The Study of Retroviral Sequences in Human Leukaemia

Thesis

How to cite:

For guidance on citations see FAQs.

© 2001 Richard Moore

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.21954/ou.ro.0000fcee

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.

oro.open.ac.uk
The Study of Retroviral Sequences

in Human Leukaemia

by

Richard Moore

A dissertation submitted for the degree of
Doctor of Philosophy
at the Open University
in the sponsoring establishment of Addenbrookes Hospital, Cambridge

November 2000

Department of Haematology
Declaration

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration. None of the work described here has been submitted for any degree, diploma or any other qualification at any University.

Richard Moore

November 2000
Acknowledgements

This feels like an Oscar ceremony so many people to thank...

Firstly I would like to thank my supervisor Abraham Karpas for his help and support throughout. Also in the Haematology department I wish to thank my second supervisor Juraj Petrik for his valuable insight and direction. In this department I received help and inspiration for many others including the many faces who passed through the Karpas laboratory. A few of whom were Solvita, Mike, Mike, Elisa, Steve, Pooja and Alex.

I am also grateful to Mike Tristem for his assistance at the embryonic stage of the project.

More recently I would especially like to thank Margaret Stanley who has been extremely understanding and supportive during my battle with this thesis. Also in Pathology I want to mention the help of all my friends and colleagues especially those of us up on the roof. A special mention to the FISH Coleman group for help with my PAC FISH.

To my friends one and all near and far you know who you are and thanks for the support and "encouragement" over the many years this has taken.

Finally for those without whom this would never have been possible. To my family and especially my parents what can I say the greatest help and support anyone could ask for thankyou. To Mel you are just amazing how do you put up with me,

Thanks.
Abstract

Retroviruses have been closely associated with cancer and leukaemia since their first discovery in the beginning of the 20th century. Many retroviruses both endogenous and exogenous have a direct causal role in leukaemia in many animal species. However in humans the only retrovirus that has been shown to cause leukaemia to date is HTLV-I with ATL. ATL generally affects those people within HTLV-I endemic areas in particular Japan, sub-Saharan Africa and the Caribbean basin. ATL cases in Europeans and Caucasians in general are very rare. In this thesis the HTLV-I sequence was isolated from a Greek Caucasian patient with ATL. The virus was subtyped and shown to novel and of the Cosmopolitan type A, the most internationally found subtype, known as the transcontinental variant. This data highlights the potential for further spread of this pathogenic virus into Western Europe and illustrates that ethnicity is not a barrier to ATL. This thesis supports the expansion of a screening programme throughout Europe including the UK.

Around 10% of the human genome is composed of sequences resulting of retrotransposition. One of the main sources of this RT activity is Human Endogenous Retrovirus-like sequences (HERV). This thesis details the isolation and characterisation of two novel HERV sequences. Their phylogeny is also considered with one HERV clustering within the HERV-I family and the second clustering in the HERV-XA family. A low copy number was estimate for both elements. PAC-FISH revealed their chromosomal locations and the relevance of these regions is discussed. Expression of both elements was analysed by RT-PCR using primers directed toward putative ORFs. This analysis revealed an expressed portion of the XA related element, spanning 880bp, in PBMC. This expression was analysed in 10 leukaemia cases and was illustrated to be higher in some AML cases. This data illustrates that even highly disrupted HERV have some coding potential and supports a possible role in leukaemogenesis for HERV in general.
Contents

Declaration i
Acknowledgements ii
Abstract iii

Chapter 1 General introduction

1.1 Retroviruses and Cancer 1
1.2 Retroviruses and Retroelements 4
1.3 HTLV I and II 15
1.4 Human Endogenous Retroviruses (HERVs) 18
1.5 Leukaemia 20

Chapter 2 Materials and Methods

2.1 Materials 26
  2.1.1 Bacterial Culture 26
  2.1.2 Enzymes 26
  2.1.3 Reagents 27
  2.1.4 Equipment 27
  2.1.5 DNA markers 28
2.2 Methods 29
  2.2.1 Peripheral Blood Mononucleocyte Isolation 29
  2.2.2 Genomic DNA Isolation 29
  2.2.3 Southern Blot 30
  2.2.4 Radiolabelled Probe Preparation 31
  2.2.5 Plasmid Preparation 32
  2.2.6 Preparation of Competent Cells 33
  2.2.7 PCR 35
  2.2.8 Electrophoresis 36
  2.2.9 DNA Purification from Agarose Gels 37
Chapter 3 Characterisation of Human T-cell Leukaemia Virus type 1 (HTLV-I) from a Caucasian ATL patient

3.1 HTLV-I and Disease
3.1.1 Adult T-cell Leukaemia (ATL)
3.1.2 Tropical Spastic Paraparesis (TSP)
3.2 The HTLV Group
3.2.1 HTLV-I Structure and Classification
3.3 Epidemiology
3.4 Case History
3.5 Results
3.5.1 Southern Blot Analysis
3.5.2 PCR Analysis of LTR
3.5.3 PCR Analysis of ENV
3.5.4 Phylogenetic Analysis of the LTR Region
3.6 Discussion

Chapter 4 Isolation and Characterisation of Two Novel HERV Fragments

4.1 HERVs
4.1.1 Class I HERV
4.1.2 Class II HERV
4.1.3 Class III HERV
4.1.4 Other HERV
4.2 Results
4.2.1 PCR
Chapter 5 Expression of Two Novel HERVs

5.1 HERV Expression
5.1.1 Clone 4-1 (HERV-E)
5.1.2 ERV3 (HERV-R)
5.1.3 RRHERV-1
5.1.4 S71
5.1.5 HRES-1
5.1.6 ERV-9
5.1.7 HERV-H
5.1.8 HERV-K
5.1.9 HERV-L
5.1.10 Potential HERV Expression

5.2 Biological Significance
5.2.1 Protection
5.2.2 Genome Plasticity
5.2.3 Disease Association

5.3 Human Leukaemia
5.3.1 Acute Myelocytic Leukaemia
5.3.2 ALL
5.3.3 Rare Forms of Acute Leukaemia
5.3.4 Chronic Lymphoid Proliferations
5.3.5 Chronic Myelocytic Leukaemia
5.3.6 Myelodysplastic Syndromes (MDS)
5.3.7 Chromosomal Abnormalities
5.3.8 Causes of Leukaemia
5.3.9 Retroviruses and Leukaemia

5.4 Results
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4.1 RNA Expression</td>
<td>164</td>
</tr>
<tr>
<td>5.4.2 Probes</td>
<td>165</td>
</tr>
<tr>
<td>5.4.3 RT-PCR</td>
<td>166</td>
</tr>
<tr>
<td>5.4.4 PAC Mapping</td>
<td>168</td>
</tr>
<tr>
<td>5.5 Discussion</td>
<td>170</td>
</tr>
</tbody>
</table>

Chapter 6 General Discussion

Chapter 7 Reference List

Appendix Publications during period of registration
Chapter 1: General Introduction

This thesis focuses on human retroviruses and their involvement in cancer, with particular reference to leukaemia. Hence an introduction to retroviruses and their long association with this disease is presented.

1.1 Retroviruses and Cancer

Environmental mutagens can contribute to the development of cancer; these mutagens include radiation chemicals and viruses. Retroviruses may play a key role in one or more of the stages in the multistep process of carcinogenesis, either as a direct mutagen, or by more subtle indirect means.

Avian and mammalian type C retroviruses have been shown to cause cancer in their natural hosts. The study of these animal retroviruses has contributed to the understanding of the genetic basis of human leukaemia and lymphoma.

Since it has been shown that adult T-cell leukaemia is caused by HTLV-I a blood screening test has been developed to prevent the viruses spread through blood transfusions. HIV-1 and 2 are also screened for to prevent their spread by the same mechanism. The possibility of discovering more retroviral involvement in human disease should not be discounted.

In birds, cats and rodents type C retroviruses have a direct causal role in leukaemia. In contrast, with the exception of HTLV-I, there is little evidence yet that the retroviruses are involved in the development of other human malignancies. Retroviral involvement may be direct or indirect. If acting directly a viral gene product or viral control sequence acts from within the tumour cell causing cell proliferation. If acting indirectly viral gene products or control sequences alter cells other than the tumour cell causing an indirect effect on the target cell, e.g. immune system modifications.

Some retroviruses have an integrated cDNA copy of a cellular proto-oncogene. These viruses are invariably replication deficient due to this insertion and
corresponding deletion of a portion of the retroviral genome. These retroviruses therefore require a second virus before they can undergo an infectious life cycle. The prototype cancer causing retrovirus studied was Rous Sarcoma Virus (RSV), isolated by Peyton Rous from chicken (Rous 1911), RSV causes soft tissue sarcomas quickly in birds and is an acute transforming virus. It is formed by a recombination between Avian Leukosis Virus (ALV) and the host genome (Stehelin et al. 1976) involving src-a mutant form of a cellular proto-oncogene. Now more than 20 host cell genes have been isolated from acutely transforming viruses. All of these are viral forms of cellular proto-oncogenes e.g. myc from avian myelocytomatosis virus (Coffin et al. 1981). Although there are no known acutely transforming viruses that affect humans, there are close homologues to most of these proto-oncogenes in the human genome.

Retroviral insertion can contribute to tumour progression e.g. insertion close to ets-1 (proto-oncogene) (Bear et al. 1989) and tpl-2 a probable serine-threonine protein kinase (Patriotis et al. 1993). In 1981 Hayward investigated avian bursal lymphomas and found that they were caused by ALV virus integration close to the myc gene (Hayward et al. 1981). Retroviruses seem to show a random integration pattern, but may in fact be directed towards actively transcribed regions of chromatin (Sandmeyer et al. 1990). Insertional mutagenesis has led to the discovery of many oncogenes (Peters 1990), including both novel genes and those previously isolated from acutely transforming viruses.

Oncogenes’ normal functions include the regulation of cell proliferation, differentiation or longevity, over expression or mutant expression of oncogenes can cause cell proliferation and deviant cellular behaviour leading to cancer. Oncogenes control cells by various mechanisms and act at various levels of regulation; they can be related to growth factors (e.g. Sis, platelet derived growth factor) (Waterfield et al. 1983); transmembrane receptors e.g. EGFR (Downward et al. 1984); signal transduction machinery e.g. Ras (Downward 1990); hormone receptors e.g. THRA thyroid hormone (Sap et al. 1986); transcription factors e.g. Jun, Fos of AP-1 complex (Kouzarides et al. 1989); or regulators of cell cycle progression e.g. Vin-1 -cyclin D2 (Hanna et al. 1993).

In Burkitt’s lymphoma the common translocations t(8;14, 8;2, 8;22), all involve the MYC locus (Croce et al. 1985) and the Philadelphia chromosome
translocation t(9;22) involves the ABL1 oncogene (Heisterkamp et al. 1983). The study of retroviruses has been instrumental in the discovery of many oncogenes and so the understanding of the biology of human leukaemia.

The study of retroviral gene knockout, such as that associated with the Friend strain of Murine Leukosis virus (MLV) has led to the isolation of tumour suppressor genes. When MLV was inserted into a p53 coding region, this resulted in the loss of p53 function in erythroleukaemia (Munroe et al. 1990) and thereby highlighted p53 as a tumour suppressor gene.

The Env glycoprotein of some murine retroviruses can interact with growth factor receptors in the host cell membrane and cause proliferation of the cell without the presence of exogenous growth factors. For example, Friend spleen focus forming virus via the erythropoietin receptor (Li et al. 1990) and Mink Cell Focus inducing (MCF) virus via the IL-2 receptor. MCF virus is formed by recombination with endogenous murine leukosis virus (Stoye et al. 1991) and can infect different cells, as the recombinant virus possesses polytropic env (Kozak et al. 1992). This virus was first observed in Akv mice after the formation of spontaneous tumours (Hartley et al. 1977).

Retroviral long terminal repeats (LTRs) can exert transcriptional control on host genes, by enhancer (Nusse et al. 1982; Payne et al. 1982; Cuypers et al. 1984), readthrough or a more distal paracrine method (Tsichlis et al. 1990). Viral LTR may also show tissue specificity and can be preferentially active in certain tissues (Stocking et al. 1985). This however is not found to be the primary factor for many virally induced tissue specific tumours (Green et al. 1988; Jolicoeur 1991; Feuer et al. 1993). Fusion of viral protein sequences to oncogenes may also influence the tissue specificity of transformation, e.g. Gag-Abl fusion (Goff et al. 1981).

However for non-acute, long latency tumours, LTRs are the major determinants of pathogenicity (Nusse et al. 1985; Fan et al. 1988; Feuer et al. 1990). Core elements in the transcriptional enhancer of the LTR seem to be the most important (Speck et al. 1990; Hallberg et al. 1991). Deletion of the downstream GC-rich region from the enhancer leads to a relaxation of the specificity of other enhancer elements and can cause multiple lineage leukaemias (Speck et al. 1990; Hallberg et al. 1991).
1.2 Retroviruses and Retroelements

This thesis focuses on retroviruses and HERV therefore an introduction to retroviral classification and structure is given below.

The RNA nature of some infectious agents was discovered at an early stage, the first retrovirus identified was Equine infectious anaemia virus in 1904 followed by avian leukosis and sarcoma viruses in 1908 and 1911 respectively. During the 1950s and 1960s the term RNA tumour viruses was adopted. However after Temin and Baltimore independently demonstrated that these viruses contain a unique enzyme which can code for a DNA copy of the viral RNA named reverse transcriptase, the viruses were subsequently named retroviruses. Retroviral LTRs were described in 1978 and the first isolated human retrovirus HTLV-I was shown to be involved in the development of adult T-cell leukaemia in 1981.

Retroelements are unique amongst viruses as they utilise a viral encoded reverse transcriptase to replicate through a DNA intermediate. Retroelements, which are transposable or movable genetic elements, function via an RNA intermediate and share the feature of having integrated by reverse transcription converting RNA to DNA. They are ubiquitous in vertebrates, invertebrates, plants, fungi and prokaryotes (Garfinkel 1992; Temin 1992). Reverse transcriptase is an ancient protein and has been conserved throughout evolution from bacteria and the mitochondria of fungi to mammals (Garfinkel 1992). In myxobacteria and some E. coli retroelements exist as single-copy genes called retrons (Temin 1989). No cellular gene for reverse transcriptase has been identified in eukaryotes only those present in retroelements, however reverse transcription is involved in the maintenance of telomeres (Lundblad et al. 1990). It is proposed that retroelements evolved by co-opting a cellular enzyme in a far distant ancestor (Temin 1980), their broad distribution and conserved amino acid homology among the reverse transcriptases, including retrons, support this theory.

Transposition increases element copy number as the process of reverse transcription creates a new copy each time, resulting in up to $10^5$ copies dispersed throughout the genome. An insertion target site is opened by causing staggered nicks
to each DNA strand leading to short direct repeats of target DNA flanking the inserted element. Many retroelements also contain an indication of an RNA intermediate e.g. a 3’ polyA tract downstream of the polyA signal.

The primary distinction for classification within retroelements is the presence or absence of an LTR. If an LTR is present the element is a retrotransposon, if LTRs are absent it is a retroposon (Hull et al. 1989). Both groups contain members that encode reverse transcriptase and those that don’t are thought to be an efficient template when the RT is supplied by another element.

Retroposons were originally designated as repetitive DNA sequences and classified by the restriction sites they contained, e.g. KpnI elements (now L1 elements) and Alu elements. On the basis of length they are divided into long interspersed elements (LINEs) and short interspersed elements (SINEs) (Singer 1982). In mammals there appears to be only one LINE per species, so the term is interchangeable with L1, for example L1Hs for L1 in Homo sapiens, SINEs do not encode RT, but LINEs do.

LINEs have been found in insects, protozoa, plants and fungi (Eickbush 1992) as well as mammals. In mammals, the L1 family has approximately $10^4$ randomly integrated, dispersed copies (Skowronski et al. 1986). L1 elements are approximately 6.5kb in length and at the 3’ end have a polyA tract resembling mRNA. Most L1 elements are disrupted by deletions and point mutations and are commonly truncated at the 5’ end (Skowronski et al. 1986). However coding-competent L1 elements containing two ORFs have been found in both mice and humans (Martin et al. 1984; Dombroski et al. 1991). The 3’ ORF encodes a protein with homology to RT (Hattori et al. 1986) and has been shown to possess RT activity (Mathias et al. 1991). L1 elements, the RT protein and a protein equating in size to that encoded by the 5’ ORF have been observed in cytoplasmic macromolecules in human carcinoma cell lines (Deragon et al. 1990) as well as in mice (Martin 1991) and may correspond to where L1 is transcribed. L1 elements contain an internal RNA polymerase II promoter 3’ to the transcription start site (Swergold 1990; Minakami et al. 1992) and promoters are also seen in analogous elements in Drosophila melanogosta (Mizrokhi et al. 1988). These promoters ensure that new copies of the element retain transcriptional ability.

The most abundant SINEs are retroposons, Alu elements are an example with $10^5$-$10^6$ copies in the human genome. Alu elements are 300bp in length with a 3’
polyA tract (Weiner et al. 1986). They have no ORF and are derived from a 7SL structural RNA (Ullu et al. 1984). Alu elements in humans are dimeric repeats of internally deleted 7SL RNAs, each ending in a polyA-rich tract. This is referred to as a processed pseudogene as it is derived from a cellular gene. Alu elements are efficient templates for RT, as is shown by their high copy number, as a comparison other pseudogenes tend to have only up to 100 copies (Vanin 1985). They also possess autonomous transcription signals in vivo (Matera et al. 1990) with internal regulatory sequences. Deletions may have enhanced these signals so Alu elements are no longer dependent on upstream signals as is the case for 7SL (Weiner et al. 1986). RNA polymerase III, which depends on a base paired nucleotide primer for initiation, can transcribe these elements, the termination signal is a tract of polyT. Therefore Alu elements may self-prime via their polyA and polyT tracts (Weiner et al. 1986), or may associate with a source of RT e.g. L1 macromolecular structures, L1-Alu fusion elements support this idea (Miyake et al. 1983).

The human genome also has 4,000-5,000 copies of SINE-R (Ono et al. 1987), these are derivatives of HERV-K, made up of a 630bp short stretch of internal 3’ HERV-K followed by an LTR, the LTR has a 370bp deletion and terminates with a polyA tract downstream of the polyA signal. At the 5’ end there is a tandemly repeated 40bp GC rich internal RNA polymerase II promoter which is not derived from HERV-K (Ono et al. 1987); SINE-R has no coding capacity.

Retrotransposons are 6-7kb in length, with two ORFs, which often overlap and are flanked by 300-600bp LTRs. The number of copies per genome varies from ten copies up to thousands (Boeke et al. 1989). The same gene homologues are best characterised for Ty1 in Saccharomyces cerevisiae and copia in Drosophila melanogaster, their expression yields a ribonucleoprotein particle within the cell (Shiba et al. 1983; Garfinkel et al. 1985; Garfinkel 1992). The 5’ ORF (A) encodes the major protein component of these particles, ORF B encodes an RT and this RT is also present in the particle along with a mRNA copy of the element (Adams et al. 1987). This mRNA appears to be a transposition intermediate (Garfinkel 1992). Elements unable to encode proteins due to mutations are still able to transpose, probably because products of coding-competent elements can act in trans on transcripts from defective elements.
Chapter 1

General Introduction

Transposon-like human elements (THE-1) are dispersed by retrotransposition (Deka et al. 1988). They have 350bp LTRs, flanked by 5bp repeats. The THE-1 consensus sequence is 2.3kb and lacks ORFs or any significant sequence homology with other retrotransposons or retroviruses (Paulson et al. 1985). They also lack any primer binding sites for plus and minus strand synthesis. There are 10,000 THE-1 in the haploid genome and another 10,000 solitary LTRs (Paulson et al. 1985). Together these sequences account for about 1% of the total genome. THE-1 was first identified by Law (Law et al. 1982), they are also known as O family repeats, LRS (low repeat sequence) and MstII repeats. THE-1 are also found in other primates genomes but not in rodents (Paulson et al. 1985). The prosimian *Galago* (Bush baby) transposon-like element for instance only has a few copies per genome (Schmid et al. 1990). Study of this element from *Galago*, found it to be flanked 5' by a common repeat sequence related to R of the THE-1 LTR and immediately 3' to this are sequences with homology to different parts of the THE-1 internal sequence (Schmid et al. 1990). This leads to the conclusion that THE-1 are a composite formed from two pre-existing genomic loci. In several cell lines extrachromosomal circular DNA containing THE-1 sequences were detected (Paulson et al. 1985; Misra et al. 1987). 10% of these forms were cloned and characterised in Hela cells and were found to be derived from a single genomic locus by recombination with adjacent cellular sequences, so were not typical (Misra et al. 1989).

Variable numbers of tandem repeats (VNTR) sequences have been found associated with THE-1 solitary LTRs. One VNTR is a tandem repeat of the internal part of the THE-1 LTR. The other is the junction region of flanking DNA (Mermer et al. 1987). There is evidence that THE-1 sequences may predispose a region to rearrange (Pizzuti et al. 1992) and hot spots have previously been linked to THE-1 (Keyeux et al. 1989; Fields et al. 1992). An example of a linked translocation was noted with a 15kb section of chromosome six (Wong et al. 1990) and there are also examples of papilloma virus integration close to THE-1 (Baker et al. 1987). THE-1 lack a defined coding strand so it is hard to predict their transcriptional orientation (Paulson et al. 1985; Paulson et al. 1987). Therefore they usually undergo external promotion (Paulson et al. 1987). However a putative promoter has been identified in one LTR (Misra et al. 1987; Fields et al. 1992) and a pseudogene found (Deka et al. 1988). THE-1 have been studied in detail (Paulson et al. 1987). A transcript was
reported embedded in a cellular sequence in the 3’ untranslated region of a mRNA coding for a calmodulin (calcium binding) gene (Deka et al. 1988). There are some reports of possible roles for THE-1 elements in gene expression (Deka et al. 1988), or as an origin of replication (Misra et al. 1989), or as nuclear matrix attachment points (Law et al. 1989), but any cellular function is still under debate.

Figure 1 An Illustration of the Classification of Human Retroelements

**Retroposon -LTR**

![Retroposon -LTR Diagram]

**Retroelement +LTR**

![Retroelement +LTR Diagram]

**Transposon like human elements THE-1**

![Transposon like human elements THE-1 Diagram]

**Solitary LTRs**

![Solitary LTRs Diagram]
Retrotransposons are evolutionary precursors to infectious exogenous retroviruses. Retroviruses are similar to retrotransposons but have a higher degree of complexity. They have the same sequence features and homologues of the two ORFs encoded in retrotransposons and also have at least one other ORF that leads to their infectivity (env) which is located between the 3' LTR and the reverse transcriptase. The two ORF homologues are the gag and the pol regions with the general retroviral structure being LTR-gag-pol-env-LTR.

The life cycle of infectious retroviruses is more complex as there is an extracellular phase. They have only been identified in vertebrate species including fish and snakes, implying they may have evolved more recently acquiring an env gene (Temin 1980; Temin 1982). Some retrotransposons e.g. gypsy families in Drosophila have a third ORF in the same position as env and although these also have the most closely related RT (Boeke et al. 1989; Xiong et al. 1990) they are non-infectious and the ORF product is unknown.

Reverse transcription occurs within a nucleoprotein complex derived from the infecting virion and can be complemented by functional proteins from another provirus, but only at virus assembly (Varmus et al. 1991). Retroviral LTRs possess both promoters and enhancers for RNA polymerase II.

Endogenous retroviruses typically occur as 1-100 copies per genome with most being disrupted (Coffin 1984). Some examples of functional endogenous retroviruses are baboon endogenous virus, endogenous feline leukaemia virus, endogenous murine leukaemia viruses and avian endogenous retroviruses (Coffin 1984; Kozak et al. 1992; Payne 1992). There are also dispersed, multicopy families with varying degrees of similarity to retroviruses, called endogenous retrovirus-like (ERV) elements, these can also be more similar to retrotransposons.

Retroviruses are obligate parasites and are found in most vertebrate species. Retrovirus-like elements have been isolated from neoplasia in reptiles (Jacobson et al. 1980), as well as bivalve molluscs (Underwood et al. 1977). Those elements isolated are homologous to those found in higher organisms. There are also several retroviruses that naturally infect fish (Pilcher et al. 1980), as well as two well-documented retroviruses of snakes, viper retrovirus (VRV) (Zeigel et al. 1969; Zeigel et al. 1971) and Corn snake retrovirus (CSRV) (Lunger et al. 1974).
Chapter 1

General Introduction

Exogenous or endogenous retroviral particles can be detected by electron microscopy as enveloped particles of about 100 nm in diameter. The particles contain dimeric polyadenylated RNA of about 7-10 kb, within a protein capsid. Also present are all the viral proteins required for integration, reverse transcriptase, RNase H, protease and integrase. Once integrated into the host genome the provirus is transcribed and translated by the cellular machinery to yield viral protein products. Viral \textit{gag} protein products package the viral RNA, the envelope is added at the budding stage encoded by viral \textit{env} and once fully assembled the virion can then be exported and infect further cells.

Exogenous retroviruses were previously divided into three subfamilies but are now divided into seven genera as follows: Spumaviruses (Chimp/Bovine syncitial (foamy) virus), Mammalian C type, Avian C-Type, Lentiviruses (SIV, HIV), D-Type, B-Type and BLV-HTLV-like. These divisions are based on amino acid sequence comparison the best fit is obtained by using the \textit{pol} sequences and the worst fit from the \textit{env} sequences alone (Xiong et al. 1990). The \textit{Env} sequence of type D retroviruses is most closely related to those of MLV-like which, as the rest of the type D retroviral genome is similar to B type retroviruses, implies that a recombination occurred to form the D type retroviral branch (Doolittle et al. 1989).

The A to D type classification derives from the particle's morphological appearance under electron microscopy.

A type retroviruses are considered endogenous as the particles are only ever found inside cells. The particles have a clear centre surrounded by a shell and are not infectious. They are often referred to as intracisternal A-type particles (IAP), VL30 is a well known ERV found in murine IAPs. As IAP are only internal they are often called retrotransposons (Kuff et al. 1988; Heidmann et al. 1991).

B type retroviruses have doughnut shaped cores at budding and eccentrically located cores in mature virions e.g. mouse mammary tumour virus.

C type retroviruses have crescent shaped cores at budding and central cores in virions e.g. RSV.

D type retroviruses are more elongated and have electron dense cores in virions e.g. Mason-Pfizer monkey virus.

Provirus related elements have been found in the genomes of many species including humans. In humans intracisternal A particles have been detected (Ono et al.
Many IAP have been observed in mice and certain cell types, e.g. B cells and their tumours (Dalton et al. 1967), as well as embryos. IAPs encoded by retrotransposons have been noted (Kuff et al. 1988).

**Figure 2** Schematic Representation of Virion Structure

1. Surface domain
2. Transmembrane domain
3. Envelope
5. Capsid
6. Nucleoprotein
7. RNA Genome
8. Protease
9. Integrase
10. Reverse Transcriptase
If the retroviral RNA is analysed in more detail, a tRNA is associated to a primer-binding site near the 5' end and is the primer for DNA binding. At the 5' end of the retroviral RNA there is a short direct repeat sequence called the R region, then comes a unique short region called U5. Next there is a primer binding site of 18 bases complementary to 3' of the specific tRNA. There is then an untranslated leader (L) and a signal for packaging (ψ) (Mann et al. 1983; Linial et al. 1990). This region contains the splice site for subgenomic env transcripts. The retroviral genes then follow in the regions gag, pro, pol, env. At the 3' end of the retroviral RNA there is polypurine (PP) AG rich primer for positive strand DNA synthesis during reverse transcription. The U3 unique region is located between PP and R signals to specify and regulate transcription and processing, the 3' R region is a direct copy of the 5' R. In the proviral DNA a long terminal repeat (LTR) is formed at each end of the provirus and corresponds to U3-R-U5, as well as an integration signal. This signal consists of TG...CA direct repeats of 4-6bp flanking the provirus (Varmus 1988; Boeke et al. 1989). A classical RNA polymerase II promoter is encoded within the LTR region so retroelements do not require external promoters for transcription. The elements have developed elaborate reverse transcription that regenerates sequences at both ends of the DNA, resulting in complete LTRs as well as retention of the promoter and so a complete copy.

The gag (for group antigen) encodes the major structural elements of the capsid, composed of 2,000-4,000 units per virion. There are usually three proteins and some retroviruses encode a fourth. The proteins are matrix (MA), which is located on the inner face of the capsid structure and is necessary for budding (Rhee et al. 1990), capsid (CA) which forms the core shell (Dickson et al. 1985) and nucleocapsid (NC) which is associated with the RNA in the core. NC is a basic protein with the zinc finger array Cysteine-Histidine (Katz et al. 1989).

The protease (pro) is situated between gag and pol. It encodes an aspartic protease, which acts functionally as a dimer (Bizub et al. 1991), cleaving both itself and mature gag and pol to yield individual functional proteins. Its three dimensional structure has been resolved (Miller et al. 1989).

The polymerase (pol) encodes enzymes required for the synthesis of viral DNA and proviral integration. These are reverse transcriptase (RT), RNase H and an integrase (IN) with the RT encoded first. RT is a RNA directed DNA polymerase.
RNase H digests the RNA from the RNA-DNA intermediate allowing double stranded DNA to be encoded. The IN trims the 3' end and cleaves the DNA target, enabling ligation and therefore integration (Brown 1990).

The envelope (*env*) region encodes the envelope proteins. *Env* transcripts are formed by splicing to give subgenomic transcripts, in most retroviruses *env* overlaps the 3' region of *pol*. However this is not the case for lentiviruses and spuma viruses. There are two Env proteins formed by cleavage and processed by cellular systems to form a heterodimer. The envelope is associated with the virion budding and facilitates attachment to the host cell and cell entry. The surface (SU) is the larger protein and is glycosylated. It is this that binds to the host-cell receptor and the region against which most neutralising antibodies are reactive. The transmembrane (TM) is a C-terminal cleavage product, it anchors the heterodimer and mediates fusion.

There is an extra region named pX in certain retroviral groups (e.g. HIV and HTLV) which encodes proteins with roles in transcriptional regulation (see chapter 3).

**Figure 3** Genomic structure of a typical retrovirus. A representation of the encoded protein products is also included
Figure 4  An illustration of the life cycle of a typical exogenous retrovirus
Retrovirus-like elements are prevalent in lower organisms and as endogenous retrovirus-like elements in the higher organisms.

Numerous retroelements have been isolated from micro-organisms (Doolittle et al. 1989; Xiong et al. 1990). Some examples are the protozoa retroposons SLACs (MAE) and CRE1 (Gabriel et al. 1990), RIME (ribosomal mobile element) and the INGI/TRS elements. There are also many examples of intron movement via reverse transcription (Levra Juillet et al. 1989). Bacterial retrons have been isolated (Temin 1989) and multi-copy single stranded DNA (msDNA) is further evidence of active reverse transcription. Some examples of retroelements found in lower organisms include the Ty1-4 elements of *S. cerevisiae*, copia of *Drosophila*, Ta1 of *Nicotiana* and Tnt1 of *Arabidopsis*. (Xiong et al. 1990).

1.3 HTLV-I and II

An emphasis is given to the HTLV group of viruses as chapter three is devoted to HTLV I sequence analysis as well as HTLV’s role in ATL.

Human T-cell Lymphotropic Virus type-I (HTLV-I) is an exogenous retrovirus and is grouped with HTLV-II, Simian T-cell Lymphotropic Virus (STLV) and Bovine Leukaemia Virus (BLV) because of their similar molecular structure.

HTLV is morphologically similar in appearance to C-type viruses but is more divergent, the structure of its genome is also somewhat different. By electron microscopy the retroviral particles have a diameter of 110-140nm. An envelope surrounds a spherical core of 80-100nm, which is assembled at budding, this envelope has barely visible projections. After budding the maturing core condenses into a central electron dense structure, caused by the cleavage of the Gag precursor by the protease. The core contains a dimeric RNA and the viral enzymatic complex of RT, protease, RNase H and integrase.

The genome is 9032bp long (Seiki et al. 1983), including LTRs. As well as the normal retroviral genes (*gag, pro, pol, env*) there is also a pX region of 2.1kb with at least four ORFs, encoding tax p40 and rex p27, as well as rex p21 of unknown function (Kiyokawa et al. 1984; Kiyokawa et al. 1985). More recently some other proteins were observed p12, p13 and p30 (or Tof), but these are less well
characterised and not necessary for cell transformation or virus replication (Derse et al. 1997).

**Figure 5** Genomic structure of HTLV-I including all the protein products encoded

HTLV-I is the only retrovirus that has been etiologically linked to human leukaemia, showing close geographical clustering with Adult T-cell Leukaemia (ATL). HTLV-I has low infectivity if any as cell free virus (Sinangil et al. 1985) and requires close cell-to-cell contact for spread. The cell surface receptor has been found to be located on chromosome 17 (Sommerfelt et al. 1988; Gavalchin et al. 1995). Many varied cell types and species appear to have this receptor, but the virus mainly infects CD4⁺ T-cells (Richardson et al. 1990) of monocyte/macrophage origin including dendritic cells (Macatonia et al. 1992). There are subtypes of HTLV divided by sequence differences from I-IV (De et al. 1991; Komurian Pradel et al. 1992; Gessain et al. 1993) for more detail see Chapter 3. ATL is a disease of adults but cases in children have been noted (de Oliveira et al. 1990) and generally there appears to be a 10-40 year latency among infected individuals before the onset of ATL. There have also been cases of ATL without HTLV-I detection but these are very rare (Shimoyama et al. 1986; de Oliveira et al. 1990). In ATL cells the HTLV
genome is often partially deleted but the pX region and the 5' LTR are usually retained suggesting their importance in disease progression. HTLV genomes are usually transcriptionally silent when within ATL cells, with a very low expression of *tax* sometimes detected by RT-PCR (Ohshima et al. 1996). HTLV-I activates primary T-cells to varying degrees, up to inducing growth independent of IL-2.

HTLV-I has also been associated with a slowly progressive neurological disorder named HTLV-I associated myelopathy/tropical spastic paraparesis (HAM/TSP). HAM/TSP patient cultured T-cells show an activated but not transformed phenotype. They show an increased proliferation in IL-2 absence but are dependent on IL-2 and periodic re stimulation or feeder cells for continual growth. HTLV-I signals may replace IL-2 receptor generated signals acting at the G1 phase of growth. HTLV contributes to but is not sufficient for transformation, it appears that Tax is the main factor involved (Grassmann et al. 1989; Tanaka et al. 1990). No chromosome specific abnormality has been associated with ATL although abnormalities are observed.

Other lymphomas have also been HTLV associated including Sézary syndrome and mycosis fungoides (Hall et al. 1991; Zucker Franklin et al. 1992). However these cases could in fact have been ATL as the symptoms are similar. There have been several cases of B-cell chronic lymphoblastic leukaemia in HTLV-I positive adults (Hendriks 1985) but in all of these cases no integration of HTLV was seen in the leukaemia cells. Continuous antigenic stimulation however could lead to proliferation (Mann et al. 1987). There are also cases of HTLV-I positive lymphoma resembling Hodgkin's disease (HD) (Duggan et al. 1988).

HTLV-II was first isolated from a T-cell line derived from a case of hairy cell leukaemia (HCL) (Kalyanaraman et al. 1982). Thereafter it was reported that a similar virus was isolated from a case of B-cell HCL (Rosenblatt et al. 1986) and a coexisting malignant CD8⁺ T-cell clone (containing HTLV-II) (Rosenblatt et al. 1988). HTLV-II was also found in a case of large granular lymphocyte leukaemia (Loughran et al. 1992), in this case it was not clear if HTLV-II was present in the leukaemic cells. There has also been an unconfirmed association with a rare chronic neurodegenerative condition (Hjelle et al. 1992). See Chapter 3 for further detail.
1.4 Human Endogenous Retroviruses (HERVs)

Chapter four of this thesis will focus on the isolation and classification of two novel HERV into their respective families therefore a brief overview of HERV is detailed.

All eukaryote genomes contain multiple copies of DNA related to infectious retroviruses (Coffin 1984). They are stable Mendelian genes and are most likely remnants of prior infections. It is thought ERVs evolved from retrotransposons (Temin 1980; Temin 1992) and therefore it is possible that some ERVs are precursors of infectious forms. ERVs can serve as a pool of genes for exogenous retroviruses to use to produce variants, giving altered host specificity and phenotype. ERVs can encode gene products that compete for or complement in trans retrovirus functions, they can also act as insertional mutagens, changing host gene regulation.

The early work on ERVs was performed on the chicken model and the first studied viruses were the Rous-associated endogenous viruses (ev-1 to ev-16), there was a unique integration site for each (Astrin 1978; Rovigatti et al. 1983). Some could synthesise infectious virions (Crittenden et al. 1983) and others had deleted genomes with subset of proteins (Astrin 1978; Crittenden et al. 1981). It was proposed that ev was involved in ontogeny, but those chickens lacking all ALV-related endogenous viruses were found to be healthy and breeding (Astrin et al. 1979). It was suggested that they gave a selective advantage on an animal exposed to infectious virus, as ev negative chickens were very susceptible to infection (Robinson et al. 1981; Crittenden et al. 1982). But further study found gag, pol, env and LTR related in the correct order in an ev' animal (Dunwiddie et al. 1986). In fact 40-100 copies of retrovirus-related sequences were found in an ev' animal, illustrating that more distantly evolved families were present.

Mouse genomes contain at least four families of endogenous sequences that are related to different classes of exogenous retroviruses (Stoye et al. 1987). Most sequences are defective with no functional gene products. Some mouse strains carry complete proviral DNAs in the germ line and induction can lead to viral spread and even leukaemia (Chattopadhyay et al. 1975; Breindl et al. 1979; Steffen et al. 1979; Chumakov et al. 1982). Most strains also have partially defective ERVs which can help replicating viruses by recombination. Mink cell focus forming viruses (MCF) are
recombinant viruses with expanding host range (Famulari 1983; Evans et al. 1985). ERVs can provide functional DNA sequences to deletion-defective exogenous viruses during replication in NIH/3T3 cells. This repair occurs by weakly homologous recombination (Schwartzberg et al. 1985; Martinelli et al. 1990). ERVs are ubiquitous in nature, they may be important in the biology of their host and in the creation of new phenotypes of exogenous retroviruses.

It has long been recognised that DNA sequences related to retroviruses are present in human genomes (Benveniste et al. 1974) and Martin first isolated these sequences in 1981. Now it is known that up to 10% of the human genome is composed of sequences resulting from reverse transcription (retrosequences) and amongst these retrosequences there are many ERV sequences containing partial or complete ORFs. In humans there are many different groups of ERV named HERVs for human ERV. A classification of these has been proposed by their pol sequence (Callahan 1988), homologies to mammalian type C retroviruses are termed Class I, homologies to type A, B and avian type C are termed Class II. The other well adopted classification is based on tRNA homology to the minus strand primer binding site immediately downstream from 5' LTR (Larsson et al. 1989), some distinctly different HERVs share the same tRNA homology and not all have this short region (18bp). There is also a lack of sequence information for all human tRNAs and as many HERVs were fixed in the genome over 40 million years ago there may be point mutations within this short region.

