TITLE OF PROJECT

The Detection and Role of Human Endogenous Retrovirus K (HML-2) In Rheumatoid Arthritis

Graham Lee Freimanis BSc. (Hons.), MSc.

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for degree of Doctor of Philosophy.

January 2008

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or for any other purposes (unless previously indicated). Save for any express acknowledgements, references and/or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

Signature…….

Date ……

Abstract

Human endogenous retroviruses are the remnants of ancient retroviral infections present within our genome. These molecular fossils show similarities with present day exogenous retroviruses but act as typical Mendelian elements that are passed vertically between generations. Despite being repeatedly linked to a number of autoimmune diseases and disorders, no conclusive proof has been identified. Rheumatoid arthritis (RA) is one such disease which has been associated with an increase in HERV expression, compared to controls.

In order to elucidate a clear role for HERVs in RA pathogenesis, autoantigens implicated in disease pathogenesis were scanned for sequence homology to retroviral genes. Such epitopes would induce antibodies cross reactive with host proteins, resulting in disease. Short peptides mimicking these regions were synthesised and the prevalence of anti-HERV antibodies was determined in RA patients and disease controls. Additionally, a novel real-time Polymerase Chain Reaction (PCR) assay was developed to accurately quantify levels of HERV-K (HML-2) *gag* expression, relative to normalised levels of housekeeping gene expression. Both serological and molecular assays showed significant increases in HERV-K (HML-2) activity in RA patients compared to disease controls with CD4+ lymphocytes harbouring the highest activity.

The real-time assay was also used to determine whether factors within the synovium could modulate HERVs, resulting in their upregulation. Exogenous viral protein expression and pro-inflammatory cytokines were shown to exert a significant modulatory effect over HERV-K (HML-2) transcription. From this data, it is clear that RA patients have increased levels of HERV-K (HML-2) *gag* activity compared to controls. Despite this it is likely that factors within the synovium such as exogenous viral expression and pro-inflammatory cytokines also influence HERV-K (HML-2) transcription possibly contributing to a role of bystander activation, i.e. being influenced by external factors, rather than actively contributing to disease processes. The exact role of HERVs in RA pathology remains elusive; however this research proposes several mechanisms by which HERV-K (HML-2) may contribute to disease.

Acknowledgements

Firstly, I would like to thank my collaborating partners in taking part in this research. These include Antony Jones, Functional genomics and proteomics lab at the University of Birmingham for the provision of their microarray scanner. My thanks also go to Professor Jun-Ichi Fujisawa (Kansei Medical University, Japan), Dr Paul Murray & Dr Joanne Flavell (University of Birmingham) and Professor Susan Marriot (Baylor College of Medicine, US) for the kind donation of the viral expression plasmids used during subsequent investigations. Additionally, thanks go to Dr Wolfgang Seifarth (University of Heidelberg, Germany) and his group for their donation and assistance using the HERV microarray gene chips.

Thanks also go to the rest of the Molecular Immunology research group and members of the biomedical researchers for their help and critique, notably to Denise Roden for her assistance with several aspects of this work.

Additional thanks go to the technical support during my time as a research, including Dr Angela Williams (for all my last minute orders) and Keith Holding, Richard Smith and Ray Ransley for their invaluable technical advice ranging from strange German card games, to small rodents and other hair related trivia.

I would also like to thank my supervisory team, Dr Paul Nelson and Dr Paul Hooley for their invaluable advice and critique over the last couple of years (particularly during the write-up period).

Finally last but not least, thanks go to my family and notably my wife, Andrea for all her help and guidance, and putting up with my endless research related ranting during this period. It is to her I dedicate this work.

TABLE OF CONTENTS

CHAPTER 4

OPTIMISATION & DEVELOPMENT OF ELISA FOR DETECTION OF ANTI-HERV-K ANTIBODIES WITHIN PATIENT SAMPLES

CHAPTER 5

DEVELOPMENT AND OPTIMISATION OF A NOVEL REAL-TIME PCR ASSAY FOR THE DETECTION AND QUANTITATION OF HERV-K GAG mRNA IN RA PATIENT SAMPLES

LIST OF FIGURES AND TABLES

Chapter 1 *Page*

26 - 77

Figures

Tables

Chapter 2 *Page*

78 - 96

Figures

Tables

Chapter 3 *Page 9 -138*

Figures

Tables

Chapter 4 Page

139 - 186

Figures

Tables

Chapter 5 Page

187 - 244

Figures

Tables

Chapter 1:

Abbreviations

Acquired immunodeficiency syndrome (AIDS)
Alcohol dehydrogenase 1 C (ADH1C)
Amino acids (aa)
American College of Rheumatology (ACR)
American Rheumatism Association (ARA)
Ankylosing spondylitis (AS)
Antigen-presenting cells (APC)
Anti-citrullinated cyclic peptide (anti-CCP)
Autoantibodies (auAbs)
Avian leukosis virus (ALV)
Baboon endogenous virus (BaEV)
B cell activating factor (BAFF)
B cell receptor (BCR)
Bovine leukaemia virus (BLV)
B lymphocyte stimulator (BlyS)
Caprine arthritis encephalitis virus (CAEV)
Cartilage - pannus junction (CPJ)
Disease modifying anti-rheumatic drugs (DMARDS)
Epitope spreading (ES)
Epstein Barr virus (EBV)
Fibroblast-like synoviocytes (FLS)
FLICE-inhibitory proteins (FLIP)
Follicular dendritic cell (FDC)
Group specific antigen (Gag)
Germinal centres (GC)
Glucose-6-phosphoisomerase (GPI)
Human immunodeficiency virus (HIV)
Human T-cell lymphotropic virus (HTLV)
Human leukocyte antigen (HLA)
Type-1 diabetes (T1D)
Integrase (IN)
Interferon-alpha (IFN-α)
Interferon-gamma (IFN-γ)
Interleukin-1 (IL-1)
Interleukin-2 receptor (IL-2r)
International Committee on the Taxonomy of Viruses (ICTV)
Intracellular adhesion molecule-1 (ICAM-1)
Jaagsiekte sheep retrovirus (JSRV)
Juvenile rheumatoid arthritis (JRA)
Koala retrovirus (KoRV)
Long terminal repeat (LTR)
Major histocompatibility complex (MHC)
Matrix (MA)
Mitogen-activated protein kinase (MAPK)
Matrix metalloproteinases (MMPs)
Mouse mammary tumour virus (MMTV)
Murine leukaemia virus (MLV)

Natural killer (NK)
Nucleotides (nt)
Open reading frames (ORF)
Oral contraceptive pill (OCP)
Peripheral blood mononuclear cells (PBMC's)
Phosphatase and tensin homolog (PTEN)
Protease (Pro)
Psoriatic arthritis (PA)
Rabbit endogenous lentivirus type-K (RELIK)
Receptor activator for nuclear factor κB ligand (RANKL)
Regulatory T cell (T_{Regs})
Reverse transcriptase (RT)
Rheumatoid arthritis (RA)
Rheumatoid arthritis synovial fibroblast (RASF)
Rheumatoid factor (RF)
Rous sarcoma virus (RSV)
Splice donor (SD)
Superantigen (sAg)
Surface protein (SU)
Systemic lupus erythematosus (SLE)
T helper cell subset/response (T_H)
Toll like receptors (TLR)
Transacting responsive sequence (TAR)
Transmembrane protein (TM)
Tumour necrosis factor – alpha (TNF-a)
Variable segment of heavy chain (V_H)

Chapter 1: Retroviruses and Rheumatoid Arthritis

1.1. Retroviruses & HERVs

1.1.1. Retroviruses

The retroviruses comprise a diverse range of viruses of vertebrates, united by a mechanism of replication comprising two unique features (Gifford *et al*., 2005). Firstly, their life cycle involves the integration of a DNA copy of the viral RNA genome (provirus) into the chromosomal DNA of their host (Figure 1.2). The second unusual feature of retroviral replication is that they have the ability to reverse the usual direction of information flow in biological systems, i.e. from DNA to RNA to protein. Consequently, the family is so named for its possession of the unique enzyme responsible for this reversal - 'reverse transcriptase' (Latin, *retro* = backwards).

1.1.2. Retroviral structure

The Retroviridae family is split into seven genera (Table 1.3) of which just two contain members that cause disease in humans - the BLV-HTLV retroviruses (e.g. Human T-cell lymphotropic virus) and the Lentiviruses (e.g. Human immunodeficiency virus - HIV) (Colmegna and Garry, 2006). Typically, retroviruses are 80-100 nm in diameter with a genome of 7-13 kilobases (kb). They are enveloped viruses, with a diploid positive strand RNA genome and a life cycle characteristically mediated by a DNA intermediate. This intervening step is attributed to the presence of an RNA-dependent DNA polymerase called 'Reverse transcriptase' (RT) which uses a single stranded RNA template to produce a double stranded DNA copy (Temin, 1962). All retroviruses can be categorised based upon their genome structure, which is either simple or complex (Figure 1.1). Simple retroviruses carry just four core genes, with complex viruses containing the core genes, in addition to a number of accessory genes, responsible for aiding and facilitating replication (Coffin J. M. *et al*., 1999). The roles and functions of the core retroviral proteins observed in simple and complex retroviruses are discussed below.

1.1.2.1. Core retroviral proteins: Gag (Group-associated antigen)

The group antigen is so called as antibodies produced against the Gag proteins cross-react with related retroviruses, of same species, but not others. Gag also acts as the precursor to the internal structural proteins of all retroviruses, and is initially synthesised as an immature polyprotein prior to being cleaved by the viral protease to produce three distinct proteins.

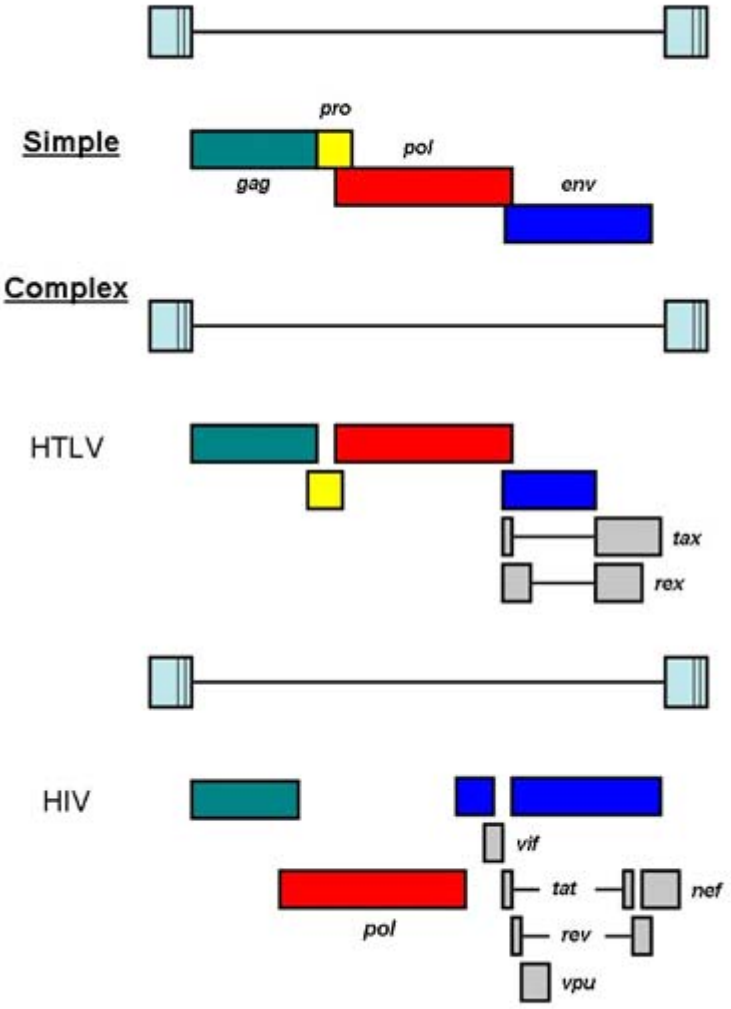


Figure 1.1 The genome structure of both simple & complex retroviruses [amended from (Coffin *et al*., 1999)]. Retrovirus genomes whether simple or complex share 3 major open reading frames. These genes code for the Gag, Pol and Env proteins. Complex retroviruses utilise mechanisms such as frame shifting and splicing to produce multiple accessory proteins from different reading frames in order to minimise their genome size, thus minimising the number of mutations potentially made by the error-prone reverse transcriptase enzyme. Accessory proteins are generally utilised to facilitate and regulate viral replication.

All Gag proteins share similar basic organisation from their amino to their carboxyl terminus: (NH2) – MA – X – CA – NC – Y – (COOH) with the minimal Gag unit being MA – CA – NC (Matrix – Capsid – Nucleocapsid). The Gag polyprotein in infectious virus self-assembles at the cell membrane to form immature virions. During maturation it is proteolytically cleaved by the protease into the Matrix, Capsid, Nucleocapsid and several low mass cleavage products (including p1, p2 and p6). In HIV-1 the matrix shell comprises of approximately 2000 copies of Ma protein which lies under the inner surface of the lipid membrane. The CA is the largest of the three domains (between 200-270aa in size) and the most antigenic, containing the highly conserved 'major homology region'. The virus's conical core, located at the centre of the particle is composed of ~2000 CA proteins (capsomeres). The core contains the viral genetic material which is in the form of 2 single strands of genomic RNA (diploid) that are stabilised through binding with the nucleocapsid proteins as a ribonucleoprotein complex in HIV. The nucleocapsid is a small basic protein (size 60-90aa) which displays a high affinity for RNA and may play a role in helping to form and stabilise the RNA dimer formation of the diploid retroviral genome (Fu *et al*., 1994). It has also been suggested to stimulate RT through facilitation of the binding of the primer tRNA to the primer binding site and promoting strand transfer, in addition to playing a central role in assembly and budding through the presence of 'assembly domains' (Wills and Craven, 1991). Morphologically the inner core has an electron dense, icosahedral vase or cone shaped appearance in HIV.

1.1.2.2. Core retroviral proteins: Pol (Polymerase)

The *pol* gene is the most conserved of all retroviral genes and one that is essential for replication (Xiong and Eickbush, 1990). It has two coding domains, leading it to act as a precursor for both the Reverse Transcriptase (RT) and Integrase (IN) proteins. During the early 1970's the controversial discovery of the RT enzyme revolutionised the field of Retrovirology (Baltimore, 1970). Being an RNA dependent DNA polymerase with RNaseH properties, it allows the degradation of DNA-RNA hybrid templates and is essential in ensuring the synthesis of DNA complementary to genomic viral RNA template (Temin and Mizutani, 1970). The reverse transcription step occurs within the capsid at a stage shortly after viral entry into the cell, prior to entry into the nucleus. In HIV, RT is initially packaged into particles as a Gag-Pol precursor, with proteolytic cleavage initially producing a homodimer of two p66

molecules. Subsequent proteolytic excision of one RNaseH domain results in the mature p66-p51 heterodimer (Turner and Summers, 1999). The IN is the second replication enzyme and is primarily concerned with the insertion of the provirus into the host cell's chromosomal DNA. The Protease (Pro) is an additional enzyme encoded between the *gag* and *pol* genes which acts late in assembly of the viral life cycle proteolytically cleaving the precursor polyprotein. PR is generally translated as a Gag-Pol fusion product produced by a ribosomal frameshift (Jacks et al, 1988). The enzyme is released from the polyprotein by an autocatalytic mechanism (Farmerie et al, 1987).

1.1.2.3. Core retroviral proteins: Env (Envelope)

The *env* gene encodes a polyprotein that is cleaved to form two components of the viral envelope - the Surface protein (SU) and the Transmembrane protein (TM). Whilst the SU is primarily responsible for receptor binding and anchoring particles to the surface of cells, the TM is thought to be linked with cellular fusion and entry. Consequently, both play central roles in determining cell and receptor tropism, e.g. CD4 and HIV. The TM consists of an N-terminal ectodomain, a transmembrane domain, and a C-terminal intra viral segment that interacts with the matrix below the lipid membrane and may play a role in fusion during both budding and cell entry, through this close association with the MA (Brody and Kern, 2004). The presence of the *env* gene distinguishes retroviruses from retrotransposons and other classes of retroelements (Mills *et al*, 2007). Figure 1.2 shows the distribution of these proteins in a typical exogenous retrovirus particle.

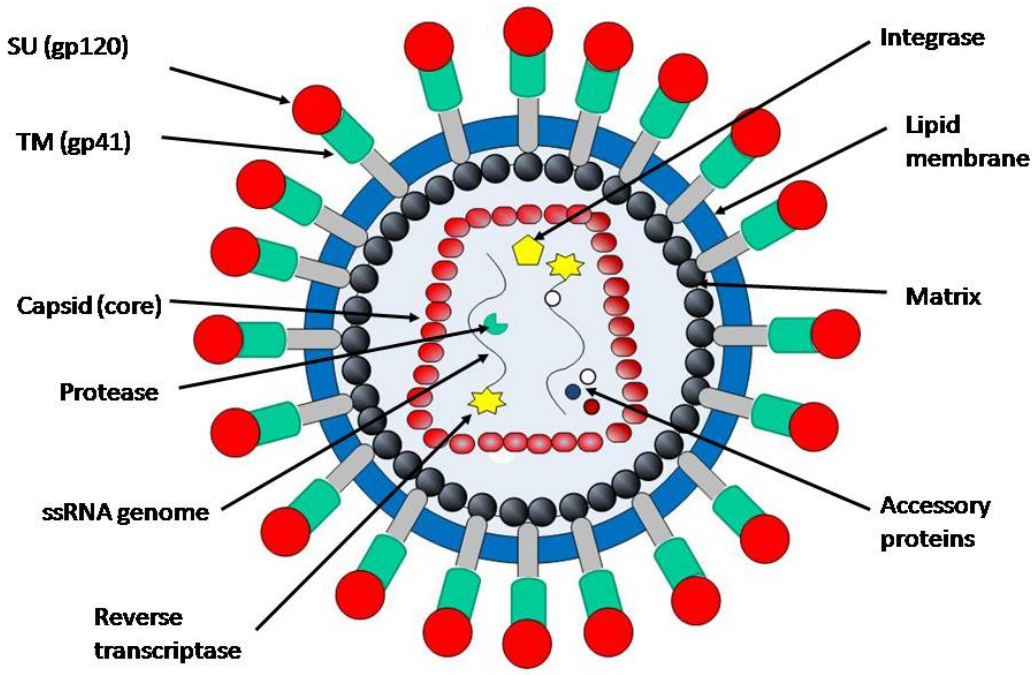


Figure 1.2 Typical structural organisation of an exogenous retrovirus particle using HIV as an example.

1.1.2.4. Other proteins

The majority of endogenous retroviruses possess a simple genome configuration, although some members of the HERV-K family possess genes that show homology to exogenous accessory proteins (Armbruester *et al*., 2002, Mayer *et al*., 2004). Using HIV as an example other accessory proteins are predominantly concerned with regulation of replication.

Tat is a regulatory protein which plays a role in stimulating both structural and replication proteins (including itself). In order to exert its effect, Tat depends upon a short sequence of nucleotides called 'TAR' (Tat responsive sequence) which can be located at the start of the viral genome and is included in all mRNA transcripts of the HIV viral genome. Whereas Tat is able to increase the production of all viral proteins, a second regulatory protein called Rev enables the integrated virus to selectively increase either regulatory or structural components of the virus. Rev also acts along two other pathways. Firstly, it is able to selectively prevent transcripts (including its own) from being translated into proteins. Secondly it is able to interact with proteins

in order to override the repressive pathways effects. Rev participates in sequence specific transport of unspliced and incompletely spliced vRNAs from the nucleus to the cytoplasm. It interacts specifically to the 'Rev response element' (RRE), which is located in unspliced viral transcripts of *env* (Malim *et al.*, 1990). Once replication occurs, interactions between Tat and Rev hold viral growth in check.

The two pathways interact with one another with Tat increasing both itself and Rev, whilst Rev acts upon itself and Tat by binding to RRE, thereby favouring the replication of structural proteins.

When expressed, HIV Nef is a 27 kDa regulatory factor that is expressed at high concentrations shortly after infection (Goldsmith *et al*., 1995) and is thought to have at least two main roles within the replication cycle, including the enhancement of viral replication and down regulation of CD4 expression upon the infected cells surface (Salghetti *et al.,* 1995). This decrease in CD4 is thought to occur through interaction of Nef and the cytoplasmic tail of the CD4 receptor. This reduction in CD4 may indirectly enhance viral replication by reducing the chances of premature Env binding to CD4 and thus being prevented from entering the immature viral particle. A downregulation of CD4 has also been suggested to prevent reinfection of nascent progeny virions after budding (Mangasarian & Trono, 1997).

The virion infectivity factor (Vif) appears to enhance the infectivity of viral particles once they have budded from the host cell membrane to infect nearby cells. HIV Vif is a 23kDa viral accessory protein required for the production of infectious virus in a cell specific manner. v*if* defective viruses were observed to be unable to grow in certain 'non-permissive' cell lines, including T cells and Macrophages, compared to wild type virus (Borman *et al*., 1995) whilst permissive cells (SupT1 and Jurkat) allowed growth of the defective viruses. Heterokaryon analysis of cells fused from both permissive and non-permissive cell lines suggested that non-permissive cells did not allow replication in the absence of Vif due to the presence of a cellular transcription factor (Madani and Kabat, 1998; Simon *et al*., 1998). This transcription factor, named APOBEC3G was identified as a member of the APOBEC family of cytidine deaminases, a group of proteins which actively deaminated cytosine residues. APOBEC3G specifically induces C to U mutations in the minus strand viral DNA, resulting in G to A changes in the coding strand thus causing mutated and nonviable virions (Lou *et al*., 2007). It is thought that Vif targets APOBEC3G for polyubiquitination and degradation (Yu *et al.,* 2003).

1.1.2.5. Long terminal repeats (LTRs)

The viral protein coding genes are bracketed by two identical stretches of sequence termed the long terminal repeats (LTR), which can each be further separated into three separate domains – U3, R and U5 (Figure 1.3). The LTRs are exact copies of one another at opposing ends of the retroviral genome and play significant roles in the initiation and transcription stages of the retroviral lifecycle (Urnovitz and Murphy, 1996). The Primer Binding Site (PBS), which is located at the end of the U5 region spans ~18 nucleotides and is complementary to the 3' end of a specific cellular tRNA. This complementarity causes the tRNA to bind to this region, where it is used to prime DNA synthesis by the viral RT. Each different group of viruses shares a specific tRNA primer for this function. As different types of retroviruses use different tRNAs as primers, a taxonomy system for classifying ERVs based upon their use of tRNA primers is used (Urnovitz and Murphy, 1996). The U3 region is similar to the U5, in that it is a unique non-coding region of length 200-1200 nt which forms the 5' end of the provirus. This sequence also plays an important role in initiation and facilitation of replication via a number of binding sites for cellular transcription factors along its length, which are able to interact with a large number of host cell factors, cytokines and cellular transcription signals (Bandziulis *et al*., 1989, Maniatis *et al*., 1987, Urnovitz and Murphy, 1996) (Figure 1.5). Consequently, the U3 region has a significant influence over the cell tropism and tissue specificity of the virus and it has become apparent that dynamics between cellular factors and viral sequences are almost as important in regulation of expression as the cellular receptors used for entry (Maniatis *et al*., 1987). The U5 region is a unique non-coding region of 75-250 nt in length and forms the 3' end of the provirus.

The 5' LTR contains enhancer and promoter sequences in its U3 region to provide signals recognised by the cellular machinery for transcription initiation, whereas the 3' LTR with its U5 region provides polyadenylation signals. This DNA copy of the retrovirus then integrates itself into the host cells DNA. At this point in the viral life cycle, the provirus may be processed by cellular replication machinery in order to express viral proteins and virions.

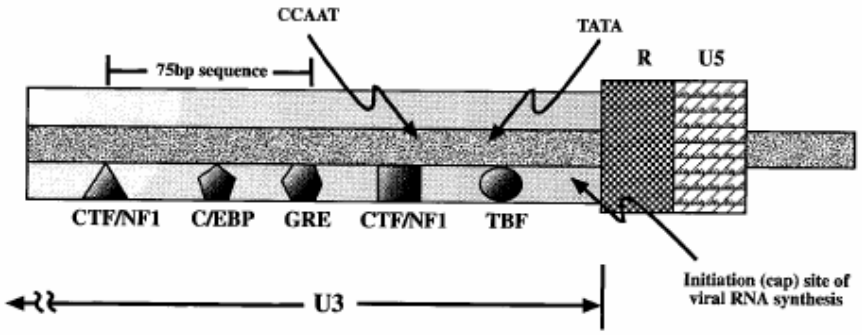


Figure 1.3 The typical structure of a retroviral LTR showing locations of transcriptional control binding sites for cellular transcription factors [diagram adapted from (Urnovitz and Murphy, 1996)]

1.1.3. Retroviral classification

The International committee on the taxonomy of viruses (ICTV) has classified the retroviridae family into seven genera, based upon both phylogenetic analysis and sequence analysis of the RT gene (Zsiros *et al*., 1999). The majority of these groups are found in piscine (epsilonretrovirinae), avian (alpharetrovirinae) and other non-primate vertebrates. Amongst those groups previously reported in humans, are the Spumavirinae, Lentivirinae and BLV-HTLV retrovirinae (formerly the Oncovirinae) (Table 1.1).

Spumaviruses are a genus of exogenous viruses that are widespread in mammals and that cause a characteristic foamy cytopathic effect when grown in primate cells. Any association with disease in humans however, is as yet unclear (Flugel *et al*., 1987). The Lentiviruses are generally involved in slow infections (those with long incubation phases), with clinical disease typically characterised by immunosuppression, e.g. HIV and Acquired immunodeficiency syndrome (AIDs) (Gifford and Tristem, 2003). The BLV-HTLV retroviruses are distinguished by their characteristic genomes and oncogenicity.

Until recently, Lentiviruses and BLTV-HTLV retroviruses were the only genera to exclusively contain exogenous retroviruses. Recently however, the discovery of the first endogenous lentivirus - Rabbit endogenous lentivirus type-K (RELIK) was reported in the European rabbit. RELIK showed evidence of *tat* and *rev* genes by sequence homology and corresponding RNA secondary structural motifs, proving that lentiviral germ-line infections can occur (Katzourakis *et al*., 2007).

1.1.4. Endogenous retroviruses: Definition & Classification

Courtesy of the recent sequencing of the human genome (Lander *et al*., 2001), we now know that just 3% is made up of protein coding sequences. The remaining 97% is made up of so-called 'junk' DNA, which may include transcriptionally active sequences (ENCODE, 2007) and ~9% of which is comprised of HERV-like sequences of the LTR element class (Lander *et al*., 2001). Endogenous retroviruses in humans were not discovered until the early 1980s (Bonner *et al*., 1982, Martin *et al*., 1981). Unfortunately, the systems of classification for both exogenous and endogenous retroviruses have since evolved separately and are poorly integrated, therefore some ERVs are clearly endogenised variants of exogenous viruses e.g. mouse mammary tumour virus (MMTV) and Jaagsiekte sheep retrovirus (JSRV) and Koala retrovirus (KoRV), all of which rather confusingly have both exogenous and endogenous forms (Salmons and Gunzburg, 1987, Palmarini *et al*., 2004, Tarlinton *et al*., 2006)

Genera		Example	Virus morphology	Genome	Endogenous form
Alpha retrovirus	Avian sarcoma & leukosis viral group	Rous Sarcoma virus	Central, spherical core "C particles"	Simple	-
Betaretrovirus	Mammalian B-type viral group	Mouse mammary tumor virus	Eccentric, spherical core "B particles"	Simple	Class II
Gammaretrovirus	Murine leukemia-related viral group	Moloney murine leukemia virus	Central, spherical core "C particles"	Simple	-
Deltaretrovirus	Human T-cell leukaemia- bovine leukaemia viral group	Human T-cell leukemia virus	Central, spherical core	Complex	Class I
Epsilon retrovirus	D-type viral group	Walleye dermal sarcoma virus	Cylindrical core "D particles"	Simple	-
Lentivirus	Lentiviruses	Human immunodeficiency virus	Cone-shaped core	Complex	-
Spumavirus	Spumaviruses	Human foamy virus	Central, spherical core	Complex	Class III

Table 1.1 Schematics of retroviral classification of the exogenous retroviridae according to the ICTV (Coffin *et al.*, 1992)

The majority of retroviruses are only able to infect somatic cells, consequently limiting infections to horizontal transmission. Rarely, instances occur however when these viruses successfully infect germ line cells, thereby colonising the host germ cells (Gifford and Tristem, 2003) allowing vertical transmission of proviral DNA to host progeny. Once the initial germ line colonisation has occurred by the founder virus, a rapid period of amplification follows before activity decreases as mutations accumulate within the open reading frames. This decline is attributable to host selectional mechanisms and the lack of a purifying selection within the retroviral coding regions. The final point at which accumulations lead to inability to express proteins is termed 'extinction' resulting in intergenic DNA. Although extinction is an inevitable occurrence for most lineages, some retain varied levels of expression for millions of years, e.g. HERV-K (Reus *et al*., 2001).

This retention of coding ability may be due to mechanisms such as complementation *in trans* (Certo *et al*., 1999, Browning *et al*., 2001, Beasley and Hu, 2002) or if the initial colonisation and subsequent amplifications conferred advantages upon the host (Mi *et al*., 2000, Andersson *et al*., 2002, Stoye, 2002). Those retroviruses that enter the germ line in this way are termed endogenous in order to distinguish them from their exogenous counterparts (Baltimore, 1975). Once in the germ line, unless insertions are disadvantageous to the host, ERVs are transmitted vertically to host progeny, increasing their frequency within the host population through mechanisms such as mutational drift and hitch-hiking (Smith and Haigh, 1974). Simultaneously ERVs may retain their capacity to replicate, resulting in further germ line insertions via retrotransposition (Cantrell *et al*., 2005) and reinfection (Belshaw *et al*., 2004), resulting in a mixed population of fixed and non-fixed elements (Gifford and Tristem, 2003). Figure 1.4 shows some of the proposed main evolutionary events which led to the insertion of the main ERV families into the primate lineage. In some species, ERVs that have both fixed and un-fixed forms include Jaagsiekte retrovirus causing lung cancer in sheep (Palmarini *et al.*, 2004), Feline leukaemia virus causing leukaemia and lymphomas in cats (Roy-Burman, 1995) and Koala endogenous retrovirus. This is of particular interest as it represents a newly emerging retrovirus that is not only threatening the species, but also in the process of undergoing endogenisation (Tarlington *et al.,* 2006).

