*, to the extent that only the first 30 bp. and two short segments of TM-encoding sequence are represented (positions 4339 to 4472, 68% homology to AKV and positions 4491 to 4543, 62.3% homology to AKV).

The 3' LTR is preceded by a polypurine tract almost identical to that of AKV (15/16 bp. correct). The 3' LTR ends at position 5075, beyond which 1191 bp. of 3' flanking DNA was sequenced. As mentioned in the previous section, the first 43 bp. of flanking sequence was related to the equivalent 3' flanking sequence of the L4 element. Beyond that, there were short, interspersed regions of homology (up to 78 bp.) to a wide variety of sequences in the Microgenie™ database. Highest and most extensive homology was to human apolipoprotein E

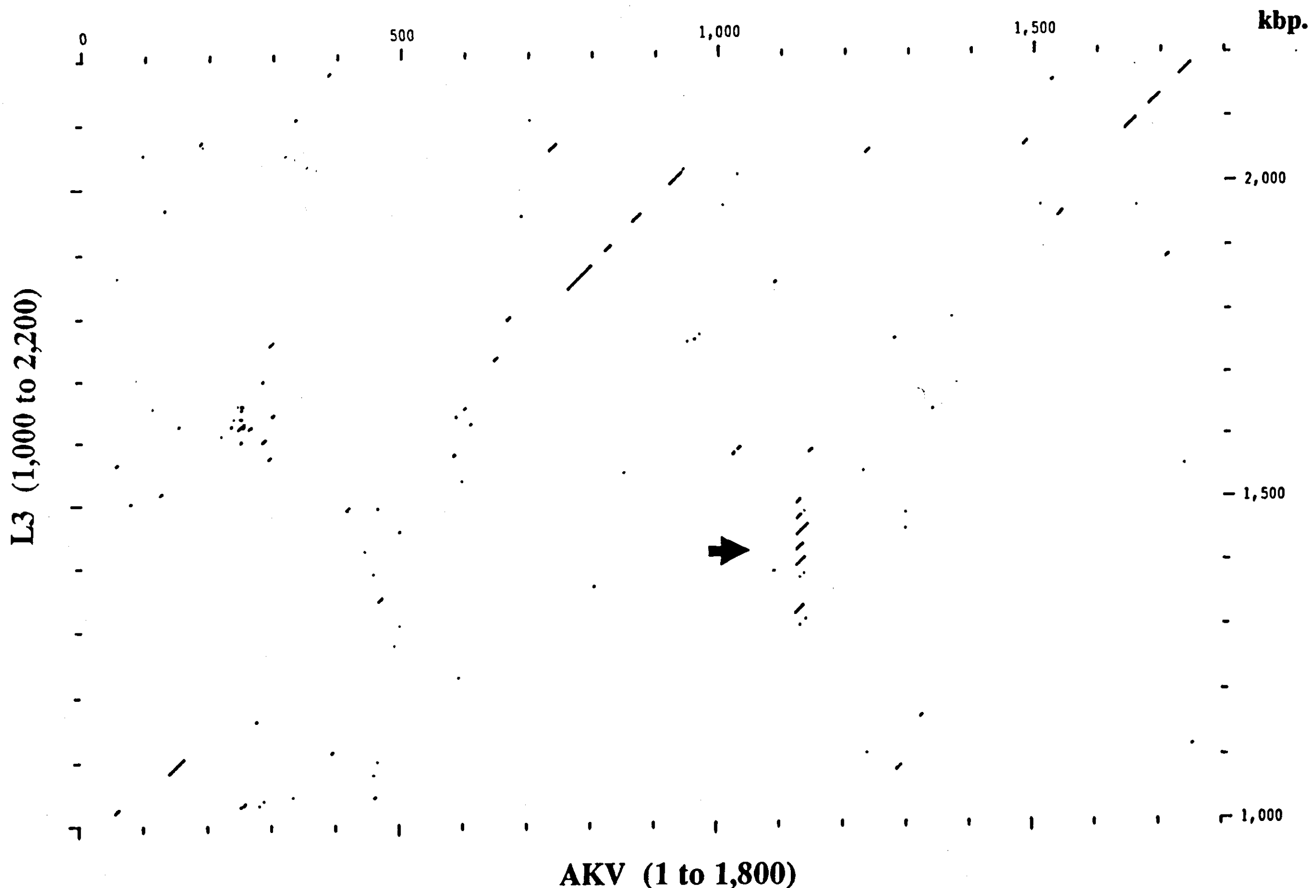


Fig. 3.15 (a): A dotplot of part of the 5' end of L3 (positions 1,000 to 2,200) versus the 5' end of AKV (positions 1 to 1,800). Regions of homology appear as broken diagonal lines. The vertical array of short diagonals (arrow) represents homology between AKV and the tandem array of L3 "minisatellites". Subsequent analysis using G.C.G. BESTFIT showed that there is a sequence in AKV *gag* (positions 1,127-1,155) 83% homologous to the L3 "minisatellite" sequences. Dotplot generated using G.C.G. COMPARE; Window: 21, Stringency: 14.

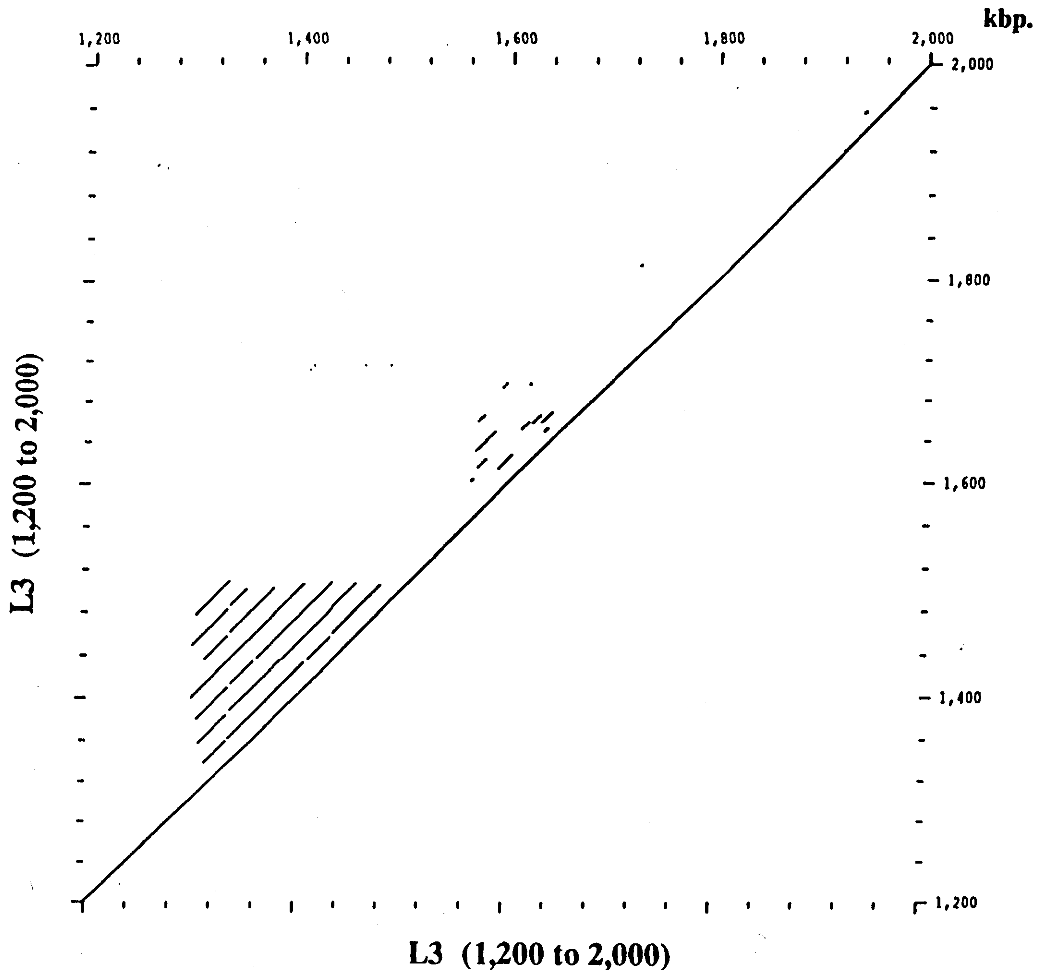


Fig. 3.15 (b): A dotplot of L3 (positions 1,200 to 2,000) compared with itself to show repetitive DNA. The two sets of tandem repeat elements are evident. Individual "minisatellites in the 5' group are highly homologous to one another and produce a strong pattern on the dotplot (*bottom left*). Homology between individual elements in the 3' group is lower; so a much less distinct pattern is produced on the dotplot (*centre*). Generated using G.C.G. COMPARE; Window size: 21, Stringency: 14.

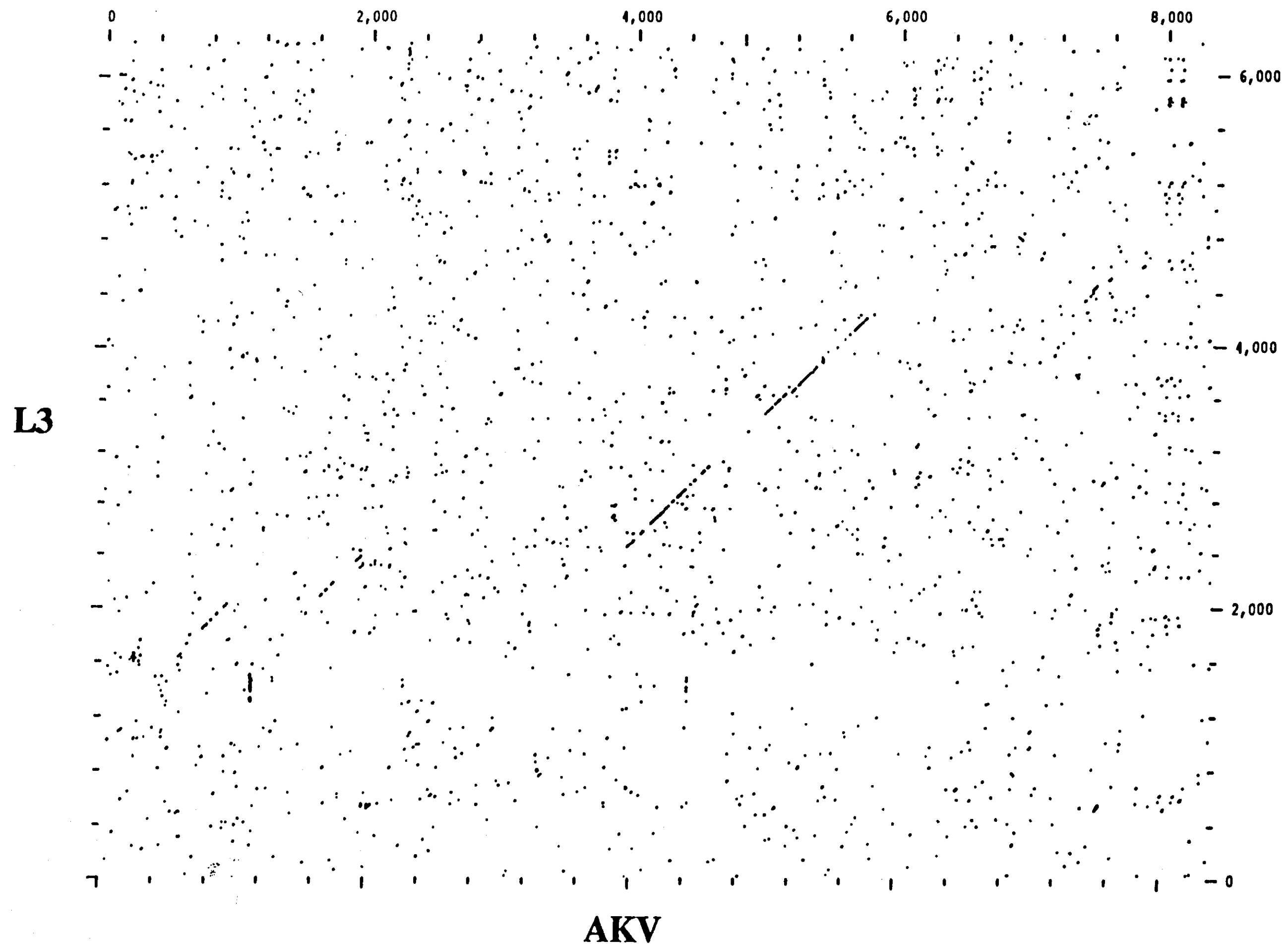


Fig. 3.15 (c): A dotplot of L3 versus AKV (Herr, 1984). The entire length of both sequences is compared. Regions of homology appear as diagonal lines. Deletions in L3 relative to AKV appear as horizontal steps between the diagonal lines. Generated using G.C.G. COMPARE; Window size: 21, Stringency: 14.

(epsilon-4 allele) gene (68.9% homology over 74 bp.), c-*fms* proto-oncogene (74.6% over 59 bp.), human c-Ha-*ras*1 proto-oncogene (61.5% homology over 78 bp.), African green monkey *Alu*-family repeats (69.2% homology over 52 bp.), *Xenopus laevis* 28S rRNA spacer repetitious region (71.4% homology over 70 bp.) and *Xenopus laevis* transcribed repetitious spacer regions in the other rRNA genes.

The region of homology to the apolipoprotein E gene overlapped that to the c-*fms* proto-oncogene. Similarly, the segment homologous to the human c-Ha-*ras*1 proto-oncogene overlapped the stretch homologous to the *Xenopus laevis* 28S rRNA spacer repetitious region.

As with L4, translation of the L3 provirus reveals frequent stop codons in all three plus strand reading frames. The longest ORF is 580 bp. in Frame 2 (positions 3768 to 4347). This ORF includes integrase-encoding sequence and spans the 5' *env* deletion. Figure 3.16 shows how homology to AKV is distributed among the three positive strand reading frames of L3.

3.6.3 COMPARISON OF L3 AND L4

Parts of L3 and L4 are closely related. This is evident from the dotplot in Fig. 3.17a and the diagram in Fig. 3.17b. L4 contains 5' and 3' *pol* sequences, short segments of *env* and a 3' LTR. L3 is more intact, containing in addition, a 5' LTR and 5' and 3' parts of *gag*. The only retroviral sequence present in L4, but absent in L3 is the 559 bp. 5' *pol* segment. Homology among the various colinear segments ranged from 81 to 92.5%, although homology between the 5' halves of the 3' LTRs fell to 38.5%. This was a consequence of a single 41 bp. imperfect direct repeat in L4 and multiple deletions, mostly in L4 compared with L3. The polypurine tracts were also quite different in the two proviral elements. Homology in the 3' halves of the LTRs was higher and extended for 43 bp. beyond their ends. Another intriguing observation was that L3 and L4 have identical deletions in *env*. Compared with AKV, both L3 and L4 lack precisely the same large 5' *env* segment and have an approximately 54 bp. deletion in the 3', TM-encoding part of the gene. A small retroposon which interrupts *pol* sequences in L3 is present at the equivalent position in L4. In L4, though, it has itself been interrupted by another retroposon (see Fig. 3.18). Perhaps a case of "....and little fleas have lesser fleas and so *ad infinitum*..." ?

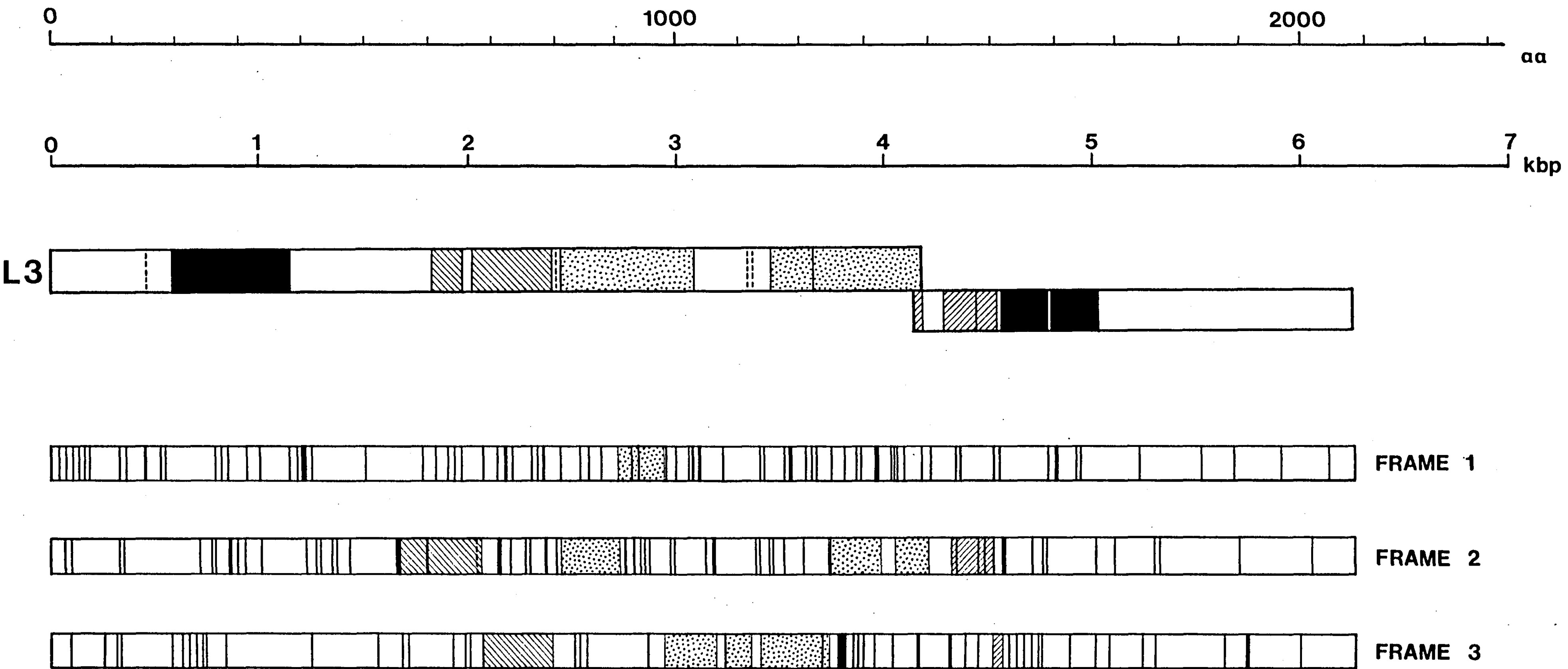


Fig. 3.16: Homology of the defective L3 proviral element to AKV at the protein level. The amino acid sequence of L3 was deduced from its nucleotide sequence for all three (+) strand reading frames. The resulting amino acid sequences were compared with that of AKV. Homology was found to be shared among the reading frames. Forward-hatched, stippled and back-hatched regions within the three reading frames represent regions of homology to AKV *gag*, *pol* and *env*, respectively. The positions of the many stop codons in each frame are marked by vertical lines.

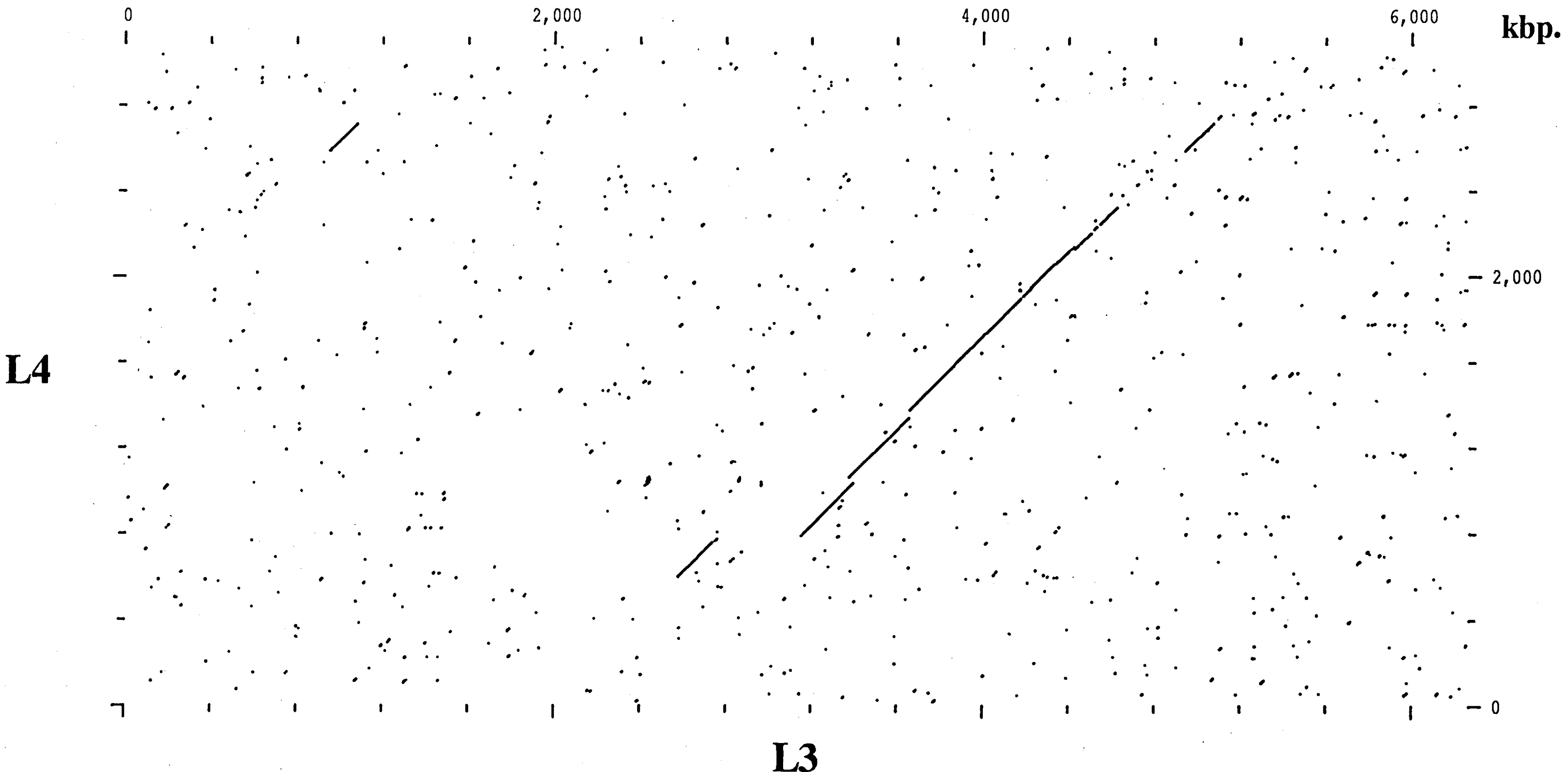


Fig. 3.17 (a): A dotplot of L3 versus L4. The entire length of both nucleotide sequences is compared. The dotplot shows the strong homology between the two sequences as a diagonal line. A large deletion in L4 relative to L3 is evident as a horizontal "step" in the diagonal (*bottom centre*). Smaller deletions or insertions in one sequence relative to the other produce the smaller irregularities in the diagonal line. The short diagonal line quite separate from the rest (*top left*) is a consequence of homology between the 5' L3 LTR and the 3' L4 LTR. Generated using G.C.G. COMPARE; Window size: 21, Stringency: 14.

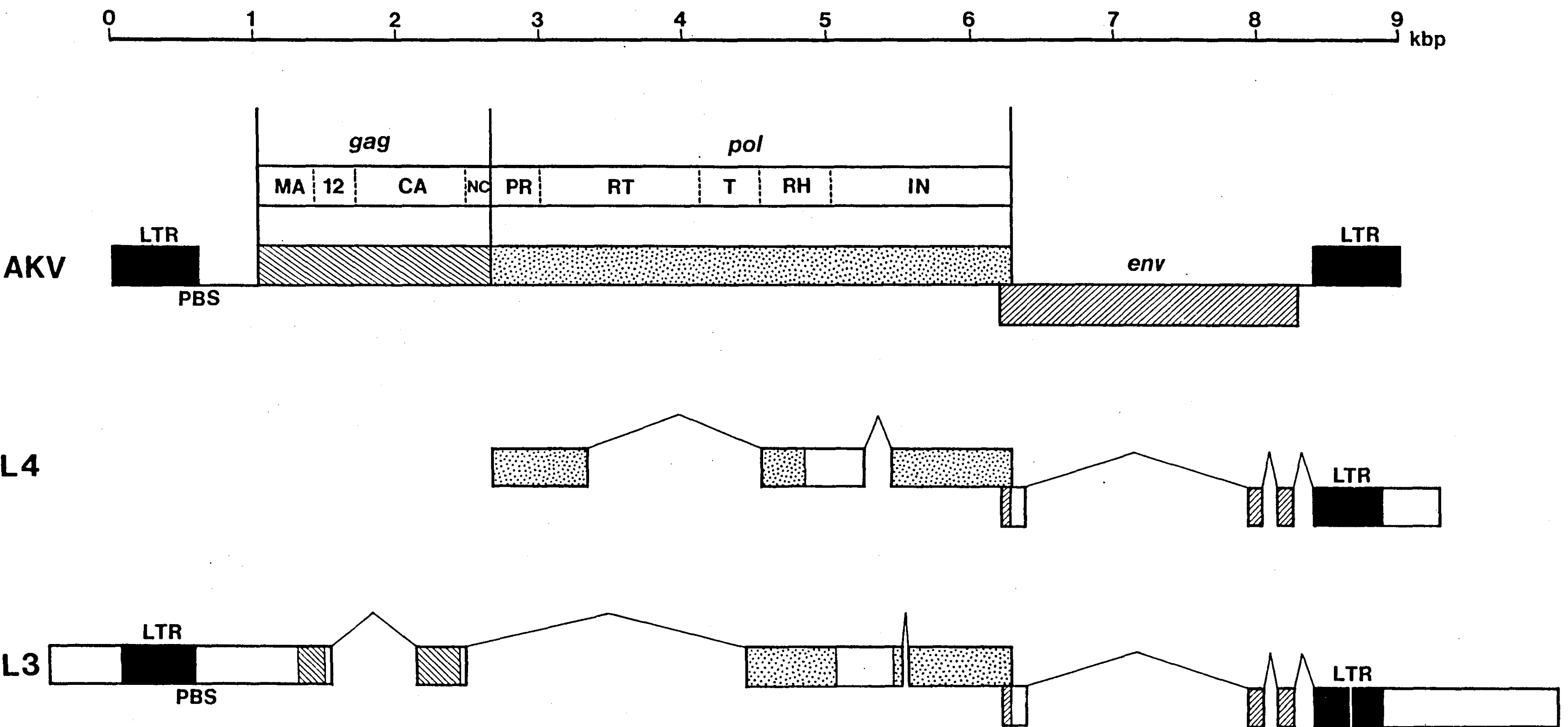


Fig. 3.17 (b): An alignment of AKV, L4 and L3 on the basis of nucleotide sequence homology. The diagram is intended to highlight the defectiveness of the two canine endogenous retroviral elements and the similarities between them. Sequences which did not have significant homology to AKV are shown as white boxes. The similarity of the *env* deletions in L3 and L4 is striking. The non-retroviral sequences in *pol* were also similar in the two proviral elements.

Abbreviations: LTR, long terminal repeat; PBS, primer binding site; MA, envelope-associated matrix protein, 12, *gag* protein of uncertain function; CA, major structural capsid protein; NC, nucleocapsid RNA-binding protein, PR, protease; RT, reverse transcriptase; T, tether; RH, ribonuclease H; IN, integrase; SU, major *env* surface glycoprotein; TM, transmembrane protein.

L3: AT**GCAAAGAT**CACTATTAAAGACTCTCGA**GCAAAGAT**TGAGCAGATCGTGGTGAGCTGCC

L4: AT**GCAAAGAT**CACTAT**TAAAGACTCTCGA**AAAAAAAATTTAAAAAAATAAATAAATAAATAAA**TAAAGACTCTCGAGCAAAGAT**CG

Fig. 3.18: Differences between L3 and L4 at a site of retroposon insertion. In L3, a small putative retroposon with 8 bp. perfect direct repeats (DRs) interrupts *pol*. The 8 bp. DRs are shown double underlined. In L4, a similar retroposon, present at the equivalent site, has been interrupted by a second retroposon with 13 bp. perfect DRs (single-underlined).

3.7 DISCUSSION

Canine 3132 lymphoma DNA was found to contain at least two closely-related retroviral proviral elements. Nucleotide sequence data analysis showed that these elements (termed L3 and L4) have several features in common with well-characterised infectious murine type-C retroviruses, but are heavily defective. The proviruses sequenced in this study could certainly not be responsible for the production of infectious virions, as was reported for the 3132 cell line by Strandstrom and Bowen in 1982. L4 was the first element to be sequenced. It consists of truncated *pol*, with interspersed cellular DNA; markedly truncated *env* and a 3' LTR. L3 is more intact, with both LTRs and segments of *gag*, *pol* and *env* represented. L3 and L4 have different deletions in *pol*; but interestingly, they were found to have three identical *env* deletions when their sequences were compared with AKV.

The extensive 5' *env* deletion present in both L3 and L4 is reminiscent of a similar defect in endogenous, *Gv-1*-responsive MuLV-related sequences found in murine genomic DNA (Wilson *et al.*, 1988). When the *env* deletion of these murine defective proviruses, which were found to have lost SU and part of TM, was compared with the available intact proviral sequence data; the deletion breakpoints were found to map to 7 bp. direct repeats in the intact sequence. Deletion was postulated to have occurred by homologous recombination between the repeated sequences. Since intact canine retroviral elements have not yet been sequenced, it is not possible to determine whether homologous recombination was responsible for the major, 5' *env* deletion found in L3 and L4. Two smaller, 3' *env* deletions were found in L3 and L4 when their nucleotide sequences were compared with that of AKV. The origin of these deletions is unclear.

Evidence indicates that L3 and L4 have been produced by a DNA amplification process rather than by conventional, separate integration of infectious retroviral proviruses. The presence of identical defects (in particular, the *env* deletions and retroposon insertions) could not be plausibly explained by independent mutation of separately-integrated intact retroviral genomes. Rather, it seems likely that defects developed in a common progenitor of these elements and that at some later time an amplification process occurred. In this scenario, the many differences between L3 and L4 must have arisen subsequent to

amplification. It is not clear which, if either, of the two elements is the more ancient (*i.e.*, which more closely resembles the ancestral form). The presence in L4 of two retroposons (one inside another) at a site where there is only one in L3 does not help in this respect. One might be tempted to argue that the common ancestor had a single retroposon at this site and that L4 has subsequently acquired a second. However, it is also possible that the ancestor had the double retroposon complement and that deletion of one retroposon has occurred in L4 by homologous recombination between the flanking direct repeat sequences.

The nature of the amplification process which has occurred cannot be deduced from the available sequence data. Had the 5' terminus of the L4 element been cloned and sequenced, a better understanding might have been achieved. The fact that 43 bp. of L4 3' flanking sequence is 60.4% homologous to its L3 counterpart suggests that the amplified unit might include the defective provirus and some flanking sequence. Alternatively, this degree of flanking sequence homology might reflect a preference by this type of retroviral element for certain target sequences during retrotransposition. However, previous work which compared numerous sequenced integration sites from several host-retrovirus combinations did not reveal any host sequence motif which was predisposed to retroviral integration (Shimotohno and Temin, 1980). Were it not for the similarities between L3 and L4 in the 3' cellular flanking sequence; retrotransposition, perhaps using RT supplied *in trans*, would be a likely mechanism to explain amplification of L3 and L4. However, since the amplified unit seems to be larger than the provirus itself, alternative mechanisms should be considered. Previous work sheds some light on this subject. Steele *et al.* (1984) found multiple copies of a family of human endogenous retroviral elements (HERV-E) which shared similar cellular flanking DNA. Interestingly, in view of the findings described in this chapter, members of one truncated group of elements had lost their LTRs and were bounded instead by tandem repetitive DNA (minisatellites). Other HERV-E elements were typical full-length proviruses. The flanking cellular sequences of some of these elements were part of the *Kpn*I family of LINE-type repeated human DNA. An intriguing possibility is that amplification of this particular group of human defective proviruses may have occurred by use of the amplification machinery of the adjacent *Kpn*I repeats. Whether LINE or SINE

elements are, indeed, involved in the amplification of mammalian endogenous retroviral sequences is yet to be determined.

In this regard, it is notable that sequences related to repetitive DNA of various vertebrate species were present 3' to the L3 proviral element. In particular, a 52 bp. length of DNA 69.2% homologous to African green monkey *Alu*-family repetitive DNA was present. *Alu*-like sequences are SINEs, which retrotranspose via an RNA polymerase III transcript. SINEs have been shown to fuse to new sequences, which can become part of the functional, retroposing unit (Rogers, 1985; Weiner *et al.*, 1986; Zelnick *et al.*, 1987). However, the fused, new sequence must lack RNA polymerase III termination signal sequences (consensus TTTT) if it is to be successfully transcribed by this mechanism. There are numerous terminators for this enzyme, on both DNA strands, in L3 and L4. Thus, amplification via an RNA polymerase III transcript is highly unlikely to have occurred. Despite the fact that RNA polymerase III terminators occur frequently throughout genomic DNA; *Alu*-like sequences have been associated with DNA amplification in other systems. For example, after amplification at the adenylate deaminase locus, nine novel junctions between amplified segments of DNA were mapped to a 2.6 kbp. A+T-rich DNA fragment. Within this fragment was a chimaera of four *Alu*-like repeat elements (Hyrien *et al.*, 1987). *Alu*-like sequences have been found at the breakpoints of several illegitimate DNA recombinations (for a review, see Meuth, 1989). A mechanism for translocation of DNA sequences located between two *Alu*-like elements has been postulated: homologous recombination between the *Alu*-like sequences might cause DNA excision to produce a free circular DNA molecule. This circular DNA might subsequently re-integrate elsewhere in the genome. However, this form of translocation would not necessarily be associated with DNA amplification. The precise nature of the association between *Alu*-like elements and DNA amplification is yet to be elucidated and the presence of such a sequence 3' to the L3 element may well be co-incidental. However, if future studies consistently show LINE- or SINE-related sequences adjacent to amplified retroviral proviral elements, a possible functional role for these sequences in retroviral DNA amplification should be investigated.