HERV sequences have been identified in four main ways:
1. Screening by low stringency hybridisation with probes from non-human endogenous and exogenous retroviruses (Martin et al. 1981; O'Connell et al. 1984; Perl et al. 1989),
2. Screening by hybridisation to oligonucleotide probes with the 3'-terminal sequence of selected tRNAs (Harada et al. 1987; Kroger et al. 1987),
3. Analysing flanking regions of unrelated genes (Mager et al. 1984; Maeda 1985) and
4. Direct amplification by PCR using redundant primers for conserved reverse transcriptase sequences (Bangham et al. 1988; Shih et al. 1989).

HERVs possess the same structure as integrated provirus with flanking LTRs of several hundred bases. Full length HERVs have homologous gag, pol and env
sequences. Most HERVs however are defective with multiple termination signals and no infectious HERVs have been found to date.

Various cellular functions of HERVs have been suggested (Temin 1985) -a role in the evolution of the genome as a source of genetic variation, a role increasing the effective target size for lethal mutations, or by acting as insertional mutagens. ERVs may be the origin of pathogenic exogenous viruses by recombination e.g. Harvey and Kirsten murine leukaemia viruses from exogenous murine leukaemia virus and endogenous VL30 elements (Chien et al. 1979; Ellis et al. 1980). HERVs have been associated with human neoplasia by producing gene products that act as transcriptional activators. ERVs in other mammals have been shown to encode proteins that compete in trans with retroviral functions. ERVs have also been implicated in the development of autoimmune disease (Krieg et al. 1992). HERV prevalence and maintenance in the human genome suggests a role in human biology. In fact recent research has indicated a role for a HERV-W encoded protein (syncytin) in forming the syncytiotrophoblast layer in placental development (Mi et al. 2000).

By analogy with other animal models these sequences should be given careful consideration. For further discussion see Chapters 4 and 5.

1.5 Leukaemia

In chapter five of this thesis expression of the isolated HERV sequences will be studied in peripheral blood samples as well as panel of leukaemia patients. Below a brief introduction to the disease and it’s association with retroviruses ingeneral in given.

Until the middle of the 19th century leukaemia was diagnosed as infection, anaemia, or even dropsy. Pathologists were the first to diagnose leukaemia as the presence of more white than red blood cells. The earliest observation was in fact in 1827 by Velpau, whom Virchow later cited in 1847 and he used the term “Leukämie” (Greek for white blood). He was also the first to use the terms “acute” and “chronic” which are still used today. Virchow used the diagnosis of pus in blood as the cause of death. In 1868 Neumann related the disease to the bone marrow and named it splenic
form or leukaemia myelogenous, this type of leukaemia is still known as myelogenous today.

Leukaemia is often referred to as cancer of the blood. The term refers to a group of closely related malignant conditions affecting the blood-forming cells in the bone marrow. In leukaemia normal control mechanisms break down and the bone marrow starts to produce large numbers of abnormal cells of one type, disrupting the normal production of blood cells. Often in leukaemia the spleen and liver will become enlarged and this may cause abdominal discomfort. The large numbers of white cells being produced in leukaemia are all abnormal which means that patients may have frequent, severe infections as well as anaemia.

Leukaemia is classified as either lymphoid or myeloid lineage and as either acute or chronic. Lymphoid and myeloid refer to the type of white cell affected. If this is a lymphocyte or lymphocyte-like cell the condition is called lymphocytic or lymphoblastic leukaemia. Myeloid leukaemias affect any of the granulocytes. The terms acute and chronic refer to the speed at which the disease progresses if it is left untreated. Acute leukaemia comes on quickly and, if not effectively treated, will rapidly progress. Chronic leukaemia is slow to develop and slow to progress, even when not treated.

Leukaemia is diagnosed by whiteness in blood due to leukaemia cells. In acute types the proliferating cells are immature, where as in chronic types the proliferating cells are more mature. The classification is performed by morphology, cytochemical, immunological, cytogenetic and molecular biology techniques. The lineage and degree of maturation of abnormal proliferating cells leads to morphology, the other techniques give more precise sub-typing. The sub class determined by clinical features can lead to differing treatments e.g. different sub-types have different response rates and survival rates. It is also important to gain as much information as possible for communication and co-operation for the development of future treatments.

Myelodysplastic Syndromes (MDS) are primary bone marrow disorders about 20% progress to become blastic leukaemia-they usually affect the elderly.
Various treatments have developed since the diagnosis of leukaemia including the use of chemicals and radioactivity. Paul Ehrlich was the first to use aniline dyes in blood as a treatment and this was the forerunner of modern chemotherapy. In 1865 Lissauer used arsenous oxide as a cure for chronic myelogenous leukaemia which led to the development of chemotherapy for individual leukaemia types. Röentgen rays were discovered in 1895 and used in 1902 by Pusey and in 1903 by Senn on leukaemia patients and led to the development of radiotherapy. In 1930 Gloar used a combination of blood transfusion, radiotherapy and chemotherapy and achieved some success in certain cases. Alkylating agents were developed in the 1940s and are now used in the preparation for bone marrow transplantation. Folic acid antagonists were developed by Farber, in particular pteropterin and aminopterin, which are used to create temporary remission in acute leukaemia of children but are not a permanent cure. This however led to purine and pyrimidine synthesis inhibitors and DNA assembly inhibitors by Hitchings and Elion including 6-mercaptopurine, 6-thioguanine and allopurinol. In the 1960s Cytorabine was introduced. More recently the adenosine analogues, 2-chloroadenosine and fludarabine, which have broad anti-leukaemic effects, were introduced. Some of the bodies natural compounds are also used to fight leukaemia e.g. ACTH (Adenocorticotropic hormone), Adrenal hormones, L-asparagine, vinca alkaloids and antibodies e.g. anthracyclines. In the 1970s cytokines were manufactured using recombinant gene technologies including α-INF for chronic leukaemia, marrow stimulatory factors (Granulocyte colony stimulating factor), immunostimulatory leukokines IL-3, IL-2 and trans retinoic acid (promyleocytes). Combination drug therapy was first used in the 1960 and 1970s and the technique of high dose and rescue has been successful. A significant percentage of acute lymphocytic leukaemias in children can now be cured whilst there are advances in both remission rates and long-term survival in many other leukaemia types. Grafting is another technique commonly employed in leukaemia treatment. This was first investigated when it was observed that living spleen and bone marrow cells could protect lethally irradiated mice. This led to the use of bone marrow grafts; matches are needed to avoid graft versus host disease (GVHD). Chromosome banding has led to the identification of changes central to chronic myelogenous leukaemia, promyelocytic and Burkitt’s lymphoma and also to the identification of
oncogenes, repressor genes etc. PCR has also been a useful tool and will lead to
future developments and further understanding.

As previously mentioned a number of retroviral infections cause cancer, some
of these are highly oncogenic e.g. RSV Avian myeloblastosis virus. Many
retroviruses have been specifically associated with various forms of leukaemia. In
chicken these include avian myeloblastosis virus, avian erythroblastosis virus, avian
myelocytomatosis MC29, mouse Abelson murine leukaemia virus (A-MLV) and both
leukaemia and carcinoma in the case of avian myelocytomatosis and carcinoma virus
MH2. These and others form the avian leukosis-sarcoma group and cause a variety of
leukotic disorders.

All highly oncogenic retroviruses are replication deficient and therefore
require a helper virus, this may be another exogenous yet harmless retrovirus, or an
endogenous retroelement that supplies the missing proteins, usually envelope proteins.
Retroviruses cause lymphoproliferative disease virus of turkeys and
reticuloendotheliosis viruses cause lymphomas and other cancers in birds (Solomon et

There are also leukaemia causing, slowly transforming retroviruses that do not
require oncogene integration and so do not require a helper virus. These endogenous
viruses can create genetic changes and become weakly oncogenic e.g. MCF in akv
mice (Laigret et al. 1988).

In rodents there are many A type retroviruses which can act as helper viruses.
There are also present some B type e.g. murine mammary tumour virus and C type
viruses that effect rodents and can cause leukaemia e.g. MLV.

There have also been some xenotropic retroviruses identified which can cause
disorders in several species (Callahan et al. 1975).

Pike and Walleye fish have a common neoplasia caused by a retrovirus.
Plasmocytoid leukaemia in Chinook salmon is also retrovirus associated (Salmon
Leukaemia virus) (Eaton et al. 1992). Numerous retroviruses are also associated with
spontaneous neoplastic/proliferative diseases of bony fish (Pilcher et al. 1980).

Corn snake CSRV (Lunger et al. 1974) is associated with lymphoid
leukaemia/lymphosarcoma in diverse snakes (Jacobson et al. 1980).
Chapter 1 General Introduction

Feline leukaemia virus (FeLV) is a member of the simple retrovirus class; there are vestigial remnants of a similar virus in human DNA implying there were epidemics in humans in the past (Stoye et al. 1985). There are also clusters of human leukaemia which could be caused by an as yet unknown viral agent (Stoye et al. 1985). FeLV is a good model as it infects a free ranging species causing leukaemia. Upon closer study there are changes to the FeLV genome among the more aggressive forms of the virus, these occur in $env$ and in the LTR enhancer elements (Fulton et al. 1990). Host genes are also found to be inserted in some cases in a similar way to RSV. A recombination with an endogenous virus found in the host germ-line leads to an expanded host range and leukaemic activity. Many of these variants also die out as the prototype still persists within the host.

Bovine leukaemia virus (BLV) causes tumours 1-8 years post infection in small numbers of affected cattle. In 1969 Miller observed viral particles, these were later shown to induce leukosis. BLV has been shown to be an exogenous retrovirus that causes enzootic bovine leukosis (Burny et al. 1985). BLV’s genomic structure has been determined (Sagata et al. 1985) and it is similar in structure and composition to HTLV with differing protein sizes. Transmission occurs as a cell associated retrovirus-it naturally infects cattle, sheep, capybara and water buffalo. In vitro infection has been shown in goats, pigs, rabbits, rhesus monkeys, chimpanzees and buffaloes (Burny et al. 1987). The envelope proteins are gp51 and gp30. The major internal protein is p24, also present are a gag p15, a p12 RNA binding and a p10 protein. Expression seems to be blocked in vivo as with HTLV (Van den Broeke et al. 1988). Using RT-PCR some expression of tax/rex has been observed in neoplastic cells (Jensen et al. 1991).

There has been an association between STLV-I and malignant leukaemia/lymphomas, this may be linked to monoclonal integration of the virus (Tsujimoto et al. 1987). Cells similar to ATLL cells have also been observed in association with STLV-I infection (Sakakibara et al. 1986).

There is suggestive evidence of an infectious agent in childhood leukaemia (Greaves et al. 1993), but no direct evidence yet. There is a higher risk in children of high socio-economic status and amongst the first born or single children. It is thought
not to be result of a widespread infectious agent but may be that delayed exposure to a common infectious agent in infancy could lead to increased risk. Some clusters of leukaemia occur in rural communities that have undergone a migration of town dwellers (Kinlen 1988; Kinlen et al. 1990), supporting this hypothesis, however leukaemia seems to be a rare outcome from infection by a common virus.

These examples raise concerns for human infections. To date the only exogenous retroviruses that have been closely associated with disease are HIV with immuno-suppression in particular AIDS and HTLV-I with ATL and HAM/TSP (see Chapter 3), although HTLV-I and II have both been linked with various other disorders.

What other potential threats lie within the genome in the form of HERVs is not known, but as knowledge of these sequences increases the likelihood of discovery of a disease causing virus increases.
Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Bacterial Culture

Bacterial strains TG2, TG1, DH5α, DH1, Inv10α (Invitrogen).

Luria Broth.

2xTY (16g bacto-tryptone, 10g bacto-yeast extract and 5g NaCl per litre).

SOB (20g bacto-tryptone, 5g bacto-yeast extract, 0.5g NaCl per litre and 2.5mM KCl, 10mM MgCl₂).

SOC (SOB and 20mM glucose).

Agar (15g added to litre of liquid media).

Ampicillin (final concentration 50-100 µg/ml).

Kanamycin (final concentration 25µg/ml).

5-Bromo-4-Chloro-3-Indoyl-β-D-Galactoside (X-gal) (Sigma).

Isopropylthio-β-D-Galactoside (IPTG) (Sigma).

2.1.2 Enzymes

2.1.2.1 Restriction Endonucleases


2.1.2.2 Polymerases

SuperTaq (HT Biotechnologies), Pfu Turbo (Stratagene), Klenow (NEB), T7 RNA polymerase (Promega), Superscript II (Gibco), Sequenase v2.0 (USB).

2.1.2.3 Modifying Enzymes

RNaseA, RNaseH, proteinase K, T4 DNA Ligase (Gibco BRL), Shrimp Alkaline Phosphatase (Amersham Life Sciences), Calf Intestinal Alkaline Phosphatase (Amersham Life Sciences).
2.1.3 Reagents

All chemicals BDH/Sigma or see appropriate method.
Whatmann 3MM blotting paper.
Hybond N nylon membrane (*Amersham*).
Hybond N+ nylon membrane (*Amersham*).
MSI scientific nylon membrane.

2.1.4 Equipment

Uvikon 930 spectrophotometer (*Konitron* instruments).
Transilluminator (*UVP*).
Photography equipment (*Anachem Kodak Digital Science*).
Image capture (*Anachem and Kodak 1D gel analysis software*).
Gene Pulser (*Bio-Rad*).
UV Cross linker, UV Stratalinker 1800 (*Stratagene*).
Film Developer (*Xomat*).
Vacuum Dryer (*Gyrovap*).

2.1.4.1 Centrifuges

MSE microcentaur.
Sorvall RC5B refrigerated centrifuge, SS34 and GS3 rotors.
Heraus Centrifuge.

2.1.4.2 Electrophoresis

Power Supply (*LKB Bromma*).
Sequencing tank (*Amersham*).
HE 99X Max Submarine Unit, flat bed gel tank (*Amersham*).
ABI 377 automated sequencer.

2.1.4.3 PCR

Techne PHC-3 Thermal Cycler.
Perkin-Elmer Gene Amp PCR system 9600.
2.1.4.4 Incubators

HB-1 hybridiser (*Techne*).

G25 Incubator Shaker (*New Brunswick*).

Grant Heat block.

2.1.5 DNA Markers

<table>
<thead>
<tr>
<th>1Kb Marker</th>
<th>λ HindIII digest</th>
<th>φx Hae III digest</th>
<th>BM VI</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(Gibco BRL)</em></td>
<td><em>(Amersham Pharmacia)</em></td>
<td><em>(HT Biotechnology)</em></td>
<td><em>(Roche)</em></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Fragment Size (bp)</th>
<th>12,216</th>
<th>11,798</th>
<th>10,180</th>
<th>7,716</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,108</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4,072</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,054</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,036</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,636</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,018</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>506, 517</td>
<td>448</td>
<td></td>
<td></td>
<td>542</td>
</tr>
</tbody>
</table>

*Hind I fragments of the vector*
2.2 Methods

2.2.1 Peripheral Blood Mononucleocyte Isolation

This protocol is intended for 50-100ml of whole blood. Throughout a Heraus centrifuge with a rotor radius of 15 cm was used.

Sample blood in EDTA or Citrate was centrifuged for 10min at 3,000 rpm at room temperature, the plasma was then harvested and frozen at -20°C. An equal volume of PBS/TNC/BSA* was added and gently mixed. Then carefully 30ml of the mixture was layered above 15 ml of Ficoll 1.077. This mixture was then centrifuged at 2,800 rpm with the break off for 25min. The buffer above the interphase was then removed and the interphase harvested. This was then washed twice with PBS/TNC/BSA by gently re suspending then centrifuging at 1,800 rpm for 7min for each centrifugation. Finally the pellet was washed with PBS and pelleted again at 1,600 rpm for 7min, these mononuclear cells could then be used for RNA (section 2.2.12) or genomic DNA isolation (section 2.2.2).

*PBS/TNC/BSA
80% Phosphate Buffered Saline (PBS).
0.38% w/v Tri-sodium Citrate (TNC).
3% w/v Bovine Serum Albumin (BSA).
Made up in distilled water.

2.2.2 Genomic DNA Isolation

Pelleted cells were washed twice with PBS before re suspension in 10mM Tris at approximately 1x10^7 cells per ml. Next 10 volumes of extraction buffer* was added and the cells gently mixed to give an even suspension, before incubation at 37°C for 30min. Next, cells were lysed by the addition of proteinase K to a final concentration of 100μg/ml and the mixture incubated at 50°C for 2 hr. Following lysis the samples were cooled to room temperature, then extracted with an equal volume of phenol equilibrated to pH 8.0, followed by an extraction with an equal volume of 1:1 phenol:chloroform and finally an extraction with an equal volume of chloroform. The genomic DNA was then precipitated from the aqueous layer using an equal volume of
propanol for 15min. The pellet was formed by centrifugation and washed with 70% ethanol before air-drying and re-suspension in an appropriate volume of water.

*Extraction Buffer*

10mM Tris HCl.
10mM EDTA pH 8.0.
0.5% sodium dodecyl sulphate (SDS).
100µg/ml RNase A.

### 2.2.3 Southern Blot

Southern blots were performed in a similar way to that described in *Maniatis* (Sambrook et al. 1989). Appropriate quantities of digested plasmid or genomic DNA were electrophoresed on an agarose gel of appropriate concentration (0.8%-1.5%). If fragments of greater than 10kb were to be blotted then the gel was first soaked for 15min in 0.25M HCl. The gel was then denatured for 30min in 1.5M NaCl, 0.5M NaOH at room temperature with constant shaking, then neutralised by shaking for 15min in 1.5M NaCl, 0.5M Tris-HCl pH 7.2, 0.001M EDTA. This neutralisation step was repeated using fresh neutralisation solution. Gels were then blotted onto a nylon membrane, either *Hybond-N*, or *MSI*. A capillary blot was set up as described in *Maniatis* (Sambrook et al. 1989), using *10X SSC as transfer buffer and left overnight. After dismantling the blot, the nylon membrane was rinsed in 6X SSC, then dried at 80°C for 60min, before being fixed using a UV cross linker. The membrane was then ready for hybridisation.

Filters were prehybridised in 6xSSC, **5x Denhardt's solution, 0.5% SDS and 100µg/ml denatured salmon sperm DNA using 0.2ml of pre hybridization solution per cm² of filter for at least one hour at 65°C. Hybridisation was for a further 4 to 18 hours in the same solution with the addition of the radioactively labelled DNA probe (see section 2.2.4).

Filters were washed using variations on the following standard protocol, depending on the stringency required. All washes were 100 to 200 ml. One wash of 2xSSC, 0.5% SDS at room temperature for 5min, one wash of 2xSSC, 0.1% SDS at
room temperature for 5min, two washes of 0.1xSSC, 0.5% SDS for 60min at 37°C then 65°C and a final wash of 0.1xSSC for 5min at room temperature were performed. The filter was then wrapped in Saran and exposed against X-ray film at -70°C for the appropriate length of time and developed using an Xomat.

Radiolabelled probes were stripped from the filter by pouring a solution of just boiled 0.1xSSC over the filter and allowing it to cool to room temperature, the effectiveness of the stripping was tested by re-exposure to X-ray film as discussed above.

*20xSSC
3M Sodium Chloride
0.3M Sodium Citrate
pH to 7.0.

**50x Denhardt’s solution
5g Ficoll type 400 (Pharmacia)
5g of BSA
5g of polyvinylpyrrolidone
Water to 500ml.

2.2.4 Radiolabelled Probe Preparation

DNA probes were labelled using the oligolabelling kit (Pharmacia Biotech) as described by the manufacturer. 25-50ng linear heat denatured DNA was incubated with 5U Klenow enzyme in the presence of 50μCi [α-32p] dCTP, 20μM each dNTP (dATP, dGTP, dTTP), supplied buffer and random hexadeoxyribonucleotides at 37°C for 1 hour. The probe was heat denatured prior to use in hybridisation solution.

Riboprobe Preparation

RNA probes were also produced by labelling linearised plasmid preparations. 1μg of linearised plasmid was labelled using the riboprobe Gemini system (Promega) following manufacturer's instructions. T7 RNA polymerase is utilised in this system to
create $^{32}$P labelled dCTP RNA probes. The reaction was left for 2 hours at 37°C for transcription. RNase free DNase 1 was used to remove any DNA by digestion at 37°C for 15min. The probe was then used as for DNA probes.

2.2.5 Plasmid Preparation

Plasmid DNA was extracted following the alkali lysis method as detailed below, or by the use of Qiagen or Wizard plasmid preparation kits, following manufacturers instructions.

An appropriate volume of E. coli cells were grown in appropriate media supplemented with antibiotic (Ampicillin 100μg/ml) overnight.

The following protocol is for a 300-500ml culture the quantities were adapted when smaller culture volumes were used. Throughout a Heraus centrifuge with a rotor radius of 15 cm was used.

The cells were centrifuged at 4,000rpm for 10min at 4°C, then re suspended in 5ml of 10mM EDTA (pH 7.4). 5ml of 0.1M NaOH, 2% SDS was added then cells were boiled for 2-5min. 2.5ml of 3M sodium acetate (pH 4.8) was added, mixed, placed on ice for 5min to precipitate linear DNA and centrifuged for 2min at 6,000rpm. The supernatant was added to 30ml ethanol then placed on ice for at least 10min and then centrifuged for 2min at 6,000rpm. The pellet was washed with 70% ethanol, dried then re suspended in 200μl of water. 40μg of RNase A was added and the tube left at room temperature for 30min. 200μl of 1.6M NaCl, 13% PEG-8000 was added, mixed and centrifuged for 5min at 6,000rpm. The supernatant was removed with a Pasteur, then the pellet was re suspended in 200μl of water. Two Phenol/Chloroform extractions were performed, followed by a Chloroform extraction. The plasmid DNA was then precipitated with 3M sodium acetate and ethanol on ice then centrifuged at 6,000rpm for 5min. The pellet was washed with 70% ethanol, dried then re suspended in 100μl water.
2.2.6 Preparation of Competent Cells

Competent cells were prepared using a calcium chloride or a rubidium chloride method, for transformation by heat shock, with the rubidium prepared cells giving a higher transformation efficiency. Electrocompetent cells were also prepared and used if an even higher efficiency was required.

2.2.6.1 Rubidium Chloride Method

A colony of *E. coli* cells from a fresh plate was picked and grown up in 10-15ml of media for 16-20 hr at 37°C. This was diluted in media by 1:10-1:30 (roughly 5ml: 100ml) and grown at 37°C to an OD$_{600}$ of 0.4-0.6. Then put in a cold 50ml centrifuge tube and placed on ice for 10-15min whilst the centrifuge was cooled to 4°C. Then the tube was centrifuged at 2,000-3,000g for 15min at 4°C. The supernatant was poured off and the tube inverted on filter paper for a few minutes to dry. The cells were then re suspended in RF1* ($\frac{1}{3}$ culture volume- i.e. 30ml) then left for 15min on ice and centrifuged as before. The supernatant was poured off and the pellet re suspended in RF2** ($\frac{1}{12.5}$ culture volume- i.e. 8ml), then left for 15min on ice. The cells were then aliquoted into 400μl aliquots, frozen in liquid N$_2$ and stored at -70°C. 200μl were used for transformations.

*RF1 (100ml)*

- RbCl 1.2g
- MnCl$_2$.4H$_2$O 0.99g
- 1M KAco. (pH 7.5) 3ml
- CaCl$_2$.2H$_2$O 0.15g
- Glycerol 15g
- pH 5.8 with 0.2M acetic acid

**RF2 (50ml)**

- RbCl 60mg
- 0.5M MOPS (pH 6.8) 1ml
- CaCl$_2$.2H$_2$O 550mg
- Glycerol 7.5g
- pH 6.8 with NaOH

Solutions were sterilised by filtration.
2.2.6.1 Calcium Chloride Method

1ml of an overnight culture was grown in 100ml of TY medium at 37°C, until OD$_{600}$ was 0.5, then left on ice for 10min. The cells were then pelleted by centrifugation at 4,000g for 10min at 4°C. The pellet was re suspended in 15ml of 100mM CaCl$_2$ left on ice for 10min, then centrifuged at 6,000g for 6min at 4°C. The pellet was re suspended in 2ml of 100mM CaCl$_2$ and 15% Glycerol. The cells were then aliquoted in 300µl aliquots and used for transformation, surplus aliquots were frozen at -70°C for future use.

2.2.6.3 Transformation

100µl of competent cells were transformed with either 10µl water (control), 9µl water and 1µl DNA, or 9.9µl water and 0.1µl DNA. The transformed cells were left on ice for 40min then heat shocked at 42°C for 2min. 900µl of TY was added to each tube before incubation at 37°C for 1hr. 50 and 100µl aliquots were then plated out onto Ampicillin containing plates (100µg/ml).

2.2.6.4 Electrotransformation of *E. coli* cells

Preparation of Cells.

A stab culture of *E. coli* was grown at 37°C in 2xTY overnight. Approximately 1ml of this culture was then used to inoculate 1 litre of 2xTY. This was grown at 37°C and shaken at 225rpm until an OD$_{600}$ of between 0.5-0.8 was obtained, ensuring that the cells were still growing in log phase. The flask was then chilled on ice for 30min followed by centrifugation at 4,000g for 15min. The cell pellet was then carefully and completely re suspended in 1 litre of cold water, followed by centrifugation as before. The cell pellet was then re suspended in 500ml of cold water and centrifuged as before. Next the pellet was re suspended in 20ml of 10% Glycerol and centrifuged as before. The pellet was then re suspended in 10% Glycerol to a final volume of 2-3ml, in order
to give a cell concentration of at least $3 \times 10^{10}$ cells /ml. Aliquots were frozen in liquid nitrogen and stored at -70°C.

**Electrotransformation.**

The 0.2cm cuvettes were pre chilled on ice for at least 5min and the Gene Pulser set to the 25μF capacitor, 2.5kV and the Pulse Controller Unit to 200Ω. The frozen electroporation cells were gently thawed on ice. DNA was added in 2-3μl to 50μl of cells, mixed and left on ice for 5min. The mixture was then transferred to the pre chilled 0.2cm cuvette. One pulse was applied using the above settings that resulted in a pulse of 12.5kV/cm with a time constant of 4-5msec. Immediately after having applied a pulse 500μl of room temperature SOC medium was added, the cells were then incubated for 60min at 37°C and shaken at 225rpm. The appropriate aliquots were then plated out on the desired plates.

### 2.2.7 PCR

#### 2.2.7.1 HERV PCR

All PCRs were performed for 25-35 cycles with some variation on the standard conditions with a heat shock of 5min at 94°C

Followed by the PCR of:

- 30sec at 94°C
- 1min at 37-50°C
- 1min at 72°C

Then followed by a final extension at 72°C for 7min.

The primers used were at 25-100pmoles per reaction:

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5' RTC RTC NAC RTA YTG 3'</td>
</tr>
<tr>
<td>nA</td>
<td>5' AG RTC RTC CAT RTA NTG 3'</td>
</tr>
<tr>
<td>B</td>
<td>5' AG NAG RTC RTC RTC NAC RTA YTG 3'</td>
</tr>
<tr>
<td>C</td>
<td>5' TN GTN GAY ACN GGN GCN 3'</td>
</tr>
<tr>
<td>C2</td>
<td>5' TN GTN GAY TCN GGN GCN 3'</td>
</tr>
<tr>
<td>MLV</td>
<td>5' YTN KTN GAY CAN GGN GCN SA 3'</td>
</tr>
<tr>
<td>Universal</td>
<td>5' GTK TTN KTN GAY CAN GGN KC 3'</td>
</tr>
</tbody>
</table>
2.2.7.2 HTLV-I PCR

All PCRs were performed for 25-30 cycles with some variation on the standard conditions with a heat shock of 5min at 94°C. Following the PCR of:

1min at 94°C
1min at 55°C
2min at 72°C

Then followed by a final extension at 72°C for 7min.

The primers used were at 25-100pmoles per reaction:

LTR1 5'TGA CAA TGA ACC ATG AGC CCC AAA TAT CCC 3'
LTR2 5'AAT TTC TCT CCT GAG AGT GCT ATA 3'
LTR3 5'TGT GTA CTA NNT TTC TCT CCT 3'
LTR4 5'TTT GAG CGG CCG CTG ACA ATG ACC ATG AGC CCC 3'
LTR5 5'ACT TAG AAT TCC GCA GTT CAG GAG GCA CCA CAG GCG 3'
ENV1 5'CCT CAA TAT TCC GCA GTT CAG GAG GCA CCA CAG GCG 3'
ENV2 5'AGA ACA GGA TAT CAA GGC CT 3'
ENV3 5'TCT AGT CGA CGC TCC AGG 3'
ENV4 5'CGT CTG TTC TGG GCA GCA TAC TGC 3'

2.2.8 Electrophoresis

Electrophoresis grade agarose (Gibco) ranging from 0.6% to 1.5% was dissolved in 1x TAE buffer* by boiling in a microwave oven for 2 minutes. The solution was cooled under running water and ethidium bromide added to a concentration of 1ug/ml. Gels were poured into gel trays fitted with combs and allowed to set for ~25 minutes. Combs were removed and gels were placed in electrophoresis running tanks filled with 1x TAE. Sample DNAs were mixed with 6x DNA loading buffer** and loaded into the wells of the gel. Electrophoresis was performed at an appropriate voltage (30-150v) and the DNA was visualised using an ultra-violet transilluminator and photographed.

* TAE Running buffer: 4.8g Tris base, 1.1ml glacial acetic acid, 2ml 0.5M EDTA (pH8.0) per litre.
**DNA loading buffer; 0.25% bromophenol blue, 0.25% Xylene cyanol FF, 40% (w/v) sucrose in water.**

### 2.2.9 DNA Purification from Agarose Gels

The GeneClean II kit (*BIO 101 Inc.*) is specifically designed to purify DNA larger than 500bp from agarose gels in TAE buffer and was performed according to the manufacturers instructions.

The QIAquick gel extraction kit (*Qiagen*) is designed to extract and purify DNA of 100bp to 10Kb from agarose gels in TAE or TBE buffer and was performed according to the manufacturers instructions.

Both kits provided high yields of pure nucleic acids for direct use in sequencing, cloning and radioactive labelling.

### 2.2.10 Ligation Reactions

Both fragment and vector DNA was digested with the appropriate restriction enzymes to generate compatible ends for cloning. If a single restriction enzyme, or a blunt ended cutter, was used to prepare the vector, the DNA was treated with calf intestinal alkaline phosphatase, according to the manufacturer’s instructions (*Gibco BRL*). This removed 5' phosphate groups and thus prevented recircularisation of the vector during ligation. T4 DNA ligase and T4 DNA ligase buffer were used (*Gibco BRL*) in a final volume of 10-20µl. The reaction was incubated at 15°C overnight.

### 2.2.11 Sequencing

Both automated and manual sequencing were used, using both double and single stranded DNA templates.
2.2.11.1 Single Stranded Template Preparation

Throughout an MSE microcentaur was used at maximum speed. A colony was picked, added to 1ml 2xTY media and grown at 37°C. In the evening 15μl were taken and added to 1.5ml 2xTY including M13K07 phage, then grown for 1 hr. at 37°C. 5-6μl of Kanamycin (25mg/ml) were then added and the cells grown for 14-18 hr. at 37°C. The cells were then pelleted by centrifugation for 10min. 1.4ml of supernatant was removed and placed in another tube, then centrifuged for 10min. 1.3ml of supernatant was then removed and to this 200μl of 20% PEG (MW 6-800), 2.5M NaCl was added. This was mixed and left for 5min at room temperature, then centrifuged for 5min. The supernatant was removed and the pellet centrifuged again. The pellet was eluted in 100μl of water and 100μl of fresh phenol, vortexed, left for 5min at room temperature, vortexed, then centrifuged. To the aqueous phase 12μl of sodium acetate (3M, pH 5.2) and 300μl of cold (-20°C) ethanol were added. The tube was left for 15min at -70°C, centrifuged, then the supernatant was removed and the pellet washed in 70% ethanol. The DNA was then pelleted by centrifugation and dried for 1hr. at room temperature, re suspended in 20μl of water and 5μl checked on a gel. 7μl aliquots were used in sequencing reactions (see section 2.2.11.3).

2.2.11.2 Double Stranded Template Preparation

Double stranded templates required denaturing before sequencing, this was performed using alkali. The DNA was re suspended in 0.2M sodium hydroxide and left at 37°C for 60min. The DNA was then ethanol precipitated with the addition of 5μl of 3M sodium acetate at pH4.8 for every 100μl of ethanol added. A 70% ethanol wash was also performed before re suspending the template in the appropriate volume of sterile water.

2.2.11.3 Manual Sequencing Reactions

Sequencing reactions were performed using the USB sequencing kit and Sequenase® version 2.0 reactions were run on a polyacrylamide gel (section 2.2.11.4).
2.2.11.4 Sequencing Gel

The eared plate was cleaned with acetone then silanated with 2-3ml of dimethyldichlorosilane. The other plate was cleaned with acetone then ethanol, then acetone again. Spacers were inserted and the plates sealed with tape. The gel mix was prepared as below:

<table>
<thead>
<tr>
<th>50ml of 0.5xTBE Gel Mix</th>
<th>50ml of 5xTBE 6% Gel Mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5ml 40% acrylamide</td>
<td>7.5ml 40% acrylamide</td>
</tr>
<tr>
<td>2.5ml 10xTBE</td>
<td>25ml 10xTBE</td>
</tr>
<tr>
<td>23g urea</td>
<td>23g urea</td>
</tr>
<tr>
<td>add water to 50ml</td>
<td>add water to 50ml</td>
</tr>
<tr>
<td>100µl TEMED and 100µl 10%</td>
<td>15µl TEMED and 15µl 10%</td>
</tr>
<tr>
<td>Ammonium persulphate to set.</td>
<td>Ammonium persulphate to set 7.5 ml</td>
</tr>
</tbody>
</table>

Most of the 0.5xTBE gel mix was drawn into 50ml syringe, then 7.5ml was drawn into a 25ml pipette followed by 7.5 ml of the 5x TBE gel mix and a couple of bubbles to give a gradient. The gradient was poured from the pipette down the centre of the plates whilst holding them at about 45°, then keeping the gel continuous, it was continued with the 0.5xTBE gel mix until the gel was complete. The comb was added, the sides of the gel clipped with bulldog clips and the gel covered with Saran wrap. It was then left to set for about 30min. The polyacrylamide gel was run at a power of around 40watts for between 2-8 hours depending on the type of reactions loaded.

2.2.11.5 Automated Sequencing

The Thermo Sequenase dye terminator cycle sequencing kit from Amersham Life Science was used and the manufacturer's instructions followed. 0.2-2 µg of template DNA was used. Samples were loaded on a polyacrylamide gel and run on an ABI 377 automated sequencer and analysed using ABI Prism 100 version 3.0.
2.2.11.6 Sequencing Primers

-40 5’ GTTTTCCCAGTCACGAC 3’
+10 5’ CAGTGTGATGGATATCTGCAGAATT 3’
-10 5’ CCGCCAGTGTGCTGGAA 3’
REV 5’ AGCGAATAACAATTTTCACACAGG 3’

2.2.12 Total RNA Preparation

RNA was isolated using the guanidium thiocyanate method, or using RNAzolB following manufacturer's instructions, purity was tested by measuring optical density at 260 and 280nm.

2.2.12.1 RNAzolB Method

Cells were centrifuged to a pellet then re suspended in RNAzolB (0.2ml per 10^6 cells). Cells were lysed by gently pipetting a few times, 0.1 volume of chloroform was then added and the sample was shaken for 15 sec. The sample was then centrifuged at 12,000g for 15min and the aqueous phase collected, to this an equal volume of propanol was added. The solution was gently mixed and the RNA left to precipitate at 4°C for 15min. The sample was next centrifuged for 15min at 12,000g then washed in 75% ethanol, followed by a centrifugation at 7,500g for 10min. The RNA pellet was then dried under vacuum (Gyrovap) and re suspended in DEPC treated water of an appropriate volume. Purity and concentration were determined by spectrophotometry.

2.2.12.2 Guanidium Thiocyanate Method

For every 1g of homogenized tissue or cells 10ml of lysis solution* was added. Followed by in proportion 1ml of 2M sodium acetate, 10ml of water saturated phenol and 2ml of chloroform/isoamyl alcohol 49:1, this was mixed thoroughly and shaken for at least 10 sec. The mixed solution was then chilled on ice for 30min, followed by a centrifugation at 10,000g for 20min at 4°C. The aqueous layer was then transferred to a fresh tube and 3 volumes of 4M sodium acetate pH7 added, the tube was then left on ice for at least one hour. The sample was then centrifuged as before, then the pellet re-
suspended in 5ml of the lysis solution*. 5ml of isopropanol were then added, mixed
and then left at −20°C for one hour. The sample was then centrifuged as before. The
supernatant was then discarded and the pellet re-suspended in 2ml of lysis solution,
2ml of isopropanol added, mixed and then left at −20°C for one hour. The sample was
then centrifuged as before and the pellet washed with 70% ethanol, centrifuged again
and the pellet dried before re-suspension in water at 65°C if necessary.

*Lysis solution:  Guanidium Thiocyanate 250g
          Autoclaved DEPC treated water 293ml
          0.75M sodium citrate pH7 17.6ml
          10% sarkosyl 26.4ml

Dissolved at 65°C
Before use 360μl of 2-β-mercaptoethanol added to 50ml of stock solution.

2.2.13 Northern Blot

Equal amounts of RNA were ethanol precipitated and re-suspended in 10μl of
suspension media*, heated to 65°C for 15min to denature the RNA, cooled on ice and
then 2μl of RNA gel loading buffer** was added. A 1% agarose gel was used, this was
made by dissolving the agarose in 62.5ml of water, then 17.5ml of formaldehyde and
20ml of 5x gel running buffer*** were added. The gel was set and placed in the
electrophoresis tank, then covered with 1x gel running buffer. The re-suspended RNA
was then loaded and the gel run at 40-150v for the appropriate length of time. As a
denaturing gel was run no further denaturing was required so after running the gel it
was rinsed in water and blotted as with a Southern blot.
The Blot was fixed and hybridised as with a Southern blot (see section 2.2.3).

*Suspension media
4 volumes 5x Gel Running buffer
7 volumes Formaldehyde
20 volumes Deionised Formamide
9 volumes Water.

***5x Gel running buffer
20.6g of 3-(N-Morholino) propane sulphonic acid in 800ml of 50mM
sodium acetate, made to pH 7.
20ml of 0.25M EDTA
Made up to 1l with water.
**RNA Gel loading buffer**
Glycerol 10ml
0.25M EDTA 80μl
Bromophenol blue 50mg
Xylene cyanol 50mg
Water to 20ml.