Endogenous retroviruses have since been found in all vertebrates investigated, ranging from mice to humans (Herniou *et al*., 1998). In humans they comprise

approximately 200,000 entities, including 230 full length proviruses, around 8,100 elements containing *pol*-related sequences and 3,661 of these with full or partial open reading frames (Lander *et al*., 2001, Seifarth *et al*., 2005). Endogenous classification follows that of exogenous retroviruses, with around 31 distinct families or lineages (Gifford, 2006, Gifford and Tristem, 2003) classified into three classes (Table 1.2), based on the sequence homology of their *pol* regions with the *pol* genes of exogenous gamma-retroviruses, beta-retroviruses and spumaviruses. Whilst many ERVs are defective, a limited number retain the potential to produce viral products and indeed form viral particles (Ogasawara *et al*., 2003), although some have been reported to be active, at least to a transcriptional level, within several tissues (Stauffer *et al*., 2004).

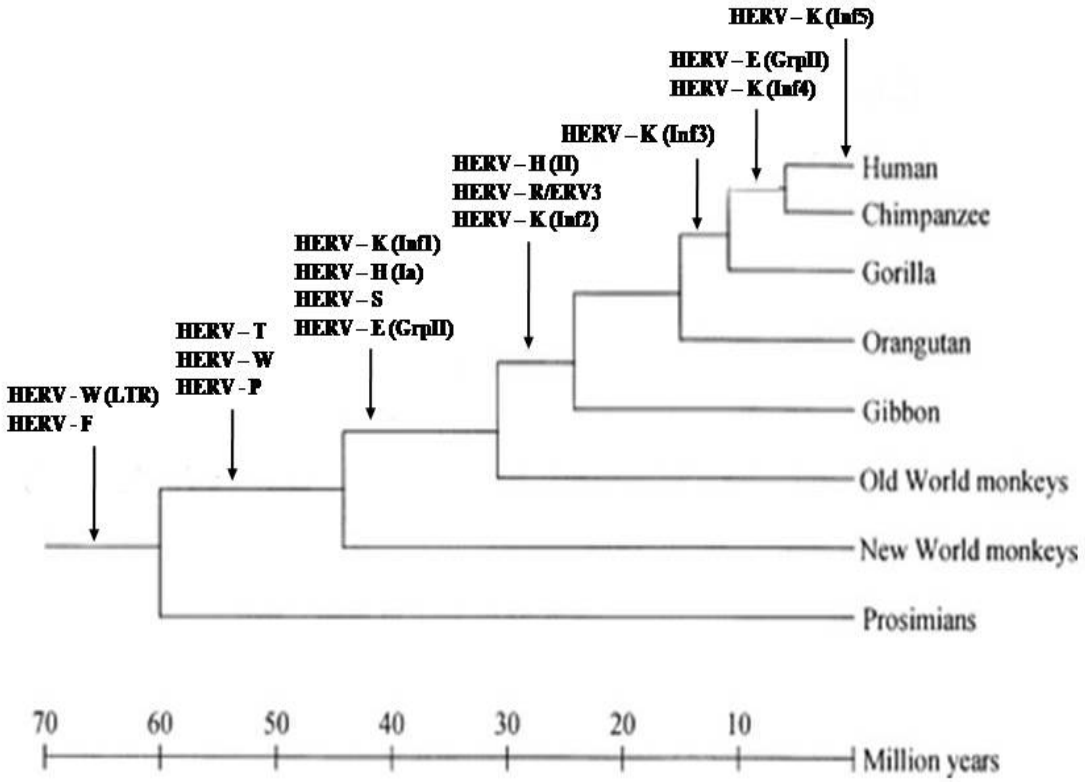


Figure 1.4 Timeline showing events resulting in the insertion of different HERV families into the primate lineage. Diagram compiled from data presented in following references: HERV-P (Yi *et al.,* 2007); HERV-F (Kjellman *et al.,* 1999); HERV-I (Lee and Kim 2006); HERV-E grp I & grp II (Yi and Kim 2007a); HERV-S (Yi *et al.,* 2004); HERV-R/ERV3 (Herve *et al.,* 2004); HERV-W (Kim *et al.,* 1999); HERV-H Ia & II (Anderssen *et al.,* 1997) & (Mager and Freeman 1995); ERV9 (Di Cristofano *et al.,* 1995); HERV-T (Yi and Kim 2007b); HERV-K (Flockerzi *et al*., 2005, Lavie *et al*., 2004, Leib-Mosch *et al*., 1993)

1.1.4.1. Class-I HERVs

Class I HERVs share sequence similarities with the gamma retroviruses such as Murine leukaemia virus (MLV). This group of endogenous viruses make up the largest class with approximately twenty members (Tristem, 2000), with the majority exhibiting genomes of simple organisation (with the exception of HERV-H and HERV-FRD which lack *env* and possess an additional *gag* respectively). Viruses of this class also exhibit varying copy numbers ranging from one (ERV3) to several hundred, although this is dependent upon the family and selection criteria used (Gifford and Tristem, 2003). Almost all class-I elements are highly defective although several families do show evidence of intact, or nearly intact open reading frames, including HERV-W (Blond *et al*., 1999, Blond *et al*., 2000), HERV-H (Lindeskog *et al*., 1998, Lindeskog *et al*., 1999) and HERV-F (Kjellman *et al*., 1999). Three families within this class show homology with MLV and Baboon endogenous virus (BaEV) in the highly conserved *pol*, *gag* and *env* regions.

1.1.4.2. Class-II HERVs

Often referred to as the 'HERV-K superfamily' Class-II elements, show the greatest sequence homology to Beta-retroviruses such as MMTV, although not all members possess a $tRNA^{lys}$ PBS (Lysine primer binding site – Section 1.1.2.5) (Tristem, 2000). All class II retroviruses (exogenous and endogenous) produce a Gag-Pol polyprotein via one or two ribosomal frameshifting sites rather than the via a stop codon as commonly observed in other retroviruses (Petropoulus, 1997). Additionally many of these viruses share a short glycine-rich region related to the G-patch domain found in many RNA-binding proteins (Aravind and Koonin, 1999). They were initially identified through hybridisation studies using the *pol* gene of MMTV and Syrian hamster A-type particles (Deen and Sweet, 1986, Ono *et al*., 1986). The complete sequence of one of these clones has since become the prototypic sequence for the HERV-K family although it lacks a 290 bp fragment from the *env* gene due to occurrence of a stop codon due to a C to T substitution (Ono *et al*., 1986). Subsequent studies using cross-hybridisation identified a further 9 subgroups within the HERV-K family termed HERV-K (NMWV1-9) (Franklin *et al*., 1988). This was followed up by a further study using degenerate primers to HERV-K10 *pol*, which identified 6

subgroups termed HERV-K (HML1-6). HERV-K (HML2) contains the prototypic sequence for the group - HERV-K10 (Medstrand and Blomberg, 1993). The subfamily has attracted recent attention for its increased numbers of insertional polymorphisms most likely resulting from recent transpositional activity (Belshaw et al., 2005). Proviruses from the HTDV/HERV-K (clone K10) (Boller *et al*., 1997) and from the HERV-K(C4) (Tassabehji *et al*., 1994) family have been fully sequenced.

Class	Examples	Organisation	Disease association
I (Gamma/ C-type)	HERV-H	LTR-*gag*-*Δpol*-LTR LTR-*gag*-*pol*-*env*-LTR	Cancer
	HERV-F	-	-
	HERV-W	LTR-*gag*-*pol*-*env*-LTR	MS, Schizophrenia, Psoriasis, Cancer
	HERV-R (ERV9)	LTR-*part gag*-*part pol*-*env*-LTR	RA
	HERV-P (HuERS-P, HuRRS-P)	LTR-*gag*-*pol*-*env*-LTR	-
	HERV-E (4-1, ERVA, NP-2)	LTR-*gag*-*pol*-*env*-LTR	Cancer, SLE, MCTD
	RRHERV-I (HERV 15, ERV3)	LTR-*Δpol*-*env*-LTR	CHB, RA
	HERV-T (S71, CRTK1, CRTK6)	*Gag*-*Δpol*-LTR	-
	HERV-I (RTVL-1)	LTR-*gag*-*pol*-*env*-LTR	-
	HERV-IP-T47D (ERV-FTD)	-	-
II (Beta, A, B, & D)	HERV-K (HML 1.1)	LTR-*gag*-*pol*-*env*-LTR	Seminoma, CML
	HERV-K10 (HML2)	LTR-*gag*-*pol*-*env*-*rec*-*np9*-LTR	Testicular cancer/ RA
	HERV-K18	LTR-*gag*-*pol*-*env*-*SAG*-LTR	JRA, Diabetes type 1
	HERV-K-T47D (HML4)	LTR-*gag*-*pol*-*env*-LTR	-
	HERV-K-NMWV2 (HML-5)	-	-
	HERV-K (HML-6)	LTR-*gag*-*pol*-*env*-LTR	-
	HERV-K-NMWV7) (HML-7)	-	-
	HERV-K-NMWV3 (HML-8)	-	-
	HERV-K-NMWV9 (HML-9)	-	-
	HERV-K C4 (HML-10)	LTR-*gag*-*pol*-*env*-LTR	-
	HERV-K113	LTR-*gag*-*pol*-*env*-LTR	SS, MS
	HERV-K115	LTR-*gag*-*pol*-*env*-LTR	-
III (Spuma)	HERV-L	LTR-*?gag*-*pol*-*?env*-LTR LTR-*gag*-*pol*-*DUT*-LTR	RA
	HERV-S (HERV18)	-	-
	HERV-U	-	-
	HERV-U3	-	-

Table 1.2 Classification of human endogenous retroviruses, coupled with coding ability of those discovered and disease associations from within the literature. Δ signifies entire gene sequenced but contains obvious deletions. ? Indicates partial sequence. SAG: Superantigen; RA: Rheumatoid Arthritis; SS: Sjogrens syndrome; MS: Multiple sclerosis; CML: Chronic myeloid leukemia;

1.1.4.3. Class-III HERVs

Class-III HERVs were originally described on the basis of sequence similarity to the Spumaretroviruses (Cordonnier *et al*., 1995). Only four families were assigned to this class (Tristem, 2000, Benit *et al*., 2001) with the HERV-L family being the largest, with a copy number of 200-500, followed closely by HERV-H (Gifford and Tristem, 2003).

1.1.5. HERV expression within humans

All human cell and tissue types show a baseline of HERV transcriptional activity (Krieg *et al*., 1992, Medstrand *et al*., 1992, Brodsky *et al*., 1993). However evidence has been building that specific families of HERVs are differentially regulated by different tissue and cell types (Sjottem *et al*., 1996, Anderssen *et al*., 1997, Baust *et al*., 2001, Schon *et al*., 2001). Class I endogenous retroviruses appear more transcriptionally active in the skin, uterus and cervix than Class II HERVs, that exhibit higher levels of transcription in Peripheral Blood Mononuclear cells (PBMC's), brain and mammary glands (Seifarth *et al*., 2005). This differential expression may be due to the presence of different cellular transcription factors. This is supported by evidence showing binding sites for such cellular factors along the length of the U3 LTR region. These sites have been suggested to play central roles in the initiation, efficiency and regulation of retroviral replication (La Mantia *et al*., 1992, Knossl *et al*., 1999, de Parseval *et al*., 1999a, Lee *et al*., 2003b). HERV transcriptional activity appears elevated in those tissues or cell types with higher rates of proliferation or metabolic activity, than observed in highly specialised non-proliferative cells, e.g. muscle cells (Seifarth *et al*., 2005). This variation emphasises the importance of cellular transcription factors and their requirement in retroviral replication.

Class I and II HERVs are the most active with a large amount of reported evidence that activity correlates with age. Class II HERVs in a number of studies have shown lower levels of activity. This has been suggested to be the result of their elimination from gene-rich regions via purifying selection through evolution (Landry *et al*., 2002). The fact that the HERV-K family are reported to be the most recent integrations into the human genome is reflected in their activity relative to other

HERVs. Distribution across the genome of HERVs was considered non-random with chromosomes 4, 19, X and Y having the highest number of HERVs with maximal concentrations identified on the Y chromosome (Villesen *et al*., 2004). HERVs were also reported to preferentially migrate across the genome to regions of chromosomes with increased levels of heterochromatin, thereby increasing their chances of being retained within the host genome (Kjellman *et al*., 1995).

1.1.6. HERVs and Host genomic evolution

Despite the fact that ~45% of the human genome is made up of millions of retroelements (Lander *et al*., 2001), evidence is only now beginning to hint at the specific roles for these sequences within the genetic evolution and function of their host.

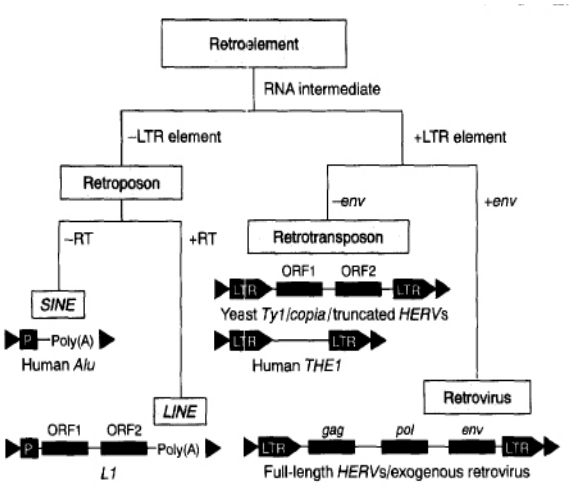


Figure 1.5 Structural classification of the retroelements. Triangles represent the short direct repeats flanking the retroelements. Internal promoters, LTR elements, open reading frames (ORF), genes and poly (A) tails are indicated. Diagram taken from Andersson *et al.* 1998.

The term retroelements is inclusive of transposons, retrotransposons and the LTR retroelements (Figure 1.5) which are all characterised by their ability to generate multiple DNA copies that can be integrated into new chromosomal positions. New evidence now suggests however, that integration no longer appears the random process it was once assumed to be with groups of retroelements exhibiting distinct patterns of distribution across the genome. Exogenous retroviruses have been shown to have a varying preference for integration sites with MLV exhibiting preferences for promoter regions and CpG islands (commonly associated with promoters in humans) (Mitchell *et al*., 2004) whilst Avian leukosis virus (ALV), avoids such sites (Weidhaas *et al*., 2000).

Similarly, complex integration patterns have also been reported for endogenous LTR retroelements. Previous investigations have reported an increased propensity, with age, for HERVs to accumulate/migrate to non-gene coding AT-rich regions, such as on the Y chromosome (Kjellman *et al*., 1995). This allosomal clustering has been attributed to its recombinatorial isolation from the remainder of the autosomal genome and its increased level of heterochromatin (Junakovic *et al*., 1998). Younger HERV families on the other hand preferentially integrate into regions of increased GC content and open chromatin although even different families have a tendency to differ significantly from one another both in prevalence in genic regions and in their orientational bias (van de Lagemaat *et al*., 2006). Additionally, HERVs have been noted to cluster in higher numbers around fragile sites, chromosomal breakpoints and recombination hotspots (Taruscio and Manuelidis, 1991, Meese *et al*., 1996, Andersson *et al*., 1998), a pattern that has also been observed in the genomes of mice (De Castro and Balagura, 1976). Class II ERVs differ from other groups as they preferentially integrate into regions of increased gene density, specifically about 5 – 20 kb upstream of gene coding sequences (Medstrand *et al*., 2002). Integrations into transcriptionally active areas of the genome such as these can increase the likelihood of host gene disruption and at these points, host selection is at its maximum, supporting the idea that insertional inactivation of genes has been strongly selected against in the lineage leading to modern humans. This is reflected in the fact that 80% of HERVs in intergenic regions are in anti-sense orientation to the gene sequence (Bushman, 2003). One example where this is observed within the human genome is the human C4 gene involved in the complement cascades (Carasso *et al*., 1976). Reports of HERVs in areas of gene transcription have thus lent credence to a new role for HERVs within the human genome i.e. one of a regulatory nature contributing to tissue specific expression. Documented examples supporting this hypothesis include Salivary amylase and HERV-E (Ting *et al*., 1992), HERV-K and *ISNL4* placental expression (Mu *et al*., 2003), a HERV LTR driving Beta1,3-galactosyltransferase beta3Gal-T5 expression (Mare and Trinchera, 2007) and an ERV9 LTR driving Alcohol dehydrogenase 1C (ADH1C) expression (Chen *et al*., 2002). Additionally, it is noteworthy that the majority of documented HERV insertions into gene coding regions, resulting in HERV driven tissue specific expression have been in those genes concerned with energy metabolism. Such evidence support a regulatory role for ERVs within the host genome (Dunn *et al*., 2003) with differential HERV expression occurring between host tissues (Seifarth *et al*., 2005).

After the initial integration of the provirus into the infected cell's genome, there follows a period of intense reintegration and reinfection in which each of the integrations has the potential to reduce host and proviral fitness; therefore subjecting the provirus to a highly purifying selectional pressure. Solitary retroviral LTRs also become abundant through this process leading to an increase in their numbers within the host genome by retrotransposition, and this can result in their influencing of host genes expression. Retroviral HERVs and LTRs may modulate cellular gene expression by a number of possible mechanisms;

1. **Alternative splicing from the 5' LTRs using cellular splice acceptors.**
2. **Initiation of cellular transcripts from 3' LTRs.**
3. **Insertion of retroelements directly into gene promoter sequence.**

A chimeric cDNA clone of a fusion transcript containing the 5' LTR sequences and leader region of a HERV-H element fused to the human c*albindin* gene (coding for a cytosolic calcium-binding protein) was identified in a cell line derived from a prostate metastasis. The HERV-H element splices to the second exon of the *calbindin* gene, introducing a different leader peptide sequence in its place, and thus altering the calcium binding motif structure of the protein (Liu and Abraham, 1991).

A HERV-K (HML-10) element was found to be involved in tissue specific expression of the insulin-like growth factor gene *INSL4* in human placenta (Bieche *et al*., 2003). In this case, the cellular gene was transcribed from the 3' LTR of a truncated provirus lacking the 3' end of the 5' LTR including the PBS. This resulted in a 10-fold upregulation of *INSL4* expression in differentiated syncytiotrophoblasts, compared to the corresponding cytotrophoblasts, suggesting a potential role in human placenta morphogenesis.

Retroelements can play a role in influencing gene regulation by inserting themselves directly into human gene promoters (Jordan *et al*., 2003). Additionally, HERVs and endogenous LTR elements have also shown the ability to be influenced by endocrine agents such as progesterone and steroidal modulation. Upregulation of HERVs in response to the presence of steroids may be particularly important in the reproduction process, particularly with the role of progesterone in the temporal regulation of the HERV-W Env protein, syncytin in the endometrium of rhesus

monkeys (Okulicz and Ace, 2003) and the development and function of the placenta (Taruscio and Mantovani, 1998). Additionally, the importance of retroviral activity within the reproductive process was eloquently illustrated by a recent investigation involving the administration of retroviral '*env'* blockers to pregnant ruminants. In this study, morphino-antisense oligonucleotides designed to inhibit the expression of enJSRV *env* mRNAs, were administered *in utero* to sheep on day 8 of pregnancy. Injection of these oligonucleotides were shown to correlate with reductions in enJSRV *env* production in the conceptus trophectoderm, which resulted in pregnancy loss by day 20, thereby illustrating the important role in placental morphogenesis ERVs may play. Authors also concluded that this role may also be observed in other mammalian species (Spencer *et al*., 2007).

A significant number of genes have been reported to have been influenced by retroviral endogenisation. Such examples include HERV-W and syncytin (Mi *et al*., 2000), apoptotic and to ubiquitination related proteins (Villesen *et al*., 2004, Campillos *et al*., 2006) and specific components of both the human immune response (van Gent *et al*., 1996, Agrawal *et al*., 1998). Retroviral influence has also led to the evolution of defenses against potential retroviral invaders. The development of the innate anti-retroviral immune response proteins of the RNA editing protein families APOBEC and Trim-5α are two recent and topical examples. Both are transcription factors that become incorporated into the nascent virions as they bud from infective cells and subsequently sabotage the retroviral genomes upon infection of the new host cells by causing promiscuous guanine-to-adenine hypermutations (Towers, 2007; Chiu and Greene, 2008). Such mechanisms have recently been reported to suppress activity of endogenous LINE-1 retro-transposition (Kinomoto *et al*., 2007). A second pathway involves Trim-5α, which acts to block retroviral replication prior to integration through specific interactions with structural components of the retroviral virion (Stremlau *et al*., 2004, Stremlau *et al*., 2006). These innate mechanisms are in addition to measures such as methylation (Yoder *et al*., 1997) and genetic mechanisms for downregulating transcription [including small interfering (si) RNA and micro (mi) RNA] (Jensen *et al*., 1999). HERV LTRs have also been identified near to genes involved in the JAK-STAT and cytokine response pathways and hence may play potential roles in their regulation and function (Vinogradova *et al*., 1997). Finally, HERV sequences in the MHC class II gene cluster were increased ten-fold in number in comparison to the remainder of the human genome (Kulski *et al*., 1999). Examples include ERV9 LTRs within the (HLA)-

DR region that may function as γ-IRE transcriptional enhancers for the *DR* genes in which they are located, (Svensson and Andersson, 1997) and the MHC class-II superantigen expressing HERV-*IDDMK$_{1,2}$22* isolated from Type-1 diabetes patients (Conrad *et al*., 1997).

HERVs at any stage may retain fitness despite their fixed mutations, through a variety of mechanisms including co-packaging or recombination between other homologous HERVs or LINE elements (Pavlicek *et al*., 2002, Gifford *et al*., 2005) or through recreation of functional hybrid genomes by gene conversion (Liao *et al*., 1997).

Of those HERVs in the genome, only 13% are intragenic, thus HERVs are generally under-represented within coding regions (Villesen *et al*., 2004) although recent findings may suggest alternative roles in influencing host gene expression for those HERVs in the 'junk DNA' (ENCODE, 2007). The majority of those genes with intragenic HERV sequences (constituting 813 genes) have been identified as being orientated in an antisense direction, again suggesting possibilities of an ancient form of antiviral defence using mechanisms such as dsRNA formation or RNA interference (Mack *et al*., 2004).

HERV gene expression can also confer a selective advantage onto its host, resulting in a purifying selection keeping the intact provirus and its reading frames. Examples of HERV coding regions that have been maintained by natural selection in the human genome are the HERV-W *env* gene (Syncytin – 1), which has led to the formation of the syncytiotrophoblast layer of the placenta (Blond *et al*., 2000, Mi *et al*., 2000, Muir *et al*., 2004). Dysregulation or disruption of the syncytin protein causes defects in the cell fusion process which can lead to development of hypertensive disorders (e.g. preeclampsia and HELLP syndrome (Villarreal, 1997). The HERV *env* encoded protein may also play a role in the immunosuppression of the maternal immune system attacking the fetus (Haraguchi *et al*., 1995, Villarreal, 1997) and in the origins of the salivary amylase gene (Meisler and Ting, 1993). There are many examples of HERV proteins possessing immunosuppressive or immunomodulatory domains including peptide domains in HERV-H *env* (Mangeney *et al*., 2001) and the HERV-E CKS–17 peptide (Cianciolo *et al*., 1985). Such immunomodulatory domains, both immunosuppressive and immunostimulatory, i.e. superantigens, have already been suggested to play central roles in the development of aspects of the immune system, such as peripheral tolerance in the thymus and the evolution of the diverse specificity of the immune response (Lewis and Wu, 1997, Agrawal *et al*., 1998, Meylan *et al*., 2005).

Expression of HERV sequences as mRNA transcripts or translated protein may also confer a selective advantage upon the host through protection against exogenous retroviruses (Larsson and Andersson, 1998), possibly acting to protect the fetus from infection in the maternal blood. Cellular resistance against infection of spleen necrosis virus has been demonstrated by transfection of cultured cells with an expression plasmid containing an HERV-W *env* insert (Ponferrada *et al*., 2003).

1.1.7. HERVs and human disease

Despite a number of studies associating HERVs with a large number of clinical conditions (Table 1.3i and 3ii) none, so far, have reported any conclusive findings, specifically implicating HERVs as the causal agents. Many of these associations exist in the form of circulating anti-retroviral antibodies and increased levels of HERV gene expression in disease samples versus controls.

Disease	HERV	Reference
Cancer		
Ovarian cancer	HERV-W HERV-E L1-retrotransposon	(Wang-Johanning *et al*., 2007) (Hu *et al*., 2006) (Menendez *et al*., 2004)
Germ cell tumour - Seminoma	HERV-K	(Rakoff-Nahoum *et al*., 2006)
Germ cell tumour - gonadoblastoma	HERV-K	(Herbst *et al*., 1999)
Gastrointestinal cancer	HERV-H	(Wentzensen *et al*., 2007)
Prostate carcinoma	HERV-E	(Wang-Johanning *et al*., 2003) (Molinaro *et al*., 2006)
Breast cancer	HERV-K	(Ejthadi *et al*., 2005)
Bladder/ Urothelial / renal cell carcinoma	LINE-1/ HERV-K	(Florl *et al*., 1999) (Jurgens *et al*., 1996)
Childhood lymphocytic leukemia	misc	(Iwabuchi *et al*., 2004)
CLL/ CML/ ALL (leukemias)	HERV-K	(Depil *et al*., 2002)
Stem cell myelopathy disorder	HERV-K	(Guasch *et al*., 2003)
Melanoma	HERV-K108 HERV-K HERV-K (MEL)	(Hirschl *et al*., 2007) (Buscher *et al*., 2006) (Buscher *et al*., 2005) (Schiavetti *et al*., 2002)
Atherosclerosis	Misc HERVs/ retroelements	(Hiltunen and Yla-Herttuala, 2003)

Table 1.3i Association of HERVs with human cancers and leukemias reported within the literature.

Autoimmune disease		
Rheumatoid arthritis	HERV-K HERV-L ERV9 ERV3/ λ4-1	(Ejtehadi *et al*., 2006, Nelson *et al*., 1999) (Seidl *et al*., 1999, Nakagawa *et al*., 1997) (Stransky *et al*., 1993, Takeuchi *et al*., 1995) (Ehlhardt *et al*., 2006)
Psoriasis	HERV-E/ K/ W/ ERV9 HERV-E	(Moles *et al*., 2005) (Bessis *et al*., 2004)
Opitz Syndrome	HERV-K/ HERV-E	(Moles *et al*., 2005)
Congenital heart block	ERV-3	(de Parseval *et al*., 1999b) (Li *et al*., 1996)
Autoimmune liver disease (Primary biliary cirrhosis)	HIAP	(Mason *et al*., 1998) (Meilof and Smeenk, 1998)
Addisons disease	LTR13	(Pani *et al*., 2002) (Gambelunghe *et al*., 2005)
Alopecia areata	HIAP	(La Placa *et al*., 2004)
Essential thrombocytopenia	HERV-K10	(Morgan and Brodsky, 2004)
Type 1 Diabetes	LTR13 HERV-K18	(Bieda *et al*., 2002) (Stoker *et al*., 1999) (Marguerat *et al*., 2004)
Juvenile rheumatoid arthritis	HERV-K	(Ogasawara *et al*., 1999)
Systemic Lupus Erythematosus	HERV-E(4-1)	(Piotrowski *et al*., 2005) (Sekigawa *et al*., 2006) (Sicat *et al*., 2005)
Sjogrens Syndrome	HIAP HERV-K113	(Garry *et al*., 1990) (Sander *et al*., 2005) (Moyes *et al*., 2005)
Mixed connective tissue disease	HERV-E (4-1)	(Hishikawa *et al*., 1997)
Multiple sclerosis	MSRV HERV-W	(Mameli *et al*., 2007) (Perron *et al*., 2005) (Antony *et al*., 2006)
Neurological		
Bipolar disorder/ Depression	HERV-K (HML-2) HERV-W	(Frank *et al*., 2005) (Weis *et al*., 2007)
Schizophrenia	HERV-K (HML-2) ERV-9	(Yolken RH, 2000) (Huang, 2006) (Weis *et al*., 2007)
Motor neuron disease	HERV-W	(Oluwole *et al*., 2007)
Miscellaneous		
Interstitial lung disease	HERV-E (4-1)	(Tamura *et al*., 1997)
Acquired immunodeficiency syndrome	Misc HERVs HERV-K	(McIntosh and Haynes, 1996) (Contreras-Galindo *et al*., 2007) (Contreras-Galindo *et al*., 2006)
Aging	Misc HERVs & retroelements	(Fuke *et al*., 2004)

Table 1.3ii Association of HERVs with human autoimmune and neurological diseases as reported previously within the literature.

1.1.8 Mechanisms by which HERVs may cause disease

HERVs may induce disease through several mechanisms including expression of superantigens and induction of cross-reacting antibodies via molecular mimicry to host proteins (Perl, 2001). These mechanisms are outlined below (Figure 1.6).

