Studies on Human Endogenous Retroviruses (HERVs) with Special Focus on ERV3

BY

ANN-CATRIN ANDERSSON
Human endogenous retroviruses (HERVs) represent approximately 7% of the human genome. This investigation was focused on one particular HERV, ERV3, with the main purpose of characterising its gene expression patterns and genomic distribution of ERV3-like sequences. Furthermore, this careful expression study should provide insights into the biological role of HERVs. The impact of HERVs in health and disease is not yet clarified. ERV3 is expressed as three envelope (env) transcripts, of which two also contain a cellular gene, H-plk (human proviral linked Krüppel). ERV3 env expression was mainly investigated at the RNA level. The gene expression of two other HERVs, HERV-K and HERV-E was analysed and compared with ERV3 activity.

Real-time PCRs were developed and in combination with in situ hybridisation, it was found that ERV3 is expressed in a tissue- and cell-specific way. High levels of ERV3 mRNA (up to six times over histone3.3) were demonstrated in placenta, sebaceous glands, foetal and adult adrenal glands, brown fat, corpus luteum, pituitary gland, thymus and testis. In monocytic cells including both normal monocytes and malignant U-937 cells, elevated ERV3 mRNA levels were observed after retinoic acid (RA)-induced differentiation. An ERV3-encoded Env protein was detected in selected cases, one following RA-treatment. In addition, several new ERV3–like sequences were discovered in the human genome.

ERV3 was found to have conserved open reading frames in contrast to other ERV3-like sequences in the human genome. This suggests that ERV3 may be involved in important cellular processes such as differentiation, cell fusion, immunomodulation and protection against infectious retroviruses. The developed techniques and obtained results will allow further studies of HERV expression to better correlate HERV activity to both normal development and disease.

Key words: human endogenous retroviruses, HERV, ERV3, expression, in situ hybridisation, real-time PCR, U-937, retinoic acid.

Ann-Catrin Andersson, Department of Genetics and Pathology, Rudbeck Laboratory, SE-751 85 Uppsala, Sweden

© Ann-Catrin Andersson 2002

ISSN 0282-7476
ISBN 91-554-5342-2
Printed in Sweden by Eklundshofs Grafiska AB, 2002
This thesis is based on the following papers, which are referred to in the text by their roman numerals:


IV: Ann-Catrin Andersson, Zhihong Yun, Erik Larsson and Jonas Blomberg. ERV3 and related sequences in humans; studies of RNA expression by real-time PCR and by *in situ* hybridisation. (manuscript).

V: Ann-Catrin Andersson, Fredrik Öberg, Anna Dimberg, Jonas Blomberg and Erik Larsson. ERV3 in relation to cell differentiation in normal and neoplastic monocytes. (manuscript).

Reprints were made with permission from the publishers.

¹ © 1996, Blackwell Publishing


³ © 2002, Elsevier Science (USA)
ABBREVIATIONS ........................................................................................................ 6

INTRODUCTION ........................................................................................................ 7
General introductory comments .................................................................................... 8
Retroviral and transposable elements in the human genome ........................................... 8
Tropism and Envelope ................................................................................................. 10
Reverse transcriptase (RT) .......................................................................................... 10
RNA Reverse transcribing viruses ............................................................................. 12
Retroviruses ................................................................................................................ 12
Classification of retroviruses / Retroviridae ............................................................... 14
Host defence mechanisms ............................................................................................. 15
Retrovirus resistance .................................................................................................... 16
Immunosuppression ........................................................................................................ 16
Human endogenous retroviruses (HERVs) ................................................................. 17
Classification of HERVs ............................................................................................... 17
Regulation of genes through HERVs .......................................................................... 18
Expression during normal development and embryogenesis ...................................... 21
Involvement of HERVs in disease .............................................................................. 21
HERVs and cancer ........................................................................................................ 22
ERV3 (HERV-R) as a model for HERVs ................................................................. 24
ERV3 products .............................................................................................................. 24
Possible biological effect of ERV3 env expression .................................................... 27
ERV3 in pathological conditions ............................................................................... 28
ERV3 and related sequences ...................................................................................... 29

THE PRESENT INVESTIGATION
Specific aims ............................................................................................................... 30
Materials and methods .............................................................................................. 31
The monocytic cell line U-937 .................................................................................... 31
In situ Hybridisation – ISH .................................................................................... 31
Quantitative real-time PCR (QPCR) ....................................................................... 32
Results and discussion ............................................................................................... 33
I High expression of ERV3 in sebaceous glands in skin .......................................... 33
II Variation of expression of ERV3 in the human monoblastic cell line U-937, clone 1, during induced differentiation .......................................................... 34
III Developmental expression of ERV3 and HERV-K in human tissues .......................... 35
IV ERV3 and related sequences in humans; studies of RNA expression by real-time PCR and in situ hybridisation .......... 37
V ERV3 and monocytic differentiation ............................................ 39

CONCLUSIONS ........................................................................ 41

GENERAL DISCUSSIONS AND FUTURE ASPECTS ............................................. 42

SUMMARY IN SWEDISH - SAMMANFATTNING PÅ SVENSKA ......................... 44

ACKNOWLEDGEMENTS ............................................................ 46

REFERENCES .............................................................................. 49
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BaEV</td>
<td>Baboon endogenous retrovirus</td>
</tr>
<tr>
<td>bp</td>
<td>base pair(s)</td>
</tr>
<tr>
<td>env</td>
<td>envelope</td>
</tr>
<tr>
<td>ERV</td>
<td>endogenous retrovirus</td>
</tr>
<tr>
<td>gag</td>
<td>group specific antigen</td>
</tr>
<tr>
<td>GaLV</td>
<td>gibbon ape leukaemia virus</td>
</tr>
<tr>
<td>HERV</td>
<td>human endogenous retrovirus</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HML</td>
<td>human MMTV-like viruses</td>
</tr>
<tr>
<td>H-plk</td>
<td>human proviral linked Krüppel</td>
</tr>
<tr>
<td>HSRV</td>
<td>human spumavirus</td>
</tr>
<tr>
<td>HTDV</td>
<td>human teratoma derived virus</td>
</tr>
<tr>
<td>HTLV</td>
<td>human T-cell lymphotrophic virus</td>
</tr>
<tr>
<td>ISH</td>
<td>in situ hybridisering</td>
</tr>
<tr>
<td>kD</td>
<td>kilo Dalton</td>
</tr>
<tr>
<td>LINE</td>
<td>long interspersed elements</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeats</td>
</tr>
<tr>
<td>MLV</td>
<td>murine leukaemia virus</td>
</tr>
<tr>
<td>MMTV</td>
<td>mouse mammary tumour virus</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PBS</td>
<td>primer binding site</td>
</tr>
<tr>
<td>pol</td>
<td>polymerase</td>
</tr>
<tr>
<td>PolyA</td>
<td>poly adenyalted</td>
</tr>
<tr>
<td>QPCR</td>
<td>quantitative PCR</td>
</tr>
<tr>
<td>RA</td>
<td>all trans retinoic acid</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcriptase</td>
</tr>
<tr>
<td>Sag</td>
<td>superantigen</td>
</tr>
<tr>
<td>SINE</td>
<td>short interspersed element</td>
</tr>
<tr>
<td>SU</td>
<td>surface-part of the envelope protein</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane-part of the envelope protein</td>
</tr>
<tr>
<td>TPA</td>
<td>tumour promoting agent (Phorbol-12-myristate-13-acetate)</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>U3</td>
<td>unique 3’sequence</td>
</tr>
<tr>
<td>U5</td>
<td>unique 5’sequence</td>
</tr>
<tr>
<td>Vit D3</td>
<td>vitamin D3</td>
</tr>
<tr>
<td>VL30</td>
<td>virus-like 30S RNA element</td>
</tr>
</tbody>
</table>
INTRODUCTION

The human T-cell lymphotropic virus 1 (HTLV-1) was discovered in 1981 and was the first identified human-tropic infectious retrovirus (Hinuma et al., 1981; Yoshida et al., 1982). Soon after, in 1983, the third human retrovirus, human immunodeficiency virus (HIV), was isolated and later found to be responsible for the acquired immunodeficiency syndrome, AIDS (Gallo and Montagnier, 1987; Montagnier, 1987). Retroviruses had been established as causative agents for disease in mice and chicken since the 1950s. The Rous sarcoma virus discovered in 1911 was among the first isolated retroviruses. Prof. Rous realised that chicken tumours (sarcomas) were transmissible between birds (Rous, 1911). After injecting a cell-free preparation of chicken sarcoma tumour tissues he was able to conclude that the transferred tumour inducing preparation must contain an infectious agent. The next important discovery in the history of retroviruses was made by Prof. Bittner, who described that susceptibility to develop mammary tumours in mice was transferred from females to their offspring via breast milk (Bittner, 1942). This so-called Bittner-factor in murine breast milk, responsible for the development of mammary tumours in certain mice was later found to be a retrovirus. Because of the correlation between the above mentioned retroviruses and tumours, these viruses were designated RNA tumour viruses or oncoviruses. Nowadays we know that the group of retroviruses in mammals is large and not only do they cause tumours but they are also associated with immunodeficiency and neurological disorders (Coffin, 1992; Temin, 1992). Therefore, the term oncovirus was replaced by retrovirus, not least because of the retroviral life cycle involving reverse transcription. In fact, there is a fifth, newly identified exogenous human retrovirus, HRV5, not yet definitively associated to any disease process (Griffiths et al., 1997).

Genomes of mammals, apart from humans, have been known to contain endogenous retroviruses (ERVs) (Weiss and Payne, 1971). It was suspected for a long time that also the human genome harboured these kinds of sequences (Todaro, 1980). The final proof of this was provided in 1981, when Martin and co-workers first identified parts of human ERVs (Martin et al., 1981). Since then, several HERV families, as well as solitary long terminal repeats (LTRs) have been characterised (La Mantia et al., 1991; O’Brien et al., 1983; O’Connell et al., 1984; Ono et al., 1986).

This thesis is focused on human endogenous retroviruses, HERVs, with special emphasis on ERV3 (HERV-R) expression, also including HERV-K (paper III) and HERV-E (paper IV). An introduction of human retroelements in general and HERVs in particular is presented, followed by a description of ERV3 and related sequences. The
results presented in papers I-V are discussed as well as possible biological functions of HERVs.

Figure 1 The frequency of transposable elements in the human genome is depicted. Reverse splicing introns are not included here, but in an unconfirmed report it is mentioned to make up 16% of the human genome (Cousineau et al., 2000). Data was deduced from the following published articles; (1) reviewed in (Prak and Kazazian, 2000); (2) (Smit, 1999); (3) (Harrison et al., 2002). Adapted from Blomberg, J. (unpublished)

General introductory comments

Retroviral and transposable elements in the human genome

The load of retroviral repetitive sequences in the human genome has, since the discovery in the 1980s, been considered as junk, selfish or parasitic DNA of no importance for either health or development of disease. (Yotsuyanagi and Szöllösi, 1984). However, newer data supports the theory that these elements have participated in shaping our genome during evolution (Baltimore, 1985; Daniel and Chilton, 1978; Temin, 1992) increasing genome plasticity (Leib-Mösch et al., 1992; Löwer et al.,
1996) and providing beneficial effects for the species (Gardner et al., 1991; Schulte et al., 1996; Sverdlov, 2000; Whitelaw and Martin, 2001).