2.2.14 RT-PCR

A reverse transcriptase reaction was performed on purified RNA for use in PCR.

First Strand cDNA Synthesis was performed using SUPERSCRIPT II (Gibco) for RT-PCR. A 20μl reaction volume was used for 1-5 μg of total RNA or 50-500 ng of mRNA. The following components were added to a nuclease-free tube:-

1 μl of Oligo (dT)12-18mer (500 μg/ml).
1-5 μg of total RNA.
Sterile, distilled water to 12 μl.

The mixture was heated to 70°C for 10min and quickly chilled on ice. Then the following was added:-

4 μl of 5x First Strand Buffer
2 μl of 0.1 M DTT
1 μl of 10 mM dNTP Mix (10 mM each dATP, dGTP, dCTP and dTTP at neutral pH)

The contents of the tube were mixed gently and incubated at 42°C for 2min.

Then 1 μl (200 units) of SUPERSCRIPT II was added and mixed by pipetting gently.

The reaction was incubated for 50min at 42°C then inactivated by heating at 70°C for 15min. The cDNA was then used as a template for PCR amplification.

However, amplification of some PCR targets (those >1 kb) required the removal of RNA complementary to the cDNA. This was removed by the addition of 1μl (2 units) of E. coli RNase H and incubating at 37°C for 20min.

β-Actin Primers:
5’ GTG GGG CGC CCC AGG CAC CA 3’
5’ CTC CTT AAT GTC ACG CAT GAT TTC 3’
Chapter 2 Materials and Methods

HERV1 primers:
5' ATGCCACACTGTCCATCCCCAC 3'
5' GTACTTTAAAGAGGGTGAGT 3'

All PCRs were performed for 25-35 cycles with some variation on the standard conditions with a heat shock of 5min at 94°C
Followed by the PCR of
- 30sec at 94°C
- 1min at 50-60°C
- 1min at 72°C

Then followed by a final extension at 72°C for 7min.
The primers used were at 25-100pmoles per reaction:

2.2.15 PAC DNA Isolation

For PAC DNA isolation a variation on the standard alkaline lysis protocol was used as follows.

A single colony was picked and grown over-night in 10ml of LB plus kanamycin (25µg/ml). This culture was then added to 200ml of LB plus kanamycin and grown at 37°C for 90min, then 0.5mM IPTG was added to stimulate P1 replication and the culture grown for a further 4-5 hours. The bacteria were pelleted by centrifugation and the pellet re-suspended in 12ml of P1* solution. 12ml of P2** solution was then added to lyse the cells. The lysate was incubated on ice until the solution cleared, then 12ml of 3M potassium acetate pH 4.8 was added and the lysate returned to ice for 15min. The pellet was then precipitated by centrifugation at 15,000 rpm in a SS34 rotor at 4°C for 30min. The supernatant was then precipitated with an equal volume of iso-propanol at -20°C for 60min. The precipitate was then pelleted by a similar centrifugation, washed with 70% ethanol and then air-dried. The pellet was re-suspended in 0.1M Tris and incubated over-night with 100µg/ml RNAse A and 50µg/ml RNAse ONE (Merck). Purification of the DNA was then performed by two rounds of phenol chloroform extraction.

*P1 15mM Tris pH 8.0 **P2 0.2M NaOH
10mM EDTA pH 8. 0.1% SDS
100µg/ml RNAse
2.2.16 BAC/PAC Mapping

Probes were labelled by nick translation using biotinylated dUTP. The reaction was incubated for 90 min at 15°C. The reaction was stopped by the addition of 1 μl of 0.5M EDTA pH 8.0 and heated to 65°C for 10 min.

The probe was then ethanol precipitated and re-suspended in 15-30 μl of hybridisation buffer and 1 μg of COT1 DNA was added to each 15 μl of probe suspension. Next the probe was heated in a water bath at 70°C for 10 min, then for 30 min in a water bath 37°C.

Normal *Vysis* target slides were used. The slide was taken from the freezer and allowed to air dry, the position of metaphases were marked with pencil marks. Then the slides were placed in a coplin jar containing 50 ml of 5 ml 20x SSC, 10 ml water and 35 ml of formamide in a water at 72°C for 3 min. Next the slides were plunged into 50 ml of ice cold ethanol in a coplin jar on ice for 3 min. Then the slides were dehydrated as follows: 70% ETOH for 3 min, 90% ETOH for 3 min and 100% ETOH for 3 min and air-dried.

For hybridisation 15 μl of probe was added to each target metaphase and each site covered with a 22x22 coverslip then sealed with cowgum. These slides were placed in a moist dark box and incubated at 37°C for at least 14 hours. Cowgum and coverslip were removed by soaking in 2xSSC.

For detection slides were passed through 4 coplin jars containing two jars of 50% formamide/1xSSC and two jars 0.1xSSC in water bath at 45°C for 5 min in each. Then the slides were placed in a coplin jar containing 50 ml of 4xSST (500 ml of 4xSSC/250 μl TWEEN20). One drop of DAPI mount was placed on each target and covered with a cover slip, this was sealed around the edge with nail varnish and allowed to dry well before analysis by fluorescence microscopy.
Chapter 3: Characterisation of Human T-cell Leukaemia Virus type 1 (HTLV-I) from a Caucasian ATL patient

HTLV-I is the only human retrovirus that has been shown to cause leukaemia, in the form of ATL. ATL however is a disease that is most commonly found in certain geographical areas and is uncommon in Caucasians in general. In this chapter a Caucasian case of ATL was studied from a Greek patient presenting in the UK. The aim of this chapter was to subtype the viral strain by PCR and sequencing in order to determine which strain was responsible for the infection. This was done to determine if it was in fact a novel subtype or a known strain that was capable of causing disease in a European Caucasian. Subtyping was performed by sequencing and phylogenetic analysis of the LTR and Env regions.

3.1 HTLV-I and Disease

3.1.1 Adult T-cell Leukaemia (ATL)

ATL was first discovered as a distinct form of leukaemia in 1977 in Kyoto Japan (Uchiyama et al. 1977). Subsequently it was also found to be endemic in some people of African origin. It is a disease that generally affects adults only; the symptoms include rashes, skin abnormalities, lymphadenopathy and hepatosplenomegaly, increased white blood cell count and abnormal lymphoid cells with polymorphic nuclei in the blood or lymph nodes. Thrombocytopenia and limited bone marrow infiltration are often present, although many cases present with only anaemia. Hypercalcemia is sometimes present and can be the cause of early death (Kiyokawa et al. 1987). There may also be some Central Nervous System (CNS) involvement. Production of parathyroid hormone related protein from ATL cells promotes bone absorption and is the major cause of hypercalcemia (Watanabe et al. 1990). General immune suppression is also observed in ATL patients (Katsuki et al.
1986). On average ATL has a long latency post infection of 20-30 years, but can present earlier (Kawano et al. 1985; Plumelle et al. 1997). Other somatic mutations are probably critical for progression accounting for the long latency (Shimoyama 1991). There is a high correlation of ATL cases with HTLV-I endemic areas (Tajima et al. 1985). In general the affected cells are T helper cells (CD4+), HTLV-I can immortalise T-cells \textit{in vitro} and has been shown to be the causative agent of ATL, accounting for more than 600 cases of ATL in Japan per year (Tajima et al. 1992). 1 in 1,000-3,000 HTLV-I carriers develop ATL in Japan (Tajima et al. 1985), some cases present early with a pre-leukaemic / smouldering state (Yamaguchi et al. 1983).

HTLV-I is monoclonally integrated into the DNA of ATL cells (Yoshida et al. 1984) with intact LTRs and nearly all ATL patients have antibodies raised against HTLV-I proteins. Virus infected cells are required for transfer of virus, as cell free virus has very low infectivity. The modes of transmission are maternal milk (Kinoshita et al. 1984), male to female sexual contact, rarely female to male (Chen et al. 1991) and blood transfusion (Okochi et al. 1984).

HTLV-I infected cells are present in maternal milk (Kinoshita et al. 1984) and vertical transmission from mother to child is the major mode of spread. This is caused by extended breast feeding (Takahashi et al. 1991; Monplaisir et al. 1993; Takezaki et al. 1997) as the IgG protective antibodies decline in the child after approximately 6 months (Takahashi et al. 1991). Around 10-25% of breast fed children of carriers themselves become HTLV-I carriers. \textit{In utero} transmission is debatable and at a very low level if it occurs at all (Katamine et al. 1994). HTLV-I incidence is rare in the cord blood of neonates (Hino et al. 1985) but can be detected more frequently by PCR (Saito et al. 1990). The overall risk of ATL has been linked to infection in early life (Murphy et al. 1989; Tajima et al. 1995), therefore it is important to prevent vertical transmission.

Male to female sexual transmission is far more efficient than female to male (Murphy et al. 1989) as HTLV-I is present in semen mononuclear cells but is absent from the vaginal mucosa. In Japan the infection rate from husband to wife is approximately 60% over 10 years but far lower from wife to husband. Worldwide 2-10% of all carriers go on to develop disease in their lifetime. Overall seroprevalence seems to be low and stable amongst children, increasing gradually with age, especially
in women over 40 in whom there seems to be no consistent explanation for the excess prevalence in older women (Takezaki et al. 1995; Maloney et al. 1998). This is probably not due to increased exposure but may be a reflection of long latency. Other plausible explanations could be an effect of hormonal or immunological changes associated with age leading to an increased susceptibility to the virus or viral proteins.

ATL cells appear similar morphologically to Sézary leukaemia cells with indented or lobulated nuclei by microscopy, hence they are referred to as flower cells. The typical antigenic markers are CD3, CD4, CD8- and CD25 (IL2-Rα) which is secreted in its soluble form (Yasuda et al. 1988). There have been some cases noted in the UK amongst immigrants from the Caribbean (Greaves et al. 1984) but cases are very rare among Caucasians in general. Expression or translation of HTLV antigens is generally absent in fresh ATL cells, but present in vitro transfected lymphocytes, implying some form of expression suppression in vivo (Franchini et al. 1984). ATL cells may escape immunological surveillance if expression is suppressed, however RT-PCR reveals a low level of tax and rex expression in ATL cells at a similar level to healthy carriers (Kinoshita et al. 1989). ATL cells express IL-2R and HLA-DR (Uchiyama et al. 1983), which is mediated by tax (McGuire et al. 1993) and can lead to leukaemogenesis. In one cytogenetic study no unique or specific chromosomal changes were noted in 107 ATL patients analysed, but abnormalities were common and far less frequent in smouldering and chronic forms (Kamada et al. 1992). This data points to the importance of non-specific progression events in the onset of disease and may account for the long latency. Integration is not at a specific site, but is clonal for each individual (Seiki et al. 1984). p53 mutations were evident in 30% of cases (Sakashita et al. 1992) and were more common in the more aggressive forms, implying the multistep nature of ATL progression.

Antibodies to HTLV are found in both carriers and ATL patients. The host immune response to HTLV-I is ineffective in vivo, but neutralising antibodies do work in vitro (Hayami et al. 1984). The vast majority of ATL cells show typical Th (CD4+, CD8-) characteristics but there have been a few cases of CD4+ and CD8+ (Ciminale et al. 2000), or CD4- and CD8+ (Jacobson et al. 1990) infected cells. A decreased expression of TCR in ATL cells is often noted (Matsuoka et al. 1986).
There were some reported cases of ATL that were negative for HTLV-I antibodies (Shimoyama et al. 1986), but these are very rare and may well be a result of ineffective screening (Nagatani et al. 2000).

ATL can be subdivided into four types; acute ATL that is resistant to chemotherapy (Kawano et al. 1985); chronic type that is long term with slow progression; smouldering which has few ATL cells in blood (Yamaguchi et al. 1983) and lymphoma type with no lymphocytosis and <1% circulating abnormal lymphocytes (Kawano et al. 1985). Chronic ATL cells are more uniform in size and nuclei shape. They are smaller than in other ATL types with no azurophilic granules and the nuclei are lobular. Smouldering ATL cases have fewer large cells with no granules or vacuoles and possess lobulated nuclei with coarse strains of chromatin.

ATL is resistant to current anti-leukaemia agents and the main treatment involves combination chemotherapy, which has limited success (Shimoyama et al. 1988). The use of monoclonal antibodies has also been tried. The overall survival median is eight months after onset (Waldmann et al. 1993).

ATL has a particularly high incidence in Southwest Japan where it was first discovered as well as cases presenting worldwide. In the Caribbean ATL was first recognised in 1982 by Catovsky (Catovsky et al. 1982). ATL cases have been reported in all HTLV-I endemic areas (Gessain et al. 1992; Blank et al. 1993; Delaporte et al. 1993; Plumelle et al. 1993; Pombo de Oliveira et al. 1995). Some sporadic cases are also seen in Europe and the USA (Gessain et al. 1990; Soriano et al. 1996; Levine et al. 1999) mainly amongst immigrant populations.

In Japan around half of all ATL patients originate from Kyushu and a quarter from big cities of which most of these patients were born in Kyushu (Tajima 1990). There is a plateau of incidence in males at 50 whereas in females incidence rates increase until 70 years of age. Even so ATL still has a male to female ratio of 3:2. The average age of onset in Japan is 57 (Tajima 1990), whereas in the Caribbean, South America and Africa the average onset age is younger at 40-45 years (Gerard et al. 1995; Hanchard 1996), which implies other co-factors are involved. One suggested co-factor is Strongyloides stercoralis infestation that may delay the onset of leukaemia; hypereosinophilia may be an indication of longer latency (Plumelle et al. 1997). The lifetime cumulative risk of ATL in carriers is estimated at 1-5% in Japan.
and Jamaica (Tajima 1990; Hanchard 1996), while a higher incidence was observed in the Noir-Marron in French Guyana (Gerard et al. 1995).

3.1.2 Tropical Spastic Paraparesis (TSP)

TSP was first described as a myelopathy in 1985 by Gessain and de The when seen in the Caribbean in endemic HTLV-I areas. In Japan a similar disease was described and named HTLV associated myelopathy (HAM) (Osame et al. 1987). They were designated as the same disease and it was referred to as HAM/TSP soon after. HAM/TSP is a slowly progressive symmetrical myelopathy involving pyramidal tracts; high titres of antibody to HTLV-I are generally noted. HAM/TSP patients have 10-100x more HTLV-I provirus than carriers (Kira et al. 1991) and 3x the percentage of polyclonally infected cells (Yoshida et al. 1989). This contrasts with ATL cases that usually have monoclonally infected tumour cells. HAM/TSP has an earlier onset than ATL and can occur within two years of infection especially after HTLV infected blood transfusion (Osame et al. 1987).

The link between HTLV-I and TSP was first described in Martinique in 1985 (Gessain et al. 1985) and confirmed in Jamaica, Colombia and Japan (Vernant et al. 1987). HAM/TSP is more frequent in females with a ratio of 3:2 in Japan and 7:2 in Martinique. It is closely associated with HTLV endemic areas. Incidences vary with geography even between areas with similar seropositivity. For example the lifetime risk of HAM/TSP in Kyushu is 0-0.25% (Kaplan et al. 1990), but in Martinique is 2-7% (Vernant et al. 1987). Incidence rates also vary across ethnic group in the same area e.g. in one large study in Zaire among five ethnic groups, HAM/TSP cases were only found amongst the Ntomba tribe (Jeannel et al. 1993).

HTLV-I has also been associated with many other diseases. These include infective dermatitis in infants and several autoimmune syndromes: idiopathic adult polmyostisis (Morgan et al. 1989), uveitis, Sjörgens syndrome, chronic arthritis (Nishioka et al. 1989), thyroiditis, T-cell bronchioalveolitis and T-lymphocytic alveolitis (Sugimoto et al. 1987; Vernant et al. 1988; Sugimoto et al. 1993). Ophthalmological complications are also seen in some HTLV positive individuals (Goto et al. 1997) as well as acute pancreatitis (Senba et al. 1991) which may be a
result of hypercalcemia as seen in ATL cases. Mycosis fungoides (Zucker Franklin et al. 1992) has been associated with HTLV. There is also a suggested link with multiple sclerosis (MS), but the link is tenuous, some HTLV sequences have been noted by PCR in MS patients (Reddy et al. 1989; Ferrante et al. 1997), but others couldn’t find any (Ehrlich et al. 1991; Myhr et al. 1994).

3.2 The HTLV Group

HTLV-I is classified by DNA sequence similarity and falls into a group of exogenous retroviruses with HTLV-II, Bovine Leukosis Virus (BLV) and Simian T-cell lymphotropic virus (STLV).

HTLV II can be distinguished from HTLV-I by PCR (Tuke et al. 1992; Heredia et al. 1996) or sequence analysis (HTLV II 60% homology with HTLV-I) and also by the use of specific monoclonal antibodies. HTLV-II is prevalent amongst intravenous drug users (IVDU) (De Rossi et al. 1991; Murphy et al. 1998) and is also endemic in Amerindian populations (Biggar et al. 1996), the native Indians of Panama (Lairmore et al. 1990) and new Mexico (Hjelle et al. 1990). HTLV-II has also been found in the Mongolians of Asia so it appears not to be a New World virus as was first thought (Zaninovic et al. 1994). There are two strains of HTLV-II, IIA and IIB that can be differentiated by LTR restriction fragment length polymorphism (RFLP) (Hall et al. 1992) and these can be further sub-divided into IIA0-IIA4 and IIB0-IIB5. The majority (69%) of IVDU infections are of the IIA0 subtype and it appears that the IIA clades evolved from IIB.

The HTLVs share an uneven nucleotide bias, HTLV-I and II both have a high C% of 34.9 and 35.6 respectively and a low G% of 19.0 and 18.2 respectively. No disease has been unequivocally linked to HTLV-II, but there are several associated diseases including chronic fatigue syndrome, with HTLV-II gag sequences detected (DeFreitas et al. 1991). In 1982 Kalyanaraman isolated HTLV-II virus from T-cell line (Mo-T) derived from a Hairy cell leukaemia patient. However there is no evidence for a link to the disease (Rosenblatt et al. 1987).

STLV-I was isolated from the peripheral blood lymphocytes of Japanese Macaques, the virus was first detected by electron microscopy (EM) (Miyoshi et al. 1982). Cell lines from seropositive monkeys and chimpanzees have been established,
genomic blots revealed a HTLV-like virus and RFLP proved that it was non-identical. Sequence analysis revealed a high homology of 90-95% to HTLV-I (Watanabe et al. 1986).

STLV seroprevalence, as with HTLV, shows regional variation. STLV infects many species, including Macaque, orangutang, African green monkey, gorilla, siamang (ape), baboons, chimpanzee and various African and Asian monkeys. New World monkeys appear not to be infected in the wild but can be infected by HTLV in vitro, including marmosets and squirrel monkeys (Kinoshita et al. 1985; Nakamura et al. 1986). Folks in 1994 suggested that there might be some New World monkeys infected with a natural STLV-II.

There has been a strong correlation with malignant leukaemia/lymphomas and STLV-I (Homma et al. 1984; Voevodin et al. 1996; Akari et al. 1998). Sakakibara in 1986 observed cells similar to ATL cells in an African Green monkey (Sakakibara et al. 1986) and STLV specific antigens were seen as well as monoclonal integration of STLV in six cases of leukaemia/lymphoma, polyclonal integration was seen in the non-leukemic infections (Tsujimoto et al. 1987). Other leukaemia/lymphoma cases have also been reported in STLV-I infected primates (McCarthy et al. 1990; Mone et al. 1992; Hubbard et al. 1993). In great apes, natural infection with STLV-I is only seen in Gorillas (Blakeslee et al. 1987). STLV-I can infect human lymphocytes in vitro (Miyoshi et al. 1983; Sommerfelt et al. 1990) but has not been seen in recent times in vivo. There is a very low level of infection by cell free STLV-I with a similar cellular host range to HTLV-I. STLV-I has also been shown to activate resting lymphocytes (Gazzolo et al. 1987; Tozawa et al. 1988).

STLV-I has a long R region in the LTR of 227 bases, this forms a stem-loop structure, the poly(A) signal is 274 bases upstream of the poly(A) site and the three 21 base direct repeats occur 5' of the TATAA signal. There is also a 10 base deletion in U3. The env and pX regions are very similar to those of HTLV-I.

It is thought that HTLV and STLV share a common ancestor- a primate virus (PTLV). Analysis of the differences can yield an estimate of divergence time. It is thought that the virus originated in the Indo-Malay region, then spread to Africa followed by the Americas and Japan (Saksena et al. 1992; Vandamme et al. 1994; Van Brussel et al. 1997).
Sequence comparisons among the virus serotypes indicate relatively recent interspecies virus transfer within the primates (chimpanzee and Baboon) including humans. There appears to have been at least three independent human to simian cross-species viral exchanges (Koralnik et al. 1994; Saksena et al. 1994; Song et al. 1994; Vandamme et al. 1994).

BLV is a good animal model for HTLV, its replication, pathogenesis and general biology are similar and it is common amongst cattle. Experimentally BLV can cause disease in goats, pigs, rabbits, rhesus macaques and chimpanzees, although whole lymphocyte transfer is required for infection (Ferrer 1980). In all 30-35% of infected individuals develop persistent lymphocytosis and only 5% develop clonal tumours. In BLV-associated tumours it is mature B-cells that are the tumour cells in general not T-cells as with HTLV. The disease model differs from the insertional mutagenesis associated with type C retrovirus (Gregoire et al. 1984) and involves the novel pX region in a similar way to HTLV. This region encodes p34 Tax and p18 Rex and these proteins trans regulate the viral LTR promoter and control the switch from small, multiple spliced to full length transcripts. BLV tax can immortalise rat fibroblasts in culture with Ha-ras, causing tumourigenic conversion (Willems et al. 1990), possibly by activating cellular promoters. In BLV-associated disease there is a slow rate of progression. This can be explained by the lack of viral expression within the tumour cells implying that BLV is an inefficient initiator.

3.2.1 HTLV-I structure and classification

ATL was first described by Takatsuki in 1977. In 1981 Hinuma et al. (Hinuma et al. 1981) discovered that the MT-1 cell line, which was derived from a patient with ATL, produced C-type retrovirus and that the viral antigen named Adult T-cell Leukaemia associated antigen (ATLA) was involved with ATL. At about the same time Miyoshi et al. reported that the retrovirus from the MT-1 cells transformed human cord T-cells giving rise to the MT-2 T-cell line (Miyoshi et al. 1981). In 1982 Yoshida isolated virus from MT-1 and MT-2 cells and named it Adult T-cell Leukaemia/Lymphoma Virus (ATLV), which Seiki then sequenced in 1983. The term HTLV was given (Poiesz et al. 1980) to a C-type retrovirus that was found in a T-cell
line that was established by Gazdar et al. (Gazdar et al. 1980) from an African patient with T-cell lymphoma (mycosis fungoides). Later these viruses were found to be the same as ATLV (Watanabe et al. 1984) and thereafter the virus was generally referred to as HTLV-I. It is possible that these lymphomas were in fact ATL as they were clinically similar (Bunn et al. 1983).

The HTLV-I genome is 9032bp long (Seiki et al. 1983) and possesses two LTRs. The structure of the HTLV genome is LTR-gag-pol-env-pX-LTR, the additional pX region separates HTLV from that of a typical retrovirus.

The LTR is 755 nt. long and is composed of U3, R and U5 of 352, 227 and 176 nt. respectively. U3 contains a TATA box and three imperfect 21nt. repeat elements called Tax responsive elements (TRE) that control proviral expression (Beimling et al. 1990). These are found at -251/-231, -203/-183 and -103/-83, and all contain a central 5'-TGACGTCA-3' which is a cAMP response element (Beimling et al. 1990). There is another Tax target of 47 nt. found at -163/-117 (Tanimura et al. 1993). Many cellular transcription factors bind the U3 LTR. U3 also contains termination and polyadenylation signals. The polyadenylation signal of AATAAAA is not 10-30 nt. upstream of the polyadenylation site as in most retroviruses, but is more than 250 nt. downstream. It is brought close by secondary structure folding resulting in the Rex-responsive element (RRE), the site of action of Rex (Felber et al. 1989). The R region is unusually long and forms most of the RRE. It also contains the capping site and the first splice donor site at nt. 119 of the mRNAs. The U5 function is not totally determined. The last 100 nt. contain a cis-acting element which represses expression of genomic mRNA but not the spliced mRNA, acting as an antagonist to Rex (Black et al. 1991; King et al. 1998).

The group antigen (gag) region encodes the core proteins. It is 8.5kb unspliced and encodes a 53-55kD precursor. This protein is cleaved by viral protease into matrix (p19), capsid (p24) and nucleocapsid (p15), of 130, 214 and 85 amino acids respectively (Nam et al. 1988), p19 is post-transcriptionally modified with the addition of myristic acid to the N terminus.

The protease region encodes a 234 amino acid protein of 14kD. This is encoded in a different reading frame to Gag. It is encoded from within the 3' end of
gag and extends into the 5' end of pol by a ribosomal frameshift (Nam et al. 1988). The protease undergoes self-maturation by catalytic self-cleavage (Louis et al. 1999).

The polymerase region encodes a 95kD protein, the largest ORF being a potential 896 amino acids formed by a second ribosomal frame shift (Nam et al. 1993). It is an Mg$^{2+}$ dependent reverse transcriptase encoded by the 5' end of pol, the RNase H and integrase are encoded further downstream.

The envelope region (env) starts on an alternate ORF that overlaps the 3' of pol. A 4.2kb single spliced mRNA encodes a 488 amino acid precursor, which is later glycosylated and is 61-68kD. It is then cleaved into a 46kD surface glycoprotein of 312 aa and a 21kD transmembrane glycoprotein of 176 aa. The N terminal of gp21 has a hydrophobic domain of 29 amino acids, the C terminal contains the transmembrane domain. Point mutations on env and the glycosylation sites disrupt envelope processing or function (Pique et al. 1990). There is only limited genetic variability in the env region between subtypes (Schulz et al. 1991) and this region may be used for classification.

The pX Region is a unique 2.1kb region situated to the 3' of env. At least four ORFs are present, ORFs III and IV are from a single polycistronic doubly spliced mRNA. This polycistronic mRNA encodes the two major proteins Tax and Rex. Tax controls transcription of viral genes (Giam et al. 1989) and some cellular genes, e.g. IL-2 receptor expression (Ruben et al. 1988). At first cellular factors control HTLV gene expression, tax is favoured over rex because of sequences surrounding the initiation codons. Later tax mediated trans-activation occurs and rex expression increases slowly to a threshold, this leads to high levels of gag/pol and env and entry to the late phase of replication and then virion assembly. Overabundance of rex seems to have an autocrine repression of viral transcription. This reduced amount of Tax and Rex protein may lead to evasion of immunosurveillance.

ORF IV encodes p40tax that is a nuclear phosphoprotein of 353 amino acids, it has the same initiation codon as the envelope but is spliced to ORF IV at nt. 6950 extending into U3 of the 3' LTR. p40tax is called a trans activator protein as it acts in trans at TREs in the U3 of the 5' LTR. It also has indirect actions mediated by cellular factors, e.g. CREB and can activate transcription factors. c-fos is also activated by tax (Alexandre et al. 1991). Tax also enhances expression of cellular genes via NFκB, including IL-2, IL-2Rα (Ruben et al. 1988), as well as GM-CSF
(Nimer 1991). These factors play a role in T-cell growth and function and imply that tax is important in the transformation of T-lymphocytes. Other genes that are activated include c-sis, c-jun, NGF, IL-6, TNF-β, globin and vimentin genes and parathyroid hormone related protein. Tax has a negative action on the β-polymerase gene, implicated in DNA repair.

The p27rex nucleolar phosphoprotein of 189 amino acids has a nucleolar targeting sequence of 19 arginine rich amino acids at its N terminus. It is encoded from the same mRNA as Tax, with initiation 56 nt. upstream on the second exon, in frame with ORF III. Rex acts post transcriptionally to regulate viral gene expression (Seiki et al. 1988). Rex increases viral expression by facilitating passage into the cytoplasm of mRNA for gag, pol and env, without Rex these stay in the nucleus. Rex is dependant on cis-acting sequences (Seiki et al. 1988) in the R region of the 3' LTR, i.e. the RRE. There are similar regions in BLV and HTLV-II and HIV-1 (Felber et al. 1989). Rex is functionally similar to HIV-1 Rev being a regulator of viral expression in the same manner (Ahmed et al. 1990).

Tax is preferentially translated over Rex as a more favourable initiator surrounds its AUG. In fact the tax initiator is surrounded by one of the optimum sequences of caccATGg, whereas rex has a poor initiator in particular having a T at the important -3 position (cTgcATGc) (Kozak 1986). Stimulation by Tax leads to increased expression of Rex, which leads to viral mRNA export to the cytoplasm allowing viral assembly and budding. Rex directly inhibits doubly spliced mRNA coding for Tax and Rex and may explain viral latency.

ORF III also encodes a p21rex of 111 amino acids that is found in the cytoplasm. It is the C terminal of Rex and is of unknown function (Koralnik et al. 1993), but may act as a trans dominant inhibitor of Rex.

ORF I and ORF II of pX are novel spliced mRNAs. They have been found by PCR and non-PCR methods in ATL patients, cell lines and HAM/TSP patients (Berneman et al. 1992; Cereseto et al. 1997). These transcripts are also found by RNase protection assay in ATL and HAM/TSP lines and PBMCs from ATL patients, so they are not PCR artefacts (Cereseto et al. 1997). Alternative splicing as with HIV, BLV and HTLV-II leads to the increased complexity of these viruses and allows the small genomes to encode more genes (Berneman et al. 1992; Cereseto et al. 1997).
Three further proteins encoded by pX are p12, p13 and p30. p12 is encoded on pX ORF I and is a 99 amino acid, 12 kD proline and leucine rich hydrophobic protein, found in cellular endomembranes (Koralnik et al. 1993). It bears structural and functional similarities to E5 of bovine papillomavirus; both bind a 16 kD part of cellular H⁺-vacuolar ATPase (Franchini et al. 1993). p12 can co-operate with E5 in a cell transformation assay (Franchini et al. 1993) and it has very deletion sensitive activity which is not due to binding (Koralnik et al. 1995). p12 also interacts directly with IL-2 R β and γ but not α (Mulloy et al. 1996) and may have a role in T-cell transformation (Mulloy et al. 1996). p12 may be involved in persistent viral infection (Collins et al. 1998) but is not necessary for immortalisation or viral replication in primary cells in vitro (Derse et al. 1997; Robek et al. 1998).

ORF II encodes two nuclear proteins of 13 and 30 kD, p13 is 87 amino acids and is located in the nuclear matrix (Koralnik et al. 1993), p30 (Tof) is found in the nucleoli (Koralnik et al. 1993). Their functions are not yet defined (Green et al. 1995), but neither gene is necessary for immortalisation or viral replication (Derse et al. 1997). However both p13 and p30 do seem to be expressed in chronic infections and are targets for immune response, as demonstrated by a CD8+ CTL response (Pique et al. 2000). This data supports the theory of a role for the proteins in a long-term chronic infection.

The HTLV group also have a region of 650 nt. between the 3’ of env and the second exon of tax/rex- it is unique to this family and is not necessary for HTLV-II replication (Green et al. 1995). This region was thought to be non-coding, however in HTLV-I ORF I and II of tax/rex have overlapping reading frames, with mRNAs seen in vivo (Cereseto et al. 1997), so now the non-translated region is reduced to only about 100 nt. A similar protein to p12, in HTLV-II has also been observed called p10 (Ciminale et al. 1996).
In ATL cases the integrated HTLV provirus is often partially deleted, but the pX and 3' LTR regions are preferentially retained implying an importance in disease progression (Korber et al. 1991).
3.3 Epidemiology

The endemic areas for HTLV-I are mainly the Caribbean, Southern Japan and certain regions in sub-Saharan Africa. HTLV-I is an old virus that has undergone vertical transmission via mother's milk and the infection of umbilical cord T-cells. Horizontal transmission as cell free virus probably never takes place. However transmission occurs with infected lymphocytes by male to female sexual contact. In the vagina there is an absence of CD4+ T-cells except during menstruation so there is little or no HTLV, therefore female to male transmission is very rare. The other main mode of transmission is via blood products in in vitro drug users (IVDU) and from blood transfusions.

Worldwide 11-20 x10^6 individuals are infected and disease in the form of ATL or HAM/TSP occurs in less than 5% of all infected individuals. As mentioned the main endemic areas are Southern Japan, intertropical Africa and the Caribbean basin (Blattner et al. 1982; Kayembe et al. 1990). There are also many minor endemic areas notably South America and the Seychelles (Arango et al. 1988; Lavanchy et al. 1991).

There is also a massive potential for further spread as there are only pockets of infected regions, an example of this is the more recent spread to North America. North and South American aboriginal groups may have carried HTLV-I across the Bering Strait when the Americas were colonised by humans.

Different strains of HTLV-I were noted in Melanesia and Australia (Gessain et al. 1991; Bastian et al. 1993) by PCR analysis. Worldwide three subtypes were recognised (Koralnik et al. 1994); the Cosmopolitan type found in many regions (Gessain et al. 1992); the Central African type found in the Zaire and Gabon (Fukasawa et al. 1987) and the Austro-Melanesian type found in Papua New Guinea, Australia and Melanesia that is the most divergent (Gessain et al. 1991; Bastian et al. 1993). A fourth subtype has now been reported in Central Africa and Pygmies that is closely related to Chimpanzee STLV-1 (Mahieux et al. 1997).

The Cosmopolitan type (Ia) is very homogeneous and probably originated in West Africa, later being transported by the slave trades and overland trade routes (Gessain et al. 1992). There are 3 subtypes within the Cosmopolitan group named A to C. B is mainly found in Japan (also India and British Colombian American.
Indians) and C is found mainly in West Africa and the Caribbean. A is widely dispersed and accounts for most of the genetic diversity (Miura et al. 1994). The virus was probably spread to the Caribbean from the slave trade in West Africa, as they are closely related. Two separate migrations to Japan may explain the presence of two sub-types (Vidal et al. 1994). In British Columbia the virus was endemic pre-European colonisation but recent travel, e.g. Japanese fishermen, or African descendants from American slaves may have led to its spread (Picard et al. 1995). Hinuma in 1986 proposed HTLV-I was transmitted amongst the Japanese since prehistoric times, this is supported by infection of the Ainu from whom the Japanese are thought to have descended (Ishida et al. 1985).

The Zairian type (lb) is generally confined to Central Africa (Gessain et al. 1992). There have however been cases reported from elsewhere e.g. a case in Sicily with this subtype (Boeri et al. 1995).

The Melanesian type (lc) was isolated from the remote areas of Papua New Guinea, the Solomon Islands and later from Australian aborigines (Gessian et al. 1991).

The proposed Central African and pygmy type (ld) contains members that are not closely related to any other African types but are closely (98%) related to Chimpanzee STLV-1 indicating recent interspecies transmission (Mahieux et al. 1997).

The Melanesian and Cosmopolitan types are from a common ancestor, so the virus probably reached Australia in ancient times as the Melanesian type shows the earliest divergence.

RFLP of LTR sequences has led to 12 profiles divided into 5 groups by geography: Cosmopolitan; Central African; Western African; Japanese and Melanesian. These are confirmed by sequence analysis of LTR (Vidal et al. 1994). The Cosmopolitan type is found in most regions, the Japanese type is most closely related to the Cosmopolitan and is found mainly in Southern Japan. This type accounts for 75% of carriers in Japan the other 25% being Cosmopolitan (Vidal et al. 1994). The Central African strains show the largest interdiversity, possibly from genetic drift in each country, but may also be the result of interspecies transmission including the infection of humans with STLV-I (Koralnik et al. 1994). The
Melanesian type is molecularly quite distant from the four others due to genetic drift over 10,000-40,000 years in the Indo-Malay region (Gessain et al. 1993).

There are small clusters of HTLV positive individuals worldwide but these are not ubiquitous (Levine et al. 1988). Highly endemic areas are the South West islands of Japan, the Caribbean and surrounding area, South America, intertropical Africa, the Middle East (Iran) and Melanesia. Some cases have been reported in Taiwan (Wang et al. 1988), India (Babu et al. 1993), China (Pan et al. 1991) and Korea (Lee et al. 1986). Incidences have been noted in the Arctic population (Robert Guroff et al. 1985), but the virus does not appear to be endemic. Africa has by far the largest reservoir with 5-10 x 10⁶ infected individuals including Senegal, Guinea Bissau, Guinea, Mali, Burkina-Fasso, Benin, Tchad, the Ivory Coast, Ghana and Nigeria in West Africa (Hunsmann et al. 1984; Saxinger et al. 1984; Verdier et al. 1989; Dumas et al. 1991). Gabon, Zaire, Cameroon, Central African Republic, Congo and Equatorial Guinea are also affected (Delaporte et al. 1989; Schrijvers et al. 1991; Garin et al. 1994; Tuppin et al. 1996). There is a high prevalence in southern Gabon (Schrijvers et al. 1991) and in northern Zaire (Garin et al. 1994) and also occasional clusters in the Seychelles (Schrijvers et al. 1991).

In the Americas there is a highly endemic population on the Caribbean islands of Haiti, Jamaica, Martinique and Trinidad (Blattner et al. 1990; Frery et al. 1991; Maloney et al. 1991; Allain et al. 1992), South American areas also have endemic regions, e.g. Tumaco, Colombia (Trujillo et al. 1992) and the Noir-Marron population of French Guyana (Gerard et al. 1995), but there are only low levels in the rest of South America. USA and Canada have a low number of carriers and there are clusters in recent immigrant populations and amongst the black population in particular (Williams et al. 1988).

In the Middle East, the Mashad region in Northern Iran is a reservoir of HTLV carriers (Nerurkar et al. 1995), emigrants from this region are now in Israel, USA and Italy. There are also sporadic cases of HTLV-I in Iraq (Denic et al. 1990) and Kuwait (Voevodin et al. 1995).

On the Pacific islands and in Melanesia there are isolated clusters of HTLV as well as in two tribes of PNG, the Solomon islands (Asher et al. 1988) and in Australian Aborigines (Bastian et al. 1993), incidence is rare on other islands (Asher et al. 1988).
The virus is not endemic in Europe, but there have been some cases in Italy (Pandolfi et al. 1985; Chironna et al. 1996). Most cases in Europe arise in immigrants from the West Indies, Africa or the Middle East (Cruickshank et al. 1989; Wyld et al. 1990; Taylor 1996). There have also been some sporadic cases in Greece, in some eastern European countries (Taylor 1996) and in Georgia (Senjuta et al. 1991). There has also been a recent report of a case of ATL in Germany without relation to an endemic area (Ellerbrok et al. 1998). Several HTLV-I positive individuals have been noted throughout Europe (Taylor 1996), including Switzerland (de Saussure et al. 1991) and Germany. In England the seroprevalence is low, but most studies have been on blood donors which automatically removes the highest risk individuals, from antenatal studies the seroprevalence may be up to 100 times higher than that first thought (Nightingale et al. 1993; Taylor 1996).