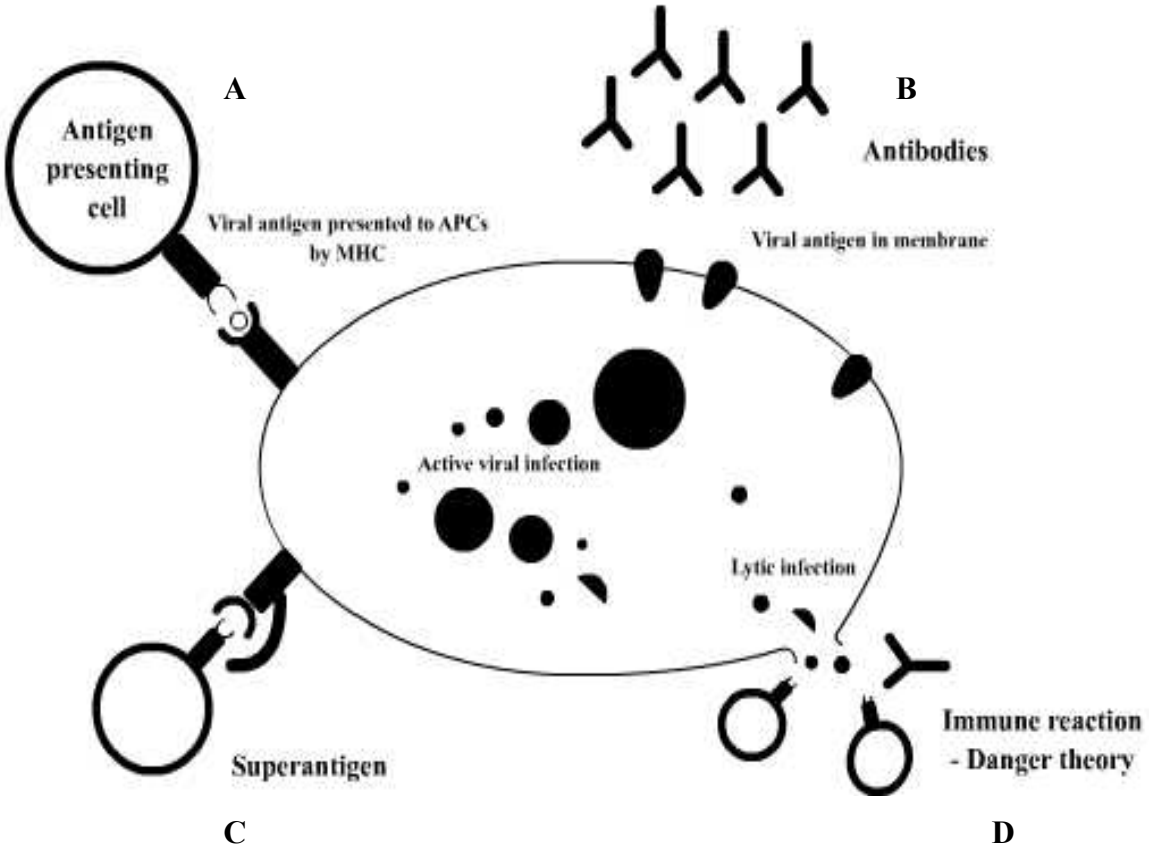


Figure 1.6 Potential mechanisms by which viruses may cause autoimmune disease. HERVs may induce disease through several mechanisms within their host. These are summarised in the above diagram and include: A) Presentation of HERV protein by APC to activate immune cells, resulting in an immune response. HERV proteins, to which the immune system may not have encountered previously may be treated as non-self, and therefore have no level of tolerance to, thereby resulting in an immune response being mounted, which may lead to disease. B) Antibodies to HERV proteins: HERV proteins may be expressed upon the surface of cells, or released from dying cells, thereby coming under immune surveillance. This may result in the initiation of an immune response in which antibdies specific for the HERV protein are produced. These may also detect HERV proteins in other cells/tissues , and target such tissues if treated as non-self or foreign. Anti-HERV specific antibodies may also come into contact with HERV proteins and bind to them forming immune complexes which then deposit within tissues, leading to activation of complement and ensuing immune reactions. Anti-HERV antibodies may also bw crossreactive with self proteins that share sequence or conformational homology with HERVs. C) Retroviruses and several bacterial species are able to bypass antigenic specificity and activate large numbers of immune cells through interaction with the hypervariable region of the TCR. This can lead to large numbers of activated lymphocytes circulating the host, increasing the changes of 'hit and run' immune responses, in addition to anergy. D) Danger theory suggests that rather than responding to exogenous signals that represent non-self, the immune system is controlled from within, responding to endogenous signals that originate from stressed or damaged cells. HERV components or proteins may be released from stressed cells and come into contact with a naïve immune response that treats components released from tissues as non-self, resulting in ensuing immune responses. APC: Antigen presenting cell; MHC: Major histocompatibility. Such a release of cellular debris may also result in the initiation of eiptope spreading.

1.1.8.1. Expression of Neo-antigens

Viral infection usually results in the production and presentation of peptides derived from viral protein, being presented to the immune system. If viral epitopes had been previously sequestered within the tissues, in immuno-privileged locations, then tolerance would not have had a chance to develop. This would result in an immune system being naïve to such epitopes and thus treating them as foreign. An immune response targeting any such epitopes within tissues would likely damage tissues, leading to disease.

1.1.8.2. Expression of super-antigens

Superantigens are proteins so far identified in retroviruses and some bacteria, that are able to bypass the normal MHC restrictive processes of T-cell stimulation resulting in the activation of 10^3 - 10^5 more T-cells than conventional antigens (Sundberg *et al*., 2002) and commonly resulting in the breakdown of peripheral tolerance and induction of clonal anergy and apoptosis (Nelson *et al*., 2004). The mechanism is based upon their capacity to bind to MHC class II proteins and Vβ chains of the T-cell receptor, regardless of antigen specificity (Huber *et al*., 1996). Such a phenomenon has frequently been observed in retroviruses (Fleischer, 1991) either in exogenous or endogenous forms, with some even suggested to be adopted into our genome through evolutionary mechanisms to aid early stages of host peripheral tolerance control (Meylan *et al*., 2005). Endogenous forms of superantigens have also been reported to be influenced by several members of the Herpesviridae, including Epstein Barr virus (EBV), which has been reported to play roles in both viral persistence (Sutkowski *et al*., 2001) and Type-1 diabetes (Conrad *et al*., 1994). Notably, the presence of EBV LMP-2A has been reported to transactivate the HERV-K18 SAg (Sutkowski *et al*., 2004).

Evidence of increased Superantigen (sAg) activity has been reported in the synovial fluid of rheumatoid arthritis patients (Paliard *et al*., 1991, Uematsu *et al*., 1991) and a HERV-K18 superantigen has been proposed as a contributory factor in the development of Juvenile Rheumatoid Arthritis (JRA) (Sicat *et al*., 2005). A HERV-K18 sAg associated with Type-1 diabetes (Conrad *et al*., 1997) and also reported as a

possible cause of Multiple Sclerosis, after HERV-W was found to bias the T-cell immune response in patients (Perron *et al*., 2001).

Viruses may also cause aberrant expression of host proteins, thus skewing the resulting immune response. An example of this is HTLV-1's ability to up regulate MHC expression on the surface of infected cells (Iwakura *et al*., 1991). This provides the 'extra boost' required for an immune event to become an autoimmune reaction in RIP-LCMV mice – an animal model for Type 1 Diabetes (Evans *et al*., 1996).

1.1.8.3. Transcriptional activation

HERVs may act as insertional mutagens or *cis*-regulatory elements causing activation, inhibition, or alternative splicing of cellular genes involved in immune function. A-type retroviruses, for example, can operate as transpositional elements and sequences within their LTRs can activate cellular genes, including those of cytokines or oncogenes. HERVs also may encode elements similar to Tat in HIV-1 and Tax in HTLV-1, both of which encode proteins capable of transactivating cellular genes. Alternatively a HERV-encoded protein may act *in trans* to regulate gene expression.

1.1.8.4. Molecular mimicry

Molecular mimicry occurs when an exogenous protein shares an amino acid sequence with that of a host component (Oldstone, 2005). One example is the association between Group A streptococcal M protein and cardiac myosin in the development of rheumatic fever, although direct evidence for this in RA is lacking (Hyrich and Inman, 2001). Viral infection results in the activation of virus specific immune cells, which may cross-react with host self-antigens which share sequence or conformational identity with the viral proteins. Despite its simplicity, such a hypothesis is deceptively difficult to prove.

1.1.8.5. Direct modulation of the host immune response

Viruses may also affect the immune response of the host directly, by nature of their cellular or tissue tropism. Viruses infecting immune cell populations can cause significant immune dysregulation, e.g. HTLV-1 (Murphy *et al*., 2004);

immunosuppression, e.g. HIV (Zandman-Goddard and Shoenfeld, 2002) and defects in cell mediated immunity, e.g. Epstein-Barr virus (EBV) (Parolini *et al*., 2000). These pathogens can also use these cells for transport to locations within the host and trigger local responses, such as in the Synovium, as observed in CAEV (Peterhans *et al*., 1999). Arthritis may be one outcome of such an impaired or abnormal immune response.

The mere presence of some viruses can also trigger immune responses of the innate immune system. Double stranded RNA (potentially from a replicating RNA virus) can initiate immune responses resulting in development of arthritis (Zare *et al*., 2004) and the release of pro-inflammatory cytokines, leading to local and systemic inflammation.

1.1.9. HERVs and their association with Rheumatoid arthritis

As with other viruses, earlier reports implicating HERVs in Rheumatoid Arthritis focus upon serological investigation. The majority of these identified anti-HIV and HTLV-1 reactive antibodies in patients (Pelton *et al*., 1988, Ziegler *et al*., 1989, Talal *et al*., 1990, Douvas *et al*., 1996), despite a lack of evidence of infection by either virus (Nelson *et al*., 1994, Nelson *et al*., 1995, Seemayer *et al*., 2002). Additionally, virus particles with retrovirus Type-C like morphology have been observed within the synovial fluid of RA patients (Stransky *et al*., 1993, Roberts *et al*., 1995).

More recently, investigations have focussed on the detection of increased levels of viral genetic material at disease sites with the development of more sensitive techniques, such as PCR. Nakagawa and colleagues used RT-PCR on synovial fluid from RA patients and healthy donors to detect increases in HERV-L, HERV-K and ERV-9 (Nakagawa *et al*., 1997). Nelson compared synovial fluid samples from RA patients against samples from healthy donors showing increases in HERV K10 compared to levels seen in controls (Nelson *et al*., 1999), with these results being confirmed by a subsequent follow up study (Davari Ejtehadi *et al*., 2005).

Other groups of HERVs such as ERV3, Lambda 4-1, HRV-5 and L1-transposons have also been implicated in RA (Takeuchi *et al*., 1995, Griffiths *et al*., 1999, Brand *et al*., 1999, Neidhart *et al*., 2000). A number of studies however have

published contradictory evidence, thereby questioning a role for HERVs in Rheumatoid arthritis (Norval *et al*., 1979, Stoye, 1999, Gaudin *et al*., 2000).

Elevated levels of HERV-K expression in patients with rheumatic diseases (including SLE, RA and SS), as opposed to healthy controls, showed increased levels of antibodies to HERV-K18 superantigen. These were predominantly in SLE patients (Herve *et al*., 2002). The HERV-K18 superantigen was originally shown to be activated by EBV infection (Sutkowski *et al*., 2001). However a similar role to that demonstrated by Herve and colleagues has been reported in Juvenile Rheumatoid Arthritis (JRA). Sicat *et al* (2005) reported significantly increased levels of K18 superantigen in JRA patients compared to controls using a semi-quantitative RT-PCR. This increase was determined to be irrespective of other factors such as EBV infection, IFN-alpha (IFN-α) levels and expansion of T cells within the peripheral blood.

Additionally, a novel endogenous retrovirus of the HERV-E family – HERV E CD5 has been identified in a small subset of B cells (B1a and B1b cells) that are involved in the production of rheumatoid factors (a disease marker for RA) (Renaudineau *et al*., 2005). Expression of the HERV-E CD5 provirus induces a second alternative cellular promoter which may result in initiation of replication shifting from exon 1 to exon 3. This leads to production of a truncated protein which loses its ability to translocate to the plasma membrane, ultimately leading to reductions in transduction of the BCR and an increase in the production of autoantibodies. HERV K (HML-2) Rec protein has also been identified within synovial tissues and cells in both RA patients and controls, although the authors concluded this was most likely not causal of the disease process (Ehlhardt *et al*., 2006). Much interest has also surrounded the finding that a HERV-K10 LTR (LTR3) was identified in the HLA class II region (Kambhu *et al*., 1990). This was investigated almost a decade later and identified as a potential disease marker for RA, being significantly more frequent in RA patients than in healthy controls (Seidl *et al*., 1999). Other studies have highlighted the potential for LTRs in the HLA-DRB1 region to play a role in disease pathology (Pascual *et al*., 2001, Pascual *et al*., 2003).

1.1.10. Other factors influencing HERVs and the development of RA

1.1.10.1. Other viral pathogens

Although this study is focussed specifically upon the potential role of HERVs in RA, one theory that is becoming more popular is that HERVs are actually influenced either directly or indirectly by the presence of exogenous viruses within the synovium. The best candidates for such a role include the Retroviruses, the Herpes viruses and Parvovirus B19 (Masuko-Hongo *et al*., 2003). Major human retroviral pathogens HIV and HTLV-I cause well characterised diseases in humans such as AIDs and T cell leukaemia respectively. HIV and HTLV-I however, also give rise to neurological complications which mimic some of the inflammatory autoimmune symptoms (e.g. HAAP and an inflammatory condition of the salivary glands resembling Sjogren's syndrome). These perhaps represent the most convincing arguments for the involvement of ERVs in autoimmune disease. A more detailed breakdown of viruses implicated in RA is shown in Table 1.2 Members from all three aforementioned viral families have been shown to interact directly with HERVs (Sutkowski *et al.,* 2004; McIntosh and Haynes 1996; Lee *et al*., 2003c).

1.1.10.2. Immune components: Inflammatory cytokines

Another aspect of autoimmune disease pathogenesis which may play a role in influencing HERVs are the inflammatory cytokines – specifically IL-1, IL-6 and TNF-α., all of which are major players in RA, in addition to being known to modulate viral expression (Walev *et al*., 1995). Increased levels of inflammatory cytokines and the oxidative conditions that they induce can also lead to somatic mutations and DNA damage. This has been shown in RA, particularly through the presence of somatic mutations in the p53 tumour suppressor gene within the synovium. This protein in particular plays an important role in surveillance of mutations within genes (Firestein *et al*., 1997, Reme *et al*., 1998, Inazuka *et al*., 2000). Additionally the presence of microsatellite instability may contribute to the expression of DNA mismatch repair enzymes – an indicator of DNA damage (Lee *et al*., 2003a). Thus, it is therefore possible that host cells infected and activated by exogenous or endogenous viruses

could migrate into the synovium releasing pro-inflammatory cytokines such as IL-6 (Mitchell, 2002) and TNF-α (Naito *et al*., 2003) or that activated synoviocytes produce cytokines (Firestein, 2003) that activate or mutate DNA leading to the activation of HERVs.

1.2. Rheumatoid arthritis

1.2.1. General Introduction

The word 'arthritis' is derived from the Greek word for 'joint' (arthron) and describes over 200 forms of arthritis affecting both young and old alike. These arthritides can be varied in presentation, being of either an inflammatory or non-inflammatory nature. The most common type of inflammatory arthritis is Rheumatoid Arthritis (RA) with a global prevalence of between 1-3% (Firestein, 2003). Other inflammatory arthritides include Ankylosing Spondylitis (AS), Gout, Psoriatic Arthritis (PA) and Systemic Lupus Erythematosus (SLE) (Klippel and Dieppe, 1997).

Although atypical symptoms of these diseases predominantly focus upon the effects of the inflammation within the joint, their overall presentation is varied and can be either local or systemic in nature (Figure 1.7).

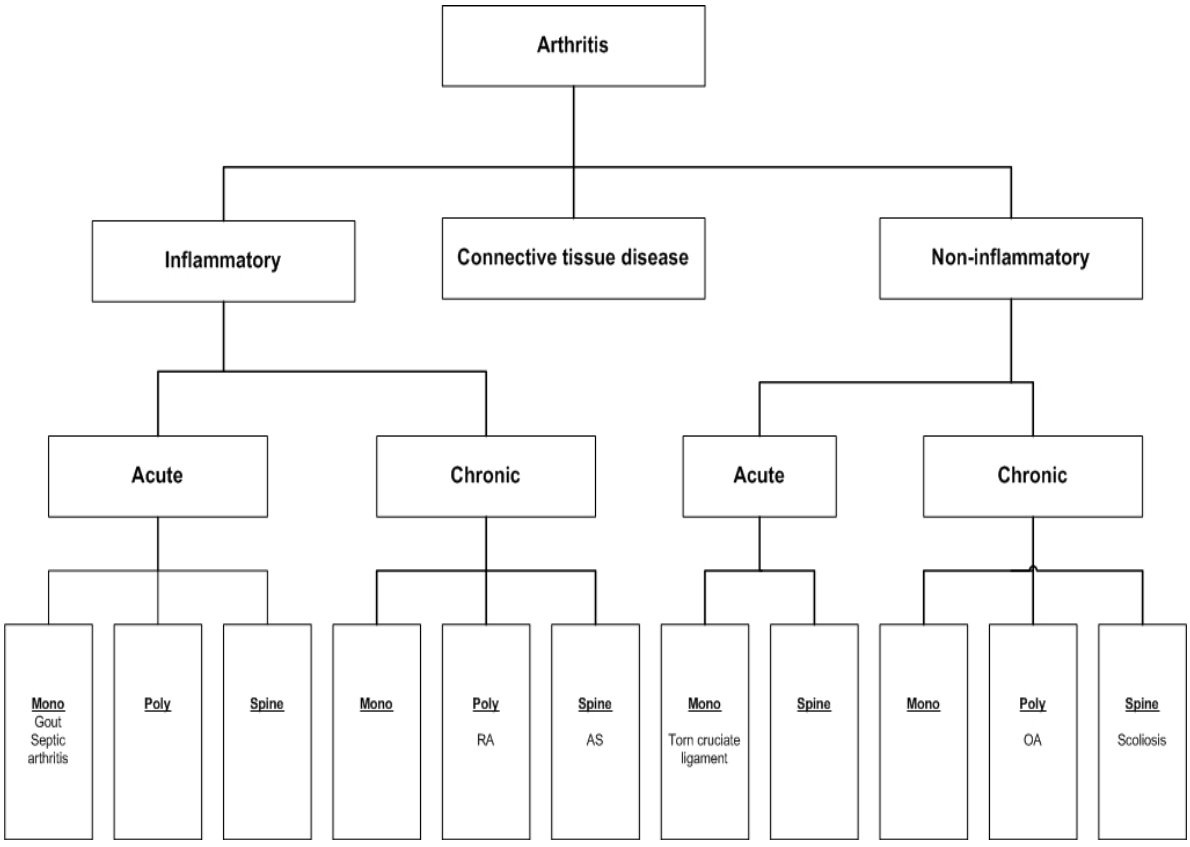


Figure 1.7 The main classification path for diagnosis of Rheumatic disease. RA: Rheumatoid Arthritis; AS: Ankylosing spondylitis; OA: Osteoarthritis; Mono: Monoarticular; Poly: Polyarticular. Taken from 'Clinical assessment of the muscularskeletal system: A handbook for medical students, Arthritis Research Campaign. 2005.

1.2.2. Inflammation

The process of inflammation represents an important physiologic response to a wide array of stimuli such as infections and tissue damage, which occurs prior to the establishment of an immune response. Despite acting independently of the ensuing immune response, the two are inextricably linked as the inflammatory process subsequently results in the infiltration of immune cells into the site of inflammation. Inflammatory responses can be localised or systemic with an acute or chronic response. The duration and intensity of a local response must be carefully regulated in order to prevent tissue damage and complement necessary tissue-repair mechanisms. In many autoimmune diseases, this is accompanied with a systemic acute phase response which typically exhibit increases in levels of acute phase proteins within the plasma, such as C-reactive protein. Either way, the inflammatory process triggers a series of non-specific events that confer protection, restricting tissue damage to the site of infection/injury (Delves and Roitt, 2000).

A localised inflammation exhibits symptoms characteristic of inflammation including swelling, redness, heat, pain and loss of function. Within minutes of injury, vascular dilation results in an increase in the volume of blood at the site, typically resulting in the symptoms of increased redness and heat. This increase in vasodilation is accompanied by increases in vascular permeability, i.e. leakage of fluid from vessels which can build up in the surrounding tissues (edema). As this leakage occurs, kinins, clotting factors and fibrinolytic systems become activated through the action of bradykinins and fibropeptides (Male *et al*., 2006).

Within hours of onset, neutrophils begin to adhere to endothelial walls of blood vessels within the vicinity of the injury site, before migrating through vessel walls into the surrounding tissue. These cells play crucial early roles in the establishment of inflammation, with numbers peaking within the first 6 hours to almost ten-fold normal levels (Pillinger and Abramson, 1995). Circulating leukocytes are coaxed from the periphery in response to the increased expression of adhesion molecules such as E and S-selectin, mediated via pro-inflammatory cytokines Interleukin-1 and Tumour Necrosis factor – alpha (IL-1 and TNF-α) (Sweeney and Firestein, 2004). Circulating neutrophils become tethered to the vascular endothelium via integrin, before migrating through the endothelial layers to the damaged tissue, following chemokine attractant gradients established at the site of injury. Upon arrival at the site of inflammation,

neutrophils upregulate increased levels of receptors for chemokines upon their surface, thereby establishing chemotaxis (Szekanecz and Koch, 2001). These can include several chemokines; complement split products, fibrinopeptides, prostaglandins and leukotrienes. Activated neutrophils express increased levels of Fc receptors for antibody/complement, thereby enhancing phagocytosis of complement/antibody coated micro-organisms. Neutrophils are also able to release reactive oxygen intermediates or reactive nitrogen intermediates in eliminating pathogens, although this also contributes to tissue injury (Corrigall and Panayi, 2002). Macrophages enter the inflamed tissues between 12-24 hours later, after following chemotactic gradients to the site. Upon arrival they become activated, thereby enhancing phagocytosis and the increased production and release of pro-inflammatory mediators and cytokines. Following activation, tissue macrophages secrete IL-1, TNF-α and IL-6, all of which propagate the inflammatory process. Subsequently, more circulating monocytes, neutrophils and lymphocytes recognise adhesion molecules and migrate through vascular endothelium into the tissue space (Szekanecz and Koch, 2000). A chronic inflammatory response may develop from an initial acute response, where antigen persists, usually resulting in a significant degree of tissue damage. This is characterised by the accumulation and activation of macrophages, which in some cases may lead to granuloma formation.

A systemic response is also able to occur simultaneously with the localised response. This leads to generalised symptoms such as induction of fever, increased hormone synthesis, e.g. hydrocortisone, increase in the production of white blood cells or the production of acute phase proteins. During such a period systemic levels of inflammatory markers such as C-reactive protein, acting to enhance complement mediated phagocytosis, may increase almost 1000-fold (Male *et al*., 2006).

1.2.3. Rheumatoid Arthritis (RA)

RA is a chronic inflammatory disease of unknown aetiology and complex multifactoral pathogenesis, affecting joints and other tissues. Global prevalence ranges between 1-3%, although some populations may have levels as high as 5% (Ferucci *et al*., 2005). As the most common inflammatory and destructive arthropathy, it has a major socio-economic impact in the UK, estimated to cost around £1.3 billion per year (Palmer and Cooper, 2006). RA is commonly associated with a group of anti-IgG autoantibodies called Rheumatoid Factors (RF) (Westwood *et al*., 2006). RA also has a

strong genetic association with certain Class II antigens, such as Human Leukocyte Antigen (HLA)-DR4, encoded within the Major Histocompatibility Complex (MHC) concerned with antigen presentation to T cells (Reviron *et al*., 2001). Although RA can develop at any age, advancing age is a strong risk factor in developing the disease with peaks occurring in the sixth to eighth decades of life (Doran *et al*., 2002). Like many other autoimmune diseases, gender also appears to play a role with a female to male patient ratio of 3 to 1 (Cutolo *et al*., 2002).

1.2.4. Pathogenesis of Rheumatoid Arthritis

Typical presentations of RA begin as a progressive, symmetrical polyarthritis in the peripheral joints. This then spreads to the larger more central joints as the disease progresses, resulting in bone damage and other systemic complications, e.g. nodules. RA is characterised by a cell-mediated immune response in the synovial joints (Figure 1.8). Most researchers agree however, that RA is most likely initiated by an antigenic/autoantigenic peptide complexed to either the rheumatoid-associated MHC Class II molecules [(HLA)-DR4 or DR1] on the surface of Antigen-Presenting Cells (APC), or to antibody leading to the formation of an immune complex (Zvaifler, 1973). Immune complexes are presented to APCs with the appropriate receptors, subsequently resulting in their activation. This leads to surface expression of activation markers, including the Interleukin-2 receptor (IL-2r), MHC Class II molecules and CD69. Expression of such markers leads to the increased production of lymphokines, including Interferon Gamma (IFN-γ) and IL-2 (Chung *et al*., 2006). Activated CD4 + T lymphocytes, through direct cell surface contact and by the release of IFN-γ, stimulate monocytes to differentiate into macrophages which play a pivotal role in disease pathology, due to numbers present within the pannus and at the Cartilage Pannus Junction (CPJ) (Chung *et al*., 2006). Macrophages also release IL-1 and TNF-α which act to stimulate mesenchymal and endothelial cells, resulting in the release of the Matrix Metalloproteinases (MMPs) such as collagenase and stromelysin, which degrade connective tissue causing joint damage (Hanaoka *et al*, 2003). IL-1 and TNF-α also act to upregulate the expression of adhesion molecules, further facilitating the recruitment of inflammatory cells into the synovial joints and thus driving the inflammatory process. IL-1 and TNF induce the Fibroblast-like Synoviocytes (FLS) to express IL-6, chemokines and MMPs, thereby further contributing to articular damage.

Additionally IL-1 mediates cartilage degradation directly by inducing MMP expression in chondrocytes. CD4+ T cells, B cells and macrophages infiltrate the synovium, leading in particular patient subsets to the formation of lymphoctyic aggregates and germinal centres (Goronzy and Weyand, 2005). The overall increase in synovial immune activity results in the activation of fibroblast-like synoviocytes in the thin acellular membrane lining the joint, causing them to proliferate producing a hyperplastic villus called a 'Pannus', which extends across the surface of the cartilage/bone assuming an almost tumour-like phenotype (Firestein, 1999). These cells also contribute to monocyte migration by increasing chemokine and IFN-γ expression (van Boekel *et al*, 2002). Where autoimmune and inflammatory mechanisms account for many of the features of RA, the Pannus is primarily responsible for the radiologically detected damage of the joint. This joint destruction may also be induced through the deposition of immune complexes upon the surface of the cartilage.

These act as points of binding for neutrophils and complement, which release proteolytic and granular contents directly onto the bone. Free immune complexes further activate complement cascades, resulting in further articular degradation (Firestein, 2003) and increases in levels of proinflammatory cytokines resulting in the upregulation of Receptor Activator for Nuclear factor κB Ligand (RANKL) upon the surface of fibroblast-like synoviocytes. RANKL induces the accelerated maturation of osteoclasts (bone-resorption cells) (Schett *et al*., 2005). A large proportion of the cartilage damage is caused by the action of proteases released by the activated fibroblast-like synoviocytes in the pannus (Sweeney and Firestein, 2004). As the cartilage disappears, fibrous adhesions are formed and the bone ends are bound together by fibrous tissue (bony ankylosis).

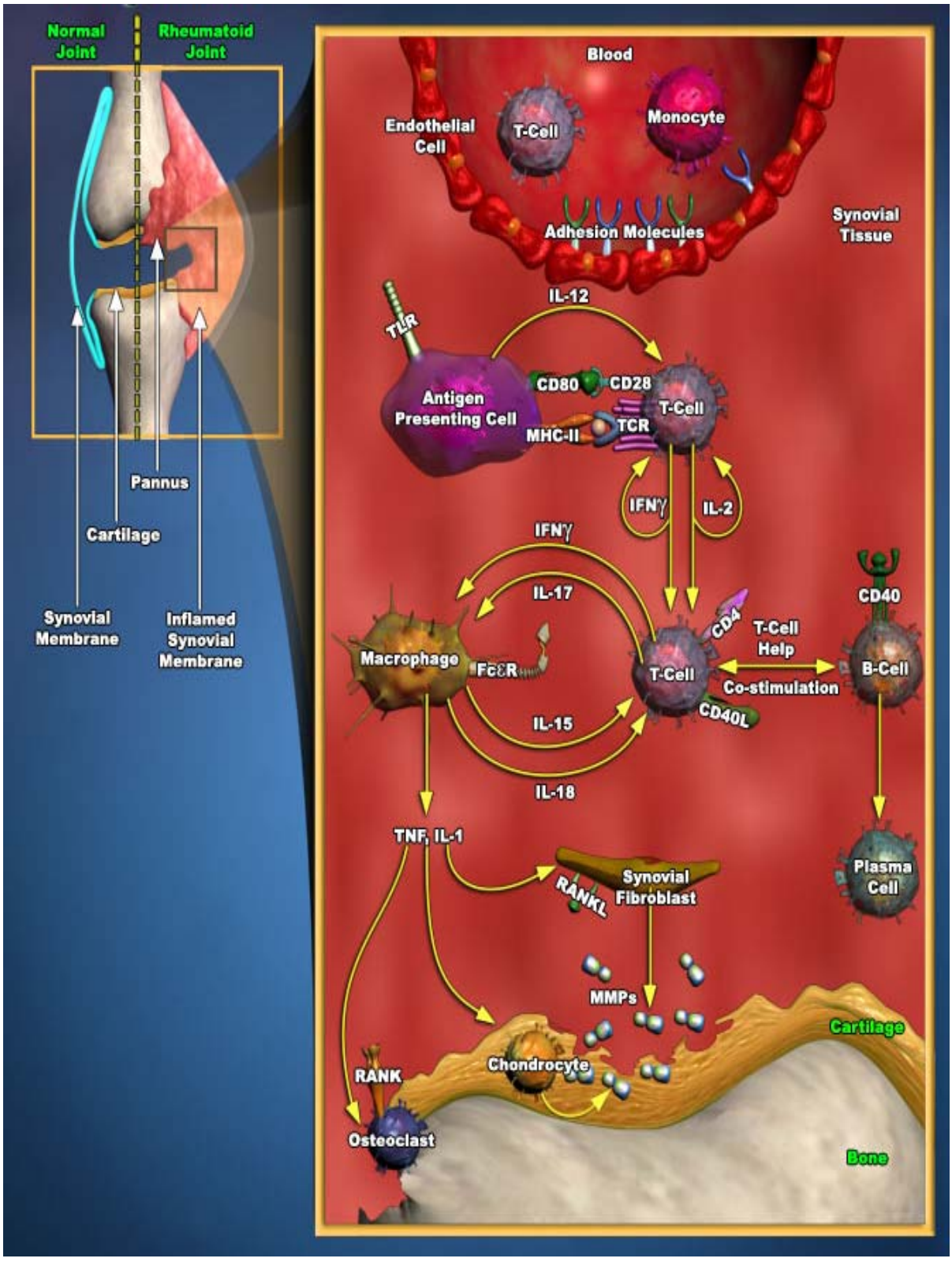


Figure 1.8 The main immunological processes and cellular interactions active in Rheumatoid Arthritis disease pathogenesis. (figure taken from http://www.ambion.com/ -Permission to use diagram given by www.ambion.com). The pathology of RA is characterised by the infiltration of several inflammatory cells in both the Pannus and joint fluid. Articular manifestations are mediated by antigen presented in the context of MHC-II to T lymphocytes. Upon activation these cells express both chemokines and proinflammatory cytokines which act to both further propagate and regulate inflammatory responses. Activated MHC-II molecules secrete regulatory cytokines and growth factors. TNF-α plays a central role in these processes through indirect mechanisms leading to bone and damage, both via induction of IL-1β and through accelerated differentiation of Osteoclasts. TNF-α also acts to increase expression of endothelial cellular adhesion molecules resulting in the recruitment of more immune cells into the joint.