It can now be concluded that a significantly large portion of the human genome is derived from repetitive mobile sequences, e.g. transposable elements (see Fig. 1). Fractions of the transposable elements are still active and found not only to express their own sequences but also in addition affect the activity of cellular genes in different ways (see Table 3 and 4). In most cases they are probably harmless or symbiotic elements, but may also act by interrupting genes through insertional mutagenesis (reviewed in Prak and Kazazian, 2000; Smit, 1999; Wilkinson, 1994). Structural features of transposable retroelements in the human genome are presented in Table 1. Retroelements are usually divided into DNA transposons and retrotransposons, because of their different mechanisms during transposition. The retrotransposon-derived sequences use RNA as an intermediate in their life cycle while transposons use DNA. Alternatively, they can be classified according to whether they are autonomous or not. Non-autonomous transposable elements are dependent on other transposable elements for their mobilisation, while the autonomous elements are able to encode proteins and, or otherwise, harbour the capacity to enable the transposition within one cell (genome).

Table 1. Structures of transposable element types in the human genome.

<table>
<thead>
<tr>
<th>Retroelements</th>
<th>Autonomous</th>
<th>Example</th>
<th>Promoter region</th>
<th>Genes/open reading frames (ORFs)</th>
<th>PolyA-signal or -tail (AAA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-LTR retrotransposons</td>
<td>yes</td>
<td>LINEs (L1)</td>
<td>5’UTR</td>
<td>ORF1 (nucleic acid binding) ORF2 (RT)</td>
<td>3’UTR+AAA</td>
</tr>
<tr>
<td>HERVs (often defective)</td>
<td>yes</td>
<td>ERV3</td>
<td>5’LTR</td>
<td><em>gag</em></td>
<td>3’LTR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><em>prt</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><em>pol</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><em>env</em></td>
<td></td>
</tr>
<tr>
<td>LTR Retrotransposons</td>
<td>yes</td>
<td>Members of HERV-H</td>
<td>5’LTR</td>
<td>HERVs without env</td>
<td>3’LTR</td>
</tr>
<tr>
<td>DNA transposons2</td>
<td>yes</td>
<td>Sleeping beauty</td>
<td>no</td>
<td>transposase</td>
<td></td>
</tr>
<tr>
<td>SINEs</td>
<td>no</td>
<td>Alu</td>
<td>RNApol. III 2</td>
<td>2 non-coding GC-rich regions</td>
<td>PolyA</td>
</tr>
<tr>
<td>Older SINEs</td>
<td>no</td>
<td>tRNA-derived</td>
<td>no</td>
<td>1 tRNA-rel. +1 non-tRNA region</td>
<td>L13’UTR +AAA</td>
</tr>
<tr>
<td>Processed pseudogenes</td>
<td>no</td>
<td>any reverse transcribed gene</td>
<td>no</td>
<td>–</td>
<td>AAA</td>
</tr>
</tbody>
</table>
Tropism and Envelope

The retrotransposons have sequences corresponding to retroviruses and ERVs with the exception of the envelope (env) gene. This means that these elements can still transpose but are restricted to one genome. The Env thus provides a possibility for a retroelement, such as a retrovirus or an ERV, to infect other cells expressing the appropriate receptor. The host range, tropism, of a retrovirus is in this way restricted to certain cells, exemplified by the interaction between HIV and CD4 positive cells. This makes it also possible for some viruses to cross species barriers and also to adapt to new species. The production of ERV encoded Env proteins in mice and chicken have been shown to block superinfection of the corresponding exogenous retrovirus (described in “Defence mechanisms” and “Retrovirus resistance). The env gene in this system works as a restriction gene and is exemplified by the Fv4 gene in mice, found independently by Gardner and Suzuki, reviewed in Gardner (Gardner et al., 1991).

Despite the absence of env, retrotransposons can form and incorporate RNA into particles within the cell due to a packaging signal in gag (Figure 2). In mice, particles encoded by retrotransposons called intracisternal or intracytoplasmic A-type particles (IAP) are often associated with the cell membrane or with intracellular membrane vesicles. However, murine retrotransposons like IAP and VL-30 are present also in infectious retroviral particles as co-packaged material, demonstrating that diverse defective retroelements can complement each other (Boeke and Stoye, 1997). Several of the retrotransposons in human DNA are in fact defective HERVs; for example, many members of the HERV-H family lack the env gene (Wilkinson et al., 1993).

Reverse transcriptase (RT)

One main characteristic feature of retrotransposons, including the HERVs, is the possession of the reverse transcriptase gene, RT, or similar genes in other retroelements. The retroviral encoded enzyme RT was discovered in 1970 (Baltimore, 1985; Temin and Mizutani, 1970), and was shown to catalyse DNA replication from a RNA template. The RT gene, included in the polymerase (pol) region of retroviruses, is often used in phylogenetic studies because of its high conservation. In 1990, Xiong and Eickbush studied the evolutionary relationship of RT-containing elements from such diverse species as animals, plants, protozoans and bacteria (Xiong and Eickbush, 1990). Results from their phylogenetic studies indicated that the probable RT ancestor is an element with a retrotransposon structure but presumably without LTRs. In addition, RNA-viruses were also analysed and proposed to share a common ancestor with the retroelements, comparing the RNA-directed RNA polymerases. Human
telomerase is another gene that has sequence similarity to RT. In *Drosophila Melanogaster* the retrotransposons may function as telomerase enzymes (reviewed (Boeke and Stoye, 1997). Like the *pol* gene of retroviruses, telomerase also provides, RT-, RNase H- and endonuclease (EN) activity. Though the sequences are similar to the non-LTR retrotransposons, telomerase uses an intrinsic and species-specific RNA template for reverse transcription. The telomerase is likely provided by an ancient retroelement (Boeke and Stoye, 1997; Smit, 1999). The origin of retroelements, and of retroviruses in particular, was discussed by Howard Temin, the discoverer of the RT enzyme. He hypothesised that retroviruses could have originated from within a genome as a result of intragenomic transpositions and then later became exogenous (Temin, 1976). However, in a recent study from 1999 the RT, at least of the non-LTR elements, can now be phylogenetically dated back to the very beginning of eucaryotes (Malik et al., 1999). This contradicts the Temin-theory.

Integration of retroelements into the host genome has been seen to occur in different ways partly depending on the kind of retroelement. Largely, retroelements seem to favour actively transcribed regions in the DNA, probably because of an open chromatin structure. The mechanisms behind these preferences for integration are however not fully understood. The chromosomal distribution of proviral integration of retroviruses is described by the following sometimes contradictory terminology: non-random, evenly distributed, clustered, to actively transcribed regions, open chromatin structures and CpG-islands (Taruscio and Manuelidis, 1991). It is often described that regions of high GC-content acquire a lot of integrations, mainly because of the DNA structure mentioned above. Alu: s and L1 elements seem to be inhomogenously distributed on human chromosomes. However, Alu: s are more often found in early replication regions and CpG-islands while the L1 are concentrated within heterochromatin and GC-poor regions. Integration site preferences are also found to be an exact match of the proviral GC-composition, following a phenomenon designated isopycnicity (Rynditch et al., 1998). Thus, GC-poor viruses such as MMTV integrate preferentially in GC-poor regions and GC-rich viruses like HTLV-1 integrate in areas with a corresponding GC content. Also the LTRs mirror the GC-content of a virus. Although, transcriptionally active regions are said to attract integration, Weidhaas et al could however not confirm this hypothesis when analysing *de novo* insertions in a model system involving avian leucosis virus, ALV, in quail cells (Weidhaas et al., 2000). HERVs are often found in Alu and GC-rich regions. Integration sites of some HERVs are mapped close to a few hereditary malformation syndrome loci, HERV-E (Taruscio and Mantovani, 1996), or close to loci involved in tumour initiation or progression, ERV9 (Svensson et al., 2001) and to chromosomal breakpoints and radiation hot spots (Svensson et al., 2001). In addition, HERV-E also shows preferences for distal
telomeric regions. This is true for about 50 % of the HERV-E copies (Taruscio and Mantovani, 1996).

**RNA Reverse transcribing viruses**

The family of reverse transcribing viruses includes, besides Retroviridae, Metaviridae and Pseudoviridae. Among the Metaviridae there are viruses like Ty3-virus hosted by Fungi and the gypsy virus that has invertebrates like Drosophila Melanogaster as host. The Pseudoviridae includes two virus species that use invertebrates as hosts, Ty1 and Drosophila Melanogaster copia virus. None of these exist in vertebrates. The opposite situation is found in Retroviridae, where all seven genera have vertebrates as hosts.

**Retroviruses**

A retrovirus is the most complete retroelement and is characterised by three defined sets of regions of genes: gag (group specific antigen), pol (polymerase), env (envelope) flanked by long terminal repeats, LTRs. Retroviruses are RNA viruses encoding and carrying the characteristic RNA-dependent DNA polymerase enzyme, reverse transcriptase (RT). The virion carries the RT enzyme together with two copies of the viral plus-strand RNA. When entering the cell, one of the two single-stranded RNA molecules, is reverse transcribed into minus-strand cDNA and then into double-stranded DNA (Fig. 2). The modified DNA copy of the retrovirus is integrated in the host genome of the cell, thus creating a provirus including LTRs. cDNA synthesis starts at the primer binding site (PBS) near the U5 (5’ unique sequence) complementary to a 18 bp sequence of one of the tRNAs. Since different types of retroviruses use different tRNAs as primers, a taxonomy system for classifying endogenous retroviruses (ERVs) of humans is based on the use of tRNA as primers (see classification of HERVs).
LTRs

Preceding reverse transcription the full-length RNA, usually 7-10 kb, has 5’ and 3’ specific ends, U5 and U3 respectively, together with a direct repeat (R). During DNA synthesis the ends form LTRs with U3-R-U5 in 5’ to 3’ direction, seen in the integrated provirus. LTRs contain among other things strong promoter and enhancer elements. More specifically, LTRs regulate initiation of transcription, contain sequences capable of binding different regulatory transcription factors such as steroid receptors, decide cleavage and polyadenylation of the mRNA and direct the integration procedure by sequences recognised by the enzyme integrase, \textit{int}.

\textbf{gag} and \textbf{pol} Proteins encoded by \textit{gag} and \textit{pol} are produced as polyproteins that are processed into individual proteins. The protease gene is located between \textit{gag} and \textit{pol}. \textit{Gag} gives rise to three or four proteins, of which at least three are known to be structural proteins involved in virus assembly when budding of particles occurs. Specifically, the \textit{gag} proteins, designated MA (matrix), CA (capsid) and NC (nucleocapsid), maintain the three dimensional structure of the virion and the binding to the RNA molecules via the \textit{packaging signal}, $\psi$.

\textbf{pol} codes for the RT, RNase H and integrase, which are all processed in the same way together with the gag proteins. Many retroviral sequences are well conserved, and one of the most conserved is the \textit{pol} gene.

\textbf{env} The \textit{env} gene is transcribed and spliced to subgenomic RNA, translated, modified by glycosylation and cleaved into two proteins, which are components of the viral envelope. Although the two Env proteins are generated by proteolytic cleavage, they still interact with one another by chemical binding. The virus envelope surface (SU) and transmembrane (TM) proteins are responsible for the attachment to the receptor on the host cell, and the entrance into it. TM in the envelope plasma membrane anchors the SU to the envelope, and in co-operation with TM, it supports the virion to fuse with the host-cell membrane.
Moreover, retroviruses can exist either in exogenous or endogenous forms or both. The main difference is that an endogenous retrovirus enters the germline, in contrast to an exogenous retrovirus that is only able to enter cells with the appropriate virus-specific receptor for that virus, usually a somatic cell. Exogenous and endogenous retroviruses sometimes co-exist in the same species, such as the murine leukaemia virus (MLV) and mouse mammary tumour virus (MMTV) together with their respective endogenous counterparts. Some retroviruses are endogenous in one species but exogenous in another. This phenomenon was seen for gibbon ape leukaemia virus (GALV) and simian sarcoma associated virus SSAV, which both are exogenous in primates (Todaro, 1980). The endogenous counterpart of GALV was discovered in the murine species *M. caroli*, and appears to have been the infectious source for GALV. There are also other traces of retroviral cross-species transmission. The endogenisation of baboon endogenous retrovirus, BaEV, and the feline endogenous RD114 is probably an ancient event, since the putative common ancestor of the infectious retrovirus has not been found. Not only do BaEV and RD114 share stretches of sequence homology but they also use the same receptor, the type-D mammalian retrovirus receptor. In addition, the same type-D mammalian retrovirus receptor is used by the newly discovered HERV-W when fusing cells in human placenta (Blond et al., 1999; Blond et al., 2000).