This chapter details a case of acute ATL presenting in Europe without any indications of links to a known HTLV-I endemic area.

**Figure 3.2** World Map illustrating HTLV-I endemic countries

HTLV shows a founder effect and high transmission within cultural local communities. It also shows marked clustering according to ethnic background as shown in Trinidad. Amongst 1025 people of African descent 3.2% were HTLV-I
infected. Of 487 people of Asian descent only 0.2% were infected and 11.4% of those of African descent infected were among the Africans in the coastal villages of Tobago implying that socio-economic and environmental co-factors affect the rate of transmission (Blattner et al. 1990). In particular this can be partially explained by the means of spread (mother to child and sexual contact) and the tendency for communities to mix mainly within their own cultural groups.

HTLV shows low sequence divergence even amongst geographically distant populations. By using envelope gp46 and LTR sequence the greatest variation is seen, but the amount varies with the length of sequence used (Nerurkar et al. 1995) - the greatest variation being located in the U3 and U5 regions of the LTR. Daughter in parent effect also has to be taken into account when assessing sequence differences due to the low mutation rates seen in HTLV.

PCR analysis of HTLV, has illustrated that inter and intrastrain genetic variability is very low (Vidal et al. 1994). There are geographic links to specific changes, mainly substitutions, but these are not disease linked (Komurian et al. 1991; Gonzalez Dunia et al. 1993; Nerurkar et al. 1994; Georges Gobinet et al. 1998).

The results of tests on sera from tropical areas tested by western blot analysis have been questioned (Gessain et al. 1995). Previous studies may have overestimated the seroprevalence of HTLV. The WHO and CDC have now redefined a positive result to contain at least one gag-encoded (p19, p24) and at least one env-encoded (gp21, gp46) response. The new generation kits are also more specific.

The seropositivity of the patient in this case was confirmed to have both gag and env positive antibody response. The patient's sera also reacted with the Japanese T-cell lines (MT-1 and MT-2) that are chronically infected with HTLV-1 as well as with the T-cell line Karpas 1010 that was transformed by a virus that was isolated from a Caribbean patient with ATL.
3.4 Case History

The patient referred to in this chapter was a 36-year-old Caucasian European from Greece who came to the UK for treatment. She first presented whilst pregnant, complaining of a persistent cough, a raised lymphocyte count of $30 \times 10^9/l$ was noted, of which a proportion were abnormal. No clinical, radiological or biochemical abnormalities were noted. A preliminary diagnosis of Sézary syndrome was made and deoxycoformycin followed by chlorambucil administered with no improvement in her white cell count. Repeated episodes of bronchitis type symptoms and loin pain then followed. Computed tomography (CT) scans revealed extensive pulmonary infiltration, enlarged intra-abdominal lymph nodes and bilaterally enlarged kidneys, with a renal biopsy showing infiltration by T-cell lymphoma. Two courses of intravenous combination chemotherapy were given (PROMACE protocol). This cleared the renal and intra-abdominal disease but had no effect on either the pulmonary or bone marrow infiltrates. A subsequent blood count showed haemoglobin 12.0g/l, platelets $248 \times 10^9/l$ and a white cell count of $13.6 \times 10^9/l$ of which $6.5 \times 10^9/l$ were atypical lymphoid cells. Bone marrow aspirate showed 40% malignant lymphoid cells and a subsequent CT scan showed extensive pulmonary infiltration, a large paracardiac lymph node but no intra-abdominal disease. The patient died of the malignancy within 2 years of presentation.

In this laboratory Dr. Karpas isolated PBMC from the patient for culture. Three flasks were seeded one with only the patient's PBMC and the other two co-cultivated with cord white blood cells (w.b.c.) from two different umbilical cords (Karpas personal communication). The two co-cultivated samples continued to proliferate and were examined by EM. These showed HTLV-like particles (Karpas personal communication) and most cells in these cultures were also shown to express a high level of viral antigens, which were detected by patient sera and monoclonal antibodies raised against Gag protein of a Japanese isolate of HTLV-I (Karpas personal communication).
3.5 Results

3.5.1 Southern Blot Analysis

DNA was isolated from the MT-2 cell line which was transformed by a Japanese strain of HTLV-1 (Miyoshi et al. 1982), a cell line transformed by a strain of HTLV-1 isolated from an ATL patient of Caribbean origin designated Karpas 1010 (Malik et al. 1988) as well as from the cord T-cells transformed by HTLV-1 from a Caucasian case of ATL designated Karpas 1682. These DNA preparations were digested with several restriction enzymes then run out on a 1% agarose gel before Southern blotting as described (section 2.2.3). The resultant blot was then probed using $^{32}$P radioactively labelled HTLV-I DNA from the Hs35 subtype of HTLV-1 plasmid clone (Malik et al. 1988). Labelling was performed as described (section 2.2.4).

After digestion with restriction endonuclease EcoRI that does not cut within the prototypic HTLV-1 proviral sequence, a single faint band was observed within the 1682 lane (lane 5). This contrasts with MT-2, the HTLV-1 producing cell line, which has many integrated copies and produces a ladder under the same conditions (lane 4). The one band observed for 1682 was characteristic of integrated HTLV-1 and was of a height around 10kb. This indicates a clonal infection at one integration site which is the expected pattern for an ATL case (Yoshida et al. 1984). The specificity of the probe is indicated by a lack of hybridisation to a pool of similarly digested donor genomic DNA, again isolated from w.b.c. (lane 3). A small amount of HS35 plasmid was also run on the gel as a positive control (lane 2), see figure 3.3.
Figure 3.3 HTLV positive Southern blot

A: 1% agarose gel of genomic DNA digestions prior to blotting, lanes are as in B.

B: Genomic blot probed with radioactively labelled HS35. Lanes 1 and 6 1Kb ladder, lane 2 HS35 positive control, lane 3 normal donor genomic DNA, lane 4 MT2 genomic DNA, lane 5 1682 genomic DNA. All samples were digested with EcoRI.

The Southern blot shown above (Figure 3.3) illustrates integrated HTLV-I within 1682 genomic DNA. Lane 5 shows one clear band indicative of integrated virus, the linear positive control plasmid is also indicated (lane 2).
3.5.2 PCR Analysis of LTR

In order to type the virus more accurately RFLP was performed using restriction endonucleases Apal, NdeI, SacI and DraI. This resultant data was ambiguous due to the very faint nature of the bands on the Southern blot. Therefore the next step of PCR amplification was undertaken. The PCR primers were selected to amplify the LTR and ENV regions of the virus as these areas are shown to harbour the greatest sequence variation (Nerurkar et al. 1995).

PCR was performed using several sets of primers until the primer set that amplified the whole of the LTR region was chosen i.e. primers LTR 1 and LTR 4. The PCR conditions and primer sequences are given in the materials and methods (section 2.2.7.2). A band of expected size 750-760 bp was observed for the entire LTR region, illustrated in figure 3.4 below. The positive control plasmid HS35 was also PCR amplified as well as genomic DNA from the cell line MT-2. These also showed bands of a similar height corresponding to the entire LTR region. No band was observed when these primers were used on normal genomic DNA isolated from blood donors screened for HTLV antibodies, this is shown as the negative control and demonstrates the primer specificity.

Figure 3.4 PCR products from 1682 for both the complete LTR region and partial env region. Lane 1 BM VI size marker (Roche), lanes 2-5 LTR 1 and 4 primers, lanes 6-9 ENV 1 and ENV4 primers, lanes 2 and 6 negative controls (normal genomic DNA from PBL), lanes 3 and 7 HS35, lanes 4 and 8 MT-2, lanes 5 and 9 1682.
The band was then gel purified (section 2.2.9) and sequenced using the automated method described (section 2.2.11.5). The resultant sequences are given (Figure 3.5).

**Figure 3.5** Complete LTR sequence of 1682, bases differing from ATK-1 are shaded

1 U3 60
TGACAATGACCATGACCCCACAATATCCCCCAGGGCTTAGAGCCTCCCCAGTGAAAACA

61 U3 120
TTTCCGGAACAAGAAGTCTGAAAGGTACGGGCCCAGACTAGGCTCTGACGTCTCCCC

121 U3 180
CCGGAGGGCAGCTAGCCAACGGCTAGGCCCCTAGCTGATGCCCCCTGAGAGACAAA

181 U3 240
TCATAAGCTCGACCTCCGGAGGGAACCCCGCCCTTCCTCAAAACCTACTGTGCTGACAGCAT

301 U3 300
AGCCGCCCTCAGGCGTTGTACGACAAACCCTACTCACTCAAAAAATCTTTCTTAGCCAGCAT

361 R 420
ATGGCTGAATAAACCAACAGGATCTATATAAAGCCTGGAGAGTTACAGGAGGGGCTCG

541 R 600
AGAAGCTAGTACGGATCCGACGCTCCCTGTCGACCCAGCTCCCTGCTAGCTAGCTAG

601 U5 660
CCGGCTCTCCAGGCTCTGCTGACCTGCTGACCTGCTGACCTGCTGACCTGCTGACCTG

661 U5 720
CCGGCTGCGGCACTCGTGACGAGAGGGCTGCGGAGGCCAGGACAGCACAGCCA

721 U5 758
TCTATAGCAGTCTCCAGGAGAGAAGAGGGCTTAGTACACA

Following homology searches using BLAST (Altschul et al.) it was clear that this sequence was in fact very similar to other LTR regions from HTLV-1. In order to subtype the particular isolate several alignments were performed and an example is presented below (Figure 3.6).
Figure 3.6: Sequence alignment of LTR region. The isolated sequence is indicated in bold (1682). The prototypic sequence ATK-1 is given across the top. Matching bases are denoted with a dash (-). Deletions for alignment are denoted by (*). Details of sequence abbreviations are given in a separate table.
<table>
<thead>
<tr>
<th>Sequence</th>
<th>351</th>
<th>420</th>
<th>491</th>
<th>560</th>
<th>631</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATK-1</td>
<td></td>
<td>GAGGGGGCTGGCATCTCTCCTTCAACGCCGCCGCCCTACCTGAGGCCGCCATCCACGCCGTTGAGTG</td>
<td></td>
<td></td>
<td>CACCCCTTCCCTTCAATCCGACTGACGCTGCCGCTG</td>
</tr>
<tr>
<td>1682</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QU2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT IS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyg19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEL5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP-1</td>
<td></td>
<td>GCGTCTGCCGCTGGCTGGCTCCTTCTAAGCTGACGGCTGCTGGCTTACAAGTTAACGCTACG</td>
<td></td>
<td></td>
<td>CACCCCTTCCCTTCAATCCGACTGACGCTGCCGCTG</td>
</tr>
<tr>
<td>1682</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QU2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT IS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyg19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEL5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP-1</td>
<td></td>
<td>CTGAGAGGGCTGGCTGGCTCCTTCTAAGCTGACGGCTGCTGGCTTACAAGTTAACGCTACG</td>
<td></td>
<td></td>
<td>CACCCCTTCCCTTCAATCCGACTGACGCTGCCGCTG</td>
</tr>
<tr>
<td>1682</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QU2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT IS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyg19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEL5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP-1</td>
<td></td>
<td>CTGACCTGCTGCTCACTCTACCTCAGCTTTCTCGTTCTGCTGCTGGCTTACAGATCGAAAGTT</td>
<td></td>
<td></td>
<td>CACCCCTTCCCTTCAATCCGACTGACGCTGCCGCTG</td>
</tr>
<tr>
<td>1682</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QU2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT IS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyg19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEL5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP-1</td>
<td></td>
<td>CACCCCTTCCCTTCAATCCGACTGACGCTGCCGCTG</td>
<td></td>
<td></td>
<td>CACCCCTTCCCTTCAATCCGACTGACGCTGCCGCTG</td>
</tr>
<tr>
<td>1682</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QU2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT IS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyg19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEL5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>HTLV Subtype</td>
<td>Region</td>
<td>Accession number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>-----------------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATK-1</td>
<td>la (B)</td>
<td>Japan</td>
<td>J02029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1682</td>
<td>la (A)</td>
<td>Greece/UK</td>
<td>This thesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QU2</td>
<td>la (A)</td>
<td>Peru</td>
<td>Y16476</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-35</td>
<td>la (C)</td>
<td>Caribbean</td>
<td>D00294</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-2</td>
<td>la (A)</td>
<td>Japan</td>
<td>Z31660</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT-4</td>
<td>la (B)</td>
<td>Japan</td>
<td>Z31661</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR1</td>
<td>la (A)</td>
<td>Caribbean</td>
<td>K02722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT IS</td>
<td>lb</td>
<td>Democratic Republic of Congo</td>
<td>Z32527</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pyg19</td>
<td>Id</td>
<td>Central African Republic</td>
<td>L76310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEL5</td>
<td>Ic</td>
<td>Solomon Island</td>
<td>L02534</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A table illustrating the percentage differences from prototypes of each HTLV-1 strain is also presented (Table 3.1). This data is using the complete LTR region but the greatest variation is seen in U3 and U5 with the R region being more highly conserved within strain types. From this data the subtype of virus could be slotted into the Cosmopolitan type A. The greatest homology was to an isolate from Peru (QU 2) with some minor differences noted. In order to investigate this further the ENV region was also analysed.

**Table 3.1** Percentage differences within the LTR sequence of strain 1682 and various known HTLV-1 sequences.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>HTLV Subtype</th>
<th>Region</th>
<th>Accession Number</th>
<th>Percentage difference from 1682</th>
</tr>
</thead>
<tbody>
<tr>
<td>QU2</td>
<td>Ia (A)</td>
<td>Peru</td>
<td>Y16476</td>
<td>0.1</td>
</tr>
<tr>
<td>MT-2</td>
<td>Ia (A)</td>
<td>Japan</td>
<td>Z31660</td>
<td>0.3</td>
</tr>
<tr>
<td>ATK-1</td>
<td>Ia (B)</td>
<td>Japan</td>
<td>J02029</td>
<td>2.5</td>
</tr>
<tr>
<td>HS-35</td>
<td>Ia (C)</td>
<td>Caribbean</td>
<td>D00294</td>
<td>3.3</td>
</tr>
<tr>
<td>ITIS</td>
<td>Ib</td>
<td>Congo</td>
<td>Z32527</td>
<td>5.4</td>
</tr>
<tr>
<td>pyg19</td>
<td>Id</td>
<td>CAR</td>
<td>L76310</td>
<td>5.3</td>
</tr>
<tr>
<td>MEL5</td>
<td>Ic</td>
<td>Solomon Islands</td>
<td>L02534</td>
<td>9.0</td>
</tr>
</tbody>
</table>
3.5.3 PCR Analysis of ENV

ENV sequence was PCR amplified in a similar manner to the LTR region. See earlier (Figure 3.4). An expected band size was obtained with 1682 and the control HS35 of around 700bp within the env region. The bands were gel purified and the sequence determined. The complete sequence is listed below (Figure 3.7).

Figure 3.7 Partial ENV DNA sequence (C term of gp46 and N term of gp21) of 1682, bases differing from ATK-1 are shaded
ENV was aligned with various subtypes an example is shown in figure 3.8 and confirmed the alignments with the LTR region. This placed the virus firmly in the Cosmopolitan subtype with the closest homology to members of the Cosmopolitan A subtype.

**Figure 3.8** Alignment of 1682 ENV sequence (C term of gp46 and N term of gp21) with several examples of HTLV subtypes. ATK1 sequence is given across the top. Identical matches are given by a dash (−), the 1682 sequence is given in bold and deletions are indicated with (^).
Chapter 3 Characterisation of HTLV-I from a Caucasian ATL patient

<table>
<thead>
<tr>
<th>ATK1</th>
<th>ATCAAAACTCTGGACCTTTGTCAGTTAACCCTACTACAAAGCCTATTATTTGCTATGGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1682</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MT2</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>SP</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>HS 35</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MT4</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>EL</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MEL 5</td>
<td>----------------------------------------------------------------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATK1</th>
<th>CTGTATCGACGCTGCCCCACTCCACTCCACGCTCTATCTCTCCCCACGCTCTCTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1682</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MT2</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>SP</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>HS 35</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MT4</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>EL</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MEL 5</td>
<td>----------------------------------------------------------------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATK1</th>
<th>TCCATCCTCTCTTATCCTACCCCTACCCCTCTTACCCCATCGTTAGCTCTACATCGCCGCCACCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1682</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MT2</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>SP</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>HS 35</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MT4</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>EL</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MEL 5</td>
<td>----------------------------------------------------------------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATK1</th>
<th>GACGTTAATTTAATGGACCCACTGGTCTTTTGACCACAGTTAATAGCTCTCCCTCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1682</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MT2</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>SP</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>HS 35</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MT4</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>EL</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>MEL 5</td>
<td>----------------------------------------------------------------</td>
</tr>
</tbody>
</table>

- 74 -
Chapter 3 Characterisation of HTLV-I from a Caucasian ATL patient

421
ATK1 CCCCTGCTATAACTCCCTCATCTGCCCCCCTTTCTTGTGTCACCTGTTCCACCCTAGG
1682
MT2
SP
HS 35
MT4
EL
MEL 5

481
ATK1 ATCCCGCTCCCGGAGCGGTACCGGTGGCGGTCTGGCTTGTCTCCGCCCTGGCCATGGG
1682
MT2
SP
HS 35
MT4
EL
MEL 5

541
ATK1 AGCCGGAGTGGCTGGCGGGATTACCGGCTCCATGTCCCTCGCCTCAGGAAAGAGCCTCCT
1682
MT2
SP
HS 35
MT4
EL
MEL 5

601
ATK1 ACATGAGGTGGACAAAGATATTTCCCAGTTAACTCAAGCAATAGTCAAAAACCACAAAA
1682
MT2
SP
HS 35
MT4
EL
MEL 5

-75-
Limited phylogenetic analysis was performed using the Env region sequence as there was not many sequence differences noted and only a limited number of Env sequences available with corresponding LTR. An example of a tree formed using these sequences is given below and placed 1682 Env into the Ia A subtype. Presented is a neighbour joining horizontal phenogram with numbers representing percentage bootstrap support for branches. For accession numbers for sequences see Table 3.2.
3.5.4 Phylogenetic analysis of the LTR region

In order to confirm the DNA sequence alignments trees were constructed using the Phylip software package. The LTR sequence was chosen in preference because of the higher percentage of differences between the subtypes gives a more precise classification. It was demonstrated that 1682 clustered in the Cosmopolitan la subtype using this limited number of HTLV subtypes. The most common unrooted tree using the maximum likelihood method is presented below (Figures 3.9). Using the LTR region the la subtype could be more accurately subdivided and this meant that the 1682 sequence clustered well within the type A subgroup that corresponds to a transcontinental variant. The abbreviations and accession numbers are listed in a table below (Table 3.2).

Table 3.2: HTLV-1 subtypes used for tree formation.

<table>
<thead>
<tr>
<th>Name</th>
<th>Subtype</th>
<th>Geographical Origin</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1682</td>
<td>la A</td>
<td>Greece/UK</td>
<td>This thesis</td>
</tr>
<tr>
<td>MT2</td>
<td>la A</td>
<td>Japan</td>
<td>Z 31660</td>
</tr>
<tr>
<td>CR1</td>
<td>la A</td>
<td>USA</td>
<td>K 02722</td>
</tr>
<tr>
<td>QU2</td>
<td>la A</td>
<td>Peru</td>
<td>Y 16476</td>
</tr>
<tr>
<td>AMA</td>
<td>la A</td>
<td>Brazil</td>
<td>X 88871</td>
</tr>
<tr>
<td>MAQS</td>
<td>la A</td>
<td>Brazil</td>
<td>X 88876</td>
</tr>
<tr>
<td>AINU</td>
<td>la A</td>
<td>Japan</td>
<td>D 23694</td>
</tr>
<tr>
<td>CMC</td>
<td>la A</td>
<td>Taiwan</td>
<td>X 88872</td>
</tr>
<tr>
<td>ATK1</td>
<td>la B</td>
<td>Japan</td>
<td>J 02029</td>
</tr>
<tr>
<td>MT4</td>
<td>la B</td>
<td>Japan</td>
<td>Z 31661</td>
</tr>
<tr>
<td>ATL-YS</td>
<td>la B</td>
<td>USA</td>
<td>U 19949</td>
</tr>
<tr>
<td>ATM</td>
<td>la B</td>
<td>Japan</td>
<td>J 02030</td>
</tr>
<tr>
<td>H5</td>
<td>la B</td>
<td>Japan</td>
<td>M 37299</td>
</tr>
<tr>
<td>HS35</td>
<td>la C</td>
<td>Caribbean</td>
<td>D00294</td>
</tr>
<tr>
<td>GH78</td>
<td>la C</td>
<td>Ghana</td>
<td>D 23693</td>
</tr>
<tr>
<td>BO</td>
<td>la D</td>
<td>Algeria</td>
<td>U 12804</td>
</tr>
<tr>
<td>OD</td>
<td>la D</td>
<td>Mauritania</td>
<td>U 12805</td>
</tr>
<tr>
<td>ITIS</td>
<td>lb</td>
<td>Zaire</td>
<td>Z 32527</td>
</tr>
<tr>
<td>H24</td>
<td>lb</td>
<td>Cameroon</td>
<td>L 76308</td>
</tr>
<tr>
<td>MEL5</td>
<td>lc</td>
<td>Solomon Islands</td>
<td>L 02534</td>
</tr>
<tr>
<td>pyg19</td>
<td>Id</td>
<td>Central African Republic</td>
<td>L 76310</td>
</tr>
<tr>
<td>H23</td>
<td>Id</td>
<td>Cameroon</td>
<td>L 76312</td>
</tr>
</tbody>
</table>
Figure 3.9  Maximum likelihood tree rooted at MEL5 isolate derived from complete LTR sequence. In this horizontal phenogram, vertical distances are only for clarity and do not represent homology.
Figure 3.10  Maximum likelihood unrooted tree derived from complete LTR sequence
Figure 3.11 Neighbourhood Joining tree including an STLV 1 sequence, with bootstrap support for LTR sequence. Numbers represent percentage bootstrap support for branch.
3.6 Discussion

HTLV-1 and its related diseases, ATL and TSP/HAM, are a major health problem in certain regions of the world. Further study into this virus and its genome should aid our understanding of its pathology and give us an insight into Leukaemia in general. By studying the virus subtype in unusual cases, such as the one presented here in a Caucasian woman, our knowledge of this virus will increase. Subtyping may reveal new strains of virus that could have altered infectivity and so take hold in new geographical areas.

ATL is unusual as a leukaemia since no specific chromosomal mutation or translocation has yet been associated with the disease. However chromosomal aberrations are common in ATL tumour cells (Kamada et al. 1992) and further analysis of tumour cases using chromosome painting or comparative genomic hybridisation may help in the determination of progression factors.

It is clear that Tax is important in the transformation of T-cells and has influence on T-cell growth and function via the induction of the NFkB pathway. In particular growth is stimulated by the up-regulation of IL-2 production as well as GM-CSF. However Tax also influences cellular growth with the induction of numerous other factors that could aid oncogenesis. Another notable observation is the down regulation of β-polymerase, leading to a lower level of DNA repair. This in turn will lead to an increase in chromosomal abnormalities and is a plausible mechanism of progression.

It seems clear however that due to the relatively low risk of developing disease after infection (less than 5%) that other co-factors must be important. There is probably a genetic element giving a predisposition to progression, however other possibilities include other bacterial or viral infections possibly even Human Endogenous Retroviruses. Strongyloides stercoralis and hypereosinophilia possibly delay onset (Plumelle et al. 1997), therefore other infections may also play a role.

The Cosmopolitan subtype is the most globally widespread HTLV-I and contains the largest number of isolated viruses. The 1682 isolate sequence presented here differs only very slightly from a previous isolate (QU 2) from Peru (Van Dooren...
et al. 1998) and may in fact be the same strain. This strain has been subtyped as Ia (A) (the transcontinental variant) and from the data presented 1682 also falls within this group.

The discovery of a case of ATL in a Caucasian patient presenting in a non-endemic country indicates the potential for further spread of this pathogenic virus. To our knowledge all previous cases of HTLV-I infection in the UK can be traced back to endemic areas either by direct travel or via relationship with someone from an endemic area, usually Afro-Caribbean or Japanese origin where the virus is most prevalent. It is our understanding that this patient had no direct contact with an endemic area and as such indicates that the Cosmopolitan strain may be spreading into Western Europe. A recent case of a virus subtype related to Romanian cases was also reported in Germany (Ellerbrok et al. 1998) indicating a possible pocket of HTLV in central Europe. A pocket of infections may also exist in Southern Italy (Manzari et al. 1985). This evidence supports the continued screening of blood donors throughout Europe to help prevent new infections and keep track of the rates of spread within Europe and in particular in the UK, where testing has not yet been routinely implemented.
Chapter 4: Isolation and Characterisation of Two Novel HERV fragments

The aim of this chapter was to isolate novel HERV sequences using degenerate PCR primers and the screening of human libraries. Then to determine their phylogeny. An approximate copy number was also obtained. Two novel sequences were analysed in detail, further sequence information was obtained by screening a human PAC library. The phylogeny of these elements was then determined by sequence analysis of their pol region. This meant that the novel elements could be placed into their respective families. This chapter illustrates the number of undiscovered HERV present in the genome and their potential biological activity is discussed.

4.1 HERVs

A revision of the nomenclature used to describe HERVs is currently underway, but at present a variety of mechanisms are used. One well adopted method is to utilise the primer binding site (PBS) of the LTR, this PBS is complementary to human tRNA. The one letter code representing the amino acid corresponding to this tRNA is then taken as the HERV's name, for example HERV-K would contain a PBS complementary to human tRNAlysine, this is then followed by a number for each HERV within the group. This nomenclature however does have some problems as some retroviral sequences do not have a determined PBS. A second problem is that some sequences with the same PBS are highly divergent, an example being the families primed by tRNAIle that contain both HML5 and RRHERV1 (Tristem 2000). Another attempt to classify HERVs is by their homology to exogenous retroviruses Class I being those related to mammalian type C retroviruses, Class II being those related to mammalian type B or D retroviruses and a putative Class III being those related to the spumaviruses. Other nomenclature has been more haphazard, for example relying on nearby genes e.g. ADP-pol (Lyn et al. 1993) or clone number e.g. S71 (Werner et al. 1990). Recent screening of the human genome (Tristem 2000) has led to greater than 150 HERV members being documented, grouping by phylogeny
into at least 22 groups. 17 of these groups were designated as Class I, three as Class II and two as Class III. In this chapter more historical nomenclature is used.

4.1.1 Class I HERV

Class I HERVs have homology to mammalian type C retroviruses by pol sequence comparison.

The HERV-ERI superfamily, are a family of closely related elements by hybridisation back to genomic DNA, their LTR and env sequences differ and they have different primer binding sites but are all closely related. The LTR and env regions diverge fastest for infectious viruses and so give a good indication of relatedness (Shimotohno et al. 1985; Doolittle et al. 1989). The name ERI is used, as the PBSs of the first members of the group were complementary to the human Glutamic acid, Arginine and Isoleucine tRNAs.

The 4.1 (HERV-E) group is so named as the first elements isolated from this group were clones 51-1, 4-1 and 4-14. They were isolated by cross-species genomic library screening. African Green monkey DNA was analysed by cross hybridisation to MLV, a portion of this clone was highly related to the gag-pol region of BaEV, which was then used to screen a human genomic library (Martin et al. 1981), following the work of Benveniste in 1974. Clone 4-1 contained a full length HERV (8.8kb), with gag-pol-env bound by 490bp LTRs (Repaske et al. 1985). The PBS was homologous to rat glutamic acid tRNA (Steele et al. 1984), so the element was called HERV-E (Larsson et al. 1989). HERV-E LTR sequences were also identified in the human genome (Steele et al. 1984) and these are thought to occur from the recombination of two LTRs from an original full length HERV.

Clone 51-1 contained an almost complete gag and pol with 85-93% homology to 4-1 (Repaske et al. 1983). 51-1 (6kb) had no env region and was bound by novel repeat structures of 72-76bp tandemly repeated 8-13 times, with no similarity to retrovirus LTRs. These repeats seem to be unique (Steele et al. 1984).

The copy number of both 51-1 and 4-1 elements is 35-50 and they are found associated with regions of amplified DNA (Steele et al. 1984). Conserved restriction sites were detected in the 3’ flanking elements (Steele et al. 1984) and DNA to the 5’ was used as a probe to isolate multiple distinct clones of the family (Repaske et al.
1983). On Southern blots bands not corresponding to a single chromosome were observed illustrating that these clones are dispersed throughout the genome (Steele et al. 1986). Clusters of clones were seen in telomeric and to a lesser extent centromeric regions of chromosomes when mapped by in situ hybridisation (Taruscio et al. 1991).

The characterisation of the amylase gene family led to the discovery of a cluster of HERV-E elements. There is a series of tandem duplications yielding five functional amylase genes. Three genes were expressed in the salivary gland and two in the pancreas. The full length HERV-E were called ERVA1A-1C and were found 800bp 5' to each salivary gland gene in the opposite transcriptional orientation. There was also a solo LTR (ERVA2A) 5' to one of the pancreas genes (Samuelson et al. 1990). Insertion of full length HERV-E activated a promoter that directed tissue specific expression of the genes. Parotid specific regulatory sequences were found in 5' regions of the HERV-E elements at this locus and explain why deletion of proviral sequence by LTR-LTR recombination reverted to pancreatic expression (Ting et al. 1992). HERV-E was important in the evolution of this gene family, in the chimpanzee version there are less genes and it seems there are further duplications in humans (Samuelson et al. 1990). This finding illustrates how amplifications involving endogenous retroviruses can lead to species differences.

HERV-E members have been found in Old World monkeys, apes and humans. 4-14 was also found in the same location in all by PCR (Shih et al. 1991) and was integrated at least 25 million years ago (Sibley et al. 1987). This clone was amplified in the same region of DNA in humans and chimpanzees (at least 5 million years ago) (Sibley et al. 1987). Polymorphisms within humans are not detected and HERV-E is therefore an ancient and stable part of genome now (Steele et al. 1984; Taruscio et al. 1991). Successive integrations into the germ line probably occurred and some HERV-E was dispersed as single elements (Taruscio et al. 1991) as well as a divergence of LTR sequences from different clones. The average genetic drift is 5-10% by comparison of 5' and 3' LTRs within a single full length element (Repaske et al. 1985) and the observed HERV-E variation is greater than this suggesting that different isolates integrated at different stages in evolution and are now preserved.

Clone 4-1 has been completely sequenced (Repaske et al. 1985) and in frame stop codons were found in each of the coding regions. Expression of 4-1 has been
studied in various tissues and cell lines this is detailed in the next Chapter (section 5.1.1). Protein products however have not been reported.

Hs5 is a HERV that was isolated from a human library by hybridisation to FeLV 3' pol, env and LTR sequences (Levy et al. 1990). It has a distinctive map that is largely deleted with only 3' pol and env identified (Levy et al. 1990). All of the Hs5 coding regions are disrupted. These identified regions are 77% homologous overall to 4-1. The highest homology is for a fragment of LTR that is 80% homologous to a solitary HERV-E LTR (LTRpdhB).

NP-2 was found by using HERV-E pol regions as probes and was shown to be related by hybridisation (Silver et al. 1987). NP-2 has a 3' LTR plus gag, pol and env related regions. The HERV-E related cloning site is excluded along with the 5' LTR and some gag. 1.1kb has been sequenced and has 73% homology with 4-1 env and decreasing homology with other members of HERV-ERI group was seen. The only two elements found were on the Y chromosome and were associated with similar flanking sequences suggesting duplication. No transcripts were seen in normal testis, lung, colon and pancreatic carcinomas (Silver et al. 1987).

Another HERV, ERV1 was isolated by hybridisation to a chimpanzee pol fragment with homology to BaEV (Bonner et al. 1982). A group of clones from 18q22-23 (O'Brien et al. 1983; Renan et al. 1987) contained a retroviral element named ERV1 that had been there for at least 5 million years (Sibley et al. 1987) and could be older since there was cross hybridisation to ERV1 in Old World monkeys. ERV1 has gag, pol and env-related sequences as well as a putative 3'LTR, but the 5'LTR was deleted and the PBS was not characterised (Bonner et al. 1982). Of the area sequenced, the most homology was to BaEV, 4-1, ERV3 and RRHERV-1. With low stringency cross hybridisation as many as 15 related elements are seen in the human genome (O'Brien et al. 1983), any expression is doubtful.

ERV3 (HERV-R) is a single copy element that was isolated by hybridisation to a chimpanzee clone with 3' pol, env and a BaEV LTR (O'Connell et al. 1984). It spans 9.9 kb on chromosome seven and contains gag, pol and env-related regions flanked by 590bp LTRs (O'Connell et al. 1984), the PBS being most closely related to
tRNA for arginine. The Southern blot clearly shows only a single band, but 10 or more bands if lower stringency conditions are employed (O'Connell et al. 1984). HERV-R is also detected in Old World monkeys (but not gorillas) and a related element is seen in New World monkeys (Shih et al. 1991). The LTRs and env have been sequenced as well as some of the other genes, gag and pol show a close homology between HERV-R, ERV1 and clone 51-1 of HERV-E (O'Connell et al. 1984). Env is related to RRHERV-I with 68% homology over more than 2kb (Kannan et al. 1991). Gag and pol are disrupted by mutations with stop codons in the ORFs, but have homology to retroviral proteins (O'Connell et al. 1984). HERV-R expression is discussed in the next Chapter (section 5.1.2). HERV-R is the only element with homology to BaEV LTR. Others have been isolated but not characterised (Noda et al. 1982) as there are 10 bands on a genomic blot it is assumed there is a family of about 10 and HERV-R is one of this family. ERV-3 has now been studied more closely showing variable expression patterns in different human tissues (Larsson et al. 1997; Sibata et al. 1997).

RRHERV-1 was isolated inadvertently by PCR (Kannan et al. 1991), HERV-R was used to probe a teratocarcinoma cell line (from PA-1) and an RNA transcript was detected, with several other clones being subsequently isolated from a cDNA library. A 3.3kb composite was named RRHERV-I and appears to be assembled from several cDNAs (Kannan et al. 1991). It has homology with the env of HERV-R and 4-1 and the LTR of ERV1. It is primarily expressed as a spliced subgenomic mRNA and hybridised to a 4kb transcript on a northern blot (Kannan et al. 1991). Termination codons are present in the putative env and there are an estimated 20 copies present in the human genome, with a similar number in Old World monkeys making the virus over 25 million years old.

The expression of spliced transcripts may be indicative of host adaptations limiting the release of infectious virions from endogenous elements (Wilkinson et al. 1993). Another theory is that env may have adapted to serve some cellular function (Larsson et al. 1994). The low homology in env and LTR shows the different stages of evolutionary lineage of the ancient exogenous retrovirus and periodical fixation within the host genome.
Other Class I elements show a lower level of homology than HERV-ERI elements do to type C retroviruses or the other retroelements, but still have a significant homology.

RTVL-I (HERV-I) was discovered during analysis of a haptoglobin-related locus on chromosome 16 (Maeda 1985), it is a 9kb element containing 500bp LTRs as well as gag, pol and env related regions (Maeda et al. 1990). The PBS is complementary to the tRNA for isoleucine, hence the name RTVL-I (retrovirus-like) also known as HERV-I. Three separate HERV-I insertions seem to have occurred. After the evolutionary divergence of humans and chimpanzees a region containing one of these elements and a haptoglobin-related gene was deleted (McEvoy et al. 1988; Maeda et al. 1990). In humans the elements are located within the first intron of the gene and in the 3' flanking region and the elements are called RTLV-Ia and RTLV-Ic. Ia is integrated in the same location in Old World monkeys, Ib is only present in chimpanzees. All three have been sequenced (Maeda et al. 1990) and no expression was reported. There are several termination codons and short deletions present in the elements, the internal 3' region of Ia has an inserted Alu element, the 5'LTR and gag of Ic are deleted and the remaining region has two Alu insertions (Maeda et al. 1990).

This region of genomic DNA has homologous but unequal recombinations, for example the third haptoglobin-related gene in primates and its deletion in humans (Maeda et al. 1990). There are up to six tandemly repeated copies of this gene described (Maeda et al. 1986). The HERV-I elements could be directly involved in these duplications. There is also a solitary HERV-I LTR associated with another region of genomic variation, the minisatellite region 1q42-43 (Armour et al. 1989). A partial LTR was identified in the 5' regulatory sequence of a cytochrome c1 gene and may be involved in the regulation of transcription (Suzuki et al. 1990).

An element named ADP-pol was isolated by its proximity to the ADP-ribosyltransferase (NAD+) pseudogene on 13q34. A 1.5Kb portion with homology to pol was identified with the greatest homology to HERV-I but this element was defective (Lyn et al. 1993).
S71 was isolated when simian sarcoma-associated virus (SSAV) was used to
screen a human library (Leib Mosch et al. 1986). S71 is a single element found on
18q21 (Brack Werner et al. 1989) and has been present in all individuals tested so far
(Leib Mosch et al. 1986; Brack Werner et al. 1989). Under low stringency about 20
sequences related to S71 are noted on a Southern blot (Leib-Mosch et al. 1992).
These sequences are present in Old but not New World monkeys. Sequenced, short
direct repeats are used to delineate the element composed of retrovirus-related and
unrelated sequences. The 5.4kb HERV contains $gag$ and $pol$ regions and ends in a
535bp 3'LTR with 400bp at the 5' end plus 1.1kb between $gag$ and $pol$.
This 1.1kb was initially thought to have no retroviral homology but later it was found that almost
this entire region is a solitary HERV-K LTR in antisense (Leib Mosch et al. 1993).
The $gag$ and $pol$ regions are disrupted with mutations, a deletion has removed the
protease and RT regions of $pol$ and there is no 5' LTR or putative $env$. The element is
related to SSAV, but also equally related to BaEV and MuLV with the RNase H
region having homology with SSAV of 60% over 156 residues (Werner et al. 1990).
There is one short ORF of 167 amino acids at the 5' end of $gag$.

Hybridising transcripts were also seen in the leukaemia cell line K562, a
prominent 2.9kb transcript and minor 2.5, 3.6 kb transcripts (Leib-Mosch et al. 1992).
The transcripts were not detected in normal PBMCs or in several other cell lines
tested, however a cDNA containing 350bp related to the 3' end of S71 was isolated
from a human placental library (Leib-Mosch et al. 1992).