The presence of a region of tandem repeats downstream of the PBS in L3 was unexpected and interesting. The number of tandem copies, or "minisatellites", in such tandem arrays is sometimes very variable from individual to individual within a species. It is for this reason that hypervariable minisatellite regions are the basis of "DNA fingerprinting" in humans (Jeffreys *et al.*, 1985). Interestingly, probes derived from human hypervariable regions have been used to resolve canine paternity disputes (Morton *et al.*, 1987). However, when human probes are used to study canine DNA, the results are not quite as impressive as they are with human target DNA: the sequences detected in canine DNA are not as hypervariable as they are in humans. In fact, broadly similar patterns of hybridisation, especially of smaller DNA fragments, were found when human minisatellite probes were used to study individual canine DNAs. These probes were recently found to be useful in a quite different application relevant to this study. Minisatellite probes developed by Alec Jeffreys were used to probe samples of genomic DNA from several members of six isolated populations of the California Channel Island fox, *Urocyon littoralis* (Gilbert *et al.*, 1990). In this study, DNAs from members of an isolated population of foxes each had similar patterns of hybridisation to the minisatellite probes. Samples from different islands had markedly different hybridisation patterns. It was possible to deduce from the hybridisation data the phylogenetic relationships among these natural fox populations.

Although there was no evidence to suggest that the array of repeat elements in L3 would be part of a large, hypervariable family; a pilot study was conducted to determine whether a probe encompassing the two sets of repeats would detect differences between individual dogs when hybridised to canine genomic DNA digested with *Hinf*I and *Hae*III. It was also of interest to learn whether the canine probe would produce broadly similar results to those reported for the human minisatellite probes. Preliminary results indicated that the canine probe produces a complex pattern of multiple hybridising bands and can detect differences between dogs, although there were more points of similarity than difference (data not shown). Work is in progress to develop an improved probe consisting solely of the 5', highly-conserved group of repeats.

Another interesting finding was the presence of several, short, putative retroposons interspersed in L3 and L4. Retroposons have been defined as dispersed, repeated DNA elements formed by the reverse flow of genetic information from RNA to DNA (Rogers, 1983). Thus retroposons, like retroviruses, are defined by their replication strategy. Formally, it is not possible to deduce from the available sequence data the precise mechanism by which the DNAs termed "retroposon" in this chapter entered the genome. The use of this term might therefore be questioned. However, these sequences are bounded by 7-21 bp. direct repeats and each includes an A- or T-rich DNA sequence. Since these are characteristic features of mRNA-derived retroposons (Weiner *et al.*, 1986), the use of this term seems justified.

Given that 10% of the mammalian genome is thought to derive from RNA, it is perhaps not surprising that several small retroposons were present in the approximately 10 kbp. of non-coding DNA represented by L3 and L4. In human DNA, there is, on average, one 300 bp. *Alu* element every 5 kbp.; and these elements are distributed fairly randomly throughout the genome (Deininger, 1989). A short retroposon was present just 5' to the 3' LTR in the S71 human endogenous retroviral element sequenced by Werner *et al.* (1990), although this is not commented upon by the authors. Retroposons are attracted to integrate in A-rich regions (Furano *et al.*, 1986). This might explain the interruption of one mRNA-derived retroposon by another in L4. The presence of so many retroposons and other defects in these proviral elements prompts the suggestion that these are very ancient sequences or, possibly, that they are sites for accelerated mutation. Work described in the next chapter tends to support the former suggestion. The significance of the findings described in this chapter will be considered further in Chapter 7.

Chapter 4

DISTRIBUTION OF ENDOGENOUS RETROVIRAL ELEMENTS IN CANINE GENOMIC DNA

4.1 INTRODUCTION

4.2 TARGET DNAs

4.2.1 NORMAL TISSUES

4.2.2 NEOPLASTIC TISSUES

4.2.3 CULTURED CELLS

4.3 PROBE SELECTION & PREPARATION

4.3.1 PROBES DERIVED FROM L4

4.3.2 OTHER RETROVIRAL PROBES

4.4 HYBRIDISATION CONDITIONS & AUTORADIOGRAPHIC FINDINGS

4.4.1 PROBES DERIVED FROM L4

4.4.2 OTHER RETROVIRAL PROBES

4.5 DISCUSSION

Appendix 1: PRESENCE OF CERV-RELATED SEQUENCES IN WILD CANIDS.

Appendix 2: ABSENCE OF CERV-RELATED SEQUENCES IN HUMAN GENOMIC DNA.

Appendix 3: A72-E CELLS EXPRESS FeLV-B AND APPEAR NOT TO BE OF CANINE ORIGIN.

Appendix 4: *c-myc* IS REARRANGED IN 3132 DNA.

4.1 INTRODUCTION

This chapter describes how canine genomic DNA from a variety of normal and neoplastic tissues was probed for the presence of retroviral sequences. Probes derived from the newly-sequenced L4 proviral element, and well-established murine and primate retroviral probes were used. Numerous retrovirus-like elements were found in all samples of canine DNA examined. No differences in the arrangement of proviral elements were detected between any of the normal or neoplastic tissues.

The work described in Chapter 3 showed that a particular canine lymphoma cell line (3132) contains at least two defective retroviral proviruses. Having isolated and determined the nucleotide sequence of *pol* segments in one of these elements (L4), it was considered of interest to determine the distribution of such elements in normal and abnormal canine genomic DNA. It was important to discover whether the presence of retroviral sequences in 3132 DNA was a feature unique to that cell line, or a more general feature of canine DNA. The number of such elements present in the entire 3132 genome was also of interest. Since 3132 is a lymphoma cell line and canine lymphoma is a disease which has long been suspected to have a retroviral aetiology, it was decided that normal canine genomic DNA should initially be compared with DNA obtained from lymphomatous tissues. The objectives were to determine whether or not there were differences in the distribution of retroviral proviruses between the two groups; and if so, to determine if the differences were of aetiological significance. Later, a few DNA samples from dogs with diseases other than lymphoma were examined.

The techniques described in this chapter have been used successfully to characterise and determine the approximate copy number of endogenous retroviral elements in other species. Early work with chicken DNA demonstrated that individual birds varied considerably in their endogenous retroviral complement (Astrin, 1978). Indeed, such variability was, for a while, considered one of the important hallmarks of endogenous retroviruses (Coffin, 1982b). Individual wild mice were also found to vary in the number and types of endogenous retroviruses they carried (Cohen and Varmus, 1979). More recently, ancient and often defective retroviral elements have been identified in human and murine DNA (human: Steele *et al.*, 1984; Mariani-Constantini *et al.*, 1989; murine: Itin and

Keshet, 1986; Obata and Khan, 1988). These ancient elements seem to be distributed identically in the genomes of all members of their host species. Closely-related species have a very similar pattern of distribution of the elements. Slightly more distantly-related species have a less similar pattern. There is evidence that several of the elements in mice and humans have undergone an amplification process involving the provirus and some of its flanking cellular sequences (see the discussion in Chapter 3). Genomic DNA containing such elements was digested with restriction endonucleases which cut in the flanking sequences but not in the provirus itself. Agarose gel electrophoresis, denaturation, Southern transfer, hybridisation to a retroviral probe and autoradiography were carried out. Recruited (very bright) autoradiographic bands were found with several of the flanking sequence-cutting REs (Steele *et al.*, 1984; Steele *et al.*, 1986). These findings can best be explained by invoking an amplification process, in which the amplified unit is larger than the provirus itself. Since all members of a given species appear to have a virtually identical distribution of these ancient elements, there is no evidence that the amplification process is continuing apace. Nevertheless, it is an intriguing possibility that, during the process of amplification, such elements could act as insertional mutagens.

The basic materials and methods used to carry out the work in this chapter have been described in Chapter 2. The probes and target DNAs are described further below.

4.2 TARGET DNAs

Table 4.1 gives summary data of genomic DNA samples obtained from clinical cases and post mortem material.

4.2.1 NORMAL TISSUES

High molecular weight genomic DNA was extracted from the thymus and kidney of a healthy 6 week old puppy, from the kidney of a healthy retired racing greyhound and from the kidney of a healthy crossbred dog. In addition, 8 samples of DNA were extracted from normal tissues of untreated multicentric lymphoma and ALL patients: six from kidney and two from brain. The tissues were subsequently examined histologically and determined to be free of neoplastic cells. Kidney was the first choice of tissue for such extractions. If it appeared that the

Table 4.1: Details of the patients and tissues from which high molecular weight genomic DNAs were prepared.

Case No.	Signalment	Tissues Taken	Diagnosis
070137	12 yr M Crossbreed	K	renal LSA
107027	6 wk F Sheltie	K, Thymus	normal
107190	7 yr F Bull Terrier	LN	mLSA
107227	Adult M Crossbreed	LN	mLSA
107228	Adult FS Crossbreed	K	normal
107235	Adult M G. retriever	K, LN	mLSA
107311	2 yr F G. retriever	K, LN	mLSA
107318	4 yr M GSD	K, Lb	ALL
107490	7 yr M Bernese	K, Liver, Heart	histiocytosis
108386	9 yr FN Rough collie	LN	mLSA
108548	5 yr F Cross breed	LN	mLSA
108557	9 yr F GSD	LN	mLSA
108573	2 yr F Bernese	Skin	histiocytosis
108601	Adult F Lab	K, LN	mLSA
108666	6 yr FN G. retriever	K, Br, BM	ALL
109341	5 yr F Great Dane	LN	mLSA
109342	9 yr M Irish Setter	LN	mLSA
109393	13 yr M Lab	LN	mLSA
109456	Adult M Lab	LN, Br	mLSA
109533	3 yr M WHWT	LN	mLSA
109801	8 yr F Crossbreed	Thyroid	Thyroid carcinoma
NK 1	7yr M Greyhound	K	normal
PB	3 yr MN Lab	K, LN, Lb	ALL
ZB	8yr FN Afghan	LN	*mycosis fungoides*

Abbreviations: M, male; F, female; N, neutered; wk, week; yr, year; Lab, labrador retriever; G. retriever, golden retriever; GSD, German shepherd dog; Bernese, Bernese mountain dog; WHWT, West Highland white terrier; LN, lymph node; K, kidney; Lb, peripheral mononuclear cells from dogs with ALL; Br, brain; BM, bone marrow; mLSA, multicentric lymphosarcoma; ALL, acute lymphoblastic leukaemia.

kidneys might be infiltrated by neoplastic cells, then brain, free of meningeal *dura mater*, was used instead.

4.2.2 NEOPLASTIC TISSUES

13 samples of DNA were extracted from affected lymph nodes of dogs with histologically-confirmed multicentric lymphoma. DNA was extracted from peripheral blood lymphoblasts of two dogs with acute lymphoblastic leukaemia and from the bone marrow of a third. Tissues from two cases of histiocytosis of Bernese mountain dogs were examined: DNA was extracted from the skin of one dog and the heart, liver and kidney of another. In addition, single samples of neoplastic tissue were obtained from cases of renal lymphoma, *mycosis fungoides* and thyroid carcinoma.

4.2.3 CULTURED CELLS

DNA was extracted from 3132, A72-F, A72-E and MDCK cells. Extractions were carried out when the monolayer cells were just sub-confluent. DNA was extracted from 3132 cells when they were at a density of $5x10^6$/ml. A72-E was shown not to be a canine cell line (see Appendix 2).

4.3 PROBE SELECTION & PREPARATION

4.3.1 PROBES DERIVED FROM L4

After the nucleotide sequence of the *pol* segments of the L4 proviral element were determined, Microgenie™ was used to produce a map showing the restriction sites of most of the commercially available REs. This map allowed identification of restriction sites useful for the production of probes consisting exclusively of colinear canine retroviral sequence. The sequence data indicated that two probes could readily be derived. A 426 bp. *Hinf*I-*Pst*I fragment spanning positions 40 to 465 in L4 encompassed PR- and 5' RT-encoding *pol* sequences. A 559 bp. *Spe*I-*Eco*RV fragment, from positions 1238 to 1796, consisted of IN-encoding *pol* (Fig. 4.1). When L4 subclones were digested with these pairs of enzymes, fragments of the sizes predicted by the sequence data were produced. The 559 bp. *Spe*I-*Eco*RV fragment was purified from LMP agarose and subcloned into pBluescript KS (+) for large scale preparation. The 426 bp. *Hinf*I-*Pst*I fragment was double purified by electrophoresis through LMP agarose and used directly in labelling reactions. The specificity of the probes was ascertained by

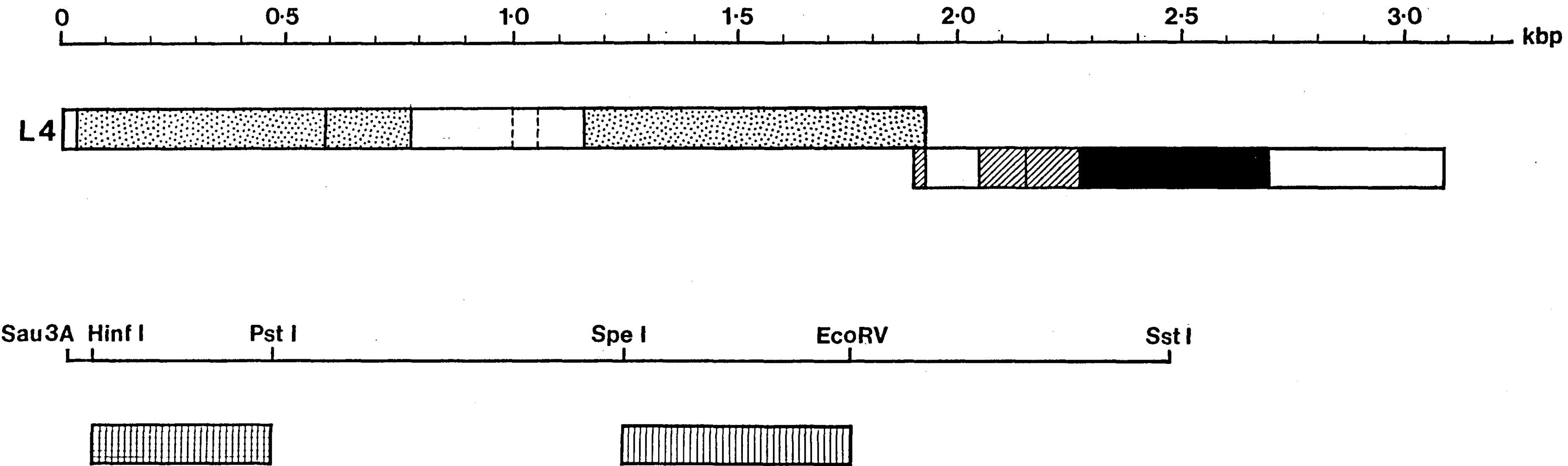


Fig. 4.1: Derivation of two retroviral *pol* probes from L4. At the top of the figure is a diagrammatic representation of L4. Below is a restriction map, showing relevant restriction sites in the segment of L4 which had been characterised at the time of probe development. The bottom line shows the relative positions of two DNA sub-fragments which were used to generate canine retroviral *pol* probes.

demonstration that each hybridised only to fragments containing its own sequence in control lanes which contained a variety of L4 subfragments.

4.3.2 OTHER RETROVIRAL PROBES

Canine genomic DNA samples were probed with four established retroviral probes: BaEV, FMuLV, HTLV1 and MMTV. The FMuLV and BaEV probes were permuted proviruses. The HTLV1 probe was a non-permuted provirus produced by use of an LTR-cleaving RE. The MMTV probe consisted of a *gag-pol* domain. The origins and nature of these are described further in Chapter 2 (2.1.9).

4.4 HYBRIDISATION CONDITIONS & AUTORADIOGRAPHIC FINDINGS

Although the majority of Southern transfer membranes were probed only once, several were probed at both high and low stringency. Membranes were first probed at high stringency, followed by low stringency. The procedures were carried out in this order, rather than the opposite way around; because it was found that cleaner autoradiographs resulted. This was particularly true of the high stringency hybridisations. HIGH STRINGENCY consisted of pre-hybridisation and hybridisation at 45°C in the presence of 50% formamide. Washing was for 2 x 90 minutes in 0.5xSSC, 0.1% SDS at 65°C; with a 10 minute, room temperature, 2xSSC rinse at the beginning and end. LOW STRINGENCY consisted of pre-hybridisation and hybridisation at 37°C in the presence of 25% formamide. Washing was for 2 x 90 minutes in 1xSSC, 0.1% SDS at 55°C; with a 10 minute, room temperature, 2xSSC rinse at the beginning and end.

4.4.1 PROBES DERIVED FROM L4

Once it was established that the L4 *Sst*I 2.5 kbp. fragment hybridised strongly to FMuLV, but before any sequence data was available, a few preliminary hybridisations were carried out using the whole 2.5 kbp. fragment as a probe. The only target DNAs available at that stage were from normal puppy kidney and thymus. These two samples were digested with *Bam*HI, *Sst*I and *Xba*I before transfer and hybridisation. These enzymes were chosen empirically, because they were known to cleave L3 and L4 close to fragments which hybridised strongly to FMuLV. The resulting autoradiograph showed a complex pattern of numerous bands with all three REs (Fig. 4.2). There were no appreciable differences

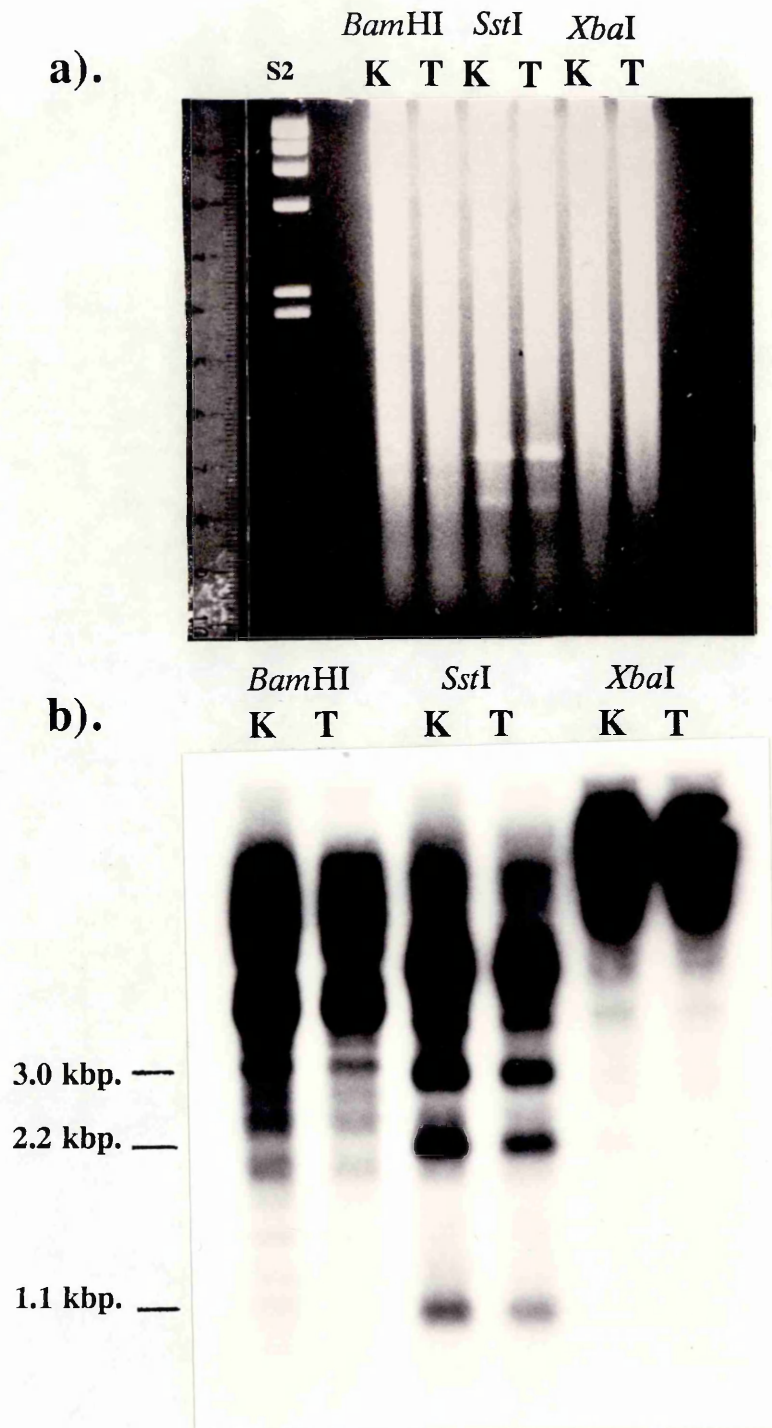
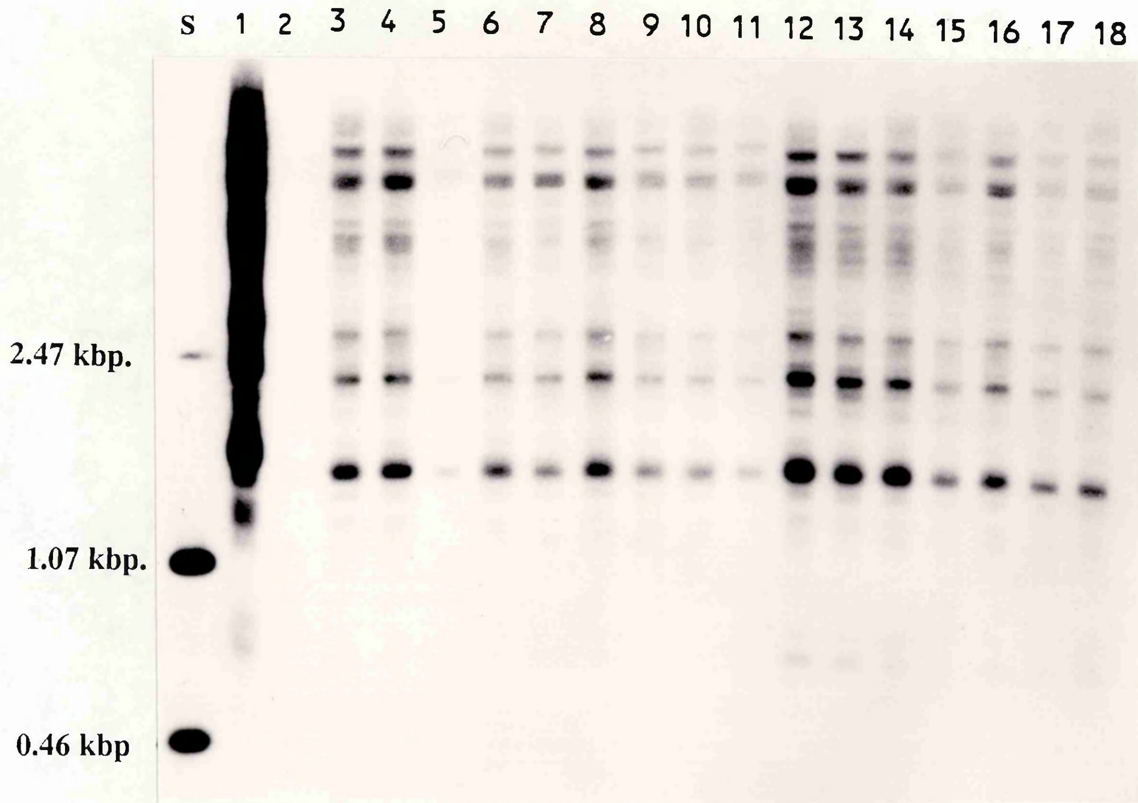


Fig. 4.2: **a).** A photograph of an ethidium bromide-stained gel, showing RE digestions of high molecular weight genomic DNAs prepared from kidney (K) and thymus (T) of a healthy 6 week old puppy (Case No. 107027). Discrete bands representing repetitive DNA elements are evident in the *Sst*I digestions. S2, DNA size marker: bacteriophage lambda DNA digested with *Hin*dIII.
b). An autoradiograph derived from the gel shown in (a). The Southern transfer membrane was probed at high stringency with a probe prepared using the L4 *Sst*I 2.5 kbp. fragment which hybridised strongly to an FMuLV probe. The bands hybridising to this probe are too numerous to count accurately. There are three prominent, recruited bands in the *Sst*I digestions (3.0, 2.2 and 1.1 kbp.). All of these bands are larger than the repetitive DNA bands seen in (a). No differences between K and T are evident.

between the two DNA samples. The total number of bands could not easily be counted, because there were so many that they tended to merge with one another. The number was certainly greater than 30 in the *Bam*HI digestions. Several bands were intensely bright, or recruited. It was considered possible at the time that the recruited bands represented internal proviral fragments; *i.e.*, that *Bam*HI, *Sst*I and *Xba*I had internally cleaved several separately-integrated proviruses in at least two sites. However, the large size of some of the recruited bands tended to argue against this. A second, identical membrane was available because samples in the gel electrophoresis set had been duplicated. This membrane was probed at high stringency with FMuLV. Unsurprisingly, probes derived from FMuLV and the L4 *Sst*I 2.5 kbp. fragment produced somewhat similar patterns of autoradiographic bands. This was particularly evident in the *Sst*I digestions, although there were far fewer bands with the FMuLV probe. The finding of multiple elements in normal canine genomic DNA which hybridised to a putative canine retroviral provirus indicated that these were endogenous retroviral sequences.

After partial L4 sequence data became available, the *Hinf*I-*Pst*I (5' *pol*) and *Spe*I-*Eco*RV (3' *pol*) probes were developed. At this stage 18 canine genomic DNA samples were available for analysis. These DNAs were digested with *Pst*I and probed with the *Hinf*I-*Pst*I (5' *pol*) probe (Fig. 4.3). In a separate experiment, the DNAs were digested with *Eco*RV and probed with the *Spe*I-*Eco*RV (3' *pol*) probe (Fig. 4.4). *Pst*I and *Eco*RV were chosen for these digestions so that in each experiment the probes would hybridise to just one fragment per proviral element. In theory, each RE would cleave proviruses immediately to one side of the sequence detected by its respective probe. This assumed that the arrangement of *Eco*RV and *Pst*I sites within all of the proviral elements in the canine genome was similar. In particular, it was important that there be no additional *Pst*I sites in the part of the viral genome spanned by the *Hinf*I-*Pst*I (5' *pol*) probe and no *Eco*RV sites in the span of the *Spe*I-*Eco*RV (3' *pol*) probe, otherwise two or more hybridising fragments would be produced per provirus. Both the 5' and 3' *pol* probes produced complex autoradiographic findings. In both hybridisations there were numerous bands, many of which were recruited. The 5' *pol* probe detected at least 26 bands in *Pst*I-digested DNA samples, 8 of which were obviously recruited (Fig. 4.3). The 3' *pol* probe detected at least 28 bands in the *Eco*RV-digested



Lane	Case No.	Tissue
1	107027	thymus
2	107027	K
3	107311	LN
4	107311	K
5	107318	Lb
6	107318	K
7	108601	LN
8	108601	K
9	108548	LN
10	108557	LN
11	PB	LN
12	PB	K
13	PB	Lb
14	108386	LN
15	107235	LN
16	107235	K
17	107490	Liver
18	107490	Heart

Fig. 4.3:

An autoradiograph of 18 canine genomic DNA samples, digested with *Pst*I and probed with the L4 *Hinf*I-*Pst*I (5' *pol*) probe. At least 26 hybridising bands are present in each lane, 8 of which are obviously recruited. Lane 1 has been grossly overloaded and Lanes 2 and 5 have been underloaded with DNA. In the lane marked S, L4 fragments known to hybridise to the probe were used as size markers: 2.47 kbp. *Sst*I fragment, 1.07 kbp. *Sst*I-*Xho*I fragment and 0.46 kbp. *Sst*I-*Pst*I fragment.

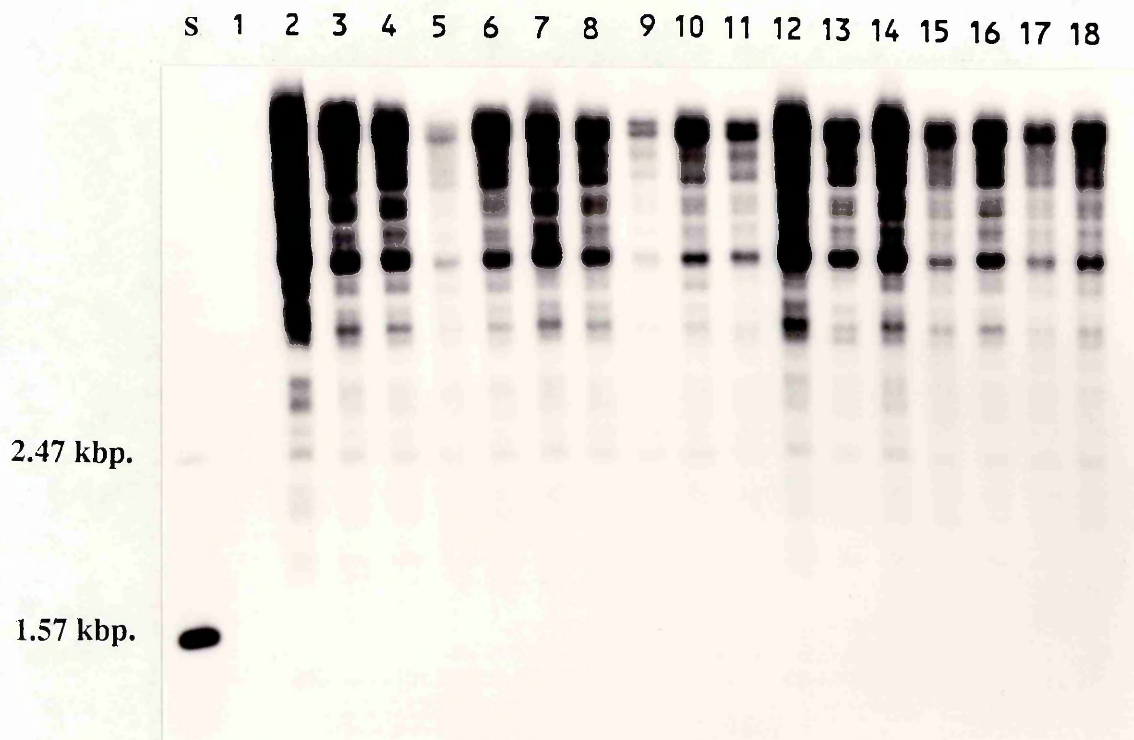
Abbreviations:

K, kidney; LN, lymph node; Lb, peripheral lymphoblastic mononuclear cells

DNA samples, with a preponderance of high molecular weight fragments (Fig. 4.4). Importantly, all DNA samples, normal and neoplastic appeared to have an identical retroviral complement. Since the probes used in these hybridisations consisted exclusively of retroviral sequences, it was concluded that the numerous autoradiographic bands represented multiple, related endogenous proviral elements.

The 3' *pol* probe was later used to probe a further 16 canine DNA samples which had been digested separately with *Eco*RI and *Xho*I. In addition, each membrane included two samples which had been used in the previous hybridisations. This allowed comparison of the new batch of samples with the old. Although each RE produced a distinct pattern of autoradiographic bands, the fundamental findings were the same in all experiments: multiple autoradiographic bands were detected, several of which were recruited. There were no detectable polymorphisms among the samples with any RE. The *Xho*I digestions were interesting, in that most of the hybridisation was to high molecular weight DNA fragments. The effect was much more pronounced than that noticed with *Eco*RV. Although the DNA smears in the ethidium bromide-stained gel of *Xho*I digestions appeared normal, the experiment was repeated in case the digestions had been incomplete. An identical result was obtained the second time. This finding is in agreement with those of Battula *et al.* (1982); who found that most BaEV M7 sequences were resistant to *Xho*I digestion. By digesting baboon DNA with *Hpa*II and *Msp*I, which distinguish methylated from non-methylated DNA sequences (see Chapter 1 section 1.2.5); these workers showed that baboon endogenous retroviral sequences are extensively methylated and are therefore cleaved poorly by the methylation-sensitive enzyme *Xho*I.

A small number of DNA samples were double-digested with *Hinf*I and *Pst*I and probed with the *Hinf*I-*Pst*I probe. A similar experiment was carried out using *Spe*I and *Eco*RV double-digestions with the *Spe*I-*Eco*RV probe. The objective was to determine the proportion of proviral elements in which these particular restriction sites were conserved. This would be one indication of how closely related were the many proviral elements detected by the two probes. Taken to an extreme, if the four restriction sites used to produce the two probes had been arranged exactly as for L4 in all of the genomic proviral elements, then a single hybridising band the same size as the probe would have been detected in each of the double-digestions. As it turned out, double-digestion with the probe-



Lane	Case No.	Tissue
1	107027	thymus
2	107027	K
3	107311	LN
4	107311	K
5	107318	Lb
6	107318	K
7	108601	LN
8	108601	K
9	108548	LN
10	108557	LN
11	PB	LN
12	PB	K
13	PB	Lb
14	108386	LN
15	107235	LN
16	107235	K
17	107490	Liver
18	107490	Heart

Fig. 4.4:

An autoradiograph of 18 canine genomic DNA samples, digested with *Eco*RV and probed with the L4 *Spe*I-*Eco*RV (3' *pol*) probe. At least 28 hybridising bands are present in each lane, with a preponderance of high molecular weight fragments. Lanes 1, 5 and 9 have been underloaded and Lane 2 has been overloaded with DNA. In the Lane marked S, L4 fragments known to hybridise to the probe were used as size markers: 2.47 kbp. *Sst*I fragment and 1.57 kbp. *Sst*I-*Xho*I fragment.