1.2.5. Clinical manifestations of RA

Like the majority of Rheumatic diseases, the clinical presentation of RA is heterogenous and characterised by a wide variation in presentation, prognosis and disease severity. Such heterogeneity makes diagnosis and treatment difficult, with clinical manifestations and course being typically characterized by periods of exacerbation and remission (Grassi *et al*., 1998).

The most common manifestations appear as a symmetrical polyarticular arthritis, whose onset occurs early in the disease course. This primarily affects the smaller diarthroidial joints, (most frequently the metacarpophalangeal and proximal interphalangeal joints, usually in a symmetrical pattern. Clinical manifestations of RA tend to depend upon the joint involved and the disease stage. Overall the disease is presented as an acute/subacute inflammatory arthritis with patterns of presentation being palindromic (episodic), monoarticular (slow and acute), extra-articular (tensynovitis and bursitis), in addition to the accompaniment of more generalised symptoms, e.g. malaise, fatigue, weight loss and fever. In order to make a diagnosis of RA, a patient must satisfy four or more of the revised criteria for at least 6 weeks, listed by the American Rheumatism Association (ARA) (Table 1.4) (Arnett *et al*., 1988). Such criteria have become useful in the classification of patient subgroups and in the assessment of disease frequency and diagnosis. The disease is not just confined to an inflammatory reaction within the joints and manifestations can include extra-articular involvement such as rheumatoid nodules. Vasculitis complications may be evident in the small vessels leading to gangrene and arteritis and cardiac involvement can include the development of nodules, resulting in valve damage. Pulmonary involvement may include interstitial fibrosis, pulmonary nodules, and pleural effusions. Overall, the course of the disease is primarily progressive with intermittent remission and worsening, but continuing disease activity, supported through the slow progression of joint symptoms which are atypical of RA. Regular radiologic studies are of crucial value in both prognosis and treatment of patients (Salaffi *et al*., 1994).

ARA revised criteria for rheumatoid arthritis classification
Morning stiffness.
Arthritis of three or more joint areas.
Arthritis of hand joints.
Symmetric arthritis.
Rheumatoid nodules.
Serum rheumatoid factor.
Radiographic changes.

Table 1.4. The revised criteria for the diagnosis of Rheumatoid Arthritis (Arnett *et al*., 1988).

1.2.6. The Immunology of different cell types involved in RA

The normal synovium is a relatively acellular structure containing a thin layer of synoviocytes. The sublining is made up of an extracellular matrix consisting of blood cells, synoviocytes and mononuclear cells. In the inflamed synovium of RA patients this differs, with patients showing an extensive infiltration of T cells and B cells into the joint. Each of these populations plays a distinct role in the disease pathology of RA (Zvaifler, 1973).

1.2.6.1. Roles of B cells in Rheumatoid Arthritis

B cells constitute an essential maintenance of the cellular and humoral protective memory and may play a pathogenic role in RA via a number of mechanisms, including the production of autoantibodies (AuAbs), activation of APC function, modulation of T cells, synthesis of cytokines and ectopic neolymphadenogenesis (Lipsky, 2001).

The evidence supporting a role for B cells in RA initially came after the finding of a large number of auAbs to self proteins in RA patients. Autoantibodies also play a role in the generation of immune complexes, which result in the activation of the complement cascade pathways (Firestein, 2003). Potential autoantigens involved in RA include Collagen II (Wooley *et al*., 1984), Sa (Menard *et al*., 2000), BiP (Berner *et al*., 2000) and p205 synovial protein (Hain *et al*., 1996). The majority of those identified however, exhibited a typical lack of specificity (Routsias *et al*., 2004). Two

autoantibodies with a notable specificity for RA however, include Rheumatoid Factor (RF) and Anti-citrullinated cyclic peptide (anti-CCP). RF are antibodies directed to determinants on the Fc portion of IgG molecules (Westwood *et al*., 2006), whereas anti-CCP recognise the presence of deiminated arginine upon the surface of proteins such as Fibrinogen, Filaggrin and Vimentin (Zhou and Menard, 2002). Interest has been generated in these two in particular due to their high sensitivity and specificity, in addition to the fact that their presence is detectable years prior to the onset of disease, suggesting that the breakdown of immunological tolerance actually begins at an early stage, with viable symptoms occurring at an advanced stage of disease (Rantapaa-Dahlqvist *et al*., 2003). The role of B cells in RA pathology is not limited to autoantibody production. They may also play a crucial role in a particular subset of patients (11%) whose synovitis is characterised by the generation of Germinal Centres (GC) (Kim and Berek, 2000). With the main cellular component of GCs being activated, they provide an environment in which B cells are able to differentiate locally into plasma cells. Furthermore, analysis of the B cell repertoire in GCs suggests a local antigen-driven immune response, with most B cells isolated from GCs in the synovium exhibiting unmutated V_H (Hypervariable) genes. This supports the hypothesis that they recently entered the synovium and were activated at the site of disease (Kim and Berek, 2000). Such mutations are limited and subsequent gene rearrangements and mutations implicate synovial autoimmunity directed against autoantigens (Clausen *et al*., 1998). It is likely that TNF-α plays a major role in the formation of lymphoid tissue (Pasparakis *et al*, 1996). Analysis in immune deficient mice has shown that B cells express lymphotoxin-α with the expression of this molecule essential for the development of Follicular Dendritic cell (FDC) networks (Fu *et al*., 1998). One suggestion has been that autoantibodies such as RF are produced as a result of B cell activation within such follicular structures in the synovial tissue. Interestingly studies using hybridoma lines with RF specificity from an RA patient showed that several lines expressed an identical V_H gene rearrangement, suggesting an increased affinity for full immunoglobulin, leading to the conclusion that RF specific B cells undergo the affinity maturation within the synovium (Randen *et al*., 1992). The further importance of B cells in these structures was shown by the treatment of these regions with anti-CD20 resulting in the destruction of such structures accompanied by the complete loss of FDC networks and disruption of T cell activation (Takemura *et al*., 2001). This supports the hypothesis that B cells play a pivotal role in synovial inflammation when GCs are

involved in pathogenesis. Work has also focussed more upon B cells expressing B lymphocyte stimulator (BlyS) or B cell activating factor (BAFF). This molecule enables the survival of plasma cells that undergo cycles of high antibody-mediated inflammation (Dorner, 2006). BlyS levels in RA are significantly increased to controls with RA patients exhibiting enhanced B cell activity (Salzer *et al*., 2005). BlyS levels in the synovial fluid are increased over that in plasma, suggesting B cell differentiation may occur at the site of disease (Dorner, 2006). The contribution of T cells appears not to be necessary for the development of B cell function. Cross-linking of immune complexes with B cell receptor (BCR) or binding of Toll-Like receptors (TLR), may result in the activation and conversion of B cells to plasma cells. Such changes are likely to result in the synthesis of proinflammatory cytokines and chemokines, including IL-1, IL-6 and TNF-α (Lau *et al*., 2005). IL-10 produced by B cells plays an important role in inducing dendritic cell activation which results in the enhanced antigen presentation and the further differentiation of B cells to plasma cells. The increase in plasma cells then results in the synthesis of autoantibodies such as RF and anti-CCP, leading to the local production of immune complexes (Firestein *et al*., 2003).

The recent discovery of a spontaneous arthritis mouse model, in which mice produce antibodies to the ubiquitous antigen, 'Glucose-6-Phosphoisomerase (GPI)', has led to a resurgence of interest in autoantibodies and immune complexes (Kouskoff *et al*., 1996) and although the mere presence of autoantibodies within serum may not be specific for RA, they could well represent a type of antibody signature for subsets of patients. Innate immunity also plays a role in antibody-mediated arthritis with mast cells and TLRs being required for the disease onset in some models (Terato *et al*., 1995, Lee *et al*., 2002, Ronaghy *et al*., 2002).

1.2.6.2. Roles of T cells in Rheumatoid arthritis

T-cells are a key component of the human adaptive response and there is now growing evidence supporting a potential role for them in RA pathogenesis. T-cells are the dominant species of lymphocyte within the synovium – most likely due to clonal expansion (Weyand, 2000). CD4+ T-cell infiltrates within the synovium typically form follicular aggregates with B cells and with CD8+ T cells occupying interfollicular tissue. In fact the synovium appears to provide an environment promoting the survival of T cells as despite a high level of expression of apoptotic markers on cells, T cell

death rates remain low (Salmon *et al*., 1997). T cells themselves within the joint are involved in a wide array of activities, including the production of pro-inflammatory cytokines, cell to cell interactions, antigen specific responses, osteoclast activation and bone resorption suggesting they may play a central role in any disease process (Firestein and Zvaifler, 2002). In addition to this, a distinct correlation of T_H1/T_H2 (T-Helper cell subsets) ratios with disease activity has also been reported (van der Graaff *et al*., 1999). Further evidence supporting the role of a T_H1 response in RA pathogenesis has been hinted at due to clinical remission of symptoms during pregnancy (Olsen and Kovacs, 2002) and decreased prevalence of allergens in RA sufferers most likely due to predominant roles of IL-2 and IFN-γ in skewing responses to that of a T_H 1 nature (Verhoef *et al*., 1998). Furthermore studies have shown increased numbers of CD4+ T cells producing IFN- γ in the synovial fluid of RA patients, compared to controls (Davis *et al*., 2001) with a distinct association of T cell activation marker CD69, with severe disease (Iannone *et al*., 1996). Circulating levels of CD69+ cells are low, leading to suggestions that T cells expressing CD69 within the synovium are likely to be activated at the site of disease. CD69+ cells may also contribute to disease via the induction of CD40 ligands, which have been shown to be responsible for the promotion of B cell proliferation, Ig production, monocyte activation and dendritic cell differentiation (MacDonald *et al*., 1997).

Much speculation has centred upon a role for T cells within RA. Such claims are further supported by the observation that elimination of T cells from rheumatoid arthritis patients results in the amelioration of symptoms, although this effect is only temporary suggesting it is not a causative role (Vaughan *et al*., 1984). Additionally, T cells have been shown to passively transfer disease from humans to immunocompetent SCID mice models. Mice with synovial grafts were able to transfer disease symptoms, which were subsequently reduced upon treatment with anti-CD2 antibodies (Klimiuk *et al*., 1999). With the relative risk of developing RA in monozygotic twins (12-62 times) and with dizygotic twins (2-17 times) higher than in unrelated individuals (Weyand, 2000), such evidence suggests a genetic basis for the disease. To date the best identified risk factor for RA has come in the correlation of HLA-DR with incidence of inflammatory arthritis (Winchester and Gregersen, 1988). Conversely a triggering antigen may not be a necessary requirement for autoimmunity due to the excessive numbers of autoreactive T cells in the synovium (Goronzy *et al*., 1998, Auger *et al*., 1998).

The Regulatory T cell (T_{Regs}) response may also be functionally compromised in RA patients. T_{regs} (CD25+/ CD4+) play the role of policing immunological T cell responses, thereby maintaining self tolerance and suppressing proliferation of effector T cells. These cells in RA patients showed an anergic phenotype of effector T cells in addition to an inability to suppress proinflammatory cytokines (Ehrenstein *et al*., 2004). They also contribute to disease pathogenesis through the conversion of the synovium into lymphoid tissue with T-cell dependent germinal centres, formed from CD4+ T cells and CD20+ B cells (Goronzy and Weyand, 2005). Costimulatory molecules play important roles in the T cell activation and the ensuing response, adding a further regulatory mechanism to the process of T cell activation. One of the more prominent T cell costimulatory receptors is CD28, which is expressed upon most T cells. It is a receptor responsible for interacting with the B7 family of proteins of which B7-1 (CD80) and B7-2 (CD86) are the best characterised. Both are expressed upon the surface of APCs and studies in knockout mice have shown their roles to be essential for a full, functional T cell response in mice (Shahinian *et al*., 1993). *In vitro* and *in vivo* studies in RA patients have also hinted at a significant role in disease activity with upregulated expression of both CD80/86 proteins in the synovium, hinting at a self sustained mechanism for T cell activation, even in the absence of a triggering antigen (Genovese *et al*., 2005).

With increased knowledge of potential T-cell pathways and their subsequent role in disease, novel therapeutic strategies have already been designed, targeting T cell autoreactivity. One example is the use of CTLA-4 to interfere in binding between CD28 and CD80/86 (Kremer *et al*., 2003). Additionally, subsets of T cells ($CD4^+CD28^-$) have been identified in patients with RA, which bypass the requirement for a costimulatory receptor. Their presence has been reported to correlate with progressive erosive disease in patients with early RA. These cells share some features with those of Natural killer (NK) cells with the surface expression of regulatory killer activating and inhibitory receptors, in addition to exhibiting cytolytic activity accompanied by the synthesis of high levels of pro-inflammatory cytokines (Goronzy *et al*., 2004).

1.2.6.3. The role of fibroblast-like synoviocytes in RA

The rheumatoid synovium harbours cell populations of synoviocytes which can be separated into two distinct subfamilies - those of monocyte lineage (macrophage like

synoviocytes) and those of fibroblast lineage (Fibroblast-like synoviocytes) (Duke *et al*., 1982). Originally these cells were thought to play roles of little or no importance in disease pathogenesis. It is now becoming apparent however, that both populations (particularly those cells of fibroblast-like origin) are actively engaged in the initiation and perpetuation of rheumatoid arthritis) and it is this feature that distinguishes RA from other inflammatory arthritides (Huber *et al*., 2006).

Fibroblast-like synoviocytes appear to be differentiated into several different subpopulations of cells within the synovium, dependent upon their location (Fassbender, 1983). These populations also differ in their secretion patterns of matrix proteins, cytokines, adhesion molecules, i.e. VCAM-1 and costimulatory molecules, e.g. CD40 (Buckley, 2003). Synoviocytes within the acellular synovial membrane lining in RA patients (RASFs) have been found to differ substantially from those found at deeper levels within synovial tissue, as they take on an activated phenotype associated with an invasive and aggressive nature (Pap *et al*., 2000b). These activated cells are able to cause disease, either by direct processes, i.e. the attachment to cartilage via an upregulation of adhesion molecules and the subsequent production of matrix digestive enzymes (Muller-Ladner *et al*., 1996), or by indirect mechanisms such as the induction and differentiation of macrophages into osteoclasts via the upregulation of RANKL, in response to proinflammatory cytokines (Nakano *et al*., 2004, Gravallese, 2002, Takayanagi *et al*., 2000).

There are a number of potential mechanisms by which these cells can be activated. These include innate immune mechanisms, signalling from neighbouring fibroblasts and also by components of the immune response. TLRs are key structures in the innate immune response for the recognition of conserved microbial patterns (Uematsu and Akira, 2006). Their activation can result in the upregulation of proinflammatory cytokines and chemokines (Pierer *et al*., 2004). Subsequent reactions lead to the establishment of a chronic inflammatory reaction within the synovium. Production of proinflammatory cytokines contributes to the activation of RASFs, in addition to growth factors and cross-talk between neighbouring fibroblasts. TNF-α is one of the major players in triggering inflammation and joint destruction with its importance demonstrated in the TNF transgenic mouse model (hTNFtg) (Keffer *et al*., 1991) in which systemic overexpression leads to the initiation of chronic synovitis, cartilage destruction and bone erosion (Zwerina *et al*., 2004). Other cytokines that may contribute to fibroblast activation include IL-1, IL-15, IL-22 and IL-21R (Jungel *et al*.,

2004, Ikeuchi *et al*., 2005). However cytokines do not play an exclusive role in activating RASFs as has been shown in SCID mouse co-implantation models (Geiler *et al*., 1994, Muller-Ladner *et al*., 1996), which showed that disease processes could still occur in the absence of both inflammatory cells and proinflammatory mediators (Muller-Ladner *et al.,*, 2003). Amongst the direct consequences of fibroblast activation is the upregulation of adhesion molecules upon their surface. This increase in expression enables the direct attachment of RASFs to articular cartilage. In RA, the β1 subfamily of integrins plays an important role. These integrins mediate the attachment of RASFs to fibronectin-rich sites on cartilage and collagen (Ishikawa *et al*., 1996). Upon binding, integrins activate the early cell cycle within the RASF, activating expression of MMPs, i.e. zinc containing endopeptidases involved in tissue remodelling. Together with oncogenes *c-fos* and *c-myc*, these all play roles in activating RASF, leading to tissue destruction (Fujiwara *et al*., 1995, Dike and Farmer, 1988).

The process of tissue degradation comprises four major stages:

- **Growth and activation.**
- **Spreading.**
- **Invasion of inflamed synovial tissue.**
- **Destruction of cartilage.**

With the overall degradation process being mediated by MMPs, cathepsins and an activated plasmin system (Trabandt *et al*., 1991, Huber *et al*., 2006), the exact mechanisms by which RASFs are activated in RA are still unknown, although it is clear that the altered aggressive phenotype of RASFs mirror specific alterations within the intracellular signalling pathways and transcriptional profiles, e.g. apoptotic profiles, upregulation of oncogenes and the downregulation of tumour suppressor genes, may all contribute to the subsequent activation of RASFs. Transcriptional profiles particularly can be significantly modified by stimulation with either T_H1 (IFN-γ) or T_H 2 (IL-4) (Parsonage *et al*., 2003). Evidence suggests that the subsequent hyperplasia within the synovial tissue occurs due to impairment in apopotic mechanisms of RASFs and macrophages. The resulting deficiency in apoptosis results in the prolonged survival of

RASFs. This is supported by upregulated levels of bcl-2, sentrin-1 and FLICE-inhibitory proteins (FLIP) identified at the sites at the cartilage/pannus junction (Catrina *et al*., 2002, Goberdhan and Wilson, 2003). This activation or changes in cellular expression may also be due to changes in tumour suppressor gene expression. Phosphatase and tensin homolog (PTEN) is involved in cell cycle arrest and apoptosis, with its expression in RA patients shown to be significantly reduced to that observed in healthy synovial fibroblasts, whilst little or no expression has also been recorded at the cartilage pannus junction (Pap *et al*., 2000a).

Although overexpression of wild type (wt) p53 is characteristic of RA, its expression has also been shown to be significantly reduced in RA patient SF cells invading articular cartilage (Seemayer *et al*., 2003). p53 is a transcription factor that plays a central role cell growth, cell cycle control, DNA repair and apoptosis and as such it has been reported to be mutated in about 55% of human cancers (Hollstein *et al.*, 1991). Overexpression of p53 occurs when DNA damage occurs within the cell, however if DNA strand breaks are severe and persistent this may result in mutations of p53 that may result in the dysregulation of apoptosis and tumor suppression mechanisms. This can prove significant in RA as inactivation or mutation of wt-p53 has already been shown to result in enhanced FLS proliferation, incease anchorage independent growth and lead to invasiveness into the cartilage (Pap *et al.*, 2001; Aupperle *et al.*, 1998). Additionally, mutant versions of p53 have been shown to lose their gene repressor activity i.e. that of MMP1,3 and 13 (Sun *et al*., 2004) L1 retrotransposons in synoviocytes have also been shown to induce Mitogen-activated protein kinase (MAPK) p38 with p38 able to induce and upregulate MMP-1, MMP-3, IL-6 and IL-8 (Suzuki *et al*., 2000).

Another theory recently proposed, suggests the inflammatory response is a result of the inability of the RASFs to switch off the resulting inflammation, thus resulting in the inappropriate retention of leucocytes within the synovium, with increased expression of surface chemokines, i.e. CXCL12, reported on RASFs (Buckley *et al*., 2000). Such increases may result in CD4+ T cells accumulating within the synovium at a point when the inflammatory response should be winding down (Buckley *et al*., 2004). This ectopic expression of chemokines has also been observed in other autoimmune manifestations (Amft *et al*., 2001), in addition to being upregulated in stromal cells in response to microbial infection (Mazzucchelli *et al*., 1999).

1.2.7. Treatment strategies in Rheumatoid arthritis

It has now been identified that joint damage occurs early in the course of disease pathogenesis in RA patients; therefore within two years of disease ~60% patients show some radiographical evidence of some degree (van der Heijde, 1995). Whilst the effectiveness of early use of disease modifying anti-rheumatic drugs (DMARDS) has been well documented, these approaches until recently have been limited to mono- or combination therapy with either immunosuppressives or DMARDs such as Prednisolone (Saag *et al*., 1996), intramuscular injections of gold salts (Myochrysine® and Solganal - which have been reported previously to work through inhibition of T cell proliferation (7932408) (Rau *et al*., 1997) and Sulphasalazine and Methotrexate (Ortendahl *et al*., 2002). With the overall efficacy and response to new drugs defined by an outcome measure of the American College of Rheumatology (Felson *et al*., 1995) in recent years, several new drugs have been developed in order to target specific points within the disease process. Through observations made in mouse models and its established upregulation of RANKL (Keffer *et al*., 1991), TNF-α plays an important role in the pathogenic mechanisms of RA (Choy and Panayi, 2001). Three agents have recently been developed in order to target and inhibit the effects of this particular cytokine. Infliximab is a chimeric anti-TNF-α antibody consisting of the antigen binding region of mouse antibody and the constant region of human antibody. It has shown the ability to bind to both soluble and membrane bound TNF-α, thereby eliminating binding to receptors and those cells expressing it upon their surface (Goldblatt and Isenberg, 2005). Etanercept is a soluble TNF-α receptor fusion protein composed of two dimers each with ligand binding portions of type-2 receptors (p75) linked to the Fc portion of human IgG1. Proteins are able to bind to both TNF-α and TNF-β, thus preventing each from interacting with their respective receptors. Adalimumab is a recombinant human IgG1 monoclonal antibody with the ability to bind TNF-α, thereby precluding binding to its receptor. The monoclonal antibody is also able to lyse cells that express the cytokine upon their surface. As a human recombinant antibody it has the advantage of avoiding the generation of anti-chimeric antibodies. Several side effects and problems have been associated with the use of anti-TNF-α therapies ranging from minor side effects (Day, 2002), to the exacerbation of infections such as histoplasmosis and tuberculosis (Nakelchik and Mangino, 2002, Kroesen *et al*., 2003) and demyelinating neurological diseases (Mohan *et al*., 2001).

IL-1 is a pro-inflammatory cytokine produced by stimulated monocytes, macrophage and synoviocytes (Dinarello, 1996). IL-1 receptor antagonists compete with IL-1 for binding to the receptor and thus subsequently downregulating IL-1 activity. IL-1 has been shown to stimulate the release of MMPs and increase the bone resorption of osteoclasts, thereby having a direct influence over bone erosion. This was shown in mice, deficient in IL-1 receptor antagonists which developed inflammatory arthritis similar to RA (Horai *et al*., 2000). It was also shown in RA patients who exhibited levels of IL-1ra lower than expected, relative to levels of circulating IL-1 (Arend *et al*., 1998).

Contrary to long standing belief that RA was a predominantly T cell mediated disease, the role and importance of B cells in disease pathogenesis was reaffirmed by the development and trials of B cell depletion therapies, most notably that of Rituximab (Leandro *et al*., 2002). B cells contribute to RA pathogenesis through a number of mechanisms (Section 1.2.6.1). Autoantibodies have the ability to drive their own production, this thereby provided the initial rationale for the proposal that the elimination of B cells may result in prolonged disease remission (Edwards *et al*., 1999). B cell development is generally defined by a series of changes in surface markers. CD20 is expressed at intermediate stages before being lost prior to the final maturation/terminal differentiation into plasma cells. Rituximab is the form of a chimeric antibody raised to human CD20 and is thought to act via complement mediated antibody dependent cell-mediated cytotoxicity, induction of apoptosis and inhibition of cell growth (Reff *et al*., 1994). Treatment results in the rapid depletion of CD20+ B cells in the peripheral blood. These are subsequently replenished by stem cells within 3-12 months after treatment (Maloney *et al*., 1997). IL-6 is produced by lymphocytes, monocytes, fibroblast synoviocytes and endothelial cells and is a major cytokine in the process of inflammation and haematopoiesis (Nishimoto and Kishimoto, 2006). Whilst evidence clearly shows a role in inflammatory processes, over expression leads to the generation of autoreactive T cells and osteoclasts, stimulation of autoantibody production and the release of inflammatory mediators (Nishimoto and Kishimoto, 2004). Atlizumab is a humanised anti-IL-6 receptor monoclonal antibody. Despite limited success at targeting T-cell pathways, i.e. anti CD4+ antibodies (Smolen and Steiner, 2003), anti-CD5 ricin toxin immunoconjugate (Olsen *et al*., 1996) and anti-CD7 therapy (Keystone, 2002) , some therapies targeting such pathways have shown promise. A successful example of interfering with T-cell regulatory pathways includes

'CTLA4lg'. CTLA4lg works via interference with the CD80/86 pathway (Section 1.2.6.2) which acts as a negative feedback loop for suppression of T cell activity and as a safeguard against autoimmunity. CTLA4lg is a fusion protein that binds to CD80/86, thereby preventing their binding to other CD28 receptors, thus activating further T cells. This approach has shown promising results in trials (Firestein, 2003). Eculizumab is a humanised monoclonal antibody, which prevents the cleavage of complement component C5 into its pro-inflammatory components (Kaplan, 2002, Smolen and Steiner, 2003). Antibodies to adhesion molecules such as Intracellular Adhesion Molecule-1 (ICAM-1) which act to block the entry of lymphocytes into the joint have also shown promising results (Mikecz *et al*., 1995).

1.2.8. Risk factors for RA

Despite intensive research into the triggering cause of RA, the underlying causal agents have remained elusive with a combination of environmental and genetic factors likely being responsible. Risk factors highlighted within the literature include age, gender, hormonal and reproductive factors and lifestyle (Goronzy, 2005). Increasing age has been reported to correlate with incidence of disease with increasing age peaking between 75-85 years (Doran *et al*, 2002). Peaks are also observed in women at menopausal age (Voigt *et al*, 1994) and this combined with observations of remissions during pregnancy suggest a role for hormonal and autocrine modulation (Masi *et al*, 1995). Evidence has been reported suggesting associations with the Oral contraceptive pill (OCP). The OCP has been suggested to have a protective effect against RA, however this has yet to be substantiated (Spector and Hochberg, 1990). Thyroid and neuroendocrine hormones can also influence RA, partly through actions in macrophages (Wilder and Elenkov, 1999). Other associations have included a positive link between smoking and incidence of RA in women (Hernandez Avila *et al*., 1990, Stolt *et al*., 2003) which has recently also been linked to the strong genetic background of a (HLA)-DRB1 genotype (Padyukov *et al*., 2004). These have additionally been linked, as with OCP to lower androgen levels – a risk factor for the onset of RA (James, 1993). Other plausible links include environmental factors including red meat (Pattison *et al*., 2004b), which acts as a susceptibility marker. Oil-rich diets, such as those in the

Mediterranean, (Cyranowski *et al*., 2004) and vitamin C (Pattison *et al*., 2004a) have been reported to have a protective effect.

Virus	Genus / family	Genome	Protein	Disease	Comment
Retroviruses (endogenous)	Dep. on family	ssRNA	Varied: Simple	Varied: None conclusive	For most part lie dormant within human genome
HTLV	HTLV-BLV (formerly Oncovirinae)	Diploid RNA	6	ATL/ L/ TSP	Distinguished by typical genome and ability to cause tumours
Rubella	Rubivirus / Togaviridae	RNA	-	GM/ PA-(adults)	-
Parvovirus B19	Erythrovirus / Parvoviridae	ssDNA	3	EI/ AC/ FI	Can form persistent infection in immunocompromised patients.
EBV	γ-Herpesvirinae/ Herpesviridae	Linear dsDNA	80-100	IM/ L	Ubiquitous virus
CMV	β- Herpesvirinae / Herpesviridae	dsDNA Linear	80-100	IM-like disease	Important in transplant recipients Risk of maternal infection
HHV-6	β- Herpesvirinae / Herpesviridae	dsDNA Linear	80-100	ES/R– children IM-like disease - (adults)	-
HSV 1/2	α-herpesvirinae/ Herpesviridae	Linear dsDNA	80-100	HW	Ubiquitous virus
Hepatitis B	Orthohepadnavirus / Hepadnaviridae	Circ. dsDNA	6	HC/ LC/ CH	Spread by blood or sexual activity
Hepatitis C	Hepatitis C-like virus / Flaviviridae	dsDNA	11	HC/ LC	Blood borne
Adenovirus	Mastadenovirus / Adenoviridae	Linear dsDNA	-	RI/ PF/ EK/ GI	Airborne, ubiquitous

Table 1.5 The main viruses that have been implicated in the pathogenesis of rheumatic disease-like symptoms within the literature (Section 1.18.1).
(Key: ATL: Adult T-cell leukaemia; L: Lymphoma; TSP: Tropical Spastic Paraparesis; EI: Erythema Infectiosum; AC: Aplastic Crisis; FI: Fetal infection; IM: Infectious Mononucleosis; ES: Exanthema Subitum; R: Roseola (Slapped cheek syndrome); HW: Herpetic warts; GM: German Measles; HC : Hepatocellular Carcinoma ; LC : Liver Cirrhosis ; CH : Chronic Hepatitis ; PA : Polyarthritis; RI: Respiratory infection; PF: Pharyngoconjunctivial Fever; EK: Epidemic Keratoconjunctivitis)

Much of the disease pathology and the immunological mechanisms at work in disease processes suggest an initial trigger is responsible for the disease. The identity of this agent has however remained elusive, despite a large number of studies focusing upon the question. Those investigated have included bacterial and fungal agents such as Proteus and Klebsiella (Tani *et al*., 1997), Candida albicans (Hida *et al*., 2007) and Mycoplasma species (Postepski *et al*., 2003). Much attention has also focused upon viral pathogens, and potential roles that they may play in either triggering or propagating disease processes (Masuko-Hongo *et al*., 2003) (Table 1.5)

1.3. Aims and Objectives

The overall aims of this research project were to test the association between the expression of HERV-K and onset of Rheumatoid Arthritis. This involved the direct elucidation of those mechanisms by which HERVs may play a role in, or trigger disease immunopathology. This was undertaken in an approach that ran parallel to the central dogma of biology itself, with investigations taking place on a DNA, RNA and protein level. This approach enabled the determination of the nature of HERV upregulation, i.e. whether it resulted in protein production, suggesting a complete open reading frame read-through, thus narrowing down potential mechanisms by which HERV-K may trigger or propagate the disease. Such approaches included:

1. **DNA**: A bioinformatic approach to the identification of putative HERV-K encoded epitopes that may play a role in molecular mimicry through shared sequence homology with disease autoantigens (Chapter 3).

2. **RNA**: The development of a real-time PCR assay and its use in the analysis of HERV *gag* gene expression in RA patients and controls and the use of the aforementioned tools and reagents in the exploration of different mechanisms by which HERVs may trigger or participate in disease pathogenesis (Chapter 5).

3. **Protein**: An immunological assay of epitope expression in patient samples comparing RA versus disease controls (Chapter 4).