HRES-1 (HTLV-I related endogenous sequence-1) was isolated by screening a
library with HTLV-I LTR and $gag$ sequences (Perl et al. 1989). Eight clones were
isolated and one was characterised in detail. Under high stringency a single fixed
locus of 1q42 was determined (Perl et al. 1989; Perl et al. 1991). This was
homologous in Old and New World monkeys but not other mammals. One other
similar element was noted and named HRES-2 (Perl et al. 1991). 5' LTR and $gag$
related regions are present in HRES-1 (Perl et al. 1989), a 684bp LTR is followed by a
PBS (12/18) of chimpanzee histidine tRNA. The $gag$ consists of 2 overlapping ORFs,
only one of which has detectable homology with other retroviral sequences and this
homology is low and limited to two short stretches of amino acids. One region is at
the amino terminal (31aa) and shows a 32% and 39% homology to p19 of HTLV-I
and II respectively. The other is 36aa in the middle of \textit{gag} and shows a 25% and 28% homology to p24 of HTLV-I and II respectively, as well as a 33% homology with the HIV-2 \textit{gag} region. A 6kb transcript containing \textit{gag}-related sequences was seen in the polyadenylated RNA of several cell lines for details see the next Chapter (section 5.1.5). Antisera directed against HTLV picked up a 28kD protein that corresponds to an HTLV-related ORF, the protein was intracellular and cytoplasmic, with some seen in nuclear bodies (Banki et al. 1992).

HERV-P families all contain a PBS corresponding to the tRNA for proline. A genomic library was screened by one group with an oligonucleotide that extends past the PBS to the conserved CA, defining the end of the LTR. Several bases of proviral homology were also added with close matches to retroviral and not tRNA sequences (Kroger et al. 1987). Another group end labelled 3' fragments of purified proline tRNA to screen a genomic library and identified retroviral hybrids by digesting with RNase A during screening, thus removing the unbound CCA tail (and radiolabel) of the probe bound to the tRNA clones (Harada et al. 1987). By these methods 3 different families were found, distinguished by LTR sequences, and were called HuERS-P1, -P2 and -P3. Two HuERS-P1 were isolated P1-1 of about 8kb and P1-2 of 7.5kb (Harada et al. 1987). The elements were sequenced including the 690bp LTR and an Alu sequence was found in the 5' LTR of P1-2. Also an internal P1-1 fragment was shown to have homologous elements in Old World monkeys but not mice and a copy number of about 10-20 (Harada et al. 1987).

A single HuERS-P2 was isolated, it contained 890bp LTRs flanking 3.2 kb, the LTRs were not related to P1 and P3 and large regions were deleted. The internal region had a related element in Old World monkeys but not mice, 20-30 elements appear present in humans (Harada et al. 1987).

Both groups isolated several P3 (Harada et al. 1987) and HuRRS-P elements (Kroger et al. 1987), one complete element was reported and two partial elements containing primarily LTR sequences (Harada et al. 1987; Kroger et al. 1987). The full element was 8.1kb with 630 bp LTRs. Over 145 residues a 62% amino acid homology with RT of MuLV was noted, but P3 had in frame termination codons. The P3 LTR fragment has related sequences in apes, Old and New World monkeys, but not mice (Harada et al. 1987). Using the \textit{gag} and \textit{env} regions there appears to be 20-
40 related elements present (Kroger et al. 1987) and the internal regions are largely uncharacterised.

ERV-9 was the first element isolated in a family, ERV-9 consists of three 95bp tandem repeats at the 5' end of a cDNA coding for a protein with zinc-finger-related domains (Pannuti et al. 1988). These repeats are commonly associated with other short repeated sequences, they are highly recurrent in the human genome and appear unrelated to other repeats. Expression is seen in the teratocarcinoma cell line Ntera2D1 and is down regulated by retinoic acid induced differentiation (La Mantia et al. 1989).

A complete retroelement sequence called pHE.1 has been isolated with gag, pol and env related sequences. The env region is partially deleted as well as the 5' terminus of the clone although the gag is complete the sequence is disrupted by mutations (La Mantia et al. 1989). pHE.1 is most related to BaEV with the pol being 54% similar over 1340bp and env being 62% similar over 300bp. There are 40 related sequences seen by pol hybridisation on a Southern blot and they appear to be present in other primates but not rodents (La Mantia et al. 1991). It is estimated that 3,000-4,000 related solitary LTRs are also present in the genome. pHE.1 pol sequence was used to isolate other members, one of which was called λ Fix 1.1 with exceptionally long 1.8kb LTRs flanking a 6kb sequence, this element contained the PBS for arginine tRNA (La Mantia et al. 1991). The 5' and 3' LTRs bear 90% similarity and have two types of novel repeats, they differ from the LTR sequence of the initial isolate in the number of repeats giving an important source of heterogeneity (La Mantia et al. 1991). An 8kb transcript was detected in undifferentiated Ntera2D1 and was downregulated when differentiation was induced by retinoic acid (La Mantia et al. 1991). Other transcripts were noted in placenta, but leukaemic and carcinoma cell lines were not analysed. ERV-9 segments were PCR amplified from glioma RNA including an immunosuppressive region of env (Lindeskog et al. 1993). Spliced RNAs of 1.5 and 2kb that remove most of the internal sequence were noted in Ntera2D1 (Lania et al. 1992). Deletion analysis of the promoter of ERV-9 LTRs showed that repeat sequences are not essential for transcription from the LTR (La Mantia et al. 1992). The LTR lacks a TATAA sequence and has an unusually short R region of 6bp, the
promoter may be regulated by Sp1 binding and another protein interacts with an initiator-like element -7 to +6 from the transcription start site (La Mantia et al. 1992).

RTLV-H (HERV-H) was the first member of this group to be isolated, it was found near to the 3' breaks of three large deletions in the β-globin gene (Mager et al. 1984; Mager et al. 1985). The PBS matched human histidine tRNA hence the name. The majority of HERV-H sequences are approximately 5.8kb and contain gag and pol sequences flanked by 400 to 500bp LTRs (Mager et al. 1984). They are similar to retrotransposons as they lack env sequences, so are sometimes grouped with retrotransposon-like elements. These are actually derivatives of less abundant retrovirus-like elements with gag-pol-env sequences. There are approximately 1000 HERV-H elements per human haploid genome and a similar number of solitary LTRs (Mager et al. 1984). A similar number are found in apes and Old World monkeys, but not in rodents (Goodchild et al. 1993). In situ hybridisation studies show the elements are widely dispersed within the genome, with clusters at 1p31 and 7q31 (Fraser et al. 1988). The locations seem to be fixed, although two siblings carried an allele where the element had deleted leaving only LTR-LTR (Mager et al. 1989). Sections of gag are 50% similar to BaEV p30 and 55-61% similar to the NC of retroviruses in HTLV phylogeny. In pol the closest homology is to MLV at 40-50% depending on the domain (Mager et al. 1987). The majority of HERV-H elements contain four shared deletions termed A, B, C, D that remove a lot of the functional pol, an undeleted element however has been detected (Hirose et al. 1993). A HERV-H element was also identified with a partial env. This had the highest homology with the TM of a type D retrovirus with 40% amino acid homology over 240 residues (Hirose et al. 1993). PCR was also used on an immunosuppressive region leading to the isolation of env from HERV-H (Lindeskog et al. 1993). HERV-H is transcribed in many cell lines for details refer to Chapter 5 (section 5.1.7). Some intact HERV-H contain ORFs greater than 1kb in pol and/or env regions (Hirose et al. 1993; Wilkinson et al. 1993). There are three subtypes of LTRs associated with HERV-H (Goodchild et al. 1993; Wilkinson et al. 1993), these are closely related in R, the U5 regions and the first 130bp of U3. The rest of U3 contains different tandemly repeated sequences that distinguish each LTR type (Goodchild et al. 1993; Wilkinson et al. 1993). Types I and II are most abundant and are ancient, with several hundred copies in humans, apes
and Old World monkeys. A third subtype la, formed from a recombination of I and II (Goodchild et al. 1993; Wilkinson et al. 1993) is not found in Old World monkeys but has a copy number of 100 in humans and higher primates (Goodchild et al. 1993). The different subtypes have functional differences in their promoter activity (Feuchter et al. 1990). In general I and II have little or no promoter activity in most cell lines, and when assayed in CAT (chloramphenicol acetyltransferase) construct transfection experiments, type la have stronger promoter activity (Feuchter et al. 1990). Other cellular sequences, apparently promoted by HERV-H LTRs, have been isolated (Feuchter et al. 1992). A chimeric cDNA was isolated from a library of two prostate metastasis-derived cell lines and is a spliced fusion between HERV-H and an exon of calbindin gene, which is expressed at the protein level (Liu et al. 1991). This leads to deregulation of calbindin expression and may have a role in malignancy.

Five members of a group designated XA34-XA38 ERV were isolated by using an ERV-9 probe on a human glioma library (Widegren et al. 1996). The largest element (XA38 ERV) was 4kb long extending from the central part of pol covering a truncated env. These elements seem to have a low copy number and show greatest homology to RTLV-H.

Recently several novel families have been isolated by searching the human genome databases (Tristem 2000). This has resulted in five new Class I families HERV-Rb, HERV-F, HERV-Fb, HERV-Z69907 and HERV-HS49C23. HERV-F and HERV-Fb are closely related to each other and to RTLV-H. The HERV-F element is 8.7Kb containing gag, pol and env related sequences, it is present on Xp11.3-4, exists with a low copy number and only encodes short ORFs. HERV-Fb may exist at higher copy numbers as two elements were isolated. HERV-Rb again exists as a low copy number. A complete element containing LTR-gag-pol-env-LTR within 8.7kb was characterised and is most closely related to FRD. This element probably doesn't encode any gene products and was found on 4q25. As the LTRs were not found the next two groups were given the designations of the cosmids that encoded them. A HERV-Z69907 element was characterised on 22q11.2 with gag, pol and env regions, however the pro was deleted and the env was largely deleted. The homology for this element was variable. The final family HERV-HS49C23 had its most complete
element on the X chromosome, it was highly defective and there were no LTRs present.

A retrovirus that has been associated with multiple sclerosis has been isolated. Initially called LM7 and later referred to as MS-associated retrovirus (MSRV). RT-PCR was used to isolate pol sequences (Perron et al. 1997). This element has now been further characterised and placed within the HERV-W family. A complete element of this was isolated via its expression in placental tissue with intact regions in gag, pol and env (Blond et al. 1999). More recent work has shown that a protein expressed from HERV-W named syncytin is in fact involved in placental morphogenesis (Mi et al. 2000). Syncytin seems to have been adopted by the genome for the purpose of inducing the formation of syncytiotrophoblasts. This demonstrates the potential for endogenous retroviral expression and true biological activity within the human genome.

Three HERV-sequences were isolated from purified T47-D particles (Seifarth et al. 1995). These particles are type B in appearance but two of the isolated sequences, ERV-FRD and ERV-FTD, are derived from pol sequences from defective Class I endogenous viruses. This data suggests that these type B particles are in fact packaging sequences originating from several endogenous retroviruses. More recent analysis of the FRD family (Tristem 2000) has led to the isolation of further elements and in fact this family appears to be one of the oldest in the genome.

4.1.2 Class II HERV

Class II HERVs have homology to exogenous mammalian retroviruses of type B or D.

As mentioned above three retroviral sequences were isolated from the T47-D cell line. The third sequence ERV-MLN (Seifarth et al. 1995) was related to HERV-K10 (described later in this section) and is a Class II HERV. This was a multicopy family and can be grouped with the HERV-K family. Later investigation led to the discovery of a PBS for tRNA<sup>lysine</sup> and therefore the name HERV-KT47D was accepted (Seifarth et al. 1998).

- 94 -
HERV-K is the most well described member of this Class. It was originally isolated by cross-hybridisation to murine mammary tumour virus (MMTV) pol-region probes. By using a gag-pol probe many related elements were seen by Southern blotting in humans (Callahan et al. 1982; Deen et al. 1986; Ono et al. 1986). The prototypic element is 9.5kb long bound by 970bp LTRs and contains a PBS for lysine tRNA (Callahan et al. 1985; Ono et al. 1986). Homology to MMTV is restricted to the pol region with a high homology here of up to 70% in some regions (Callahan et al. 1985; Deen et al. 1986; Ono et al. 1986). Homology with Class II squirrel monkey retrovirus (SMRV) LTRs was seen by cross-hybridisation, other regions of homology were also seen with avian RSV, murine IAP and SMRV (Callahan et al. 1985; Ono et al. 1986). There are approximately 50 copies of HERV-K in the human haploid genome (Callahan et al. 1985; Deen et al. 1986; Ono et al. 1986). Some expression occurs and transcripts of various sizes are seen for details refer to the next Chapter (section 5.1.8).

A member of this group, HERV-K10, has very few mutations in its coding regions (Ono et al. 1986). The gag region contains only one frameshift and the env region has one non-sense mutation. The protease region is intact encoding 344 amino acids, the pol region encodes 937 amino acids, with a 290bp deletion removing the C terminal and some of the N terminal of env. The 5' and 3' LTRs are almost identical with only two nucleotide differences over 968bp. All of this implies a relatively recent insertion of HERV-K10.

HERV-K related sequences have been found in great apes and Old World monkeys but not in more distant species. These sequences are dispersed throughout the genome but are found at fixed and largely conserved locations in both humans and apes implying more ancient integration (Horn et al. 1986; Marianna-Constantini et al. 1989). There is an insertion in chimpanzee that is not present in human showing that the retrovirus was active after human-chimpanzee divergence (Craigie 1992). There is evidence that some elements are associated with tandemly repeated genomic sequences (Horn et al. 1986) and one (NMWV 1E) is specifically rearranged in some breast carcinoma cell lines (May et al. 1989).

As well as full length HERV-K, up to 25,000 solitary HERV-K LTRs are present in the human genome and these are expressed as parts of cellular transcripts in
a variety of tissues (Leib Mosch et al. 1993). This copy number doesn’t include the repetitive SINE-R family as the probe used can distinguish between the two groups of elements (Ono et al. 1987). A few have been isolated and are derived from LTR-LTR recombinations (Kambhu et al. 1990; Liu et al. 1990; Leib Mosch et al. 1993). These LTRs show non-random distribution (Leib Mosch et al. 1993) implying that the DNA amplification mechanism was not HERV-K retrotransposition, also two HERV-K related loci were noted on Xq28 within 160kb of each other (Sedlacek et al. 1993). It is suggested that HERV in general may tend to associate with recombination hot spots, evidence supporting this includes HERV-K LTRs association with HLA-DQ (Kambhu et al. 1990) and SINE-R association with the complement C2 VNTR locus (Zhu et al. 1992).

Several other Class II HERV have been isolated. 100 clones were isolated and divided into 9 groups by pol cross-hybridisation, the largest group of 64 being closely related to HERV-K (Franklin et al. 1988). Another researcher used PCR to isolate 245bp fragments of RT (Medstand et al. 1993), leading to six groups named HML-1 to HML-6 on sequence comparison. HML-2 was found to be approximately 90% identical to HERV-K. The most divergent was HML-6 which was 55-60% identical to HERV-K group or MMTV and was more closely related to IAP (Medstand et al. 1993). Southern blot studies showed that HML-1, 2, 3 and 6 were the most abundant with 60-80 copies per human genome, RNA expression was shown in several tissues (Medstand et al. 1993).

Rearrangements within HERV lead to heterogeneity, these are commonly deletions (Deen et al. 1986; May et al. 1986; Ono et al. 1986). They can also differ in LTR sequence, e.g. a 6kb element containing 430bp LTRs has been isolated called NMWV4. It had a lysine tRNA PBS, but the codon was UUU not CUU as with other HERV-K (May et al. 1986).
4.1.3 Class III HERV

More recently several HERVs have been isolated that show sequence homology to foamy viruses rather than type C or type B or D exogenous viruses, hence these have been denoted as Class III. The HERV-L group fall into this Class. HERV-L was first isolated in 1995 (Cordonnier et al. 1995) from placental RNA. Further analysis revealed it to be related to foamy virus and that around 100 copies are present in the genome. The expressed region was a large portion of pol. Earlier this year (Tristem 2000) identified a further Class III family named HERV-S, the longest member of which was found on the X chromosome. This element is 6.7kb in length with no major gene products able to be expressed due to multiple stop codons and frame shifts. This family is most closely related to HERV-L and seven elements were found in the 7% of the genome screened.

4.1.4 Other HERV

Retrovirus-like particles have been detected in human sources not known to be infected by virus, e.g. placenta (Dirksen et al. 1977; Johnson et al. 1990), milk (Moore et al. 1971; Vaidya 1973), breast cancer biopsies (Axel et al. 1972), monocytes from myeloproliferative disease (Boyd et al. 1989), multiple sclerosis patients (Haahr et al. 1992), breast cancer cell lines (McGrath et al. 1974; Keydar et al. 1984) and teratocarcinoma cell lines (Boller et al. 1983; Bronson et al. 1984; Lower et al. 1987). Retrovirus-related antigens have also been detected from other similar sources but since an infectious retrovirus was not shown with these particles they appear to be encoded by HERV. Some were encoded from a single HERV, but others were encoded by complementation and by one or more families. Rare functional ERV elements within a family of defective elements have been discovered in the genomes of other species (Reeves et al. 1984) and also among the retroelements in humans (Matera et al. 1990; Dombroski et al. 1991; Deininger et al. 1992).

There may be other HERV families yet to be elucidated as PCR has yielded different RT related sequences (Bangham et al. 1988; Shih et al. 1989; Medstand et al. 1993) of great variety. However only Class II HML groups (Medstand et al. 1993) were identified by the analysis of RT-PCR fragments (Shih et al. 1989; Wichman et
al. 1992). Hybridisations with Mason-Pfizer monkey virus imply a distantly related element in humans that is yet to be isolated (Barker et al. 1985).

HERV elements that are closely related to exogenous human retroviruses have also not yet been successfully isolated, for HIV the attempt was wholly unsuccessful (Horwitz et al. 1992) and only small segments related to HTLV-I have been found (Mager et al. 1987; Perl et al. 1989).

A HERV with proposed structure similar to murine sarcoma virus on the basis of hybridisation was also noted (Zabarovsky et al. 1983; Chumakov et al. 1985), but analysis of its LTR-U5 implied that it was Alu derived.
4.2 RESULTS

4.2.1 PCR

The reverse transcriptase region was chosen as a primary screen due to its presence only in retroviral sequences. This region had also been successfully employed by others for HERV isolation and was found to have the greatest homology across the HERV families (Tristem 1996). The other alternative was the retroviral LTR which is also viral specific but this region is highly divergent and creates problems in designing primers. There are also many LTR only elements present in the human genome and these may have been amplified.

Primers were designed to cover the majority of the pol region of most HERVs. This was done by sequence alignment of numerous HERVs from different classes as well as other retroviral pol sequences, including exogenous retroviruses. Several other groups have performed similar alignments (Xiong et al. 1990) and this has led to the establishment of seven regions within the reverse transcriptase. Using this designation a conserved domain can be seen in region five, which incorporates either the YMDD or QYVDD at the amino acid level of all retroviruses. A second region for the reverse primer was chosen that shows significant conservation, this region corresponds to the protease active site and contains the LDTGA domain. This results in a PCR product containing the majority of the reverse transcriptase sequence (domains 1-5) as previously reported (Tristem 1996).
### Figure 4.1
Alignment of various retroviral sequences around the primer sites

<table>
<thead>
<tr>
<th>Retrovirus</th>
<th>Protease Domain</th>
<th>Reverse Transcriptase</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>MuLV</td>
<td>QPVTFLEVTDGAQH</td>
<td>DLILQYVDDLLLA</td>
<td>NP040333</td>
</tr>
<tr>
<td>MoMLV</td>
<td>QPVTFLEVTDGAQH</td>
<td>DLILQYVDDLLLA</td>
<td>NC001501</td>
</tr>
<tr>
<td>ERV-9 (pHE.1)</td>
<td>QEIDPELEVDGAFAF</td>
<td>GTLILQYVDDLLLAT</td>
<td>X57147</td>
</tr>
<tr>
<td>HERV-W</td>
<td>QEVNCELVDGAFAF</td>
<td>DTLILQYVDDLLLVA</td>
<td>AF009668</td>
</tr>
<tr>
<td>RTLV-I (HERV-I)</td>
<td>BLETPELDGAAC</td>
<td>QLCILQYVDDHHSG</td>
<td>M92067</td>
</tr>
<tr>
<td>ADP.pol</td>
<td>EKMTFLEVTDGAC</td>
<td>ETQLILQYVDDHHSG</td>
<td>L14752</td>
</tr>
<tr>
<td>4-1 (HERV-E)</td>
<td>KDIDFLEVTDGAEH</td>
<td>GCVILQYVDDHLGH</td>
<td>M10976</td>
</tr>
<tr>
<td>HERV-K10</td>
<td>KQFEGLEVTDGADV</td>
<td>DCYCHYDDEHHCAA</td>
<td>M14123</td>
</tr>
<tr>
<td>BLV</td>
<td>NQALSLEVTDGAEN</td>
<td>QSLILVSYYHDHHIAS</td>
<td>AAF97917</td>
</tr>
</tbody>
</table>

For degenerate primer production, isosine was used to represent any base, the actual sequences used are given in Chapter 2 (section 2.2.7.1). These pairs of primers corresponded to the MLV (murine leukaemia virus) group, a universal group that demonstrated cross type amplification and a more specific to HERV pair of primers.

Using the degenerate primers, PCRs were performed on a pool of genomic human DNA extracted from PBMC. PCR conditions were as described (section 2.2.7.1). Resultant PCRs were run on 0.8% agarose gel and resulted in major bands of around 800bp and 600bp. Several other bands were also present (Figure 4.2). Arrows indicate the 2 major bands.

### Figure 4.2
Degenerate Primer HERV PCR
Chapter 4 Isolation and characterisation of two novel HERV fragments

Lane 1 ΦX HaeIII digest, lane 2 1Kb ladder, lanes 3-5 primer A, lanes 6-8 primer nA, lanes 9-11 primer B, lanes 3, 6 and 9 primer C2, lanes 4, 7 and 10 primer MLV, lanes 5, 8 and 11 primer U.

Expected band sizes from the alignments were around 800bp. Therefore these bands were gel purified in preference. However several other bands were observed corresponding to shorter amplified sequences. These could have been due to partial pol sequences that are incorporated in the genome or may in fact be due to mispriming. This is possible as degenerate primers were used at a relatively low annealing temperature.

4.2.2 Cloning and Selection

The desired amplified band was gel purified using the Qiagen gel purification kit in preparation for cloning. The inserts were ligated into pCR2.1 plasmid from Invitrogen using T4 DNA ligase. This plasmid system utilises a feature of Taq polymerase, which leaves a 3' A overhang on all amplified products. Therefore the plasmid incorporates a 5' T overhang allowing sticky end ligation of PCR products. Competent E. coli cells were then transformed, as described earlier (section 2.2.6), with the ligation products and the transformants selected by growth on ampicillin plates. The ampicillin resistance is conferred by the plasmid backbone, therefore any untransformed bacteria die. A second selection is by the use of X-gal driven blue/white colour, when the vector contains an insert the β-galactosidase gene is interrupted. This results in there being no active β-galactosidase so the X-gal substrate cannot be broken down. However in the re-ligated plasmid without insert β-galactosidase is produced, the X-gal substrate is cleaved and a blue colour is released, therefore these colonies appear blue. Only white colonies were picked, then grown and the plasmids isolated as detailed (section 2.2.5).

These clones were digested using the restriction endonuclease EcoRI that releases the insert from the plasmid backbone. A gel photograph of some positive clones is illustrated below; the arrow indicates a size of 800bp (Figure 4.3). The positive clones were then re-amplified using the degenerate primers to check the specificity of the insert.
Figure 4.3  EcoRI Digest of Clones

Lanes 1 and 6 1Kb ladder, lanes 2 and 3 HERV1 clone, lane 2 EcoRI digest, lane 3 undigested. Lanes 4 and 5 HERV2 clone, lane 4 EcoRI digest, lane 5 undigested.

The next stage was to sequence the inserts using both the -40 forward primer and the rev reverse primer. -10 and +10 primers were also used to obtain more sequence information. These primers were all present on the vector and as such could be used to sequence all inserts; details of the primers are given earlier (section 2.2.11.6). A figure illustrating the various sequencing methods used is shown below (Figure 4.4).
Figure 4.4 Sequencing of retroviral clones.

A: An example of manual sequencing

B: An example of single stranded DNA preparation for sequencing

C: An example of an electrophoreogram from an automated sequencing run.
Over 500 clones were sequenced in this way yielding a surprising number of nonsense sequences. These candidates were narrowed to around 35 sequences that, on further analysis, could be grouped into 12 different sequences. Of these 12 with some homology to known retroviral *pol* sequences two had very high homology yet appeared novel and were chosen for further study. Homology was determined by BLAST searching (Altschul et al. 1990).

These sequences named HERV1 and HERV2 are given below (Figures 4.5 and 4.6).

**Figure 4.5** Partial HERV1 Sequence

```
1 60
c t c g t c g a t a c c g g a g c a t t a c c c a c g t c t a c g c c c t a c c t a c t g t g a
LVDTGATYCSLPQSIGPTC
61 120
tcagaattgtcgctttgttggaacgcagctcgtccccatggttcgtggtgccatgggctacacactga
SELSVVGLDGLVWCWPATQ
121 180
ttcgatttcgcagctaaaggaacacagcccttcagtcattcctctcctctcgttaagccacactgt
FCELRNTAFSHSFLVMPHC
181 240
catccccactctcttgaaagagatatactcacaagacacccccctccccacacacagcatat
PSPLLGLRDILTKH**?
241 300
PLPPQHI
301 360
actcgttgcacacagctcataacctacaagťttggatataacatcgccagctctctctgc
LVATQLNPDVDWTDISPIIA
361 420
agcagatctcacaataacagcttaagttccagccactttccccagttcataacccacatcc
ADLTIQLKSKAFVPQVOYPIP
421 480
cgaacacgcacaaagggctcacaacacccctgcgccccaaagactcagaaggctggactaat
EHATKGLKPVAQRLKQSGL
481 540
agaccatcacaattccacataaacaacacacaatactccccataaagacttcgaggactac
RPSNSPYNTPILPIKKSDGS
541 600
atatagggtggtccaggacctaagagttatacaaccaggcaactttgagccagttcactccgct
YRLVQDRLRVINQALEPVHVPV
581 640
cgtccccgcaccaatatacccttacatctctctctctctctctccccacacacccccacacacac
VPDPYTLHISLISNTTHTY
```
Chapter 4 Isolation and characterisation of two novel HERV fragments

** Denotes a putative deletion.
? Denotes a putative frame shift.
- Denotes a stop codon.
A theoretical translation is given.

Figure 4.6 Partial HERV2 Sequence

```
1   60
cttgtagatagccggcgtctggtagctcagctgtgttccttgctctggtcagtaactgtg
LVDSGAACSSVCFLPPGVTC 61
agctcagaagagcctcatctcttaagggggttaggaggggttaagggcagatactct
SSEELILSGLKGEGFKAKL 121
120
agagccaccgaggttagatacaagaccatagtaactcttcattctctgtagatcata
ESTEVRYQDHSHTHLQPFLLII 181
180
cagctggacacccactactagggagacactctgtctttgagctctagtaacacc
PEAGTNLLGLRDLML***RPCNT 241
240
ccaatactgcgaactcacagacagcctactcagtacgctctctgtctctgtctcatgtcata
PILPVKTDSYRLVQDRLRAI 301
360
cacacaaatagctccagactccacccacttgtttttctctagctctccacaccacttttaagc
NQIVQTTHPIVPVLTPFLSK 361
360
gatccatataatcataataggttgtactgtaaatagttagggtagtcctttggcagttgc
IPYNHQWFTVIDLKDAPGMS 421
480
cctggctagagcagcctataattctctttgatattggaggtgttccttggagcagttgc
LAEDSDIFAFEFEDPHSGR 481
540
acaaacatatggtttctctacacagtctctacaccactttttttttgg
QQYQWMVLPQGFTDSPNLF 540
```

- 105 -
** Denotes deletion.

? Denotes hypothetical frame shift.

- Denotes a stop codon.

A theoretical translation is given.

In order to confirm the presence of these sequences in the genome a Southern blot of normal genomic DNA was performed. This was then probed with radioactively labelled HERV1 or HERV2 insert. No bands were present after short exposure but if the blot was exposed to film for a long period (2 weeks) a smear containing at least 2 bands was observed. An example for HERV1 is presented below (Figure 4.7).

**Figure 4.7** Genomic Blot Probed by HERV1

The genomic DNA was digested with EcoRI and at least 2 bands are visible. Size indicators are given by arrows.
Probe specificity was demonstrated with the use of vector plasmid sequences and a positive control of a small amount of insert run. However the signal was very weak for both HERVs and in particular for HERV2 (data not shown) indicating a low copy number within the genome. BLAST searches of the databases also revealed no isolated sequences with a great enough homology to indicate other closely related members of a similar family implying that these sequences are present at low copy number. PAC screening as detailed below however revealed numerous hybridising sequences that gave a better indication of copy number. Further investigation will reveal the true copy number within the human genome in particular as the human genome project releases further sequence information.

4.2.3 PAC screening

To obtain further sequence data from the isolated HERVs a human PAC library was screened.

A human genomic library was obtained from the HGMP as spotted filter arrays. The library was formed as BAC PAC clones. The PAC library used was RPCI11. Pieter de Jong and his group at the Roswell Park Cancer Institute in Buffalo constructed this library. The vector is pCYPAC2. The DNA source is a normal male blood donor and the average insert size 110Kb. The library consists of approximately 120,000 clones in 314 384-well microtitre plates with 25% of the clones lacking insert (Ioannou et al. 1994; Ioannou et al. 1996). The whole genome is represented at least twice within this library.

HERV inserts were gel purified and radioactively labelled by random priming. These labelled probes were then hybridised to the spotted filters and positive pairs of spots determined (Figure 4.8 A and B).

After the elimination of known false positives, due mainly to inserts containing repeat DNA sequences, around 30 positive clones were chosen from the filter arrays. Due to the reported covering of the genome twice in this library a copy number of less than 5 for each element over the whole genome can be assumed.

These 30 clones were ordered from the HGMP and the PACs isolated as described (section 2.2.14). These PACs were digested with numerous restriction
endonucleases then run on 0.8% agarose gels. (Figure 4.8 C). The gels were then Southern blotted and probed using the labelled HERV fragments (Figure 4.8 D and E). This led to the selection of clones with positive bands of 2-5kb. The most promising candidates were digested and run on agarose gels again for gel purification.

**Figure 4.8** PAC Screening

![Image of PAC Screening](image-url)
Chapter 4  Isolation and characterisation of two novel HERV fragments

Figure 4.8:
A: Southern blot using HERV2 probe against PAC grid
B: Southern blot using HERV1 probe against PAC grid
C: BamHI digest of selected PAC clones
D: Southern blot of BamHI digests of PAC clones probed with HERV1
E: Southern blot of BamHI digests of PAC clones probed with HERV2

1 to 23 inclusive various BamHI digested PAC clones, 24 1Kb ladder (Gibco), 25 λ HindIII digest.

The most promising digests are shown (Figure 4.8). BamHI digestion of clone 27K22 led to the release of a band of 3.8Kb corresponding to a band that hybridised strongly to the HERV sequence HERV1. A digestion with BamHI of 26J13 led to the release of a band of 2.7Kb that hybridised strongly with HERV2. These bands were subsequently gel purified and cloned into bluescript. The ligation was performed by utilising the cohesive 5' overhangs created from the enzyme digestions. Again a blue/white screen was performed to isolate clones containing inserts. The entire insert was then sequenced from both ends initially; the sequencing was completed by primer walking along the insert. The resultant sequences are given below.

Figure 4.9  Sequence of HERV1

```
1 accttttcacgaaccctgatatttcatcaacagctcaagaaccctgtccccaatccacc
   T F S S N P D I Y I N R L K N L S Q S T
61  aaactctctttgcatgatatctgttatctctctgaatatactcttgtacccccagaagcat
   K L S W H D I Y V I L S N T L L P Q K H
121 agattgaaagaaagactggattgcgcacacaggtgccgaatgacaccgacacagt
   R L K R K D W I A A Q A C E - S A S T L
181 ccctattaggcacacacaggtgccgttgcaccattctaggtatatctctttcggaggttccc
   P Y - A H R G C C S I L G I S L P I V P
241 actcagagctgatcagaccgggtgctgcataactacatgaccacattctcctactcagcgcgt
   T H R P E P G V A - L H D H I P Y L R A
301 Gag (MHR)
tagtaacactgtccatatggatgtaatattgaaagctcaagaggtgtcctcaggcggcg
   - H C P Y G C N I E S V K R L S Q R A
```
Isolation and characterisation of two novel HERV fragments

361 Gag (MHR)

gataaagatccacccgctattcccttcacccaaggaaggctttgaaagataataccaccac

421

dknptafpshtlealyth

481

agatatccctcagactctgtaaacttctgctgactgtaatcgctctctaagttacatc
edtpqtpternelnlnlalvkvis

541

acaggggtacagtgaattataattcgtatataagccccctggacactttgtgcctagtaacaagcttc

601
tcgctgcatccattttgcaacaccacatcctgctgctccaggaacaatacctgccctatacca

LAASICTQPTLAVQGHNRPYT

661

ctaatccagctggccaaattgatcagcaagatgtgtagrcgcaagtccccgctctgctgtagt

721

lgnlakcsspsvsrsplllv

781

cctaatccacagaggattaaactgagcaccagggctccagctccatccatctctctctgtatc

791

Protease

agagccctcgcgtaacccctctttaatgggaggctagccagtctatttcctctctgcgtagaccgg

eprvltlmggpyvlpvdtg

961

gcatctctctgcaqctctccacagctctattcggctctcactgtgtgcagataattgtccgt

1021

atycslpqssigptcselsv

1081
tgtaggactcgtggtgtgtgtgctgccatggctacacaactattctgttgttcggagct

1141
caggaaacagccctccagctattccctctctgcttacagcaactgtgctcatccacaccacttt

1201

rnafahtsfsflvmphcsppll

1261
tacagcctatcagggcattttccacagatctcctccagttttgatccacactttcaca

1321

[Domain1]
aaggggtccaaaccgctggccaaagactcaagcagtccggactataaagacccatcctcatt

kglkpvaqrlkqsglirpshen
<table>
<thead>
<tr>
<th>1381</th>
<th>Domain1</th>
<th>]</th>
<th>Domain2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ccaccatacaacacaccaaatctccccataaaagaagtcgcagttgagcatatataggtggctc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPYNTPILPILPKKSDGSSYRLV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1441</td>
<td>Domain2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aggccttaaggtatatcaaccagggcatgcttcctgcccagttcggacccat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QDLRVINQALEPVPVVPVP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1501</td>
<td>]</td>
<td>[</td>
<td>Domain3</td>
</tr>
<tr>
<td>atacccctacatctctctcatctctctcaacacaccacaccactacactgcaatgaactaa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YTLHSLISSNTTHYTIAEL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1561</td>
<td>Domain3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aagagtcctttctccaccatctccccctcatctctgattcctccaacaaaaagtcctttgctctactct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KDAPFLHSDSQU LP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1621</td>
<td>]</td>
<td>[</td>
<td>Domain4</td>
</tr>
<tr>
<td>ggagcataacacactcactctccacagctcactgactttgcaccttcgttctcctcaagga</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTENDTQSQQLTWTLPQG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1681</td>
<td>Domain4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttcaggatatagctcttttcacctggcagctctagcctccataacccctttaccttaaacc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQDSLHFFRQVLAQLDASNL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1741</td>
<td>]</td>
<td>[</td>
<td>Domain5</td>
</tr>
<tr>
<td>ttccccctagccatctctttccataacgctcgagcctccacctttttatagacagccataaa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSPLSHLLQYVDDLLLYSPAI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1801</td>
<td>]</td>
<td>[</td>
<td>Domain6</td>
</tr>
<tr>
<td>acatctccgcgacactcactctggacactttttgcaccttcgttcttgaaaataagggattcaagaa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NISRTHGTLHLLLLTNGYR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1861</td>
<td>Domain6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttctctctataaagcacaacagctcttcataacccctttacctgttggtggtaccacacgtct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISPIKQHQHSTPNATYLGVPR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1921</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ctccccatgtcacagcctagctccttcggtatatgtctaatattagacagccttaactccac</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPMSSRAMTPVSAILDSTP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1981</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cttctttaaagtagtactaaatactttttatttagttcatgcctttatcaggtbatgga</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSLK--ILSFLGLSGLIKVW</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttccgagctttcactctcccagcagcagccgtcttataaacaacacgcagggccacccagacata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPSFTPLAQPLYKTAKGPSH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tctctctataacatcactacaacatcactgctcctacctcggcagactccaatcaccctccc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILVNS--HTILYLFRLQITL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2161</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttccgagccctgcgtttttctcttacccaaagctctccacaacactttttactttctctgtgcca</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTAPALFNPVSNQFPTLYPA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2221</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aaataagcctaagaataagcgtttggtattcatttccacagcagaagttagaatctcctctacttt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KKSRLIALGLSQQK--NPPT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2281</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tctctccccgtatctcagctctgctctgatcaattgaccaactacgcagaaagtttgccccctgctt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSPVSRSDQLTNHVKGWPC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2341</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gtattactgtactgcacacacgtcactctttctagccccctacagcagcacaagttgcacactttcag</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CITV?TAHAALLALQSKMHSFS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4  
Isolation and characterisation of two novel HERV fragments

2401
ccta atgc aactatatct cagtcc cacacaactatatgtgccttc cagtgacgttgcattgagagc cattgag

2461
tctgctcc accc agc gacagacatt catgtgc cagagcccc cgccttt ttaaa agat ccc aagtt

2521
ttgcat ggg cc aagt gttcctct cag ccc aag cat cctactctgc atctctg cctactctctctctctcctctctctctctctctctctctcctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc...
Chapter 4  Isolation and characterisation of two novel HERV fragments

3861  Alu
gcctggccctccaaagtgctgg

Alu inserts at 3' end are underlined and a translation of the putative pol and gag are
given. The protease active site is labelled as are the seven domains of reverse
transcriptase. The conserved MHR region of Gag is also denoted. Frameshifts are
labelled with "?" and stop codons are labelled with "-.

The isolated HERV1 sequence covers a portion of gag and almost the entire
pol followed by an Alu repeat. The pol region covers the protease region and domains
1-7 of RT all of which are labelled. There was no env sequence present and an Alu
element is inserted at the 3'.

The putative Gag MHR region begins at nucleotide 345. The protease active
site starts at nucleotide 885. The RT domain one extends from nucleotide 1353 and
the end of domain seven is nucleotide 1907.

Frame shifts to maintain retroviral homology were at nucleotides 433, 769,
1169, 1605, 1876, 2224, 2352, 2520, 2608, 2705, 2741, 2779, 2934 and 2976. Stop
codons were found at nucleotides 103, 187, 268, 301, 304, 553, 559, 588, 760, 856,
931, 1006, 1993, 1996, 2116, 2266, 2599, 2605 and 2752. There was also a deletion
with respect to HERV-XA sequence at nucleotide 1169.