The distribution of endogenous retroviruses in vertebrates is widespread and includes most vertebrate species (Herniou et al., 1998). The diversity of these viruses was investigated and some viruses were shown to cluster in phylogenetic groups, indicating that cross-species events seldom occur, at least in some groups. However, Martin and Tristem have reported two cases of quite recent horizontal transmissions (Martin et al., 1999). One of those interspecies transmissions was a retrovirus closely related to the porcine endogenous retroviruses (PERVs). The existence of an ERV in one species does not necessarily mean that the host range is restricted to that particular species, hence giving rise to xenotropism.

*Classification of retroviruses /Retroviridae*

Retrovirus classification was initially based on the morphology of the mature virion, during maturation and assembly of particles at the cell membrane (Coffin, 1992). Accordingly, the retroviruses were designated A-, B- and C-type retroviruses etc. Currently, the classification is based on sequence identity in the pol region, but also according to certain functional properties or genomic organisation in simple and complex structure. The seven genera of Retroviridae are presented in Table 2, also including a type species virus representing each genus as stated by the International Committee on Taxonomy of Viruses (ICTV) (van Regenmortel et al., 2000). A recent
and representative phylogenetic tree based on the pol gene, including examples of all retroviral groups, was published together with two newly discovered porcine ERVs, belonging to the Betaretrovirus genus (Ericsson et al., 2001). The Lentivirus HIV is a complex retrovirus with additional genes such as tat and rev.

Table 2. Taxonomic structure of the Retroviridae family

<table>
<thead>
<tr>
<th>Genera</th>
<th>Type species</th>
<th>Assigned abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpharetrovirus</td>
<td>Avian leukosis virus</td>
<td>ALV</td>
</tr>
<tr>
<td>Betaretrovirus</td>
<td>Mouse mammary tumor virus</td>
<td>MMTV</td>
</tr>
<tr>
<td>Gammaretrovirus</td>
<td>Murine leukemia virus</td>
<td>MLV</td>
</tr>
<tr>
<td>Deltaretrovirus</td>
<td>Bovine leukemia virus</td>
<td>BLV</td>
</tr>
<tr>
<td>Epsilonretrovirus</td>
<td>Walley dermal sarcoma virus</td>
<td>WDSV</td>
</tr>
<tr>
<td>Lentivirus</td>
<td>Human immunodeficiency virus 1</td>
<td>HIV-1</td>
</tr>
<tr>
<td>Spumavirus</td>
<td>Human foamy virus</td>
<td>HFV</td>
</tr>
</tbody>
</table>

Host defence mechanisms

As already described, retroviruses in some species exist in both endogenous and exogenous forms. In these cases discussed below, it is proposed that this is an example of an endogenisation process that persists until resistance is fully achieved, as seen in the example of BaEV in baboons and endogenous feline RD114 in cats (Löwer et al., 1996; Todaro, 1980). Perhaps ERVs early in evolution once protected human ancestors from exogenous retroviruses. This has in the literature been called intracellular immunisation (Baltimore, 1988). However, that time could be recent, considering that two HERV-K viruses are found only in a part of the population (Turner et al., 2001), further discussed in “HERVs and Cancer”. It is speculated that HIV, for example, in the future will become endogenised as a protection from HIV infection. The only exogenous retroviruses that are known to infect humans today are Delta-Lenti-and the newly discovered HRV5, belonging to the Betaretroviruses (Griffiths et al., 1997). There exist a few zoonotic infections with Spumaviruses originating from primates. The fifth, newly identified human retrovirus, HRV5, is still investigated for possible correlation to disease (Griffiths et al., 1999; Rigby et al., 1998). However, Spuma-like viruses are found as ERVs in humans (HERV-L) (Benit et al., 1999), but there are no known endogenous Lentiviruses.
Retrovirus resistance

Receptor interference, or receptor blockage, is carried out either by blockage or down-regulation of the cellular receptors that otherwise would provide an entrance for the exogenous retrovirus (Rubin, 1960). It occurs in chronically infected cells to prevent superinfection by a retrovirus using the same receptor (described in Weiss, 1993). The pattern of interference has been used to classify receptor groups (pseudotyping) and to determine the host range of retroviruses (Sommerfelt and Weiss, 1990). Like env from exogenous retroviruses, endogenous Env can also block receptors through interference, and this phenomenon was first described in chicken cells (Vogt, 1967; Weiss, 1967). Soluble Env are produced, secreted and attached to the receptor, thus blocking the receptor and protecting the cell from superinfection.

Another blocking mechanism observed in the murine system, is when capsid protein from the endogenous MuERV-L blocks the post-entry retroviral replication. MuERV-L–like sequences exist in humans (HERV-L), and could account for a similar mechanism (Best et al., 1996).

In mice, expression of the endogenous MMTV exclude certain V beta clones of T-cells minimising the cells that harbour the receptor for the exogenous MMTV. The reason behind the T-cell activation and subsequent depletion of T cell clones, is the superantigen, sag, encoded by an open reading frame (ORF) located in the LTR of the endogenous MMTV.

Immunosuppression

It is since long described that the TM part of Env, the prototype p15E, has immunosuppressive properties. The responsible amino acids have been limited to 17 and the peptide is hence designated CKS 17. In a rat model of yolk sac tumour, Lindvall et al. could demonstrate that the tumours expressed p15E-like proteins, and that tumour growth was inhibited by treatment with an anti p15E antibody (Lindvall and Sjögren, 1991). Mangeney and Heidmann elegantly showed in a tumour transplantation model that cells expressing the TM part of the MLV virus escape the immune mediated rejection (Mangeney and Heidmann, 1998), also mentioned under “HERVs and cancer”. A similar immunosuppressive motif is present also in filoviruses such as Marburg-Ebola, and shows a high conservation (Becker, 1996). The p15E protein, or parts of it, has among others the following documented effects on cells: induces cyclic AMP (Haraguchi et al., 1995), inhibits chemotaxis (Benomar et al.,
binds to transforming growth factor β-receptor and inhibition of DNA synthesis (Huang and Huang, 1998). The exact mechanisms behind this TM protein induced immunosuppression are still not known and remains to be established.

**Human endogenous retroviruses (HERVs)**

The integrated proviral form of an exogenous retrovirus has the same fundamental structure as an ERV. Clearly, most HERVs are endogenised exogenous retroviruses, once integrated in our genomes during primate evolution. Integration of HERVs into the germline is thought to have occurred 2 to about 70 million years ago. Thus, they are present in all human cells and inherited according to expected Mendelian laws. Although many of the HERVs are defective, some still have ORFs that are free from deletions and mutations. From these sequences several types of expression are seen, from subgenomic and full-length RNA transcripts (Kato et al., 1987; Katsumata et al., 1998; Leib-Mösch et al., 1992; Medstrand et al., 1992; Shiroma et al., 2001; Sibata et al., 1997; Tamura et al., 1997; Tönjes et al., 1996), to complete retroviral particles with polymerase- and protease-activity (Berkhout et al., 1999; Mueller-Lantzsch et al., 1993; Towler et al., 1998; Tönjes et al., 1997a). The maintenance of complete and undefective ERVs in the human genome indicates that HERVs might play an important role in biological processes.

**Classification of HERVs**

HERV families exist in broad diversity concerning their relation to exogenous retroviruses, copy number and expression. There are different systems to classify HERVs. The earliest classification was based on the tRNA specificity of the PBS by adding the one-letter code for the specific amino acid to HERV (Cohen and Larsson, 1988; Larsson et al., 1989). For example, one of the most studied HERVs, HERV-K, uses a lysine tRNA (K) as the specific primer. Because of the confusion caused when two viruses use the same tRNA primer, this system was complemented by other ways to designate HERVs. Other classifications of HERVs are based on their sequence identity to known exogenous retroviruses and sometimes with additional information according to copy number. Class I HERVs includes MLV- and BaEV-like viruses that belong to the *Gammaretroviruses*, former C-type retroviruses. Class II on the other hand consists of the MMTV-like viruses, also designated human MMTV-like viruses, the HML-group (Andersson et al., 1999; Medstrand and Blomberg, 1993). The large HML-groups cluster with the exogenous *Betaretroviruses* in phylogenetic analysis. Class III consists of HERV-L and related.
Regulation of genes through HERVs

Deduced from table 3 and 4, are many examples of HERVs affecting cellular genes, often by their LTRs. There seems to be a complex interplay between regulatory elements and coding sequences (Whitelaw and Martin, 2001). Retroelements have been shown to affect regulation of genes usually by providing promoter-enhancer function or alternative splicing. Some retroelements involved in gene regulation were discovered in Drosophila Melanogaster. The I-element existing in D. Melanogaster have been shown
to regulate and to limit integration of new I-element through a mechanism described as homology dependent co-suppression (Jensen et al., 1999)