Chapter 2

Chapter 2: Materials & Methods

All reagents were purchased from Sigma, UK unless otherwise stated.

2.1. Ethics applications

Ethical approval involving work with NHS patient samples, during the project was sought from the relevant Local Regional Ethics Committees (LRECs) (Table 2.1). Ethical approval for selected project protocols was also granted from internal University ethics committees.

Committee	Reason for Application	Site/Sample type	Date Approved
Wolverhampton LREC	EPE	New Cross/ SF, PB	09/04
University of Wolverhampton	Internal ethics – ST	RES20a/b	02/04
Wolverhampton LREC	EPE	New Cross/ ST	04/05
Wolverhampton LREC	SA	New Cross/ IBD PS	11/06

Table 2.1 Ethics applications submitted for work carried out during this research project. (Key: SF – Synovial Fluid; PB – Peripheral Blood; ST – Synovial Tissue; SA – Substantial Amendment; SSE – Site Specific Extension; EPE - Existing Project Extension; IBD PS– Irritable Bowel Disease patient serum.).

2.2. Collection and processing of patient samples

2.2.1. Collection of patient samples

The ethics committees of the New Cross Hospital, Wolverhampton granted ethical approval for the collection of patients' samples (project number 985). Specific inclusion criteria stipulated that all patients with RA must be in attendance at the Rheumatology clinic, New Cross Hospital, Wolverhampton, and have fulfilled the diagnostic criteria set by the American College of Rheumatology (ACR). No exclusion criteria were set for this investigation. Blood was also taken from other non-RA patients in attendance at the clinic (IBD, SLE and OA) and all samples were taken by staff working at the clinic. 10ml of venous blood were collected in sterile preservative-free heparin from each patient after informed consent had been given. Blood samples were also obtained from healthy volunteers of known age and sex. In order to

maximise accuracy of the investigation all blood samples were processed within 3 hours. All steps were performed under aseptic conditions.

2.2.2. Isolation of PBMN cells from whole venous blood

Mononuclear cells were separated from 10ml whole blood using a Ficoll gradient method. Anti-coagulated venous blood was layered onto 7ml Histopaque®-1077 in a 20ml sterile universal. This was then centrifuged at 400 *x g* for 30 min at room temperature. During centrifugation, erythrocytes and granulocytes were aggregated by polysucrose and rapidly sedimented at the base of the tube whilst lymphocytes and mononuclear cells remained at the plasma-Histopaque®-1077 interface (Figure 2.1). The upper layer (plasma) ending approximately '0.5cm' (see Fig.2.1), above the central opaque interface (containing mononuclear cells) was aspirated and transferred to a sterile universal using a Pasteur pipette. Plasma was stored at -70ºC until required in serological investigations. The opaque interface was also transferred to a sterile centrifuge tube. These cells were resuspended in 1 ml of TRI reagent for RNA extraction (Chomczynski and Sacchi, 1987).

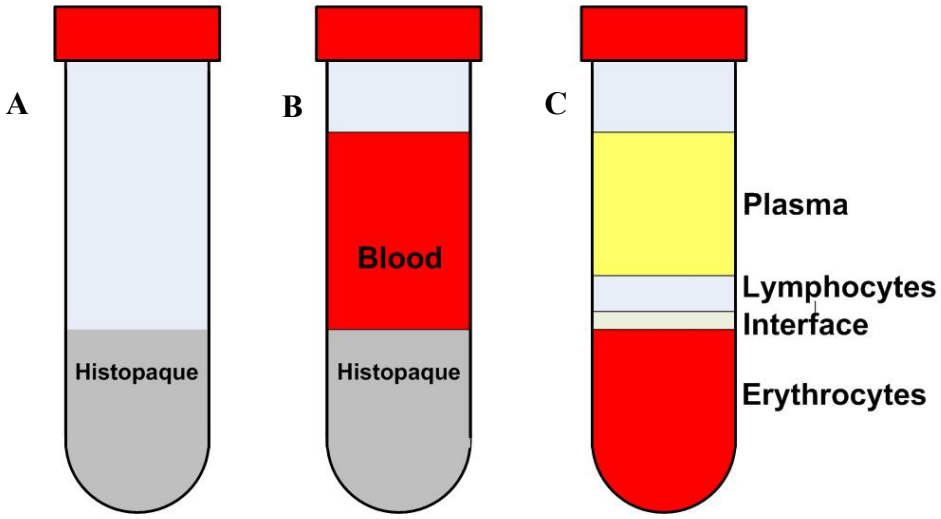


Figure 2.1 Principle of gradient separation procedure used to separate different cell populations from patient blood. A) Histo-Paque was placed at the bottom of a tube, B) Venous blood was layered on top of the Histopaque. C) After centrifugation the following layers were visible in the tube, from top to bottom: Plasma and other constituents; Mono-nuclear cells (PBMC/MNC); Histo-paque, erythrocytes & granulocytes (sometimes pelleted at the tubes base).

2.2.3. Isolation of cells from synovial fluid

Between 10-20 ml of synovial fluid were aspirated from the knee joints of RA and OA patients and kept on ice in sterile universals. Synovial fluid was diluted in 10ml 1x PBS before being centrifuged at 350 × g for 10 min. The supernatant was discarded and the cell pellet washed 2-3 times with PBS. In order to prevent red blood cell (RBC) contamination, cell pellets were incubated in lysis buffer (0.144 M NH_4Cl, 1mM $NaHCO_3$) for 15-20 min at room temperature before centrifuging at 350 *x g* for a further 10 min. Cells were resuspended in 1ml TRI reagent for RNA extraction (Chomsacchi method) (2.8.2).

2.3. Separation of specific cell populations using paramagnetic beads

For a small number of new samples, different species of lymphocytes were separated from one another to ascertain levels of HERV-K (HML-2) in different cell populations. This was done using a series of washes with paramagnetic beads coated with antibody specific for CD19 (B-cells), CD4 and CD8 (T-cells). The Ficoll interface containing mononuclear cells was separated into three equal aliquots. Each aliquot was then mixed with dynabeads specific for the cell population being extracted i.e. B-cell (CD-19). Dynabeads (25ul/ml^{-1} sample) were added to the interphase and incubated at 4ºC under agitation before incubating at room temperature for 2 min with the MPC Dynal magnet (Dynal technologies, UK). After separation the supernatant was discarded and cells/dynabeads washed 5 times with PBS/ 0.1% BSA. Beads and cells were separated from one another using commercially available Detachabead® kits (Dynal technologies, UK). Cells and Detachabead® reagents were mixed and incubated together at room temperature for 45 min before placing upon the MPC Dynal magnet, for a further 5 min were decanted into a sterile centrifuge tube and washed three times in RPMI/1% FCS before centrifuging for 8 min at 500 *x g*. Cell pellets were finally resuspended in appropriate volumes of RPMI/1% FCS. These cells were then lysed in RLT buffer for RNA extraction using the QIAAmp kits (Qiagen, UK).

2.4. Bioinformatic analysis

Reference sequences of potential autoantigens were identified in rheumatic diseases and their protein sequence was extracted from the NCBI/ Genbank online database (www.ncbi.nlm.nih/). Sequences were edited accordingly using a standard text editor.

2.4.1. Primary and Secondary structure analysis

Protein residues were analysed via *in silico* analysis of physicochemical parameters, using normalised scales derived from previous experimental data, in order to elucidate the antigenic structure of each autoantigen. Predictions were made using computerised algorithms for Hydrophilicity (Hopp and Woods, 1981), Residue polarity (Grantham, 1974), Solvent accessibility (Janin, 1979) and flexibility index (Bhaskaran and Ponnuswamy, 1988). These analyses were performed with the bio-computing software programmes from http://www.expasy.ch/cgi-bin/protscale.pl. Window lengths on these programs were set to '9 residues' to mirror the length of the peptide being produced. Thresholds were designated as the top 25% of values for each protein. All residues with values within this top 25% were considered to be over the threshold level, and therefore antigenic. All studies were run in parallel with, and results compared to, a previously published B-cell epitope prediction algorithm called BCEPRED (http://www.imtech.res.in/raghava/bcepred/) that has been claimed to reach 80% prediction accuracy (Saha and Raghava, 2004).

Protein secondary structure analysis including the elucidation and mapping of Beta-turns was examined using the neural network tool 'Betatpred2' server at http://www.imtech.res.in/raghava/betatpred2/ (Kaur and Raghava 2003) and the PSIPRED tool (McGuffin *et al.,,* 2000) available online from http://globin.bio.warwick.ac.uk/psipred/. The PSIPRED program had repeatedly scored highly in prediction accuracy using the CASP4 and EVA projects (http://cubic.bioc.columbia.edu/eva.html). Protein sequence alignments were performed using an online alignment tool LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html).

2.5. Peptide and bleed synthesis

A single peptide 'PLSK' (VGPLESKPRGPSPLSA) was synthesised manually using F_{moc} solid phase synthesis. Synthesis was carried out on a 0.2 mM scale. Kaiser tests were employed to verify coupling and deprotection (Chan *et al*., 2000). Peptide identity was verified by molecular weight using Matrix assisted laser desorption ionisation time of flight mass spectrometry (MALDI-TOF) (Kratos Analytic Kompact probe, UK). The peptide was freeze dried and resuspended in sterile phosphate buffered at 10mM concentration and stored at -20°C. Peptides 'GfPN1kpr' (GPSESKPRGTSPLPAG) and 'GfPN2eip' (PCPKEIPKESKNTEVL) and specific antiserum were synthesised commercially (Severn Biotech, UK).

Antibodies specific for these two peptides (identified in Chapter 3 - GfPN1kpr and GfPN2eip) were raised commercially (home office license no.: 40\2154). These peptides were conjugated to the carrier - Purified protein derivative (PPD) prior to inoculation into two rabbits in order to obtain polyclonal antibodies to the peptide. Bleeds were harvested at specific time intervals including pre-immunisation, 7, 14, 21 and 28 days post-immunisation. Polyclonal serum was subsequently used for optimisation and comparison purposes. The final optimized assay and reagents were then used to detect anti-HERV antibodies in RA patient serum and disease controls.

2.7. Serological assays

2.6.1. Indirect ELISA

Wells on a 96 well plate were coated with 400ng/50µl of peptide and incubated overnight at 37°C. Plates were washed three times with PBS using a Well washer 4 MK2 (Thermo lab systems, Waltham, MA) and blocked with 2% Bovine serum albumin (BSA) in PBS for one hour at 37°C. After three washes with PBS/ Tween20 (Tw20) (0.1%) primary antibody was diluted in PBS/ Tw20 (0.05%)/ BSA (2%) and 50µl was added to the wells and incubated at 37°C for one hour. After washing three times with PBS/ TW20 (0.1%), conjugate (Dako, UK) was diluted appropriately according to manufacturers instructions in PBS/ TW20 (0.1%) / BSA (2%) and incubated for one hour at 37°C in the dark. Plates were washed with PBS/ TW20

(0.1%) before 3, 3', 5, 5'-Tetramethyl Benzidine (TMB) liquid substrate system was added to each well. Reactions were stopped after 5 minutes using 2N Hydrochloric acid. Absorbance was measured at 450nm using a Multiscan MS spectrophotometer (Thermo lab systems, Waltham, MA).

2.6.2. Inhibition ELISA

An inhibition assay was performed as described previously (Westwood & Hay, 2000). Standard curves were produced for each antibody of interest using the indirect ELISA protocol above. Analysis of curves was used to predict the dilution giving an absorbance of 0.5. This antibody dilution was then made up in diluent with PBS/ Tw20 (0.05%)/ BSA (2%) and incubated with 0.5 $\log_{10}$ dilutions of the appropriate or control peptide for 1 hour at room temperature. Incubated preparations were transferred to a plate pre-coated with the 400ng/50 µl of corresponding peptide, with each sample repeated in triplicate. After washing three times with PBS/ TW20 (0.1%) the appropriate conjugate (Dako) was diluted in PBS/ TW20 (0.1%)/ BSA (2%) according to manufacturers instructions and incubated for one hour at 37°C in the dark. Plates were washed with PBS/ TW20 (0.1%) before 3, 3', 5, 5'-Tetramethyl Benzidine (TMB) liquid substrate system was added to each well. Reactions were stopped after 5 min using 2N Hydrochloric acid. Absorbance was read at 450nm using a Multiscan MS spectrophotometer (Thermo lab systems, Waltham, MA).

2.6.3. Semi quantitative RF assays

The RA latex test kit (Plasmatec, UK) was used in order to test serum samples for the presence of rheumatoid factor (RF). After allowing the kit reagents and patient serum samples to equilibrate to room temperature upon the test slide using disposable sterile pipettes. The latex reagent was placed next to the serum and both were mixed using the spatula provided. RF was quantitatively determined in patient samples by comparing agglutination at different serum dilutions (Table 2.2). Four sterile eppendorf tubes were labelled and 100 µl of saline provided, was placed in each tube. In the first tube, 100 µl of patient serum was also added and the two were briefly mixed. Serial dilutions of serum were then established in the remaining tubes using 100 µl of diluted patient sample. These were then tested as described above with latex reagent upon the

test slide. Positive and negative controls supplied by the kit were used with each sample.

Dilutions	1/2	1/4	1/8	1/16
Sample serum	**100 µl**	**-**	**-**	**-**
Saline	**100 µl**	**100 µl**	**100 µl**	**100 µl**
	→	**100 µl**		
		→	**100 µl**	
			→	**100 µl**
Sample volume	**100 µl**	**100 µl**	**100 µl**	**100 µl**
8xNo. Of dilution	**8 x 2**	**8 x 4**	**8 x 8**	**8 x 16**
RF titre (I.U/ml)	**16**	**32**	**64**	**128**

Table 2.2Table showing the dilution strategy required for quantitation of rheumatoid factor in patient serum

2.6.4. Immunofluoresence

Cells were cultured at appropriate seeding concentrations (e.g. 5 x 10^6) in BD Falcon Polystyrene vessel 8 chamber culture slides (Becton Dickinson) with 500µl fresh supplemented medium (see Appendix III) at 37ºC with 5% CO_2 until the monolayer grew to between‘70-90%’ confluency. Spent medium was then discarded and cover slips washed with PBS without Ca^{2+}/Mg^{2+} before incubating in 2ml of Methanol at room temperature. Methanol was removed using disposable Pasteur pipettes and monolayers were incubated a further 2 min in 2ml Acetone to fix cells. Acetone was removed and fixed cells were washed twice in 2ml 1 x PBS without Ca_{2+}/Mg_{2+}. Cover slips were blocked by incubating in 2ml 2% BSA for 1 hour in order to prevent non-specific binding. Primary antibody was diluted to appropriate concentration using 1 x PBS/ BSA (2%) before adding 250 µl to each culture chambers and incubating at room temperature for 1 hour under agitation. Cells were washed three times with PBS without Ca^{2+}/Mg^{2+} for 5 min under agitation. Slides were then

treated with a secondary conjugate (with a goat anti-rabbit – FITC) (Dako, UK). After staining, chamber sides were removed using the tool provided before a single drop of anti-photobleach agent containing DAPI counter-stain (Vectashield™) was added to the slide. Sterile glass coverslips were placed over the fixed and stained cell monolayers and sealed with nail varnish before drying for 2 - 5 min in the absence of light. Slides were then checked for immunofluoresence using the confocal microscope (Carl Zeiss, US) using the FITC and DAPI filters.

2.7. Cell culture

A diverse number of cell types were grown for work within the project ranging from cancer cell lines to lymphocytes and primary cells derived from disease tissue. Cells were grown in conditions dictated by their characteristics i.e. adherent, semi-adherent or suspension. Medium was made up freshly and supplemented accordingly dependent upon the requirements of each cell line (for specific passaging conditions and reagents refer to Appendix III).

2.7.1. Adherent cells

Cultures were monitored using an inverted microscope (Olympus UK, UK) to assess both confluency and the presence of bacterial or fungal contaminants. Spent medium was removed using disposable sterile glass pipettes and the monolayer washed with a volume equivalent to half the culture medium of PBS without Ca^{2+}/ Mg^{2+}. Cells were washed before adding Trypsin/ EDTA to the washed monolayer using 1ml per 25cm^2 of surface area. The flask was tilted to allow the whole monolayer to come in contact with trypsin before incubating at 37ºC for 2-10 min or until cells detached. Cells were resuspended in 8ml fresh serum containing medium to inactivate the trypsin (at this stage a small volume was removed for a cell count when required). For those cultures able to grow in the presence of diluted trypsin (e.g. HEK-293, MCF7), an appropriate number of cells were transferred to a new flask of pre-warmed medium. In cultures sensitive to trypsin (e.g. Tera-1), cells were transferred to a sterile universal and spun at low speeds (110 *x g*) for 5 min. Medium was discarded and the cell pellet resuspended in an appropriate volume of fresh medium. This was then transferred

appropriately to new flasks of pre-warmed medium and incubated at 37ºC with 5% CO_2 until culture reached 70-90% confluency.

2.7.2. Semi-adherent cells

Semi-adherent cell lines such as B95-8 cell lines show characteristics of a mixed culture with a proportion of cells adherent whilst others remain in suspension. The degree of confluency was determined using an inverted phase contrast microscope (Olympus UK, UK). Spent medium was decanted into a sterile centrifuge tube and retained for further use. The monolayer was then washed with PBS without Ca^{2+}/Mg^{2+} using 1-2 ml per 25cm^2 of flask surface area. Trypsin was then added at 1 ml per 25cm^2 surface area into the flask, taking care to immerse the whole monolayer before decanting excess trypsin. The flask was then placed in an incubator for 2-10 min or until cells detached. Trypsinised cells were transferred to the universal containing spent medium and suspended cells, before being diluted appropriately to seed new flasks (at this point cell counts were performed when required). Appropriate dilutions of cells were added to new flasks which were incubated at 37ºC with 5% CO_2 until the monolayer reached 70-90% confluency. This was repeated generally every 72-96 hours, although this was dependent upon the cell line involved.

2.7.3. Suspension cultures

Generally, cell lines derived from blood cells such as lymphocytes grow in suspension. Cultures were viewed using an inverted phase contrast microscope with healthy cells appearing bright, round and refractile under a phase contrast inverted light microscope. Cell counts were performed upon small aliquots of the culture before the flask contents was diluted appropriately in fresh supplemented medium and transferred to new sterile flasks containing fresh supplemented medium.

2.7.4. Resuscitation of frozen cell lines

Cells were stored between usage at -180ºC in a solution of 60% (v/v) RPMI-1640, 10% (v/v) dimethyl sulphoxide (DMSO) and 30% (v/v) fetal bovine serum (PAA, UK). Aliquots of 5x10^6 cells were defrosted by placing in vials of water at room temperature, and then added to 8ml of supplemented media (warmed to 37ºC) in order

to dilute out DMSO. Cells were then incubated overnight before media was changed in order to eliminate any further toxicity due to residual DMSO. Media in each flask was replaced every 48 hours and cells were grown until 70%-90% confluent, approximately every 96-120 hours.

2.7.5. Cell viability assays

The effect of treatments upon cellular viability in *in vitro* cellular models was monitored using the CellTiter 96 AQ_{ueous} cell proliferation assay (Promega, UK) (Berglund et al., 2005). Once all reagents had thawed to ambient temperature, 20ul of Celltiter 96 AQ_{ueous} one solution reagent (Promega, UK) was added to wells on a 96 well plate containing samples in 100ul culture medium. Plates were incubated at 37ºC with 5% CO_2 for 1-4 hours before fluoresence was read on a 96 well plate reader (Thermo lab systems, Waltham, MA) and absorbances were plotted onto a standard curve in order to calculate cell viabilities.

2.8. Molecular assays

2.8.1. Methodology for RNA extraction

All patient samples collected during the project or in previous studies had RNA templates extracted using the Chomczynski and Sacchi methodology (Section 2.8.2). Those samples collected during the latter stages of the project (i.e. primarily cell separations, treatments and *in vitro* models) were extracted using commercially available RNeasy kits (Qiagen, UK).

2.8.2. Chomczynski and Sacchi RNA extraction

After treatment, cells were collected and washed in PBS and were subsequently immersed in Tri®-reagent for 5 min under agitation. Chloroform was added at a 5:1 ratio before incubating at ambient temperature, resulting in phase separation. Layers were further defined by centrifugation at 10,000 rpm for 20 min at 4˚C. After layering, three phases were evident (Figure 2.1).

- **Colourless upper aqueous phase (containing RNA).**
- **Opaque interphase (containing DNA).**
- **Red organic phase (containing protein).**

The RNA phase was incubated in the presence of isopropanol (1:1 Tri®-reagent) before being incubated overnight at 4˚C. Samples were then centrifuged at 12,000 xg for 20 minutes at 4˚C and the supernatant subsequently removed. The resulting pellet was then washed in 75% ethanol before being re-centrifuged for 5 minutes at 4˚C. The supernatant was discarded and the pellet was allowed to air-dry before being re-suspended for quantification in 85ul of nuclease free dH_20, 5ul RNase inhibitor (5U) (Promega, UK) to inhibit the action of any RNases and 10µl RNase-free DNase (1U) (Promega, UK) to remove any contamination by genomic DNA. The sample was then incubated at 37˚C for 30 min to aid dissolution and action of DNase. Samples were then either snap-frozen on dry ice and stored at -80˚C or quantified.

2.8.3. Assessing RNA integrity and DNA contamination of RNA samples

Before using the total RNA in subsequent reverse transcription, it was checked for structural damage and contamination with genomic DNA (Fig. 2.3). The RNA was analysed by agarose gel electrophoresis using 1% agarose-Loening gel, containing Ethidium bromide (0.5 mg/l EtBr). The electrophoresis was performed in 1 X Loening buffer (718 mM Tris base, 680 mM Sodium dihydrogen orthophosphate, 20 mM EDTA; pH 7.7) at 70V for one hour. Gels were visualised with 'Grabber' software (Perceptics, Knoxville, TN) using a BioVision UV gel imager (Perceptics, Knoxville, TN).

2.8.4. Quantification of total RNA

The concentration of total RNA was determined using a spectrophotometer (Geneflow, UK). Once set to measure RNA the instrument was zeroed using 70 µl of nuclease-free water. RNA (7µl) was appropriately diluted 1:10 with nuclease-free water (63 µl) and gently mixed before transferring to a quartz microcuvette. The cuvette was inserted into the Genova (Geneflow, UK) and the RNA concentration (µg/ml) and A_{260}/A_{280} ratio (purity) were measured.

2.8.5. Two – step quantitative Real-time RT-PCR

cDNA template was synthesised using 1 µg of total RNA and was converted into cDNA using oligodT primers and the avian myeloblastosis virus (AMV) RT enzyme (Promega, UK). RT cycles consisted of an initial step of 5 mins at 72°C incubating template with random primers (Promega). This was followed by addition of mastermix containing Magnesium ions (1mM), Reverse transcription buffer, dNTPs and sterile nuclease-free dH_20. This was then incubated at 42°C for 1 hour and 20 min before heating to 95°C for 5 min. cDNA was diluted 1:5 prior to running in the real-time assay. Reagents for the gene of interest (GOI) and housekeeping genes (HKG) were mixed with cDNA in separate 0.5ml eppendorf tubes (Table 2.3). These included 9 µl of sterile nuclease-free dH_20, 0.5 µl HERV-K10 *gag1* set1 primers (nM), 2.5 µl of template (diluted 1:5 in dH20) and 12.5 µl Absolute SYBR green master mix (Abgene, UK). Once mixed, reagents were transferred to a 96-well plastic plate (Abgene, UK) and capped using plastic strips (Abgene, UK) before being sealed using plate sealers (Abgene, UK) preventing possibility of evaporation. Final optimised cycling parameters are listed below (Table 2.4).

Component	Volume (µl)	Final concentration
SYBR Green Master Mix (Abgene)	12.5	1x
Forward Primer (70nM)	0.5	nM
Reverse Primer (70nM)	0.5	nM
Template	2.5	<500ng/reaction
Nuclease Free Water	9	
Total volume	25	

Table 2.3 Constituent reagents making up the mastermix for HERV-K10 *gag1* real-time PCR.

Cycle	Step	Repeats	Temp	Time (mins:secs)
1	**1**	**1x**	**95**	**15:00**
2	**1**		**95**	**00:30**
	2	**40x**	**56**	**00:30**
	3		**72**	**00:30**
	Data collection and real-time analysis enabled.			
3	**1**	**1x**	**56**	**01:00**
4	**1**	**80x**	**56**	**00:10**
Increase set point temperature after cycle 2 by 0.5ºC Melt curve data collection and analysis enabled.				

Table 2.4 Cycling parameters for HERV-K(HML-2) *gag1* qRT-PCR

2.8.6. Primer design: A comparison of 3 different programmes for primer design

All primers were designed using the reference sequence for HERV-K10 (Ono *et al.,* 1986)(accession no: M14123) in the NCBI Nucleotide sequence database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide&itool=toolbar). The performance of three different primer design programs was assessed based upon the success rates of predicted primers. The performance of Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi) (Korwin-Kossakowska *et al.,* 2006), PrimerSelect (LaserGene version 5.0; DNAstar Inc, Madison, WI) (Breslauer *et al.,* 1986, Freier *et al.,* 1986, Baldino *et al.,* 1989, Rychlik *et al.,* 1990) and Beacon designer version 5.0 (Premier Biosoft Int., Palo Alto, CA) (Ram and Shanker 2005) were compared and contrasted. Specificity of all primer sets were confirmed using melt-curve analysis.

2.8.7. Determination of primer efficiency & specificity

Primer reaction efficiency was determined by standard curve analysis. cDNA was diluted appropriately over a ten-fold dilution series ranging from 1ng down to 1pg. All dilutions were made using nuclease free water (Promega). Four different concentrations were run in triplicate in order to construct a standard curve. 'iCycler' software 'iQ' was used to calculate primer efficiency using the standard curve (Figure 5.6).

Primer specificity was confirmed by purifying three PCR products using RNA from several different cell lines (well 1- Tera-1, well 5- HEK-293 and well 6- synovial fibroblasts) using Minielute gel purification kits (Qiagen, UK). DNA sequencing was performed by the Sequencing service (School of Life Sciences, University of Dundee, Scotland, www.dnaseq.co.uk) using Applied Biosystems Big-Dye version 3.1. Chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer. Sequencing was carried out using gene specific primers unless otherwise stated.

DNA sequences of PCR products were then run through the BLAST sequence database to identify most similar matches.

2.8.8. Assay sensitivity & Cloning

PCR product was run in a 2% agarose gel, and the 120bp amplicon cut out under exposure to long wave UV light. This was then purified using MiniElute gel extraction kit (Qiagen, UK, UK) and cloned into the pGEM-T cloning vector (Promega, UK) and transformed into JM-109 competent cells (Promega, UK). Transformed bacteria were incubated overnight before blue-white screening identified colonies positive for the K10PCR-pGEM-T plasmid. Positive colonies were then incubated overnight, at 37°C in LB broth. Plasmids were extracted and purified using a QIAPREP mini kit (Qiagen, UK) and the purified plasmid preps were run in an agarose gel prior to digestion with restriction enzymes to confirm the identity of the insert. Once the presence of insert within the plasmid had been confirmed 100ng of plasmid/insert was aliquoted, and serial diluted ten-fold to give nine dilutions ranging from 100ng to 1fg K10PCR-pGEM. A QPCR reaction incorporating all optimised conditions was run with the nine plasmid dilutions and negative controls (NTC, Plasmid no insert, Irrelevant plasmid – TOPO (Invitrogen) in triplicate.

2.8.9. Data analysis, Quantification and Normalisation.

Fluorescence data was copied and pasted from the 'iQ' software (BioRad) directly into Excel. The freeware programme LinRegPCR (Ramakers *et al*., 2003) was then utilised for analysis and comparison of PCR reaction efficiencies during all investigations. Validated results showed PCR efficiencies between '1.7 and 2.1', with correlations '>0.999'. Results outside of these margins were discarded and repeated. Final $2^{-\Delta\Delta C_T}$ values were calculated by determining average C_T values over repeated runs and applying these normalised values to the relative quantitation formula. Statistics and graphical analysis were performed using Prism 4 (Graphpad software, CA).

2.8.10. Confirmation of PCR product from patient samples

In order to confirm the identity of PCR products and specificity of the real-time PCR for HERV-K10 *gag1* the PCR reaction mix was run in a 2% agarose gel after amplification. The PCR product band was cut from the gel under long wave UV light

using a scalpel and the purified using the Minielute PCR purification kit (Qiagen, UK). DNA sequencing was performed by the Sequencing service (School of Life Sciences, University of Dundee, Scotland, www.dnaseq.co.uk) using Applied Biosystems Big-Dye version 3.1 chemistry, on an, 'Applied Biosystems model 3730', automated capillary DNA sequencer. Sequencing was carried out using gene specific primers unless otherwise stated.

2.9. HERV Microarray analysis

2.9.1. Sample RNA: Total RNA was extracted and purified from cultured and treated cells using the RNeasy RNA extraction kits (Qiagen, UK) according to manufacturer's instructions. Eluted RNA was checked for integrity by running in a 1% gel and quantified using a spectrophotometer.

2.9.2. Reverse transcription and PCR labelling: DNA-free mRNA (50 – 100ng) was reverse transcribed as described previously (section 2.8.5). CY3-labeled DNA probes were synthesised by PCR using primer mixes MOP-1 and MOP-2 as described by (Seifarth *et al*., 2003). Amplified DNA fragments were purified using MinElute biospin columns (Qiagen, UK), ethanol precipitated and air dried before being redissolved in 25 µl of buffer containing 3 x SSC (1 x SSC containing 0.15 M NaCl and 0.0015 M sodium citrate), 0.5 % sodium dodecyl sulfate, 50 % formamide and 50 mM sodium phosphate buffer (pH 7.4).

2.9.3. Mixed oligonucleotide primers (MOP) for multiplex PCR: For amplification of RT-related sequences, two primer cocktails, MOP-1 and MOP-2 were used. Primer sequences correspond to highly conserved regions present within the reverse transcriptase gene of all known human retroviruses and related animal viruses. All anti-sense primers were modified with Cy3 fluorochrome. Therefore all specific amplification products generated in the polymerase chain reaction were labelled with Cy3 at the 5' end. Set MOP-1 is a complex mixture of 16 primers, which preferentially amplify human and betaretroviral reverse transcriptase genes, whereas the MOP-2 primer mixture (12 primers) allows the amplification of human and mammalian gammaretroviral RT sequences and RT sequences of various human exogenous retroviruses, such as HIV, HTLV and Spuma viruses. For each MOP primer set, a

separate PCR reaction was performed. As an internal standard for the assessment of RNA quality, a mixed primer set for primer amplification of five housekeeping genes (HKG mixture: ubiquitin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), α-tubulin, β-Actin, hypoxanthine-phosphoribosyltransferase (HPRT) was designed. HKG primers were added to each MOP-1 and MOP-2 PCR reaction.