Figure 4.10  Sequence of HERV2

<table>
<thead>
<tr>
<th>Domain1</th>
<th>[ Domain2</th>
</tr>
</thead>
<tbody>
<tr>
<td>841</td>
<td>Domain3</td>
</tr>
<tr>
<td>361</td>
<td></td>
</tr>
</tbody>
</table>

-113-
Chapter 4  Isolation and characterisation of two novel HERV fragments

421 Domain3 [ ]
  cctggtcagacagcagatatatggcttttggaattggagaccaccccactcagggcggaa
  LAEDSDIFAFEWEDPHSGRK

481 Domain4
  acaacaaatcatatcaggttgcttgccctcaagggtcacagatcctccctaatctttttgg
  QQYQWMVPQLPQFTDSPNLFG

541 Domain5 [Domain5]
  ccaaatatttagaacaagattcttagaaaaagttgtctacccagaaacaaatatgctctctca
  QILEQVLEKVVIPEQICLLQ

601 Domain6 [ ]
  gtacatggatgattcctcatatccgccgagaaatatggaaacacaactggtctcagcac
  YMDDLIEGESTQLGPST

661 Domain7
  tcataattaaaaaccctcactgacgctcagcataagctacaagctcacaatgtaagctg
  HLFNHLQVEGLQVSKHKLCE721 [ ]

781
ggtggaatgggttaataatctatcagctcatgtccctccctcataagctacaagctcacaatgtaagctgcatgtagtaagctacaagctcacaatgtaagctg
  AEPVVKYLGKLMISAGKTRIG

841
gaaattctctgcttcagcatctcgacgctcagctcacaatgtaagctacaagctcacaatgtaagctgcatgtagtaagctacaagctcacaatgtaagctg
  KFLGGLVYCHLGMNSHALS

901
taagatttcatatcagaaaaactctgctccgagactctgacgctcatactggtgaccctcaca
  KILYEKLAPGHPQHILWTSK

961
tagctctggatagatcacaagaaacactttaagaaagataagctttataacaagctcatacgtcttttgcctcacaatgtaagctacaagctcacaatgtaagctg
  CLDEIKEKLKEEPLITAHALAT

1021
tacctctctggagtataaatcctaccccttaattttgtagctcagctcataatgtaagctacaagctcacaatgtaagctgcatgtagtaagctacaagctcacaatgtaagctg
  LPSLEYTIHLFVTVDIVAL1081

1141
gctctctatctcagaaagcattgaggtccctctccagctcataatgtaagctacaagctcacaatgtaagctgcatgtagtaagctacaagctcacaatgtaagctg
  GVLFGQEHGWSWQPFKLFLSKL

1201
tgataattggtcaccctggtctgcccggctgtttatttatattgcagctacactgctggat
  L-IVTRGPRCI-SNAHTAG1261

1201
tgatagagaggttagtagagagcataacccctctgctcagagaaaagataagctacaagctcacaatgtaagctgcatgtagtaagctacaagctcacaatgtaagctg
  LDEECRKLTFG-KLSTPLS

1321
aagttgagcagcaagctttaaatataacaaacgtggcagccaggaacacacagggaggagaaaaattttt
  QVRUTAILNKTAGRPTESRIL1381

1381
aagttgacttgggtatttttcatatatagaaacagaaagagacactttatgttttaataatcataatg
  KYWISLLKKD?KEHLCLNL1381

1441
dacacagactcaccgcagcaggggtctttgaaaagctctgctctgcttcagagacagggcttctact
  DNQTHRPRFDLEKLPFFKTGLH1441

1521
tagatgggaatgtttagttaaagaaacttagggagagagacaacactgagatgtatatcattaa
  LFVDGSSEMI-GKRHNVYP1521

1521
  tagatgggaatgtttagttaaagaaacttagggagagagacaacactgagatgtatatcattaa
  IDGEMPLVWLRTEARKLPNTW
Chapter 4  Isolation and characterisation of two novel HERV fragments

1581
tgctcaatctttgtgagcttttttcagccaagccataaagacttcgaaatcaaga
  A Q S C E L P S F S Q A I K D L Q N Q E

1641
atgaactatctacagcagcatgtatgctattagagtctacatcacaatattttgaaat
  - T I Y T D S M Y A F R V T H T F - K I

1701
ttggactcaatgagggttagctgtagcagcaaaaggttagtgaacacttttccataaaaaacattgtat
  W T Q - G L S D S K G Y E L L H K N L I

1761	attcaagtagcagagaaaattcgccatatctgtgtgccagag
  I Q V M E N L Q L P Q E I A H I C V P R

1821
gcataaaataaagggttagcctctccgtagtggtgaggctagctacatcagcaggctcccgaag
  H Q I R L S F E C - G N D L S D Q I P K

1881
acaagcgcacactcccctccgaaacactccgctcttttaacttaaaccctgttttaacctctatt
  Q A A I P S N T P I F Y L T P C L S P F

1941	tacctgcacattgcttttattcctgccctttgtaaaagcaagtataaaggctagctgaggacaaggc
  T A T C L F S S L V K Q V - G Y V G T A

2001
[g LTR
ggagatcaccatatggctagctcttagaccacaaattatattgttataagcggtaaa]

2061
[ LTR
aataagctgttaaatttaacagcgagattgtttaatacagtaagaaataacactttcaca

2121
 LTR
aggttgagcttttattaagtttcctttgtcttttgtctccctgcctttcaagggcagacttccct

2181
 LTR
tacctcctgtgtctcctctgctgtgtaaacaacottccccgcagctgtatctatagag

2241
 LTR
cccacattccccaaactgccaaccactctgtgatattaccctttgcacataaatggccttccct

2301
 LTR
cggcacaacctgtacctttcctgccctccccagcaattgtaatcacattcatgcacatttccaag

2361
 LTR
ttagcaacccggcttcaggcttgatagcaggtgccaactgtagtcaatggagtccagata

2421
 LTR
cagtagccagggacaagctgcgcttgtagacataaaaccctctgtttttcttttgccgggtgtgc

2481
 LTR
ttcctcgcacagcaccaccaacttataagggagacacattaagaaatttgccttggctgta

2541[LTR]
  [ Alu
tagacc  atgattttttttttttagacgcgttttttgctctttgtgcccaggggtggaatgc

2601
  [ Alu
aaggcgatactcgcctacaccaacactctctgctccttagttgccacaagactttcactgcc

2661
  [ Alu
tcggcctgccagatgagctgagttataaaacagttgccacacagccctggctaattttgtat

2721
  [ Alu
ttttagtagacagcggttttcctcagtgtgctaggctgtctcatactccgcacctcag

2781
  [ Alu
gtttctgccctgcctgccctccccaaagttgtggattacagggacactcgtgccc

2841 Alu ]
agctcat
A putative translation of the *pol* is given and the 3' LTR is underlined. The Alu repeat is also labelled. Domains 1-7 of reverse transcriptase are noted. Frame shifts are labelled with "?" and stop codons are denoted with "-".

The isolated HERV2 sequence begins with the protease active site. Nucleotide 225 marks the beginning of domain one of the reverse transcriptase, nucleotide 758 denotes the end of domain seven. The 3’ LTR covers nucleotides 2035-2547, this is followed by an Alu element. Frame shifts to maintain retroviral homology are at nucleotides 259, 988, 1270, 1350 and 1545. Stop codons are encoded at nucleotides 233, 814, 820, 895, 994, 1003, 1102, 1144, 1177, 1234, 1490, 1540, 1643, 1692, 1710, 1850 and 1980. Deletions with respect to HERV-Ia are found at nucleotides 233 and just prior to the 3’ LTR. This second deletion has removed the entire *env* region.

### 4.2.4 Homology Searching

The obtained sequences were investigated by BLAST searching (Altusch 1990) and sequence alignment packages including multialin, clustal W and pileup.

#### 4.2.4.1 HERV1

The isolated sequence covered the entire *pol* region with a small deletion by comparison with XA. A substantial portion of *gag* sequence was also present and three Alu elements were inserted at the 3’ end of *pol*. This seems to have led to the loss of *env* genes. The genes are interrupted with several stop codons and frame shifts, however there are some viable shorter ORFs that will be discussed later. 5’ to the *pol* sequence there is a disrupted protease and before that a portion of the *gag*. In total 3.0kb of retroviral sequence was obtained. No LTRs were found so dating the integration of this element is difficult. The short adjacent genomic sequence was used in a search for chromosomal location but this was unproductive. However the genome sequencing project is progressing at such a rate that this information should be obtainable in the near future. Instead PAC FISH was employed (see Chapter 5).
The greatest homology for this sequence was seen to the HERV-XA family and next to the HERV-H family.

Using a portion of the originally PCR amplified region of pol (starting with the QYP amino acid domain prior to Domain one of RT and extending to the primer site in Domain five), homology with the HERV-XA element (Tristem 2000) was 77% and the same region showed around 60% homology with RTLV H. The table following this section shows the relatedness to exogenous viruses over this same region. The most homologous viruses were MuLV, FeLV and BaEV, which are all closely related to each other, with up to 48% homology over this region. Therefore this sequence appears to be a Class I retrovirus with the highest homology to type C exogenous retroviruses.

Due to the high homology with the XA sequences already obtained and the clustering within this group it is likely that this isolate is in fact a member of the HERV-XA group. Although if this isolated element is related it must be on the edge of the group as the homology is under 80% at the greatest. Isolation of a similar element and especially intact LTRs will shed more light on the taxonomy, but at present it is most appropriate to include this element within the HERV-XA cluster.

An alignment of the most conserved region of pol is included.

<table>
<thead>
<tr>
<th>Retrovirus Name</th>
<th>Percentage Homology to HERV1</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV-XA</td>
<td>77%</td>
<td>U29659</td>
</tr>
<tr>
<td>RTLV H (HERV-H)</td>
<td>58%</td>
<td>D11078</td>
</tr>
<tr>
<td>HERV-F</td>
<td>53%</td>
<td>Z94277</td>
</tr>
<tr>
<td>MSRV (HERV-W)</td>
<td>48%</td>
<td>AF009668</td>
</tr>
<tr>
<td>FeLV</td>
<td>48%</td>
<td>L06140</td>
</tr>
<tr>
<td>MuLV</td>
<td>48%</td>
<td>U13766</td>
</tr>
<tr>
<td>PERV B43</td>
<td>47%</td>
<td>AJ133818</td>
</tr>
<tr>
<td>BaEV</td>
<td>45%</td>
<td>D10032</td>
</tr>
<tr>
<td>HERV-K10</td>
<td>26%</td>
<td>M14123</td>
</tr>
</tbody>
</table>
Figure 4.11  Alignment of HERV1 putative translation of a region of pol

<table>
<thead>
<tr>
<th>HERV1</th>
<th>HERVXA</th>
<th>HERVH</th>
<th>BaEV</th>
<th>FERVB43</th>
<th>HERVK</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVHPVDPYTLISLISXNSTTHYTAIELKDAFTIPLSNDXSP---V---DN---Q---P---N---L---</td>
<td>--MG--L--^--Q--GLPSPAM--IPKDWPLIIID--C------AEQDCEK--IPA--NKEPATR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2.4.2  HERV2

HERV2 showed the greatest homology to RTLV I (HERV-I) sequences (79%) and can probably be considered as a member of that group. The pol region contained a small deletion of around 200bp towards the 5' that explained the truncated PCR product of around 600bp. From further analysis the majority of the pol region was determined as well as a 3' LTR. It appears that the env sequence is completely deleted in this element and the pol region is followed immediately by the 3'LTR, this is then followed by a single Alu element. The 5' LTR was not obtained and is probably also deleted. As with HERV1, the coding sequence contained numerous stop codons and frame shifts and is unlikely to express major transcripts. An alignment of the most conserved region of pol is included. The table below shows the percentage homology over 135 amino acids from the 3' end of deleted region of Pol to the PCR primer site. This sequence can be classed as Class I as the most homology was to the MuLV group of type C viruses.
### Table 4.1

<table>
<thead>
<tr>
<th>Retrovirus Name</th>
<th>Percentage Homology to HERV2</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTLV la (HERV-I)</td>
<td>79%</td>
<td>M92067</td>
</tr>
<tr>
<td>ERV 9</td>
<td>51%</td>
<td>X57147</td>
</tr>
<tr>
<td>HERV-Rb</td>
<td>48%</td>
<td>AC004045</td>
</tr>
<tr>
<td>MuLV</td>
<td>48%</td>
<td>U13766</td>
</tr>
<tr>
<td>BaEV</td>
<td>48%</td>
<td>D10032</td>
</tr>
<tr>
<td>GALV</td>
<td>47%</td>
<td>U60065</td>
</tr>
<tr>
<td>Sus Scrofa</td>
<td>47%</td>
<td>AF038600</td>
</tr>
<tr>
<td>MoMLV</td>
<td>42%</td>
<td>PO3355</td>
</tr>
<tr>
<td>RRHERV-I</td>
<td>38%</td>
<td>AC002992</td>
</tr>
<tr>
<td>HERV-K10</td>
<td>29%</td>
<td>M14123</td>
</tr>
</tbody>
</table>

**Figure 4.12**  
Alignment of HERV2 putative translation of a region of pol

<table>
<thead>
<tr>
<th>HERV2</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV2</td>
<td><strong>PCNTPILPVK?TDG*SYRLVQDLRAINQTV</strong></td>
</tr>
<tr>
<td>HERV1</td>
<td>QPFIPLEGMLGLKPIERSLINDGLLEPCMS-Y-------- KS------K-------T-</td>
</tr>
<tr>
<td>ERV9</td>
<td>-Y-R--^AHK--QN--VKH--KAQ--VRK--S-------E- QRP^--------L--EA--</td>
</tr>
<tr>
<td>HERVrb</td>
<td>-Y-LKK--LVEIQPALIWFLQY--R--Q--SYD--F--KPHSHK--F--H--DT--</td>
</tr>
<tr>
<td>MuLV</td>
<td>-Y-MSH--AR--I--H-QR--LDQ--I-V--Q--W--L--KPGTKD--P-------EV--Kr--</td>
</tr>
<tr>
<td>HERVk</td>
<td>-W-L-KQKLBA--HLLANEQLEK--HI--SF--W-S-VFVIQKKS--^KWHTLT--V--AVI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HERV2</th>
<th>61</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV2</td>
<td>QTTHPIVPVLTFTFLSKIPYNNQWFTVIDLKDAF^GMSLAEDS<em>DIFAFEWEDPHSGR</em>KQ</td>
</tr>
<tr>
<td>HERV1</td>
<td>---N-V--NPYTI------------------------WAC^----E-R-T---------QL-X^--</td>
</tr>
<tr>
<td>ERV9</td>
<td>ISLY--V--NPYTL--Q--ESEA------L------FCVP--HSD--QPL-----^--THN^TS</td>
</tr>
<tr>
<td>HERVrb</td>
<td>EDI------ANPYTMFASL--KD--E------L------FCIPVEIE--QLL------T--ETAA^QF</td>
</tr>
<tr>
<td>GALV</td>
<td>-DI--T--NPYNL--SL--PS--T--YS--L------FCLK--HPN--QSL------K--EK^NTG</td>
</tr>
<tr>
<td>MuLV</td>
<td>EDI------T--NPYNL--GL--PS--Y--L------FCRL--HPT--QPL------RD--EM--I^S-</td>
</tr>
<tr>
<td>HERVk</td>
<td>-PFG--LQ---G--L--SPAM--KDPW'^LI----F--FTIP----QDCEK----TIPAINNKRAT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HERV2</th>
<th>121</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV2</td>
<td>QTQNMVLPQGFTPDSNPNLFQILEQVLEKVVIPEQ^ICLLQYMDI</td>
</tr>
<tr>
<td>HERV1</td>
<td>W--------M--------D--SV--K--L--------</td>
</tr>
<tr>
<td>ERV9</td>
<td>-LT---------R--H--A--AKD--GHFB^SPGTLV--V--L</td>
</tr>
<tr>
<td>HERVrb</td>
<td>--C--T-------KN--SI--EA--I--DEA--RLQLENG^V--V--L</td>
</tr>
<tr>
<td>GALV</td>
<td>-LT-TR--KN--T--DEA--HRD--ALFRAHNPQVK--V--L</td>
</tr>
<tr>
<td>MuLV</td>
<td>-LTWTR------KN--T--DEA--HRD--ADFR--QHPDLI--V--L</td>
</tr>
<tr>
<td>HERVk</td>
<td>RF-WK-------MLN--TICQTFVGRA--QP--REKFSDCYIIH—I---I</td>
</tr>
</tbody>
</table>

---

119
In order to further establish the phylogeny of the isolated HERV-elements alignment with previously detailed HERV sequences was undertaken. Reverse transcriptase domains 1-5 were used, the same region as detailed above. When using the majority of HERV families reported HERVI clustered with HERV-XA and HERV2 clustered with HERV-I as mentioned previously. An example of an alignment used is given below.

**Figure 4.13 Alignment of HERV**

```
    1  50
S .VTSTHQYHL PGGHTE. ITK TIKKLEELQI VSGTHRYPYH ........FPYW
L XXIJKH.YHH PGGIAB.ISA TIKDKLNAVG VIPITSFLN ....SPFW
S71 .PVQVQKXS V SQQRERINP HQQQLQTQGI LT.CQSWNT ......PFF
E ........... PV LREALEGIQV HLCITLTFRI VPCQSPFNT ....PIL
RRHERVI .LVXVCQGPL PKEVIEGITQ HLNRLTXQQGL IVKCKSASN ....PIL
R .LVXVRQYPF PXEAIKXGCK HLDWLYKHR LVRQCSFRT ....PIL
I .VVRRKQFFI PEGMLGLK P ILESQINDGL LEPCMSYPNT ....PIL
2 ................. .PCNT .......PIL
ADP ............... GL LKPCMSYPNT ....PVL
XA ................. FPPLQYQYPL PXH TKGL. P SAXLTDQQS...XTNPYPNN ......PIL
1 .FPQV.QYPI PPHATKGLKP VAQLKQOSGL IRPSNSPYNT ........PIL
H .YPAQCOQPI POHALQG LKP ITLRLEHL LKFIPNSYPN PVL
F .AIKPVQGPL NPQNGGQP KP KISXQKGPL NLTHSPNPT ........PIL
W .FPYRTYQYPL RPEALQGQKQ QDKAQLQGK VLPFQSPCNT ....PVL
ERV9 .FPYRTQYP FPHELQGKLPQ KVKQGKLP PVLKFQSPCNT ....PVL
Z .THVRQGPL PREAQGVXQ LIAKFNLYRL QPCQGQPCNT ....PVL
Rb .YPCRQGQFL QLEAKGLQSP LIKXKTHQP IS-CNPSNT ....PIL
FRD ............... QVHLQGKLPQ KVKQGKLP PVLKFQSPCNT ....PVL
HS .LPQXIQYAI PQQVALISLP KISXQKGPL NLTHSPNPT ....PVL
F .PLSQXQYAI PQQVALISLP KISXQKGPL NLTHSPNPT ....PVL
K .PVWNNQGPL PQQKLEAHGL QDKAQLQGK VLPFQSPCNT ....PVL
    51
S PVRKPDGTO ....... WRMVTDY RELNATAPPL HAAVPSIMDF MEHLLTEGQ
L PVRKQGTO ....... WRMVTDY RELNATAPPL HAAVPSIMDF MEHLLTEGQ
S71 .PVQK ... R TDYWYVQO Q TEVNQIVO Q HPTVNPYTL LRLLPPE.HT
E PVPK ... R TKDNYRQO Q RLNNQATVL HPTVNPYTL LRLLPPE.HT
RRHERVI PVKHP ....... NGEYRQVQFD WVKANATVI YAIVNYPNT GMIPAB.AM
R PAKXSQPGP SNVRPFQO Q RANFRATAI HLVTWNLTLT MLIPAT..AT
I PTVKSPGDM ....... YLVQOQ PLVQAVQIPL HPYVFPNYPYTL LPLPSS..TT
2 .PV.KTDG.S ....... YRLQOQ RAINQINVQQT HPYFVLPFPF LSKIPY.HQ
ADP PTVKPDGDM ....... YLVQOQ PLVQAVQIPL HPYVFPNYPYTL LPLPSS..TT
XA PIKXSGDM ....... YLVQOQ PLVQAVQIPL HPYVFPNYPYTL LPLPSS..TT
1 PIKXSDG.S ....... YRLQOQ RAINQALVPL HPYFPDPFTT MSLISN..TT
H PVLQGDM ....... YRLQOQ RAINQALVPL HPYFPDPFTT MSLISN..TT
F PIKXSGDM ....... YRLQOQ RAINQALVPL HPYFPDPFTT MSLISN..TT
W GVRKNGQ ....... WRQLGQO Q RINEAVPL YPAVSNPFYT LSLIPER.AE
ERV9 EVQRQ ....... WRQLGQO RLINEAVPL YPAVSNPFYT LSLIPER.AE
Z PVKXNQOQ ....... YKVRQOQ QMVNEAAPPY VP.YPVHTI LTQVSED.AN
Rb PLKXPHSHQ ....... YKRQVQO Q RAINQCVVL HPYVFPNYPYTL LPLPSS..TT
FRD TVKXSMQ ....... YRQLGQO Q MVNEAAPPY VP.YPVHTI LTQVSED.AN
HS PIKXPDG ....... RXGWRFXQDL SEIKYYIIP FSVXNPNAL LPLPSS..TT
```
### Chapter 4 Isolation and characterisation of two novel HERV fragments

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fb PVKLDGA... YHLVXDL RLLIQAVLPV CPVPPYTYTL LSTIPS.D TT</td>
<td></td>
</tr>
</tbody>
</table>
Table of Accession numbers used for the alignment.

<table>
<thead>
<tr>
<th>HERV</th>
<th>Abbreviation</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV1</td>
<td>1</td>
<td>This Thesis</td>
</tr>
<tr>
<td>HERV2</td>
<td>2</td>
<td>This Thesis</td>
</tr>
<tr>
<td>HERV-S</td>
<td>S</td>
<td>AC 004385</td>
</tr>
<tr>
<td>HERV-L</td>
<td>L</td>
<td>X 89211</td>
</tr>
<tr>
<td>HERV-S71</td>
<td>S71</td>
<td>AC 002992</td>
</tr>
<tr>
<td>Clone 4-1 (HERV-E)</td>
<td>E</td>
<td>M 10976</td>
</tr>
<tr>
<td>RRHERV-I</td>
<td>RRHERVI</td>
<td>AC 002992</td>
</tr>
<tr>
<td>ERV 3 (HERV-R)</td>
<td>R</td>
<td>AC 004609</td>
</tr>
<tr>
<td>RTLV Ia (HERV-I)</td>
<td>Ia</td>
<td>M 92067</td>
</tr>
<tr>
<td>ADP pol (HERV-ADP)</td>
<td>ADP</td>
<td>L 14752</td>
</tr>
<tr>
<td>HERV-XA</td>
<td>XA</td>
<td>AC 000378</td>
</tr>
<tr>
<td>RGH2 (HERV-H)</td>
<td>H</td>
<td>D 11078</td>
</tr>
<tr>
<td>HERV-F</td>
<td>F</td>
<td>Z 94277</td>
</tr>
<tr>
<td>MSRV (HERV-W)</td>
<td>W</td>
<td>AF 009668</td>
</tr>
<tr>
<td>ERV9 (pHE.1)</td>
<td>ERV9</td>
<td>X 57147</td>
</tr>
<tr>
<td>HERV-Z69907</td>
<td>Z</td>
<td>Z 69907</td>
</tr>
<tr>
<td>HERV-Rb</td>
<td>Rb</td>
<td>AC 004045</td>
</tr>
<tr>
<td>HERV-FRD</td>
<td>FRD</td>
<td>U 27240</td>
</tr>
<tr>
<td>HERV-HS49C23</td>
<td>HS</td>
<td>Z 93019</td>
</tr>
<tr>
<td>HERV-Fb</td>
<td>Fb</td>
<td>Z 83745</td>
</tr>
<tr>
<td>HuRRS-P (HERV-P)</td>
<td>P</td>
<td>Z 98255</td>
</tr>
<tr>
<td>HERV-K10</td>
<td>K</td>
<td>M 14123</td>
</tr>
</tbody>
</table>

For true phylogeny tree formation was necessary. Trees were produced using the Phylip package and in particular Protpars, yielding the most parsimonious tree. The alignment in figure 4.13 was used for confirmation of the isolated HERV clustering.
The isolated sequences HERV1 and HERV2 are shaded.

The next page shows the phenogram with bootstrap support for the branches.
Chapter 4

Isolation and characterisation of two novel HERV fragments
From this phylogenetic analysis using the reverse transcriptase region HERV1 clustered consistently with HERV-XA and HERV2 clustered with HERV-I. In order to analyse this more closely further phylogenetic studies were undertaken using members of each family as well as representative members of closely related families.

### 4.2.5.1 HERV1 Phylogeny

Due to the limited sequence information available for most of the XA elements a more 3’ region of the pol sequence was used for the phylogenetic studies. An example of the resultant alignment is given below. HERV-H, HERV-F and HERV-Fb sequences were also included in the alignments.

**Figure 4.15  Alignment of XA elements**

<table>
<thead>
<tr>
<th></th>
<th>HERVFb</th>
<th>HERVF</th>
<th>HERV1</th>
<th>HERVXA</th>
<th>XA34</th>
<th>XA35</th>
<th>XA36</th>
<th>XA37</th>
<th>XA38</th>
<th>HERVH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PPTXFDGS</td>
<td>APLDFTGSS</td>
<td>DDQLFDGST</td>
<td>NDQFLFDGS</td>
<td>DDQLFDGS</td>
<td>DDQLFDGS</td>
<td>DDQLFDGS</td>
<td>DDQLFDGS</td>
<td>DDQLFDGS</td>
<td>DHTWFIDGSS</td>
</tr>
<tr>
<td>50</td>
<td>FLHQGRWHAQ</td>
<td>SKNHPL.QAG</td>
<td>SRPARIQDQC</td>
<td>SRPAGL.QNC</td>
<td>SRPTSYPPDC</td>
<td>SGPTS.QPNQ</td>
<td>SGPTG.QPNQ</td>
<td>SRATGCPCN</td>
<td>SKSVNL.QX</td>
<td>TRPRNHRSP .</td>
</tr>
<tr>
<td></td>
<td>YAIIVSSTPH</td>
<td>TAITEGHHDD</td>
<td>WICSCFTLTK</td>
<td>WICSCFS.TK</td>
<td>WICSCFP.TK</td>
<td>WYIICPF .. K</td>
<td>WICSSFS.TB</td>
<td>WICSSRS.T</td>
<td>WICSTN.QX</td>
<td>WICSTN.QX</td>
</tr>
<tr>
<td></td>
<td>TH ........ TI</td>
<td>TH ........ TI</td>
<td>X ........ L</td>
<td>X ........ L</td>
<td>X ........ L</td>
<td>X ........ L</td>
<td>X ........ L</td>
<td>X ........ L</td>
<td>X ........ L</td>
<td>X ........ L</td>
</tr>
<tr>
<td></td>
<td>EANLLLLSTT</td>
<td>EAAPLPFGTS</td>
<td>KPGLYLSCS</td>
<td>KPGPYLQ.TS</td>
<td>KPSFYPG.TS</td>
<td>KPSFYPG.TS</td>
<td>KPSFYPG.TS</td>
<td>KPSFYPG.TS</td>
<td>KPSFYPG.TS</td>
<td>KPSFYPG.TS</td>
</tr>
<tr>
<td>51</td>
<td>SKKAELVALP</td>
<td>SQEAEVALI</td>
<td>PKKAELTALT</td>
<td>PKKAELTALT</td>
<td>SQKAELIALT</td>
<td>SQKAELIALT</td>
<td>SQKAELIALT</td>
<td>. KKAELIALT</td>
<td>SQKAELIALT</td>
<td>AGYAELVALI</td>
</tr>
<tr>
<td>100</td>
<td>QALTLAGQP</td>
<td>RAQTLMNKK</td>
<td>RALHLSKGKQ</td>
<td>RALHLSKGKQ</td>
<td>RALNLSKGKR</td>
<td>RALNLSKGKR</td>
<td>RALNLSKGKR</td>
<td>RALTLSKGKK</td>
<td>RALTLSKGKR</td>
<td>QALTLAGGLT</td>
</tr>
<tr>
<td></td>
<td>............INI</td>
<td>. . . . . . . . . . . . . . . .</td>
<td>. . . . . . . . . . . . . . . .</td>
<td>. . . . . . . . . . . . . . . .</td>
<td>. . . . . . . . . . . . . . . .</td>
<td>. . . . . . . . . . . . . . . .</td>
<td>. . . . . . . . . . . . . . . .</td>
<td>. . . . . . . . . . . . . . . .</td>
<td>. . . . . . . . . . . . . . . .</td>
<td>. . . . . . . . . . . . . . . .</td>
</tr>
<tr>
<td></td>
<td>YNSRYTFYI</td>
<td>YNSKYAYNV</td>
<td>YCYNSKAYHI</td>
<td>YCYNSKAYHI</td>
<td>YCYNSKAYHI</td>
<td>YCYNSKAYHI</td>
<td>YCYNSKAYHI</td>
<td>YCYNSKAYHI</td>
<td>YCYNSKAYHI</td>
<td>YCYNSKAYHI</td>
</tr>
<tr>
<td></td>
<td>VYSHSSIWKK</td>
<td>IHYNAGIQWSE</td>
<td>HPASHAIWQK</td>
<td>HPASHAIWQK</td>
<td>HPASHAIWQK</td>
<td>HPASHAIWQK</td>
<td>HPASHAIWQK</td>
<td>HPASHAIWQK</td>
<td>HPASHAIWQK</td>
<td>HPASHAIWQK</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
</tr>
<tr>
<td>129</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
<td>AINGSLISKL</td>
</tr>
<tr>
<td></td>
<td>LQAARLYQK</td>
<td>LQARLPGK</td>
<td>LQARLPGK</td>
<td>LQARLPGK</td>
<td>LQARLPGK</td>
<td>LQARLPGK</td>
<td>LQARLPGK</td>
<td>LQARLPGK</td>
<td>LQARLPGK</td>
<td>LQARLPGK</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
<td>ITNSRLIYLN</td>
</tr>
<tr>
<td></td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
</tr>
<tr>
<td></td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
<td>ITNSPLIYQL</td>
</tr>
<tr>
<td></td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
<td>VKNGPLIYQL</td>
</tr>
<tr>
<td></td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
<td>LQAARLPGK</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
<td>RGFLIAKNTP</td>
</tr>
<tr>
<td></td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
<td>INASLRKLC</td>
</tr>
<tr>
<td></td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
<td>SRLYFQRK</td>
</tr>
</tbody>
</table>
Table of accession numbers used in the alignment.

<table>
<thead>
<tr>
<th>HERV</th>
<th>Abbreviation</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV1</td>
<td>HERV1</td>
<td>This Thesis</td>
</tr>
<tr>
<td>HERV-XA</td>
<td>HERVXA</td>
<td>AC 000378</td>
</tr>
<tr>
<td>XA 34</td>
<td>XA34</td>
<td>U 29659</td>
</tr>
<tr>
<td>XA 35</td>
<td>XA35</td>
<td>U 37054</td>
</tr>
<tr>
<td>XA 36</td>
<td>XA36</td>
<td>U 37067</td>
</tr>
<tr>
<td>XA 37</td>
<td>XA37</td>
<td>U 29658</td>
</tr>
<tr>
<td>XA 38</td>
<td>XA38</td>
<td>U 37066</td>
</tr>
<tr>
<td>RGH2 (HERV-H)</td>
<td>HERVH</td>
<td>D 11078</td>
</tr>
<tr>
<td>HERV-F</td>
<td>HERVF</td>
<td>Z 94277</td>
</tr>
<tr>
<td>HERV-Fb</td>
<td>HERVFb</td>
<td>Z 83745</td>
</tr>
</tbody>
</table>

**Figure 4.16** Horizontal phenogram of XA family

```
               HERV F
                  
                  HERV Fb
                  
                  XA 37
                  
                  XA 38
                  
                  XA 35
                  
                  XA 34
                  
                  XA 36
                  
                  HERV XA
                  
                  RGH2
```

10 PAM
Chapter 4  
Isolation and characterisation of two novel HERV fragments

HERV1 consistently clustered within the XA family and was most closely related to HERV-XA isolated by Tristem (Tristem 2000). An example of a phenogram is given (Figure 4.16). These alignments also reinforce the current evolutionary groupings as described by Tristem (Tristem 2000). This places HERV1 within a separate XA family distinct from the other HERV-F families and most closely related to the HERV-H family.

4.2.5.2  HERV2 Phylogeny

HERV2 alignments were performed against HERV-I family members as well as the closely related families HERV-ADP, HERV-R and HERV-XA. The reverse transcriptase region was used for the analysis. A table of the accession numbers used is given followed by an example alignment (Figure 4.17) and a representative horizontal phenogram (Figure 4.18).

Table of accession numbers used in the alignment.

<table>
<thead>
<tr>
<th>HERV</th>
<th>Abbreviation</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV2</td>
<td>2</td>
<td>This Thesis</td>
</tr>
<tr>
<td>RTLV Ic</td>
<td>Ic</td>
<td>M 92069</td>
</tr>
<tr>
<td>HERV-I 4682</td>
<td>4682</td>
<td>AC 004682</td>
</tr>
<tr>
<td>RTLV Ia (HERV-I)</td>
<td>Ia</td>
<td>M 92067</td>
</tr>
<tr>
<td>HERV-I 4074</td>
<td>4074</td>
<td>AC 004074</td>
</tr>
<tr>
<td>HERV-I 3100</td>
<td>3100</td>
<td>AC 003100</td>
</tr>
<tr>
<td>HERV-I 4210</td>
<td>4210</td>
<td>AC 004210</td>
</tr>
<tr>
<td>HERV-I 21878</td>
<td>21878</td>
<td>AL 021878</td>
</tr>
<tr>
<td>HERV-I 394</td>
<td>394</td>
<td>AC 000394</td>
</tr>
<tr>
<td>ADP pol (HERV-ADP)</td>
<td>ADP</td>
<td>L 14752</td>
</tr>
<tr>
<td>HERV-XA</td>
<td>XA</td>
<td>AC 000378</td>
</tr>
<tr>
<td>ERV 3 (HERV-R)</td>
<td>R</td>
<td>AC 004609</td>
</tr>
</tbody>
</table>
Figure 4.17 Alignment of HERV-I Family
Figure 4.18  Horizontal phenogram of HERV-I Family

HERV R

HERV XA

HERV ADP

AC004210

AC004682

HERV Ic

AC000394

AL021878

AC003100

AC003100

HERV 2

AC004074

HERV Ia
Phylogenetic support of a neighbour joining tree with bootstrap values applied.

Following phylogenetic analysis HERV2 can be clearly clustered within the HERV-I family with its closest relative being a HERV found by genome database searching (Tristem 2000). The trees formed all consistently placed HERV2 within this family.
4.3 Discussion

This chapter has described the isolation of two HERV fragments both encoding large regions of Pol. Both elements contain numerous frameshifts and stop codons within the coding region and can be assumed to be non functional. These elements were isolated from normal blood donors and illustrate the number of elements still to be discovered within the human genome. Soon the sequencing of the entire genome will be completed leaving a huge database for the determination of many more HERVs some of which may prove to be functional or at least partially expressed as with HERV-W and syncytin. The first analysis of around 7% of the genome revealed fully 150 novel sequences and the discovery of six new families (Tristem 2000).

Both HERVs reported here appear novel by BLAST searching and show only around 80% homology to the closest known retroviral sequences.

HERV1 showed closest homology to HERV-XA at just under 80% and can, for now, be placed within that family. Discovery of further elements within the XA group may mean a second branch and a further family including HERV1. The other elements within the XA group are very closely related (XA34-38 and HERV-XA) so this theory has some backing.

HERV2 showed closest homology to the HERV-I group including the full element RTLV Ia. Again this element can be considered a member of the HERV-I family by consistent phylogenetic clustering.

The copy numbers of both elements is assumed to be low due to the weak signal by Southern blotting and the fact that they have not been isolated previously, despite various probing of the human genome with probes derived from retroviral sequences. The relatively weak homology with all exogenous retroviruses may in part explain this, with the greatest homology seen to the MuLV group at under 50% in the most homologous region. In particular the signal for HERV2 was very faint, suggesting only a couple of elements present in the genome. The signal for HERV1 was detectable and at least 2 bands could be differentiated suggesting at least one other closely related HERV.
Both of these elements were disrupted with the insertion of Alu elements which is not uncommon and in fact was seen in the XA34 isolation as well. That element was also a recombination with ERV9 (Widegren et al. 1996).

The sequence of HERV1 contained a portion of \textit{gag} that included the relatively conserved MHR region (this has been marked on the translation in figure 4.9). In total around 3kb of sequence was isolated and a further attempt was made to locate the 5′ LTR this proved unsuccessful with the adjacent region on the PAC (27K22) containing no further retroviral sequences. The protease region was sequenced and the active site is indicated in figure 4.9. Virtually the entire \textit{pol} region was determined and the reverse transcriptase domains 1-7 were noted. No \textit{env} sequences were found but this region appears to be disrupted with the insertion of Alu elements.

The sequence of HERV2 started with \textit{pol} sequences and only a small portion of the protease was also present. As for HERV1 all seven domains within reverse transcriptase were noted. An Alu repeat again disrupted the \textit{env} region but this was followed closely by a retroviral LTR, which was homologous with those of the HERV-I group so is assumed to be associated with HERV2. Neither the \textit{gag} nor 5′ LTR sequences were determined for this element.

Any expression of retroviral fragments must be taken seriously, as retroviruses are very adept at recombination and co-operation. In particular ERVs in animal models often co-operate to create a disease state often leukaemia. With this in mind and with retroviruses’ close association with leukaemia in mammals (see Chapter 1) and even humans via HTLV-1. Expression of these disrupted elements in various leukaemia samples as well as in normal lymphocytes was undertaken, this will be summarised in the next chapter.
This chapter aimed to assess any potential expression of the two novel HERV as well as the integration sites. The aim was to demonstrate expression potential in even mutated HERV and therefore emphasise a potential biological role for these elements and other known or unknown disrupted elements. The most successful method was via RT-PCR. One element was found to be weakly expressed in normal PBMCs so this expression was further analysed in several leukaemia cases. Any variable expression within leukaemia cases could imply a biological activity of HERV.

5.1 HERV Expression

Retrovirus-like sequences are ubiquitous amongst animals. They have been isolated from reptiles (Jacobson et al. 1980) as well as bivalve molluscs (Underwood et al. 1977) through to humans including most organisms in-between.

In humans most retroviral elements are non-functional containing multiple stop codons, deletions and insertions within the putative coding regions. However many cases of HERV expression have been previously reported in particular in tumour cell lines.