### Table 3

<table>
<thead>
<tr>
<th>Retroelements</th>
<th>Examples of documented effects</th>
</tr>
</thead>
</table>
| **Non-LTR retrotransposons**   | Insertion of L1 in Myc locus in a breast cancer (Morse et al., 1988)  
Insertion of L1 in the APC gene in a colon cancer (Miki et al., 1992)  
Antisense L1-promoter driving transcription of cellular genes (Speek, 2001)  
Exon shuffling by L1 retrotransposition (Moran et al., 1999)                                                                                      |
| **HERVs**                      | Homologous recombination causing deletion of AZFa (Sun et al., 2000)  
Fusion of cells in placenta by HERV-W Env, Syncytin (Mi et al., 2000)  
Providing tissue-specific expression of:  
–pleiotrophin in placenta (Schulte et al., 1996)  
–amylase in parotid gland (Ting et al., 1992)  
Contributing to allelic variation, ERV9 insertion into the MHC-region (Årvidsson et al., 1995; Svensson and Andersson, 1997)  
Phylogenetical tools/ molecular clocks (Andersson et al., 1998b)  
Function as cellular promoters, driving expression of:  
–human proviral linked H-plk by ERV3 (Kato et al., 1990)  
–Zn finger protein ZNF80 by ERV9 (Di Cristofano et al., 1995)  
HERV-H Env expressed by murine tumour cells escape immune rejection (Mangeney et al., 2001)                                                                 |
| **Solitary LTRs**              | Function as cellular promoters, driving expression of:  
–phospholipase A2-related gene by RTVL-H (Feuchter-Murthy et al., 1993)  
Alternative promoter for:  
–Endothelin B receptor (Medstrand et al., 2001)  
–Apolipoprotein C-I (Medstrand et al., 2001)  
Alternative splicing of human Leptin receptor (Kapitonov and Jurka, 1999)  
Providing poly A signal to cytochrome c1 (Suzuki et al., 1990)                                                                                     |
| **SINEs**                      | Alu insertion causing  
–neurofibromatosis NF1 (Wallace et al., 1991)  
–Fukuyama-type congenital muscular dystrophy (Kobayashi et al., 1998)  
Alu increasing transcription and decreasing translation of a human Zn finger gene ZNF177 (Landry et al., 2001)                                      |
<table>
<thead>
<tr>
<th>Retroelements</th>
<th>Examples of speculated effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-LTR retrotransposons</td>
<td>Origin of Telomerase? (Malik et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>Implications in rheumatoid arthritis (RA)? (Neidhart et al., 2000)</td>
</tr>
<tr>
<td>HERVs</td>
<td>Transcriptional fusion between genes (Kowalski et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>Regulating the transcription of the human β-globin genes? (Long et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>Abnormal gene activation via HERV LTRs (Taruscio and Mantovani, 1998b)</td>
</tr>
<tr>
<td></td>
<td>Antisense transcription of genes (Schneider et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>–Schizophrenia and psychosis? (Crow, 1984; Deb-Rinker et al., 1999; Hart et al., 1999;</td>
</tr>
<tr>
<td></td>
<td>Karlsson et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>–Autoimmune diseases? (Christensen et al., 2000; Conrad et al., 1997; Perron et al., 1997;</td>
</tr>
<tr>
<td></td>
<td>Stauffer et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>–not supportive; (Lan et al., 1998; Löwer et al., 1998; Murphy et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>Provide new cell tropism for HIV? (An et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>Provide enzyme activity for the cell or for HIV?</td>
</tr>
<tr>
<td></td>
<td>–dUTPase (Harris et al., 2000)</td>
</tr>
<tr>
<td></td>
<td>–protease activity (Mueller-Lantzsch et al., 1993; Towler et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>–RT-activity (Berkhout et al., 1999; Simpson et al., 1996)</td>
</tr>
<tr>
<td></td>
<td>Recombination with retroviral vectors in gene therapy (Chong et al., 1998; Patience et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>Recombination with pig endogenous retrovirus (PERV) in xenotransplantation? (Le Tissier et al., 1997; Patience et al., 1997)</td>
</tr>
<tr>
<td></td>
<td>Insertional polymorphism-infectious HERVs? (Turner et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>Creates genome instability through</td>
</tr>
<tr>
<td></td>
<td>–chromosomal breakpoints? (Svensson et al., 2001; Taruscio and Manuelidis, 1991)</td>
</tr>
<tr>
<td></td>
<td>–radiation hot spots? (Taruscio and Mantovani, 1996)</td>
</tr>
<tr>
<td></td>
<td>Advantageous with species-specific integrations? (Schulte et al., 1996)</td>
</tr>
<tr>
<td></td>
<td>Tumour cells with Env escape the immune system? (Mangeney et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>Anti-HERV-K antibodies as tumour marker for seminomas? (Sauter et al., 1996)</td>
</tr>
<tr>
<td></td>
<td>Local immunosuppression during pregnancy? (Taruscio and Mantovani, 1998a; Villarreal, 1997)</td>
</tr>
<tr>
<td></td>
<td>Involvement of Env in differentiation, placental function and retrovirus protection? (Andersson et al., 1996a; Boyd et al., 1993; Lin et al., 2000)</td>
</tr>
<tr>
<td></td>
<td>Transactivation through HERV-K Rec protein? (Boese et al., 2000)</td>
</tr>
<tr>
<td>Solitary LTRs</td>
<td>Providing tissue-specific expression in gene therapy? (Schön et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>The proposed events above involving LTRs</td>
</tr>
<tr>
<td>LTR retrotransposons</td>
<td>The proposed events above involving HERVs, except for env-related events</td>
</tr>
<tr>
<td>Alu (SINE family)</td>
<td>Stress-related function of Alu through induced protein translation (reviewed in Smit, 1999)</td>
</tr>
</tbody>
</table>
Expression during normal development and embryogenesis

Some of the HERV genes are still intact and expressed, although the majority are mutated containing in-frame stop codons. The most abundant expression of different HERVs is seen in placenta and embryonic tissues, or in other reproductive tissues or cells, such as testis and oocytes. Not only proteins but also whole particles are seen (Kättstrom et al., 1989; Larsson et al., 1981; Nilsson et al., 1981; Yotsuyanagi and Szöllösi, 1984), suggesting a normal function for these proteins in this environment (Levy, 1986; Mwenda, 1994; Taruscio and Mantovani, 1998a). Recently, a HERV-encoded protein, denoted Syncytin, was shown to fuse trophoblasts in human placenta. Syncytin was cloned from placenta and was later found to be part of the envelope gene in the HERV-W genome (Mi et al., 2000). Except from the effect reported for the syncytin protein, other involvements of HERVs during placental or normal development are speculative. In theory, if HERV proteins could influence immunosuppression locally in placenta, it could provide a foeto-maternal immunosuppression during pregnancy (Larsson et al., 1994; Taruscio and Mantovani, 1998a; Villarreal, 1997). Furthermore, particles or env-proteins in placenta tissue could act as protection against exogenous retroviruses through receptor interference, i.e., receptor blockage. It has been proposed that ERV3 together with other HERVs are associated with differentiation and ERV3 also with fusion processes during placental development (Boyd et al., 1993; Larsson et al., 1996; Lin et al., 1999; Venables et al., 1995). This could implicate an involvement in growth regulation or differentiation of the cells in placenta. The broad expression of HERVs in embryonic tissues is likely to be sufficient for induction immunological tolerance towards HERV-encoded proteins.

Involvement of HERVs in disease

The pathogenic roles of HERVs have been heavily debated throughout the years and are still controversial. Soon after the review published by Dr. R. Löwer in Trends in Microbiology, with the teasing title “The pathogenic potential of endogenous retroviruses: facts and fantasies” (Löwer, 1999) and several comments were published in the subsequent issues of that journal (Mager, 1999; Stoye, 1999). Based on the knowledge about exogenous retroviruses and their interaction with the immune system, immunological disorders have been in focus for speculations. For example, the debate is still active concerning the possible involvement of HERVs in insulin dependent diabetes mellitus, IDDM (Conrad, 1998; Conrad et al., 1997; Stauffer et al., 2001). Several subsequent reports have been unable to provide experimental evidence supporting this notion (Badenhoop et al., 1999; Lapatschek et al., 2000; Löwer et al., 1998; Murphy et al., 1998).
Anti-HERV antibodies and retroviral nucleic acid are detected, mainly in blood, in several autoimmune diseases such as systemic lupus and multiple sclerosis MS. However, this does not prove if HERV expression is a cause or a consequence. The exact mechanisms behind these proposals are unknown but would probably involve violation of tolerance to HERV proteins or expression of superantigens (Posnett and Yarilina, 2001; Stauffer et al., 2001; Sutkowski et al., 2001). The outstanding questions are (i) what is the consequence if only partial tolerance is achieved or (ii) if tolerance is lost due to events such as expression of proteins in a tissue not normally expressing this protein or crossreactivity against proteins of exogenous retroviruses? Perron and co-workers found a retrovirus that was associated with multiple sclerosis (MS). This retrovirus was designated MSRV (Perron et al., 1997) and are now referred to as HERV-W. It is not excluded, so far, that this might be a new intact exogenous retrovirus. In addition, there is a newly established correlation between the HLA-DQ locus with HERV-K(HML-6) LTRs (DQ-LTR3) and rheumatoid arthritis (Seidl et al., 1999). Loss of a normal HERV-function, for example a protein, could also have pathogenic effects. Schizophrenia is another disease that has been associated with HERVs. Antiretroviral antibodies and HERV nucleic acids in particles have been reported in persons suffering from schizophrenia (Deb-Rinker et al., 1999; Hart et al., 1999; Karlsson et al., 2001) Anti-HERV IgM-antibodies are often seen in healthy individuals, and switch to IgG in HIV infected and SLE patients (Lawoko et al., 2000)

**HERVs and cancer**

Association between infectious retroviruses and cancer is well established and are exemplified in the introductory chapter (reviewed in (Weiss, 1992). The mechanisms of retroviral involvement in tumourigenesis are firstly: acute transformation of cells (insertional mutagenesis) and secondly; transactivation involving retroviral proteins such as Rev and Rex of HIV and HTLV, respectively. Regarding HERVs and cancer, there is no clear case of a human cancer initiated and caused solely by a HERV. There is now established that the development of human tumours is a multistep process in which activation of oncogenes and inactivation of tumour suppressor genes are critical events (Cho and Vogelstein, 1992; Hanahan and Weinberg, 2000).

If HERVs are involved in these processes, they would probably contribute with only one step in this multistep process. One such step would be through insertional mutagenesis, like the example for another retroelement, L1 and the *de novo* insertion in the APC gene in a colon carcinoma (Miki et al., 1992). Retrotranspositions of engineered L1 elements has been shown to occur *in vitro* in human HeLa cells (Moran
et al., 1999). Not much is known regarding the frequency of retrotransposition of HERVs in the human genome. However, a recent study revealed two new HERV-K variants, HERV-K113 and HERV-K115, that only exist in 16 and 29% of the population, implicating that these could be infectious and hence also cause insertional mutagenesis (Turner et al., 2001).

Expression of several HERVs are seen both in normal and tumour tissues. Some HERVs are expressed in tumour tissues, either detected as mRNA or indirectly by the presence of anti-retroviral antibodies, however without any known connection to the tumourigensis. A few examples will be presented below, involving the human teratocarcinoma derived virus HTDV/HERV-K belonging to the HML-2-group (Boller et al., 1993). This virus has an additional gene Rec (former cORF) that has similar transactivating property as Rex from the HTLV-1. Rec was described in 1993 and proved that HERV-K had a genomic structure comparable with that of a complex retroviruses (Löwer et al., 1993). It was later found that only type 2 genomes of HTDV/HERV-K (HML-2) could produce the 14kD Rec protein, and similar to Rev related proteins it locates in the nucleolus (Löwer et al., 1995). The possible connection between Rec and tumours is not only because of the expression of Rec in the teratocarcinoma cell line, but also because with its association with the promyelocytic leukaemia Zn-finger protein, PLZF (Boese et al., 2000). It is proposed that Rec could interfere with the normal function of PLZF. Since PLZF is involved in spermatogenesis, at least in mice, Rec could hereby contribute to for example germ-cell tumours such as seminomas (Boese et al., 2000). The function of Rec is to transport viral RNA from the nucleus to the cytoplasm involving also cellular proteins (Magin et al., 2000; Magin et al., 1999).

Antibodies against HERV-K Gag proteins and to a lesser extent Env were detected in patient sera in different pathological conditions but especially high in patients with seminomas (Sauter et al., 1995). Sera were used to immunologically localise HERV-K proteins to the seminoma tumour cells. (Sauter et al., 1996). It was concluded that anti-HERV-K antibodies could be detected almost exclusively in seminoma patients. Based on that it has been implicated that the presence of those antibodies could be used in diagnostic screening for seminomas. However, due to the low incident of this tumour type, it will probably not be used as such (Goedert et al., 1999).

The production of TM proteins by tumour cells could suppress the immune system and facilitate tumour progression. In fact, several human tumours produce p15E-like (TM) proteins, such as head and neck cancer, among which the relapses of tumours have been monitored by presence of p15E (Simons et al., 1994; Tas et al., 1993). It is thus
advantageous for a tumour cell to express retroviral TM proteins. This has been shown in a murine system both for the MLV p15E gene and the env gene of HERV-H, when expressed in cells transplanted to an allogeneic donor. The under normal conditions rejected tumour cells were in this experimental system able to escape the immune system of the donor mice (Mangeney et al., 2001; Mangeney and Heidmann, 1998).

ERV3 (HERV-R) as a model for HERVs

ERV3 is a complete 9.6kb human provirus belonging to the Gammaretroviruses. Originally, it was detected by O’Connell and Cohen when hybridising a human genomic library, with two nonhuman retroviral probes under low stringency conditions (O’Connell et al., 1984; O’Connell and Cohen, 1984). The retroviral probes represented LTR and pol sequences from BaEV. Based on analysis of the LTR sequences it was found that this provirus resembles class I, the former mammalian type-C viruses and is present only in primates. Instead of, like classical mammalian Gammaretroviruses, having a PBS corresponding to proline tRNA, ERV3 PBS is complementary to arginine tRNA, i.e., R. Together with HERV-E, it is a member of the HERV-ERI superfamily (Wilkinson, 1994). Surprisingly, there is also a HERV-R in Repbase (Jurka, 2000) that does not correspond to ERV3 but is 100% identical with BaEV (Tamura et al., 1981) and is therefore not a HERV (from a search performed in February 2002). As a consequence of that, the term ERV3 will be used in this thesis instead of HERV-R, in order to avoid confusion.