2.9.4. <u>Hybridisation procedure:</u> Hybridisation was carried out under the standardised conditions. All handling steps involving fluorochromes were carried out in dimmed light conditions. Prehybridisation of glass slides was performed in Falcon 50 ml tubes (Greiner Bio-One Ltd., Stroudwater, UK) in 6 x SSC, 0.5% sodium dodecyl sulfate (SDS), 1% bovine serum albumin (BSA) at 45ºC for at least 2 hours. Slides were then washed five times (1 min each) in water under agitation. Subsequently, slides were dried by a short centrifugation step (500 x g) for 1 min at RT). For hybridisation, a '22 x 22' mm glass lifter slip (Erie Scientific Co.) was taped to the microarray. Subsequently, 25 µl of the denatured (5 mins at 95 ºC) Cy3-labeled probe were applied to the small space between the coverslip and slide and sealed with Fixogum rubber cement (Marabu Gmbh, Tamm, Germany). Sealed slides were incubated in a humid chamber for 16 hours at 42 ºC. Hybridised slides were then washed three times without agitation (10 min each) at 22 ºC in 1 x SSC, 0.5 x SSC and 0.1 x SSC, respectively. After a brief centrifugation step (500 x g for 1 min at RT) the hybridised microarrays were stored in the dark at room temperature until scanning.

2.9.5. Multiplex PCR

Cy3-labelled samples were synthesised by PCR in a total volume of 50 µl using components and reaction conditions listed below (Table 2.5 and Table 2.6). Aliquots of amplification products were analysed on 2.5% TBE agarose gels stained with ethidium bromide. Amplification products of both MOP-1/HKG and MOP-2/HKG were combined and stored at -20 until hybridisation. Immediately before hybridisation step, Cy3-labelled DNA was ethanol precipitated, air dried and redissolved in 25µl of hybridisation buffer containing 3x SSC, 0.5% SDS, 50% formamide, 50mM sodium phosphate buffer and adjusted to pH 7.4.

Component	MOP-1/HKG (µl) (x1 reaction)	MOP-2/HKG (µl) (x1 reaction)	Final concentration
Nuclease free water	35	35	-
10x PCR buffer	5	5	1x
dNTP's	0.5	0.5	25mM
MOP-1 (β – retroviruses)	2	-	50µM
MOP-2 (γ – retroviruses)	-	2	50µM
HKG	1	1	50µM
$MgCl_2$	3	3	3mM
cDNA template	3	3	-
Enzyme	0.4	0.4	25mM
Total volume		**50**	

Table 2.5 Components of the HERV microarray amplification reaction mix

2.9.6. Chip scanning and visualisation

Microarrays were scanned using a GenePix 4100 (Molecular Devices, Sunnydale, CA) using recommended settings for Cy3 fluorochrome. High resolution images were saved as bitmap (*.BMP) file format (24-bit) and further analysed using GenePix Pro (version 6.0.1.22) and Photoshop CS (Adobe, San Jose, CA).

Cycle	Step	Repeats	Temp	Time (mins:secs)
1	1		94	00:30
	2	3x	45	03:00
	3		72	01:00
2	1		94	00:30
	2	30x	50	02:00
	3		72	01:00
3	4	1	72	10:00
4	5		4	ω

Table 2.6 Cycling parameters for the PCR amplification programme for synthesis of Cy3 labelled PCR products for HERV-specific microarray.

2.10. Transfection

2×10^4 cells (HEK-293 and human fibroblast-like synoviocytes) were seeded in 500µl of supplemented growth medium (see Appendix III for reagents and growth conditions) until 50-80 % confluent. For each sample 0.2-0.4µg DNA was mixed in 25µl Opti-MEM without serum (Invitrogen, UK). Between 0.5-5µl of lipofectamine (Invitrogen, UK) was diluted in 25µl Opti-MEM medium (without serum) before combining both aliquots. These were mixed gently before incubating for 15-45 min at room temperature. 0.2ml of transfection complexes were then added to cells mixing gently in addition to 0.2ml growth medium. These cells were then incubated for 12 hours at 37°C in a CO_2 incubator. After incubation, 0.4ml growth medium containing 2X normal serum concentration was added to the cells. Transgene activity was tested for 24-72 hours post transfection as appropriate for cell type and vector.

Chapter 3

3.0 Abbreviations

Amino acids (aa)
Antigen: antibody (Ag: Ab)
Artificial neural network (ANN)
Autoantigens (AuAgs)
Caprine arthritis encephalitis virus (CAEV)
Congenital heart block (CHB)
Complementarity determining regions (CDRs)
Cyclic citrullinated peptide (CCP)
Epitope spreading (ES)
Experimental allergic encephalomyelitis (EAE)
Glutamic acid decarboxylase (GAD65)
Gold standard (GS)
Human immunodeficiency virus (HIV)
Human T cell lymphotropic virus-I (HTLV-I)
Immunoglobulin (Ig)
Major histocompatibility complex (MHC)
Monoclonal rheumatoid factor antibody (mRFAB)
National centre for biotechnology information (NCBI)
Non-obese diabetic mice (NOD)
Position - specific iterative basic local alignment search tool (PSI-BLAST)
Position - specific iterative prediction (PSI-PRED)
Rheumatoid arthritis (RA)
Rheumatoid factor (RF)
Sjogrens syndrome (SS)
Sliding Window (SW)
Sub-sclerosing (SScl)
Systemic lupus erythematosus (SLE)

Chapter 3 – Bioinformatic analysis: An investigation of the potential for RA disease pathology via molecular mimicry with HERVs

3.1 Autoimmune diseases

Autoimmune diseases can be classified broadly based upon the distribution of their pathologies. Organ-specific autoimmunity is characterised by autoimmune reactivity that is directed against antigens found in a specific organ or tissue, e.g. autoimmune thyroiditis. Organ-nonspecific autoimmunity, by contrast, is characterised by a disease pathology that is not organ specific, but rather directed against ubiquitous cellular antigens, e.g. such as that observed in Systemic lupus erythematosus (SLE) (Routsias *et al*., 2004).

Rheumatoid Arthritis (RA) is one of the major systemic autoimmune diseases whose patients exhibit elements of both organ specific and systemic autoimmunity in its symptoms (Firestein, 2003). Despite a number of bacterial and viral pathogens having been implicated (Masuko-Hongo *et al*., 2003) an infectious trigger has thus far remained elusive. RA patients typically present a diverse array of autoantibodies, targeting a large number of autoantigens, found both systemically and locally within the joint (Corrigall and Panayi, 2002). The presence of many autoantibodies overlapping in a number of closely related syndromes, makes specific diagnosis difficult and it is this mix of auto-antibodies, specific for both systemic and cellular components, that makes the focussed nature of the immunopathology within the joint so intriguing (Griesmacher and Peichl, 2001).

The mapping of the targets of auto-antibodies is useful in the diagnosis of patients as diagnostic criteria or disease markers (Table 3.1) (Arnett *et al*., 1996, Tzioufas *et al*., 2000) and in the classification of clinical subsets (Kuwana *et al*., 1993). Elucidation of such targets also contribute to the understanding of how tolerance is lost the nature of subsequent mechanisms leading to this loss, i.e. molecular mimicry and epitope spreading (Albert and Inman, 1999, Powell and Black, 2001). Such knowledge can also aid the development of novel assays, incorporating synthetic peptides as antigens (Fournel and Muller, 2003) or vaccines (Chacon *et al*., 2003).

Disease	Autoantibody/ Autoantigen	Diagnostic/Prognostic implication	Reference
RA	**RF & CCP**	Poor prognosis & severe radiological damage	(Corrigall and Panayi, 2002) (Kroot *et al*., 2000)
SLE	**Ribosomal P protein**	Neurological complications	(Arnett *et al*., 1996), (Tzioufas *et al*., 2000)
	Anti-Ro/SSA	Indicative of subcutaneous lupus, or neonatal lupus in pregnant women	(Conrad and Schlosler, 2002)
	Anti-histone antibodies & Anti-dsDNA	Presence suggests drug-induced SLE Associated with nephritis & increased disease activity	(Shen *et al*., 1998) (Lefkowith *et al*., 1996) (Raman, 2002) (Rahman and Hiepe, 2002)
SS	**SS-A/Ro SS-B/La**	Link to CHB and neonatal lupus in pregnant mothers	(Scofield, 2004) (Tzioufas *et al*., 2002)
SScl	**DNA-Topoisomerase I**	Disease marker & associated with PIF.	(Rizou *et al*., 2000)

Table 3.1 Examples of autoantigens that feature important roles in the diagnosis of their respective disease. Their presence is important in the distinguishing of several disease subsets from their parent diseases, e.g. SLE and RA. SScl: Subacute Sclerosis. RF: Rheumatoid Factor. CHB: Congenital heart block. CCP: Cyclic citrullinated peptide.

3.1.1 Epitopes & Epitope mapping

Antibodies are proteins produced by B-cells in response to immunogenic substances such as viruses, vaccines and other allergens. B-cell or antibody-related epitopes are the molecular structures within the antigen that make specific contacts with residues of soluble and membrane bound antibody. There are several different types of epitopes; however, for the purposes of this investigation two broad distinctions will be made.

- **Linear (Continuous) epitopes.**
- **Conformational (Discontinuous) epitopes.**

Continuous epitopes typically consist of several consecutive amino acid residues, i.e. predicted directly from the protein's primary structure, whereas discontinuous epitopes refer to those of a more conformational or spatial nature that form when distant regions of a protein come together as protein folds and mature (Routsias *et al*., 2004).

Discontinuous epitope predictions rely upon analysis of secondary and tertiary structure of proteins. No single method is currently capable of identifying all epitopes on a protein; however, large numbers of techniques are available for the mapping and study of such regions. These range from the use of phage peptide libraries (Smith *et al*., 2004) to recombinant autoantigen expression in *in vivo* systems such as baculovirus and bacterial hosts, although each method has its own advantages and disadvantages (Schmitt and Papisch, 2002). The Pepscan methodology presents an alternative to recombinant technology, involving the synthesis of thousands of synthetic peptides adhered to polyethylene rods (Fournel and Muller, 2003) and despite it lacking the reproducibility problems associated with recombinant autoantigens, it can only be used to detect continuous epitopes and may be beyond the financial capabilities of some laboratories. It is here that the use of '*in silico*' epitope prediction methods in the design of synthetic peptides, presents a cost effective alternative. Techniques in *in silico* epitope mapping primarily use immunoinformatics tools – derived from the branch of bioinformatics, charged with the analysis of biological data, with reference to immunological processes (Flower, 2003). Immunoinformatics describes both T cell and B-cell orientated applications, ranging from the structural modelling of Major histocompatibility complex (MHC) (Schueler-Furman *et al*., 1998) to B-cell epitope mapping and modelling of immunoglobulin (Martin *et al*., 1989). The field, however, has been somewhat uneven in its coverage with the majority of the research focussed upon T-cell epitopes, whilst B-cell epitope analysis has been comparatively neglected (Flower, 2003). This T-cell bias has been largely due to the fact that B-cell epitope prediction is significantly complicated by the conformational aspects of antibody binding. As a result even the best predictions have fared only marginally better than random (Blythe and Flower, 2005).

The key to the identification and use of synthetic peptides in investigating protein antigenicity is therefore to predict those sequences that play key roles in the interactions with antibodies. Such mapping studies are important, as they provide information defining subsets of patients within disease groups (Shen *et al*., 1998, Routsias *et al*., 2004). Additionally, knowledge of sites where antibodies bind on proteins may suggest the mechanisms responsible for the generation of the antibodies themselves.

3.1.2 Mechanisms of Antibody: Antigen binding

The exact mechanisms at work governing Antigen: Antibody binding and recognition remain unknown despite intense scrutiny; however such answers are fundamental to our understanding of the workings of the mammalian immune system. At present most of the evidence for Antigen: Antibody (Ag: Ab) binding has been derived from the analysis of data sets obtained from experimental investigations. X-ray crystallography studies elucidating the structures of Ag: Ab complexes have provided valuable information of their molecular architecture and shed considerable light upon potential influencing factors, including the identities of contacting residues, the number of H-bonds formed and conformational changes occurring within both proteins. Additionally, residues at the Ab:Ag interface have been observed to be predominantly aromatic, i.e. Tyrosine and Tryptophan, with a notable absence of charged residues such as Aspartic acid, Glutamate and Lysine (Lo Conte *et al*., 1999). Increased levels of Arginine have also been reported, an observation attributed to its ability to form multiple bonds (Bogan and Thorn, 1998).

Antibody molecules themselves are homodimers composed of two identical polypeptide chains of approximately 450 amino acids (heavy chains – H) which are covalently linked to identical polypeptide chains of 250 amino acids (Light chains – L). Based upon amino acid sequences, H and L chains may be divided into N-terminal variable (V) and C-terminal portions. Each of the V_H and V_L domains have been broken up into three segments or loops that connect the β-strands and are highly variable in length and nature between different antibodies. These loops are called 'Complementarity determining regions' (CDRs).

Although the exact mechanisms at work in the process of molecular recognition are still unclear, it has become apparent that the antigen binds directly to CDRs; thereby these regions exert a major influence over binding site conformation. Of the 6 CDRs, the binding of as few as two can result in recognition (Decanniere *et al*., 1999). A number of properties governing normal protein: protein interactions also have influence over this interface with both antibodies and antigens demonstrating a high degree of shape and chemical complementarity. These include hydrophobicity, charged residues and side-chains, and cleft binding conformation (Lo Conte *et al*., 1999). The degree of geometric match between the 2 juxtaposed surfaces has also been reported to play an important role, although the matching observed between proteins indicates poorer shape

correlation in comparison to other protein: protein interactions (Zuurbier and Huijing, 1993). This has been suggested to be due to the lack of evolutionary optimisation compared to other protein: protein interactions. This may help explain the evolution of additional mechanisms such as affinity maturation (Jones and Thornton, 1996).

Generally the combining sites of antibodies recognising larger proteins tend to be relatively planar in comparison to other protein: protein binding surfaces (Jones and Thornton, 1996). Antibodies recognising smaller antigens such as peptides and DNA, tending to exhibit a groove antigen contacting surface, whilst haptens bind via distinct cavities (Rees *et al*., 1994). A common feature for the majority of anti-peptide antibodies studied has been that they allow peptides to adopt beta-turn motifs once in the binding site. This has been observed for type-I (Rini *et al*., 1992), Type-II (Stanfield *et al*., 1990) and multiple tight beta-turns (Garcia *et al*., 1992). The importance of water within the binding process is also apparent within the literature, emphasised through the fact that it is present in all but one of the published X-ray crystallography studies of Ag:Ab complexes (Muller *et al*., 1998). Its presence has been suggested to improve fit and neutralise impaired H-bond groups and also compensate for the lack of evolutionary optimisation between the Ag: Ab interface.

One final factor which aids the structural recognition of antigens is the changes that occur in the conformational structure of both antigen and antibody upon binding. The term 'induced fit' was first coined to describe the induction of an active state in an enzyme by a substrate molecule (Koshland, 1958). In terms of an Ag: Ab relationship, this theory implies that the binding reaction between the two induces conformational changes leading to a better complementarity or 'fit' between the two proteins. Changes in both proteins upon formation of the protein complex have been documented previously. These vary from the small movements within the CDR loops (<3Å), slight shifts in orientation of the V_L and V_H domains (equivalent to up to 3Å), to somewhat larger changes including CDR loop displacement (of up to 7Å), which have been described upon binding to DNA and peptides (Tormo *et al*., 1994). These movements have been suggested to help play a role in affinity maturation processes (Sundberg and Mariuzza, 2002). An alternative theory to the induced fit is that of 'conformational selection' although the exact mechanisms at work are still unclear (Berger *et al*., 1999).

3.1.3 B-cell epitope mapping

From previously published work four main assumptions were followed in this analysis.

- ***B-cell epitopes can contain sequential or non-sequential amino acids (Routsias et al., 2004)***

- ***B-cell epitopes on native proteins generally are composed of hydrophilic amino acids on the protein surface that are topographically accessible to membrane bound or free antibody (Scofield and Harley, 1991).***

- ***B-cell epitopes tend to be located in flexible regions of an immunogen and display site mobility (Ponnuswamy et al., 1980).***

- ***Complex proteins contain multiple overlapping B-cell epitopes, some of which are immunodominant (James and Harley, 1998).***

Epitope mapping predictions were constructed using bioinformatic tools that process information on certain physico-chemical properties of amino acid residues, i.e. hydrophilicity, polarity, etc. All of the tools employed algorithms that operated using the 'sliding window' principle. Predictions using combined algorithms available on the ExPASy server were made using the four physicochemical properties described previously (Section 2.4). These were also later combined with beta-turn analysis in an attempt to improve prediction accuracy. The performance of ExPASy was compared and contrasted with a newer prediction server – BCEPRED. This program had the advantage of choice of up to seven different algorithms including hydrophilicity, flexibility/mobility, accessibility, polarity, exposed surface and beta-turns, all using the sliding windows algorithms. In addition to incorporating secondary structure analysis, the tabular output of this method eliminated the requirement for output to be formatted manually, thereby saving time and resources.

A significant number of studies have highlighted that hydrophilicity, accessibility and flexibility are the best criteria for identifying these regions of antigenicity (Parker *et al*., 1986, Strynadka *et al*., 1988, Pellequer and Westhof, 1993, Hopp, 1994, Alix, 1999). Other properties cited as also being useful have included beta-turns (Pellequer and Westhof, 1993), antigenicity (Welling *et al*., 1985) and hydrophobicity (Beardsley *et al*., 2003).

3.1.3.1.1 Hydrophilicity

Hydrophilicity was the first criterion identified to correlate with antigenicity and its popularity has since been reflected in its common usage throughout the literature in the prediction of epitopes (Galbraith *et al*., 2000, Chacon *et al*., 2003, Beardsley *et al*., 2003). Hopp & Woods, (1981) derived a set of parameters to be used to predict hydrophilicity, based upon previously published solvent accessibility studies (Levitt, 1976). These were designed to identify exposed regions of mixed charge, i.e. polar, with few or no large hydrophobic residues. Typically, regions that are highly exposed with lesser charge were more likely to bind antibodies, compared to those regions which were highly exposed, but devoid of charge (Hopp, 1993).

A second data set of hydrophilicity values was published by Parker *et al*, (1986) who claimed increased prediction accuracy. These values were based on experimental retention times of the 20 amino acid residues in an HPLC reverse phase column. The accuracy of predictions was also reported to increase further when used in conjunction with accessibility and flexibility (Parker *et al*., 1986).

3.1.3.2 Flexibility

Flexibility or mobility has been cited in the literature to correlate better with antigenicity than other properties (Westhof *et al*., 1984, Karplus and Schulz, 1985). The main source of flexibility within a protein occurs about the Φ and Ψ bonds. Large residue side chains have a significant effect upon the flexibility of their respective residues and thus, the flexibility of an amino acid can have profound effects upon the protein. This was confirmed in further work by Vihinen *et al*, (1994) who suggested that flexibility played a crucial role in the binding, and allostery of proteins.

The ExPASy tool used the parameter index derived by Bhaskaran *et al* (1988), who had reformulated a previously published differential equation model used to predict residue flexibility. The old model could not account for fluctuations from the predicted values for each amino acid, due to the presence and behaviour of residues within secondary structures. The new flexibility index was developed by categorising residues based on their mode of activity within 2^0 structure and was normalised to a value between '0-1' (one being maximum flexibility, and 0 being indicative of no flexibility). The index was validated against 19 proteins of known structure.

3.1.3.3.1 Accessibility

Antigenic epitopes require good accessibility on protein surfaces in order to guarantee their recognition and interaction with antibodies and other immune cells. Therefore, the accessibility or positioning of a residue at the surface was the third criterion used for the selection of immunodominant regions. The bioinformatic tool on ExPASy operates on values derived from the x-ray and solvent accessibility analysis of the structures of 22 proteins. Parameters were then calculated using derived formulae. The resulting values represented the partition coefficient of each amino acid residue between the surface and the inside volume of the protein (Janin, 1979).

These formulae have since been used successfully in studies investigating antigenicity (Emini *et al*., 1985) and in numerous database/computer tools (Parker *et al*., 1986, Saha and Raghava, 2004).

3.1.3.3.2 Polarity

Studies of the accessibility of amino acid residues to solvents in several proteins have already confirmed the suggestion that those residues of increased polarity are most frequently found on the surface of proteins (Lee and Richards, 1971, Shrake and Rupley, 1973). Therefore, it seems logical that the more polar residues are likely to be located on the surfaces of proteins. This coupled with the fact that the majority of immunodominant sites are found on the surface of the proteins (Hopp, 1993), suggested that polarity would be a good indicator of surface epitopes. Long runs of charged and polar residues have been reported in autoantigens involved in rheumatic disease (Brendel *et al*., 1991), including CREST in CENP-B and DSc in Topoisomerase-I (D'Arpa *et al*., 1988). A recently published study (Saha and Raghava, 2004) also concluded that the combination of more criteria involved in the selection of epitopes, results in increased accuracy of predictions. For these reasons, polarity was identified as a fourth additional criterion, with which to use to predict surface epitopes.

Table 3.2 shows the classification of different amino acid residues based upon their polar nature.

Group	Amino Acid
Non-polar/ Hydrophobic	Val, Leu, Ile, Ala, Phe, Trp, Met, Tyr, Cys
Polar/ Hydrophilic	Ser, Thr, Lys, Arg, His, Asp, Glu, Asn, Gln
Charged	Gly, Pro

Table 3.2 Examples of amino acid categorisation of different amino acid residues based upon their polar properties [taken from (Krane and Raymer., 2002)].

3.1.4 The 'Sliding window' technique

All the algorithms used in this investigation to predict the physico-chemical properties of specific residues in the primary structure of a protein, were based upon the 'Sliding window' (SW) principle. The main concept behind this technique has been used previously for purposes ranging from electronics to marketing and weather forecasting with the only requirement being an experimentally derived data set for each property analysed. After a window size has been determined (default size of 9) the window progresses along the length of the query sequence. The amino acid at the centre of each window has its value calculated by assigning it the average of all parameter values of residues within the window (Figure 3.1). Despite an accuracy level of ~60%, and a low level of sensitivity and specificity the sliding windows main strength is its robustness (Claverie and Notredame, 2006). It is for this reason that the majority of the computer-based methods, available on the internet, for elucidating primary structure use this principle behind their calculations.

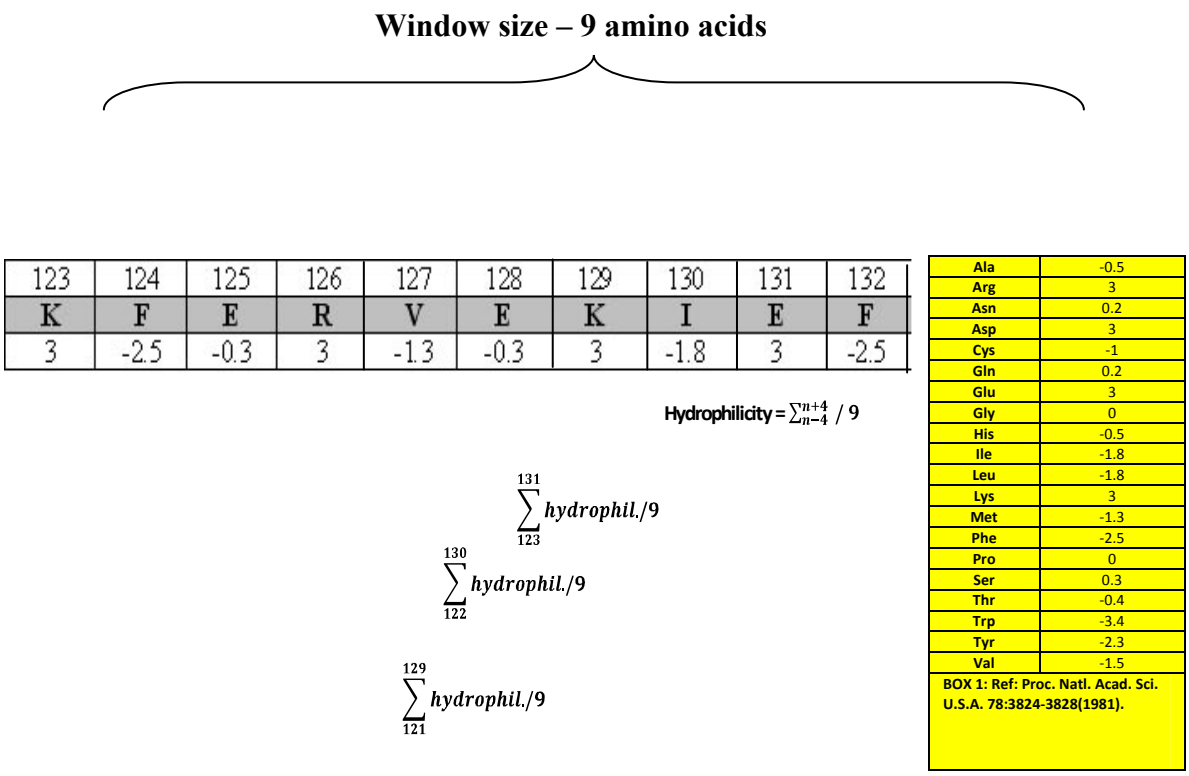


Figure.3.1 The mechanism of the 'sliding windows' mechanism in action. Box 1 (right/yellow) contains data set for amino acid hydrophilicity values determined by experimental investigation. A default window of 9 amino acids is selected for analysis. Values for hydrophilicity (Box 1) are applied to the relevant amino acid residues within the window. An average of all 9 hydrophilicity values are calculated and assigned to the residues at the centre of the window, i.e. Valine at position 127. The window then advances one position with the average of all in the window is again assigned to the amino acid at its centre, this time glutamic acid at position 128 (Claverie and Notredame, 2006).

3.1.5 Importance of threshold values

It has been suggested that each globular protein contains a finite number of epitopes, each confined to a discete, yet accessible area and pre-determined by overall protein conformation. This has been shown to be incorrect, with the majority of the surface on some globular proteins being potentially antigenic in nature (Newman, 1982). It has thus become important in epitope mapping studies to be able to distinguish between major and minor 'false' epitopes. Threshold values therefore are not just critical to the speed of the analysis but play a central role in the accuracy and specificity of the predictions. The higher the threshold values, the fewer epitopes predicted and the more stringent the selection conditions, however this may result in the missing of some less antigenic epitopes. The lower the threshold values, the less stringent the selection conditions, however this results in a large increase in 'false positive' epitopes, i.e. those of limited antigenicity that play little role in the resulting immune response. Threshold values were therefore determined in order to distinguish

between the small subset of those epitopes that were likely to be with the immunodominant immune response.

3.1.6 Epitope Spreading

Threshold values play an important role in the process of epitope prediction through the identification and distinction of primary epitopes from secondary epitopes which may play lesser (if any) roles in disease pathogenesis. The term 'epitope spreading' (ES) was first coined in the early 1990's to describe the ability of the B and T-cell immune response to diversify at the level of specificity from a single determinant to many sites on the autoantigen (Lehmann *et al*., 1992). Such diversity would be advantageous for any immune system, which has evolved to efficiently attack as many targets as possible. The mechanism describes the development of immune responses against endogenous epitopes, secondary to the release of self antigen during either chronic autoimmune or infectious responses. These responses can be located either upon the same (intramolecular) or upon different proteins (Intermolecular) with secondary epitopes often being cryptic or hidden epitopes located upon the same protein or as dominant epitopes on neighbouring molecules (Powell and Black, 2001). Such a lack of tolerance to self-antigens in immuno-priveledged sites has been reported through occurrence of autoantibodies to sperm antigens and ocular lens crystalline after tissue trauma. This phenomenon may play a key role in the normal response against infectious agents and have been described previously during viral infection (Milich *et al*., 1987). During HBV infection it has been reported that the priming of murine helper T cells can drive the production of antibodies targeting both HBcAg (core) and HBsAg (surface) providing primed mice are boosted with intact virus. When mice were primed with separate antigen preparations, ES did not occur suggesting that both proteins must be associated for the phenomenon to occur. Additionally, it may account for the natural recovery from acute autoimmune phenomena, resulting in an amelioration of symptoms as observed in Lewis rats and adjuvant arthritis (Moudgil, 1998).

Studies in two animal models of multiple sclerosis, experimental autoimmune encephalomyelitis (EAE) and Theiler's murine encephalalitogenic induced demyelinating disease (TMEV-IDD) have shown conclusively that epitope spreading plays a pathological role in ongoing disease and that by blocking this process by induction of tolerance to myelin epitopes or blocking T cell costimulation, this blocks

ongoing clinical disease. ES with the murine model of relapsing experimental autoimmune encephalomyelitis (R-EAE) being one well characterised example. Other models include that observed in non obese diabetic mice (NOD) – a mouse model for Type-1 diabetes, which develop immune responses to GAD65 (a 65kDa isoform of glutamic acid decarboxylase), also develop responses also targeting pancreatic islet cells (Tisch et al., 1993). Induction of tolerance to GAD65, leads to the amelioration of the autoreactive response to pancreatic β cell antigens (Tian et al., 1996). Evidence for the incidence of ES is also evident in human diseases, notably involving clusters of target molecules such as those forming macromolecular complexes. Ro60 is one such molecule which was shown through the immunisation of rabbits to induce antibodies specific for multiple epitopes on both Ro60 and La (a macromolecular complex which frequently binds to Ro60). Additionally, antibodies that were cross reactive with epitopes on other common spliceosomal proteins were also observed (Smets and Sokal, 2002) suggesting that a loss of tolerance to a single epitope could induce an autoimmune response involving epitopes on other proteins potentially leading to immune responses typically observed in patients with rheumatic diseases. Other studies have also reported mechanisms which may result in the occurrence of such phenomena which include a conserved 'driver' epitope or motif which may act as an initiator. Although ES has been observed to involve both B and T cell components, the simultaneous involvement of both has not yet been reported (James and Harley, 1998).