Limits in copy number are evident in all HERV in comparison to other retroelements suggesting some cellular regulation. Gag and Env proteins direct the virus to cellular export and therefore exclusion from the retrotransposition pathway and this could limit the element’s copy number. Blocking of receptors by secreted Env may hinder the re-infection of virus producing cells, and therefore re-integration, limiting copy number. Although copy number is generally limited for complete HERV there are also many LTR only elements derived from LTR-LTR recombination. Their existence implies frequent spread of HERV sequences and suggests that whole elements may be harmful.
5.1.1 Clone 4-1 (HERV-E)

Clone 4-1 env is disrupted with stop codons, however a 1.3 kb ORF was detected in the env region covering almost the entire surface domain. Other transcripts, in particular polyadenylated 3.0 and 1.7 kb transcripts, were seen in placenta, normal colon, spleen and liver (Rabson et al. 1983; Gattoni Celli et al. 1986), and lower levels of 3.6 and 2.2 kb were also observed in placenta (Rabson et al. 1985). Analysis of a placental cDNA library is consistent with the 3.0 and 3.6 kb transcripts as spliced env (Rabson et al. 1985). The 1.7 and 2.2 kb transcripts also appear to result from splicing events analogous with their larger counterparts in elements containing two deletions. One deletion removes 222bp in the surface domain, the other is a 938bp deletion including all of the TM and some of the U3 region of the 3' LTR (Rabson et al. 1985). The 3.0 and 1.7 kb transcripts are also detected in colon and breast carcinoma cell lines (Rabson et al. 1985). Increased expression of the 1.7 kb transcript is seen in primary colon tumours and colon cell lines (Gattoni Celli et al. 1986). Others have reported full-length HERV-E expression as well as the 1.7 kb transcript in human brain (Taruscio et al. 1991).

Further LTR specific transcripts of 3.6kb were detected in normal colon mucosa and colon carcinoma cell lines (Gattoni Celli et al. 1986), these contained transcribed cellular sequences polyadenylated by HERV-E LTR illustrating the use of HERV LTR as promoter sequences (Rabson et al. 1985). Two cDNAs, jnLTR8 and jnLTR22, polyadenylated by HERV-E LTR were isolated from a lung carcinoma cDNA library (Tomita et al. 1990).

Using PCR HERV-E expression was also found in PBMCs (Medstrand et al. 1992) demonstrating activity in normal tissue.

5.1.2 ERV3 (HERV-R)

ERV-3 has now been closely studied and shows variable expression patterns in different human tissues (Larsson et al. 1997; Sibata et al. 1997).

The env region of ERV3 encodes a 1.9kb intact ORF encoding the SU and most of the TM (Cohen et al. 1985). There are polyadenylated 3.5, 7.3 and 9kb mRNAs with similar splicing to retroviral subgenomic transcripts evident (Kato et al. 1991).
1987). In 1990 Kato showed that an ERV3 element is located in the 5' region of, or within, a Kruppel-related gene (Zinc finger), H-plk, but there is a stop codon in the \textit{env} region that may stop the gene's translation (Kato et al. 1990).

Further screening showed that the chorionic villi expressed 3-50x higher levels of HERV-R than other tissues tested and was the only tissue to express the 7.3, 3.5 and 9 kb transcripts widely in some only the 3.5 kb transcript was seen (Cohen et al. 1988). HERV-R \textit{env} transcripts were located to the syncytiotrophoblasts in the placenta (Larsson et al. 1994), proteins were also detected by antibodies derived against synthetic peptides. However no RNA was seen in 6 choriocarcinoma cell lines tested (Cohen et al. 1988; Kato et al. 1988), nor in hydatidiform, a cystic outgrowth from chorionic villi.

5.1.3 RRHERV-1

RRHERV-1 was first isolated from PA-1 teratocarcinoma cell line as a 3.3kb cDNA with homology to 4-1 (LTR) and HERV-R \textit{env}). A 4kb transcript was also observed on a northern blot (Kannan et al. 1991).

5.1.4 S71

Transcripts of S71 were not detected in normal PBMCs or several other cell lines tested, however a cDNA containing 350bp related to the 3' end of S71 was isolated from a human placental library (Leib-Mosch et al. 1992).

5.1.5 HRES-1

HRES-1 expression has been noted as a 6kb transcript containing \textit{gag}-related sequences in the polyadenylated RNA of several cell lines, including MA-T (T-cell), HL60 (promyleocytic leukaemia), W7A and W7B melanoma, but not in K562 (leukaemia) or U937 (monocyte lymphoma) (Perl et al. 1989). Antisera directed against HTLV picked up a 28kD protein that corresponds to an HTLV-related ORF. The protein was intracellular and cytoplasmic, with some seen in nuclear bodies (Banki et al. 1992).
5.1.6  **ERV-9**

ERV-9 expression is seen in the teratocarcinoma cell line Ntera2D1 and is downregulated by retinoic acid induced differentiation (La Mantia et al. 1989). pHE.1 expression as a full length 8kb transcript was detected in undifferentiated Ntera2D1 (La Mantia et al. 1991), other transcripts were noted in placenta. ERV-9 segments were PCR amplified from glioma RNA including an immunosuppressive region of env (Lindeskog et al. 1993). Spliced RNAs of 1.5 and 2kb that remove most of the internal sequence were also noted in Ntera2D1 (Lania et al. 1992).

5.1.7  **HERV-H**

HERV-H is transcribed in many cell lines, with the highest expression from teratocarcinoma and germ cell tumours (Hirose et al. 1993; Wilkinson et al. 1993). Expression is also seen in normal placenta (Johansen et al. 1989) and amniotic or chorionic membranes (Wilkinson et al. 1990). Lower levels of expression are observed by PCR in PBMCs (Medstrand et al. 1992) and by northern blot in some normal tissues. There are complex expression patterns in cell lines due to the high copy number (Wilkinson et al. 1990). The most abundant transcript is 5.6kb, which is the unit length from a deleted element, a spliced 3.7kb transcript is also common (Wilkinson et al. 1990). The region removed in the case of the 3.7kb transcript extends from 140bp after the 5' LTR to a cluster of splice acceptor sites just upstream of the pol domain (Wilkinson et al. 1990) i.e. removing gag and leaving pol. Expression is reduced in Ntera2D1 cells after induction by retinoic acid (Wilkinson et al. 1990). Transcription of intact HERV-H was detected in teratocarcinoma and testicular tumour cell lines (Hirose et al. 1993; Wilkinson et al. 1993), with a putative unit-length 8.6kb transcript detected. Env transcripts were seen in Teral cells (Wilkinson et al. 1993), smaller spliced and 10kb or larger transcripts including cellular fusion products were also detected (Hirose et al. 1993; Wilkinson et al. 1993). Endogenous expression of HERV-H LTRs also differs among different cell lines, LTR types I and II have very low endogenous expression or are expressed only in teratocarcinoma cell lines whereas Ia is readily detectable (Goodchild et al. 1993; Wilkinson et al. 1993). All LTR types can be activated by SV40 large T antigen and increase 5-30 fold but not all individual elements are responsive (Feuchter et al. 1992). This is due to transcription factors and is not a direct effect (Feuchter et al. 1992).
1992). cDNAs containing HERV-H LTRs and cellular RNA transcripts have been found, but their functions are unknown and the expression is limited (Goodchild et al. 1992; Feuchter Murthy et al. 1993). HERV-H expression has also been noted in both normal leukocytes and T-cell leukaemia cell lines (Lindeskog et al. 1997).

5.1.8 HERV-K

HERV-K is probably the most highly expressed of the HERVs studied to date. Retroviral particles have been observed by electron microscopy in syncytiotrophoblasts budding illustrating the existence of functional HERV (Mwenda 1993). Retrovirus-like particles (RVLP) have also been reported in more than 20 testicular tumour cell lines, embryonic carcinomas or teratocarcinomas (Bronson et al. 1984; Lower et al. 1984). HERV-K in particular is highly expressed within teratocarcinoma cell lines and has been associated with RVLP (Boller et al. 1993; Lower et al. 1993). The pol region is intact and found in particle preparations from teratocarcinoma cell supernatant, a 160kD protein with RT activity is observed (Lower et al. 1984). However cellular expression is very low implying the presence of a cellular inhibitor of reverse transcriptase (Lower et al. 1987). Env expression, although low, can be detected by RT-PCR (Tonjes et al. 1996). In teratocarcinoma cell lines env expression is very low level and may be one factor in the lack of particle infectivity.

HERV-K transcripts of various sizes are seen, including a full-length 8.8kb transcript, in several cell lines (Ono et al. 1987). In T47D, a breast carcinoma cell line, expression was induced by estradiol then progesterone (Ono et al. 1987) showing that the LTR has hormone responsive sequences this is similar to MMTV. In 1988 Franklin observed 7.0, 8.0, 8.3, 10 and 12kb transcripts as well as many smaller transcripts in cell lines and placenta, with placenta having the highest levels. HERV-K transcripts have also been noted in colon tumours and normal colon mucosa by northern blot analysis (Moshier et al. 1986) as well as in PBMC and leukocytes by RT-PCR (Medstrand et al. 1992; Brodsky et al. 1993). Some HERV-K expression was seen in teratocarcinoma cell lines (Lower et al. 1993). A full length 8.6 kb transcript and smaller spliced 3.3, 1.8 and 1.5 kb transcripts have been observed by northern blots. Retroviral-like particles are also produced by these lines in culture (Boller et al. 1993). The 3.3kb transcript includes env and is smaller due to the
removal of most of the interior retroviral regions (Lower et al. 1993). The 1.8kb transcript is a doubly spliced cORF (central). This shorter transcript encodes a 12kD putative protein that has basic regions similar to the RNA binding domains of rex, rev and tat of HTLV-I and HIV (Lower et al. 1993).

A member of the HERV-K group, HERV-K10, is relatively complete. Some coding competent, intact HERV-K10 gag and env mRNAs have been isolated using RT-PCR from GH cells (Lower et al. 1993). A gag region has also been cloned by PCR from genomic DNA and expressed in E. coli as a fusion protein (Mueller-Lantzsch et al. 1993). The HERV-K portion of the fusion protein was 73kD as predicted. When all of the gag region, including the protease homologue at the 3’end, was expressed the protein showed an auto-proteolytic generation of a smaller protein seen by anti-gag antisera (Mueller-Lantzsch et al. 1993). The smaller protein was not seen when the protease region was deleted.

HERV-K antibodies are very low in normal sera but are measurable in leukaemias, pregnancy and testicular tumours (Tonjes et al. 1996). HERV-K envelope protein is recognised by 12% of normal people as illustrated by the existence of antibody that recognises it (Vogetseder et al. 1993). This is notable as HERV antigens could be considered self, therefore the raising of antibodies is not expected and may explain some of the association with autoimmunity.

HERV-K probably encodes retrovirus-like particles as mentioned above. An example is HTDV (human teratocarcinoma derived virus), which exists in high numbers in GH. These particles are specifically stained by HERV-K gag antiserum (Boller et al. 1993). Western blotting of GH cell membrane shows that HERV-K gag antiserum reacts primarily with a 30kD protein, equivalent to the size of the putative core p30 of HERV-K (Boller et al. 1993; Mueller-Lantzsch et al. 1993). Expression of the HERV-K gag gene is sufficient for particle formation when transfected into mammalian cells, arrested in budding stage, as shown by Korbmacher (Korbmacher et al. 1993). However proper cleavage of gag may not occur as collapsed or condensed cores are not observed as is seen with mature exogenous retroviral virions (Mueller-Lantzsch et al. 1993).

Solitary HERV-K LTRs are present in the human genome and these are expressed as parts of cellular transcripts in a variety of tissues (Leib Mosch et al. 1993). HERV-K LTRs show different activity in different tissues, they appear to be
primed in embryonic cell lines but not in other cell lines e.g. hepatocarcinoma due to mutations within the LTR making them only weakly inducible. This implies a cellular selection for tissue type expression and implies a cellular function of the LTRs in these tissues.

The HML group of viruses is closely related to HERV-K. Southern blot studies showed that HML-1, 2, 3 and 6 were the most abundant with 60-80 copies per human genome, transcripts were observed in several tissues (Medstand et al. 1993).

5.1.9 HERV-L

HERV-L was first isolated from placental RNA and the expressed region was a large portion of pol (Cordonnier et al. 1995).

5.1.10 Potential HERV Expression

Retrovirus-like particles have been detected in human sources not known to be infected by exogenous virus, e.g. placenta (Dirksen et al. 1977; Johnson et al. 1990), milk (Moore et al. 1971; Vaidya 1973), breast cancer biopsies (Axel et al. 1972), monocytes from myeloproliferative disease (Boyd et al. 1989), multiple sclerosis patients (Haahr et al. 1992), breast cancer cell lines (McGrath et al. 1974; Keydar et al.) and teratocarcinoma cell lines (Boller et al. 1983; Bronson et al. 1984; Lower et al. 1987). Many of these cases may now be explained by HERV infections as detailed above, in particular functional HERV-K. Retrovirus-related antigens have also been detected from similar sources and infectious retroviruses were not shown with these particles, they appeared to be encoded by HERV. Some were encoded from a single HERV, but others were encoded by complementation and by one or more families. Rare functional ERV elements within a family of defective elements have been discovered in the genomes of other species (Reeves et al. 1984) and also among the retroelements in humans (Matera et al. 1990; Dombroski et al. 1991; Deininger et al. 1992).

Recent searching of the genome database has led to the discovery of 150 novel HERV elements and any expression of these is yet to be studied (Tristem 2000). There may be other functional HERV ORFs yet to be elucidated and as more of the human genome is sequenced daily these will soon come to light.
5.2 Biological Significance

HERV expression has been detailed in the previous section (section 5.1). The impact of this expression is a source of much debate, however several instances of true biological activity of retrovirus-like sequences have been reported. In humans intracisternal A particles have been detected illustrating true coding potential (Ono et al. 1986).

HERV LTRs are found bound to cellular genes and have been shown to control cellular gene regulation e.g. K in HLA-DQ (Kambhu et al. 1990). An insertion of HERV 4-1 in the amylase E promoter has resulted in tissue specific expression (Samuelson et al. 1990).

One HERV-W encoded protein (syncytin) appears to have been co-opted by the cell for normal cellular function promoting formation of the syncytiotrophoblast layer in placenta (Mi et al. 2000).

5.2.1 Protection

At first glance HERV sequences may seem irrelevant in terms of the biological activity of the cell due to their rapid deletion and mutation rate. However many cases of intact ORFs are reported and in particular the LTRs are often still intact. It is the LTR that encodes the pathogenic potential of many HERVs as can be compared with MMTV and JSRV. For example integration also has the potential to cause adjacent expression enhancement via the LTR (Kambhu et al. 1990).

Endogenous retroviruses may offer cellular protection against infection by a related pathogenic virus. The case of BaEV may be an example of this as it appears to have become completely endogenous. The mechanism of protection may be receptor interference and super antigen depletion of susceptible host cells. HERV gene products may also modulate immune functions and protect against superinfection by related exogenous viruses. This is seen with endogenous retroviruses in mice (Choi et al. 1991; Gardner et al. 1991). Further evidence to support this theory is the fact that most HERVs appear to exist in a stable balance with the host as measured by a stable copy number.
5.2.2 Genome Plasticity

HERVs may function within the genome to provide some plasticity which can be achieved by various mechanisms. HERVs provide variation within the genome through reverse transcription. Up to 10% of the genome is derived from reverse transcribed sequences and HERV is the main source of reverse transcriptase and therefore this variation. HERV also provide a source of allelic variation and may be used as markers of evolution (Arvidsson et al. 1995; Svensson et al. 1995). HERV accumulate in hypervariable regions and in hot spots of recombination, it is possible that HERV sequences may induce chromosomal recombination by hybridisation with each other.

5.2.3 Disease Association

5.2.3.1 Insertion

Retrovirus-like elements can cause disease by insertion into the genome, several cases of this have been documented, e.g. (Munroe et al. 1990). Somatic cells with insertion can lead to cell death, but if the insertion disrupts a tumour suppressor gene or causes over expression of an oncogene, when inserted into their promoter, carcinogenesis may result. In animal models there have been many examples of ERV causing this effect, e.g. (Munroe et al. 1990).

In humans transpositions of LINEs have been detected as de novo insertions into Factor VIII genes leading to haemophilia type A in two cases (Kazazian et al. 1988). Colon cancer cases caused by disruption of the tumour suppressor adenomatous polyposis coli gene have also been noted (Miki et al. 1992). Insertion into the myc gene was seen in a breast cancer case (Morse et al. 1988). A disruption of exon 48 of the dystrophin gene by an L1 element was observed in a case of muscular dystrophy (Holmes et al. 1994).

A chromosomal aberration via recombination of closely related sequences is another possible mechanism of disease induction by HERV. A possible example of this is an S71 element on 18q21 involving the bcl-2 translocation t(14:18) in B cell lymphomas and follicular tumours. This locus is also involved in some breast and colon cancers. ERV1 is also found at the same locus.
5.2.3.2 Gain of Function

Although a HERV may be disrupted the LTRs often retain activity, if the element or even the active LTRs alone then insert elsewhere in the genome they can act as a gene switch. Readthrough expression of cellular DNA sequences can also occur causing the LTR induction of adjacent cellular genes. Cellular sequences promoted by HERV-H LTRs have been isolated (Feuchter et al. 1992). An example is a spliced fusion between HERV-H and an exon of calbindin gene. This fusion is expressed at the protein level (Liu et al. 1991). It leads to deregulation of calbindin expression and may have a role in malignancy as these were isolated from the cDNA of 2 prostate metastasis cell lines.

5.2.3.3 Immune System Effect

Retroviruses have been shown to encode super antigens within their Env proteins. An example of this is SAG on endogenous MMTV which can lead to an increase in T cells and increased tumour causing potential (Golovkina et al. 1994; Xu et al. 1996). A similar putative superantigen is encoded in a conserved region of HERV, namely the TM region of Env. A peptide (p15E) encoded in this region has been shown to be immunosuppressive (Cianciolo et al. 1985). Some replication defective elements can still encode this region and it is actively expressed in some leukaemia cell lines, possibly aiding escape from elimination by the immune system (Cianciolo et al. 1984).

A HERV super antigen encoded by env of HERV-K10 has been linked with insulin dependent diabetes mellitus (IDDM) (Conrad et al. 1997). This antigen may be a marker of IDDM rather than the initial trigger of the disease causing an amplification of immune response (Benoist et al. 1997). Expression of HERV Gag protein is also reported as determined by antibody in sera of leukaemia and autoimmune disease patients compared to normals (Krieg et al. 1992).

5.2.3.4 Source of Isolation

By considering the source of the HERV one can make an association with disease. Many HERV were isolated by analysis of tumours or tumour cell lines either directly or from cDNAs. An example of this is the breast carcinoma cell line T47D, which has yielded three HERV sequences. All three were isolated from purified T47-
D particles demonstrating their biological activity (Seifarth et al. 1995). These particles are type B in appearance but two of the isolated sequences, ERV-FRD and ERV-FTD, are derived from \textit{pol} sequences from defective Class I endogenous viruses, the third was ERV-MLN, a Class II HERV with similarity to HERV-K10.

The HERV HRES-1 with some similarity to HTLV-1 has been isolated as a transcriptionally active element from numerous sources including MA-T cells, melanoma, placenta and EBV transformed peripheral blood B-lymphocytes (Perl et al. 1989).

Antibodies, specific to HTLV p19, that may be cross-reacting with HERV, are present in choriocarcinoma as well as early placental tissue (Suni et al. 1984). Endogenous retroviral expression detected by antibody detection of p30 was noted in placenta as early as 1984 (Jerabek et al. 1984).

A retrovirus has been isolated from multiple sclerosis patients. It was initially called LM7 and later referred to as MS-associated retrovirus (MSRV). RT-PCR was used to isolate \textit{pol} sequences (Perron et al. 1997) and a complete element of this was isolated via its expression in placental tissue with intact regions in \textit{gag}, \textit{pol} and \textit{env} (Blond et al. 1999) and it has now been placed in the HERV-W family.
5.3 Human Leukaemia

As has been previously discussed human leukaemias can by roughly divided into lymphoid or myeloid lineage and chronic or acute type. However leukaemia cases can be further sub-divided for more accurate treatment control, patient response and survival rate. A second reason for sub-typing is to provide a standard classification for ease of communication and co-operation between centres of treatment and research. The basis for classification is primarily lineage and degree of maturation of abnormal proliferating cells. Further diagnosis can be made using cytochemical, immunological, cytogenetic and molecular techniques. Below is a table illustrating the sub-types as designated by the French-American-British (FAB) Co-operation.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Abbreviation</th>
<th>FAB Code</th>
<th>Leukaemia Sub-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myeloid lineage</td>
<td>AML</td>
<td>M0</td>
<td>Myeloblastic w/o cytologic maturation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>Myeloblastic with minimal cytologic maturation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>Myeloblastic with significant cytologic maturation</td>
</tr>
<tr>
<td></td>
<td>APL</td>
<td>M3</td>
<td>Acute promyelocytic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3variant</td>
<td>Unusual hypogranular form</td>
</tr>
<tr>
<td>Myeloid and Monocytic</td>
<td>AMML</td>
<td>M4</td>
<td>Acute myelomonocytic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4eo</td>
<td>M4 with eosinophilic maturation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4baso</td>
<td>M4 with basophilic maturation</td>
</tr>
<tr>
<td>Monocytic</td>
<td>AMoL</td>
<td>M5a</td>
<td>Acute monoblastic with poor differentiation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M5b</td>
<td>M5 with more differentiation</td>
</tr>
<tr>
<td>Erythroid and Myeloid</td>
<td>AEL</td>
<td>M6</td>
<td>Acute erythroid</td>
</tr>
<tr>
<td>Megakaryoblastic</td>
<td>AML</td>
<td>M7</td>
<td>Acute megakaryoblastic</td>
</tr>
</tbody>
</table>
## Expression of Two Novel HERVs

<table>
<thead>
<tr>
<th>Mixed lineage</th>
<th>AMLL</th>
<th>Acute mixed lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progenitor cells</td>
<td>AUL</td>
<td>Acute undifferentiated</td>
</tr>
<tr>
<td>Lymphoblastic</td>
<td>ALL</td>
<td>L1 Childhood Acute Lymphoblastic with homogenous blast cells</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L2 Adult Acute Lymphoblastic with more heterogeneity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L3 Burkitt Like Acute Lymphoblastic</td>
</tr>
<tr>
<td>Mast Cell</td>
<td></td>
<td>Acute Mast cell leukaemia</td>
</tr>
<tr>
<td>Granulocytic</td>
<td>AML</td>
<td>Granulocytic Sarcoma</td>
</tr>
<tr>
<td>Mixed lineage</td>
<td>AML/ALL</td>
<td>Acute mixed-lineage</td>
</tr>
<tr>
<td>Myeloid</td>
<td>t-AML</td>
<td>Therapy related acute myeloid</td>
</tr>
<tr>
<td>Lymphocytic</td>
<td>CLL</td>
<td>B-CLL Chronic Lymphocytic leukaemia of B-lymphocytes</td>
</tr>
<tr>
<td></td>
<td>T-PLL</td>
<td>Chronic Lymphocytic leukaemia of T-lymphocytes</td>
</tr>
<tr>
<td></td>
<td>PLL</td>
<td>B-PLL Prolymphocytic of B-cell lineage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T-PLL Prolymphocytic of T-cell lineage</td>
</tr>
<tr>
<td></td>
<td>HCL</td>
<td>Hairy cell leukaemia</td>
</tr>
<tr>
<td>Plasma Cell</td>
<td>PCL</td>
<td>Plasma cell leukaemia</td>
</tr>
<tr>
<td>Lymphocytic</td>
<td>ATL</td>
<td>Adult T-cell leukaemia</td>
</tr>
<tr>
<td>Myelocytic</td>
<td>CML</td>
<td>Chronic Myelocytic leukaemia</td>
</tr>
<tr>
<td></td>
<td>CMML</td>
<td>Chronic Myelomonocytic</td>
</tr>
</tbody>
</table>
5.3.1 Acute Myelocytic Leukaemia

Acute myeloid leukaemia and acute monocytic leukaemia (AML M5) especially are of particular relevance to this thesis due to the observed differential HERV expression within an AML M5 patient.

AML has previously been associated with retroviral activity. In a study of AML cells from mice over-expression of the IAP early transposon (Etn) RNA was observed as compared to other tumour cell types (Tanaka et al. 2001). In the mouse common retroviral integration points have been mapped, homologues of these have also been found in humans and their locations mapped. Intriguingly these loci in humans have been associated with leukaemia including AML (Van Cong et al. 1989). In humans retroviral neutralising antibodies have been detected in AML and AML M5 cases (Kiss et al. 1989). In a Chinese study AML cells expressing retroviral antigens and displaying RT activity and particle production were observed, this was not detected in normals (Xu et al. 1998). HERV K expression has been studied in leukaemia and a restricted expression pattern was observed with a far larger variety of HERV K expressed in normal cells (Brodsky et al. 1993). This data emphasises a potential biological role for HERV in AML and supports the analysis undertaken.

In 1955 AML was seen as always fatal but in 1965 it was questioned whether treatment should be provided and now complete remission is achieved in most younger adults. AML is less common in children, it is usually a disease of the old but with a small incidence in younger men. AML accounts for 40% of all leukaemias in the Western world. It is radiotherapy associated and there is an increase in cases amongst survivors of the Japanese atomic bombs. Benzene and other alkylating agents lead to secondary AML. AML is typed by FAB using morphology or immunotyping. It is clonal in nature with the abnormalities repeated in all leukemic cells. AML accounts for 15-20% of childhood leukaemia and treatment is similar to that for adults. Radiation, alkylating agents and topoisomerase -2 inhibitors all lead to an increased AML risk. There is a high risk among monozygotic twins and some genetic syndromes and heritable disorders, including Bloom’s syndrome, Fanconi’s anemia and congenital bone marrow failure, Down’s syndrome carry a 10-20x
increased risk. Patients who are less than two years old have a greater proportion of M4 or M5, granulocytic sarcomas, higher CNS involvement and a high leukocyte count.

**AML M0**
These cases involve only immature cells and the myeloid lineage can be detected by expression of myeloid antigens. These cases show a poor response to combination remission induction therapy.

**AML M1**
In these cases more than 90% of non-erythroid cells are poorly differentiated myeloblasts, some cells have a few fine azurophilic granules.

**AML M2**
For this classification bone marrow must have a certain percentage of blasts of between 30-90% of the total bone marrow, with monocytic precursors making up less than 20%.

**AML M3**
Acute promyelocytic leukaemia is so-called due to the proliferation of promyelocytes defined by 10% in adults and 4-8% in children of total bone marrow. Many proliferating cells have heavy azurophilic granulations and nuclei shape is varied. There is a variant of M3 with minimal granulation that accounts for less than a third of cases.

**AML M4**
M4 accounts for around 20-30% of adult AML cases. The granulocytes and monocytes are present in varying proportions with the bone marrow monocyte proportion varying between 20 and 80%. M4 with eosinophilia has an eosinophil content in the bone marrow of between 1 and 30%, these contain eosinophilic granules and basophilic large granules. This variant has a good chance of complete remission.

**AML M5**

AML is the result of a transformation of a multipotent haematopoietic progenitor cell. M5 shows abnormal differentiation restricted to monocytic lineage. There are two types; one involving poorly differentiated monocytes named M5a that is more common in children and the M5b subtype involving well differentiated monocytes. In
type A they are poorly differentiated with less than 20% of monoblasts in bone marrow being promonocytes or monocytes and this variant is more common in children. In type B more than 20% of the abnormal cells are promonocytes or monocytes in the bone marrow.

Various specific chromosomal abnormalities have been associated AML M5, the most common is the involvement of the MLL region at 11q23 (Berger et al. 1980). This region spans 100kb and contains 21 exons and encodes several zinc finger domains (Rowley et al. 1990). One common mutation seems to occur via the fusion of Alu elements (Strout et al. 1998). 11q23 involvement is seen in other types of AML but is most common in M4/M5 (Satake et al. 1999).

Specific translocation partners for 11q23 have been noted, one of these is t(9;11)(p22;q23) occurring in 5-6% of cases of M5. This translocation results in an HRX-AF9 fusion forming a nuclear protein (Djabali et al. 1992; Nakamura et al. 1993; Odero et al. 2000). A second common translocation is t(10;11)(p12;q23) forming an AF10 fusion, this is more common in AML M5 than other AML (Gore et al. 2000).

The 11q14 region encoding a clathrin assembly lymphoid-myeloid leukaemia gene has also been associated with AML. In particular a translocation also involving 10 was observed. This t(10;11) has been seen in the U937 leukaemia cell line (Dreyling et al. 1996) as well as in cases of AML (Bohlander et al. 2000; Carlson et al. 2000).

Another noted translocation is t(10;16)(q22;p13) found in some AML 5a. This translocation involves the 16p13.3 region encoding the Creb binding protein (Borrow et al. 1996) and is seen in AML M4/5. An 11q23 (MLL) fusion to this CBP has also been noted (Panagopoulos et al. 2000))

A rarer yet specific translocation is t(8;16)(p11;p13) which is seen in M4/M5 rather than other AML subtypes (Borrow et al. 1996). This fusion protein is thought to encode a chromatin acetyltransferase and may be involved in gene activation. An RT-PCR readout has been developed for the CBP-MOZ fusion transcript encoded by this translocation (Panagopoulos et al. 2000). This occurs in less than 1% of M5 cases and is characteristic of a particular subtype with extensive erythrophagocytosis and thrombocytosis (Heim et al. 1987).
Other non-random aberrations have been noted these are mostly numerical changes and deletions including +8, +21, -5 or 5q-, -7 or 7q-, 9q-, 20q- all are more common in secondary AML and following MDS (Rowley 1990). Loss of chromosomes 5 or 7 or large deletions within them seem important as they are seen in 16% of de novo AML and 80% of therapy associated (Dabaja et al. 1999).

AML M6
Acute erythroleukaemia accounts for 4-5% of all AML and can occur de novo or as a progression from a myelodysplastic syndrome. Greater than 50% of erythroid precursors show dysmorphic features and there are more than 30% myeloblasts present in the bone marrow. This type of AML is a clonal disorder from a multipotent stem cell with broad myeloid potential.

AML M7
Acute megakaryoblastic leukaemia accounts for only 1% of AML and is difficult to diagnose via the bone marrow. In the blood there is different red cell morphology and small myeloblasts. Megakaryoblasts are pleiomorphic ranging from very small to quite large with up to three nuclei.

5.3.2 ALL

Acute lymphoblastic leukaemia arises from a clonal proliferation of malignantly transformed lymphoid progenitors in the bone marrow. ALL is most common in paediatric populations accounting for 80% of acute leukaemia cases compared with 20% of acute cases of adults. Infiltration of other organs is often noted especially liver, spleen, lymph and CNS.

5.3.2.1 Childhood ALL
Therapy for the majority of cases of childhood ALL has now been successfully achieved. This type of leukaemia is the most common form in children, accounting for 25% of new paediatric cancer cases or a population risk of 4/100,000/year, the peak of cases is at three to four years. There are marked geographical frequency differences and age distributions - B-cell variant is more common in industrialised countries, although apparent regional clusters are not statistically significant. Specific genetic changes in lymphoid progenitor cells lead to dysregulation of cellular
proliferation, clonal expansion and ALL. There are three types by FAB classification
L1 in 85%, L2 in 14% and L3 in 1-2% of cases. By using monoclonal antibodies the
B or T-cell lineage can be determined, 80-85% are from B-cell precursors and 15-20%
from T-cell precursors. By using CD markers and monoclonal antibodies B type can
be subdivided into lymphoid stem cell, early pre-B, pre-B and mature B. The T
lineage is subdivided using TCRs and monoclonals into stage I early, stage II
intermediate (midthymic) and stage III late. Most T-cell leukaemias are early,
lymphomas tend to be mid or late.

5.3.2.2 ALL in Adults

There is a low incidence of ALL in early to mid life, increasing after the age of
50 to 15% of all acute leukaemias. Treatment is usually by cyclical combination
therapy to induce complete remission, then followed by post remission intensive
chemotherapy that may involve a bone marrow transplant. Maintenance
chemotherapy using antimetabolites and CNS treatment (intrathecal chemotherapy
and cranial irradiation) is also prescribed. In mature B-ALL short intensive
chemotherapy without maintenance is effective. Philadelphia chromosome occurs in
20-30% of ALL cases. Only one third of older (>60 years) are cured but the complete
remission rate increases to 80% with younger cases. T-ALL is more common than in
children (25%), the pre-T lineage is more frequent in adults. T-ALL tends to affect
males and is associated with high wbc count, CNS involvement and medistinal
tumour and has a poor prognosis. This has improved recently with the use of CPA
and ara-C. B-ALL accounts for 75% of adult ALL cases, with early pre-B being more
frequent than in childhood ALL. Common ALL is the most frequent type with an
overall survival rate of 25-35%, the presence of Philadelphia chromosome being a
poor prognostic marker. Pre-B has similar survival figures, early pre-B has poor
prognosis with 60-70% having the t(4;11) translocation. Mature-B ALL occurs more
commonly in males and is associated with lymphadenopathy, abdominal tumour,
renal, bone and CNS involvement.

ALL L1

Childhood ALL is typified by distinctive homogeneous blast cells of small size and
high nucleus to cytoplasmic (N:C) ratio.
ALL L2
Adult ALL shows greater cellular heterogeneity, over 20% of cells are large with nuclear clefting or indentation. The N:C ration is lower than in L1 with nucleoli presence and more cytoplasm.

ALL L3
Burkitt-like ALL is rarer accounting for only 2-5% of all ALL cases. The cells involved are identical to those of Burkitt's lymphoma, being large and heterogeneous.

5.3.3 Rare Forms of Acute Leukaemia

5.3.3.1 Acute Mast Cell Leukaemia
This leukaemia is of monocyte-macrophage lineage and is not genetically related to basophilic granulocytes. 15% of patients with malignant systemic mastocytosis develop pure mast cell leukaemia. The white blood cell count contains more than 30% mast cells.

5.3.3.2 Granulocytic Sarcoma
Granulocytic sarcoma occurs in some myelogenous tumours and 3% of AML cases. The sarcomas appear green in dilute acid and are therefore called chloromas, they are also known as myeloblastoma. These are most common in children and can arise before signs in blood or bone marrow, or at relapse.
5.3.3.3 Acute Undifferentiated Leukaemia (AUL)

The systemic use of lymphoid and myeloid monoclonal antibodies and flow cytometry has led to a reduced number of unclassified leukaemias. However the definition of undifferentiated leukaemia is a complete lack of morphologic differentiation by light microscopy. There is some lineage heterogeneity in acute leukaemia and some can be biphenotypic i.e. both myeloid and B lineage.

5.3.3.4 Acute Mixed-Lineage Leukaemia (AML/L)

Stass reported a large study of acute leukaemia (Stass et al. 1986). 10% of ALL cases had myeloid antigens and 25% of AML cases had lymphoid antigens. By using EM myeloid cells are seen in ALL or unclassified leukaemia leading to a definition of acute mixed lineage. A lineage switch is possible and is more frequent in children, especially from lymphoid to myeloid at relapse.

5.3.3.5 Acute Mixed Leukaemia

This is defined by flow cytometry with two or more phenotypically different clones proliferating together. It usually occurs in blast crisis of CML or chemotherapy related cases.

5.3.3.6 AML with Trilineage Dysplasia

Monodysplastic syndrome leads to overt AML in 20-30% of cases. Cases of AML with trilineage monodysplastic features (TMDS) have a poorer remission than AML without.

5.3.3.7 Therapy-Related Acute Myeloid Leukaemia (t-AML)

AML can result following alkylating agent or radiation treatment of a neoplasm. Usually more than 30% of bone marrow blasts are abnormal. In most cases the same chromosome abnormalities are seen as in MDS.
5.3.4 Chronic Lymphoid Proliferations

5.3.4.1 Chronic Lymphocytic Leukaemia

CLL is a chronic proliferation of malignant B-lymphocytes, they appear small and well differentiated by light microscopy expressing B markers on the cell surface. The numbers increase because of prolonged survival, altered apoptosis and increased proliferation leading to 5-10 x10^9/l for diagnosis. The B-lymphocytes have a rim of cytoplasm and are small with a high N:C ratio. Infiltrate of the bone marrow occurs in varying degrees. In the early stages the bone marrow and blood are affected, later the lymph nodes, liver and spleen, which leads to lymphadenopathy, splenomegaly and hepatomegaly and later bone marrow failure usually occurs. CLL from clonal T-lymphocytes occurs but is rare, CLL is also rare in people less than 30 and generally has a male to female ratio of 3:7 in all ages. Incidence rates vary from county to county implying viral or genetic factors. Further diagnosis can be made by looking at the CD markers using antibodies, as a specific profile is present for CLL. The stage of CLL can be divided by its seriousness, some patients survive for less than one year and others for over 20 years post diagnosis. The staging goes from 0 at the least serious to IV at the most:

0 Lymphocytosis and bone marrow involvement.
I Node enlargement.
II Splenomegaly or hepatomegaly.
III Anemia.
IV Thrombocytopenia.

Bone marrow biopsy pattern is also studied, if it is diffuse then that is a poor sign. Progression of CLL to acute blastic leukaemia is rare at less than 1%. If it does happen then it is generally to L2 ALL. More common in CLL patients is the development of large cell lymphoma.

5.3.4.2 Prolymphocytic Leukaemia

PLL cases usually have a high wbc count and splenomegaly without lymphadenopathy. High-density surface immunoglobulins surround the prolymphocytes. More than 55% of all wbc's and more than 70% of all lymphocytes
are prolymphocytes in PLL cases. These cells are large with round nuclei, low N:C ratios and basophilic cytoplasm.

5.3.4.3 Hairy Cell Leukaemia

HCL or leukaemic reticuloendotheliosis is characterised by pancytopenia and splenomegaly with hairy cells in peripheral blood. Hairy cells are so called due to numerous cytoplasmic projections of irregular villi. They are large cells with a low N:C ratio. The bone marrow is generally involved and there is a remission rate of 90% with the use of therapy. It is a rare form of leukaemia accounting for 2% of all leukaemias, with a high male to female ratio of 5:1 and HCL generally affects 30-75 year olds. The particular type of B-cell (hairy cell) which is not normally seen can be induced in CLL patients by Bryostatin 1. Cytokines are important for HCL cells, in vivo HCL cells release TNF, which may inhibit bone marrow precursor growth. In vitro TNF addition leads to increased survival of HCL cells. IL-2R (CD 25) is also expressed by HCL cells and is the low affinity receptor, the medium affinity (p75) IL-2 receptor is also expressed at lower levels. Diagnosis is by the morphology of the hairy cells, they are B-lymphocytes, and have open dispersed chromatin, pale cytoplasm and agranular, multiple cytoplasmic hair-like projections. The cells are at a late stage of maturity most similar to a plasma cell. Variant HC is an intermediate between prolymphocytes and typical HCs.

5.3.4.4 Splenic Lymphoma with Villious Lymphocytes

SLVL has a wbc count of 3-40x10^9/l. This type can be differentiated from HCL by a small monoclonal immunoglobulin band in serum and urine, as well as different circulating cells from HCs. The villi are often shorter and commonly only at one pole with the N:C ratio being higher and the immunology being similar to B-PLL.