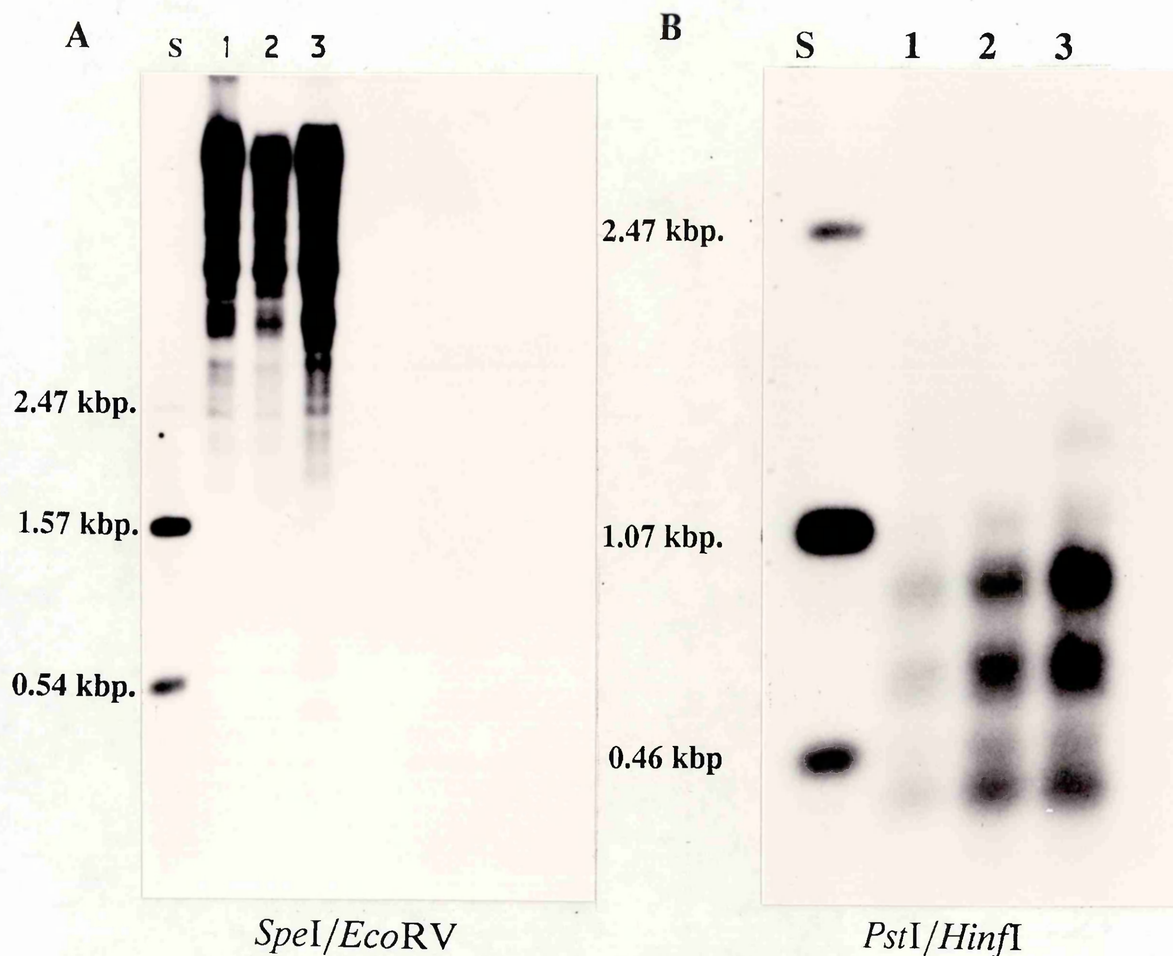
Abbreviations:

K, kidney; LN, lymph node; Lb, peripheral lymphoblastic mononuclear cells

generating enzymes simplified the pattern of autoradiographic bands considerably for one probe, but not the other. Double digestion with *Hinf*I and *Pst*I produced a dramatically simplified pattern, but not to the extent of producing a single band (Fig. 4.5). There were just 4 bands, the smallest of which was the same size as the probe fragment. This indicated that the proviral elements detected by the *Hinf*I-*Pst*I probe are members of a closely-related, but non-identical family. Double digestion with *Spe*I and *Eco*RV resulted in a complex pattern of high molecular weight bands, indicating that these restriction sites are not present in many of the endogenous proviral elements (Fig. 4.5). The fact that these particular sites are not well-conserved may reflect a greater variability among the proviruses in the IN-encoding region as compared with the PR-encoding segment of *pol*. Alternatively, it may merely be that L4 differs from the majority of proviral elements in having one or both of these sites.

4.4.2 OTHER RETROVIRAL PROBES

36 samples of canine genomic DNA, from normal, neoplastic and cultured cells, were digested with *Eco*RI and separately probed at high and low stringency with BaEV, FMuLV, HTLV1 and MMTV. Three DNA samples (normal greyhound kidney, 3132 and case no. 107311 lymphomatous lymph node) were digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Sst*I and *Xba*I and probed in the same way. The objectives were to screen a relatively large number of DNA samples for retrovirus-like sequences and to check a smaller number of representative samples for RE polymorphisms. Compared with the striking autoradiographs produced by use of the canine retroviral probes, the results were relatively lacklustre. There was little or no hybridisation by any of the samples to the MMTV *gag-pol* probe (data not shown, because it was a blank autoradiograph). Despite repeated hybridisation experiments using fresh membranes, the FMuLV probe produced very high background hybridisation during this particular series of experiments and this resulted in a technical failure. Autoradiographic findings of a previous, successful hybridisation experiment, in which different DNAs and digestions were used, are presented for interest in Fig. 4.6. There was an interesting pattern of low intensity hybridisation to the HTLV-1 probe with a prominent 7.5 kbp. recruited band in the *Eco*RI digestions (Fig. 4.7). No DNA polymorphisms were found by



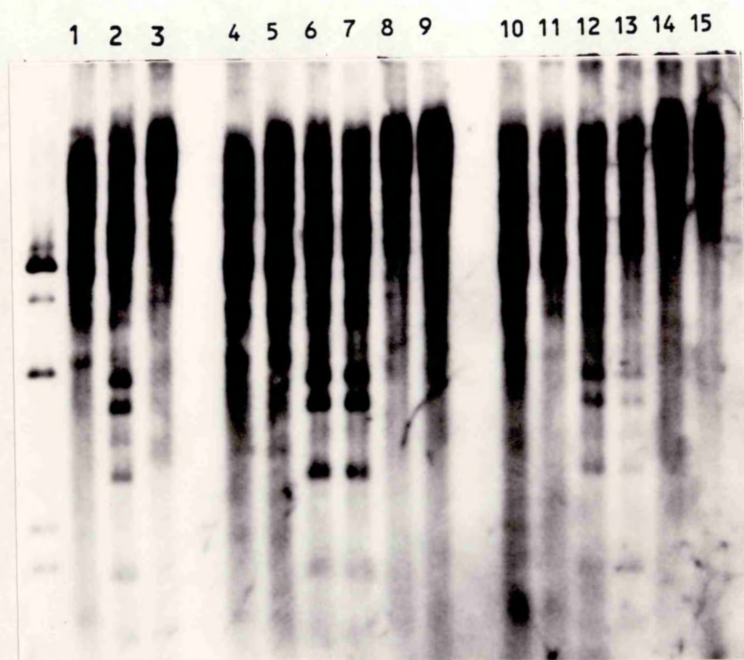
*Spe*I/*Eco*RV *Pst*I/*Hinf*I

Lane	Sample
1	107027 K
2	107311 LN
3	3132

Fig. 4.5: a). Southern analysis of *SpeI/Eco*RV-digested canine genomic DNAs. The transfer membrane was probed with the L4 *SpeI/Eco*RV (3' *pol*) probe. Probe-length fragments (0.54 kbp.) were not detected; suggesting that few proviral elements in canine genomic DNA have the same arrangement of *Spe*I and *Eco*RV sites as L4.

b). Southern analysis of *HinfI/Pst*I-digested canine genomic DNAs. The Southern transfer membrane was probed with the L4 *HinfI/Pst*I (5' *pol*) probe. The complex pattern of bands seen when samples were digested with *Pst*I alone (Fig. 4.3) has been dramatically simplified. Fragments of probe length (426 bp.) are present; suggesting that some proviral elements in canine genomic DNA may have the same arrangement of *Pst*I and *Hinf*I sites as L4.

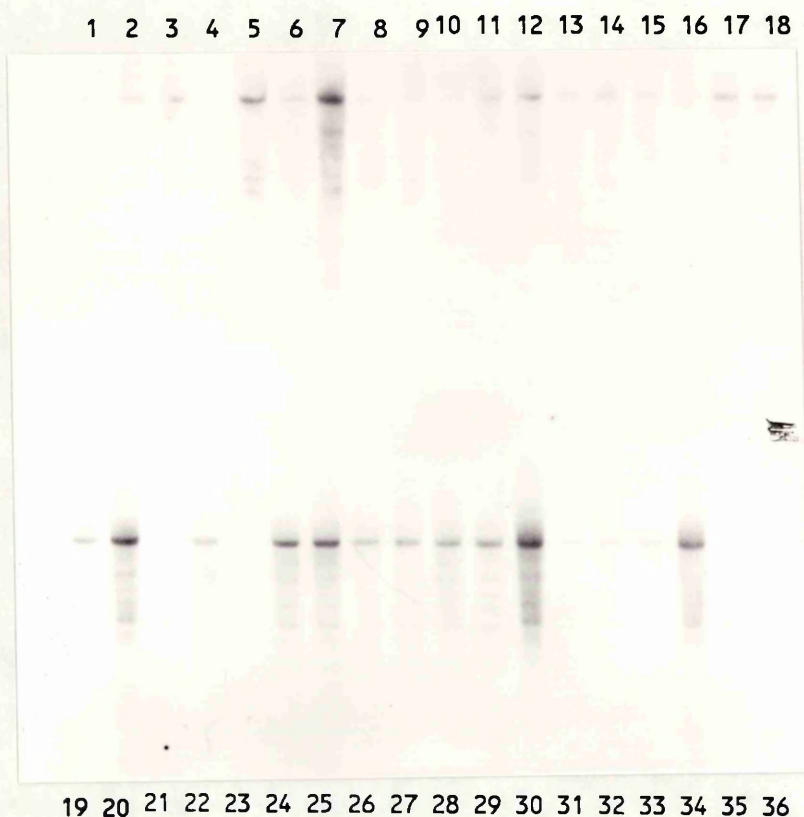
Abbreviation: S, a lane containing L4 fragments as convenient size markers: 2.47 kbp. *Sst*I fragment, 1.57 kbp. *Sst*I-*Xho*I fragment, 1.07 kbp. *Sst*I-*Xho*I fragment, 0.54 kbp. *Spe*I/*Eco*RV fragment and 0.46 kbp. *Sst*I-*Pst*I fragment. (Probes hybridise only to fragments of which they form part).



Lane	Case No.	Tissue	Digestion
1	107027	K	*Bam*HI
2	107027	K	*Sst*I
3	107027	K	*Xba*I
4	107311	K	*Bam*HI
5	107311	LN	*Bam*HI
6	107311	K	*Sst*I
7	107311	LN	*Sst*I
8	107311	K	*Xba*I
9	107311	LN	*Xba*I
10	107318	K	*Bam*HI
11	107318	Lb	*Bam*HI
12	107318	K	*Sst*I
13	107318	Lb	*Sst*I
14	107318	K	*Xba*I
15	107318	Lb	*Xba*I

Fig. 4.6: Southern analysis of canine genomic DNAs probed with FMuLV. The membrane was probed at low stringency. The *Sst*I digestions shows a pattern of recruited bands. No RE polymorphisms are apparent.

Abbreviations: K, kidney; LN, lymph node; Lb, peripheral lymphoblastic mononuclear cells



Lane	Case No.	Tissue
1	NK 1	K
2	107027	K
3	107027	Thymus
4	N/A	A 72-F
5	N/A	3132
6	107311	K
7	107311	LN
8	107318	K
9	107318	Lb
10	108601	K
11	108601	LN
12	107235	K
13	107235	LN
14	107490	Heart
15	107490	K
16	107490	Liver
17	PB	K
18	PB	LN

Lane	Case No.	Tissue
19	PB	Lb
20	108666	BM
21	108666	Brain
22	108666	K
23	109456	Brain
24	109456	LN (biopsy)
25	109456	LN (PM)
26	108548	LN
27	108557	LN
28	109341	LN
29	109342	LN
30	109393	LN
31	109533	LN
32	107227	LN
33	107228	K
34	70137	K
35	108386	LN
36	107190	LN

Fig. 4.7: Southern analysis of 36 *Eco*RI-digested canine genomic DNAs probed with HTLV-1. A recruited 7.5 kbp. band was evident in each lane of the original autoradiograph (some have been lost during photography). High stringency hybridisation; one week exposure of autoradiographic film to the transfer membrane.

Abbreviations: BM, bone marrow; K, kidney; LN, lymph node; Lb, peripheral lymphoblastic mononuclear cells; PM, post mortem.

use of these three retroviral probes. The BaEV probe produced slightly different results (Fig. 4.8).

When *Eco*RI-digested canine DNAs were probed at low stringency, there were multiple hybridising bands, as with HTLV-1 (and previous experiments with FMuLV). At high stringency, however, a single, 6.7 kbp. band was present in all lanes. The intensity of hybridisation to this DNA fragment was consistent with it being present at single copy in the haploid genome. In one of the lymphomatous lymph node samples (case no. 109393), there were two bands: the 6.7 kbp. band and an extra one of 6.9 kbp. It was thought that the DNA fragment present in all tested canine DNAs might represent a canine endogenous proviral element closely related to BaEV, perhaps equivalent to the feline RD114 endogenous virus. The extra band in case no. 109393 DNA was also of interest. Consequently, the work described in Chapter 6 was carried out to isolate and determine the nucleotide sequence of this hybridising fragment. It was subsequently determined not to be a retroviral element.

4.5 DISCUSSION

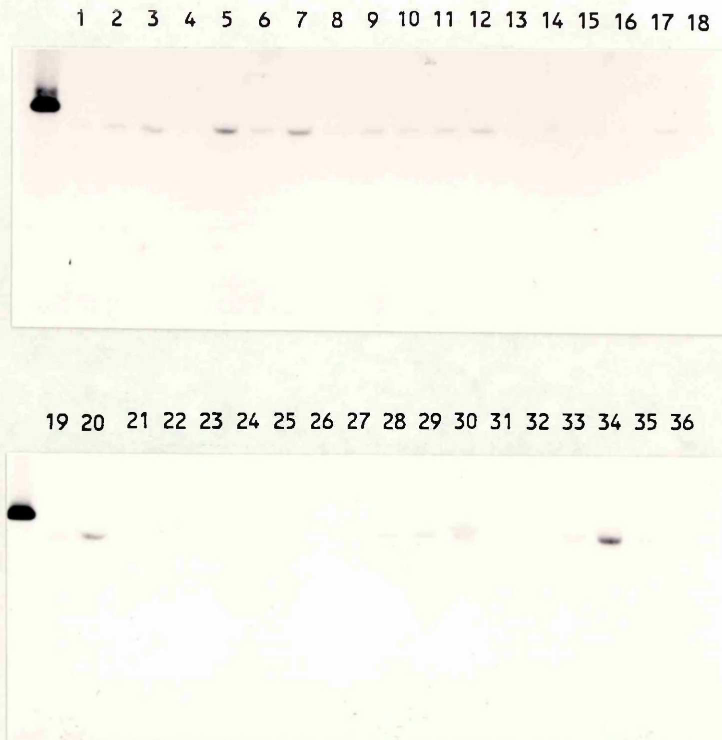
In this study, all tested samples of canine DNA were found to contain an apparently identical complement of numerous retrovirus-like elements. 3132 DNA was no different in this respect from any other sample, showing that the presence of retroviral sequences is a general feature of canine DNA rather than being a unique feature of the 3132 cell line. By definition then, these are canine *endogenous* retroviral sequences. Since autoradiographic bands were so numerous that they tended to merge, the total number of proviral elements could not be determined. Another factor which hampered estimation of the total number of proviruses was the fact that several of the autoradiographic bands were highly recruited. It is likely that the total number of proviral elements in this family exceeds 100. Recruitment of small DNA fragments most probably represented release of identical internal fragments from multiple proviral elements. The large recruited fragments were more likely a consequence of one or more DNA amplification processes involving segments of DNA larger than the provirus itself.

These findings parallel those of Steele *et al.* (1984, 1986). In their studies of the human endogenous virus family HERV-E, these workers probed human

Fig. 4.8: **a).** Southern analysis of three canine genomic DNAs (NK 1, 3132 and 107311 LN) digested with a variety of REs and probed at high stringency with BaEV. The extreme left lane contains 40pg of the BaEV plasmid insert DNA from which the probe was prepared. There are several hybridising fragments in each digestion; of which, a 6.5 kbp. *Eco*RI fragment hybridises most strongly.

Key: Each of the following set of three lanes contains NK 1, 3132 and 107311 LN, in that order. *Lanes 1-3, Bam*HI; *Lanes 4-6, Eco*RI; *Lanes 7-9, Hin*dIII; *Lanes 10-12, Pst*I; *Lanes 13-15, Sst*I; *Lanes 16-18, Xho*I.

b).



Lane	Case No.	Tissue	Lane	Case No.	Tissue
1	NK 1	K	19	PB	Lb
2	107027	K	20	108666	BM
3	107027	Thymus	21	108666	Brain
4	N/A	A 72-F	22	108666	K
5	N/A	3132	23	109456	Brain
6	107311	K	24	109456	LN (biopsy)
7	107311	LN	25	109456	LN (PM)
8	107318	K	26	108548	LN
9	107318	Lb	27	108557	LN
10	108601	K	28	109341	LN
11	108601	LN	29	109342	LN
12	107235	K	30	109393	LN
13	107235	LN	31	109533	LN
14	107490	Heart	32	107227	LN
15	107490	K	33	107228	K
16	107490	Liver	34	70137	K
17	PB	K	35	108386	LN
18	PB	LN	36	107190	LN

Fig. 4.8: b). Southern analysis of 36 *Eco*RI-digested canine genomic DNAs probed with BaEV. A strongly-hybridising 6.5 kbp. fragment was detected in each lane of the original autoradiograph (some detail has been lost during photography). There is an additional 6.7 kbp. hybridising band in Lane 30 (109393 LN). The extreme left lane of both rows contains 40pg of the BaEV plasmid insert DNA from which the probe was prepared. High stringency hybridisation. Transfer membrane exposed to autoradiographic film for one week.

genomic DNA from five unrelated individuals with a HERV-E *pol-env* probe. They found no restriction enzyme polymorphisms among the five human DNAs in terms of the sizes of retrovirus-like DNA fragments or the apparent extent of recruitment. This was still true when REs known not to cut within viral sequence (*i.e.* flanking sequence cutters) were used in the Southern analyses. Later, this group probed human and chimpanzee DNA with a HERV-E flanking sequence probe (Steele *et al.*, 1986). A similar array of amplified DNA was present in the two primates. Steele *et al.* (1986) therefore proposed that an amplification process involving proviral and flanking cellular DNA sequences occurred before evolutionary divergence of these two primate species. Because they found no evidence of RE polymorphisms when probing human DNAs with HERV-E probes, they suggested that continued amplification or rearrangement of this particular family of HERVs was not occurring. The same seems to be true for the canine retroviral sequences studied here; although formally, it would be difficult to exclude the possibility that amplification is continuing at a slow rate to produce ever brighter recruited bands, without necessarily changing the arrangement of the bands. A study carried out very recently to investigate the arrangement of endogenous retroviral elements in genomic DNA from wild canids lends further support to the view that these amplified, defective retroviral sequences are ancient. This work is described in Appendix 1 of this chapter.

The dramatic simplification of the hybridisation pattern produced with the 5' *pol* probe when *Hinf*I and *Pst*I were used in a double digestion of genomic DNA compared with *Pst*I alone provided a vivid demonstration that these canine retroviral elements are members of closely related families. The initial pattern of at least 26 hybridising bands found in the *Pst*I digestion was resolved to just four bands in the double digestion. The smallest of these four bands was the same size as the *Hinf*I-*Pst*I fragment used to generate the probe. Since the probe was derived from a non-deleted, colinear retroviral segment, the three larger bands in the double-digestion autoradiograph might represent groups of retroviral DNA fragments closely related to the L4 element, but with specific DNA insertions or duplications. Alternatively, there might be families of reiterated proviral elements with conserved but quite different arrangements of *Pst*I and *Hinf*I sites in *pol* compared with L4.

Once it was realised that L3 and L4 represented canine endogenous retroviral elements, rather than sequences specific to 3132 DNA, the discovery in canine DNA of sequences distantly related to FMuLV was predictable. After all,

it was an FMuLV probe which had been used in the initial selection of the L3 and L4 clones. The HTLV-1 probe produced results similar to those obtained with FMuLV, although the intensity of hybridisation was much lower. Most likely these probes were detecting sequences highly conserved among retroviruses. The failure to detect MMTV-related sequences in the canine genome was unexpected given the interest in B-type oncoviruses as a possible cause of mammary neoplasia in this species (M. Brearley, personal communication). MMTV-related sequences have been found in a variety of other mammalian species, including humans (May and Westley, 1986; Ono, 1986). A rather unlikely explanation for the failure to detect MMTV-related DNA in the canine genome is that the probe consisted only of a *gag-pol* domain. It is possible that only MMTV *env*- or LTR-related sequences are present in canine genomic DNA. However, MMTV *gag-pol* probes have been used successfully to detect related sequences in human DNA (May and Westley, 1986). The unusual findings with the BaEV probe will be discussed in detail in Chapter 6.

This study provides no evidence to suggest that canine endogenous retroviral elements (CERVs) have an aetiological role in lymphoma or any other disease. Tissues from normal animals, normal tissues from tumour-bearing patients and neoplastic tissues had indistinguishable patterns of proviral elements. It would have been intriguing if evidence of independent endogenous retroviral pathogenicity had been found, since there are few clear examples of this in nature. Endogenous MMTV proviruses have been unequivocally shown to cause late-onset mammary tumours in certain strains of mice. In some species, for example cats and mice; recombination between endogenous proviral sequences and related, infectious, exogenous viruses can occur, leading to enhanced retroviral pathogenicity. These examples constitute the most conspicuous pathogenic roles of endogenous proviruses discovered to date. In other species, for example humans, there are no known exogenous retroviruses closely related to their endogenous proviruses. No definitive pathological significance whatsoever has so far been found for endogenous viral sequences in such species. The dog resembles the human in this respect. That said, it is important to recognise that a single proviral translocation can have dramatic oncogenic potential. The number of proviral elements detected on each autoradiograph in this study was so large that subtle changes of that kind would almost certainly have been missed.

The work described in this chapter has provided some answers to elementary questions about the arrangement of retroviral sequences in the canine genome. This work was carried out at a stage in the project when only canine retroviral *pol* sequences had been characterised. For the future, it is planned to probe canine genomic DNA with canine retroviral *env*, LTR and cellular flanking sequence probes. This will help better to characterise the retroviral sequences present in the canine genome and to understand the amplification machinery which has had such an impact upon their distribution.

Appendix 1: PRESENCE OF CERV-related SEQUENCES IN GENOMIC DNA OF WILD CANIDS.

INTRODUCTION

A preliminary study was carried out very recently to investigate the arrangement of CERV-related sequences in genomic DNA samples from a variety of non-domestic canid species. The primary objective of this study was to determine if sequences related to L3 and L4 were present in genomic DNA from a variety of canid relatives of the domestic dog, *Canis familiaris*. If such sequences were found, it would be of interest to compare their arrangement with that found in the domestic dog. Work described in Chapters 3 and 4 tended to suggest that L3 and L4 are ancient sequences. Therefore, it was anticipated that close relatives of *Canis familiaris*, such as the grey wolf (*Canis lupus*) would have a similar arrangement of these proviral elements to that found in the dog. Whether more distantly related canids would have distinctive patterns of hybridisation was unknown.

By studying species related to *Canis familiaris*, it was hoped that insights into the evolutionary history of the L3/L4 family of retroviral elements might be gained. It was also considered possible that these ancient sequences might provide useful information concerning canid phylogeny. Frisby *et al.* (1979) showed that it was not possible to establish phylogenetic relationships among galliform birds on the basis of the presence or absence of the endogenous provirus, RAV-0 (see Chapter 1, section 1.2.3). This was presumably because of independent germline infection of unrelated galliform species by the exogenous predecessor of RAV-0. However, in contrast to L3 and L4, RAV-0 is an intact provirus, capable of producing infectious virus. It has therefore, presumably, been incorporated into the galliform genome relatively recently in evolutionary terms. The L3/L4 family seems much more ancient and might therefore be less prone to such recent, complicating influences.

The murine GLN family of defective endogenous retroviral elements is present in genomic DNA of mice and other rodents, but not in non-rodent DNA (Itin and Keshet, 1986; Obata and Khan, 1988). A HERV-K (CUU)-related genome is present at the same chromosomal integration site in humans, chimpanzees, gorillas and orang utans; but not in ape lar gibbons. Thus, it seems that these ancient murine and primate sequences might potentially be used to answer questions about phylogenetic relationships in the distant past. Whether the L3/L4 family of canine elements might also be useful in this respect seemed worthy of investigation.

MATERIALS AND METHODS

The basic methods used to carry out this work were described in Chapter 2. 34 samples of high molecular weight genomic DNA from 19 species of canid were kindly provided by Dr R.K. Wayne of the Department of Biology, University of California, Los Angeles. Details of the geographic origin of each sample are provided in Table 4.A1.1. The samples were provided as precipitates in 70% ethanol. The samples were stored at -20°C for one hour, centrifuged (14 krpm, Eppendorf centrifuge, 10 minutes) and washed in 70% ethanol. After repeat centrifugation, the pellets were air dried and resuspended overnight at 4°C in 1xTE (pH 7.6). Spectrophotometric DNA quantitation revealed that some of the samples contained less than 20ug of DNA. These samples were divided into two aliquots and half was used in each RE digestion (see below). For samples which contained more than 20ug of DNA, 10ug was used in each digestion. Separate aliquots of the samples were digested with *Pst*I and *Eco*RV. The digested DNAs were subjected to Southern analysis: the *Pst*I-digested samples were probed at high stringency with the *Hinf*I-*Pst*I L4 probe and the *Eco*RV-digested samples with the *Spe*I-*Eco*RV L4 probe (see section 4.3.1 for a description of these probes).

RESULTS

Examination of the ethidium bromide-stained gels prior to transfer revealed that several of the DNA samples were severely degraded. Fig. 4.A1.1 shows the autoradiograph obtained after the *Pst*I-digested samples were probed at high stringency with the *Hinf*I-*Pst*I L4 probe. Fig. 4.A1.2 shows the corresponding autoradiograph for the *Eco*RV-digested samples and the *Spe*I-*Eco*RV L4 probe. These preliminary results are technically poor for the following reasons:-

1). Unequal amounts of DNA were present in the different lanes.
2). Totally inadequate amounts of DNA were present in some lanes.
3). Degradation of several of the samples had occurred.

Clearly, a much improved technical result could be produced by exclusion of the degraded samples and alteration of the amounts of DNA in the various digestions. There was not time to carry out this work prior to submission of the thesis, but it is planned for the future.

Despite these reservations, this preliminary work was felt to be of sufficient interest to include in the thesis. It is evident from Fig. 4.A1.1 that *Canis domesticus* (domestic dog), *C. lupus* (grey wolf), *C. latrans* (coyote), *C. rufus* and *C. simensis* have an indistinguishable pattern of hybridisation to the *Hinf*I-*Pst*I (5' *pol*) L4 probe. The slight difference in the apparent sizes of the hybridising elements in *C.*

Lane	Number	Species	Geographic origin
1	NK 1	*Canis familiaris*	Scotland
2	300	*Canis lupus*	Canada
3	251	*Canis lupus*	Canada
4	339	*Canis lupus*	Canada
5	51	*Canis lupus*	Canada
6	157	*Canis lupus*	Minnesota
7	E4	*Canis lupus*	Sweden
8	F15	*Canis lupus*	Mexico
9	390f	*Canis rufus*	Captive
10	299m	*Canis rufus*	Captive
11	3209	*Canis latrans*	California
12	4513	*Canis latrans*	Texas
13	4011	*Canis latrans*	Maine
14	4036	*Canis latrans*	Michigan
15	312	*Canis latrans*	Kansas
16	H3	*Canis simensis*	Ethiopia
17	W7	*Canis simensis*	Ethiopia
18	DD4	*Canis familiaris*	Ethiopia
19	1	*Urocyon cinereoargenteus*	California
20	38	*Urocyon littoralis*	Santa Rosa Island, CA
21	26	*Urocyon littoralis*	San Nicolas Island
22	Z4	*Lycaon pictus*	S.Africa
23	7105	*Vulpes macrotis*	Colorado
24	4206	*Vulpes macrotis*	N/A
25	35	*Alopex lagopus*	Captive
26	48	*Vulpes vulpes*	Israel
27	5	*Canis adustus*	Kenya
28	6	*Canis mesomelas*	Kenya
29	1 male	*Canis aureus*	Israel
30	633	*Canis aureus*	Kenya
31	104986	*Speothos venaticus*	NZP, Brazil
32	1	*Cerdocyon thous*	Brazil
33	N/A	*Ateolocynus microtis*	Brazil
34	N/A	*Cuon alpinus*	China
35	N/A	*Nycteruetes procyonides*	Japan
36	NK 1	*Canis familiaris*	Scotland

Table 4.A1.1: Details of the wild canid species studied and their geographic origins. In addition, this table serves as a key to the autoradiographs in Figures 4.A1.1 and 4.A1.2.

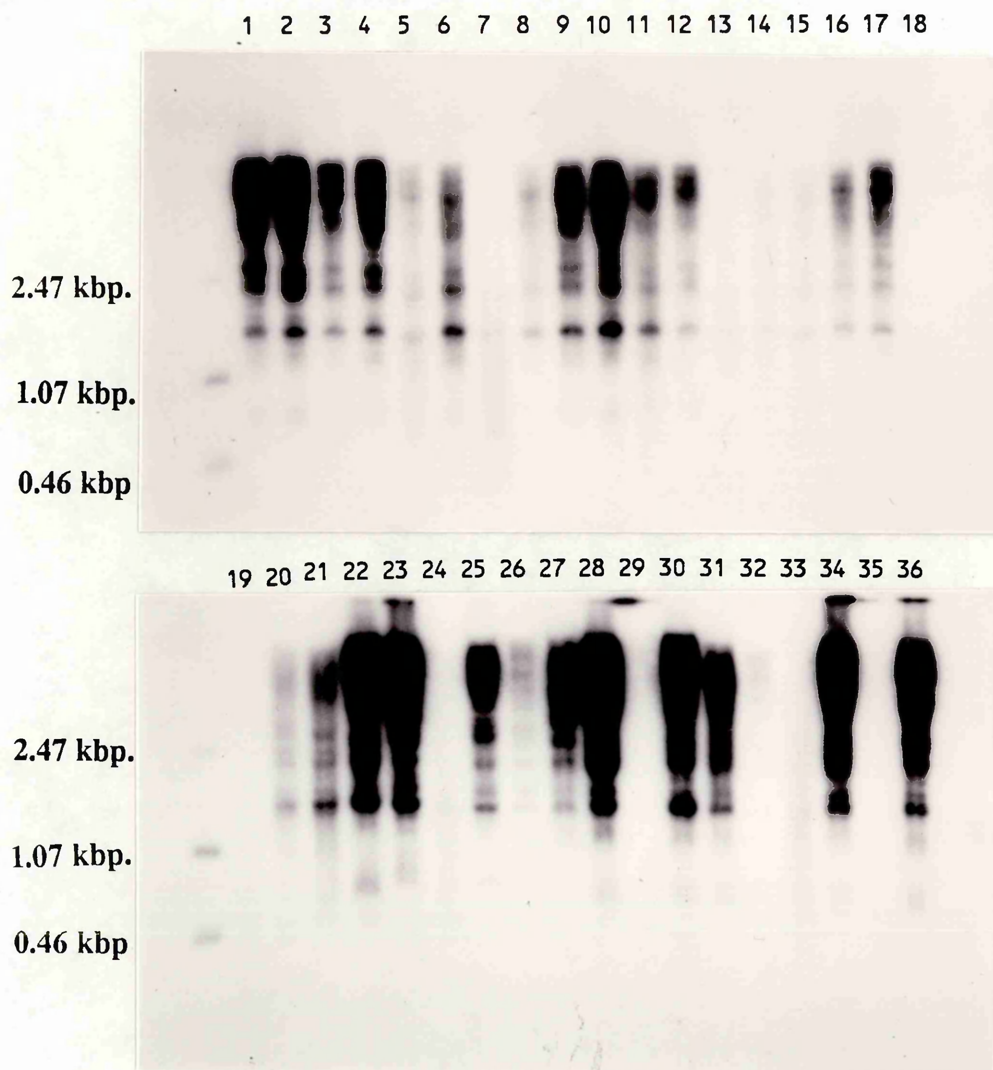


Fig. 4.A1.1: *See Table 1 for key.* Southern analysis of 36 *Pst*I-digested genomic DNAs from a variety of canid species, including the domestic dog. The Southern transfer membrane was probed with the L4 *Pst*I/*Hinf*I (5' *pol*) probe. The extreme left hand lane in each row contains L4 fragments as convenient size markers: 2.47 kbp. *Sst*I fragment, 1.57 kbp. *Sst*I-*Xho*I fragment, 1.07 kbp. *Sst*I-*Xho*I fragment, 0.54 kbp. *Spe*I/*Eco*RV fragment and 0.46 kbp. *Sst*I-*Pst*I fragment. Probes hybridise only to fragments of which they form part.