ERV3 products

Although being a complete provirus, due to in-frame termination-codons in gag and pol, it is replication defective. In most individuals, ERV3 env is free from stop codons and is expressed as Poly-A mRNAs (Cohen et al., 1985). In 1998, de Parseval and Rasmussen independently discovered a nucleotide polymorphism that results in a stop codon at nt 1352 in the original sequence deposited by Cohen 1985 (Genbank Acc no M12140.1) (Cohen et al., 1985; de Parseval and Heidmann, 1998; Rasmussen and Clausen, 1998). De Parseval investigated the frequency of hetero- and homozygosity for this C to T transition, and could conclude that 1% of the population carries the stop codon. Prior to this report, the ERV3 Env was proposed to play a role in the development of normal placenta. (Boyd et al., 1993; Kato et al., 1987; Larsson et al., 1994). To test this hypothesis, individuals carrying the stop codon were investigated regarding fertility and normal pregnancy. No disturbances were observed concerning fertility and placenta formation. From these results, the authors concluded that ERV3 Env is not vital for placental development in humans. In-vitro translated Env
containing the stop codon produces a 25kD protein (de Parseval and Heidmann, 1998), instead of the full length Env (65kD) as observed by Venables, (Venables et al., 1995). Since the ERV3 env contains seven glycosylation sites as reported by Cohen in 1985, the molecular weight of the protein was predicted to 68kD. From Western blot data using an anti ERV3 TM polyclonal antibody it is nevertheless believed that the protein is unglycosylated (Venables et al., 1995). However, in the report from de Parseval it was noted that a nucleotide (G at 796) is missing in the original report of Cohen. As a consequence of this insertion, the probable initiator codon (Met-ATG) is moved upstream, resulting in a more presumable start signal with three met within 18 nucleotides, starting at nt 689.

![Figure 4](image_url)

**Figure 4** Genetic map of ERV3 locus at 7q11 (clone AC073120) and structures of the three splice variants of transcripts from the env region. Exons are represented by thin bars and intervening sequences, in between, by lines. Abbreviations: LTR, long terminal repeat; kb, kilobase pairs; gag, group specific antigen; pol, polymerase; env, envelope; H-plk, human proviral linked Krüppel. Modified from Cohen M, 1988.

env is expressed in three different splice variants of sizes; 3.5, 7.3 and 9.0 kb (Cohen et al., 1985). None of these transcripts contains gag-pol sequences, which surprised Kato and co-workers, and they speculated if positive or negative factors could promote this effective splicing. Still, this has not been clarified. The two larger transcripts contain cellular sequences, in addition to viral sequences (Kato et al., 1987). Transcription of the 7.3 and 9.0 kb continues through the env, 3´LTR and into a cellular gene designated H-plk (human proviral linked Krüppel) (Kato et al., 1990). Krüppel is a transcription factor found in Drosophila Melanogaster. Members of this large group of zinc finger proteins are able to bind DNA and act as, for example a transcriptional repressor. In humans, there are 500 to 1000 members of the Krüppel-related protein family. Thus, it is probably the largest nucleic acid-binding family. A function of H-plk is still not
known and is further discussed below. All three transcripts are seen in human placenta tissue, thymus, ovary and testis (see Figure 5 for Northern blot data). In other tissues only the 3.5 or 9.0 kb transcript or both are present, and at different levels (Andersson et al., 1998a).

Figure 5. Multiple tissue Northern blot analyses of normal human tissues. Tissue samples are indicated above. Sizes derived from an RNA marker are indicated to the right of the filter. This figure is published as Figure 2 in the issue of International Journal of Oncology 12: 309-313 (1998), (Andersson et al., 1998a). Reprint of this figure was made with permission from the publisher.
Possible biological effects of ERV3 env expression

ERV3 was once suspected to contain a tumour suppressor gene. This was based on the high expression observed in chorion of normal full-term placenta and the total absence of ERV3 mRNA expression in choriocarcinomas and in semimalignant hydatidiform moles (Cohen et al., 1988; Kato et al., 1988). Neither hypermethylation nor deletion of the ERV3 locus could explain the lack of expression in these pathological conditions of placental development (Kato et al., 1988). These data were initially based on Northern blot analysis. Later we reported, based on ISH data, that the expression of ERV3 env mRNA was localised to the syncytiotrophoblasts in the chorion villi (Larsson et al., 1994). ISH with the ERV3 env probe revealed comparable high expression in the small number of syncytiotrophoblasts present in choriocarcinoma as compared to normal placental syncytiotrophoblasts. Choriocarcinoma cells contain immature trophoblasts and very few differentiated syncytiotrophoblasts. The loss of ERV3 expression could thus be a consequence rather than a cause of choriocarcinoma, or alternatively could ERV3 be involved in development and differentiation as discussed further below. Syncytin, the Env protein of HERV-W, is also expressed in the trophoblasts of placenta, and has been suggested to induce cell fusion (Mi et al., 2000). In addition, the amount of syncytin protein has been observed to decrease and relocate from the syncytial villi to other parts of the placenta in women with pre-eclampsia (Lee et al., 2001). If this relocation of syncytin is causally involved in this disease remains to be elucidated.

ERV3 env expression has been associated to cellular differentiation in two model systems of induced in vitro differentiation; e.g. the monocytic cell line U-937 cells and the trophoblastic cell line BeWo (Boyd et al., 1993; Larsson et al., 1996; Lin et al., 1999). The first observation proposing an involvement of ERV3 in differentiation was reported by Boyd et al (Boyd et al., 1993). In that study, freshly isolated trophoblasts from placenta were grown in vitro and monitored for ERV3 expression, during their spontaneous fusion to syncytiotrophoblasts. Elevated ERV3 expression and cell fusion occurred in parallel during differentiation. This was later confirmed by Lin and co-workers in 1999 (Lin et al., 1999). They used the trophoblastic cell line BeWo as a model system in which they could show that BeWo cells were able to differentiate in vitro by adding forskolin, thus inducing beta-HCG production and formation of multinucleated cells. Recently, Lin et al investigated this phenomenon one step further by transfecting the cell line BeWo with an ERV3 env expression vector. These transfected cells were found to both fuse and produce beta-HCG (Lin et al., 2000). One obstacle regarding a function of ERV3 Env is that the TM of ERV3 Env is truncated and it is speculated if it at all could anchor in the plasma membrane.
(Cohen et al., 1985). Anyhow, if the protein because of this truncation is secreted, it could still be of importance to the producing cells in their interaction with surrounding cells.

The ERV3 expression data obtained from the cell line U-937 support the hypothesis that ERV3 expression, or ERV3 regulated expression of env-H-plk transcripts, correlates with certain stages of maturation during monocytic differentiation (Åbrink et al., 1998). Krüppel-related zinc finger proteins like H-plk, or its splice variant ZNF13, with KRAB (Krüppel associated box) domains, often function as repressors of transcription. The DNA binding zinc finger region seems to be conserved and most likely translated (Åbrink et al., 1998). The demethylation of ERV3 in U-937 cells, preceding ERV3 env and H-plk transcripts, could imply that these putative proteins regulate haematopoietic and monocyte development. Finally, the specific regulation of the ERV3 locus in a cell type with monocytic phenotype suggests that ERV3 could be involved in, or regulating differentiation, cell fusion and immunomodulation including immunosuppression.

ERV3 in pathological conditions

ERV3 has been investigated in relation to different pathological conditions such as, congenital heart block (CHB), different skin disorders, MS, rheumatoid arthritis and different tumours (Andersson et al., 1996a; Katsumata et al., 1998; Katsumata et al., 1999; Li et al., 1996; Rasmussen et al., 1996; Takeuchi et al., 1995). Most of these are mRNA expression studies and are still inconclusive. Using the bacterially expressed TM protein of ERV3 together with anti ERV3 TM antibodies in an ELISA, Li and co-workers reported that mothers giving birth to children with CHB have higher titre of anti-ERV3 Env antibodies than mothers giving birth to healthy children (Li et al., 1996). The hypothesis that ERV3 expression in foetal heart could induce maternal autoantibodies in CHB disease was further analysed in a special case study. A woman suffering from systemic lupus erythematos (SLE) with two consecutive pregnancies, of which one involved CHB, was analysed for presence of four different autoantibodies including anti ERV3 TM (Horsfall et al., 1998). Together with the autoantibody against the 60kD Ro-antigen, the ERV3 Env titre was higher in the case with CHB compared to the healthy one. It was speculated if the mothers were homozygous for the stop-codon of ERV3 env, the full-length variant could be immunogenic if expressed from the paternal gene in the case of CHB. This could however be discarded in a follow-up study by de Parseval, who showed that none of the 12 investigated mothers was homozygous for this ERV3 mutation (de Parseval et al., 1999).
Since the first report stating that ERV3 is a single-copy locus, many ERV3 related sequences have been found (Benit et al., 2001; Kjellman et al., 1995; Tristem, 2000). In spite of this, ERV3 is still often described as a single-copy gene. In year 2000, Kim et al. mapped ERV3 to a single locus on chromosome 7q11.2 by radiation hybrid mapping, and in 2001 Yuan used ERV3 as a quantitative cell marker in real-time PCR (Kim et al., 2000; Yuan et al., 2001). Especially the env gene, and particularly the TM sequence, is abundant among ERV3 related sequences. Although there are many sequences that resemble ERV3 in env, maximum degree of identity between ERV3 and ERV3-like sequences was 70% in env (data not shown). Several of the proviruses are defective and lack either LTRs or for examples gag genes, as described in paper IV (Figure 3C in paper IV).
THE PRESENT INVESTIGATION

Specific aims

The main purpose of the present investigation was to characterise the ERV3 env gene expression patterns and genomic distribution of ERV3-like sequences.

I To investigate if the specific expression of ERV3 env in human sebaceous glands correlates to age, sex, skin neoplasia or anatomical location.

II and V
To analyse the expression of ERV3 env during induced differentiation in the human monoblastic cell line U-937 (paper II) and subsequently to study the possible role of ERV3 in differentiation of normal and neoplastic monocytes induced by RA (paper V).

III To monitor the ERV3 and HERV-K mRNA expression in normal developing tissues.

IV To characterise ERV3 expression pattern in normal tissues regarding cell type- and tissue-specificity as well as quantified mRNA levels compared with a related HERV, HERV-E. In order to achieve this aim; to develop quantitative real-time PCRs for ERV3 and HERV-E. To investigate the genomic distribution of ERV3-like sequences in the human genome.
Materials and methods

In this thesis, several well established molecular and cell biological methods are used and described in their respective paper. Two methods are central and described in more detail.

The monocytic cell line U-937

The cell line U-937 has a central role in this thesis as a model system. The cell line was originally established in Uppsala already in 1976 (Sundström and Nilsson, 1976) and has several unique characteristics, which makes it suitable to study several, aspects of HERV-biology. The cells have a phenotype comparable to normal monoblasts. Moreover the cells can be induced in vitro to differentiate with several substances such as interferon-γ, vitamin A, vitamin D3 and T-cell derived cytokines (Öberg et al., 1993b). From the parental U-937 line a variety of sublines has now been established (Dimberg et al., 2002; Dimberg et al., 2000). During the years a considerable amount of knowledge has been established concerning regulation of U-937 cells which will allow direct correlations of HERV activity to differentiation (Öberg et al., 1993a). From previous work (others and ours) it is known that several HERVs are upregulated after treatment with differentiating substances (Johnston et al., 2001; Larsson et al., 1996). The U-937 system thus allows controlled studies of HERV expression in relation to differentiation including systems in which the HERV expression is modified. As a normal control to U-937 cells, CD 14 positive peripheral blood monocytes can be used as in this thesis, paper V.