Many investigators of ES have focussed upon lupus, a disease involving a large number of auto-antibodies, to self components including those of the spliceosome and nucleosome (McClain *et al*., 2002). One spliceosome component in particular – the Sm-B/B' protein, possesses two sequences which have generated much interest – PPPGMRPP & PPPGIRGP. These motifs were highlighted by studies using the MRL lpr/lpr murine model (Williams *et al*., 1990) and by patient sera containing anti-Sm autoantibodies (James and Harley, 1992). Studies of several serial lupus patient sera over several years showed that these proline-rich motifs were the first epitopes to which patients mounted a response. Over time these responses diversified significantly, with one female patients sera eventually showing reactivity with 17 different groups of peptides from Sm-B/B' (James *et al*., 1995). The same phenomenon has been observed in rabbits immunised with the PPPGMRPP peptides. These rabbits' sera were cross-reactive with the PPPGIRGP peptides, in addition to up to 74 others (many of these

being to the same regions of Sm-B/B' as that seen in humans). Similar results were also observed in several mouse models (James and Harley, 1998).

3.1.7 Secondary structure: The significance of Beta-turns

Knowledge of secondary structure can be central to understanding a complex protein structure, in addition to further validating initial primary structure predictions (Creighton, 1992). The secondary structure of a protein describes the assignment of helices and sheets to the primary structure (produced by the hydrogen-bonding patterns of the main chain). This generally includes:

- The alpha helices.
- Beta-turns and beta strands.
- Coiled coil regions.

As proteins fold upon themselves, assuming their native structures, this results in the formation of small groups of amino acids producing patterns with one another on a local scale. It is the location and direction of these regular structures that make up a protein's secondary structure. This process is governed by a number of factors including the different side chains on each amino acid and their effects, in terms of steric collision, polarity, and the formation of H-bonds between different residues.

Secondary structure elements such as alpha helices and beta-sheets provide the rigid and stable framework of the protein; however Beta-turns have been deemed important in the prediction of antigenicity. Beta-turns were originally identified in model building studies (Venkatachalam, 1968) and tend to consist of 4 consecutive residues (i, i+1, i+2 and i+3) that are not present within the alpha-helix. They share a difference, between i and i+3, of <7Å (Rose *et al*., 1985) with turns resulting in a change in the direction of the polypeptide chain of up

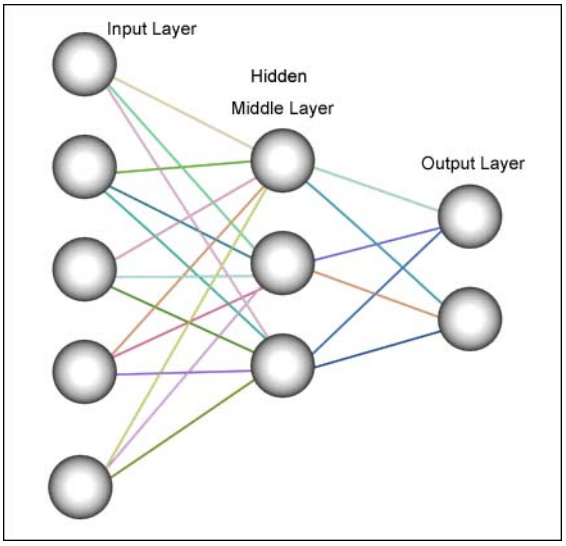


Figure 3.2. The general structure of an Artificial neural network. The network is made up of distinct layers composed of a series of nodes within each layer. The structure is designed to be analogous to that of the neurons within the human brain. Information is passed between processing nodes. As PSIPRED is a forward feed network the movement of the information is always forward through the network. Diagram taken from (Claverie and Notredame, 2006).

to 180° (Chou and Fasman, 1974). This reversal may also be accompanied by the formation of an H-bond between the carboxyl and amino groups, between the i and i+3 residues (Nemethy and Scheraga, 1980).

These turns play important roles in the globular structure of a protein in terms of structural and functional roles, tending to be involved in the formation of highly compact structures in proteins as well as acting as binding sites for their ligands (Dong *et al*., 2003). As a result, they tend to occur at exposed surfaces, thereby representing antigenic sites or those involved in mediating transient bonding with other proteins (Branden and Tooze., 1999). These are one of the fundamental properties of an ideal epitope (Alix, 1999) and exist as U-like structures, formed when a beta strand reverses direction in an anti-parallel beta-sheet. Additionally, Beta-turns were selected as they have been previously reported to occur on or adjacent to antigenic regions (Chacon *et al*., 2003).

An Artificial neural network (ANN) was recruited in order to reliably predict the location of Beta-turns in protein autoantigens, important in RA. Artificial neural networks are information processing paradigms, whose structure has been inspired by the structure of the mammalian brain. Essentially ANNs are a collection of interconnected mathematical models designed to emulate a biological synapse network. This results in a large number of interconnected processing elements analogous to neurons, tied together with weighted connections analogous to synapses. ANNs also show the ability to learn and are trained to identify patterns in their output that follow from those within their 'training' data set.

ANNs operate using two feed-forward neural networks, each with a single hidden layer. These two networks operate independently, but in succession for each turn to be predicted (Figure 3.2). The input into this first network is the position specific matrices generated by the Position-Specific iterative Basic local alignment search tool (PSI-BLAST) analysis (Altschul *et al*., 1997). This information is then used to predict location of Beta-turns in the amino acid sequence. The data produced by the first network is validated using the second network. It does this by comparing the output from the first network with the secondary structure predictions produced by Position Specific Iterative Prediction (PSI-PRED) (Jones, 1999). This further validates the output, increasing its potential accuracy. PSI-PRED is a publicly available protein secondary structure prediction programme, located at (http://bioinf.cs.ucl.ac.uk/psipred/). PSIPRED is used to detect distant homologues of a

query sequence, and thus generate a position specific scoring matrix as part of the prediction process (Jones, 1999) and it is this matrix which acts as a direct input into the networks. The BetaPred2 is an online resource available for the prediction of beta-turns (http://www.imtech.res.in/raghava/betatpred2/). BetaPred2 consists of 2 forward-feed networks (describing the direction in which information travels through the network), with a standard back propagation learning algorithm which acts to minimise the difference between the actual output of the network and the desired output for patterns in the training set. The training of the database was performed on a set of 426 non-homologous protein chains by 7-fold cross-validation. The first network is responsible for sequence to structure conversion and trained in the occurrence of various residues in a specific size of sliding window (9 residues), correlated with the output of the central residue. The predicted turn/non-turn output of the first net is then incorporated along with predicted secondary structure information in the second filtered structure-to structure network. This is trained to filter the unreasonably isolated residues predicted as turns. Both networks have single output units, i.e. 1 for turns, 0 for non-turns. Input into the networks occurs in the form of multiple protein alignments. Predictions from these alignments, rather than a single sequence improves prediction accuracy (Cuff and Barton, 1999). These are submitted as results using the PSI-BLAST server.

3.1.8 Alignment of epitopes

Once the epitopes for both retroviruses and putative autoantigens had been identified, they were aligned to identify sequence homology and thus potential molecular mimicry. The LALIGN programme (available publicly from http://www.ch.embnet.org/software/LALIGN_form.html) was used as it was able to show results from local alignments between two specific input sequences. LALIGN passed query sequences through a substitution matrix such as BLOSUM 50 and aligned peptide sequences in a manner in which the conservation score was highest.

3.1.9 The potential for molecular mimicry between HERVs & Autoantigens in Rheumatoid Arthritis

Little is known surrounding the origins of RA and its underlying causes although it has been suggested that both environmental and genetic elements may be significant contributors to triggering the resultant immunopathology. Such a response could evolve to encompass other host proteins through mechanisms such as epitope spreading and molecular mimicry, resulting in disease pathology. Through the identification of regions of homology between the primary structure of HERVs and autoantigens involved in disease pathogenesis, this may help identify early events in disease development prior to the subsequent autoimmune response and clinical disease. After optimising the novel bioinformatic technique of epitope prediction, it was applied to map epitopes on 13 different families of retroviruses (Table 3.3) and 26 autoantigens (Table 3.4), all previously associated with RA disease pathogenesis. Two additional non-human 'control' sequences were included in the analysis in order to prevent bias within the results.

Virus	Protein	Accession number	Reference
HERV-K10	Gag1 & 2/ Pol/ Env	M14123	(Ono *et al*., 1986)
HERV-K (T47D)	Gag1/ Gag2/ Pol/ Env	AF020092	(Seifarth *et al*., 1998)
HERV-E (4-1)	Gag1/ Gag2/ Pol/ Env	M10976	(Repaske *et al*., 1985)
ERV3	Pol/ Env	M12140	(Cohen *et al*., 1985)
ERV9	-	X57147	(La Mantia *et al*., 1991)
HERV-L	-	X89211	(Tristem, 2000)
HRES-1	P15/ P25	X16660	(Bengtsson *et al*., 1996)
HIV	Gag-Pol/ Pro/ Int/ RT/ Pol	NP_057849	(Coffin J. M. *et al*., 1999)
HTLV-I	GP1/ Tax/ Rex/ Env	NC_001436	(Coffin J. M. *et al*., 1999)
HRV-5	Gag/ Pro/ Pol	AAM81188	(Griffiths *et al*., 1999)
CAEV	Gag/ Pol/ Env/ S/ Q	NP_040938/ NP_040939/ NP_040940/	(Saltarelli *et al*., 1990)
Controls			
Arabidopsis thaliana	Reverse Transcriptase	AAL06422	(Wright and Voytas, 2002)
Arabidopsis thaliana	Gypsy/ Ty3 element	AAG51464	GenScan+**

Table 3.3 HERV families implicated/associated with RA pathogenesis within the literature. These families were analysed using immuno-informatic techniques outlined in Section 3.3. Reference sequences for all viruses were identified and extracted from the NCBI sequence database.

**** - GeneScan+ is a gene prediction programme running at http://CCR-081.mit.edu/GENSCAN.html. Two potential non-human negative controls were included from plants.**

HIV: Human Immunodeficiency virus. HTLV-I: Human T cell lymphotrophic virus-I. CAEV: Caprine arthritis encephalitis virus.

Autoantigen	Accession number	Reference
Collagen II	NP_001835	(Wang *et al*., 2002)
Aggrecan	P16112	(Zou *et al*., 2003)
Cartilage GP-39	P36222	(Steenbakkers *et al*., 2003)
Filaggrin	XP_048104	(Simon *et al*., 1993)
Sa	AAH53370	(Goldbach-Mansky *et al*., 2000)
Collagen I	NP_000079	(Wang *et al*., 2002)
Glucose Phosphate Isomerase (GPI)	P06744	(Herve *et al*., 2003)
Biglycan	NP_001702	(Polgar *et al*., 2003)
Decorin	P_598013	(Polgar *et al*., 2003)
MMP1	NP_002412	(ter Steege *et al*., 2003)
La	P08865	(McNeilage *et al*., 1990)
DEK	NP_003463	(Adams *et al*., 2003)
LRP-2	NP_004516	(Ooka *et al*., 2003)
Heat Shock protein 90 (HSP90)	NP_005339	(Minota *et al*., 1988)
PADI 4	Q9UM07	(Suzuki *et al*., 2003)
hRNPA2	NP_002128	(Fritsch *et al*., 2002)
hRNP A1	NP_112533	(Fritsch *et al*., 2002)
Calpastatin	NP_775083/	(Mimori and Tanaka, 2002)
Chondroitin Sulphate	O14594	(Chang *et al*., 2005)
BiP	Q9HG01	(Bodman-Smith *et al*., 2003)
YKL-39	Q15782	(Sekine *et al*., 2001)
RA-A47	NP_001226	(Hattori *et al*., 2000)
Jo-1	NP_002100	(Routsias *et al*., 2004)
Ro52/SSA	NP_003132	(Eggleton *et al*., 2000)
Nucleolin	NP_003132	(Hiscox, 2002)
Calreticulin	NP_004334	(Itoh and Reichlin, 1992)

Table 3.4 Host autoantigens implicated in rheumatic diseases that were included in this epitope mapping study. All sequences used for putative autoantigens were registered reference sequences on the NCBI public database. All autoantigens within the table were epitope mapped, before epitope and whole protein sequences were aligned with virus sequences listed in Table 3.3.

GP-39: Glycoprotein 39kDa; MMP-1: Matrix metalloproteinase 1; LRP-2: Low Density lipoprotein-related protein 2; PADI-4: Protein-Arginine deiminase type-4; hRNPA2: Heterogenous nuclear ribonucleoprotein A2; hRNPA1: Heterogenous nuclear riboprotein A1; BiP: Immunoglobulin heavy chain-binding protein; YKL-39: Chitinase 3-like protein 2 precursor; RA-A47: Rheumatoid Arthritis related antigen 47kDa.

3.1.10 Aims and Objectives

The aims of this chapter were:

- To determine the success rates of the current process of epitope prediction, and to adapt this in order to maximise this predictive accuracy.

- To use this novel optimised methodology to predict potential targets of autoantibodies which may be important in the early events of RA disease immunopathology.

- To correlate these epitopes with those regions of shared sequence between both groups

- To short-list those epitopes that would be the most suitable candidates for molecular mimicry between HERVs and autoantigens, for synthesis as short peptides for incorporation into an ELISA for screening patient blood.

3.2 Methods

3.2.1 *In silico* analysis

Reference sequences of potential autoantigens were identified in rheumatic diseases and their protein sequence was extracted from the NCBI/ Genbank online database (www.ncbi.nlm.nih/). Sequences were edited accordingly using a standard text editor.

3.2.2 Primary and Secondary structure analysis

Protein residues were analysed via *in silico* analysis of physicochemical parameters, using normalised scales derived from previous experimental data, in order to elucidate the antigenic structure of each autoantigen. Predictions were made using computerised algorithms for hydrophilicity (Hopp and Woods 1981), residue polarity (Grantham 1974), solvent accessibility (Janin 1979) and flexibility index (Bhaskaran and Ponnuswamy 1988). These analyses were performed with the bio-computing software programmes (available from http://www.expasy.ch/cgi-bin/protscale.pl). Window lengths on these programs were set to '9 residues' to mirror the length of the peptide being produced. The top 25% of peaks for each protein were considered to be over the threshold level (Parker *et al*, 1986) and therefore the antigenicity of each protein was assessed on an individual basis. Where four or more properties (out of five) were above the threshold, the residue was considered antigenic. All studies were run in parallel with, and results compared to, a previously published B-cell epitope prediction algorithm called BCEPRED (http://www.imtech.res.in/raghava/bcepred/) that has been reported to have a prediction accuracy of 80% (Saha and Raghava, 2004). Recommended defaults were used with all predictions made using this database.

Protein secondary structure analysis including the elucidation and mapping of alpha helices and Beta-turns was examined using the neural network tool Betatpred2 server (http://www.imtech.res.in/raghava/betatpred2/) (Kaur and Raghava 2003) and the PSIPRED tool (McGuffin *et al*. 2000) (http://globin.bio.warwick.ac.uk/psipred/). The PSIPRED program had repeatedly scored highly in prediction accuracy using the CASP4 and EVA projects (http://cubic.bioc.columbia.edu/eva.html). All alignments

between the HERV and autoantigen epitopes were performed using the online tool LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html). Peptides were chosen to be synthesized on the basis of levels of antigenicity and sequence homology to autoantigens implicated in RA. The best candidate peptides for involvement in molecular mimicry were selected.

3.2.3 Molecular modelling

All proposed peptides were visualized using the Deep View Swiss Pdb Viewer (vers.3.7 – SP5) (http://www.expasy.org/spdbv/). This program was used at default settings, with all protein sequences inputted in FASTA format.

Figure 3.3 summarises the process of selecting potential epitopes from a database of both nucleotide and protein sequences stored on the NCBI database.

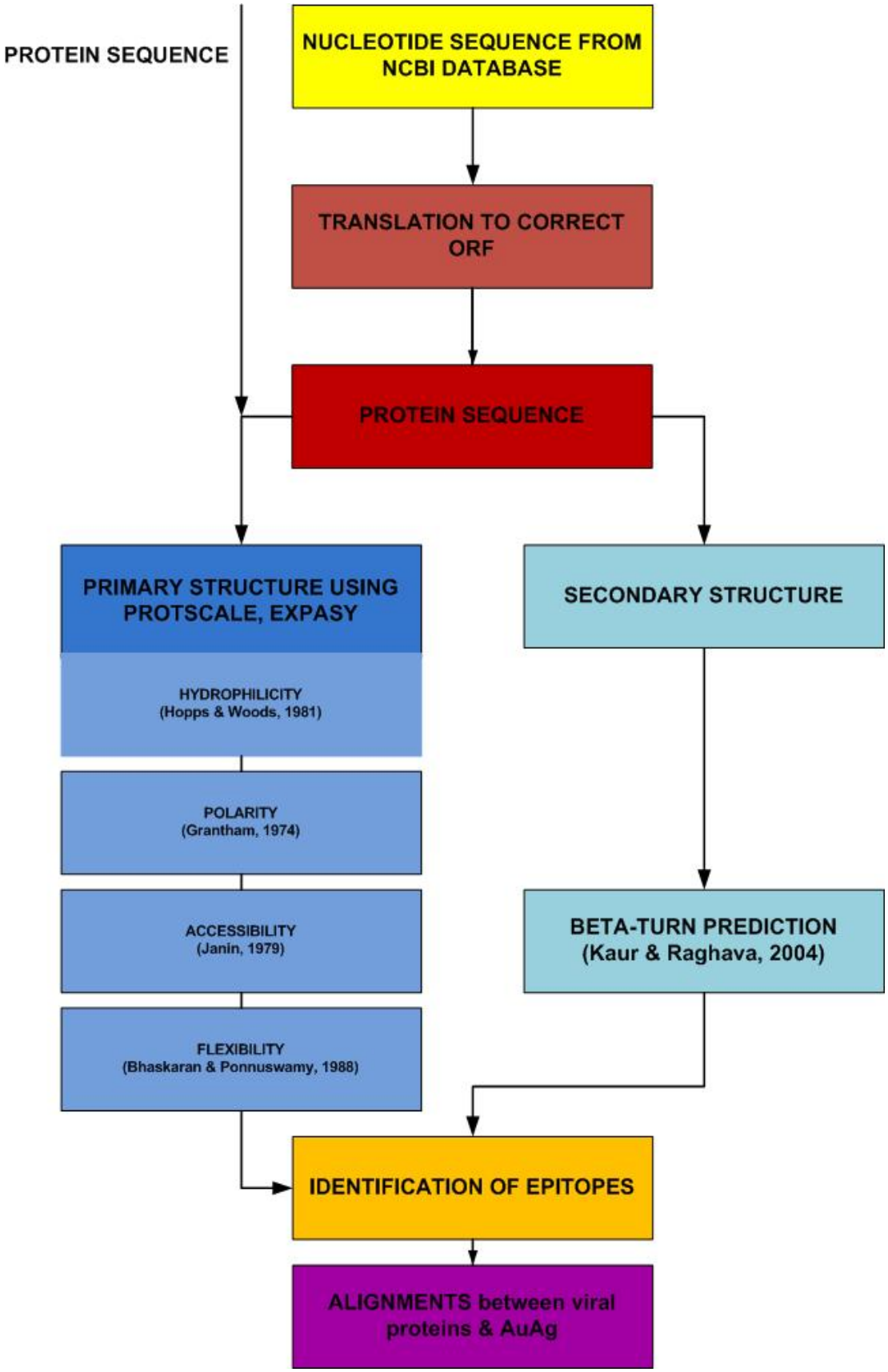


Figure 3.3 The main steps involved in the process of epitope mapping of viral and host proteins, and determining any regions of homologous sequence shared between them. Steps 1-3 (nucleotide sequence, translation of nucleotide sequence and determination of ORF are not essential steps if the protein sequence/Open Reading Frame have been experimentally determined.

3.3 Results

3.3.1 Selection of static or dynamic threshold values for ExPASy analysis

The retrospective prediction of known epitopes of autoantigens involved in RA disease pathogenesis was included within the auspices of this study as a measure of independent validation of the approach described here. 7 autoantigens involved in rheumatic disease were selected at random. A review of the literature was also carried out compiling all experimentally derived continuous epitopes published previously on these 7 autoantigens. These 21 experimental derived epitopes formed a data set of Gold Standards (GS) that were subsequently used to optimise and assess the performance and accuracy of both ExPASy and BCEPRED methodologies. Table 3.5 shows details for each GS epitope identified within the literature that formed part of this dataset.

Epitope	Autoantigen	Accession	Disease	Details	Reference
1	Calreticulin	AAB51176	RA	-	Eggleton *et al*., 2000
2				6/16 SLE	
3				11/16 SLE	
4				10/16 SLE	
5	CENP	NP_001800	RA	Reactivity with > 85% anti-ACA/ 17/19 patient sera	Muro *et al*., 1996
6				Pepscan showed 18/9 patient sera	Mahler *et al*., 2000
7	60 Ro/SS-A	NP_004591	SLE	Reacted with anti-Ro monospecific sera	Routsias *et al*., 1996
8				Immunodominant profile/ detected preferentially by SS	
9				Octa-peptides showed reactivity to anti-Ro/SSA serum.	Scofield and Hurley, 1991
10	La/SSB	NP_003133	SLE	> 70% of 122 sera samples reactive	Yiannaki *et al*., 1998
11					
12					
13				>90% of 122 sera samples reactive	
14				Most reactive to anti-SSB/La serum	Kohsaka *et al*., 1990
15				Reactivity varied – may contain several epitopes	
16	Ribosomal P protein	NP_000995	SLE	10-20% SLE serum	Elkon *et al*., 1986
17	PM/Scl-100	CAA46904	Sclerosis	Peptide based mutational analysis	Bluthner *et al*., 2000
18	Topoisomerase-1	NP_003277	SScl	Lie in proximity to beta-turns/ reactive against 81 patient sera	Rizou *et al*., 2000
19					
20					
21					

Table 3.5 Details and references of those experimentally derived epitopes selected as gold standards for the grading and validation of B cell epitope prediction methodologies outlined within this chapter. RA: Rheumatoid Arthritis; SLE: Systemic lupus erythematosus; SScl: Sub-acute sclerosis.

Using this dataset, the epitope prediction methodology was optimised for maximum accuracy. The first step in this optimisation that was considered was investigating the role of threshold values within the prediction framework and their impact upon resultant prediction output. Using 4 physicochemical properties the prediction success rates using static threshold values (Table 3.6) and dynamic thresholds (highest 25% of residues) were compared. As the results show (Table 3.7) the methodology using static threshold values performed poorly in contrast to that with the dynamic threshold. Static values achieved just 14.3% success in predicting experimentally derived epitopes in comparison to the dynamic methodology, which correctly predicted 41%. From this analysis it was concluded that dynamic threshold values would achieve greater accuracy in final epitope predictions.

Physico-chemical property	Original threshold value
Hydrophilicity (Hopp & Woods)	1.5
Polarity (Grantham)	10
Accessibility (Janin)	7
Flexibility (Ponnuswamy & Bhaskaran)	0.5

Table 3.6 Static threshold values used in initial/early epitope mapping studies. These thresholds were the 'default' values reported by those experimental studies that first determined the data sets. References for studies: Hydrophilicity (Hopp & Woods, 1982), Polarity (Grantham 1974), Accessibility (Janin 1979) and Flexibility (Ponnuswamy & Bhaskaran 1988).

Epitope	Published epitope	ExPASy Static	ExPASy Dynamic
1	^{34}LSSGKFYGDEEKDKG^{48}	×	×
2	^{142}KDIRCKDDEFTHLYT^{156}		×
3	^{193}DPDASKPEDWDERAK^{207}	×	×
4	^{253}EYKGEWKPRQIDNPDYKGTWIHPEIDNPEY^{282}		
5	^{3}PRRRSRKPEAPRRRS^{17}		×
6	^{22}PTPGPSRRGPSLGASSH^{38}		
7	^{169}TKYKQRNGWSHKDLLRSHLKP^{190}		
8	^{211}ELYKEKALSVETEKLLKYLEAV^{232}		×
9	^{485}EYRKKMDI^{492}		×
10	^{147}HKAFKGSI^{154}		
11	^{291}NGNLQLRNKEVT^{302}		
12	^{301}VTWEVLEGEVEKEALKKI^{318}		
13	^{349}GSGKGKVQFQGKKTKF^{364}		
14	^{88}KIRRSPSKPLPEVT^{101}		×
15	^{283}LGKDANNGNLQLRNRNKEVTWEV-LEGEVEKEALKKIIEDQQESLNKWKSK-GRRFKG^{338}		×
16	^{90} KKEEKKEESEEEDEDMGFGLFD^{111}	×	×
17	^{231}LDVPPALADFIHQQR^{245}		
18	^{205}WWEEERYPEGIKWKFLEHKG^{224}		
19	^{349}RIANFKIEPPGLFRGRGNHP^{368}		
20	^{397}PGHKWKEVRHDNKVTWLVSW^{416}		
21	^{517}ELDGQEYVVEFDFLGKDSIR^{536}		
	Total successfully predicted (out of 21)	**3**	**9**
	% Prediction success (Sensitivity) (successful predictions / total no. published epitopes)	**14.3% (3/21)**	**43% (9/21)**

Table 3.7 The impact of incorporating dynamic threshold values into the epitope prediction process. The first column (static) shows the success of predicting GS epitopes when using static threshold values, i.e. default values listed in Table 3.6. The second column shows the success of predicting GS epitopes when using dynamic thresholds, i.e. the highest 25% of values. Analysis was carried out using just four physico-chemical properties – Hydrophilicity (Hopp & Woods, 1982), Polarity (Grantham 1974), Accessibility (Janin 1979) and Flexibility (Ponnuswamy & Bhaskaran 1988). X = Match between predicted and published/gold standard epitope. Blank boxes indicate no match between the predicted peptides in this investigation and those previously published within the literature.

3.3.2 Optimisation of dynamic threshold values for ExPASy analysis

The next step was to determine the best threshold values, which would result in the highest sensitivity and specificity. High threshold values would be stringent, but may result in 'missing' an epitope. Lower values would be more promiscuous, resulting in the prediction of more false positives, leading to the methodology becoming more time and labour intensive. Using the reactivity of a number of rheumatoid factor monoclonal antibodies specific for the Immunoglobulin (Ig) (IgG1) Fc region (Section 3.2.3), a new dataset of experimentally defined epitopes was utilised (*unpublished data*). In addition to the 4 physico-chemical properties of primary structure, the presence of beta-turns was also included within this analysis. Figure 3.4 shows a graphical breakdown of how residues were considered under or over thresholds with threshold values of 25%, 30% and 40% investigated. Figure 3.4A shows the hydrophilicity profile of IgG1 Fc region. Figures 3.4B-D show how the top 25-40% of residues in the plot was selected for further analysis.

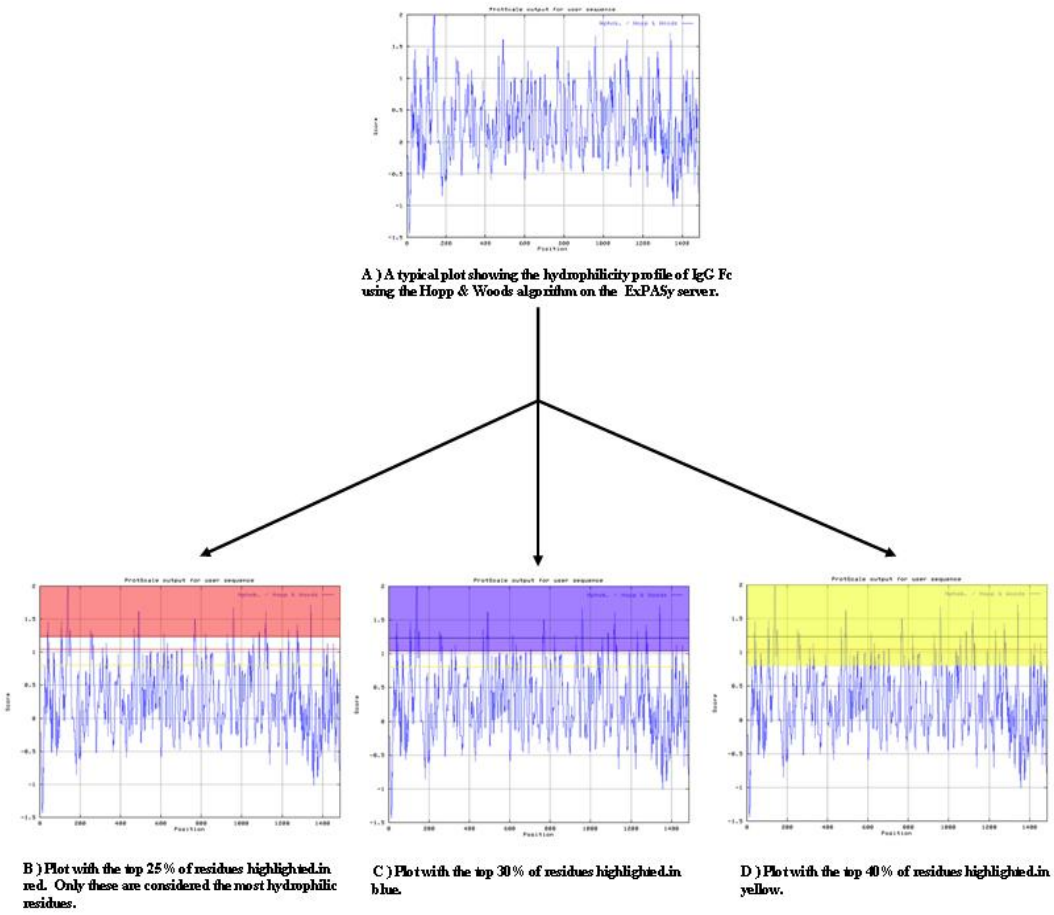


Figure 3.4 The principles of selecting the top percentage of residues when using dynamic threshold values. Plot A shows a typical plot produced by the hydrophilicity algorithm available at http://www.expasy.au. Once obtained a threshold is chosen by selecting the top 25% (Plot B), 30% (Plot C) or 40% (Plot D). Once the threshold was been set, only those residues within the shaded areas were considered to be important for analysis, i.e is hydrophilic in nature.

Due to the relatively small size of the IgG1 Fc region, i.e. 223 amino acids, there are relatively few additional epitopes predicted by the 40% threshold, in comparison to the lower 25% (Table 3.8). The higher threshold (40%) resulted in 6 additionally predicted epitopes. This effect was protein-specific as analyses of larger proteins such as Filaggrin (4061 aa) have shown. Epitope mapping of Filaggrin resulted in 25 epitopes (using 25%), which increased to 80 epitopes (using a threshold of 40%) (*Data not shown*).