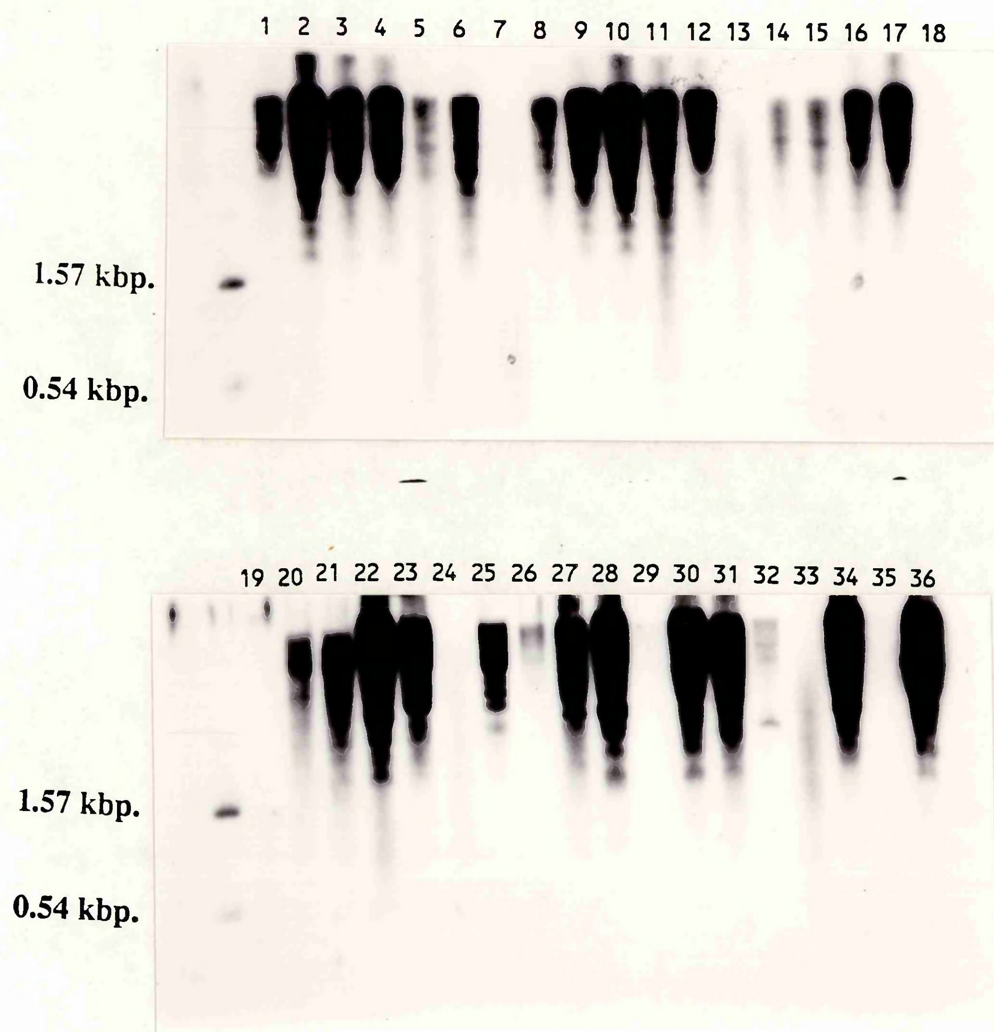


Fig. 4.A1.2: *See Table 1 for key.* Southern analysis of 36 *Eco*RV-digested genomic DNAs from a variety of canid species, including the domestic dog. The Southern transfer membrane was probed with the L4 *Spe*I-*Eco*RV (3' *pol*) probe. The extreme left hand lane in each row contains L4 fragments as convenient size markers: 2.47 kbp. *Sst*I fragment, 1.57 kbp. *Sst*I-*Xho*I fragment, 1.07 kbp. *Sst*I-*Xho*I fragment, 0.54 kbp. *Spe*I/*Eco*RV fragment and 0.46 kbp. *Sst*I-*Pst*I fragment. Insufficient of the 2.47 kbp. *Sst*I fragment is present to produce a signal. Probes hybridise only to fragments of which they form part.

domesticus DNA compared with the other species probably reflects a slightly different salt concentration in this sample compared with the others. Examination of an autoradiograph exposed to the same transfer membrane for a shorter time period (not shown) indicated that DNAs from *Lycaon pictus* (African wild dog), *Canis mesomelas* (Black-backed jackel) and *Canis aureus* (golden jackel) had a very similar pattern of hybridisation to that of the domestic dog. All of the other species in this study had a broadly similar hybridisation pattern to that of the domestic dog. In particular, all had a prominent hybridising element of 2.4 kbp. in length. Because of the poor quality of some of the DNA samples, it was difficult to identify with confidence specific differences in hybridisation pattern between the species. There appeared to be slight differences in the relative intensities and sizes of some of the intermediate-sized hybridising fragments in *Urocyon littoralis*, *Alopex lagopus*, *Cuon alpinus* and interestingly, *Canis adustus*. Further studies will be required, better to define these apparent differences.

Examination of Fig. 4.A1.2 yields little further information. However, it confirms that 3' *pol* sequences are present in genomic DNA of most, if not all, of the species examined. In these experiments, absence of hybridisation to the *pol* probes was correlated in every case with severe DNA degradation. It is therefore likely that all of the species examined harbour endogenous retroviral *pol* sequences in their genomes.

DISCUSSION

This study showed that endogenous retroviral elements related to L3 and L4 are present in the genomic DNA of several wild canid species. As expected, when DNAs from more than one individual of a species were compared, no differences were found. The domestic dog and several of its close relatives (*C. lupus*, *C. latrans*, *C. rufus* and *C. simensis*) have indistinguishable patterns of hybridisation to the L4 *pol* probes, both in terms of the sizes of hybridising DNA fragments and the extent of recruitment. The fact that the recruited bands had the same relative intensities indicates that the DNA amplification process which is presumed to have produced recruitment occurred prior to the evolutionary divergence of these closely-related canid species. This evolutionary divergence is thought to have occurred about 3 million years ago (Wayne and O'Brien, 1987).

Less closely related canids had a very similar pattern of hybridisation to the *pol* probes. For example, *C. mesomelas*, which is thought to have diverged from other wolf-like canids about 6 million years ago (Wayne and O'Brien, 1987), has a hybridisation pattern which is very similar to that of the domestic dog. *Alopex lagopus* diverged from the wolf-like canids about 9 million years ago, yet its

hybridisation pattern to the L4 *pol* probes is broadly similar to that of the domestic dog. This work therefore confirms that the L3/L4 family of elements are ancient and that the DNA amplification process which has had a major impact upon their distribution was under way at least 9 million years ago. These findings are wholly consistent with those described in Chapters 3 and 4. The highly defective nature of L3 and L4 suggested that these were ancient sequences, as did the identical arrangement of related elements in every sample of canine DNA examined. The findings closely parallel those of Steele *et al.*, 1986; who showed that amplification and dispersion of the HERV-E family of proviral elements occurred prior to the evolutionary divergence of humans from chimpanzees, more than 8 million years ago. More extensive Southern analysis of the L3/L4 family of ancient proviral elements may reveal insights into canid phylogeny. This work is planned for the future, in collaboration with Dr R.K. Wayne.

Appendix 2: ABSENCE OF CERV-RELATED SEQUENCES IN HUMAN GENOMIC DNA.

INTRODUCTION

A study was carried out to determine whether CERV-related sequences were present in human genomic DNA. Occasionally, diverse species are found to harbour in their genomes closely-related endogenous retroviruses. The most quoted example is that of the feline endogenous virus RD114, which is very closely related to the anthropoid primate endogenous virus, BaEV (Benveniste and Todaro, 1974b). To explain this similarity, it has been suggested that the anthropoid virus was horizontally transmitted to an ancestral cat about 5 to 10 million years ago (Benveniste and Todaro, 1974b). More often, it is found that endogenous viral sequences of a given species hybridise only to DNA of a few closely-related species. Given the close relationship between humans and dogs, it was considered of interest to determine if human DNA contained L4 *pol*-related sequences. The study was, in any case, a prerequisite to the work described in Appendix 3.

MATERIALS AND METHODS

Genomic DNA samples from human placenta and human MRC-5 cells were digested with *Hinf*I and *Pst*I and probed at high and low stringency with the canine L4 5' *pol* probe. In addition to the usual probe DNA control, three digested canine genomic DNA samples were included on each hybridisation membrane.

RESULTS

The samples of canine DNA produced an apparently identical pattern of hybridisation to the canine retroviral *pol* probe, consistent with previous results. There was no detectable hybridisation to placental or MRC-5 human genomic DNA.

DISCUSSION

This brief study showed that DNA sequences closely related to the canine endogenous retroviral family represented by L4 are not present in human genomic

DNA. The putative canine retrovirus released by the cell line A72-E could thus be studied in isolation by infection of human MRC-5 cells.

It is no surprise that human genomic DNA was not found to contain sequences which hybridise to the L4 probe. The L3/L4 family of elements are ancient. It is probable that an infectious, exogenous virus related to L4 and capable of infecting human cells has not been excreted by dogs during the period of companionship of these two species.

Appendix 3: A72-E cells express FeLV-B and appear not to be of canine origin.

INTRODUCTION

This study was carried out to determine if the infectious retrovirus released by the cell line A72-E was related to the canine endogenous retroviral family represented by L4. The history of the A72-E cell line is poorly documented. It arose during repeated passage of canine A72-F cells, which are fibroblastic. A72-E cells have epithelioid morphology. The virus which they produce is known to infect human cells. It was planned to infect human MRC-5 cells with the A72-E virus and extract infected human genomic DNA. Human DNA does not contain endogenous retroviral sequences which hybridise at high stringency with the L4 5' *pol* probe (Appendix 2). Therefore, by use of this system, the A72-E proviral sequence could be studied in isolation (*i.e.*, away from the multitudinous canine endogenous sequences).

MATERIALS AND METHODS

Cell culture and RDDP assays were carried out as described in Chapter 2 (sections 2.2.3 and 2.4.5). RDDP-positive A72-E supernatants were used to infect MRC-5 cells. MRC-5 cells were allowed to grow for 3 weeks after initial infection, splitting as necessary, to allow the retroviral infection to spread through the culture. Genomic DNA was extracted from sub-confluent, RDDP-positive A72-E cells and MRC-5 cells and also from uninfected MRC-5 cells. After digestion with a variety of REs, the DNAs were probed at high stringency with the 5' and 3' L4 *pol* probes. Each hybridisation membrane included control lanes which contained normal canine DNA. Subsequently, a membrane was probed with an exogenous FeLV U_3 probe. An ELISA test for detection of FeLV antigen was carried out in the laboratory of Professor Oswald Jarrett. Subgroup determination was carried out by M.Golder in that laboratory by a reverse interference test (See 2.2.3).

RESULTS

Canine DNA on each membrane showed a normal pattern of hybridisation to the L4 *pol* probe. There was no hybridisation to A72-E, infected MRC-5 or uninfected MRC-5 DNA. Since every previous sample of canine DNA examined

had contained an apparently identical complement of multiple endogenous retroviral elements, it was concluded that A72-E was not a canine cell line. It was not surprising, therefore, that proviruses of the A72-E virus, integrated in MRC-5, did not hybridise to the L4 element.

Hybridisation of an exogenous FeLV U_3 probe to a membrane which included canine and A72-E DNA produced multiple autoradiographic bands in the A72-E lane only. This indicated that the A72-E virus was FeLV or was a virus very closely related to FeLV. An ELISA test confirmed the presence of FeLV CA (p27) in A72-E culture supernatants. A reverse interference test showed that the A 72-E virus was FeLV of subgroup B.

DISCUSSION

A72-E cells are not of canine origin and they express FeLV-B, rather than a novel canine retrovirus. It is most likely that A72-F cells were contaminated and overgrown by FeLV-infected cells of another species. Since FeLV-B is polytropic, the species of origin of A72-E cells cannot be assumed to be feline. In fact, the species of origin of A72-E has not yet been determined. By examination of the arrangement of c-*myc*-hybridising DNA fragments, it appears that A72-E cells are not feline, murine or human. This work is described in Appendix 4.

Appendix 4: c-*myc* IS REARRANGED IN 3132 DNA.

INTRODUCTION

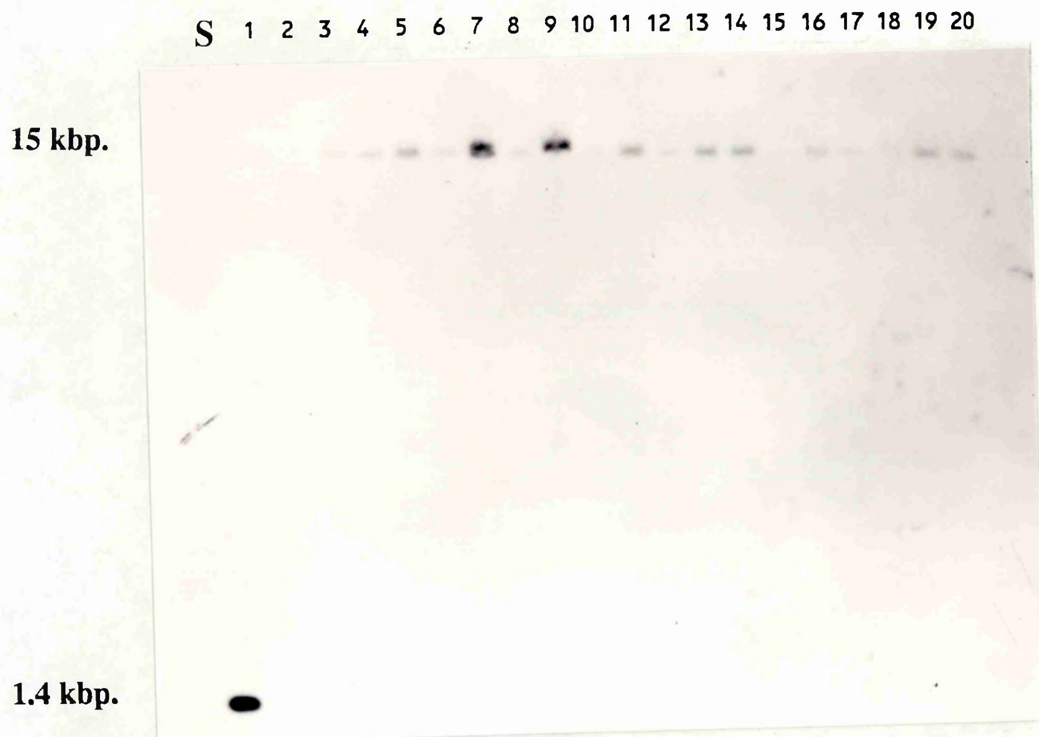
The proto-oncogene c-*myc* is important in the pathogenesis of many naturally-occurring tumours. Oncogenic interactions between c-*myc* and a retrovirus have been documented in several species. In the canine transmissible venereal tumour (TVT), a 1.8 kbp. non-viral retroposon is inserted 5' to the first exon of c-*myc* (Katzir *et al.*, 1985). These authors conjectured that retroposon-induced changes in TVT c-*myc* expression might have caused neoplastic transformation. A preliminary study to investigate the arrangement of c-*myc* in canine lymphoma genomic DNA was carried out. A secondary objective of this study was to compare the arrangement of c-*myc* in A72-E DNA with that in canine, A72F, 3132 and feline genomic DNA. The work of Katzir *et al.* (1985) had shown that *Eco*RI was a suitable RE for initial investigation of c-*myc* arrangement in canine DNA. Using this RE and two c-*myc* probes (one encompassing exon 1 and the other encompassing exons 2 and 3) these workers consistently detected 7.5 kbp. and 15 kbp. hybridising bands in normal canine genomic DNA. Consequently, *Eco*RI was used in this study.

MATERIALS AND METHODS

Eighteen canine genomic DNA samples derived from a variety of normal and neoplastic tissues were digested with *Eco*RI and subjected to Southern analysis at high stringency. 3132, A 72-F, A 72-E and FEA DNAs were included among the samples. The probe was a human c-*myc* probe, encompassing exons 1, 2 and 3. The probe was kindly provided by M. Riggio and D.A. Spandidos.

RESULTS

The pattern of hybridisation to the c-*myc* probe was identical for all canine samples except 3132. A 15 kbp. hybridising fragment was present in all canine samples. In addition, 3132 DNA had a 16.5 kbp. band (see Fig. 4.A4.1). When the amount of DNA on the membrane was taken into consideration, the intensity of hybridisation of all bands was consistent with presence at single copy in the haploid



Lane	Case No.	Tissue
1	N/A	A 72-E
2	N/A	FEA
3	NK 1	K
4	107027	K
5	107027	Thymus
6	N/A	A 72-F
7	N/A	3132
8	107311	K
9	107311	LN
10	107318	K
11	107318	Lb
12	108601	K
13	108601	LN
14	107235	K
15	107235	LN
16	107490	Heart
17	107490	K
18	107490	Liver
19	PB	K
20	PB	LN

Fig. 4.A4.1:
Southern analysis of 20 *Eco*RI-digested genomic DNAs. The Southern transfer membrane was probed at high stringency with a human c-*myc* probe encompassing exons 1, 2 and 3. Each canine DNA sample has a 15 kbp. hybridising fragment. Prolonged exposure showed that the FEA DNA had a hybridising fragment of about 11 kbp. A 72-E DNA contained a hybridising fragment which appeared slightly smaller than its canine counterparts. 3132 DNA, in Lane 7, has an extra c-*myc*-hybridising band of about 16.5 kbp. A 1.4 kbp. *Eco*RI-*Cla*I plasmid insert fragment encompassing human c-*myc* exon 3 is included on the membrane as a size marker and positive control in Lane S.
Abbreviations: K, kidney; LN, lymph node; Lb, peripheral lymphoblastic mononuclear cells

genome. The 7.5 kbp. band detected by Katzir *et al.* (1985) was not apparent. This may be a reflection of the higher stringency of hybridisation used in this experiment. Canine and feline DNA had slightly differently-sized, single c-*myc*-hybridising bands. A72-E DNA differed from both. Its hybridisation pattern was similar, but not identical to that of canine DNA.

DISCUSSION

The presence of a second, slightly larger c-*myc*-hybridising band in 3132 DNA compared with all other canine DNA samples was interesting. Since enhanced c-*myc* expression has been associated with immortalisation of cells, this finding may be of relevance to the establishment of 3132 cells in culture. Normally, canine lymphoma cells can be maintained in culture for only a few weeks. The 3132 cell line has been established and maintained for a decade. Sub-cloning and DNA sequencing of the aberrant 3132 c-*myc* fragment would help elucidate the significance of these findings and perhaps provide a mechanism to explain the observed change. An investigation of c-*myc* expression in 3132 would also be helpful. A simple chromosomal duplication would not explain the findings with 3132 DNA, since that would change the intensity of the hybridising band, but not its size. A DNA insertion within the span of the *Eco*RI sites of one of the two normal loci of the 3132 diploid genome might have occurred. Alternatively, a translocation of part of one of the c-*myc* loci is a possible mechanism. Partly because of difficulties in handling very viscous high molecular weight DNA, non-identical amounts of DNA were loaded into each lane of the gel used in this experiment. It is therefore uncertain whether the enlarged c-*myc*-hybridising band in 3132 represents one of the two normal loci of the diploid genome, or is a product of DNA amplification. In future, this problem could be resolved by further dilution of genomic DNA samples or quantitation of digested DNA immediately prior to gel loading.

The fact that all normal and lymphomatous canine DNA samples except 3132 had a normal c-*myc* arrangement in *Eco*RI-digested DNA should not be over-interpreted. Although this finding provides no evidence to suggest that c-*myc* is involved in the aetiology of canine lymphoma, a much more extensive study would be necessary to rule that out. In particular, the use of a variety of rarely-cutting

REs and pulsed-field DNA electrophoresis would allow examination of DNA segments larger than those investigated in this preliminary study.

A72-E DNA did not resemble feline, murine or human DNA in terms of c-*myc* arrangement. Although the pattern was similar to that of canine DNA, A72-E was already known not to be a canine cell line (Appendix 3). The species of origin of A72-E thus remains unresolved. It is possible, of course, that A72-E is a feline, human or murine cell line, the c-*myc* loci of which have been rearranged. However, it is rather unlikely that both loci would be identically rearranged. Although the problem was intriguing, it was not considered to be crucial to the main thrust of this project. The question was therefore left for the future.

Chapter 5

RNA EXPRESSION OF CANINE ENDOGENOUS RETROVIRAL ELEMENTS INCLUDING CELL CULTURE AND INDUCTION STUDIES

5.1 INTRODUCTION

5.2 TARGET RNAs

- 5.2.1 NORMAL TISSUES
- 5.2.2 NEOPLASTIC TISSUES
- 5.2.3 CULTURED CELLS

5.3 PROBES AND HYBRIDISATION CONDITIONS

5.4 AUTORADIOGRAPHIC FINDINGS

5.5 DISCUSSION

5.1 INTRODUCTION

This chapter describes a preliminary investigation into the RNA expression of canine endogenous retroviral elements. Total RNA was extracted from a variety of normal and neoplastic tissues and probed for the presence of retroviral *pol* sequences using probes derived from the L4 proviral element. Canine RNA samples were also probed with BaEV, FMuLV, HTLV-1 and MMTV. No hybridisation was found to the murine or primate retroviral probes. The canine *pol* probes detected a consistent pattern of short (< 1 kb.) hybridising molecules in a small minority of the RNA samples. Both normal and neoplastic tissues were represented in the group of samples which hybridised to the canine *pol* probes. Attempts to induce endogenous retroviral expression in 3132, A72-F and MDCK canine cells by growth in the presence of 5-iododeoxyuridine and 5-azacytidine were unsuccessful.

Having identified numerous proviral elements in canine genomic DNA, it was considered of interest to determine whether any of these elements were expressed as RNA. Previous work in other species had shown that, in general, endogenous retroviral elements are expressed in a similar manner to normal cellular genes. Much of the time they are transcriptionally silent, but they may become activated in certain tissues or at certain developmental stages of the host cell. Expression depends upon structural features of the provirus, particularly in the LTR regulatory sequences, but it is also dependent upon regulatory controls imposed by the host cell. The extent of proviral DNA methylation, dictated by cellular DNA methylase activity, has an important modulatory effect upon expression. The particular chromosomal integration site of the provirus may also have a positive or negative influence. Expression is not limited to complete proviruses: although defective proviruses cannot produce infectious viral particles, they frequently express sub-genomic mRNA and partial viral protein products. Even highly defective proviruses (such as L3 and L4), with frequent stop codons in all reading frames, can be expressed at the RNA level. Interestingly, the most highly expressed defective proviruses are those which express *env* and little else (Coffin, 1982b). This finding has been interpreted by some investigators as evidence that endogenous retroviral *env* expression confers a selective advantage

on the host, perhaps by preventing exogenous retroviral infection (Crittenden, *et al.*, 1982).

At the time when the work described in this chapter was carried out a selection of normal and neoplastic canine tissues were available for analysis. In retrospect, the available tissues (mostly kidney, brain, lymphomatous lymph node and peripheral mononuclear cells) were not necessarily ideal for an investigation of endogenous retroviral expression. Normal mammalian tissues which have recently been found spontaneously to express endogenous retroviral RNA, proteins and particles include placenta, epididymis, oocytes and mammary epithelium (for a review, see Larrson *et al.*, 1989). The predominance of reproductive tissues in this list is intriguing and may indicate a significant role for activated steroid receptor-responsive enhancer elements in retroviral expression. RNA from these reproductive tissues was not available at the time when this work was done. Nevertheless, it was considered of interest to determine whether retroviral RNA was expressed in the canine tissues available for analysis. The study was intended to show if there were differences in endogenous retroviral RNA expression between lymphomatous and normal tissues. This was of special interest, since a lymphoma cell line had been previously reported to produce an infectious retrovirus (cell line 3132, Strandstrom and Rimaila-Parnanen, 1979).

A variety of cultured normal and neoplastic mammalian and avian cell lines have been reported spontaneously to express endogenous proviral elements. Other, non-producing cell lines can be induced to express endogenous retroviruses. Physical agents which have been shown to induce expression include X-irradiation (Weiss *et al.*, 1971) and anoxia (Anderson *et al.*, 1988). A wide variety of chemical agents, including pyrimidine analogues, are even more effective (reviewed by Coffin, 1982b). 5-azacytidine (5-AZA), which causes DNA demethylation, is a particularly effective inducer of retroviral expression (Groudine *et al.*, 1981). 5-iododeoxyuridine (IdU) is another agent which has been used successfully for this purpose in the past (Teich *et al.*, 1973). It was decided that three established canine cell lines (3132, A72-F and MDCK) should be probed for retroviral RNA expression before and after separate exposure to IdU and 5-AZA.

5.2 TARGET RNAs

Total RNA was extracted from all tissues by the guanidinium thiocyanate-phenol-chloroform extraction method described by Chomczynski and Sacchi (1987). Details of the method are given in Chapter 2 (section 2.3.3 b). Table 5.1 gives summary details of the patients from which RNA samples were obtained.

5.2.1 NORMAL TISSUES

RNA was extracted from the thymus and kidney of a healthy 6 week old puppy, from the kidney of a healthy retired racing greyhound and from the liver of a mixed breed dog. In addition, 8 RNA samples were extracted from normal tissues of untreated multicentric lymphoma and ALL patients: six from kidney and two from brain. The tissues were subsequently examined histologically and shown to be free of neoplastic cells. Kidney was the first choice of tissue for such extractions. If it appeared that the kidneys were infiltrated by neoplastic cells, then brain, free of meningeal *dura mater*, was used instead.

5.2.2 NEOPLASTIC TISSUES

RNA was extracted from 11 affected lymph nodes of dogs with histologically-confirmed multicentric lymphoma and from a variety of tissues from three dogs with acute lymphoblastic leukaemia. Tissues from two cases of histiocytosis were examined: RNA was extracted from the skin of one dog and the liver of another. Samples of neoplastic tissue were obtained from single cases of renal lymphoma and thyroid carcinoma.

5.2.3 CULTURED CELLS

RNA was extracted from IdU-induced, 5-AZA-induced and uninduced 3132, A72-F and MDCK cells. Extractions were carried out when the monolayer cells were just sub-confluent by direct addition of $0.15ml/cm^2$ RNAzol™ to the culture flasks. 3132 cells were at a density of approximately $5x10^6/ml$ when they were pelleted and PBS-washed for RNA extraction. IdU was used at a concentration of 20ug/ml and 5-AZA at 3uM in all experiments. Induction protocols were as described in Chapter 2 (2.2.3).

5.3 PROBES AND HYBRIDISATION CONDITIONS

The probes used to carry out the work of this chapter and the hybridisation stringency parameters were the same as those described in Chapter 4. BaEV,

Table 5.1: Details of the patients and tissues from which total RNA was prepared.

Case No.	Signalment	Tissues Taken	Diagnosis
107027	6 wk F Sheltie	K, Thy	normal
107190	7 yr F Bull Terrier	K	mLSA
107228	Adult FN Crossbreed	Liver	normal
107235	Adult M G. retriever	K, LN	mLSA
107311	2 yr F G. retriever	K, LN	mLSA
107318	4 yr M GSD	BM, K, Lb, LN, Thy	ALL
107490	7 yr M Bernese	Liver	histiocytosis
108386	9 yr FN Rough collie	LN	mLSA
108548	5 yr F Crossbreed	LN	mLSA
108557	9 yr F GSD	LN	mLSA
108573	2 yr F Bernese	Skin	histiocytosis
108601	Adult F Lab	K, LN	mLSA
108666	6yr FN G. retriever	BM, Br, Lb, LN, Sp	ALL
109341	5yr F Great Dane	LN	mLSA
109342	9 yr M Irish Setter	LN	mLSA
109393	13 yr M Lab	LN	mLSA
109456	Adult M Lab	LN, Br	mLSA
109533	3 yr M WHWT	LN, Sp	mLSA
109801	8 yr F Crossbreed	Thyroid gland	Thyroid carcinoma
NK 1	7 yr M Greyhound	K	normal
PB	3 yr MN Lab	K	ALL

Abbreviations: M, male; F, female; N, neutered; wk, week; yr, year; Lab, labrador retriever; G. retriever, golden retriever; GSD, German shepherd dog; Bernese, Bernese mountain dog; WHWT, West Highland white terrier; LN, lymph node; K, kidney; Lb, peripheral mononuclear cells from dogs with ALL; Br, brain; Thy, thymus; BM, bone marrow; Sp, spleen; mLSA, multicentric lymphosarcoma; ALL, acute lymphoblastic leukaemia.

FMuLV, HTLV-1 and MMTV *gag-pol* probes were used at high and low stringency. The two probes derived from L4 *pol* (*Hinf*I/*Pst*I and *Spe*I/*Eco*RV fragments) were used at high stringency. One additional probe, encompassing L4 *env* sequences and part of the 3' LTR was prepared and used to probe the various induced and uninduced cell lines. It was not used to probe the RNAs from clinical material. This probe was prepared from a 530 bp. *Pvu*II/*Sst*I sub-fragment of L4 spanning positions 1949 to 2478 (Fig. 5.1). This fragment was purified from LMP agarose and subcloned into pBluescript KS (+) for probe preparation. All probes were labelled using the Multiprime labelling system of Amersham International plc., column purified from unincorporated label and used initially at an activity of $1x10^6$ cpm/ml of hybridisation buffer. When negative results were obtained, hybridisation experiments were repeated at low stringency with up to $5x10^6$ cpm of probe/ml of hybridisation buffer in an attempt to detect low-abundance mRNAs. Washing conditions differed somewhat from those described in Chapter 4. Each membrane was washed for 10 minutes in 2xSSC at room temperature to remove excess, unbound probe, followed by 3 washes of 20 minutes each at 65^oC in 0.5xSSC, 0.1% SDS. A final 10 minute rinse in 2xSSC at room temperature was carried out before autoradiography.

5.4 AUTORADIOGRAPHIC FINDINGS

Retroviral mRNA was not detected in extracts of A72-F, 3132 or MDCK cells by any of the retroviral probes: murine, primate or canine, including the new canine probe. This was equally true for RNA from uninduced, IdU-induced and 5-AZA-induced cells. The strategies intended to induce retroviral expression in these cell lines were therefore unfruitful. Autoradiographic film was exposed to the probed Northern transfer membranes for three weeks. This revealed only modest hybridisation to the ribosomal RNA (rRNA) molecules which predominate in extracted total RNA.

The murine and primate retroviral probes did not detect retroviral mRNA in any of the normal or neoplastic tissues examined. This was true for hybridisations at both high and low stringency, even when $5x10^6$ cpm of probe/ml of hybridisation buffer was used. Again, the only hybridisation evident after long

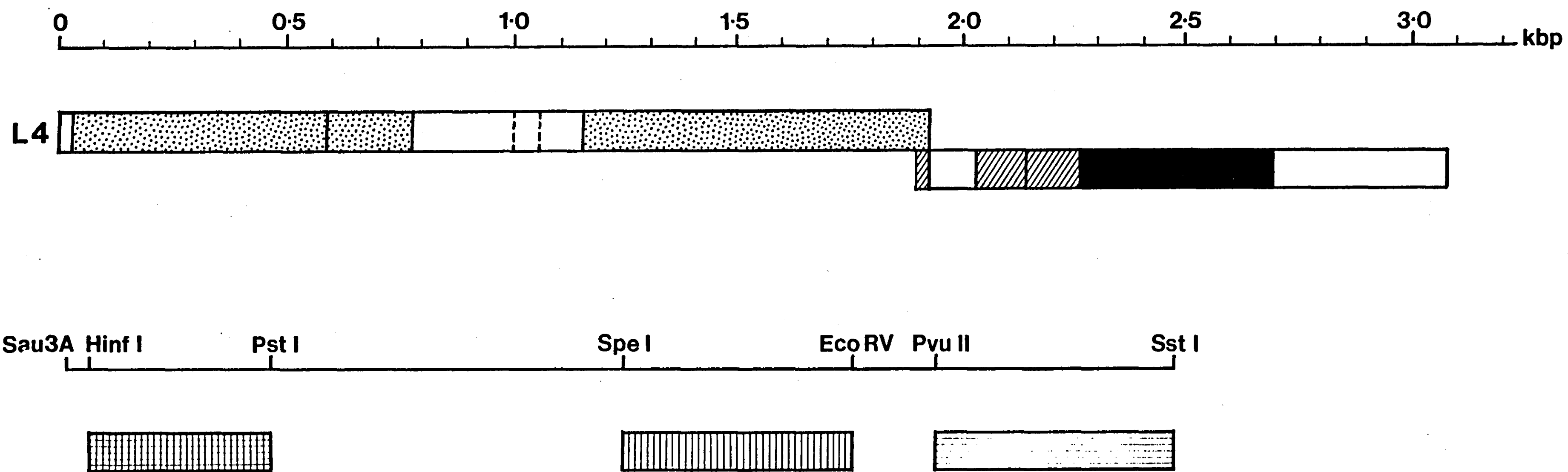
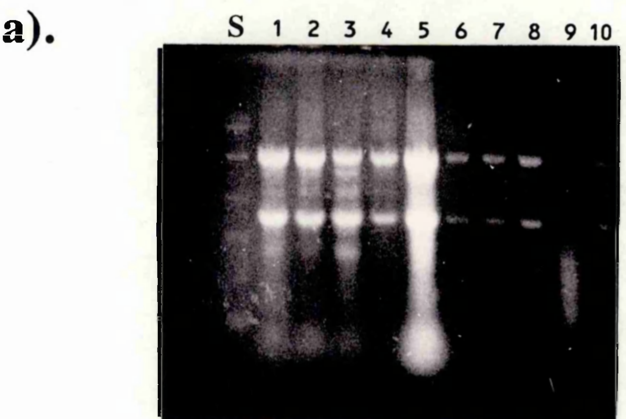


Fig. 5.1: Derivation of two retroviral *pol* probes and an *env*-LTR probe from L4. At the top of the figure is a diagrammatic representation of L4. Below is a restriction map, showing relevant restriction sites in the segment of L4 which had been characterised at the time of probe development. The bottom line shows the relative positions of three DNA sub-fragments which were used to generate the canine retroviral *pol* and *env*-LTR probes.

term exposure of the membranes to autoradiographic film was to rRNA and was of very modest intensity.