In Situ Hybridisation (ISH)

This technique identifies cells in which a specific RNA message is produced. This can be used in combination with Northern blot analysis. The latter technique does not allow identification of cell-type specific expression in a complex tissue. This means that a tissue can be negative in a Northern blot because of few cells present expressing the transcript of interest, but positive in ISH. When using radioactive labelled probe in ISH, the positive signal is detected as black spots (silver grains) generated from radiation of the probe into the X-ray photographic film layer on the slide.

Formalin fixed and paraffin embedded tissues were used for ISH in this study. The paraffin embedded tissues were sectioned (4 µm thick) and mounted on 3-aminopropyltriethoxysilane-coated slides (Sigma, St Louis, MO). After solving the paraffin in the sections and replacing it first with xylene and then with decreasing
dilutions of ethanol, the hydrated slides were put in 70% ethanol. Sections were then pre-treated with 0.2 M HCl for ten minutes and permeabilised with 2 mg/ml Proteinase-K (Merck, Darmstadt, Germany) at 37°C for 15 min prior to hybridisation. Tissue sections were hybridised with 35S-UTP labelled RNA probe produced as described in paper III. Hybridisation continued overnight at 56°C and samples were washed in 2x standard saline citrate and 50% formamide prior to treatment with RNase A (Boehringer Mannheim, Germany, 100 mg/ml 37°C for 30 min). Application of liquid photographic emulsion (KODAK) NTB2 diluted 1:1 in distilled water was followed by exposure at 4°C for 2-4 weeks. Slides were developed and counterstained with Mayer's hematoxylin and mounted with Pertex (Histolab Products AB, Gothenburgh, Sweden). Photos are taken in a Zeiss microscope through brightfield.

Quantitative real-time PCR (QPCR)

A quantitative real-time PCR method is used in paper IV and V. This sensitive PCR technique is frequently used for clinical diagnostic in detecting for example viral copies or the BCR-ABL fusion transcript in chronic myeloid leukaemia patients (Barbany et al., 2000). It allows calculation of RNA copy-number existing in the initial sample and quantification and comparing gene expression at the mRNA level. To achieve this, the sample values are normalised to an internal control, reference such as 18srRNA, GAPDH or histone3.3 (Bustin, 2000) (Ambion Technical Bulletin 151). This technique is based on fluorescence originating from the labelled reporter-quencher probes (R-Q probes). During every amplification cycle with a positive template present, the amplification product is detected by fluorescence. The R-Q probe contains a fluorophore at the 5’ end and a quencher at the 3’ end in order to only release fluorescence when the probe is replaced during the 5’nuclease assay. When the R-Q probe is displaced during amplification of the target sequence it is separated from its quencher and fluorescence is detected.

A standard curve can be created by producing a serial dilution from a stock solution containing known copies of the sequence of interest (template). As long as the standard curve shows a correlation coefficient, an R-value, of 0.970—1.00 it is possible to read the sample value from the standard curve by plotting the Ct (cycle threshold) values as a function of the logarithm of the copy number. A problem regarding amplification of sequences that exist in our genomes is the risk of contamination from genomic DNA. Treating the RNA samples with DNase prior to cDNA synthesis decreases that risk. All subsequent cDNA reactions were performed as either RT+ or RT-, thus creating another DNA contamination control in the RT- sample.
Primers are preferentially designed after the probe, which should have 10-12°C higher melting temperature, Tm, than the primer pair. The probe is designed to anneal 5-20 nucleotides from one of the primers in the target sequence between the primer pair. All R-Q Probes used in these studies were labelled with 6-FAM as a reporter at their 5’ end and an internal dark quencher attached to a thymidine in a middle position of the probe sequence. A TaqMan® Universal PCR master mix (2x) (Applied Biosystems) were used, containing optimised buffer components, AmpliTaq Gold® DNA Polymerase, AmpErase® uracil-N-glycosylase (UNG) and dNTPs with dUTP. The enzyme UNG prevents carryover contamination from earlier PCR products by degrading double stranded dU-containing DNA. After the first 10 min hold at 95°C, the cycle temperatures in a real-time PCR usually are 95°C for 15 sec and 60°C for 1 min for 55 cycles.

Results and discussion

I  High expression of ERV3 in sebaceous glands in skin

During a screening investigation of different normal human tissues by in situ hybridisation (ISH), high levels of ERV3 mRNA were found in normal skin. This tissue was found suitable for further studies. The results were obtained by ISH (as described in material and methods, paper I) using a 1.7 kb 35S-labeled RNA probe detecting all three transcripts from the env gene. It was concluded that the hybridisation was specific, since hybridisation with a sense probe was always negative. Clear tissue- and cell specificity was also observed. Only cells from the sebaceous gland in the dermal layer of skin, or in dermoid cysts of ovaries, expressed high levels of ERV3 env mRNA.

Sebaceous glands are composed of cells forming lobules and undergoing holocrine secretion. In the periphery of the lobule the cells are undifferentiated and dividing. As the cells move towards the centre, they differentiate and undergo holocrine secretion, i. e., whole cells are disintegrated and the content is accumulating into the gland and transported to the surface of the skin. This type of secretory activity is dependant on hormones, especially androgens. Androgen receptors are present in, besides the sebaceous gland, the dermal papillae of the hair follicles and the sweat glands (Choudhry et al., 1992). However, the ERV3 env expression did not follow the androgen receptor distribution. In order to study if variations in steroid hormones reflect differences in ERV3 expression, the following parameters were chosen: sex, age and anatomical location of the glands. Selected skin biopsies, including various skin
tumours were analysed. In addition, no difference in expression levels of ERV3 mRNA related to either sex or age was demonstrated. Though the amount of glands varied in different samples, the mRNA levels were estimated to be the same in all sebaceous gland investigated. The anatomical location of the glands did not affect the levels of ERV3 expression. Sebaceous glands of dermoid cysts contained comparable high levels of viral mRNA as the sebaceous glands of the skin. Slightly elevated mRNA levels were also found in differentiated keratinocytes. None of the investigated common skin tumours such as naevi, squamous cancer or basalioma were associated with ERV3 expression, except for a single tumour derived from sebaceous glands, Nevus sebaceous. To our knowledge, no other HERV is expressed in this specific way in human skin.

It is already known that some ERVs and HERVs could be transcriptionally activated by hormones, such as sex steroids and glucocorticoids (Ono et al., 1987; Sluyser et al., 1983). These hormonal effects are probably due to hormone-responsive sequences in the LTRs. Thus, the specific transcriptional regulation of some HERVs probably depends on the LTR sequences. In the LTRs of ERV3, the regulatory sequences constitute potential responsive motifs that might be responsible for the tissue-specific expression of env-H-plk spliced transcripts. The potential functions of these proteins are still unknown. The tissue-specific expression of ERV3 transcripts could hypothetically be correlated to certain functions such as receptor interference, mainly in skin, and cell fusion and development in placenta. The sebum of the sebaceous glands is transported to the surface of the skin and thus supplies bacteriostatic and fungicidal effects as well as water protection. Since the sebum probably contains the putative env protein of ERV3, it is suggested that this protein might have a function in protection against retroviral infections. Identification and characterisation of a putative ERV3 receptor(s) is not done but highly warranted.

II Variation of expression of ERV3 in the human monoblastic cell line U-937, clone 1, during induced differentiation

As early as in 1988, Cohen and co-workers examined several human cell lines for expression of ERV3 env mRNA. The results revealed that U-937 cells were positive (Cohen et al., 1988; Kato et al., 1988) and the levels of transcripts were estimated to be 30% of that detected in placenta. U-937 cells were chosen as a model system to study variations of ERV3 expression, since this cell type is involved in immunomodulation and can differentiate after treatment with certain differentiating agents. When induction of differentiation of U-937 cells occurs, the ERV3 env mRNA transcript levels increase when the cells reach a more differentiated phenotype. The agents used to induce
differentiation in this system were Vit D3 (vitamin D3), RA (retinoic acid) and TPA (PMA; Phorbol-12-myristate-13-acetate) (see materials and methods paper II), all of which give a specific phenotype (Öberg et al., 1993b). Northern blot analysis revealed the same expression pattern regardless of the induced phenotype, where two of the three ERV3 env transcripts were seen (3.5 and 9kb). Within 4 h, ERV3 env mRNA increased slightly and it further increased after 24 h, reaching a peak at 72 h. However, a different pattern of ERV3 expression was seen during treatment with IFN-γ, which leads to activation, instead of differentiation, of the U-937 cells. Such activation induced a peak of env mRNA within 4 h, which then declined constantly from 24 to 72 h, post activation. The increased expression levels of ERV3 mRNA were confirmed by ISH on cells, differentiated with Vit D3, on cytospin slides (Fig. 6 paper II). Furthermore, polyclonal antibodies raised against a bacterially expressed fusion protein (described in material and methods paper II), recognised a 65 kD protein, as predicted based on the sequence, in U-937 cells differentiated with TPA and detected by Western blot (Venables et al., 1995). The levels of the putative ERV3 env protein correlated with the variation levels of ERV3 mRNA. Another polyclonal antibody, raised in the same way, recognised the putative ERV3 env protein also in acetone/methanol (v/v) fixed U-937 cells differentiated with TPA and detected by immunofluorescence. The difference in basal levels of ERV3 mRNA, as seen in the Northern blots, probably depends on culture conditions such as fluctuations in serum contents.

Based on these findings it can be concluded that U-937 is a good model system for further studies to explain the possible differentiation and immunomodulatory effects of SU and TM proteins encoded from ERV3 env.

III Developmental expression of ERV3 and HERV-K in human tissues

From previous studies it can be concluded that most human tissue types express ERV3 mRNA (Andersson et al., 1998a; Cohen et al., 1988) and since the elevated expression demonstrated a remarkable tissue specificity, it was proposed that ERV3 could have a potential role in differentiation (Larsson et al., 1996; Lin et al., 1999). One way to study differentiation naturally would be to examine human development in foetal tissues. In this study (paper III) we examined ERV3 env expression at mRNA level in three foetuses, and at protein level in one post-partum foetal adrenal specimen. The material used in this investigation was obtained from routine cases referred to the Department of Pathology University Hospital, Uppsala, Sweden, for histopathological analysis. The three whole foetuses that were examined by in situ hybridisation (ISH) originated from specimens sent in after acute surgery for ectopic pregnancies (tubular pregnancy). The freshly frozen adrenal gland examined by immunofluorescence (IF)
was from a case of unexpected death, ten days after birth (see material and methods, paper III).

The strongest hybridisation signal with the ERV3 *env* probe was seen in placenta and in adrenal gland tissues. The ERV3 *env* protein was also detected by IF on sections from a fresh frozen post-partum adrenal (Fig. 2G in paper III). The levels of ERV3 *env* mRNA expression were semi-quantitatively estimated after ISH and the organs were ranked in low (-) to high (+++) expressing tissues. Tissue from placenta, showing the strongest ERV3 mRNA signal, was used as positive control, and set at (+++). Tissues from the adrenal gland showed similar high levels (+++). Tissues and organs with moderate elevated expression levels (++) were: developing tongue and certain structures in the primitive brain, such as Rathke's pouch. Most of the other tissues such as heart, gut, developing kidney, bone (columna vertebralis) and liver showed low ERV3 env levels. We were unable to identify skin in the foetal material and we could thus not confirm our previous findings in adult skin (presented in paper I).