5.3.4.5 Leukaemic Phase of non-Hodgkin’s Lymphoma.

Non-Hodgkin’s lymphoma may present as or evolve to a leukaemic phase (lymphosarcoma cell leukaemia) similar to CLL. The most common type to do this is follicular lymphoma with circulating peripheral blood cells. Circulating lymphoma cells are cleaved and smaller than in CLL cases with high N:C ration and a thin rim of cytoplasm.
5.3.4.6 Intermediate Non-Hodgkin’s Lymphoma

These cases are associated with lymphoma cells in peripheral blood. The cells are pleomorphic with slight nuclear irregularities, indentations or clefts and are similar to mixed CLL.

5.3.4.7 Lymphoplasmocytic Lymphoma

These cases have circulating lymphoid cells and may have a serum monoclonal immunoglobulin band.

5.3.4.8 Plasma Cell Leukaemia

2% of patients with multiple myeloma may, at terminal phase, have plasma cells in peripheral blood. This is distinguishable from primary plasma cell leukaemia which is usually an acute disease and resembles acute leukaemia. The FAB has designated two groups of plasma cell leukaemia, one with small, heterogeneous cells ranging from lymphocytes with basophilic cytoplasm to plasma cells and a second group with blast like cells. Immunophenotyping and serum investigation can differentiate between the sub-types.

5.3.4.9 Mature T-Cell Proliferations

These proliferations are easily identified by immunophenotyping and morphology.

5.3.4.9.1 T-Chronic Lymphocytic Leukaemia (Large Granular Cell Lymphocytosis)

In T-CLL the lymphocyte count is greater than $5 \times 10^9/l$ for more than 6 months. The cells are generally large granular lymphocytes and account for 50-95% of peripheral blood lymphocytes. Rearrangements of TCR genes leading to clonal proliferation are common with bone marrow variably involved. Cytopenias are also common probably caused by bone marrow suppression and factors from malignant cells.
5.3.4.9.2 T-Prolymphocytic Leukaemia

20% of PLL are of the T-lineage and they are very hypercellular with lymphocytosis of greater than 100x10^9/l. PLL is generally aggressive with lymphadenopathy and splenomegaly being common. Normally T-PLL cases have similar morphology to B-PLL so immunology is used to distinguish them. The bone marrow is diffusely infiltrated and there is often fibrosis.

5.3.4.9.3 Sézary's Syndrome

This syndrome is a generalised ex-foliate erythroderma with atypical T lymphocytes in the blood. There are two variants Sézary cells and Lutzner-like cells. Sézary cells are large with near tetraploid chromosome number and are less common. Lutzner cells are smaller and resemble cells of mycosis fungoides, all are CD3 positive and CD8 negative.

5.3.5 Chronic Myelocytic Leukaemia

CML is a clonal disorder of multipotent hematopoietic stem cells with malignant proliferation in bone marrow of granulocytic and megakaryocytic cells. There are more mitotic divisions than normal and cellular life span is longer. The Philadelphia chromosome translocation is common involving the ABL gene from 9(q34) and chromosome 22 adjacent to BCR gene in band q11. This translocation leads to a hybrid gene BCR/ABL1 encoding a 210kD fusion protein with tyrosine kinase activity, this protein plays a role in controlling cell growth in CML. T lymphocytes appear spared and B cells are involved in 25% of patients.

There are three phases of disease chronic or indolent (1 to 3.5 years) that is controllable, followed by accelerated then acute or blastic phase within 6 months that is terminal. 20% of cases develop blastic phase without an accelerated phase and another 20% die from complications before blastic phase is reached.

5.3.5.1 Chronic Indolent Phase

The first phase of CML is characterised by granulocytic leukocytosis, elevated platelets in blood and hyperplasia of the bone marrow. Few if any erythrocytic abnormalities are present. In order to differentiate between the proliferative phase of
chronic myelomonocytic leukaemia (CMML) and atypical CML the following criteria are used. In the blood CMML monocytes are greater than 10%, less than 10% immature granulocytes and less than 2% basophils. In CML cases there are 2-10% monocytes, 10-20% immature granulocytes and less than 2% basophils.

5.3.5.2 Accelerated Phase

This phase has less than 30% blasts and the bone marrow may show increased reticulin formation with dysplastic features of megakaryocytic and/or erythrocytes. In the peripheral blood platelets and haemoglobin numbers are decreasing and wbc count is rising, fever and splenic pain are also common.

5.3.5.3 Acute Phase

80% of patients with CML progress to acute or blastic phase with a bone marrow and blood picture similar to that of acute leukaemia including greater than 30% blasts. In 75% of patients the morphology is similar to that of myeloid variants of AML (M0 to M7), the remaining cases resemble ALL.

5.3.6 Myelodysplastic Syndromes (MDS)

MDS covers many alternative terms including pre-leukaemia, smouldering acute leukaemia, refractory anemia, refractory anemia with excessive blasts, smouldering myeloid states and dyshaematopoiesis. Only 20% of MDS terminate in overt AML. Primary acquired MDS and secondary MDS often occur after exposure to radiation, chemotherapeutic drugs or toxins. MDS is most common in the elderly with over 90% of all cases affecting those over 60 years old and in children, when progression to myeloid leukaemia is usually rapid. MDS has been standardised through FAB co-operative into five sub-groups. Diagnosis is based on the presence of dyspoiesis of cells, myelocytic, erythrocytic and megakaryocytic cells in the bone marrow and cytopenia in blood.

5.3.6.1 Refractory anemia

RA is an MDS sub-type characterised by less than 11 RBC g/l and is always refractory with a reticulocyte count of less than 10x10^9/l. Blast cells are generally
absent or less than 1% in a blood smear and less than 5% in the bone marrow. RA accounts for 17% of MDS and there is leukaemic transformation in 10% of patients.

5.3.6.2 Refractory anemia with ringed Sideroblasts

This is similar to RA but with greater than 15% ringed sideroblasts in the bone marrow, bone marrow blasts make up less than 5%. It rarely evolves to chronic myelomonocytic leukaemia in pure RAS. RAS-MDS with iron overload and dysplastic features has a worse prognosis.

5.3.6.3 RA with Excess of Blasts

This form of RA has obvious dysmyelopoietic features in the bone marrow. The blood has less than 5% blasts and the bone marrow 5-20% blasts. There is more frequent progression to AML than for RA or RAS with 30% leukaemic transformation.

5.3.6.4 RAEB in transformation

When the blast cells make up between 20% and 30% in the bone marrow or Auer rods are present in blasts the cases are referred to as RAEB in transformation. 50-80% of these cases progress to overt leukaemia.

5.3.6.5 Chronic Myelomonocytic Leukaemia (CMML)

CMML has a similar morphology to RA with leukaemic progression similar to RAEB at around 20%. Peripheral monocytosis is present with a count of greater than 1x10^9/l and there is also a bone marrow proliferation of monocytes. Clinical course is related to percentage of blasts, leukocyte count and degree of monocytosis.

CMML with less than 5% blasts and few or no dysplastic myeloid blasts is the least likely to progress followed by CMML with 5-20% blasts and clear dysmyeloid cells. The most likely is referred to as CMML in transformation, as there is a high progression rate to Acute MML and this has between 20% and 30% bone marrow blasts.
5.3.7 **Chromosomal Abnormalities**

Recurring and consistent chromosomal abnormalities have been identified in the majority of leukaemia types as reviewed in 1994 (Rabbitts 1994). These abnormalities can take three forms: deletions, inversions or translocations and may be specific, affecting a tumour type, or idiopathic, only affecting a particular patient.

Leukaemia causing deletions usually result in the loss of a tumour suppressor gene that leads to a clonal growth advantage in a pluripotent haemopoietic cell.

In cases where a translocation or inversion has occurred this falls into two categories. The first type involves a T-cell receptor or immunoglobulin protein relocating within close proximity to a proto-oncogene, thus leading to activation of the oncogene. A well characterised example of this is in Burkitt's lymphoma where the c-MYC proto-oncogene is brought close to an immunoglobulin or TCR in the joining or diversity segment resulting in MYC activation. The specific translocation t(8;14)(q24;q32) is the most common occurring in 90% of cases.

The second group of translocations or inversions lead to breaks within the coding regions of two separate genes. These cases lead to the production of fusion proteins, often transcription factors are involved in the fusion implying altered transcription is important in tumour progression. A well studied example of a tumour specific fusion protein is BCR and cABL on the Philadelphia chromosome in CML cases. Tumour specific fusions are the most common chromosomal abnormalities seen in leukaemia cases and transcription factors appear to be the most common target (Rabbitts 1994). The fusion proteins are possible therapy targets as they are specific to the tumour cells themselves.

5.3.8 **Causes of Leukaemia**

Leukaemia like all cancer is a multistep disease, rare mutation within a haemopoietic stem cell leads to leukaemia but it is not an on or off switch. Several subsequent steps conferring growth advantage are required for leukaemic progression. The role of viruses has already been discussed and the possible role of HERVs is addressed in the next section. The other main initiators of leukaemia are chemicals, radiation and a hereditary basis.
5.3.8.1 Chemicals

Myelotoxic agents that can cause types of leukaemia include benzene and toluene and are common in an industrial setting causing an increased risk of AML. There is also an increased risk of AML amongst smokers possibly due to benzene. Other risk chemicals are alkylating agents, cytotoxic alkaloids and antibiotics e.g. chloramphenicol. Chemicals most commonly cause AML either de novo or via multiple myeloma, they can also cause Waldenström’s macroglobulinemia and acute leukaemia. AML often occurs following chemotherapy for another form of leukaemia or unrelated disease. Drugs may be activating a latent leukaemogenic virus, or cause the induction of chromosome abnormalities, or it may be a reparative response to bone marrow hypoplasia or aplasia, or just a result of general immunosuppression of the host. Other forms of leukaemia can develop following chemotherapy but at a very low incidence and is probably coincidentally as sporadic form.

5.3.8.2 Radiation

An excess of leukaemia amongst radiologists was noted in the 1940s, two famous victims were Marie Curie and her daughter Irene. Studies of Japanese atomic bomb survivors have shown a definite link. Radiation causes DNA strand breaks, which, when they are double stranded can lead to chromosomal rearrangements and then to leukaemia e.g. Philadelphia chromosome in CML or ALL. The bone marrow has high radiation sensitivity and is reflected in the types of leukaemia generally caused by radiation exposure, there are few cases of CLL, many of CML and AML is common. ALL is only seen in younger individuals and no cases of ATL are radiation linked. There is a risk of secondary leukaemia (AML or CML) after radiotherapy or after repeated X-rays.

5.3.8.3 Hereditary Basis

In animals the hereditary basis plays an important role in leukaemia. This may be due to later genetic rearrangements because of inherited characteristics favouring leukaemogenesis, for example defective responses to environmental carcinogens and infections, DNA repair and detoxification of mutagens. There is a possible increased familial risk associated with CLL and there is a very weak association with CML. There appears to be a high susceptibility to HTLV associated ATL amongst some
Japanese families. A link between childhood common ALL and HLA-DPB1 has been established implying an infectious aetiology (Taylor et al. 1995). Down’s syndrome containing trisomy 21 have a 20x greater risk of developing leukaemia.

5.3.9 Retroviruses and Leukaemia

The association between retroviruses and leukaemia is very well established. Numerous mammalian retroviruses both exogenous and endogenous have been shown to cause leukaemia including those affecting primates (for more information see Chapter 1). Chapter 3 detailed the case of the exogenous retrovirus HTLV-1 and ATL.

Integration of HERV may have great importance in leukaemia as specific chromosomal abnormalities often present in human leukaemias causing changes in expression of cellular genes (Rabbitts 1994).

Virus-like particles in EBV negative leukaemic cell lines derived from AML have been reported implying a potential for HERV particle production (Karpas et al. 1980).

A HRES-1 transcript of 6kb has been isolated from the RNA of cell lines derived from leukaemia including MA-T cells, HL-60 promyelocytic leukaemia cells and MOLT-4 T-cell leukaemia illustrating active expression in leukaemia of this HERV (Perl et al. 1989).

Antibodies to a protein p30, similar to SSAV, have been detected in a case of human CML suggesting endogenous HERV activation in this patient as compared with normal, in this case a sibling (Derks et al. 1982). A protein related to SSAV gp70 can also be detected in the sera of normal leukocytes and those of leukaemia patients. It is proposed that the presence of such antibodies is detrimental to the prognosis of acute leukaemias or CML in blast crisis reducing survival time and implying that HERV expression may accelerate leukaemia progression (Hehlmann et al. 1984).

The complete sequence of the HERV SMRV-H, now a member of the HERV-K family, was isolated from a lymphoblastoid cell line of ALL origin (Oda et al. 1988). This line produced D type particles as classified by EM.
HERV 4-1 mRNA transcripts are detectable in a T cell acute lymphocytic leukaemia cell line (8402) as well as colon and placental tissue (Gattoni Celli et al. 1986).

HERV-K10 expression is very low in normal sera but proteins have been detected in certain tissues. Antibodies to HERV-K Gag proteins have been seen in leukaemias, pregnancy and testicular tumours and the Gag protein has been further placed by immunoperoxidase detection to the cytoplasm of cells (Sauter et al. 1995). HERV-K pol expression has also been observed in leukocytes by PCR and certain transcripts were preferentially expressed in leukaemia cases (Brodsky et al. 1993).

The HERV-S71 that is closely related to simian sarcoma associated virus (SSAV) has been shown to be positioned at 18q21 which is involved in bcl-2 chromosomal translocations t(14:18) in B cell lymphomas and follicular lymphomas. S71 hybridising transcripts were also seen in the leukaemia cell line K562. A prominent 2.9kb transcript and minor 2.5, 3.6kb transcripts were observed (Leib-Mosch et al. 1992).

The existence of antibodies to the conserved retroviral protein p15E is further evidence for HERV protein expression in leukaemia cell lines (Cianciolo et al. 1985).

Gag related proteins and p15E related proteins have also been noted by antibodies in sera of leukaemia patients and leukaemia cell lines as well as in autoimmune disease (Krieg et al. 1990).

HERV-H expression has also been noted in both normal leukocytes and T-cell leukaemia cell lines (Lindeskog et al. 1997).

There are vestigial remnants of animal leukaemia causing viruses in human DNA, for example FeLV related sequences. This suggests previous epidemics in humans in the past (Stoye et al. 1985). There are also clusters of human leukaemia which could be caused by an as yet unknown viral agent (Stoye et al. 1985). One potential candidate is childhood ALL that appears to be caused by an infectious agent (Kinlen 1988; Kinlen et al. 1990; Greaves et al. 1993) however leukaemia seems to be a rare outcome from infection by a common virus.

Another consideration for the leukaemic potential of ERVs is that of recombination with an exogenous virus. An example of this in the murine system is Harvey and Kirsten murine leukaemia viruses from exogenous murine leukaemia
virus and endogenous VL30 elements (Chien et al. 1979; Ellis et al. 1980). This capacity for recombination makes any expression of HERVs potentially pathogenic.

In the previous sections the expression of human retroviral elements has been detailed along with the possible involvement with various disease states. Looking at the level of disease caused in other mammalian systems it seems unlikely that humans have escaped the influence of further retroviruses and therefore the hunt for pathogenic viruses should continue. Leukaemia is a sensible disease to investigate any retroviral involvement due to the close historical association of leukaemia and retroviruses.
5.4 Results

5.4.1 RNA Expression

In order to try to establish any expression of HERV1 or HERV2 total RNA was extracted from peripheral blood lymphocytes from 10 normal blood donors by the protocol previously detailed (section 2.2.12). Total RNA was also extracted from the cells of 10 leukaemia/lymphoma patients these are detailed below and were all obtained from Dr. Abraham Karpas.

Table 5.1 Leukaemia cases analysed.

<table>
<thead>
<tr>
<th>Case Number</th>
<th>Abbreviation</th>
<th>Leukaemia Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>394</td>
<td>Follicular</td>
<td>Follicular Lymphoma Ascites</td>
</tr>
<tr>
<td>442</td>
<td>AML</td>
<td>Acute Myeloblastic Leukaemia</td>
</tr>
<tr>
<td>494</td>
<td>Lymphoma</td>
<td>Lymphoma Pleural Effusion</td>
</tr>
<tr>
<td>792</td>
<td>AML</td>
<td>Acute Myeloblastic Leukaemia</td>
</tr>
<tr>
<td>944</td>
<td>AMoL</td>
<td>Acute Monocytic Leukaemia</td>
</tr>
<tr>
<td>946</td>
<td>AML</td>
<td>Acute Myeloblastic Leukaemia</td>
</tr>
<tr>
<td>992</td>
<td>PCL</td>
<td>Plasma Cell Leukaemia</td>
</tr>
<tr>
<td>1081</td>
<td>AMML</td>
<td>Acute Myelomonocytic Leukaemia</td>
</tr>
<tr>
<td>1154</td>
<td>CML</td>
<td>Chronic Myeloblastic Leukaemia</td>
</tr>
<tr>
<td>1369</td>
<td>AMML</td>
<td>Acute Myelomonocytic Leukaemia</td>
</tr>
</tbody>
</table>

All 20 total RNA samples were run on 0.8% agarose gel to separate the transcripts. All RNAs were quantified by OD at 260 and 280nm to determine purity and concentration. All 20 RNAs had a 260 to 280 ratio of 1.9-2.0 indicating pure RNA preparations. Approximately 20µg of total RNA was loaded per lane as measured by OD_{260}. This gel was then northern blotted as previously described (section 2.2.13). The resultant filter was then ready for probing by the isolated HERV elements. In order to confirm probe activity and specificity a small amount of unlabelled probe was run on the gel as a positive control and a small amount of vector
as a negative control. To give an estimate of RNA loading a β-Actin probe was also used on the same gel after stripping the filter (Figure 5.1).

**Figure 5.1** Northern Blot

![Northern Blot](image)

Northern blot of total RNA probed with a radiolabeled β-actin probe, the expected height of 1.8Kb is indicated by an arrow. The intensity indicates the equal loading of RNA. Lanes 1-10 are normal donors, lane 11 and 12 are the positive control HERV inserts. Lanes 13-22 are the leukaemia cases 394, 442, 494, 792, 944, 946, 992, 1081, 1154, 1369 for details see table (Table 5.1).

**5.4.2 Probes**

HERV1 and HERV2 were radiolabelled by random priming as detailed earlier (section 2.2.4). Only cloned inserts were used and the first screen was performed using the smaller fragments.

No hybridising transcripts were seen using either probe by this technique although the positive and negative controls both worked correctly indicating that the probes were specific and correctly labelled. The RNA also appeared to have blotted successfully as shown by the β-Actin probe (Figure 5.1). Alteration of stringency of washing and hybridisation conditions did not reveal any transcripts and only resulted in higher background (data not shown).

In an attempt to improve sensitivity riboprobes were produced as described (section 2.2.4). RNA-RNA species are more stable than RNA-DNA and therefore riboprobes should be more specific and reduce background so more stringent washing
can be employed. However in this case riboprobes were also unsuccessful and no specific transcripts were observed.

Non-radioactive labelling of the probes was also employed utilising the digoxygenin system, incorporating a digoxygenin labelled dUTP instead of P\textsuperscript{32} labelled dCTP, however the background appeared even higher using this technique so radioactive labelling was preferred. Radioactive labelling also had the advantage of repeated probe stripping with little loss of signal as measured by the positive control.

From this data it seemed that any expression was at a very low level if at all in leukocytes. However this was not unexpected due to the low copy number as determined by a very weak Southern blot signal and the number of stop codons present in the coding region. Therefore an attempt at RT-PCR was undertaken.

5.4.3 RT-PCR

After analysis of the obtained sequences of HERV1 and HERV2 no significant ORFs were noted from HERV2 in the correct reading frame for retroviral peptides, therefore efforts were concentrated on HERV1. Figure 5.2 below shows potential coding regions for HERV1 and 2. Vertical lines represent stop codons; ATG start codons are shown by half lines.

**Figure 5.2** ORF maps of HERV1 (A) and HERV2 (B).
Using this analysis several PCR primers were designed covering potential ORFs. All of the primer sets were confirmed to be functional by testing on the HERV clone. Their specificity was tested by using vector sequences as a template and in all cases bands of specific size were obtained by using the HERV clone as a template (data not shown).

One pair of primers was successful yielding faint bands in both normal blood lymphocytes and in leukaemia patients (Figure 5.3). This primer set covered bases 1116 to 1995. Although the RT-PCR was not quantitative a slightly higher expression level appeared to be present in the acute leukaemia cases compared to the normals and the lymphoma cases, in particular with the AML cases 442, 792, 944. The primers used are given in Chapter 2 (section 2.2.14).

Figure 5.3 RT-PCR

Panel A shows RT-PCR using HERV1 primers as given in Chapter 2, the arrow indicates the expected size of approximately 880bp.

Panel B shows the β-Actin RT-PCR using the β-Actin primers given in Chapter 2 (section 2.2.14) the arrow indicates the expected size of 540bp. Lane 21 is a positive β Actin control PCR using the same primers.
Lanes 1-10 are leukaemia cases 442, 792, 944, 946, 992, 394, 494, 1081, 1154, 1369 for details see table above (Table 5.1). Lanes 11-20 are normal RNA samples. Negative controls were also performed using RNA prior to RT reaction these controls were all negative.

In order to confirm the transcript sequence direct sequencing of the PCR products was performed. This confirmed the expected sequence of HERV1 in all the analysed cases. Where the amplified signal was weak this was re-amplified to yield a stronger signal for sequencing.

5.4.4 PAC Mapping

In leukaemia cases specific chromosomal locations appear to be important for the advance of disease (Rabbitts 1994). Many of these locations encode specific oncogenes or tumour suppressors and these genes appear to change their expression in leukaemia cases. Therefore the putative chromosomal location of any HERV associated with leukaemia was of interest. In order to study this localisation PAC DNA was utilised. The clones from which HERV1 and 2 were isolated were labelled with fluorochromes by nick translation as described (section 2.2.15). These probes were then used to hybridise to normal metaphase spreads (Vysis), examples of the resultant hybridisations are shown below (Figure 5.4). Image Capture software, SmartCapture 2 (Digital Scientific) was used to capture images, the banding analysis was performed using the Quipps analyser package (Vysis).

This led to a chromosomal location as indicated by the arrows. For 26J13 the band hybridised to as indicated by pairs of spots on both chromosome partners was 12q21. This gives the integration point of HERV2, as this was the PAC from which the sequence was derived. A similar process for HERV1 revealed an ambiguous result as indicated by 2 pairs of arrows (Figure 5.4B). This implies strong sequence homology over this PAC DNA on two chromosomes and therefore an integration site on 3q21 or 1p31.
Figure A: PAC FISH of 26J13 to a normal metaphase spread.

Figure B: PAC FISH of 27K22 to a normal metaphase spread.
5.5 Discussion

This Chapter has detailed the isolation of a transcript from the highly disrupted HERV1 element. This is the first report of expression of a member of the HERV-XA family as previous members were only isolated as genomic inserts. The existence of transcripts from even a disrupted element suggests that more intact members of this family may be present in the genome with greater levels of expression and possible biological actions. Other cases of HERV with biological activity have been discussed earlier and support this theory.

The potential involvement of this HERV1 transcript in leukaemia cases was investigated by RT-PCR and in most cases analysed the expression was very low as in normal peripheral blood RNA. However in three AML cases there was an apparent enhanced expression, although this was not tested quantitatively. In particular cell line 944 derived from an AMoL case gave a strong positive signal. This cannot be explained by an overloading of RNA as the β-Actin RT-PCR was at a similar level to the other samples tested. It would be of interest to investigate the integration site within this case and any possible involvement in a chromosomal translocation that may have led to its over expression.

The HERV1 transcript covers an area of approximately 800bp. This extends from the 3' portion of the protease into Domain 3 of the reverse transcriptase. It is unlikely that this transcript encodes a functional protein but taking into account cooperation with other retroelements this possibility cannot be discounted. The reason for the truncated nature of any peptide would be a frame shift with respect to a putative retroviral sequence at base 1609 leading to a stop codon soon after.

PAC screening led to the identification of potential integration sites. Integration close to any transcribed gene could cause activation of the element but integration close to transcription factors, immunogloblin-like genes, oncogenes or tumour suppressors is of most note in the context of leukaemic studies. The integration site for HERV2 was determined as 12q21. The region was investigated and several genes are encoded on 12q21 these include PAR4 that interacts with the WT1 (Wilms tumour suppressor). It is unlikely that HERV2 has integrated close to any coding region however due to its lack of expression. For HERV1 two loci were determined 1p31 and 3q21 this finding was confirmed by studying 10-15 metaphase
spreads of which an example is shown (Figure 5.4). To determine which is the correct integration site PCR primers for known genes within each region could be used. These primers could then be used for amplification of the known genes using PAC 27K22 as a template. However both loci were investigated for known genes. Many active genes are present on both loci. On 1p31 the most relevant gene is IL-12 receptor beta 2, integration close to this gene would lead to over expression of the HERV in peripheral blood. 3q21 as a locus has been associated with chromosomal translocations in some AML cases (Wieser et al. 2000) and therefore is of great interest as an integration site for HERV1. The transcription factor GATA-2 is located in this region (Wieser et al. 2000). Other genes present at this locus include CD80 and CD86 (B7-1 and B7-2) which are both regulatory signals for T lymphocytes and include immunoglobulin domains. These domains are often involved in genetic rearrangements. A further gene worthy of note is MCM 2 or CDCL1 (cell division cycle like 1) that is involved in the onset of DNA replication. Integration of HERV1 in proximity to any of these genes may influence HERV1 expression patterns and could possibly influence leukaemogenesis.
Chapter 6: General Discussion

Throughout this thesis links between human retroviruses and leukaemia have been reinforced. HTLV-I is the only human retrovirus that has so far been shown to cause leukaemia in the form ATL.

Chapter three detailed the case of a European Caucasian woman with ATL. A cell line had previously been derived from this patient and this line, Karpas 1682 was studied. The aim of the work discussed in Chapter 3 was to isolate and type the virus in order to determine any possible new strain of HTLV-I. The primary objective was to detect viral DNA, as anti-viral antibody presence had already been illustrated. The first method of detection was Southern blotting which resulted in a single band implying a monoclonal integration as is expected in ATL cases. Following the successful detection of HTLV-I sequences the intention was to perform multiple restriction digestions in order to perform RFLP analysis. RFLP would have led to accurate virus sub-typing by comparison with previously published restriction patterns due to the highly conserved sequence of HTLV. However this direct Southern blotting approach yielded ambiguous results mainly due to very faint signals. Therefore a PCR amplification approach was undertaken. As the greatest variability in HTLV sequence is seen in LTR and ENV regions these were amplified preferentially (Nerurkar et al. 1995). This approach was successful in amplifying both the LTR and ENV regions, they were sequenced as detailed and phylogenetic analysis performed. By using the LTR sequence of 1682 the closest homology was to an isolate from Peru (QU2), the 1682 virus therefore clustered consistently in the Cosmopolitan IaA subtype. This finding illustrates the spread of HTLV-Ia to European Caucasians and highlights the fact that ethnicity is no barrier to this pathogenic virus. Although at present in Europe the number cases of ATL and HAM/TSP are very low, this case illustrates that the virus is spreading and expanding its territory. This conclusion is drawn because in the case discussed there was no known direct contact with endemic areas. It is also worthy of note that amongst the immigrant populations of the USA HTLV is taking hold and the number of disease cases is increasing. To prevent this spread continuing into Europe an expansion of the screening programme should be considered using the reliable antibody test now
available. This could have a drastic effect on case numbers by enabling recommendations on the shortening of breast-feeding periods by carrier mothers and the use of barrier contraception. In particular in the USA it is unsurprising that amongst immigrants of Japanese and African decent HTLV infections have been detected. Therefore to prevent further spread through blood transfusion a blood screening programme has been implemented. Likewise the French also introduced testing of blood donors mainly because of the increasing numbers of individuals of African origin. Unfortunately in the UK blood donors are not tested for HTLV-I inspite of the fact that infection with the virus can be found in people of Caribbean origin. Pregnant ladies of West Indian origin in particular should be screened because of the high risk of transmission via breast feeding.

Co-operation between retroviruses is commonly observed in animal models, for example MCF in mice (Stoye et al. 1991). In fact in the human breast cancer cell line, T47D, both B and D type retroelements are associated with particles (Seifarth et al. 1995). In humans HTLV, as an active exogenous retrovirus, may co-operate with HERV. This is a possible mechanism of HERV activation and would allow disrupted elements or previously harmless elements to become biologically active and is a line of enquiry that may warrant further investigation.

Chapter four of this thesis detailed the isolation and characterisation of two novel HERV elements. A PCR approach using degenerate oligonucleotide primers was employed. The reverse transcriptase region was chosen for amplification due to its presence exclusively in the retroviral genome. Positive bands were isolated by TA cloning and following clone screening two novel fragments were identified and selected for further analysis. After screening a human BAC-PAC library with these fragments further sequence data was obtained. The pol region of these elements was then used for phylogenetic analysis.

The PAC screening also gave a rough indication of copy number. Southern blotting suggested a low copy number and this was supported by the PAC data. The library used is reported to cover the entire genome at least twice (Ioannou et al. 1996). Screening with either HERV fragment led to only 15-20 positive clones for each
element giving a copy number of 5-10 per element throughout the entire human genome.

The age or time of integration of these individual elements is difficult to establish, as 5' LTR sequences were not isolated for either HERV. The length of time within the genome can usually be calculated by comparing the 3' and 5' LTRs and then making a calculation using the expected mutation rate. However the highly disrupted nature of these elements does imply an ancient integration.

Of these two characterised novel HERV elements the shorter clone of around 2.8kb, named throughout as HERV2, consisted almost entirely of a pol region. The element contained a protease active site and was followed by all seven regions of reverse transcriptase. There appears to have been a large deletion of the env sequences therefore the 3' LTR followed pol immediately without any env sequence in between. An Alu sequence followed the LTR and the reverse transcriptase region was used in phylogenetic analysis of this element. HERV2 clustered consistently within the HERV-I family with a homology of around 80% by amino acid similarity. The 5' LTR was not isolated but it is likely that this element is in fact a HERV-I. No expression of HERV2 was observed despite extensive RT-PCR. This element was very highly disrupted with many frame shifts and stop codons. The ORF map analysis revealed only short potential coding stretches and no transcripts could be amplified covering any of these regions. It can be concluded that this novel element, like many HERV, is transcriptionally inactive in lymphocytes. Expression in other tissues however cannot be discounted. Other members of this family have previously been isolated with intact coding regions (Maeda et al. 1990) and the isolation of this previously undetected sequence illustrates the prevalence of HERV in general.

The longer HERV element, referred to as HERV1 throughout, contained the 3' region of gag and almost the entire pol region. The 3' HERV sequences are deleted including the integrase domain of pol and the entire env as well as the 3' LTR. This deletion seems to have coincided with a multiple Alu insertion that is present at the 3' after 3kb of HERV sequence.

A theoretical translation of HERV1 sequence revealed the conserved Gag MHR region, a protease active site as well as all seven domains of the reverse
transcriptase. This region showed the closest homology to previously reported XA elements and due to the limited sequence availability of these XA elements a more 3' region of pol was used for alignment and phylogenetic analysis. This data clearly clustered HERV1 with other XA elements with a homology of up to 80% by amino acid similarity. The closest relative was HERV-XA, as isolated by Tristem (Tristem 2000) and HERV1 consistently clustered with this element. The homology was 80% at the greatest (with HERV-XA) and may even imply a sub-family but until more elements of this family are fully isolated, including LTRs, this would be presumptive. The HERV-H family, represented by RGH2, was the next closest HERV family followed by HERV-F and HERV-Fb, all of which clustered together. The expected PBS for HERV1 is phenylalanine by comparison with other known XA elements, however unfortunately no 5' LTR was found. This is noteworthy as the most closely related family has a histidine PBS, this reinforces the existence of a separate family named XA rather than HERV1 being a distal member of HERV-H or HERV-F.

Significant homology of HERV1 with exogenous retroviruses was also observed. The most homologous were MuLV and GaEV with up to 50% homology over the Pol region. Over 45% homology was also observed with a porcine endogenous retroviruses PERV B43. This observation is of note because of potential xenotransplantation trials from pig to man. The close similarity of active PERV to disrupted HERV elements could raise concern. PERV have the potential to infect human cells directly (Patience et al. 1997) and a risk of disease in particular in immunocompromised transplant patients is a real possibility. A further risk however could be with retroviral co-operation or complementation between PERV proteins and disrupted, yet expressed, HERVs. This thesis has detailed expression of a highly disrupted HERV with some homology to PERV and more complete elements may have biological and disease implications via co-operation with PERV. This risk, although small, should be considered when assessing the safety of xenotransplantation.

It is a matter of debate whether HERV are remnants of previous exogenous infections or are evolutionary precursors of exogenous retroviruses or both. HERV may well have evolved directly from RT-like sequences in ancient history, but now they can provide a reservoir of retroviral coding sequences for complementation with infectious retroviruses or as a base for the production of novel exogenous agents. As
such these sequences should be carefully considered and any expression and coding potential noted.

No expression data has been reported for HERV-I elements to date but they have been implicated in genomic rearrangements (Maeda et al. 1986). There is also evidence that LTR sequences associated with cellular genes may control their expression (Armour et al. 1989). Recently expression of XA34 has been reported in placental tissue (Kjellman et al. 1999).

The ORF map for HERV2 revealed very few potential coding ORFs and only very small stretches within the putative coding frames. RT-PCR was performed using pairs of primers covering all potential coding regions and all RT-PCRs were negative for this element.

For HERV1 however there was one ORF in the +2 frame covering around 900 base pairs. This ORF was mainly in the correct reading frame for retroviral homology. Primers designed to cover this region were positive in RT-PCR reactions in both normal peripheral blood lymphocytes and in various leukaemia cases. One case of AML M5 gave a repeatably strong signal although this was not tested quantitatively, the same starting total RNA was used and the PCR conditions were identical. It would be worthwhile to test this sample in particular using real time RT-PCR in the Taqman system.

This thesis reports the first expression of a XA family member in peripheral blood as well as in leukaemia patients. This study has illustrated the coding potential of even highly disrupted HERV elements, Taqman RT-PCR analysis may lead to the discovery of many more HERV transcripts as well as a putative less disrupted XA element with true biological activity.

AMoL (AML M5) is an acute leukaemia type with a poor survival rate and rapid progression. Over expression of HERV1 in this cell line may only be indicative of altered expression due to a random chromosomal breakage, however aid in progression cannot be discounted. A general influence of this HERV is unlikely as demonstrated by the similar very low levels of expression seen in both leukaemic
patients and normal peripheral blood lymphocytes. Although a possible role in normal PBMC is possible.

PAC FISH revealed the integration locus of both HERV1 and HERV2; analysis of patient samples with chromosomal painting will reveal any translocations in this region and therefore any potential involvement in chromosomal rearrangements. In fact the locus 3q21, the integration site for HERV1, has been highlighted as a breakpoint in a subset of AML and MDS cases (Wieser et al. 2000). It is suggested that a transcription factor GATA-2 is important in these translocations (Wieser et al. 2000), integration of HERV1 close to this transcription factor could result in over expression of HERV1. Conversely HERV1 could in fact play a role in the translocation process itself by encouraging homologous recombination between similar HERV sequences.

The data presented here throws up many future lines of enquiry. Quantitative expression studies of HERV using real time RT-PCR via the Taqman would be valuable. Leukaemia patient expression in comparison to normal peripheral blood, with M5 cases in particular as well as cell lines, would be revealing. HERV expression in other tissues in particularly placental/germ line tissues should also be studied as expression in these tissues has been previously reported.

A search for other related elements by lower stringency hybridisation could be undertaken with the aim of discovering intact, less disrupted elements.

A more thorough search of the sequence databases as the human genome is sequenced will lead to the isolation of more intact HERV in all families. The expression of these elements could then be studied in detail.

HERV expression analysis in conjunction with cytogenetic studies would shed light on any potential involvement in breakpoints as well as any altered expression levels. Correlation of expression changes with PAC FISH screening could yield any involvement in translocations or integrations close to transcription factors, immunoglobulins or oncogenes and therefore any involvement in tumour progression.

RNA in situ expression studies would also be valuable for cellular location of transcripts and isolation to the tumour cells.

HERV expression could also be studied via transcriptomes using microarray technology yielding tissue expression of intact HERV or multiple disrupted elements.
This could lead to any tissue specific tumour involvement. Microarray technology would also enable a direct comparison of normal and leukaemia samples.

In summary this thesis has reported the isolation of two novel HERV sequences and illustrated expression of a transcript from one element. This expression is the first demonstrated in peripheral blood for an XA element. Continued search may lead to intact elements with involvement in leukaemogenesis. This thesis also supports continued vigilance in the fight against potential further spread of the pathogenic retrovirus HTLV-I from its endemic areas into Western Europe and the USA.
Chapter 7: Reference List


Matsuoka, M., Hattori, T., Chosa, T., Tsuda, H., Kuwata, S., Yoshida, M.,
Uchiyama, T., et al. (1986). “T3 surface molecules on adult T cell leukemia cells are
mouse mammary tumor virus that are specific to MCF-7 breast cancer cells.” Cancer-
Res. 49(14): 3879-83.
“Spontaneous malignant lymphoma and leukemia in a simian T-lymphotropic virus
“Replication of oncornavirus-like particle in human breast carcinoma cell line, MCF-
“Influence of human T-cell leukemia virus type I tax and rex on interleukin-2 gene
Transcriptase Encoding Human Endogenous Retroviral Sequences Similar to Type A
and Type B Retroviruses: Differential Transcription in Normal Human Tissues.”
endogenous retroviral sequences in peripheral blood mononuclear cells of healthy
repeats is associated with tandemly repetitive DNA in the human genome.” Proc-Natl-
(2000). “Syncytin is a captive retroviral envelope protein involved in human placental
Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsumo, J., Kinzler, K. W.,
insertion of L1 sequence in a colon cancer.” Cancer-Res. 52(3): 643-5.
structure of a retroviral protease proves relationship to aspartic protease family.”
Minakami, R., Kurose, K., Etoh, K., Furuhashi, Y., Hattori, M. and Sakaki, Y.
transcription and a nuclear factor(s) binding to the element.” Nucleic-Acids-Res.
20(12): 3139-45.
mediates production of an extrachromosomal circular DNA containing a transposon-
extrachromosomal circular DNA copies of the human transposable element THE-1 are
related predominantly to a single type of family member.” J-Mol-Biol. 196(2): 233-
43.


Chapter 7 Reference List


Appendix: Publications During Period of Registration

Journal articles


Conference Presentations


Appendix

  "The life cycle of canine oral papillomavirus: histopathology, immunohistochemistry, RNA and DNA in situ hybridisation of experimental infections, including studies of L1, E4, E7 and E2 protein expression." 18th International Papillomavirus Conference Barcelona. Poster Presentation.