The L4 *pol* probes (both 5' and 3') detected a consistent pattern of five or more autoradiographic bands of varying intensity in a minority (9/35) of the tissue extracts (Figs. 5.2 and 5.3). These autoradiographic bands corresponded to RNA molecules of 0.9, 0.74, 0.54, 0.39 and 0.2 kb. in length. In 2 samples (107318 bone marrow and thymus), the bands were partially obscured by high background hybridisation across the entire molecular size range. Disconcertingly, the hybridising bands were present in some RNA samples which, judging from the rRNA bands on ethidium bromide-stained gels, were heavily degraded (107318 bone marrow and thymus , 109533 spleen). In view of this last finding, it was considered probable that the hybridising activity represented contaminating DNA within some of the RNA preparations rather than mRNA. However, it was not clear why contaminating DNA should produce such a consistent, clean, banded pattern in several different samples. Sheared high molecular weight DNA would normally produce a smear across the entire molecular size range, rather than a consistent pattern of distinct bands. Although contamination was not evident from the $O.D._{260}/O.D._{280}$ ratios, a very small amount of DNA would have been sufficient to produce a strong hybridisation signal, given the copy number of L4-related proviral elements in canine genomic DNA. Repeated extraction of the RNA samples with RNAzol™ or RNase-free DNase treatment of the samples followed by repetition of the hybridisation experiments might have elucidated the nature of the bands. However, the unpromising findings in the induction study, the fact that markedly degraded RNA samples showed the pattern of bands and time constraints mitigated against urgent pursuit of this question. Studies are in progress to investigate the expression of retroviral elements in canine placenta. Parallel with the new work, the samples used to carry out the work of this chapter will be re-extracted and re-probed to resolve the remaining questions.

a).



Lane	Case No.	Tissue
1	108573	Skin
2	108601	LN
3	108601	K
4	108666	LN
5	108666	BM
6	109342	LN
7	109456	LN
8	109456	Brain
9	109533	Spleen
10	PB	K

Abbreviations: K, kidney; LN, lymph node; BM, bone marrow.

b).

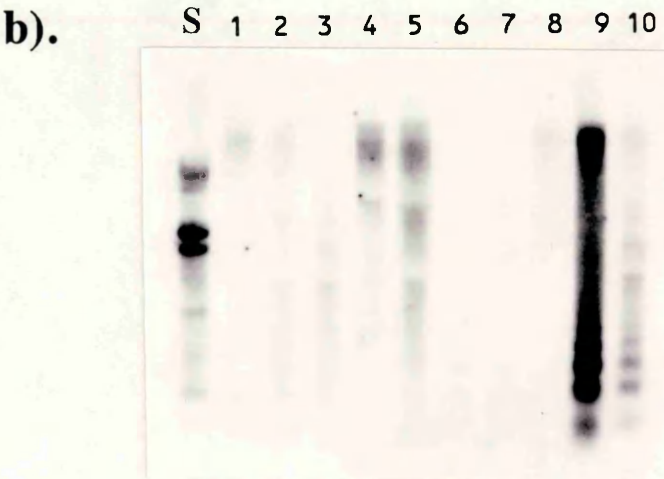
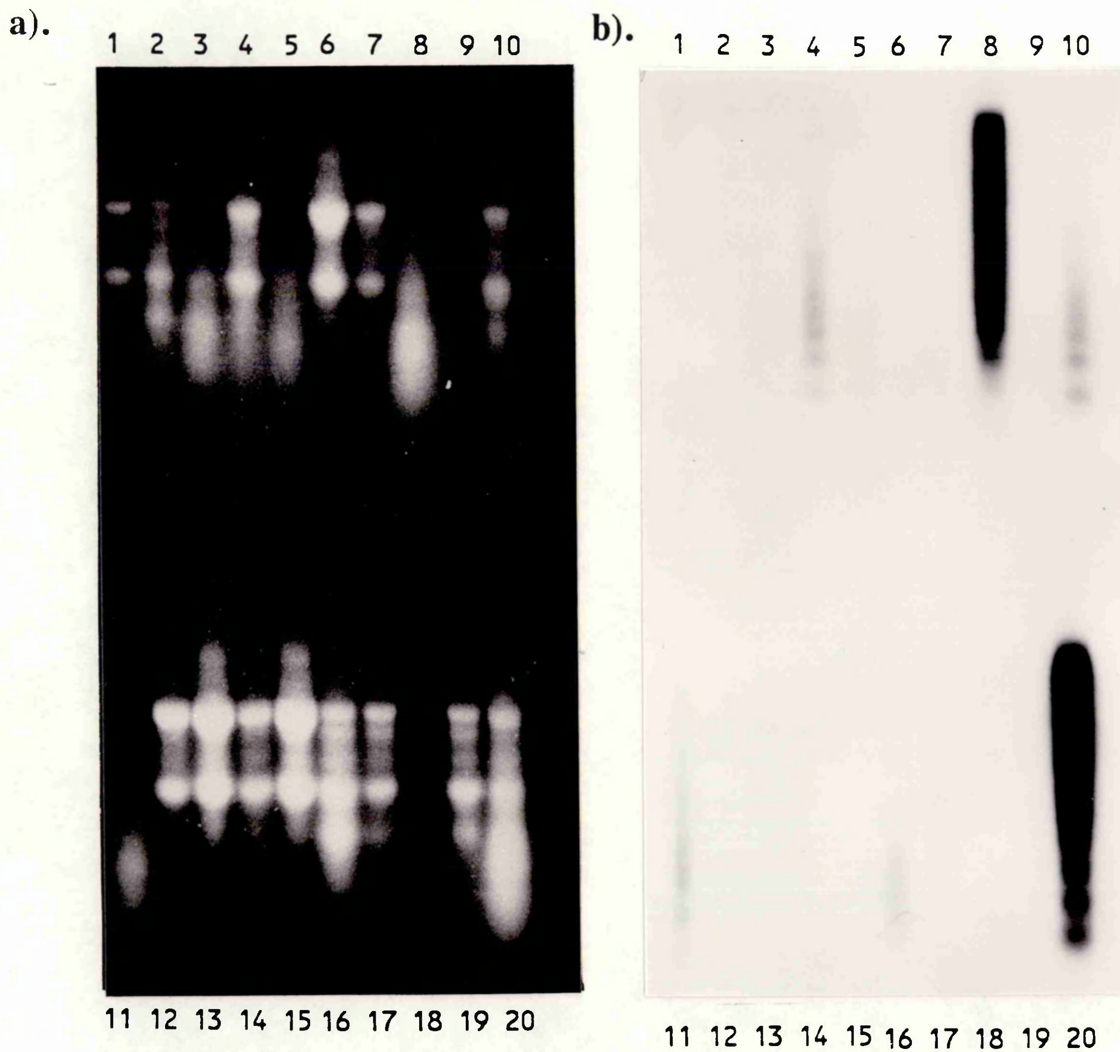


Fig. 5.2: **a).** An ethidium bromide-stained gel showing total RNAs prepared from a variety of normal and neoplastic canine tissues. Two intensely-staining rRNA bands are evident in each sample; except in Lane 9, which contains degraded RNA. The size marker lane (S) contains ss RNA molecules of 9.5, 7.5, 4.4, 2.4, 1.4 and 0.24 kb. in length.

b). An autoradiograph obtained by Northern analysis of the total RNAs shown in (a). The transfer membrane was probed with the L4 *Spe*I/*Eco*RV (3' *pol*) probe. Some fragments in the size marker lane have hybridised to the probe. The significance of this finding is unclear. Five or more small hybridising fragments are present in lanes 9 and 10, producing a banded pattern. Since lane 9 contains degraded RNA, these bands may represent contaminating DNA in the samples.



Lane	Case No.	Tissue
1	109456	LN
2	107027	K
3	107311	K
4	109801	Thyroid
5	107490	Liver
6	107318	LN
7	108666	Brain
8	107318	BM
9	109533	Spleen
10	108386	LN
11	107318	Thymus
12	109341	LN
13	109393	LN
14	109533	LN
15	107318	Lb
16	107318	K
17	107235	K
18	107027	Thymus
19	108601	K
20	107235	LN

Abbreviations: BM, bone marrow; K, kidney; Lb, peripheral lymphoblastic mononuclear cells; LN, lymph node.

Fig. 5.3: Repetition and extension of the work shown in Fig. 5.2.
a). An ethidium bromide-stained gel showing 20 samples of total RNA prepared from a variety of normal and neoplastic canine tissues. Three of the samples were shown in Fig. 5.2. Lanes 3, 5, 8, 16, 18 and 20 contain degraded RNA. Lanes 9 and 18 are markedly underloaded, or have become fully degraded.
b). An autoradiograph derived by Northern analysis of the total RNAs shown in (a). The transfer membrane was probed with the L4 *Spe*I/*Eco*RV (3' *pol*) probe. Hybridising fragments are present in lanes 4, 8, 10, 11, 16 and 20. All of these lanes except lane 4 contain samples showing evidence of degradation. A sample which previously hybridised strongly to the probe (109533 Spleen; see Fig. 5.2) has now become fully degraded. It seems likely that hybridisation is due to factors other than the presence of retroviral RNA.

5.5 DISCUSSION

The preliminary study described in this chapter does not provide convincing evidence to suggest that canine endogenous retroviral elements are expressed as mRNA in the tissues and cultured cells examined. Neither did the induction strategies employed succeed in causing expression. The nature of the hybridising activity found in a minority of the samples is presently obscure. These preliminary negative findings should be interpreted with caution.

In humans, a wide variety of defective endogenous retroviral elements have been identified, several of which are expressed as mRNA (Larsson *et al.*, 1989). The expression of HERV-R (or ERV3) has been particularly well characterised (Kato *et al.*, 1987). Interestingly, the transcription products of HERV-R are all spliced and lack 5.9 kb. of proviral sequence, including *gag* and most of *pol*. Three spliced, *env*-containing HERV-R mRNAs (9 kb., 7.3 kb. and 3.5 kb. in length) are highly expressed in human placental chorionic villi during the first trimester of pregnancy and at term. The 3.5 kb. mRNA is a typical sub-genomic, spliced transcript with its termination and polyadenylation site in the 3' LTR. The two larger mRNAs have a 5' end identical to the 3.5 kb. transcript, but extend through the 3' LTR and are spliced again about 370 bp. beyond the LTR. The downstream splice acceptor sequences have not yet been fully characterised, but each of the larger HERV-R mRNAs must contain several kilobases of human cellular sequences which are physically separated from the provirus. Most other human tissues express the 9 and 3.5 kb. mRNAs at a level of about 10% of that of placenta. The 7.3 kb. mRNA is apparently unique to placenta. No differences were noticed in HERV-R expression between normal and malignant tissues. Another quite different HERV, HERV-K (CUU), expresses *env*-containing mRNAs of 20, 14 and 6.8 kb. in melanoma cells (Ono *et al.*, 1987). It is very likely that these transcripts are a consequence of splicing events, as occurs in the case of HERV-R.

The study described in this chapter used three full length retroviral probes (BaEV, FMuLV and HTLV-1) and a *gag-pol* probe (MMTV) derived from retroviruses of species other than the dog. Two of the three canine retroviral probes encompassed only *pol* sequences. The *env*-3' LTR probe derived from L4 was used to probe RNAs from the three cell lines in the induction study. It was not

used to probe the 35 RNA samples obtained from clinical material. This was because the *env*-3' LTR probe was developed much later in the project than the L4 *pol* probes. It is therefore possible that some of the 35 canine tissues examined did not express L4-related *gag* or *pol*, but expressed L4-related, *env*-containing transcripts with insufficient homology to BaEV, FMuLV or HTLV-1 to allow detectable hybridisation. Absence of *gag* and *pol* expression might occur as a result of unusual splicing events, as with HERV-R. It is planned to test for the presence of L4-related, *env*-containing transcripts by probing a variety of canine RNA samples with the L4 *env*-3'LTR probe, or an improved derivative of it. Unfortunately, both L3 and L4 have very large deletions in *env*, so an optimal *env* probe cannot be immediately derived from either of them. In view of the recent, promising findings in other species, it is planned to probe RNAs extracted from canine term placentas for endogenous retroviral expression.

A clear finding of this work is that, neither before nor after the induction strategies, did A72-F, MDCK or 3132 cells express mRNA related to the retroviral probes used. In addition to the four non-canine retroviral probes, these cells were probed with L4 *pol*, TM-encoding *env* and LTR sequences. It is likely that if probe-related retroviral mRNA had been present in the samples, it would have been detected: the quality of RNA used in this part of the study was consistently excellent. The high quality of the RNA probably resulted from direct addition of RNAzol™ to the growing monolayer or pelleted suspension cells. This allowed instant and complete access of RNAzol™ to the cells. The failure to induce expression of L3- or L4-related retroviral mRNA may be explained by consideration of the multifactorial nature of eukaryotic gene expression. DNA methylation is only one of several factors which influence expression. In other species, certain cell lines are permissive for retroviral expression, others are not; irrespective of DNA methylation status. Presumably, certain cell lines produce regulatory factors required for retroviral expression. Enhancer and promoter sequences in the 5' LTR are critical for retroviral RNA expression. The canine endogenous retroviral elements for which sequence data is available (L3 and L4) do not have intact, recognisable, correctly-positioned enhancer or promoter sequences in their LTRs, although the L3 5' LTR does have such sequences in abberant positions. Even if RNA transcription were to begin, premature

termination and polyadenylation might occur, producing undetected, very short molecules. There are many AATAAA, polyadenylation signal sequences in L3 and L4, several of which are associated with retroposon tails. Although this signal sequence is not solely responsible for polyadenylation of eukaryotic genes, it is possible that some of the polyadenylation signals in L3, L4 and their relatives may be in a position which allows them to be active.

The failure to detect expression of BaEV-, FMuLV-, HTLV-1- and MMTV *gag-pol*-related mRNA in canine tissues might reflect absence of expression of canine retroviral elements related to these probes. Alternatively, failure to detect hybridisation might have been a consequence of relatively low target-probe homology and low levels of expression. In a future study, more promising tissues could be probed. The use of increased amounts of target RNA and poly-(A) selection would improve the sensitivity of the technique and might allow detection of lower levels of expression.

The significance of the banded, L4 *pol*-hybridising activity in nine of the RNA samples derived from clinical material is yet to be determined. At present it seems most likely that the banding was caused by contaminating DNA, given that it was present in two degraded RNA samples. The reason for the banding remains obscure. The banding was superficially reminiscent of a pattern occasionally seen when Southern transfer membranes are probed with radiolabelled, plasmid-derived insert fragments. This pattern is caused by hybridisation of plasmid sequences in the probe to contaminating plasmids in genomic DNAs extracted from bacterially-infected mammalian tissues. However, there are usually only one or two bands in such circumstances. The tissues used in this project were obtained at biopsy or very soon after death and showed no sign of bacterial infection. The very consistent pattern of bands in a number of samples also tends to argue against this explanation. Identical plasmid complements among multiple, independently-infected tissues would be unlikely. If further analysis shows that the hybridising activity is mRNA rather than DNA, then the uniform pattern of expression of short mRNA molecules in a minority of tissue samples would be consistent with what is known of the canine endogenous retroviral complement. L3 and L4 are heavily defective and truncated. Any mRNAs derived from them would be likely to include short molecules. All canine tissues apparently have an identical

arrangement of L4-related proviral elements. It is reasonable to suppose that, if stimulated to express these identical elements, different tissues might have a broadly similar pattern of expression. However, this certainly does not imply that differences in the pattern of expression would be impossible.

It was anticipated that this work might lead to the detection of retroviral genome-length mRNA expression and perhaps to the discovery of a novel infectious canine retrovirus. Instead, this study has provided no convincing evidence of endogenous retroviral expression in the dog. However, this was only a preliminary study and the anticipated study of canine placental mRNA may prove more fruitful.

Chapter 6

CHARACTERISATION OF A DNA SEQUENCE PRESENT BOTH IN CANINE GENOMIC DNA AND IN A CLONED BABOON ENDOGENOUS VIRUS

6.1 INTRODUCTION

6.2 MATERIALS AND METHODS

6.2.1 PROBLEMS WITH HIGH BACKGROUND HYBRIDISATION

6.3 CHARACTERISATION OF THE BaEV-HYBRIDISING ELEMENT IN CANINE DNA

6.3.1 RESTRICTION MAPPING OF THE LAMBDA GT10 INSERT

6.3.2 IDENTIFICATION OF HYBRIDISING SUB-FRAGMENTS

6.3.3 SUB-CLONING AND SEQUENCING STRATEGY

6.3.4 SEQUENCE DATA ANALYSIS

6.4 CHARACTERISATION OF THE HYBRIDISING DNA SEQUENCE IN THE BaEV PROBE

6.4.1 RESTRICTION MAPPING OF BaEV

6.4.2 BACK PROBING THE BaEV PROBE

6.4.3 BaEV SUB-CLONING AND SEQUENCING STRATEGY

6.4.4 DNA SEQUENCE DATA ANALYSIS

6.5 DISCUSSION

6.1 INTRODUCTION

This chapter describes how canine DNA which hybridised strongly to a BaEV M7 probe was selected from a genomic DNA library. When the hybridising canine DNA sub-fragment was sequenced, it was found not to be retroviral. Subsequent back-probing of the BaEV M7 clone identified an identical non-viral sequence upstream of the PBS. It is proposed that the BaEV clone under study acquired this segment of canine DNA at some time in the past, during *in vitro* culture in canine cells.

In Chapter 4 it was reported that when every sample of *Eco*RI-digested canine genomic DNA was probed with the BaEV probe at high stringency, a single, 6.7 kbp. autoradiographic band was found. The intensity of hybridisation to this DNA fragment was consistent with it being present at single copy in the canine haploid genome. In one of the lymphomatous lymph node samples (case no. 109393), there was an additional 6.9 kbp. hybridising fragment. It was thought that the DNA fragment present in all tested canine DNAs might represent a canine endogenous proviral element closely related to BaEV. The extra band in case no. 109393 DNA was also of interest. Consequently, a bacteriophage lambda DNA library was made from normal canine kidney. The intention was to probe the library with BaEV, isolate the hybridising DNA element and determine its nucleotide sequence. If the findings with normal kidney DNA had proved interesting, attention would have been turned to the case no. 109393 genomic DNA. It was anticipated that a novel, possibly non-defective canine endogenous retroviral element would be discovered; since hybridisation to the BaEV probe at high stringency produced a quite different pattern from that produced by the L4 family of defective retroviral elements. As it turned out, the hybridising canine DNA was non-retroviral. Back-probing of the BaEV M7 clone used in probe production, revealed evidence of a recombinational event involving the solitary LTR of this particular BaEV M7 clone.

6.2 MATERIALS AND METHODS

The materials and methods used to carry out the work described in this chapter were detailed in Chapter 2. Because the canine DNA fragment of interest was of a particular, known length, a size-selected library was prepared. Only

*Eco*RI-digested DNA fragments of 6 to 7 kbp. were excised and extracted from the LMP electrophoresis gel for cloning. The fact that one of the DNA size markers was 6,557 bp in length facilitated this procedure. The obvious advantage of the selection procedure was that it concentrated the fragments of interest within the heterogeneous population of molecules cloned into the library. This was intended to expedite screening.

6.2.1 PROBLEMS WITH HIGH BACKGROUND HYBRIDISATION

Although Lambda GT10 was the cloning vector used successfully to carry out the work described in this chapter, it was not used from the outset. Initially, a canine genomic DNA library was constructed using the vector Lambda Zap II, purchased from Stratagene (Short *et al.*, 1988). This insertion vector seemed ideal, in that it can accept upto 10 kbp. of foreign DNA at six unique cloning sites, one of which is *Eco*RI. Fragments which have been cloned into this vector can be excised by a helper phage and recircularised to generate subclones in the pBluescript SK (-) phagemid vector. Unfortunately, very high background plaque hybridisation was present on autoradiographs when this vector was used. All plaques, recombinant and non-recombinant, hybridised to the probe. The reason for this was not immediately apparent. A purified full length BaEV insert fragment had been used to generate the probe for these experiments. On ethidium bromide-stained gels, this fragment appeared entirely free of vector contamination. However, it is notoriously difficult to remove vector DNA completely from insert by electrophoresis. It was thought that a small amount of contaminating vector DNA in the probe had hybridised to pBluescript sequences in the Lambda Zap II arms to produce the unacceptably high background. To circumvent this problem, a second library in Lambda GT10 was prepared and probed. This vector was chosen because it contains an *Eco*RI cloning site which can accept up to 7.6 kbp. of insert DNA and, importantly, it does not contain plasmid sequences. Frustratingly, high background plaque hybridisation remained a problem with the new vector. At this stage, suspicions about the BaEV M7 clone were raised. Details about the origin of this clone were unforthcoming, despite repeated attempts to elicit this information. It was considered possible that the BaEV M7 insert might contain remnant bacteriophage lambda sequences from previous cloning experiments. To test this hypothesis, 100ug of sheared *Hin*dIII-digested, non-recombinant

bacteriophage lambda DNA (an aliquot of a DNA size marker) was added to the salmon testis DNA normally used to block non-specific membrane binding in hybridisation experiments (see 2.1.10). The mixture was denatured by boiling and added to the pre-hybridisation buffer. The objective was to promote hybridisation of the lambda arms on the membrane to a large excess of sheared, non-radiolabelled lambda DNA. This was intended to block binding to the postulated, contaminating radiolabelled lambda sequences in the probe. This strategy was highly effective. There was next to no background hybridisation when the membranes were probed at high stringency in the presence of an excess of *Hin*dIII-digested lambda DNA. A genuinely-hybridising recombinant plaque was quickly identified and triple plaque purified. Phage DNA was prepared on a large scale.

6.3 CHARACTERISATION OF THE BaEV-HYBRIDISING ELEMENT IN CANINE DNA

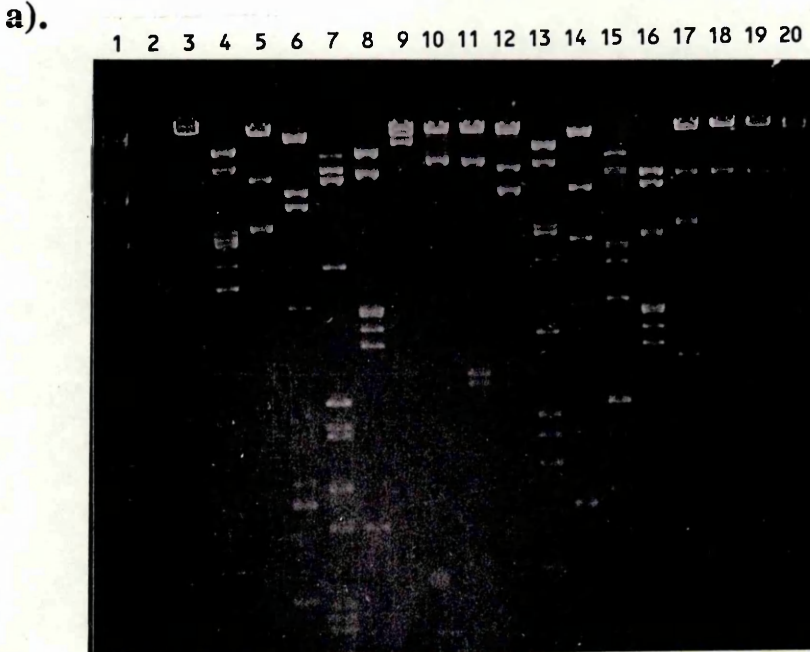
6.3.1 RESTRICTION MAPPING OF THE LAMBDA GT10 INSERT

After large scale DNA preparation, the recombinant phage genome was digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Pvu*II, *Sal*I, *Sst*I, *Xba*I and *Xho*I. Double digestions of the construct with *Eco*RI and each of the other enzymes were also carried out and run on the same electrophoresis gel (Fig. 6.1). A preliminary restriction map of the 6.7 kbp. DNA insert was thus produced. Later, the entire 6.7 kbp. *Eco*RI phage insert was cloned into pBluescript KS (+). This construct was digested with the aforementioned REs and all possible double digestion combinations. By this means, a detailed restriction map of the lambda GT10 insert was deduced (Fig. 6.2).

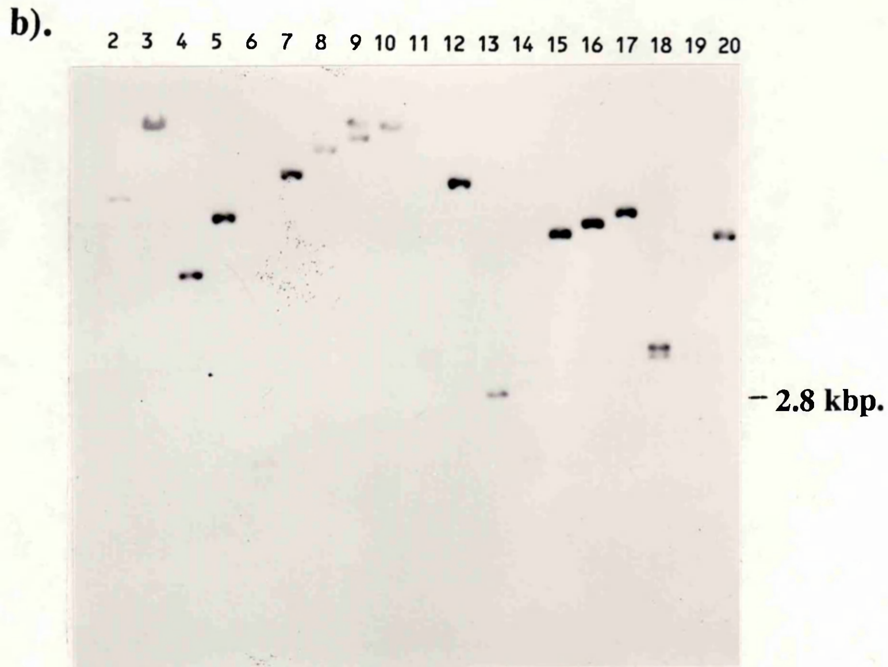
6.3.2 IDENTIFICATION OF HYBRIDISING SUB-FRAGMENTS

The DNA fragments in all gels used for deduction of the lambda insert restriction map were denatured, transferred to nylon and probed with BaEV. A variety of restriction fragments hybridised to the probe, of which a 2.8 kbp. *Eco*RI/*Bam*HI fragment at the right hand end of the lambda insert was initially most promising (Fig. 6.1). It was the only fragment in that particular digestion to hybridise to the probe *i.e.*, all of the hybridising activity was limited to that fragment.

a).



b).



Lane	RE(s)	Lane	RE(s)
1	size marker	11	*Xba*I
2	BaEV insert	12	*Xho*I
3	Uncut	13	*Eco*RI+*Bam*HI
4	*Bam*HI	14	*Eco*RI+*Hin*dIII
5	*Eco*RI	15	*Eco*RI+*Pst*I
6	*Hin*dIII	16	*Eco*RI+*Pvu*II
7	*Pst*I	17	*Eco*RI+*Sal*I
8	*Pvu*II	18	*Eco*RI+*Sst*I
9	*Sal*I	19	*Eco*RI+*Xba*I
10	*Sst*I	20	*Eco*RI+*Xho*I

Fig. 6.1: **a).** An ethidium bromide-stained gel showing single and double RE digestions of the recombinant Lambda GT10 clone which contained 6.7 kbp. of canine DNA homologous to our BaEV probe. A preliminary restriction map of the Lambda insert was deduced from these digestions. The DNA size marker was bacteriophage lambda DNA digested with *Hin*dIII.
b). An autoradiograph derived from the gel in (a). The Southern transfer membrane was probed with our BaEV clone. In the *Eco*RI/*Bam*HI digestion, BaEV hybridisation is limited to a 2.8 kbp. fragment (Lane 13). This fragment was chosen for sub-cloning.

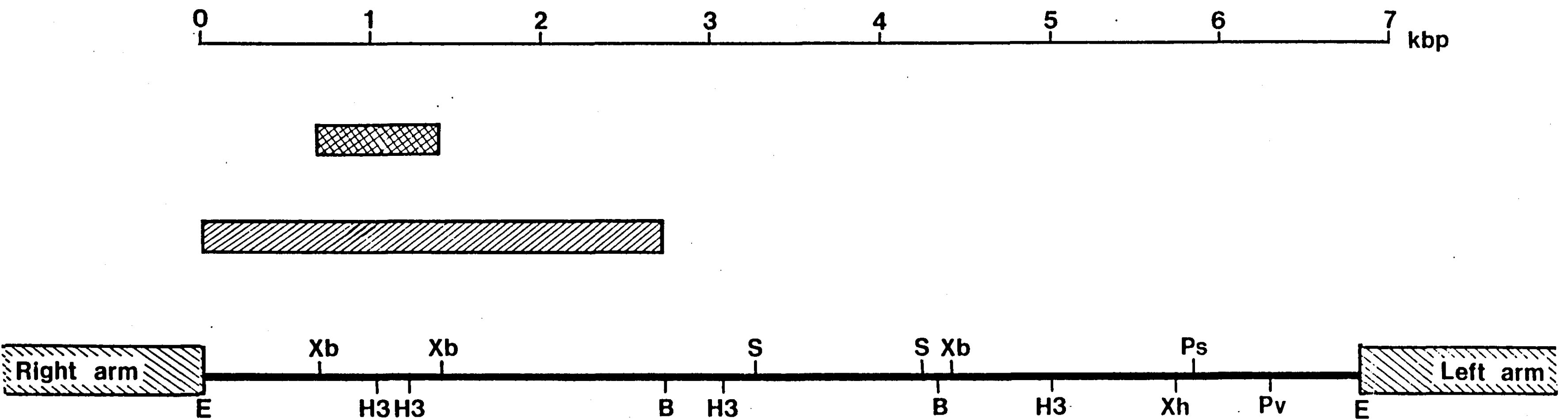


Fig. 6.2: A restriction map of the 6.7 kbp. *Eco*RI insert within the left and right arms of the bacteriophage lambda cloning vector, GT10. The map was deduced using the restriction data shown in Fig. 6.1. In addition, data from further restriction digestions was used. The diagram is orientated with the right lambda arm on the left to facilitate correlation with subsequent diagrams. The large hatched box indicates a 2.8 kbp. *Eco*RI-*Bam*HI fragment which accounted for all of the BaEV-hybridising activity when the insert was digested with *Eco*RI and *Bam*HI . The small cross-hatched box represents a 0.7 kbp. *Xba*I sub-fragment which was subsequently shown to account for all detectable BaEV-hybridisation within the clone.

Abbreviations: B, *Bam*HI; E, *Eco*RI; H3, *Hin*dIII; Ps, PstI; Pv, PvuII; S, SstI; Xb, XbaI; Xh, XhoI.

The 2.8 kbp. *Eco*RI/*Bam*HI fragment was consequently sub-cloned into pBluescript KS (+) and the resulting construct was digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Xba*I and double digestion combinations to free its component sub-fragments. After denaturation, transfer to nylon and probing of these DNA fragments, it was evident that BaEV-hybridising activity was limited to a 0.7 kbp. *Xba*I sub-fragment (shown crosshatched in Fig. 6.2). Clearly, the findings at this stage were disappointing and somewhat reminiscent of a Russian doll. A further disappointment was the fact that an alternative retroviral probe (FMuLV) did not hybridise to the 0.7 kbp. *Xba*I fragment , or to any of the other components of the 2.8 kbp. *Eco*RI/*Bam*HI fragment (data not shown because it was a blank autoradiograph). Nevertheless, it was decided that the nucleotide sequence of this *Xba*I fragment would be determined.

6.3.3 SUB-CLONING AND SEQUENCING STRATEGY

Several sub-clones were made to reduce the requirement for expensive sequencing oligonucleotides. Fig. 6.3 shows the restriction sites used to generate sub-clones. The extent of nucleotide sequences determined by individual sequencing reactions are represented as arrows. Rightward-pointing arrows represent sequences on the upper DNA strand, those pointing to the left represent sequences on the lower DNA strand. In addition to the 702 bp. *Xba*I fragment, a further 284 bp. of the directly adjacent 1.4 kbp. *Xba*I/*Bam*HI fragment was sequenced. Therefore 986 bp. of contiguous DNA sequence data was obtained (Fig. 6.4). The data from other, isolated sequencing reactions was not relevant to the findings in the BaEV M7 clone and is not presented.

6.3.4 SEQUENCE DATA ANALYSIS

The 986 bp. DNA element derived from canine normal kidney (henceforth called NK986) was compared with sequences in the Microgenie™ database using the MAKESEARCH software programme. Short segments of homology were found on the upper DNA strand to several mammalian genes, including the human coagulation factor IX gene and the human beta globin region on chromosome 11. Homology to these sequences was between 70 and 80%, but only over segments of 60 to 70 bp. in length. NK986 was translated in all 6 reading frames. Frequent stop codons were present in all frames.