The pattern of ERV3 expression in foetal material was similar to that obtained when studying the expression of a mouse retrotransposon, virus-like 30S RNA element (VL30) (Adams et al., 1988; French and Norton, 1997). The tissues investigated for VL30 expression in mouse embryonic tissues, included male and female gonads, adrenal cortex and placenta all of which are endocrine organs that synthesise and secrete steroids (Schiff et al., 1991). To investigate if the similarities in regulation between ERV3 and VL30 could be found in their respective LTRs, a search for transcription factor binding sites was made. The program RGSiteScan was chosen, using the version that existed in July 2001 available at URL http://wwwmgs.bionet.nsc.ru/mgs/programs/yura/ReCropScanStart.html. Of 28 potential binding sites of transcription factors in ERV3 5´LTR, 21 were shared with VL30. As a consequence of the high expression of VL30 in steroidogenic tissues in mice and of ERV3 in the steroid producing adrenal gland, we expected that the binding sites for the androgen receptor (AR) could be of importance for the transcription. However, though it was the most common binding site, we also found AR sites even in a higher degree in the HERV-K 5´LTRs, despite the low expression of HERV-K. Actually, two different HERV-K LTRs were investigated (and had 13 and 15 AR sites, compared with VL30 and ERV3 that had 9 and 13, respectively. Since HERV-K transcripts (Rec and Pol/Int) were absent in all tissues except placenta, the tissue-specific expression is probably not dependent on the AR binding sites.

To see if this pattern was a general phenomenon for HERVs, the sections were also hybridised with probes specific to the HTDV/HERV-K genome. HERV-K is an
interesting HERV-family, being the most biologically active (Tönjes et al., 1996). It is also one of the youngest HERVs and is therefore probably more actively transcribed. Interestingly, HERV-K has a complex genome, like that of Lentiviruses (Löwer et al., 1993). In fact, an accessory gene, Rec (former eORF) has been characterised and found to resemble the HIV accessory gene rev (Löwer et al., 1995; Magin et al., 1999). Rec has been found to encode a nuclear localisation signal (Magin et al., 2000) and Rec is localised in the nucleolus of cells from the human teratocarcinoma derived particle producing cell line, Tera-1. The integrase gene (pol/int) and rec were investigated for expression pattern in foetal tissues by in situ hybridisation. Both the rec and the pol/int probes had low signal:noise ratio, and the hybridisation signal was difficult to separate from the negative control (hybridisation with sense probe).

Regarding HERV-K, a tissue-specific expression was seen only for Rec and only in placenta. HERV-K mRNA expression is probably expressed in most foetal tissues but at a low level. These results were consistent with Northern blot analyses where no distinct mRNA pattern could be found in foetal heart, brain, lung, liver and kidney samples (not shown). In addition, it should be stressed that both ERV3 and HERV-K expression patterns may differ during other parts of the gestational period.

**IV**

ERV3 and related sequences in humans; studies of RNA expression by real-time PCR and in situ hybridisation

As discussed in paper I and III, the expression of ERV3 env transcripts is tissue-specific with elevated levels especially in the sebaceous glands, placenta and in the adrenal gland. When we found a high amount of ERV3 env in the sebaceous glands in skin we speculated that this indicated that the receptor interference phenomenon could also exist in humans as a defence. This was before the polymorphism, causing the stop codon resulting in a truncated protein in ERV3 env, was detected. However, the data from Lin and co-workers suggest that ERV3 can induce both fusion of cells and certain differentiation-related markers, when transiently expressed in the trophoblastic cell line BeWo. They also speculated if the truncated form of ERV3 Env, at least in theory, could be functional. In addition, a similar Env protein from another related HERV locus could replace ERV3 env in persons with the truncated variant. As a consequence of this it became relevant to investigate expression of both ERV3 and ERV3-related sequences in several organs and at a quantitative level.

Initially, ISH was performed, as in paper I and III, using a $^{35}$S labelled ERV3 env riboprobe. Anonymous archive material from most organs was selected, of which all cases were older than six years. Additional tissues were included from tissues that were
found positive by ISH. We have earlier published that ERV3 env, using multiple tissue Northern (MTN) blots of normal tissues, is expressed in variable levels and in different combination of transcripts in several tissues (figure 5) (Andersson et al., 1998a). In the present study, paper IV, ISH was used to identify the specific cell types expressing ERV3, and to demonstrate tissue variation of expression. In addition to the previous results observed by Northern blot, we could demonstrate elevated levels in brown fat, pituitary gland (adeno-part) and corpus luteum.

Further questions addressed in this paper are:
Which sequences are responsible for the positive ERV3 env ISH results? Is ERV3 really a single-copy gene as reported before (Cohen et al., 1985; O’Connell et al., 1984)?

Several BLAST searches were performed to find ERV3 related sequences in Genbank. Searches based on protein sequences from SU and TM revealed many related sequences, both genomic and expressed sequence tags (ESTs). Among the similar sequences we found the related and well-documented HERV-E, which is highly related to ERV3 in the TM region (Figure 3A and B, paper IV). Several ERV3 related sequences were earlier reported by Kjellman and co-workers, who found that the majority of these were located on the Y chromosome (Kjellman et al., 1995). In this study we could confirm that many ERV3-related are on the Y chromosome, and that the accession number for the clone containing the original ERV3 (M12140.1), from chromosome 7, is AC073210.8.

To be able to monitor variations and amounts of expression, QPCRs for ERV3, HERV-E and the reference gene histone 3.3 were developed. Histone 3.3 has previously been used as such (for example (Andersson et al., 1996b), and was first reported in 1993 (Medstrand and Blomberg, 1993). Histone was chosen because of its expression independent of the cell cycle (Wu, Wells). Andersson M-L and co-workers also reported that histone works as an excellent reference for HERV expression in EST databases (Andersson et al., 2001). Based on the data from Anderssons report, we can conclude that histone and most HERV transcripts occur at a similar frequency. The SU region was chosen for designing the primer pairs and probes in order to develop a specific PCR where the two HERVs would not crossreact. An ERV3 real-time PCR has been used by Yuan et al. in 2001, as a cell quantitation marker of human cells since it is described to exist in one copy per haploid genome (Yuan et al., 2001). Although there seem to exist many ERV3-like sequences (Fig 3 and 4), their PCR primers, nt 1051-1185, seem to pick up only one proviral sequence. The ERV3 primers we designed are further down stream at nt1341-1484. The forward primer of ERV3 (nt1341-1350 in M12140.1) was placed in the polymorphic region where a C to T transition is present at
nt 1352. Nevertheless, by running the PCR reaction at 54°C, also a sequence with the stop codon will be amplified by this PCR.

QPCR results from the human total RNA master panel II from Clontech showed low expression in placenta for both ERV3 and HERV-E. Placenta, has been described as a high level expressing tissue for the investigated HERVs. Skeletal muscle showed high levels for ERV3 and HERV-E. These results are contradictory to previous ISH and Northern blot results (Andersson et al., 1998a; Cohen et al., 1988). Thus, additional samples from these specific tissues were collected from the human tumour bank, Department of Clinical Pathology, University Hospital of Uppsala, and analysed. In Figure 5 and 6 in paper IV, the additional samples are marked with the symbol ‘o’. When comparing previous results with that of the RNA panel, it can be discussed if the placental samples from the panel were taken from parts containing only decidua with no trophoblastic cells. If that was the case, it would explain the low expression of ERV3. Also the skeletal muscle showed a changed pattern of expression.

Future investigation using HERV-arrays with representative HERV genes from multiple classes will provide a comprehensive and complete analysis.

V  

ERV3 and monocytic differentiation

In paper II, we reported that the ERV3 env transcripts also follow the RA-induced differentiation of the U-937 monocytes (Larsson et al., 1996). In this paper (V) we wanted to examine the role of ERV3 in monocytic differentiation, initially by investigation of normal monocytes and their relation to RA-induced differentiation. Total RNA was isolated and reverse transcribed to generate cDNA, in the same fashion as in paper IV. QPCRs were then carried out as described (paper IV) using histone 3.3 as the reference gene.

The isolation of normal monocytes was made by selection of CD14 positive cells from Ficoll Paque Plus® separated mononuclear cells from peripheral blood. It turned out that after 72h of cultivation with RA-treatment, these cells also showed increased levels of ERV3. Thus, it can be concluded that both normal and neoplastic (U-937) monocytes express elevated levels of ERV3 env transcripts after addition of RA. These results are in line with those published by Johnston et al in late 2001. They reported that several HERVs, HERV-W, HERV-H and HERV-K, are either activated or have increased mRNA levels during differentiation or activation of peripheral monocytes and U-937 cells (Johnston et al., 2001). Instead of env, they analysed the activity of the pol gene by
measuring mRNA levels and RT-activity of culture supernatants presumably in particles. The intention of their study was to investigate how monocyte/macrophage activation and the subsequent expression of four HERV families influence inflammatory brain diseases. They suggested that the elevated levels, or activated genes, of HERVs are markers for immune activity and are not specifically associated to any distinct disease.

RA or all trans retinoic acid, ATRA, is used in the clinical situation for treatment of some leukaemia’s with especially good results in acute promyelocytic leukaemia. This in vivo differentiation process and how it is regulated has been investigated by others with U-937 as a model system (Öberg et al., 1993a; Öberg et al., 1993b). Recently, Dimberg and co-workers were able to inhibit the cell cycle arrest usually achieved during Vit D3 and RA-induced differentiation of U-937 cells. Using two different sublines of U-937, each containing a phosphorylation defective Stat1, Dimberg could show that both the serine and the tyrosine phosphorylation sites are important for differentiation related cell cycle arrest (Dimberg et al., 2002; Dimberg et al., 2000). One of these sublines, serine 727 phosphorylation-deficient Stat1 (S727A.4), together with a subline containing the empty vector pCIneo, were chosen to perform RA induced differentiation. ERV3 env mRNAs were analysed in control cells and in S727A.4 cells at 72 h of culture with or without RA. Though the S727A.4 cells did not show growth arrest, they still responded to the RA-treatment by increasing the level of ERV3 env transcripts.

The isolated monocytes express CD14 and are already by definition partly differentiated compared to untreated U-937 cells, which hold a phenotype corresponding to a monoblast in the bone marrow (undifferentiated). In spite of that, CD14 positive cells from peripheral blood are indeed considered a relevant control cell type.

Because of the few cells obtained in the isolation procedure, we did not analyse these monocytes for cell cycle arrest as we did with the cells from the U-937 sublines. Presumably, the isolated cells already had reached a cell cycle arrest, considering that they are already CD14 positive. Still, both cell types, CD14 positive and U-937, reacts to RA-treatment in the same manner. Moreover, it cannot be excluded that ERV3 is involved in monocyte differentiation from these results. RA-induced cell-cycle arrested S727A.4 cells can still express p21, usually associated with differentiation, as shown by Dimberg et al 2002, though most of the differentiation-related proteins are down-regulated (Dimberg et al., 2002). Based on these findings we suggest that ERV3 most probably is not involved in the pathway downstream Stat1.
CONCLUSIONS

The following main conclusions can be made in this thesis;

The expression of ERV3 in a tissue specific way has been documented by three techniques each supporting each other and demonstrating: size and levels of the transcripts (Northern blot), cell-type localisation (ISH), and quantitative levels of transcriptional activity (quantitative real-time PCR). Confirmation of protein expression was made on selected tissues with immunofluorescence and Western blot analysis using anti-ERV3 TM antibodies.

It is concluded that there are many ERV3-like sequences in the human genome. ERV3 has the most conserved ORFs among the newly found ERV3-related sequences. Furthermore, proline (P) and isoleucine (I) were identified as PBS in sequences other than ERV3, which has an arginine (R) PBS. HERV-R in Repbase was found to be identical to BaEV. Taken together, the former designation ERV3 (HERV-R) is proposed to be designated ERV3.

ERV3 is expressed in several organs but in elevated levels in placenta, foetal and adult adrenal gland (all layers), sebaceous glands, brown fat, the adeno-part of the pituitary gland, thymus, corpus luteum and testis.