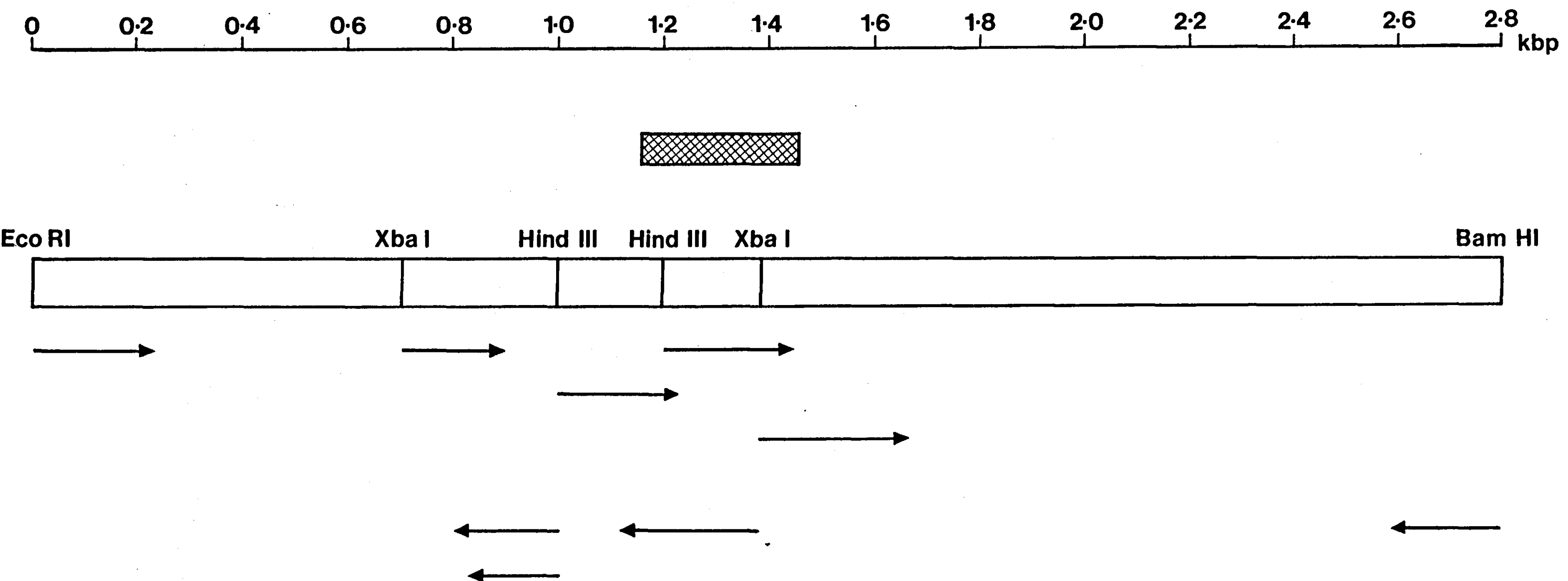


Fig. 6.3: Sub-cloning and sequencing strategy for the BaEV-hybridising component of the lambda GT10 insert. The diagram represents a 2.8 kbp. *Eco*RI-*Bam*HI sub-fragment which was originally at the right hand end of the lambda insert. Restriction sites used in the generation of sub-clones are shown. Specific fragments cloned into the pBluescript KS (+) sequencing vector were: 2.8 kbp. *Eco*RI-*Bam*HI, 1.4 kbp. *Xba*I-*Bam*HI, 1.0 kbp. *Eco*RI-*Hin*dIII, 0.7 kbp. *Xba*I, 0.21 kbp. *Hin*dIII and 0.17 kbp. *Hin*dIII-*Xba*I fragments. Leftward-pointing arrows indicate the extent of individual DNA sequencing reactions on the upper DNA strand. Rightward-pointing arrows indicate the same for the lower DNA strand. The cross-hatched box represents that part of this canine genomic DNA sequence which was subsequently found to be present as a contaminant in the LTR of our BaEV clone.

```
1   TCTAGATGAC AGCACTTAAC ACGTTGAGTA TTACTATAGG ATGAACCATT GGAAATTGAC AATACTGAGC CTTTTTAGAT CCCCCCAAAA TGGCAATCCT

101 CTGTGATTCA ACCCAATACA TGTATTCTAA TTGATTTGCA CGTATTCATC ACACTATCCC TGACACTCAG CTATAAGTAG AACTATGGTG TTATCCTTCC

201 TAGTTATAGT AAGGAGACCA AAGCTCACTG TGCTTGTGTG ACTCTCTCAA GGTCACACAG TAAGTTGAGG AGTCAGGATG GGATTCAAAT CCAGGCACTC

301 TGGCTTCCAA GCTTACTCCC TAAACCACAG ATTATTCTCT GGAGAGAGCC CAGGCCTGCA TTCTCAGATG GCGTTTCTCT AAAATGAGGC AGGACAACCA

401 GCCTTTGGCC AAGAAGGAAC TGGGGCTATT GCCTTCACTC TGGTAAAGTC ATAACAGACT TCACAAAGCC CACATATTCA ACCACTTCCT TTAATGGGTC

501 AGTCCTTATC TGCAATCAGA GAAAGCTTTA CATCATATTA TTAGAAGTCC TCTTTGAAGC TGTATTGACA CTTCAACCAG TTTTGAATAG TTCCAAATGA

601 GCGACTTAGA GCTTCAGGCC AATCAGGAAG AAGATTCCCA AGCTCTTCTC AACTCAGGTG CGGGTTCCTG AAATGGCCAC CCAATTTCAT CTGGCTTCTA

701 GACTCTGTCT CTTTGATCAT GTGTGATCTC ACAGGCTGGG TAATCTTCCC AGAACTCCCA TTTTCATTAT GTCATTCTTG GAACTGATAA TAGCTCCCCA

801 GTGCCTGATG GTCACATGCA AACTCTTGAA TCTGGCTTCT AAGGAACTCC TTTACTTGGC CTGTGCTTCC TTGGCCCCCT TCGGTCTCAC AGGTTCCTGA

901 TGGCTATCCT TTCCAGGAAG CCCATTCCCA CTTAATGCCC ACTTCCGCCA TTAATCCACT ATTTATGGCT TTGCTCCTTG AGTTCC
```

Fig. 6.4: NK986 sequence data.
The nucleotide sequence of a 702 bp. *Xba*I fragment (cross-hatched in Fig. 6.2) and 284 bp. of the adjacent 1.4 kbp. *Xba*I-*Bam*HI fragment was determined. The 986 bp. of contiguous sequence data thus obtained was compared with sequences in the Microgenie database. Short regions of homology to a variety of mammalian genes was found. NK986 has frequent stop codons in all six reading frames.

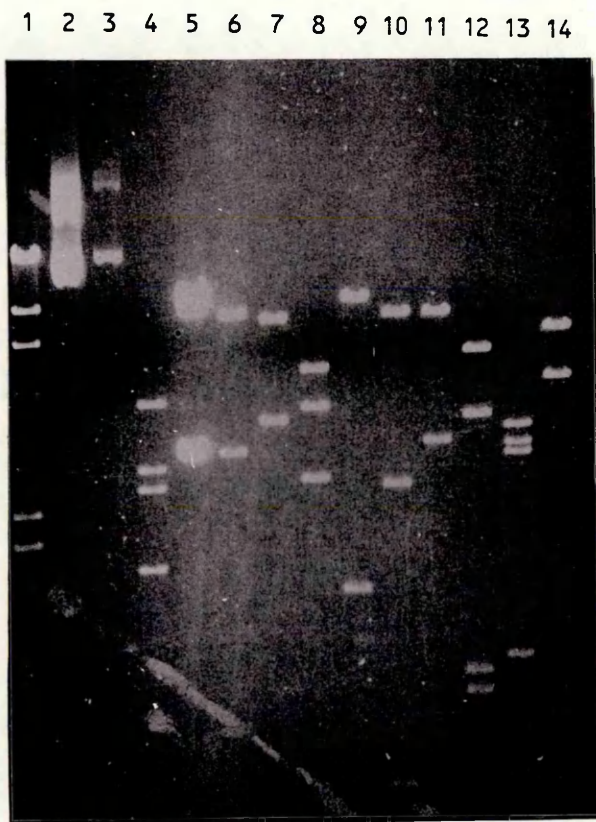
Evidently, part of NK986 hybridised to the BaEV M7 probe. Despite this, the canine sequence bore no resemblance to any of the retroviral sequences in a large database (Microgenie™ and GenEMBL), including the entire genome of BaEV itself (Kato *et al.*, 1987). It seemed very likely that the BaEV M7 clone had transduced a segment of cellular DNA and that this cellular sequence had hybridised to a single copy element in canine genomic DNA.

6.4 CHARACTERISATION OF THE HYBRIDISING DNA SEQUENCE IN THE BaEV PROBE

6.4.1 RESTRICTION MAPPING OF BaEV

The BaEV M7 construct originally provided was a 9 kbp., permuted retroviral provirus arranged as an *Eco*RI insert in pBR322. This construct was digested with *Bam*HI, *Eco*RI, *Eco*RV, *Hin*dIII, *Pst*I, *Pvu*II, *Sal*I, *Sst*I, *Xba*I and *Xho*I. Double digestions of the construct with *Eco*RI and each of the other enzymes were also carried out and run on the same electrophoresis gel. A preliminary restriction map of the BaEV insert was thus produced. For convenience, and to increase the yield of large scale plasmid DNA preparations, the intact 9 kbp. BaEV insert was cloned into the *Eco*RI site of pBluescript KS (+). This new construct was digested with the aforementioned REs, with the addition of *Bgl*II and the omission of *Eco*RV and *Sst*I. *Bgl*II was included because its restriction sites in a BaEV M7 clone have been previously reported (Tamura, 1983). *Eco*RV and *Sst*I were omitted for the converse reason. All possible double digestion combinations were also carried out (Fig. 6.5). Unsurprisingly, there were no differences in restriction pattern between the original BaEV insert in pBR322 and the insert in the new pBluescript construct.

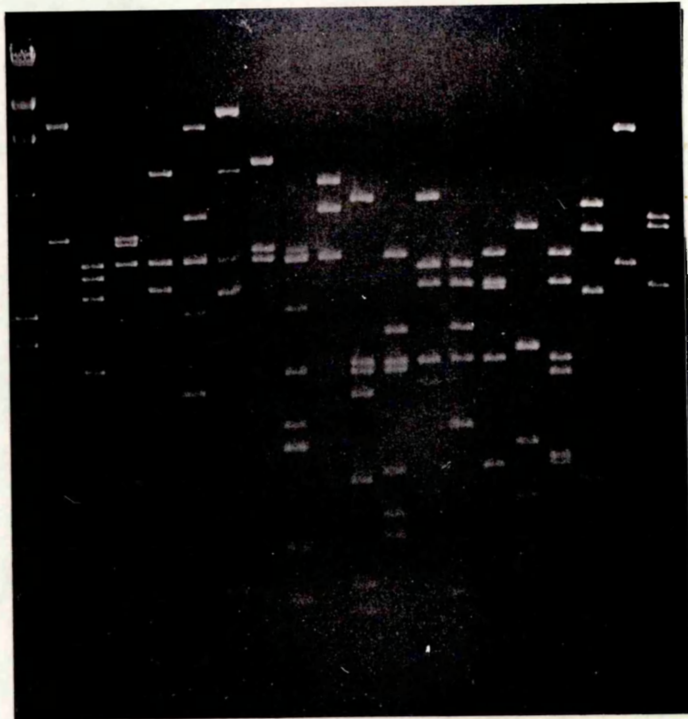
A detailed restriction map of the BaEV insert was deduced and compared with published maps (Cohen *et al.*, 1980; Cohen *et al.*, 1981; Tamura, 1983) and with a map derived from published sequence data (Kato *et al.*, 1987). The maps were broadly similar, but with several obvious points of difference (Fig. 6.6). The published maps included two LTRs, whereas the clone in our hands appeared to have only one. The evidence for this related to the *Xho*I restriction patterns of the various clones. *Xho*I was found to cleave in the LTRs of previously characterised BaEV clones. There were two or three *Xho*I sites in all of the published maps, but



Lane	RE(s)	Lane	RE(s)
1	size marker	8	*Hin*dIII
2	Uncut	9	*Pst*I
3	Uncut	10	*Pvu*II
4	*Bam*HI	11	*Sal*I
5	*Eco*RI	12	*Sst*I
6	*Eco*RI	13	*Xba*I
7	*Eco*RV	14	*Xho*I

Fig. 6.5: a). Single RE digestions of the BaEV-pBluescript KS (+) construct. Lanes 2 and 6 were grossly overloaded.

DNA size marker is bacteriophage lambda DNA digested with *Hin*dIII.

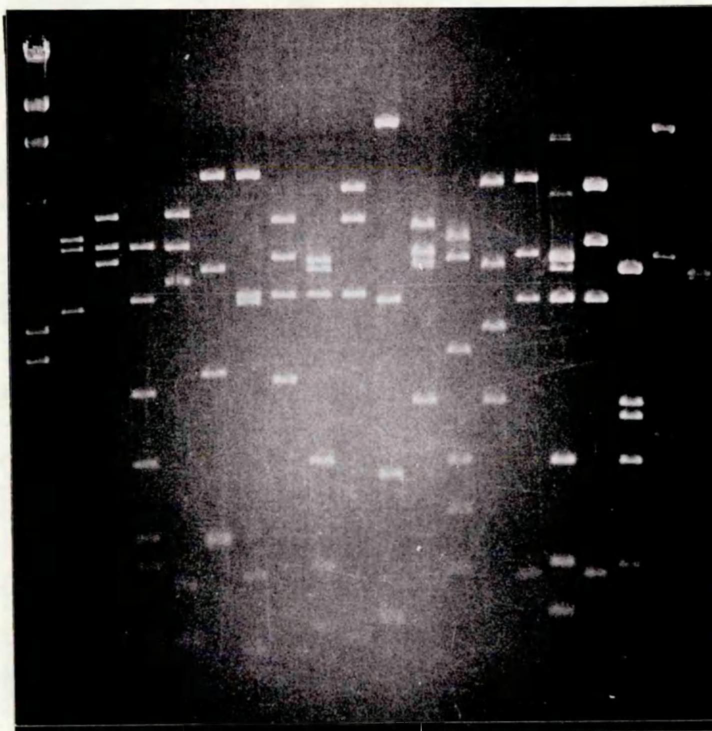


Lane	RE(s)	Lane	RE(s)
1	size marker	11	*BamHI* + *Bgl*II
2	*Bgl*II	12	*BamHI* + *Hin*dIII
3	*Eco*RI + *BamHI*	13	*BamHI* + *Pst*I
4	*Eco*RI + *Bgl*II	14	*BamHI* + *Pvu*II
5	*Eco*RI + *Hin*dIII	15	*BamHI* + *Sal*I
6	*Eco*RI + *Pst*I	16	*BamHI* + *Xba*I
7	*Eco*RI + *Pvu*II	17	*BamHI* + *Xho*I
8	*Eco*RI + *Sal*I	18	*Bgl*II + *Hin*dIII
9	*Eco*RI + *Xba*I	19	*Bgl*II + *Pst*I
10	*Eco*RI + *Xho*I	20	*Bgl*II + *Pvu*II

Fig. 6.5: **b).** Double RE digestions of the BaEV-pBluescript KS (+) construct. These digestions were used to generate a detailed restriction map of our BaEV insert.

DNA size marker is bacteriophage lambda DNA digested with *Hin*dIII.

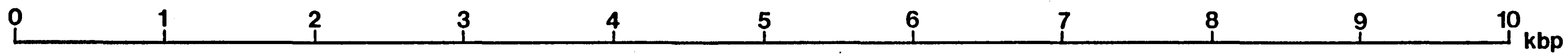
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20



Lane	RE(s)	Lane	RE(s)
1	size marker	11	*Pst*I + *Pvu*II
2	*Bgl*II + *Pvu*II	12	*Pst*I + *Sal*I
3	*Bgl*II + *Sal*I	13	*Pst*I + *Xba*I
4	*Bgl*II + *Xba*I	14	*Pst*I + *Xho*I
5	*Bgl*II + *Xho*I	15	*Pvu*II + *Sal*I
6	*Hin*dIII + *Pst*I	16	*Pvu*II + *Xba*I
7	*Hin*dIII + *Pvu*II	17	*Pvu*II + *Xho*I
8	*Hin*dIII + *Sal*I	18	*Sal*I + *Xba*I
9	*Hin*dIII + *Xba*I	19	*Sal*I + *Xho*I
10	*Hin*dIII + *Xho*I	20	*Xba*I + *Xho*I

Fig. 6.5: c). Further double RE digestions of the BaEV-pBluescript KS (+) construct. These digestions were used to generate a detailed restriction map of our BaEV insert.

DNA size marker is bacteriophage lambda DNA digested with *Hin*dIII.



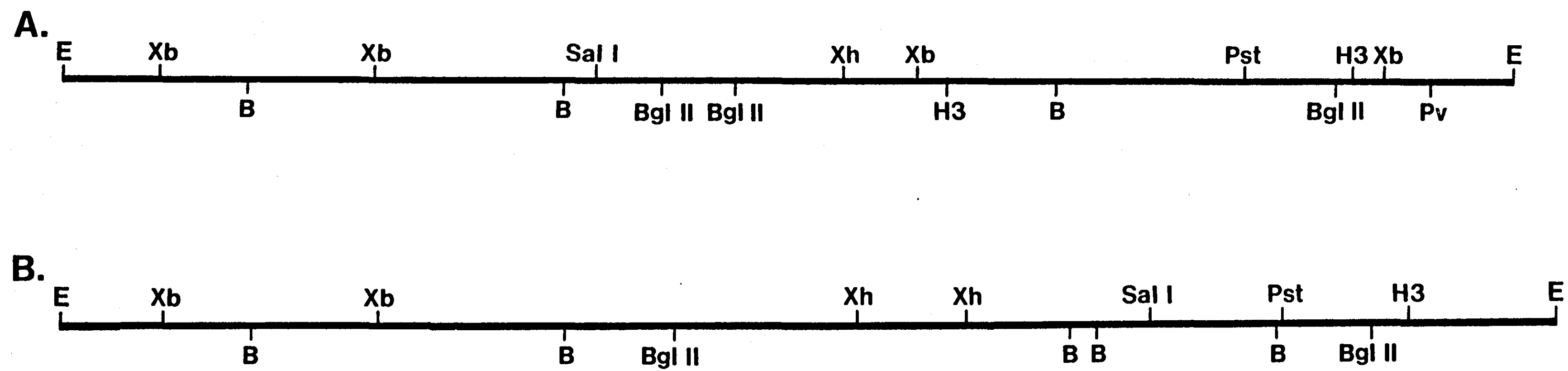


Fig. 6.6: A. Restriction map of the BaEV clone used for probe generation in our laboratory. The map was deduced from the double RE digestions shown in Fig. 6.5. The restriction map is compared with that of a published BaEV M7 permuted form (Tamura, 1983). Compared with the clone described in this publication (shown in B.), our clone lacked an *Xho*I site, indicating that this clone has only one LTR. There was also an abnormally positioned *Sal*I site, additional *Bgl*II, *Hin*dIII and *Xba*I sites and two absent *Bam*HI sites.
The cross-hatched box indicates the 0.9 kbp. *Xba*I-*Bam*HI DNA fragment which hybridised most strongly to the canine NK702 probe.

Abbreviations: B, BamHI; E, EcoRI; H3, HindIII; Pst, PstI; Pv, PvuII; Xb, XbaI; Xh, XhoI.

only one in our clone. Other features of our clone which differed from published maps included an abnormally positioned *Sal*I site, additional *Bgl*II, *Hin*dIII and X*ba*I sites and two missing *Bam*HI sites.

6.4.2 BACK-PROBING THE BaEV PROBE

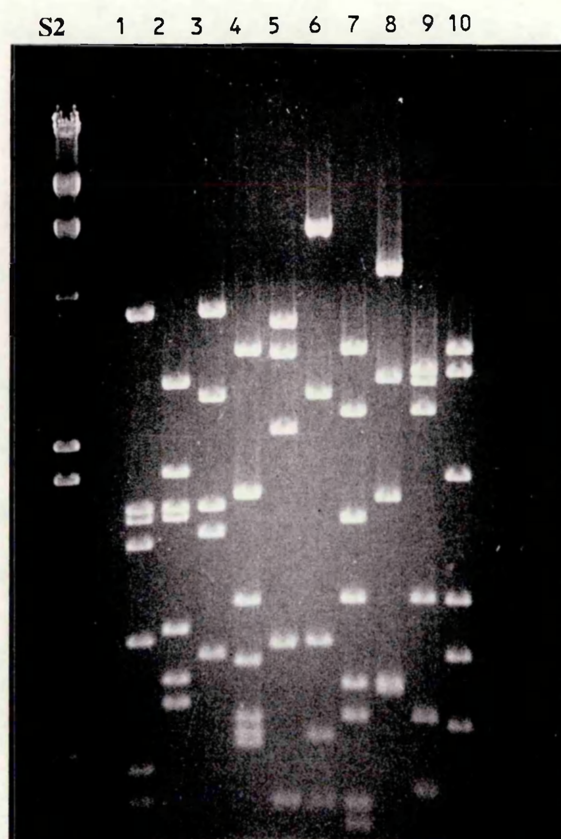
The 702 bp. *Xba*I canine DNA fragment (part of NK986, henceforth to be called NK702) which had hybridised to the BaEV clone was available as a sub-clone in pBluescript. The insert was freed by *Xba*I digestion and purified from LMP agarose for probe preparation. The BaEV-pBluescript construct was single- and double-digested with a variety of REs and the resulting DNA fragments were separated by electrophoresis and transferred to nylon. The membranes were probed at high stringency with radiolabelled NK702. The resulting autoradiographs (Fig. 6.7) showed that sequences within the BaEV clone which hybridised to the canine DNA were located in an approximately 1 kbp. DNA region just downstream (in terms of a normal retroviral genome) of the LTR *Xho*I site.

6.4.3 BaEV SUB-CLONING AND SEQUENCING STRATEGY

A 901 bp. *Xba*I/*Bam*HI fragment (cross-hatched in Fig. 6.6), which hybridised strongly to NK702 was sub-cloned into pBluescript KS (+) for sequencing. The adjacent 1.7 kbp. *Sal*I/*Xba*I fragment was similarly subcloned. The nucleotide sequence of the entire *Xba*I/*Bam*HI fragment was determined, using universal primers and two custom-synthesised oligonucleotides (Figs. 6.8 & 6.9). The ends of the *Sal*I/*Xba*I fragment were sequenced using universal primers. In total, 1058 bp. of contiguous DNA was sequenced. In addition, 215 bp. at the *Sal*I end of the *Sal*I/*Xba*I fragment was sequenced (not shown in Fig. 6.8).

6.4.4 DNA SEQUENCE DATA ANALYSIS

The contiguous sequence data from our BaEV clone is presented in Fig. 6.9(a). The diagram in Fig. 6.8 shows the structures represented by this DNA sequence. These were established by comparison of the new sequence with published BaEV sequence data and with NK986. The 1058 bp. segment of our BaEV clone was shown to consist, in order, of LTR (3' U_3 & R components), 286 bp. of interspersed foreign DNA (100% homologous to NK986; shown in Fig. 6.9(b) and a short reversed segment of p15-encoding *gag*; the last of which terminated immediately upstream of the PBS. The PBS, leader sequence and 5'



Lane	RE(s)
1	*BamHI* + *Bgl*II
2	*BamHI* + *Hin*dIII
3	*BamHI* + *Pst*I
4	*BamHI* + *Xba*I
5	*Bgl*II + *Hin*dIII
6	*Bgl*II + *Pst*I
7	*Bgl*II + *Xba*I
8	*Hin*dIII + *Pst*I
9	*Hin*dIII + *Xba*I
10	*Pst*I + *Xba*I

Fig. 6.7(a): A photograph of an ethidium bromide-stained gel, showing double RE digestions of the BaEV-pBluescript KS (+) construct. S2, DNA size marker, bacteriophage lambda DNA digested with HindIII.

Lane	RE(s)
1	*BamHI* + *Bgl*II
2	*BamHI* + *Hin*dIII
3	*BamHI* + *Pst*I
4	*BamHI* + *Xba*I
5	*Bgl*II + *Hin*dIII
6	*Bgl*II + *Pst*I
7	*Bgl*II + *Xba*I
8	*Hin*dIII + *Pst*I
9	*Hin*dIII + *Xba*I
10	*Pst*I + *Xba*I

Fig. 6.7(b): An autoradiograph derived from the gel in (a). The Southern transfer membrane was probed at high stringency with a multiprime probe derived from NK702. Sequences in BaEV which hybridised to the NK702 probe were located to the right of the *Xho*I site, with the sequence orientated as in Fig. 6.6. A 0.9 kbp. *Xba*I-*Bam*HI fragment in Lane 4 was the smallest fragment to hybridise strongly to the probe. This fragment was cloned into pBluescript for sequencing.

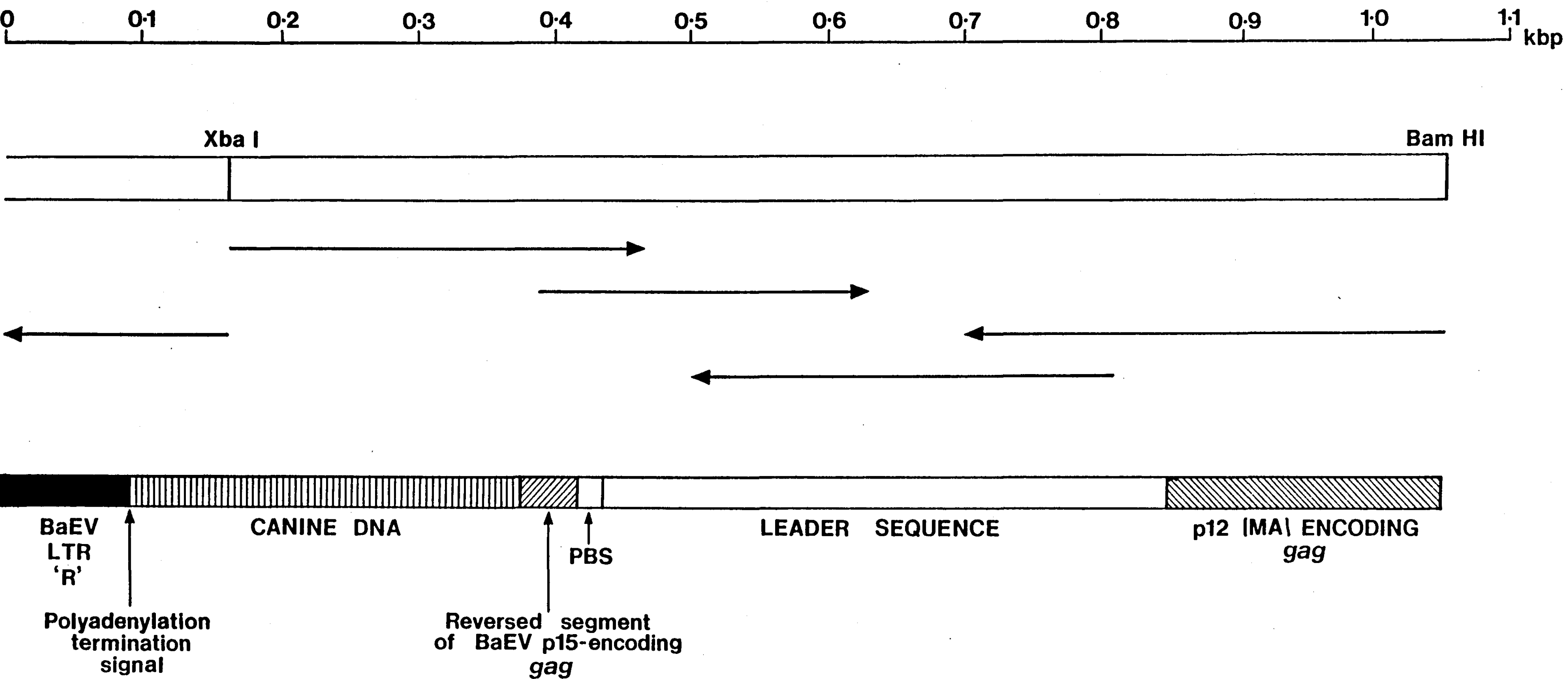


Fig. 6.8: Summary diagram to show sub-cloning and sequencing strategy for part of BaEV. The top of the diagram shows the *Xba*I and *Bam*HI sites used in cloning. The 901 bp. *Xba*I-*Bam*HI fragment was sequenced by use of the universal primers and two custom-synthesised oligonucleotide primers. Leftward pointing arrows indicate the extent of individual DNA sequencing reactions on the upper DNA strand. Rightward-pointing arrows indicate the same for the lower DNA strand. The diagram at the bottom shows the interruption of the BaEV LTR by 286 bp. of DNA 100% homologous to a segment of non-coding canine DNA. Downstream of this; and immediately adjacent to the PBS is 44 bp. of reversed BaEV *gag*.

a).

```
   1  TAACGCTATA AAAGAGATGT ATCGCCACAA ATCGGGGCTC TCTCTCCCTC ATTCTTTCCA GGTGGGAGGG CCCTGGTGCA CCAGTAAACG ACTTGAAAAT

 101  GGGAGTTCTG GGAAGATTAC CCAGCCTGTG AGATCACACA TGATCAAAGA GACAGAGTCT AGAAGCCAGA TGAAATTGGG TGGCCATTTC AGGAACCCGC

 201  ACCTGAGTTG AGAAGAGCTT GGGAATCTTC TTCCTGATTG GCCTGAAGCT CTAAGTCGCT CATTTGGAAC TATTCAAAAC TGGTTGAAGT GTCAATACAG

 301  CTTCAAAGAG GACTTCTAAT AATATGATGT AAAGCTTTCT CTGATTGCAG ATAAGGACTG ACCCATTAAA GGAAGTGGTC TTGGAAGAGG AGGTCTCGCT

 401  GCGACTCATC TGGGAGTACG TGGGGGCTCG TCGGGATTTG AGAGGCGGCC AGAGGACACC GGCCCCTTTT CCTTTTCGGC AGAAACCGCG CGGCCCGGCC

 501  ACCGGTGGCG GACGGACGAC AGGACGACTC TTGTGTGTAG TCAGGTACTT TACTTTTGTT TTGTCTTTAA TCTCTGAGGT CGGCCACCCT CCGTAGGAGT

 601  GTAGAGGGAG GACAGACGTG TCCTGGACCC TCACACTCCG ACCCCGGGGG ACGCCCCGGC GGTCGTCTGG AGGAAGGCTG ATGACACCGT CAGCCTCCTC

 701  AAATCTGAAG GCAGGTTCCC CCTGCCATCT GAATCACTTG TAGTACTTTG GCGCCATTCT CTGGCCGCGC GGCTCCTCTG TTTTTGTCTG GTTTGTGTTG

 801  TTACTGTTGT CTTATTTATA TGTGTGTTGA GCCTAAGGAC GGGACAATGG GACAGACGCT AACAACTCCT CTATCTTTGA CTTTGACACA CTTTTCAGAC

 901  GTCCGGCCAG AGCCCACAAT CTTTCCGTAG GAGTCCGAAA AGGACGATGG CAAACCTTTC TGCTCGTCCG AGTGGCCCAC CCTTCATGTC GGGTGGCCCC

1001  GGGACGGAAC TTTTGACCTC TCCGTTATTT TGCAGGTTAA GACGAAGGTA ATGGATCC
```

b).

```
 94 TGAAAATGGGAGTTCTGGGAAGATTACCCAGCCTGTGAGATCACACATGATCAAAGAGACAGAGTCTAGAAGCCAGATGAAATTGGGTGGCCATTTCAGG
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
766 TGAAAATGGGAGTTCTGGGAAGATTACCCAGCCTGTGAGATCACACATGATCAAAGAGACAGAGTCTAGAAGCCAGATGAAATTGGGTGGCCATTTCAGG

194 AACCCGCACCTGAGTTGAGAAGAGCTTGGGAATCTTCTTCCTGATTGGCCTGAAGCTCTAAGTCGCTCATTTGGAACTATTCAAAACTGGTTGAAGTGTC
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
666 AACCCGCACCTGAGTTGAGAAGAGCTTGGGAATCTTCTTCCTGATTGGCCTGAAGCTCTAAGTCGCTCATTTGGAACTATTCAAAACTGGTTGAAGTGTC

294 AATACAGCTTCAAAGAGGACTTCTAATAATATGATGTAAAGCTTTCTCTGATTGCAGATAAGGACTGACCCATTAAAGGAAGTGGT 379
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
566 AATACAGCTTCAAAGAGGACTTCTAATAATATGATGTAAAGCTTTCTCTGATTGCAGATAAGGACTGACCCATTAAAGGAAGTGGT 481
```

Fig. 6.9: **a).** 1058 bp. of nucleotide sequence from the LTR, leader sequence and 5' *gag* of the BaEV clone under study. The sequence shown encompasses the regions shown in the summary diagram in Fig. 6.8. Canine DNA begins 94 bp. into the sequence.

b). Alignment of BaEV with NK986. BaEV is shown above NK986 in this alignment. 286 bp. are identical between the two sequences. These regions of identity are arranged in opposite orientations in the two sequences.

gag up to the end of available sequence data were 96.6% homologous to, and co-linear with a published BaEV *gag-pol* sequence (Tamura, 1983). 211 bp. of the 215 bp. at the *Sal*I end of the *Sal*I/*Xba*I fragment was found to be 93.4% homologous to BaEV *env* (sequence as determined by Kato *et al.*, 1987).

The parts of the BaEV LTR which were sequenced included the TATA box, capping site and polyadenylation signal. The LTR R component was interrupted by canine DNA very close to its 3' end, in the middle of the termination signal sequence. U_5 was absent. The 286 bp. of interspersed canine DNA was orientated in the opposite direction to NK986. It included the short DNA segment homologous to the human beta globin region on chromosome 11. Between the interspersed canine DNA and the PBS was 44 bp. of BaEV p15-encoding *gag*, in the opposite orientation to the LTR and other downstream retroviral sequences. This reversed segment of *gag* ended precisely at the beginning of the PBS, which contained a single deletion relative to the 21 bp. $tRNA^{pro}$ PBS reported by Tamura (1983).

To investigate whether the abnormal juxtaposition of sequences in the LTR as described in the previous paragraph was produced by homologous recombinational events, adjacent and overlapping junctional sequences of the three component elements were compared. BaEV LTR R:U_5 (positions 420 to 520) was compared with NK986 (lower strand, positions 720 to 800). NK986 (positions 450 to 500) was compared with BaEV p15-encoding *gag* (positions 1370 to 1430). Lastly, BaEV *gag* (lower strand, positions 1330 to 1400) was compared with BaEV U_5:PBS (positions 520 to 580). There were no segments of high sequence homology at the junctions in the three comparisons. There was thus no evidence to suggest that homologous recombination had occurred in the LTR of our BaEV clone.