ERV3 is expressed in relation to RA-induced differentiation in U-937 cells. The expression is not a direct effect of the malignant phenotype since also normal CD14 positive monocytes respond by increased levels after RA treatment. The cell line U-937 has been documented as a good model system for further studies of HERVs.

Real-time PCRs for histone 3.3, HERV-E and ERV3 were developed.

Some of the HERV-E multi-copy members are also expressed in most tissues but at significantly lower levels.
GENERAL DISCUSSION AND FUTURE ASPECTS

The results presented in this thesis have been discussed in connection with the presentation of the individual papers. In the following paragraphs a few comments will be made regarding important fields (basic research and clinical implications) in which increased knowledge of HERVs will be a central issue.

1. Concerns regarding xenotransplantation and gene therapy using retroviral vectors.

2. HERVs in relation to normal differentiation and different diseases.

The possibility to use xenografts (especially from pigs) to overcome the shortage of available human organs for transplantation has been an attractive alternative. However the discovery and characterisation of pig endogenous retroviruses (PERVs) -belonging to Gammaretroviruses- led to a more pessimistic view regarding this subject. Without doubt PERVs can infect human cells in vitro (Patience et al., 1997), and poses a specific concern as human tropic infectious agents following pig-to-man xenotransplantation. Additional concerns have been raised regarding the possibilities of recombinations between HERV and PERV sequences and of the cross-packaging of HERV and PERV mRNAs in infected recipients of pig xenografts. Thus, the current investigation which is a comprehensive and rigorous analysis of HERVs of the Gammaretrovirus type should provide useful knowledge also for the risk assessment of endogenous retrovirus in pig-to-man xenotransplantation. Knowledge of the genomic complexity of Gammaretrovirus HERVs, e.g. ERV3 will be critical in following analysis of possible recombinations both in the in vivo situation and in well chosen model systems of xenotransplantation.

The same holds true for gene therapy when retroviral vectors are used. A cross-packaging of ERV sequences with the vector can occur (Chong et al., 1998) (Patience et al., 1998) and might change the tropism of the vector and even induce diseases. Thus an increased knowledge of HERV structure and activity will provide information that can be used to minimise those risks.

Other important fields where increased knowledge of HERVs will be critical are differentiation, initiation and progression of different diseases. Several observations presented in this thesis merit further studies before the hypothesis that ERV3 and related sequences are of importance for fundamental processes, such as differentiation
and specialisation, can be confirmed or refused. Just to highlight a few observations regarding ERV3:

a. ERV3 specific expression in both foetal and adult adrenal glands could indicate fundamental importance of ERV3 expression for development of steroid dependent organs

b. The specific expression of ERV3 during holocrine secretion could indicate both a coupling to differentiation and to receptor interference.

c. The high ERV3 expression in steroid dependent organs such as corpus luteum, testis and syncytiotrophoblasts could mean that ERV3 expression is regulated by changes in hormonal levels.

The role of HERVs in the pathogenesis of different diseases is uncertain. However, modern technologies available today including microdissection in combination with QPCRs and different array-techniques will allow simultaneous and detailed analysis of the activity of several HERVs in different stages of human diseases.

Finally the U-937 cell line should be highlighted as an excellent model system in which new techniques can be applied to monitor HERV expression in relation to differentiation also when the activity of different HERVs is modified.
Endogena retrovirus - våra inneboende virus

Är de till nytta och/eller skada för människan?


De virus som vi redan bär på verkar inte vara farliga för oss människor. Istället kan vi se dem som spår efter urgamla infektioner som våra förfäder fick för flera miljoner år sedan. Det som är intressant med dessa “fotspår” är att en del fortfarande kan tillverka ämnen (virusproteiner) som behövs för att ett virus ska kunna förflytta sig, och således infektera en ny individ. En del forskningsresultat tyder på att de endogena virusen skulle kunna vara till nytta för oss. Efter att ha levt i symbios med “våra” virus så länge som flera miljoner år, kan det vara så att vi människor till slut dragit nytta av en del av virusets egenskaper.

I ett försök att komma närmare en lösning på vad retrovirus gör i arvsmassan hos människan har jag försökt besvara följande frågor;

1. Var finns virus i arvsmassan och hur många är de?
2. I vilka organ och vävnader är de aktiva (när tillverkar de virusämnen)? Hur mycket virusproteiner tillverkas i jämförelse med andra kända proteiner?
3. Vid vilka hälsotillstånd är de aktiva?
4. Kan svaren från ovanstående frågor ge en vägledning om de inneboende virus kan spela någon viktig roll för oss? Är de sjukdomsalstrande eller behöver vi dem för överlevnad?

För att ta reda på om virus kan starta sjukdomar måste det först finnas uppgifter om vad som är normalt. Resultat jag fått fram i avhandlingen visar bl. a. hur aktiva virus är i friska individer och i vilka kroppsorgan. Det ger en bakgrundskunskap som är viktig.
t.ex. vid eventuella framtida transplantationer. Kunskapen behövs för att kunna bedöma risken för att grisens inneboende virus kan smitta människa.
ACKNOWLEDGEMENTS

This work was carried out at the department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University Hospital. Financial support was provided by the Swedish Cancer Society, Lions Uppsala, and Selanders Foundation.

I would like to express my sincere gratitude to:

Prof. Erik Larsson, my supervisor, for your genuine enthusiasm concerning science and particularly endogenous retroviruses, for encouragement, for contributing to the other 50% of our research group, and for the many times you trusted me to represent it;

Prof. Jonas Blomberg, my part-time mentor for the last two years, for sharing your wide knowledge about retroviruses and bioinformatics; having time for discussions and for providing excellent virology seminars;

Associate Prof. Göran Andersson, my molecular biology-mentor, who, luckily, returned to Sweden, for co-authorship and good “HERV”-collaboration;

my collaborators and co-authors in alphabetical order;

Dr. Maurice Cohen, for the opportunity to visit you and your research group at Abbott Laboratories, Chicago; many thanks to you and your family;

Prof. Patrick Venables and PhD. Tracey Mitchell, for inviting me to the Kennedy Institute of Rheumatology, London, and for their hospitality, welcoming me in their home during my visits;

Prof. Dixie Mager and PhD. Patrik Medstrand for their kind permission to use the HERV-tree in this thesis (Figure 3);

all people, from the Clinic and the University, who I had the pleasure to meet during my years at the unit of Pathology, Uppsala, among others; Görel, Ulrika L, Erik B., Pernilla E Ulf T. and Micke K;

William and Viktor for a variety of computer assistance, whenever needed, William also for have being a nice room mate;
Rehné E, Gunilla T, Gunilla Å, and especially Eva B, for all sorts of help concerning administration;

Frank Bittkowski for teaching micro photographing;

Annika Hermansson and Marianne Kastemar for your expertise in cell culture, among other things;

Sten-Sture lab; Anna D, Anna H, Lina, Inger, Karolina, Pernilla M, Fredrik, Thomas, Helena JW, Helen, Karin and Simon; for Friday gatherings and good company; especially I would like to thank Prof. Kenneth Nilsson for inviting me to the haematology lab, and together with Anna Dimberg, Fredrik Öberg and Inger Karlberg, having fruitful collaborations;

the Virology seminar group at the department of Clinical Virology; Patric, Anna F, Zhihong, Göran, Amal, Nahla, Dimitrij, Widar and others for a friendly atmosphere;

former members of “hörnfamiljen” The Busch-Larsson-lab, Ingrid Backlund, Eva Wahlund, Kenneth Wester, Lila Shokohideh, Ann-Christin Melén Boudin, Katarina Engström and Levent Akyürek for immunohistochemical expertise, friendship and a lot of humour throughout the years;

Kenneth Wester, for fruitful collaborations and co-authorship, a lot of fun conversations and for becoming a dear friend during our (sometimes long) way becoming PhD students;

Ann-Christin Melén Boudin, my sectioning mentor and sincere friend, for good times in the lab, horseback riding, and great hospitality from you and your family;

the “chicks” Mia and Marie for your happy lab-mood, and Mia also for being my WW friend;

Åsa, Anna D, Göran and Thomas S for pep talks and company during late nights in the lab; Åsa also for our superb habit of having “fruktstunder”;

my room mates from “Kardborren”; Göran, Åsa, Magnus, Jessica and Lina;

the poker(?) club; Monica, Lotta, Susanne and the NY exile-members Lene and Ulrica for sharing fun and unpredictable evenings;

Carolina Rydin, Lene Uhrbom and Ulrica Westermark, all of which I first met in the lab, for dear friendship and fabulous family-dinners with a lot of bumping around in the kitchen;

Anna-Carin, Lena, and Karin, for friendship, countless hospital-lunches; and together with Dan, Peter and Erik, for a lot of fun and memorable times;
Maria and Viktoria, for “togetherness” and for keeping my dialect;

my relatives Mimmi, Ginge, Ecke, Kajsa Torbjörn, Liselotte, Lina, Ylva and Therese for pleasant family events at Tallmas and elsewhere;

the Lorant family, Gizella, Stefan, Eva, Urban, and Margit for welcoming me in their family and for taking so good care of me;

my parents Boel and Rolf for your love and encouragement; my brother Niklas for being just the brother I want; Anna-Karin, my dear sister in law, for taking good care of Niklas and Lovisa;

Tomas, with lots of love, for your support and for being the one you are.
REFERENCES


Chong, H., Starkey, W., and Vile, R. G. (1998). A replication-competent retrovirus arising from a split-function packaging cell line was generated by recombination events between the vector, one of the packaging constructs, and endogenous retroviral sequences, J Virol 72, 2663-2670.

implications for the hormonal regulation of hair growth, sebaceous glands and sweat glands, J Endocrinol 133, 467-475.


Conrad, B. (1998). In response to Murphy et al., Löwer et al., and Lan et al., Cell 95, discussion 16.


which there was congenital complete heart block and one in which the fetus was healthy,
Arthritis Rheum 41, 2079-2080.
envelope proteins possesses a transforming growth factor-beta activity, J Biol Chem 273,
4815-4818.
homology-dependent gene silencing, Nat Genet 21, 209-212.
Monocyte activation and differentiation augment human endogenous retrovirus expression:
implications for inflammatory brain diseases, Ann Neurol 50, 434-442.
Trends Genet 16, 418-420.
Kapitonov, V. V., and Jurka, J. (1999). The long terminal repeat of an endogenous retrovirus
induces alternative splicing and encodes an additional carboxy-terminal sequence in the
Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with
schizophrenia, Proc Natl Acad Sci U S A 98, 4634-4639.
retrovirus is correlated with choriocarcinoma, Int J Cancer 41, 380-385.
Kato, N., Pfeifer-Ohlsson, S., Kato, M., Larsson, E., Rydner, J., Ohlsson, R., and Cohen, M.
two of the three ERV3 mRNAs contain human cellular sequences, J Virol 61, 2182-2191.
down regulated in choriocarcinoma encode a zinc finger protein related to Kruppel, Mol
Cell Biol 10, 4401-4405.
(1998). Tissue-specific high-level expression of human endogenous retrovirus-R in the
retrovirus-R in human vascular endothelial cells, Clin Immunol 93, 75-80.
Kim, H., Crow, T. J., and Hyun, B. H. (2000). Assignment of the endogenous retrovirus HERV-
R (ERV3) to human chromosome 7q11.2 by radiation hybrid mapping, Cytogenet Cell
Genet 89, 10.
endogenous retroviruses, Gene 161, 163-170.
Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., Nomura, Y., Segawa,
causes Fukuyama-type congenital muscular dystrophy, Nature 394, 388-392.
Kowalski, P. E., Freeman, J. D., and Mager, D. L. (1999). Intergenic splicing between a HERV-
H endogenous retrovirus and two adjacent human genes, Genomics 57, 371-379.