6.5 DISCUSSION

The finding of 286 bp. of canine DNA in our BaEV M7 clone came as a considerable surprise. After it was shown that NK986 was non-retroviral, it was anticipated that a segment of transduced cellular sequence, probably of baboon origin, would be found in the BaEV clone to explain its hybridisation to canine genomic DNA. It was certainly not predicted that the cellular sequence would be

canine. To explain how canine DNA became interspersed in our BaEV clone, it is helpful to assume that its predecessor was, at some time in the past, grown in canine cells. This is a reasonable assumption, since BaEV grows well in canine cells and there is a published report which describes the growth of BaEV M7 in a canine thymic cell line (Benveniste and Todaro, 1974a). Clearly, it is most unsatisfactory that the origin and history of our BaEV clone remain obscure. Under the circumstances it is only possible to discuss some of the results reported in this chapter conjecturally.

The detection of contaminating canine DNA and disordered *gag* in the LTR of our BaEV probe was a cause for considerable concern: a cloning artifact in our laboratory was suspected. However, careful consideration of laboratory procedures made it difficult to see how a cloning artifact could have arisen. During subcloning experiments, BaEV and canine DNA fragments were never run in the same preparative gels. The gel sets were always washed between experiments. Importantly, the unusual BaEV-canine DNA hybridisation pattern had been present from the very start of the experiments, before any BaEV subcloning was done. In any case, the arrangement of canine and disordered BaEV *gag* sequences immediately upstream of the PBS did not resemble a plasmid-associated cloning artifact. There were no restriction sites near the termini of the disordered or foreign DNA and it was considered most unlikely that a cloning artifact would co-incidentally involve a DNA segment terminating precisely adjacent to the PBS. It was thought far more likely that the arrangement of canine and disordered BaEV *gag* sequences in our BaEV clone was the consequence of one or more RT-mediated or other viral recombinational events which occurred while the virus was grown in canine tissue culture. Alternatively, the insertion of canine DNA might have occurred during initial cloning of the virus from the canine cell line. Comparison of LTR, NK986, p15-encoding *gag* and PBS junctional sequences failed to provide evidence for a homologous recombinational event to explain the rearrangements. However, homologous recombination may well have occurred to explain the loss of one of the LTRs, as predicted by the restriction map of our clone. The precise nature of the recombinational events which have occurred remains uncertain.

The intercalated canine DNA in our BaEV clone was almost certainly responsible for the single autoradiographic band found when every canine genomic DNA sample was probed at high stringency with the BaEV probe. It is likely that the foreign DNA segment in our BaEV clone was derived from a single copy element of the canine genome. Although part of the interspersed canine DNA resembled a human beta globin gene, it was either defective or it was not part of a protein-coding region since it contained multiple stop codons in all reading frames Examination of a larger segment of canine DNA (NK986) which encompassed the interspersed sequence did not provide further information as to its origin. No bacteriophage lambda DNA was found in the parts of our BaEV clone for which sequence data was obtained. If phage DNA had been found, it would have helped to explain the initial problem of high background plaque hybridisation to the BaEV probe and the resolution of this problem by inclusion of denatured *Hin*dIII-digested phage lambda DNA in the prehybridisation buffer. It is likely that bacteriophage lambda DNA is present as part of the BaEV insert, although hybridisation experiments were not done to confirm this. If phage lambda DNA were present, it would indicate that this BaEV clone has, at some previous time, been cloned in a lambda library. The intercalation of canine DNA in the BaEV LTR might have occurred as a consequence of recombination between phage genomes.

Some of the abnormalities in the restriction map of our BaEV clone can be explained in light of the sequence data. The abnormally-positioned *Sal*I site in our clone was generated by a point mutation, changing the sequence of published clones (5'-GTGGAC-3') to the *Sal*I recognition sequence (5'-GTCGAC-3'). The absence of a normally-positioned *Sal*I site was not explained because sequencing was not carried out in that region of the BaEV insert. The additional *Hin*dIII and *Xba*I sites were easily explained, since they were within the intercalated canine DNA. The additional *Bgl*II site and missing *Bam*HI sites were not explained.

The initial objective of the work described in this chapter was to characterise a BaEV-hybridising canine DNA fragment, with the hope of finding a novel, possibly non-defective canine endogenous retroviral element. Instead, contaminating canine DNA was found in the probe and it seems that canine genomic DNA does not, after all, contain sequences closely related to BaEV. This

work vividly demonstrates the importance of using only well-characterised probes in hybridisation experiments. Had it been realised that our BaEV clone was derived from a virus grown in canine cells, it would not have been used to probe canine DNA.

Chapter 7

GENERAL DISCUSSION

At the outset, the primary aim of this project was to identify and characterise canine endogenous retroviral elements by the application of molecular biological techniques. It was anticipated that a molecular biological approach would lead to the identification of endogenous retroviral elements in the canine genome because such elements have been found in the genomes of all vertebrates in which they have been actively sought. The presence of retroviral sequences in canine DNA was, in any case, predicted by previous liquid hybridisation studies (Bonner and Todaro, 1979; see Chapter 1, section 1.3.2). The intention, in seeking to elucidate the canine endogenous retroviral complement was to allow a better understanding of previous experimental findings which tended to indicate the existence of one or more canine retroviruses (Chapman *et al.*, 1967; Seman *et al.*, 1967; Rudolph, 1971; Onions, 1980; Tomley *et al.*, 1983; Sykes *et al.*, 1985). Had an infectious canine retrovirus been found, it would have been particularly important to investigate its aetiological significance.

The work described in this thesis has shown that canine DNA contains numerous elements which hybridise to established retroviral probes. The arrangement of these elements in all of the DNA samples examined was identical. Two of these endogenous retroviral elements (termed L3 and L4) were sub-cloned from the canine lymphoma cell line 3132 and their nucleotide sequence was determined. They were found to be closely-related, highly defective proviral elements; homologous to the murine leukaemia virus, AKV. The two elements had several distinctive defects in common. Most notably, they had identical, substantial *env* deletions. The presence of shared defects tended to indicate that one or both of the elements had been involved in a DNA amplification process. Since the 3' flanking cellular DNA of the two elements was similar, it was thought that the amplified unit might include the provirus and some flanking sequence.

Further evidence that canine endogenous proviral elements have been involved in a DNA amplification process came from Southern analysis of canine genomic DNAs probed with canine retroviral *pol* probes. Several of the larger autoradiographic bands were markedly recruited. Although this finding was reminiscent of that seen when identical internal fragments are released from separately-integrated proviruses; the large size of some of the recruited fragments

precluded this explanation. Recruited fragments larger than 10 kbp. could not be generated by internal cleavage of intact proviruses. Rather, it seems that amplification of DNA segments consisting of proviral and flanking cellular DNA sequences must have occurred to explain these recruited bands. The fact that several species of canid, including the domestic dog, have an apparently identical arrangement of proviral elements and a similar pattern of recruitment suggests that these are ancient sequences.

Northern analysis of a variety of canine tissues provided little evidence to suggest that these retroviral elements are expressed as RNA. Neither was it possible to induce retroviral expression from three canine cell lines using established techniques. It is perhaps not surprising that RNA expression of sequences closely related to L3 and L4 was not detected, given the marked defectiveness of these sequences.

The findings described in this thesis closely parallel those of other workers involved in the study of multicopy families of endogenous retroviral elements of other species. The genomic DNA of several mammalian species has been shown to contain numerous copies of apparently ancient retroviral elements; some of which were discussed in Chapters 3 and 4. The genomic arrangement of these ancient elements is apparently identical in all members of the host species. Other closely-related species have a very similar arrangement of the elements. There is evidence that amplification and chromosomal dispersion of endogenous retroviral sequences has occurred in all species which have been intensively studied; including baboons (Battula *et al.*, 1982), cats (Soe *et al.*, 1983), chimpanzees (Steele *et al.*, 1986), humans (Larsson *et al.*, 1989; Steele *et al.*, 1986) and mice (Itin and Keshet, 1986; Obata and Khan, 1988). Amplification and chromosomal dispersion of endogenous retroviral elements may therefore be a universal feature of vertebrate genomic DNA. Further aspects of DNA amplification, as it relates to endogenous retroviral elements, will be considered later in the discussion.

Whether the findings described in this thesis have any direct bearing upon previous experimental work concerning retroviral expression in *canine* tissues is uncertain. Previous experimental findings concerning canine retroviral expression

were described in Chapter 1, section 1.3.2. It is unquestionable that the highly defective elements sequenced in this study could not be responsible for production of virus particles or RT activity. The question is whether or not it is reasonable to assume that there are more intact relatives of L3 and L4 in the canine genome which might be capable of directing expression of retroviral particles or RT. Superficially, such an assumption seems reasonable, given the large number of L3/L4-related proviral elements found in the canine genome. L3 and L4 differ considerably in their *pol* defects and it is unlikely that these two elements, selected at random, happen by chance to be the most intact members of the family. However, L3 and L4 are members of an ancient endogenous retroviral family, as revealed by the study of wild canid DNAs described in Chapter 4, Appendix 1. Unless there are selective pressures tending to maintain in an intact state certain members of this family, but not others; it is unlikely that functional proviruses of this family would have survived to the present day. It is difficult to envision a selection pressure which would apply to one or more members of the family, but not to others. One possibility is that the presence of one or a very few intact proviruses confers a selective advantage upon the host, but that no further advantage is gained by the presence of several more. Members of a multicopy family of proviruses could be disabled by mutation with no deleterious effect until very few intact elements remained; then a selective pressure would begin to apply, tending to maintain the remaining element or elements in an intact state. No examples of this form of differential selection could be found in the literature.

A consideration which is relevant to the previous experimental findings concerning retroviral expression in canine tissues is that canine genomic DNA, like that of the human (Larsson *et al.*, 1989); may contain several distinct endogenous retroviral families. Some families may include complete, replication-competent elements. The presence of such putative elements could explain the retroviral expression which has been demonstrated in various canine neoplastic tissues by EM, RT assay and the detection of RNA-containing particles of oncoviral density (Chapman *et al.*, 1967; Seman *et al.*, 1967; Rudolph, 1971; Onions, 1980; Tomley *et al.*, 1983; Sykes *et al.*, 1985). The significance of retroviral expression in *neoplastic* tissues is moot. Since retroviruses are known to have oncogenic potential in other species, it has been suggested that retroviral expression in canine

tumours might be of aetiological significance. However, attempts to transmit such a putative aetiologically-significant canine retrovirus have been largely unsuccessful (Madewell and Theilen, 1987). Particles of retroviral density and RT activity have also been found in normal canine tissues and secretions (Tomley *et al.*, 1983; Littlewood *et al.*, 1984). It is possible that retroviral expression in canine neoplastic tissues reflects expression of endogenous retroviral sequences as a paraphenomenon, rather than the presence of an aetiologically-significant virus. Cellular regulatory mechanisms which control gene expression may be sufficiently disrupted by neoplastic transformation to allow expression of proviral elements which are normally transcriptionally silent or which produce mRNA that usually remains untranslated.

Although the presence in canine DNA of multiple endogenous retroviral elements may explain some of the phenomena previously attributed to a putative infectious virus, this study in no way rules out the possibility that a horizontally-transmissible, aetiologically-significant, exogenous canine retrovirus will be discovered in the future. However, the considerable number of endogenous retroviral elements present in the canine genome necessitate great caution when ascribing aetiological significance to any putative canine oncovirus. To prove an oncogenic role for a canine retrovirus will require fulfilment of Koch's second, third and fourth postulates[1]. Fulfilment of Koch's first postulate is unnecessary, since it is generally accepted that there are many non-infectious causes of neoplastic transformation.

While discussing putative canine oncoviruses, it is of interest briefly to ponder the significance of the infectious canine retrovirus reported to have been released by the cell line 3132 (Strandstrom and Rimaila-Parnanen, 1979). This is of particular interest, because L3 and L4 were sub-cloned from 3132 genomic DNA. Strandstrom and Rimaila-Parnanen (1979) described the release of retroviral particles by the canine lymphoma cell line, 3132. Their discovery

---

1 **Koch's postulates** define the steps required to establish a microorganism as the cause of a disease: (1) it must be found in all cases of the disease; (2) it must be isolated from the host and grown in pure culture; (3) it must reproduce the disease when introduced into a susceptible host; and (4) it must be found to be present in the experimental host so infected.

became considerably more exciting when it was shown that the virus was infectious for canine and human cells (Strandstrom and Bowen, 1982). Unfortunately, subsequent, repeated efforts to induce expression of a retrovirus from 3132 cells have been unsuccessful (D.E. Onions, personal communication and personal observations). The significance of this virus has not been satisfactorily elucidated. Since the virus was associated with a lymphoproliferative disorder; it has been implied that the 3132 virus might have been a canine oncovirus. However, Koch's third postulate was never fulfiled for this virus, so this possibility remains unproven. A second possibility is that the 3132 virus was a contaminating virus from another species. Thirdly, it is possible that the 3132 virus represented a transiently-expressed, infectious, endogenous retrovirus.

The work described in Chapter 4 showed no apparent differences in the arrangement of retroviral proviral elements in 3132 genomic DNA compared with any of the other canine DNA samples. This was equally true for the L4 *pol* probes and for BaEV, FMuLV, HTLV-1 and MMTV *gag-pol* probes applied at high and low stringency. If the cell line had been subjected to repeated rounds of retroviral infection, one might have expected to find evidence of recently-integrated proviruses in the various Southern analyses. No such evidence was found. Neither was it possible to detect RNA expression of a retrovirus using the same probes. Retroviral RNA expression was not induced by treatment of 3132 cells with 5-AZA or IdU.

One possible explanation for the failure to detect evidence of recent proviral integration events in 3132 DNA is that the retrovirus under investigation was insufficiently homologous to the applied probes to produce detectable hybridisation at the stringencies used. However, a variety of retroviral probes were used at both high and low stringencies. Strandstrom and Bowen (1982) showed that the 3132 virus contained a major protein antigenically-related to BaEV. Although antigenic-relatedness does not correlate directly with nucleotide sequence homology, the two are often related. It is therefore noteworthy that the BaEV probe did not detect any differences, by Northern or Southern analyses, between 3132 and the other canine samples. Clearly, the results of this part of the study were compromised by the presence of a short segment of contaminating canine cellular DNA within our BaEV probe. Despite this, it is likely that if

distinctive, BaEV-related retroviral sequences were present within 3132 genomic DNA and absent from other canine DNA samples, the difference would have been detected.

Persisting with the assumption that recently-integrated, distinctive retroviral proviruses were present in the 3132 genomic DNA under study; then a second explanation for the failure to detect them is that they might have been obscured by high background hybridisation, particularly when the L4 *pol*-derived probes were used. With so many hybridising bands, it would have been easy to miss a few additional proviral elements. An obvious third explanation for the failure to detect recently-integrated proviruses is that they were not present in the particular 3132 cells under study. These cells might never have been infected with the virus. Perhaps infection had a deleterious effect upon the host cells, so that infected cells tended to die out.

Failure to detect or induce mRNA retroviral expression by 3132 cells was consistent with the fact that this cell line no longer seems to express retroviral particles or RT activity (Onions, personal communication). In the absence of a specific 3132 retroviral probe, it will be difficult to pursue this elusive virus further. The work described in this thesis unfortunately provides no specific information as to its origin or significance.

Having reviewed the aims and principal findings of this work and shown how it relates to previous studies; it is worth considering the significance of what has been found. What is the significance of the discovery of highly defective canine endogenous retroviral elements, in the absence of an infectious virus? Has this study been no more than an exercise in genetic palaeontology?

It is probably true that the L3/L4 family of elements are amplified relics of a former age. Their intact, replication-competent ancestors may have been conventional, horizontally-transmissible viruses. Alternatively, the ancestral virus may have entered the canid germ line in the same way that RD 114 is thought to have entered the felid line (Benveniste and Todaro, 1974b) and viruses of this family may never have been horizontally-transmitted between canids. Ultimately, it may not be possible to deduce the origin or history of this ancient viral family from the remnants which have survived to the present day. The relatively high

homology of L3 and L4 to a murine virus (AKV) suggests that ancestors of the L4 family may have been transmitted to ancestral canids from rodents, in the same way that FeLV is thought to have been transmitted to felids (Teich, 1982). The fact that the PBS of L3 was complementary to $tRNA^{pro}(2)$ puts this virus family squarely in the middle of the modern type-C oncoviruses. A question which will probably remain unanswered, is whether the putative, infectious ancestral viruses were ever of aetiological significance in canid diseases.

This work has extended that of others, in showing that the genome of yet another mammalian species is littered with numerous defective retroviral proviruses. These sequences are evidence of the considerable role played by retroviruses and retrovirus-like elements in the continuous and dynamic remodelling of the eukaryotic genome. A striking feature of the sequences characterised in this study, relevant to genome remodelling; is their involvement in a DNA amplification process. It is evident that after becoming highly defective, members of the L3/L4 family of retroviral elements have been considerably amplified in canine genomic DNA. The almost identical defects in L3 and L4, together with the high 3' flanking nucleotide sequence homology attest to this. The dramatic simplification of the pattern of hybridisation to an L4 *pol* probe obtained when genomic DNA was double-digested with two L4 internally-cleaving REs provides further compelling evidence for the relatedness of the L4 family of elements and suggests a previous amplification process. Yet the amplification process seems presently to be proceeding at an extremely slow rate, if at all. If amplification were continuing at a substantial rate, one would expect to find differences in proviral arrangement between disparate canine breeds. Yet there were no apparent differences in the proviral complement among 36 genomic DNA samples from 24 dogs. A canid species thought to have diverged from the domestic dog 9 million years ago (*Alopex lagopus*; Wayne and O'Brien, 1987) had a similar pattern of hybridising bands, both in terms of fragment size and degree of recruitment, to that of the domestic dog. These findings parallel those of Steele *et al.* (1984, 1986), who showed that HERV-E sequences and associated flanking DNA were amplified in the human genome and that five individual humans had an apparently identical arrangement of proviruses. They also showed that chimpanzees have a similar array of amplified DNA and conjectured that

amplification occurred before evolutionary divergence of humans and chimpanzees and had since ceased.

It is intriguing to consider the mechanism by which a *transient* period of amplification might have occurred. If progeny amplified copies retained the ability to transpose, one would predict an exponential increase in copy number. It seems much more likely that the many copies of L3- and L4-related elements reflect proliferation of a very few ancestral elements. Subsequent disablement of these very few "transposition-competent" elements could then be invoked as an explanation for the transience of the amplification process. A theory, similar in part, has recently been proposed to explain the concerted evolution of LINE elements. It has been suggested that most murine LINE elements are derived from duplicative transposition of one to three active LINE sequences or "molecular drivers" (Dover, 1982; Martin *et al.*, 1985; Hardies *et al.*, 1986). These molecular drivers would seem to have much in common with the postulated "transposition-competent" retroviral elements mentioned above. However, in contrast to the situation with HERV-E and, perhaps, the L4 family of elements; there is evidence that one or more of the murine LINE molecular drivers has been active in the recent past (Bellis *et al.*, 1987).

Amplification of endogenous retroviral sequences together with segments of flanking cellular DNA has been reported in mammalian species other than primates and dogs and may be a general phenomenon. Copies of endogenous FeLV in feline genomic DNA were found to have conserved restriction sites known to be in flanking cellular DNA (Soe *et al.*, 1983). In Chinese hamster cells, recently-integrated Rous sarcoma proviral DNA and flanking cellular sequences underwent both amplification and dispersion to different chromosomes (Hillova *et al.*, 1983). There is strong evidence to suggest that members of the murine GLN family of ancient proviral elements have undergone a similar amplification process; although, formally, it has not been shown that individual elements have identical defects, or share flanking cellular sequences (Itin and Keshet, 1986; Obata and Khan, 1988). The mechanism by which amplification of proviral and flanking cellular sequences occurs is unknown. The clustering of chicken *ev* proviruses on chromosome 1 prompted the suggestion that replication might involve an amplification process distinct from reverse transcription and random

chromosomal integration (Tereba, 1983). It is possible that mechanisms similar to those which amplify the copy number of proto-oncogenes or drug resistance genes may have a role in the production of such clusters of proviral elements. However, the majority of proviruses appear to be dispersed among the chromosomes. Moloney MuLV proviral DNA, in which the *env* gene was replaced with a selectable marker, was recently shown to undergo occasional intracellular retrotransposition (Heidmann *et al.*, 1988).

Conventional retrotransposition could certainly explain proviral amplification, but does not explain why flanking cellular sequences are included in many of the amplified units. Perhaps transcription of the progenitors of these amplified provirus/flanking DNA composite sequences is initiated by a strong RNA polymerase II promoter in cellular DNA external to the provirus, rather than at the usual U_3-R junction. This promoter would have to allow strong germline expression if heritable sequence amplification were to occur. Transcription might then proceed through the provirus (on the plus or minus DNA strand) and into 3' flanking sequence until the next termination and polyadenylation signal was reached. Reverse transcription and DNA integration of the composite transcript might subsequently occur, in the manner of a LINE-type retroposon. Assuming transcription by RNA polymerase II, the "parental" promoter would be external to the transcribed sequence and would not be carried over to progeny sequences. These would therefore not be expected to inherit transcriptional activity. If this mechanism were correct, one would expect to find a poly-A tract on one strand or the other of flanking DNA some distance from the amplified proviral sequence. No such tract was found for L3 or L4; but this might merely be because an inadequate length of flanking DNA was sequenced. Alternatively, other poorly understood mechanisms might be responsible for endogenous proviral amplification. For example, a possible role for SINE sequences in DNA amplification was discussed in Chapter 3. Whatever the mechanism, the process of endogenous retroviral DNA amplification and dispersion is likely to be of biological significance. Retrotransposon-induced insertional mutagenesis may have a deleterious or fatal effect in individual animals. However, by keeping the genome in a state of flux, DNA amplification and dispersion may have the

advantageous effect of enhancing genome plasticity and thereby speeding the evolution of a species.

Given that the L4 family of elements do not appear to have recently been involved in a DNA amplification process, it is likely that their biological significance, if any, would depend upon mRNA or protein expression. No evidence was found for RNA expression, and the frequent stop codons in all reading frames of L3 and L4 preclude expression of all but very small proteins from these particular elements. Of course, it is likely that elements more intact than L3 and L4 exist, but evidence for their expression was not found in this study. As discussed in Chapter 5, failure to detect retroviral expression , despite the use of a variety of retroviral probes may have resulted from selection of inappropriate target tissues. Future studies should cast light on this issue.

An unexpected finding of this work was the presence of nine tandem copies of a 25 bp. minisatellite downstream of the PBS in L3. The possible value of this finding was discussed in Chapter 3. In collaboration with Dr R.K. Wayne of the Department of Biology, University of California; this minisatellite array will be used to probe genomic DNA samples from members of small populations of genetically-isolated wild canid species. It is anticipated that the resulting restriction fragment profiles will allow estimation of relative genetic variability and reconstruction of the evolutionary relationships of these natural populations (for a description of similar previous work, see Gilbert *et al.*, 1990). It may also be possible to develop from the minisatellite array a probe which will allow the resolution of canine paternity disputes. An exciting possibility is that a probe developed from the canine minisatellite array might detect novel, hypervariable sequences in the human genome. However, preliminary results show that a probe encompassing the two canine minisatellite arrays in L3 does not detect sequences in human genomic DNA.

It has been suggested that retrotransposons are very ancient elements and may have been part of the genomes of some of the earliest DNA life forms (Doolittle *et al.*, 1989). As such, they might be viewed as the earliest parasites; or, perhaps, as sequences which played a vital role in early stages of the evolution of

life. Retroviruses are thought to have evolved from mobile genetic elements, such as retrotransposons (Shimotohno *et al.*, 1980; Temin, 1980; Doolittle *et al.*, 1989). The processes which led to the acquisition of the retroviral DNA sequences essential for extracellular existence must have been remarkably complex, but may have occurred over a very long time period. Eventually, descendents of these primordial retroviruses have repopulated vertebrate genomes by infection of germ line cells, to the extent that the genomes of all vertebrate species probably harbour hundreds to thousands of proviral elements. What is more, the process of endogenous retroviral assimilation is continuing. Indeed, there is likely to be a continual temporal flux of proviruses through any given vertebrate germ line; with new, intact proviruses replacing heavily-defective elements as they are deleted. This dynamic remodelling process is further complicated by the intriguing and prominent influence that DNA amplification seems to have had upon the distribution of endogenous retroviral elements. The experimental findings described in this thesis provide but a glance at a snapshot of this intricate process.

There has been much discussion concerning possible beneficial functions provided to the host organism by the intact endogenous retroviruses (ERVs) incorporated in its genome. Chief among these is the possibility that postembryonic ERV expression might sensitize the host's immune system to antigens shared by the ERV and a potential retroviral pathogen, thus providing protective immunity. Benveniste and Todaro (1976) suggested that the preservation for more than 30 million years of BaEV-related proviruses in the genomic DNA of anthropoid apes indicates that these viral DNA sequences confer a selective advantage upon the host. It is intriguing to ponder one possible implication of this argument: that one or more exogenous retroviral pathogens related to BaEV might have lived in uneasy equilibrium with anthropoid apes for more than 30 million years. It is quite a different question to ask whether the substantial amount of amplified *defective* retroviral DNA present in the vertebrate genome confers any selective advantage upon the host beyond the rather pedestrian provision of supplementary genetic material. Currently, the weight of evidence would tend to suggest that these ubiquitous sequences are non-essential and of no aetiological significance in vertebrate diseases.

REFERENCES

Anderson, G.R., Stoler, D.L., Scott, J.P. and Farkas, B.K. (1988) Induction of VL30 expression as a response to anoxic stress. pp. 265-274 In: Banbury Report 30: Eukaryotic elements as mutagenic agents. Cold Spring Harbor Laboratory, New York.

Andrewes, C.H. (1939) Latent virus infections and their possible relevance to the cancer problem. *Proc. R. Soc. Med.* **33**: 75-86.

Astrin, S.M. (1978) Endogenous viral genes of the White Leghorn chicken: Common site of residence and sites associated with specific phenotypes of viral gene expression. *Proc. Natl. Acad. Sci. USA.* **75**: 5941-5945.

Astrin, S.M., Buss, E.G. and Hayward, W.S. (1979) Endogenous viral genes are non-essential in the chicken. *Nature* **282**: 339-341.

Bachmann, B.J. (1972) Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol. Rev.* **36**: 525-557.

Battula, N., Hager, G.L. and Todaro, G.J. (1982) Organization of type C viral DNA sequences endogenous to baboons: analysis with cloned viral DNA. *J. Virol.* **41**: 583-592.

Bellis, M., Jubier-Maurin, V., Dod, B., Vahlerbergne, F., Laurent, A., Senglat, C., Bonhomme, F. and Roizes, G. (1987) Distribution of two recently inserted long interspersed elements of the L1 repeat family at the Alb and beta h 3 loci in wild mice. *Mol. Biol. Evol.* **4**: 351-363.

Bentvelzen, P., Daams, J.H., Hageman, P. and Calafat, J. (1970) Genetic transmission of viruses that incite mammary tumor in mice. *Proc. Natl. Acad. Sci. USA* **67**: 377-384.

Benveniste, R.E. and Todaro, G.J. (1974a) Multiple divergent copies of endogenous C-type virogenes in mammalian cells. *Nature* **252**: 170-173.

Benveniste, R.E. and Todaro, G.J. (1974b) Evolution of C-type viral genes: Inheritance of exogenously acquired viral genes. *Nature* **252**: 456-459.

Benveniste, R.E. and Todaro, G.J. (1976) Evolution of type C viral genes: evidence for an Asian origin of man. *Nature* **261**: 101-108.

Bernhard, W. (1958) Electron microscopy of tumor cells and tumor viruses. A review. *Cancer Research* **18**: 491-509.

Bernhard, W. (1960) The detection and study of tumor viruses with the electron microscope. *Cancer Research* **20**: 712-727.

Bernhard, W. and Bauer, A. (1955) Mise en evidence de corpuscules d'aspect virusal dans des tumeurs mammaires de la souris. Etude au microscope electronique. *C. R. Acad. Sci.* **240**: 1380-1382.

Berry, B.T., Ghosh, A.K., Kumar, V., Spodick, D.A. and Roy-Burman, P. (1988) Structure and function of endogenous feline leukemia virus long terminal repeats and adjoining regions. *J. Virol.* **62**: 3631-3641.

Binn, L.N. (1980) Establishment of a canine cell line: derivation, characterization and viral spectrum. *Amer. J. Vet. Res.* **41**: 855-860.

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

Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985) Ty elements transpose through an RNA intermediate. *Cell* **40**: 491-500.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**: 95-113.

Bolton, E.T. & McCarthy, B.J. (1962) A general method for the isolation of RNA complementary to DNA. *Proc. Natl. Acad. Sci. USA* **48**: 1390-1397.

Bonner, T.I., Brenner, D.J., Neufeld, B.R. and Britten, R.J. (1973) Reduction in the rate of DNA reassociation by sequence divergence. *J. Mol. Biol.* **81**: 123-135.

Bonner, T.I., O'Connell, C and Cohen, M. (1982) Cloned endogenous retroviral sequences from human DNA. *Proc. Natl. Acad. Sci. USA.* **79**: 4709-4713.

Bonner, T.I. and Todaro, G.J. (1979) Carnivores have sequences in their DNA distantly related to the primate endogenous virus, MAC-1. *Virology* **94**: 224-227.

Bouche, J.P. (1981) The effect of spermidine on endonuclease inhibition by agarose contaminants. *Analytical Biochem.* **115**: 42-45.

Britten, R.J. and Davidson, E.H. (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Quart. Rev. Biol.* **46**: 111-138.

Brundage-Anguish, L.J. and Quimby, F.W. (submitted for publication) Identification of LQ-1 binding protein as canine Thy-1. *Devel. Compar. Immunol.*

Callahan, R. (1988) Two families of human endogenous retroviral genomes. pp. 91-100 In: Banbury Report 30: Eukaryotic elements as mutagenic agents. Cold Spring Harbor Laboratory, New York.

Callahan, R., Chiu, I.M., Wong, J.F.H., Tronick, S.R., Roe, B., Aaronson, S.A. and Schlom, J. (1985) A new class of endogenous human retroviral genomes. *Science* **228**: 1208-1211.

Callahan, R., Drohan, W., Tronick, S. and Schlom, J. (1982) Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. *Proc. Natl. Acad. Sci. USA.* **79**: 5503-5507.

Chapman, A.L., Bopp, W.J., Brightwell, A.S., Cohen, H., Nielson, A.H., Gravelle, C.R. and Werder, A.A. (1967) Preliminary report on virus-like particles in canine leukemia and derived cell cultures. *Cancer Res.* **27**: 18-25.

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

Chen, E.Y. and Seeberg, P.H. (1985) Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* **4**: 165-170.

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochem.* **162**: 156-159.

Cianciolo, G.J., Copeland, T.D., Oroszlan, S. and Snyderman, R. (1985) Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* **230**: 453-455.

Coffin, J. (1982a) Structure of the retroviral genome. pp. 261-368 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Text. Cold Spring Harbour Laboratory, New York.

Coffin, J. (1982b) Endogenous viruses. pp. 1109-1204 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Text. Cold Spring Harbour Laboratory, New York.

Coffin, J. (1985) Genome structure. pp. 17-74 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Supplement. Cold Spring Harbour Laboratory, New York.

Cohen, J.C. and Varmus, H.E. (1979) Endogenous mammary tumor virus DNA varies among wild mice and segregates during inbreeding. *Nature* **278**: 418-423.

Cohen, M., Nicolson, M.O., McAllister, R.M., Shure, M., Davidson, N., Rice, N. and Gilden, R.V. (1980) Baboon endogenous virus genome. 1. Restriction enzyme map of the unintegrated DNA genome of a primate retrovirus. *J. Virol.* **34**: 28-39.

Cohen, M., Rein, A., Stephens, R.M., O'Connell, C., Gilden, R.V., Shure, M., Nicolson, M.O., McAllister, R.M. and Davidson, N. (1981) Baboon endogenous virus genome: molecular cloning and structural characterisation of nondefective viral genomes from DNA of a baboon cell strain. *Proc. Natl. Acad. Sci. USA.* **78**: 5207-5211.

Cooper, G.M. and Silverman, L. (1978) Linkage of the endogenous avian leukosis virus genome of virus-producing chicken cells to inhibitory cellular DNA sequences. *Cell* **15**: 573-577.

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

Crittenden, L.B., Fadly, A.M. and Smith, E.J. (1982) Effect of endogenous leukosis virus genes on response to infection with avian leukosis and reticuloendotheliosis viruses. *Avian Dis.* **26**: 279-294.

Cullen, B.R., Lomedico, P.T. and Ju, G. (1984) Transcriptional interference in avian retroviruses: implications for the promoter insertion model of leukemogenesis. *Nature* **307**: 241-244.

Davis, B.R., Brightman, B.K., Chandy, K.G. and Fan, H. (1987) Characterization of a preleukemic state induced by Moloney murine leukemia virus: Evidence for two infection events during leukemogenesis. *Proc. Natl. Acad. Sci. USA* **84**: 4875-4879.

Dalgleish, A.G., Beverley, P.C.L., Clapham, P.R., Crawford, D.H., Greaves, M.F. and Weiss, R.A. (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**: 763-767.

Darlington, C.D. (1948) The plasmagene theory of the origin of cancer. *Br. J. Cancer* **2**: 118-126.

Deininger, P.L. (1989) SINEs: short interspersed repeated DNA elements in higher eukaryotes. pp. 619-636 In: D.E. Berg and M.M. Howe (editors), Mobile DNA. American Society for Microbiology, Washington, D.C.

De Rossi, A., Franchini, G., Aldovini, A., Del Mistro, A., Chieco-Bianci, L., Gallo, R.C. and Wong-Staal, F. (1986) Differential response to the cytopathic effects of human T-lymphotropic virus type III (HTLV-III) superinfection in T4$^+$ (helper) and T8$^+$ (suppressor) T-cell clones transformed by HTLV-I. *Proc. Natl. Acad. Sci. USA* **83**: 4297-4301.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**: 387-395.

Dickson, C., Eisenman, R., Fan, H., Hunter, E. and Teich, N. (1982) Protein biosynthesis and assembly. pp 513-648 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Text. Cold Spring Harbour Laboratory, New York.

Dickson, C., Eisenman, R. and Fan, H. (1985) Protein biosynthesis and assembly. pp 135-146 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Supplement. Cold Spring Harbour Laboratory, New York.

Doolittle, R.F., Feng, D.F., Johnson, M.S. and McClure, M.A. (1989) Origins and evolutionary relationships of retroviruses. *Quart. Rev. Biol.* **64**: 1-30.

Doolittle, R.F., Feng, D.F., McClure, M.A. and Johnson, M.S. (1990) Retrovirus phylogeny and evolution. *Curr. Top. Microbiol. Immunol.* **157**: 1-18.

Dougherty, R.M. and DiStefano, H.S. (1966) Lack of relationship between infection with avian leukosis virus and the presence of COFAL antigen in chick embryos. *Virology* **29**: 586-595.

Dover, G. (1982) Molecular drive: a cohesive mode of species evolution. *Nature* **299**: 111-117.

Etzerodt, M., Mikkelsen, T., Pedersen, F.S., Kjeldgaard, N.O. and Jorgensen, P. (1984) The nucleotide sequence of the Akv murine leukemia virus genome. Virology 134: 196-207.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochem.* **132**: 6-13.

Feinberg, A.P. and Vogelstein, B. (1984) Addendum: A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochem.* **137**: 266-267.

Finnegan, D.J. (1983) Retroviruses and transposable elements - which came first? *Nature* **302**: 105-106.

Fraser, C., Humphries, R. and Mager, D. (1988) Chromosomal distribution of the RTVL-H family of human endogenous retrovirus-like sequences. *Genomics* **2**: 280-287.

Frisby, D.P., Weiss, R.A., Roussel, M. and Stehelin, D. (1979) The distribution of endogenous retrovirus sequences in the DNA of galliform birds does not coincide with avian phylogenetic relationships. *Cell* **17**: 623-634.

Fulton, R., Forrest, D., McFarlane, R., Onions, D. and Neil, J.C. (1987) Retroviral transduction of T-cell antigen receptor Beta chain and *myc* genes. *Nature* **326**: 190-194.

Furano, A.V., Somerville, C.C., Tsichlis, P.N. and D'Ambrosio, E. (1986) Target sites for transposition of rat long interspersed repeated DNA elements (LINEs) are not random. *Nucl. Acids Res.* **14**: 3717-3727.

Furuichi, T., Inouye, S. and Inouye, M. (1987) Biosynthesis and structure of stable branched RNA covalently linked to the 5' end of multicopy single-stranded DNA of *Stigmatella aurantiaca*. *Cell* **48**: 55-62.

Gilbert, D.A., Lehman, N., O'Brien, S.J. and Wayne, R.K. (1990) Genetic fingerprinting reflects population differentiation in the Californian Channel Island fox. *Nature* **344**: 764-767.

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

Govidan, M.V., Spiess, E. and Majors, J. (1982) purified glucocorticoid receptor-hormone complex from rat liver cytosol binds specifically to cloned mouse mammary tumor virus long terminal repeats *in vitro*. *Proc. Natl. Acad. Sci. USA* **79**: 5157-5161.

Greenlee, P.G., Filippa, D.A., Quimby, F.W., Patnaik, A.K., Calvano, S.E., Matus, R.E., Kimmel, M., Hurvitz, A.I. and Lieberman, P.H. 1990 Lymphomas in dogs: A morphological, immunological and clinical study. *Cancer* **66**: 480-490.

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

Harada, F., Tsukada, N. and Kato, N. (1987) Isolation of three kinds of human endogenous retrovirus-like sequences using $tRNA^{pro}$ as a probe. *Nucl. Acids Res.* 15: 9153-9162.

Hardies, S.C., Martin, S.L., Voliva, C.F., Hutchison III, C.A. and Edgell, M.H. (1986) An analysis of replacement and synonymous changes in the rodent L1 repeat family. *Mol. Biol. Evol.* **3**: 109-125.

Hartley, J.W. and Rowe, W.P. (1976) Naturally-occurring murine leukemia viruses in wild mice: Characterization of a new "amphotropic" class. *J. Virol.* **19**: 19-25.

Heidmann, T., Heidmann, O. and Nicolas, J.-F. (1988) An indicator gene to demonstrate intracellular transposition of defective retroviruses. *Proc. Natl. Acad. Sci. USA.* **85**: 2219-2223.

Herman, S.A. and Coffin, J.M. (1986) Differential transcription from the long terminal repeats of integrated avian leukosis virus DNA. *J. Virol.* **60**: 497-505.

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

Hettle, S.J. (1985) PhD thesis, Glasgow University.

Hillova, J., Hill, M., Belehradek, J., Mariage-Samson, R. and Marx, M. (1983) Amplification of the provirus region in Rous sarcoma virus-transformed Chinese hamster cells and segregation of the amplified copies in somatic cell hybrids. *Virology* **128**: 1-15.

Ho, D.D., Pomerantz, R.J. and Kaplan, J.C. (1987) Pathogenesis of infection with human immunodeficiency virus. *New England J. of Med.* **317**: 278-286.

Hoyt, M.A., Knight, D.M., Das, A., Miller, H.I. and Echols, H. (1982) Control of phage lambda development by stability and synthesis of *c*II protein: Role of the viral *c*III and host *HflA*, *himA* and *himD* genes. *Cell* **31:** 565-573.

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

Holland, C.A., Hartley, J.W., Rowe, W.P. and Hopkins, N. (1985) At least four genes contribute to the leukemogenicity of murine retrovirus MCF 247 in AKR mice. *J. Virol.* **53**: 158-165.

Hutchison III, C.A., Hardies, S.C., Loeb, D.D., Shehee, W.R. and Edgell, M.H. (1989) pp. 593-617 In: D.E. Berg and M.M. Howe (editors), Mobile DNA. American Society for Microbiology, Washington, D.C.

Huynh, T.V., Young, R.A. and Davis R.W. (1985) Constructing and screening cDNA libraries in Lambda GT10 and Lambda GT11. pp. 49-78 In: D.M. Glover (Editor) DNA Cloning - A practical approach. Volume 1. IRL Press at Oxford University Press.

Hyrien, O., Debatisse, M., Buitin, G. and de Saint Vincent, B.R. (1987) A hotspot for novel amplification joints in a mosaic of *Alu*-like repeats and palindromic A+T-rich DNA. *EMBO J.* **7**: 407-417.

Inouye, S., Hsu, M,-Y, Eagle, S. and Inouye, M. (1989) Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. *Cell* **56**: 709-717.

Itin, A. and Keshet, E. (1986) A novel retroviruslike family in mouse DNA. *J. Virol.* **59**: 301-307.

Jarrett, O. (1989) Retroviridae. pp. 166-213 In: J.S. Porterfield (editor) Andrewes' Viruses of Vertebrates. Fifth Edition, Balliere Tindall.

Jarrett, O., Hardy, W.D., Golder, M.C. and Hay, D. (1978) The frequency of occurrence of feline leukaemia virus subgroups in cats. *Int. J. Cancer* **21**: 334-337.

Jarrett, O., Laird, H.M. and Hay, D. (1973) Determinants of the host range of feline leukaemia viruses. *J. Gen. Virol.* **20**: 169-175.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Hypervariable "minisatellite" regions in human DNA. *Nature* **314**: 67-73.

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

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

Johnson, M.S., McClure, M.A., Feng, D.-F., Gray, J. and Doolittle, R.F. (1986) Computer analysis of retroviral *pol* genes: Assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc. Natl. Acad. Sci. USA.* **83**: 7648-7652.

Jones, P.A. and Taylor, S.M. (1980) Cellular differentiation, cytidine analogs and DNA methylation. *Cell* **20**: 85-93.

Kabat, D. (1989) Molecular biology of Friend viral erythroleukemia. *Curr. Top. Microbiol. Immunol.* **148**: 1-42.

Kakuk, T.J., Hinz, R.W., Langham, R.F. and Conner, G.H. (1968) Experimental transmission of canine malignant lymphoma to a beagle neonate. *Cancer Res.* **28**: 716-723.

Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Novel bacteriophage lambda cloning vector. *Proc. Natl. Acad. Sci. USA* **77**: 5172-5176.

Karn, J., Matthes, H.W.D., Gait, M.J. and Brenner, S. (1984) A new selective phage cloning vector, lambda 2001, with sites for *Xba* I, *Bam*HI, *Hin*dIII, *Eco*RI, *Sst*I and *Xho*I. *Gene* **32**: 217-224.

Kato, S., Matsuo, K., Nishimura, N., Takahashi, N. and Takano, T. (1987) The entire nucleotide sequence of baboon endogenous virus DNA. *Jpn. J. Genet* **62**:127-137.

Kato, N., Pfeifer-Ohlsson, S., Kato, M., Larrson, E., Rydnert, J., Ohlsson, R. and Cohen, M. (1987) Tissue-specific expression of human provirus ERV3 mRNA in human placenta: two of the three ERV3 mRNAs contain human cellular sequences. *J. Virol.* **61**: 2182-2191.

Kato, N., Larrson, E. and Cohen, M. (1988) Absence of expression of a human endogenous retrovirus is correlated with choriocarcinoma. *Int. J. Cancer* **41**: 380-385.

Katzir, N., Rechavi, G., Cohen, J.B., Unger, T., Simoni, F., Segal, S., Cohen, D. and Givol, D. (1985) "Retroposon" insertion into the cellular oncogene c-myc in canine transmissible venereal tumor. *Proc. Natl. Acad. Sci. USA.* **82**: 1054-1058.

Keshet, E. and Temin, H.M. (1979) Cell killing by spleen necrosis virus is correlated with a transient accumulation of spleen necrosis virus DNA. *J. Virol.* **31**: 376-388.

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

Kiyokawa, T., Yamaguchi, K., Takeya, M., Takahashi, K., Watanabe, T., Matsumoto, T., Lee, S.Y. and Takatsuki, K. (1987) Hypercalcemia and osteoclast proliferation in adult T-cell leukemia. *Cancer* **59**: 1187-1191.

Kozak, C.A. and O'Neill, R.R. (1987) Diverse wild mouse origins of xenotropic, mink cell focus-forming and two types of ecotropic proviral genes. *J. Virol.* **61**: 3082-3088.

Kuff, E.L., Feenstra, A., Lueders, K., Smith, L., Hawley, R., Hozumi, N. and Shulman, M. (1983) Intracisternal A-particle genes as moveable elements in the mouse genome. *Proc. Natl. Acad. Sci. USA* **80**: 1992-1996.

Kumar, D.V., Berry, B.T. and Roy-Burman, P. (1989) Nucleotide sequence and distinctive characteristics of the *env* gene of endogenous feline leukemia provirus. *J. Virol.* **63**: 2379-2384.

Lambert, M.E., McDonald, J.F. and Weinstein, I.B. (1988) Editors: Banbury Report 30: Eukaryotic transposable elements as mutagenic agents. Cold Spring Harbor Laboratory, New York.

Lampson, B.C., Inouye, M.and Inouye, S. (1989) Reverse transcriptase with concomitant ribonuclease H activity in a cell-free system of branched RNA-linked msDNA of *Myxococcus xanthus*. *Cell* 56: 709-717.

Larrson, E., Kato, N. and Cohen, M. (1989) Human endogenous proviruses. *Curr. Top. Microbiol. Immunol.* **148**: 115-132.

Leib-Mösch, C., Brack, R., Werner, T., Erfle, V. and Hehlmann, R. (1986) Isolation of an SSAV-related endogenous sequence from human DNA. *Virology* **155**: 666-677.

Leis, J., Baltimore, D., Bishop, J.M., Coffin, J., Fleissner, E., Goff, S.P., Oroszlan, S., Robinson, H., Skalka, A.M., Temin, H.M. and Vogt, V. (1988) Standardized and simplified nomenclature for proteins common to all retroviruses. *J. Virol.* **62**: 1808-1809.

Littlewood, T.D., Tomley, F.M. and Owen, L.N. (1984) Preliminary report on the presence of RNA-dependent DNA polymerase in canine milk. *Research in Veterinary Science* **37**: 97-100.

Lieberman, M. and Kaplan, H.S. (1959) Leukemogenic activity of filtrates from radiation-induced lymphoid tumors of mice. *Science* **130**: 387-388.

Linial, M. and Groudine, M. (1985) Transcription of three *c-myc* exons is enhanced in chicken bursal lymphoma cell lines. *Proc. Natl. Acad. Sci. USA* **82**: 53-57.

Lombard, L.S., Moloney, J.B. and Rickard, C.G. (1963) Transmissible canine mastocytoma. *Ann. N.Y. Acad. Sci.* **108**: 1086-1105.

Lowy, D.R., Rowe, W.P., Teich, N. and Hartley, J.W. (1971) Murine leukemia virus: High frequency activation *in vitro* by 5-iododeoxyuridine and 5-bromodeoxyuridine. *Science* **174**: 155-156.

MacEwen, E.G. and Young, K.M. (1989) Canine lymphoma and lymphoid leukemias. pp. 380-393 In: S.J. Withrow and E.G. MacEwen (editors) Clinical Veterinary Oncology, J.B. Lippincott Company, Philadelphia, USA.

Maddon, P.J., Dalgleish, A.G., McDougal, J.S., Clapham, P.R., Weiss, R.A. and Axel, R. (1986) The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**: 333-348.

Madewell, B.R. and Theilen, G.H. (1987) Lymphoma and Leukemia: Canine. pp. 392-407 In: G.H. Theilen and B.R. Madewell (editors) Veterinary Cancer Medicine, Second Edition, Lea & Febiger, Philadelphia, USA.

Maeda, N. (1985) Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair. The haptoglobin-related gene contains a retrovirus-like element. *J. Biol. Chem.* **260**: 6698-6709.

Mager, D. and Henthorn, P. (1984) Identification of a retrovirus-like repetitive element in human DNA. *Proc. Natl. Acad. Sci. USA.* **81**: 7510-7514.

Maizel, J.V. and Lenk, R.P. (1981) Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA.* **78**: 7665-7669.

Mariani-Constantini, R., Horn, T.M. and Callahan, R. (1989) Ancestry of a human endogenous retrovirus family. *J. Virol.* **63**: 4982-4985.

Marsh, L.J., Erfle, M. and Wykes, E.J. (1984) The pIC plasmid and bacteriophage vectors with versatile cloning sites for recombinant selection by insertional inactivation. *Gene* **32**: 481-485.

Marsh, M. (1984) The entry of enveloped viruses into cells by endocytosis. *Biochem J.* **218**: 1-10.

Martin, G.S. (1986) The *erb*B gene and the EGF receptor. *Cancer Surveys* **5**: 199-219.

Martin, M.A., Bryan, T., Rasheed, S. and Khan, A.S. (1981) Identification and cloning of endogenous retroviral sequences present in human DNA. *Proc. Natl. Acad. Sci. USA.* **78**: 4892-4896.

Martin, S.L., Voliva, C.F., Hardies, S.C., Edgell, M.H. and Hutchison III, C.A. (1985) Tempo and mode of concerted evolution in the L1 repeat family of mice. *Mol. Biol. Evol.* **2**: 127-140.

Mathes, L.E., Olsen, R.G. and Hebebrand, L.C. (1979) Immunosuppressive properties of virus polypeptide, a 15,000 dalton protein, from feline leukaemia virus. *Cancer Res.* **39**: 950-955.

Matthews, R.E. (1979) Classification and nomenclatures of viruses. *Intervirology* **12**: 234-238.

May, F.E.B. and Westley, B.R. (1986) Structure of a human retroviral sequence related to mouse mammary tumor virus. *J. Virol.* **60**: 743-749.

McClure, M.A., Johnson, M.S., Feng, D.-F. and Doolittle, R.F. (1988) Sequence comparisons of retroviral proteins: Relative rates of change and general phylogeny. *Proc. Natl. Acad. Sci. USA.* **85**: 2469-2473.

McClure, M.O., Sommerfelt, M.A., Marsh, M. and Weiss, R.A. (1990) The pH independence of mammalian retroviral infection. *J. Gen. Virol.* **71**: 767-773.

McDonald, J.F., Strand, D.J., Brown, M.R., Paskewitz, S.M., Csink, A.K. and Voss, S.H. (1988) Evidence of host-mediated regulation of retroviral element expression at the posttranscriptional level. pp. 219-234 In: Banbury Report 30: Eukaryotic elements as mutagenic agents. Cold Spring Harbor Laboratory, New York.

Meinkoth, J. and Wahl, G. (1984) Hybridization of nucleic acids immobilised on solid supports. *Analytical Biochem.* **138**: 267-284.

Meuth, M. (1989) Illegitimate recombination in mammalian cells. pp. 833-860 In: D.E. Berg and M.M. Howe (editors), Mobile DNA. American Society for Microbiology, Washington, D.C.

Moldovanu, G., Moore, A.E., Friedman, M. and Miller, D.G. (1966) Cellular transmission of lymphosarcoma in dogs. *Nature* **210**: 1342-1343.

Mölling, K., Bolognesi, D.P., Bauer, H., Büsen, W., Plassmann, H.W. and Hausen, P. (1971) Association of viral reverse transcriptase with an enzyme degrading the RNA moiety of RNA-DNA hybrids. *Nature New Biology* **234**: 240-243.

Morton, D.B., Yaxley, R.E., Patel, I., Jeffreys, A.J., Howes, S.J. and Debenham, P.G. (1987) Use of DNA fingerprint analysis in identification of the sire. *Vet. Record* **121**: 592-593.

Motokura, T., Fukomoto, S., Matsumoto, T., Takahashi, S., Fujita, A., Yamashita, T., Igarashi, T. and Ogata, E. (1989) Parathyroid hormone-related protein in adult T-cell leukemia-lymphoma. *Annals of Internal Medicine* **111**: 484-488.

Mount, S.M. and Rubin, G.M. (1985) Complete nucleotide sequence of the *Drosophila* transposable element copia: Homology between copia and retroviral proteins. *Mol. Cell. Biol.* **5**: 1630-1638.

Niman, H.L., Gardner, M.B., Stephenson, J.R. and Roy-Burman, P. (1980) Differential expression of two distinct endogenous retrovirus genomes in developing tissues of the domestic cat. *J. Natl. Cancer Inst.* **64**: 587-594.

O'Connell, C., O'Brien, S., Nash, W.G. and Cohen, C. (1984) ERV 3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. *Virology* **138**: 225-235.

O'Neill, R.R., Hartley, J.W., Repaske, R. and Kozak, C.A. (1987) Amphotropic proviral envelope sequences are absent from the *Mus* germ line. *J. Virol.* **61**: 2225-2231.

Obata, M.M. and Khan, A.S. (1988) Structure, distribution and expression of an ancient murine endogenous retroviruslike DNA family. *J. Virol.* **62**: 4381-4386.

Oliff, A.I., Hager, G.L., Chang, E.H., Scolnick, E.M., Chan, H.W. and Lowy, D.R. (1980) Transfection of molecularly cloned Friend murine leukemia virus DNA yields a highly leukemogenic helper-independent type-C virus. *J. Virol.* **33**: 475-486.

Onions, D.E. (1980) RNA-dependent DNA polymerase activity in canine lymphosarcoma. *Europ. J. Cancer* **16**: 345-350.

Onions, D.E. (1984) A prospective study of familial canine lymphosarcoma. *J. Natl. Cancer Inst.* **72**: 909-912.

Ono, M. (1986) Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J. Virol.* **58**: 937-944.

Ono, M., Kawakami, M. and Ushikubo, H. (1987) Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer line T47D. *J. Virol.* **61**: 2059-2062.

Overbough, J., Riedel, N., Hoover, E.A. and Mullins, J.I. (1988) Transduction of endogenous envelope genes by feline leukaemia virus *in vitro*. *Nature* **332**: 731-734.

Panganiban, A.T. (1985) Retroviral DNA integration. *Cell* **42**: 5-6.

Paulson, K.E., Deka, N., Schmid, C.W., Misra, R., Schindler, C.W., Rush, M.G., Kadyk, L. and Leinwand, L. (1985) A transposon-like element in human DNA. *Nature* **316**: 359-361.

Payne, L.N. and Chubb, R.C. (1968) Studies on the nature and genetic control of an antigen in normal chick embryos which reacts in the COFAL test. *J. Gen. Virol.* **3**: 379-391.

Peters, G. (1989) Oncogenes at viral integration sites. pp 23-66 In: D.M. Glover and B.D. Hames (editors), Oncogenes. IRL Press at Oxford University Press.

Post, J.E., Noronha, F. and Rickard, C.G. (1970) Canine mast cell leukemia. *Bibl. Haem.* **36**: 425-429.

Queen, C. and Korn,L.J. (1984) A comprehensive sequence analysis programme for the IBM personal computer. *Nucl. Acids Res.* **12** 581-599.

Raines, M.A., Lewis, W.G., Crittenden, L.B. and Kung, H.J. (1985) c-*erb*B activation in avian leukosis virus-induced erythroblastosis: clustered integration sites and the arrangement of provirus in the c-*erb*B alleles. *Proc. Natl. Acad. Sci. USA.* **82**: 2287-2291.

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

Ruscetti, S. and Wolff, L. (1984) Spleen focus-forming virus: relationship of an altered envelope gene to the development of rapid erythroleukemia. *Curr. Top. Microbiol. Immunol.* **112**: 21-44.

Reeves, R.H. and O'Brien, S.J. (1984) Molecular genetic characterization of the RD 114 gene family of endogenous feline retroviral sequences. *J. Virol.* **52**: 164-171.

Repaske, R., Steele, P., O'Neill, R., Rabson A. and Martin, M. (1985) Nucleotide sequence of a full-length human endogenous retroviral segment. *J. Virol.* **54**: 764-772.

Rogers, J. (1983) Retroposons defined. *Nature* **301**: 460.

Rogers, J. (1985) The origin and evolution of retroposons. *Int. Rev. Cytol.* **93**: 187-279.

Rosen, C.A., Sodroski, J.G., Goh, W.C., Dayton, A.I., Lippke, J. and Haseltine, W.A. (1986) Post-transcriptional regulation accounts for the *trans*-activation of the human T-lymphotropic virus type III. *Nature* **319**: 555-559.

Rudolph, R. (1971) Virusähnliche strukturen in tumorzellen bei lymphatischer leukose des hundes. *Berl. Münch. tierärztl. Wschr.* **84**: 68-70.

Ruegg, C.L., Monell, C.R. and Strand, M. (1989) Inhibition of lymphoproliferation by a synthetic peptide with sequence identity to gp41 of Human Immunodeficiency Virus type I *J. Virol.* **63**: 3257-3260.

Russell, P.H., and Jarrett, O. (1976) An improved assay for feline leukaemia virus pseudotypes of murine sarcoma virus. *J. Gen. Virol.* **31**: 139-143.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989a) Bacteriophage lambda growth, purification and DNA extraction. pp 2.60-2.81 In: Molecular Cloning: A laboratory manual. 2nd Edition. Cold Spring Harbour Laboratory Press, New York.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989b) Extraction and purification of plasmid DNA. pp 1.21-1.52 In: Molecular Cloning: A laboratory manual. 2nd Edition. Cold Spring Harbour Laboratory Press, New York.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989c) Sequencing by the Sanger dideoxy-mediated chain-termination method. pp 13.42-13.77 In: Molecular Cloning: A laboratory manual. 2nd Edition. Cold Spring Harbour Laboratory Press, New York.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.

Schneider, R. (1983) Comparison of age- and sex-specific incidence rate patterns of the leukemia complex in the cat and the dog. *J. Natl. Cancer Inst.* **70**: 971-977.

Sarma, P.S. and Log, T. (1973) Subgroup classification of feline leukemia and sarcoma viruses by viral interference and neutralization tests. *Virology* **54**: 160-169.

Scheidereit, C. and Beato, M. (1984) Contacts between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus. *Proc. Natl. Acad. Sci. USA* **81**: 3029-3033.

Schmidt, D.M., Sidhu, N.K., Cianciolo, G.J. and Snyderman, R. (1987) Recombinant hydrophilic region of murine retroviral protein p15E inhibits stimulated T-lymphocyte proliferation. *Proc. Natl. Acad. Sci. USA* **84**: 7290-7294.

Schwartzberg, P., Collicelli, J. and Goff, S.P. (1985) Recombination between a defective retrovirus and homologous sequences in host DNA: Reversion by patch repair. *J. Virol.* **53**: 719-726.

Seman, G., Proenca, G., Guillon, J.C. and Moraillon, R. (1967) Particules d'aspect viral dans les cellules du lymphosarcome du chien. *Bull. Acad. Vet. (France)* **40**: 211-214.

Shiba, T and Saigo, K. (1983) Retrovirus-like particles containing RNA homologous to the transposable element copia in *Drosophila melanogaster*. *Nature* **302**: 119-124.

Shimotohno, K., Mitzutani, S. and Temin, H.M. (1980) Sequence of retrovirus provirus resembles that of bacterial transposable elements.; *Nature* **285**: 550-554.

Shimotohno, K. and Temin, H.M. (1980) No apparent nucleotide specificity in cellular DNA juxtaposed to retrovirus proviruses. *Proc. Natl. Acad. Sci. USA.* **77**: 7357-7361.

Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. *Nucl. Acids Res.* **16**: 7583-7600.

Smith, T.F., Srinivasen, A., Schochetman, G., Marcus, M. and Myers, G. (1988) The phylogenetic history of immunodeficiency viruses. *Nature* **333**: 573-575.

Smith, T.F. and Waterman, M.S. (1981) Comparison of biosequences. *Advances in Applied Mathematics* **2**: 482-489.

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

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

Steele, P.E., Martin, M.A., Rabson, A.B., Bryan, T. and O'Brien, S. (1986) Amplification and chromosomal dispersion of human endogenous retroviral sequences. *J. Virol.* **59**: 545-550.

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

Steffen, D. and Weinberg, R.A. (1978) The integrated genome of murine leukemia virus. *Cell* **15**: 1003-1010.

Stein, B.S., Gowda, S.D., Lifson, J.D., Penhallow, R.C., Bensch, K.G. and Engelman, E.G. (1987) pH independent HIV entry into CD4-positive T-cells via virus envelope fusion to the plasma membrane. *Cell* **49**: 659-668.

Stewart, M.A., Warnock, M., Wheeler, A., Wilkie, N., Mullins, J.I., Onions, D.E. and Neil, J.C. (1986) Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and long terminal repeat and evidence for the recombinational origin of subgroup B viruses. *J. Virol.* **58**: 825-834.

Stoye, J. and Coffin, J. (1985) Endogenous viruses. pp. 357-404 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Supplement. Cold Spring Harbour Laboratory, New York.

Stoye, J. and Coffin, J. (1988) Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. *J. Virol.* **62**: 168-175.

Strandstrom, H.V. and Bowen, J.M. (1982) Canine leukemia-lymphoma complex: A model for human haematopoeitic malignancies. pp. 447-450 In: D.S. Yohn and J.R. Blakeslee (editors) Advances in Comparative Leukemia Research 1981, Elsevier North Holland, Inc, New York, USA.

Strandstrom, H.V. and Rimaila-Parnanen, E. (1979) Canine atypical malignant lymphoma. *Am. J. Vet. Res.* **40**: 1033-1034.

Sykes, G.P., King, J.M. and Cooper, B.C. (1985) Retrovirus-like particles associated with myeloproliferative disease in the dog. *J. Comp. Pathol.* **95**: 559-564.

Summers, D.K. and Sherratt, D.J. (1988) Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the *cer* site. *The EMBO Journal* 7: 851-858.

Swanstrom, R. and Vogt, P.K. (1990) Retroviruses: Strategies of replication. (Editors) *Curr. Top. Microbiol. Immunol.* **157.**

Tabor, S. and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* **84**: 4767-4771.

Tamura, T. (1983) Provirus of baboon endogenous virus: nucleotide sequence of the *gag-pol* region. *J. Virol.* **47**: 137-145

Teich, N. (1982) Taxonomy of retroviruses. pp. 25-208 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Text. Cold Spring Harbour Laboratory, New York.

Teich, N. (1985) Taxonomy of retroviruses. pp. 1-16 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Supplement. Cold Spring Harbour Laboratory, New York.

Teich, N., Lowy, D.R., Hartley, J.W. and Rowe, W.P. (1973) Studies of the mechanism of induction of infectious murine leukemia virus from AKR mouse embryo cell lines by 5-iododeoxyuridine and 5-bromodeoxyuridine. *Virology* **51**: 163-173.

Teich, N., Wyke, J., Mak, T., Bernstein A. and Hardy, W. (1982) Pathogenesis of retrovirus-induced disease. pp. 785-998 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Text. Cold Spring Harbour Laboratory, New York.

Teich, N., Wyke, J. and Kaplan, P. (1985) Pathogenesis of retrovirus-induced disease. pp. 187-248 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Supplement. Cold Spring Harbour Laboratory, New York.

Temin, H.M. (1970) Malignant transformation of cells by viruses. *Perspec. Biol. Med.* **14**: 11-26.

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

Tomley, F.M., Armstrong, S.J., Mahy, B.W.J. and Owen, L.N. (1983) Reverse transcriptase activity and particles of retroviral density in cultured canine lymphosarcoma supernatants. *Br. J. Cancer* **47**: 277-284.

Temin, H.M. (1985) Reverse transcription in the eukaryotic genome: Retroviruses, pararetroviruses, retrotransposons and retrotranscripts. *Mol. Biol. Evol.* **2**: 455-468.

Temin, H.M. and Baltimore, D. (1972) RNA directed DNA synthesis and RNA tumor viruses. *Adv. Virus Res.* **17**: 129-186.

Tereba, A. (1983) Assymetric chromosomal distribution of endogenous retroviral loci in chickens and mice. *Curr. Top. Microbiol. Immunol.* **107**: 29-50.

Terwilliger, E., Sodroski, J.G., Rosen, C.A. and Haseltine, W.A. (1986) Effects of mutations within the 3' *orf* open reading frame region of human T-lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity. *J. Virol.* **60**: 754-760.

Varmus, H. (1988) Regulation of HIV and HTLV gene expression. *Genes Dev.* **2**: 1055-1062.

Varmus, H. and Brown, P. (1989) Retroviruses. pp. 53-108 In: D.E. Berg and M.M. Howe (editors), Mobile DNA. American Society for Microbiology, Washington, D.C.

Varmus, H. and Swanstrom, R. (1982) Replication of retroviruses. pp. 369-512 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Text. Cold Spring Harbour Laboratory, New York.

Varmus, H. and Swanstrom, R. (1985) Replication of retroviruses. pp. 75-134 In: R. Weiss, N. Teich, H. Varmus and J. Coffin (editors), Molecular Biology of Tumor viruses: RNA Tumor Viruses 2nd Edition/Supplement. Cold Spring Harbour Laboratory, New York.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (1987) The extraordinary diversity of eukaryotic viruses. pp. 898-959 In: Molecular biology of the gene. Fourth edition. Benjamin Cummings Publishing Co., Menlo Park, California.

Wayne, R.K. and O'Brien, S.J. (1987) Allozyme divergence within the canidae. *Syst. Zool.* **36**: 339-355.

Weickert, M.J. and Chambliss, G.H. (1989) Acid phenol minipreps make excellent sequencing templates. *Editorial Comments, United States Biochemical Corp.* **16**: No.2 5-6.

Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) Nonviral retroposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. *Ann. Rev. Biochem.* **55**: 631-661.

Weir, E.C. (1990) Mechanisms of hypercalcemia in canine lymphosarcoma and apocrine cell adenocarcinoma of the anal sac. pp. 921-924 In: Proceedings of the Eighth Annual Veterinary Medical Forum, Washington, D.C., May 1990. American College of Veterinary Internal Medicine.

Weir, E.C., Burtis, W.J., Morris, C.A., Brady, T.G. and Insogna, K.L. (1988) Isolation of 16,000-dalton parathyroid hormone-like proteins from two animal tumors causing humoral hypercalcemia of malignancy. *Endocrinology* **123**: 2744-2751.

Weiss, R.A., Friis, R.R., Katz, E. and Vogt, P.K. (1971) Induction of avian tumor viruses in normal cells by physical and chemical carcinogens. *Virology* **46**: 920-938.

Werner, T., Brack-Werner, R., Leib-Mösch, C., Backhaus, H., Erfle, V. and Hehlmann, R. (1990) S71 is a phylogenetically distinct human endogenous retroviral element with structural and sequence homology to simian sarcoma virus. *Virology* **174**: 225-238.

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

Wilson, M.C., Policastro, P.F. and Fredholm, M. (1988) Regulation of expression and transposition of murine endogenous retroviral elements. pp. 131-144 In: Banbury Report 30: Eukaryotic elements as mutagenic agents. Cold Spring Harbor Laboratory, New York.

Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L. and Treiber, G. (1970) Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large scale virus purification. *Virology* **40**: 734-744.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.

Yokoyama, S. and Gojobori, T. (1987) Molecular evolution and phylogeny of the human AIDS viruses LAV, HTLV-III and ARV. *J. Mol. Evol.* **24**: 330-336.

Zasloff, M., Ginder, G.D. and Felsenfeld, G. (1978) A new method for the purification and identification of covalently closed circular DNA molecules. *Nucl. Acids Res.* **5**: 1139-1152.

Zelnick, C.R., Burks, D.J. and Duncan, C.H. (1987) Mobilization of genomic sequences by flanking *Alu*-type repetitive DNA. *Nucl. Acids Res.* **15**: 10437-10453.

GLASGOW
UNIVERSITY
LIBRARY