Location of fragment	Peptide sequence	No. Prop. over threshold	Method Anal.
25%			
290-304 (295-301)	VHNAKTKPREEQYNS	5	**ExPASy**
337-351 (340-347)	PIEKTISKAKGQPRE	5	
355-370 (359-366)	YTLPPSRDELTKNQVS	4	
386-401 (389-397)	EWESNGQPENNYKTTP	4	
269-283 (272-279)	VVDVSHEDPEVKFNW	4	
	Total (5/4)	5 (2/3)	
30%			
292-306 (295-303)	NAKTKPREEQYNSTY	5	**ExPASy**
339-353 (342-350)	EKTISKAKGQPREPQ	5	
269-283 (272-280)	VVDVSHEDPEVKFNW	5	
317-331 (320-328)	QDWLNGKEYKCKVSN	5	
387-401 (390-398)	WESNGQPENNYKTTP	5	
356-369 (359-366)	TLPPSRDELTKNQV	4	
413-427 (416-424)	YSKLTVDKSRWQQGN	4	
246-260 (249-257)	VFLFPPKPKDTLMIS	4	
	Total (5/4)	8 (5/3)	
40%			
292-306 (295-304)	NAKTKPREEQYNSTY	5	**ExPASy**
339-353 (342-350)	EKTISKAKGQPREPQ	5	
268-282 (272-279)	VVVDVSHEDPEVKFN	5	
413-427 (416-424)	YSKLTVDKSRWQQGN	5	
317-331 (320-328)	QDWLNGKEYKCKVSN	5	
386-400 (388-398)	EWESNGQPENNYKTT	5	
355-369 (358-366)	YTLPPSRDELTKNQV	4	
330-344 (333-341)	SNKALPAPIEKTISK	4	
397-411 (402-406)	YKTTPPVLDSDGPFF	4	
439-449 (444-449)	HNHYTQKSLSLSPGK	4	
247-261 (248-257)	FLFPPKPKDTLMISR	4	
	Total (5/4)	11 (6/5)	

Table 3.8 Differences in epitopes predicted using different threshold cut-off values using IgG1 Fc region (AAD38158) as model.

In terms of prediction success, i.e. the correlation of predictions and experimental data generated from the Pepscan analysis, the *in silico* analysis showed a high degree of accuracy. The majority of those extra antigenic regions were those in which five properties were in excess of thresholds.

Table 3.9 shows the breakdown of the accuracy of the prediction method in comparison to each Monoclonal Rheumatoid factor antibody (mRFAB) in the Pepscan analysis. Using 25% threshold, on average 70% of the predictions agreed with those identified in vitro. These ranged from 80% observed in polyclonal RF and mRFAB, to 60% of those epitopes targeted by monoclonal antibody and anti-Fc mAb. Using thresholds of the top 30% of values, in all but one case the accuracy decreased. Monoclonal RF was the exception to this, showing a 2.5% increase in correct predictions correlating with antibody binding to particular peptides. The final threshold of 40% (Figure 3.6D) showed a further reduction in prediction accuracy. This showed the largest decrease in accuracy to an average 43.8% (a drop of 15.6%). This is unsurprising as there was a larger gap between % thresholds than the previous two. The antibody that showed the least level of correlation was the anti-Fc mAb, which saw the percentage prediction accuracy almost halve above 25%. The best correlations were observed with the polyclonal rheumatoid factor. There was a clear peak in prediction accuracy when using just the top 25% of values (attaining 70% accuracy). When using the highest 40% of values to predict which regions were epitopes, overall accuracy (number agreeing with experimental results/total number predicted) was reduced to 43.8%. Following this analysis it was decided to incorporate the 25% thresholds into the technique/analysis, in order to efficiently identify regions of antigenicity, in addition to limiting the number of false-positives generated.

	Percentage Accuracy (total correctly predicted/total number predicted)		
Type of Rheumatoid factor	**25**	**30**	**40**
Polyclonal RF	80%	75%	58.5%
Monoclonal RF	60%	62.5%	50%
Anti-Fc mAb	60%	37.5%	25%
mRFAB	80%	62.5%	42%
Overall accuracy	**70%**	**59.4%**	**43.8%**

Table 3.9 Percentage accuracy of epitope mapping predictions made using different threshold levels. Peptide functionality was confirmed using polyclonal and monoclonal rheumatoid factor, in addition to monoclonal antibodies and antibodies targeting the IgG Fc region. (RF: Rheumatoid Factor; mAb: Monoclonal antibody)

3.3.3 The importance of secondary structure within epitope mapping

The impact of incorporating secondary structural data upon prediction results is shown in Table 3.10. Overall the inclusion of beta-turns resulted in a 28% increase in prediction accuracy when applied to the identification of the 21 GS epitopes in conjunction with dynamic thresholds using ExPASy. Secondary structure resulted in six additional epitopes being identified (on six different autoantigens).

Epitope	Published epitope	ExPASy	ExPASy + BetaPred2
1	^{34}LSSGKFYGDEEKDKG^{48}	×	×
2	^{142}KDIRCKDDEFTHLYT^{156}	×	×
3	^{193}DPDASKPEDWDERAK^{207}	×	×
4	^{253}EYKGEWKPRQIDNPDYKGTWIHPEIDNPEY^{282}		×
5	^{3}PRRRSRKPEAPRRRS^{17}	×	×
6	^{22}PTPGPSRRGPSLGASSH^{38}		×
7	^{169}TKYKQRNGWSHKDLLRSHLKP^{190}		×
8	^{211}ELYKEKALSVETEKLLKYLEAV^{232}	×	×
9	^{485}EYRKKMDI^{492}	×	×
10	^{147}HKAFKGSI^{154}		
11	^{291}NGNLQLRNKEVT^{302}		
12	^{301}VTWEVLEGEVEKEALKKI^{318}		
13	^{349}GSGKGKVQFQGKKTKF^{364}		×
14	^{88}KIRRSPSKPLPEVT^{101}	×	×
15	^{283}LGKDANNGNLQLRNRNKEVTWEV-LEGEVEKEALKKIIEDQQESLNKWKSK-GRRFKG^{338}	×	×
16	^{90} KKEEKKEESEEEDEDMGFGLFD^{111}	×	×
17	^{231}LDVPPALADFIHQQR^{245}		×
18	^{205}WWEEERYPEGIKWKFLEHKG^{224}		×
19	^{349}RIANFKIEPPGLFRGRGNHP^{368}		
20	^{397}PGHKWKEVRHDNKVTWLVSW^{416}		
21	^{517}ELDGQEYVVEFDFLGKDSIR^{536}		
	Total successfully predicted (out of 21)	9	15
	% Prediction success (Sensitivity) (successful predictions / total no. published epitopes)	43% (9/21)	71% (15/21)

Table 3.10 The importance of secondary structure, i.e. beta turns, in epitope prediction is summarised here. This summary shows the increase in success of predicting experimentally determined epitopes when secondary structure is taken into consideration. Data in the first column was predicted using ONLY the 4 physico-chemical properties of primary structure. Predictions within the second column were generated using BOTH primary structure (4 physico-chemical properties) and the locations of secondary structure motifs, i.e. beta-turns specifically. In order for an epitope to be predicted at least 4 properties must be over the threshold value for that protein. X = Match between predicted and published epitopes. Blank boxes indicate no match between the predicted peptides in this investigation and those in the literature.

3.3.4 Comparison of BCEPRED vs. ExPASy

The combined performances of the algorithms used in ExPASy and BetaPred2 were compared to those utilised within a recently developed online prediction database – BCEPRED (Saha, 2003). The gold standard dataset were used to grade each method upon its performance in the prediction of each of the 21 GS epitopes. Recently optimised dynamic thresholds of 25% were used as cut-offs in ExPASy analysis. BCEPRED analysis used a default threshold value of 2.37. Table 3.11 shows a comparison of algorithms employed by each prediction methodology.

When the performance of both prediction programs was compared, BCEPRED exhibited a 14% increase in prediction accuracy over ExPASy, predicting 18/21 GC epitopes (Table 3.12). Despite this, BCEPRED also predicted almost twice the number of epitopes of that of ExPASY. The majority would have been considered weaker/ insignificant regions by ExPASy. As a result, this showed a reduction in the algorithm's sensitivity and specificity compared to ExPASy.

ExPASy	BCEPRED
Primary structure analysis	
Hydrophilicity (Hopp and Woods, 1981)	Hydrophilicity (Parker *et al*., 1986)
Polarity (Grantham, 1974)	Flexibility (Karplus, 1985)
Flexibility (Janin, 1979)	Accessibility (Emini, 1985)
Accessibility (Bhaskaran and Ponnuswamy, 1988)	Antigenic propensity (Kolaskar, 1990)
	Exposed surface (Janin, 1978)
	Polarity (Ponnuswamy, 1980)
Secondary structure analysis	
*B-Turns – using **Betapred2 (ANN)** (Kaur and Raghava, 2004) http://www.imtech.res.in/raghava/betatpred2/	Turns (Pellequer, 1993)

Table 3.11 Approaches taken by each ExPASy and BCEPRED databases, when making predictions of B cell epitopes.

Epitope	Published epitope	ExPASy & BetaPred2	BCEPRED
1	^{34}LSSGKFYGDEEKDKG^{48}	×	×
2	^{142}KDIRCKDDEFTHLYT^{156}	×	×
3	^{193}DPDASKPEDWDERAK^{207}	×	×
4	^{253}EYKGEWKPRQIDNPDYKGTWIHPEIDNPEY^{282}	×	×
5	^{3}PRRRSRKPEAPRRRS^{17}	×	×
6	^{22}PTPGPSRRGPSLGASSH^{38}	×	
7	^{169}TKYKQRNGWSHKDLLRSHLKP^{190}	×	×
8	^{211}ELYKEKALSVETEKLLKYLEAV^{232}	×	×
9	^{485}EYRKKMDI^{492}	×	×
10	^{147}HKAFKGSI^{154}		
11	^{291}NGNLQLRNKEVT^{302}		×
12	^{301}VTWEVLEGEVEKEALKKI^{318}		×
13	^{349}GSGKGKVQFQGKKTKF^{364}	×	×
14	^{88}KIRRSPSKPLPEVT^{101}	×	×
15	^{283}LGKDANNGNLQLRNRNKEVTWEV-LEGEVEKEALKKIIEDQQESLNKWKSK-GRRFKG^{338}	×	×
16	^{90} KKEEKKEESEEEDEDMGFGLFD^{111}	×	×
17	^{231}LDVPPALADFIHQQR^{245}	×	×
18	^{205}WWEEERYPEGIKWKFLEHKG^{224}	×	×
19	^{349}RIANFKIEPPGLFRGRGNHP^{368}		×
20	^{397}PGHKWKEVRHDNKVTWLVSW^{416}		×
21	^{517}ELDGQEYVVEFDFLGKDSIR^{536}		
	Total successfully predicted (out of 21)	**15**	**18**
	% Prediction success (successful predictions / total no. published epitopes)	**71% (15/21)**	**85% (18/21)**
	Method Sensitivity (Best epitopes predicted (5) correlating with total correct predictions)	**47% (7/15)**	**22% (4/18)**
	Overall Specificity (number of correct epitopes/ Total number of epitopes predicted)	**25.4% (15/59)**	**17% (18/105)**

Table 3.12 Summary of the levels of success of two methods of epitope prediction, in predicting 22 pre-published linear epitopes. X = Match between predicted and published epitopes.. Blank boxes indicate no match between the predicted peptides in this investigation and those in the literature.

3.3.5 Identification of candidates for testing the potential role of HERVs in molecular mimicry

When the maximum level of prediction accuracy was achieved, the technique was applied to predict potential epitopes present upon 26 autoantigens, and 13 retroviruses associated with RA. Upon prediction, resulting epitopes were then aligned with one another in order to detect regions of primary structure homology. Table 3.13 shows the best alignments between epitopes of both HERV and autoantigen derived epitopes.

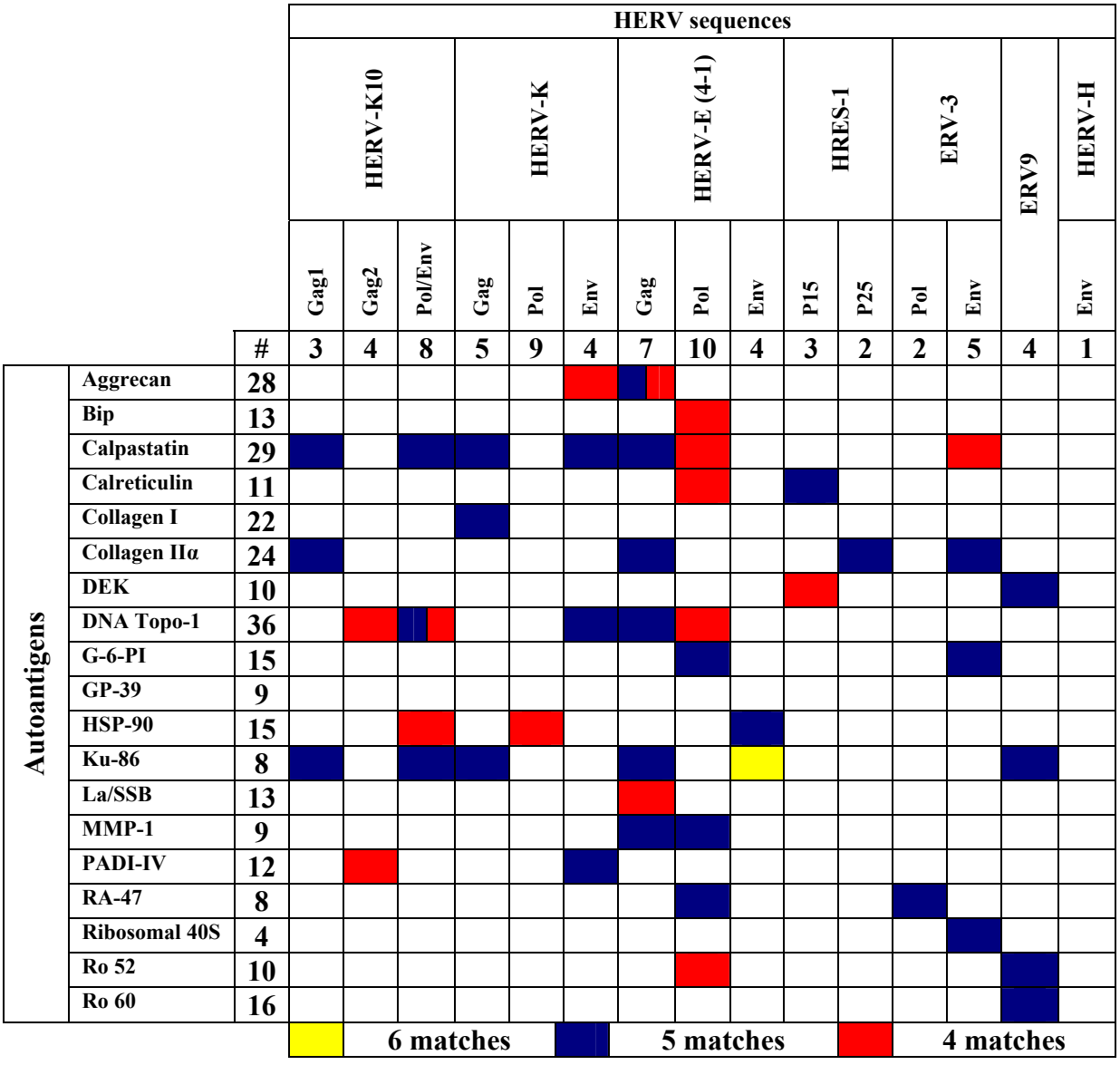


Table 3.13 Level of homologous sequences shared between epitopes of autoantigens and HERVs alike. Yellow shading indicates matches of 6 residues or more. Blue indicates matches of 5 residues and red shading shows homology of 4 amino acids.

Those shown are the best candidates for molecular mimicry as epitopes from both groups of proteins share sequence homology (more than 4 consecutive amino acids). Unexpectedly, more matches of 5 consecutive residues were identified in comparison to 4 consecutive residues. HERV K and HERV-E (4-1) both showed the most homology between their epitopes, whilst HERV-H showed the least.

3.3.6 Incorporation of Negative controls into the investigation

Often small data sets contain a strong bias for both positive and negative data, since data was selected using a stringent set of rules (Donnes and Elofsson, 2002). One way to lower potential bias and its influence over a set of results was to include some random epitopes, to which it was assumed there would be little or no binding. Thus, potential epitopes predicted on autoantigens and HERV proteins alike were also aligned with two presumed negative control peptides selected from epitopes predicted from the two control sequences – retroviral elements identified in plants (Table 3.14). Of those epitopes that matched, the most consecutive homologous residues equalled 3 (Table 3.15).

Organism	Gene	Location	Sequence
Gypsy/Ty3 element (AAG51464)	-	5-20	RSSQEEDVKGKGVEKE
		56-70	TAKRVTEGEKVLDRS
		290-303	YRPVKEVEQKSDHL
		340-353	EVLSDDDHEQKPMP
		369-382	MGVKGTVDKRDLFI
		1031-1045	SWSQDILRRKSKIVV
		199-213	MAELKKLQETDGLVE
		278-291	TWSQKGTRSGGSYR
		414-427	AVADGRKLNVDGQI*
		516-531	CVREVVSDEEQEIGSI
		588-602	YVVHQKDEIDKIVQD
		677-691	QVRMDPDDIQKTAFK
		737-751	LIYSSSIEEHKEHLR
		775-789	ISAREIETDPAKIQA
		929-943	FIIKTDQRSLKYLLE
		1062-1076	SGMGGRSGRDASHQR
		1401-1414	YILERKLVKRQGRA
Athila 4 (AAL06422)	**Reverse transcriptase**	33-47	GGMTVVKNEKDELI
		207-222	SEKGIEVDKGKVEVMM
		59-72	DYRKLNAASRKDHF

Table 3.14 The potential epitopes mapped out on the surface of two plant retroviral elements. Epitopes were mapped out using the ExPASy/BetaPred2 methodology with 25% threshold. The two epitopes shaded in yellow indicate those used to align with autoantigen epitopes in order to eliminate bias in the investigation. *: Negcont1 used in serological investigations.

This occurred with epitopes on 5 autoantigens including Collagen I and Collagen II, as well as DEK, DNA Topoisomerase and Ku-80. Only those matches where there were 5 or more consecutive residues were deemed significant, therefore it was assumed there was minimal bias within the investigation structure.

	Autoantigen																		
	Aggrecan	Bip	Calpastatin	Calreticul;in	Collagen I	Collagen II	DEK	DNA -Topol	G-6-PI	GP-39	HSP-90	Ku-86	La/SSB	MMP-1	PADI-IV	RA-A47	Ribosomal	Ro 52	Ro 60
	28	13	29	11	22	24	10	36	15	9	15	8	13	9	12	8	4	10	16
Negcont1																			
Negcont2																			

5 residues matching | 4 residues matching | 3 residues matching

Table 3.15 Sequence homology between epitopes from host autoantigens and negative control epitope sequences. Alignments were carried out using the LALIGN programme. Blue indicates 5 consecutive residues matching, Red shows 4 consecutive residues matching and yellow indicates 3 consecutive residues matching.

3.3.7 Short listing of candidates for synthesis as short peptides

Those epitopes that exhibited the most homology to autoantigen epitopes were then short-listed for progressing to phase two of the investigation, i.e. synthesis of peptide and screening of patient blood. Table 3.16 shows those epitopes most likely to be involved in playing a role in molecular mimicry and thus triggering RA disease pathogenesis. Novel peptides of notable interest include 'GfPN1kpr' which showed homology to epitopes on Collagen II and 'RA-HK-g-1', which was a significantly antigenic epitope on HERV K. Table 3.17 shows a selection of peptides, which showed no particular disease association but were considered to be novel and of interest. 'GfPN2eip', for example, was a predicted epitope which aligned with the HERV-K superantigen region to which antibodies have been reported in patients with rheumatic diseases (Herve *et al*. 2002).

Disease	Peptide name	Comments	Peptide sequence
RA	**HERV K10 Gag1 (GfPN1kpr)**	▪ 4th most antigenic peptide from K10 Gag1. ▪ Flexible region - beta turn structure. ▪ Best alignment = Collagen II 1076-1096.	**PSESKPRGTSPLPAG**
RA	**HERV K10 Pol/Env (RA-HK10PE-1)**	▪ 4th most antigenic peptide for Pol/Env. ▪ High Polar and accessible region. ▪ Best alignment = DNA Topoisomerase I 18-36 (~6).	**DLTESLDKHKHKKLQ**
RA	**HERV K10 Pol/env**	▪ Epitope on Calpastatin binds twice to HERV K10 Pol/env.	**GSPTAAGKKTEKEESTEV LKA**
RA	**ERV3 Env (RA-ERV3E-1)**	▪ 2nd most antigenic region of ERV3. ▪ Located at N-terminal of Beta-turn region – increased flexibility. ▪ High levels of Polarity and accessibility.	**LWRGKSNNSESPHPSP**
RA	**HERV K Gag (RA-HK-g-1)**	• 2nd most antigenic peptide on K Gag. • Located on Beta turn region.	**VGPLESKPRGPSPLS**

Table 3.16 Peptides selected as best candidates for testing molecular mimicry between HERVs and autoantigens associated with RA.

Disease	Peptide name	Comments	Peptide sequence
MISC	**HERV K10 Pol/Env (GFPN2eip)**	▪ Located in super antigen region of Env. ▪ 2nd best epitope on K Env. ▪ Good overall hydrophilicity, accessibility and polarity.	**PCPKEIPKESKNTEV**
MISC	**Collagen II (CII-1)**	▪ 2nd most antigenic epitope in Collagen II. ▪ High hydrophilicity and Polarity values. ▪ Highly flexible - long Beta-turn region. ▪ Similarity to two regions of HRES-1 p25 gene.	**GSPGPAGPTGKQGDR GEAGA**
MISC	**HERV K10 Pol/Env (HK10PE-2)**	▪ 2nd most antigenic peptide for K10 Pol/Env. ▪ Increased Polarity and flexibility. ▪ Best Alignment - AHNAK.	**QIENRKIKPQKIEIR**
MISC	**Ku 86/90 (165-187) (Ku-1)**	▪ Best alignment - HERV protein and Autoantigen epitopes. ▪ Significantly high hydrophilicity and polarity levels.	**LPF SLGKEDGSGDR**
MISC	**Heat Shock protein 90 (693-708) (HSP90-1)**	▪ High hydrophlicity, accessibility and polarity values. ▪ Located on beta-turn region. ▪ Best alignment: E (4-1) Gag.	**KLGLGIDEDDPTADD**

Table 3.17 Peptides selected as best candidates for implicating HERVs in triggering Rheumatic disease by molecular mimicry. These peptides were selected for their homology to epitope and antigenic regions on autoantigens involved in rheumatic disease.

3.4. Discussion

3.4.1 Prediction of epitopes

During the course of this investigation a method for the prediction of potential B cell epitopes was optimised. This method was used to epitope map those autoantigens and HERVs previously associated with triggering rheumatoid arthritis. For the purpose of this analysis, we compiled a panel of 21 gold standard (GS) epitopes, from 7 autoantigens implicated in systemic rheumatic disease pathogenesis. All GS epitopes were continuous in nature and had been published in the literature previously after being determined by experimental methodologies. These 21 epitopes formed a dataset of positive controls which could be used in to determine prediction accuracy reliably. The existing methodology of B-cell epitope prediction used in-house incorporated the 4 physico-chemical properties of hydrophilicity, polarity, flexibility and Accessibility and based analysis upon 4 experimentally derived scales (Hopp and Woods, 1981, Grantham, 1974, Janin, 1979, Bhaskaran & Ponnuswamy, 1988). These were selected as they had been reported previously to be successful (Pellequer, 1991). Previous studies had also shown that overall predictive accuracy was increased when predictions from more than one property were combined (Alix, 1999) although a number of previous studies have ignored this, making predictions based upon single algorithms (Galbraith *et al*., 2000).

Using the predictive power of combined properties, initial studies were carried out using default cut-off values i.e. values generated by individual algorithms which suggest residues is not significantly important. These original values were as dictated by authors of original experimental data scales (Table 3.6). It became apparent that predictions made using these cut-off values, showed limited accuracy (Table 3.7). As a result, dynamic threshold/cut-off values were trialled which resulted in an accuracy of 43% (an increase of 28% with 9/21 GS epitopes being predicted). These markers of significance assessed proteins upon an individual basis rather than as a marker for all proteins. Additionally, 25% was shown as the most optimal cut-off value when applied to the IgG1 Fc region. This was backed up by experimental data from four separate monoclonal antibodies (Table 3.8 – 9). This data showed that although more epitopes were detected using 40% threshold values (Filaggrin alone generated in excess of 80 individual epitopes), overall accuracy (total positive epitopes/total number of

predictions) was increased when using 25% cut-off values (43.8% and 70% respectively).

Additionally, incorporation of protein secondary structure also led to 28% increase in predictive accuracy (Table 3.10). Locations of Beta-turns were made using the BetaPRED2 forward feed neural network, with PSI-BLAST matrix results, combined with secondary structure data as input. Results were further refined in the second layer in order to determine location of secondary structure motifs (Kaur and Raghava, 2003). Such methodologies had been reported previously to attain accuracy levels of 80%, compared to the 60 – 65% achieved via sliding windows algorithm based methodologies (CASP4, 2000). Overall combining primary and secondary factors in the prediction of epitopes in the GS dataset resulted in the correct prediction of 15/21 (71%) compared to 9/21 (43%) based upon just primary.

Once fully optimised the performance of the novel prediction methodology was compared with that of a recently developed and publicly available online server – BCEPRED (Saha, 2003). BCEPRED differed from the ExPASy methodology in that it had 6 criteria (Table 3.11), all of which were algorithms based upon the sliding window principle. Overall when assessed using the GS dataset, BCEPRED predicted 85% (18/21) of epitopes compared to the 71% (15/21) managed by ExPASy. Despite this, when assessed in terms of sensitivity and specificity ExPASy fared better than BCEPRED, overall predicting fewer, and more antigenic epitopes (Table 3.12). These would be less likely to be eliminated from early stages of prediction due to their increased antigenicity. Although results initially suggest the increased potential for the ExPASy method in contrast to that of BCEPRED, it should be taken into consideration that this analysis was based upon a small dataset of 21 epitopes. Often small data sets encourage a strong bias for the results (Donnes and Elofsson, 2002) Any future work should take this into consideration, increasing the size of this dataset before further analysis.

One of the main problems with such an approach is that so far, the measure of error in predicting B cell epitopes is unknown, with very few studies addressing this problem (Blythe and Flower, 2005, Greenbaum *et al*., 2007). Those investigations which have tackled this issue have had to overcome difficulties such as how to differentiate between true positive, true negatives, false positives and false negatives when much of the field is theoretical (Baldi *et al*., 2000). This is further complicated by mechanisms such as epitope spreading (Section 3.1.6) that are characterised by

antibodies to secondary epitopes which appear at different stages during a response (James and Harley, 1998). Despite many of the associations linking ES to disease, it is likely that it operates as a vital component of the protective immune response and it has been well recognised that over time there is a progressive and ordered accumulation of antibodies generated to infectious agents. Many of these serological profiles such as that in HIV, form crucial parts of patient management. In Hepatitis B virus (HBV) infection, T-cells primed with the nucleocapsid (core) protein of HBV (HbcAg) are able to drive antibody production against the HBV surface protein (Milich *et al*., 1987). ES may also play a role in the resolution of/protection against autoimmunity. Evidence of such a role has been observed in animal models including that of adjuvant arthritis in Lewis rats (Moudgil *et al*., 1997) and the natural recovery of patients suffering from acute autoimmune phenomenon such as Guillain-Barre syndrome (Powell and Black 2001). Despite this evidence however, there are cases where successful immune responses are mounted against infectious agents, in the absence of ES (Whitton *et al*., 1989, Klavinskis *et al*., 1989, Levely *et al*., 1991, Cole *et al*., 1995, Mikszta *et al*., 1997).

3.4.2 Selection of epitopes for peptide synthesis

Once optimised, the technique for predicting antigenic epitopes was applied to the identification of epitopes on 26 autoantigens and 13 endogenous retroviruses. After identification, epitopes originating from autoantigens were aligned with those from retroviral proteins, using a local alignment program to identify regions of shared homology (Table 3.13 – 3.14). Those regions of shared sequence were then highlighted and short listed as candidates for synthesis into short polypeptides and for subsequent incorporation into ELISA for screening of patient blood for antibodies to their relevant epitope. Of those peptides short-listed (Section 3.3.7), two were selected for synthesis - GfPN1kpr (Figure 3.6A/B) and GfPN2eip. GfPN1kpr showed the best alignment with Collagen II. Collagen II is a major autoantigen in RA (Wang *et al*, 2002), with strong evidence observed in animal models in mice (Brand *et al*., 2004). This epitope was in the top four most antigenic epitopes on HERV-K10 Gag, in addition to being located next to a Beta-turn, a potential marker of antigenic regions. GfPN2eip was selected due to its location within the superantigen region of HERV-K. Super antigens are molecular protein's capable of activating large numbers of immune cells, leading to

clonal exhaustion, senescence and autoreactivity (Ivars, 2007). Such a region had been previously implicated in development of rheumatic disease pathogenesis (Herve *et al,* 2004). Additionally, Eip was located within an antigenic region making it the second most antigenic region identified on HERV-K.

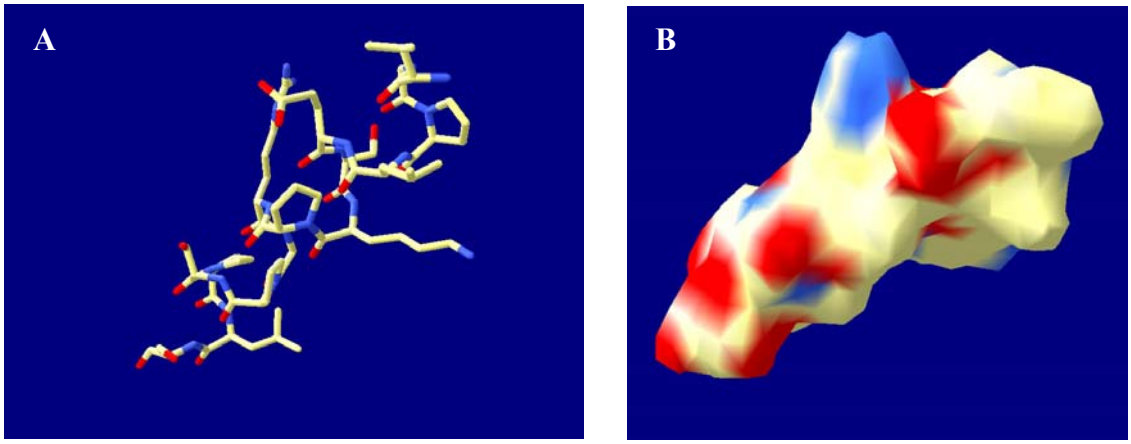


Figure 3.5 (A) A graphical representation of the peptide – GfPN1kpr. The molecular backbone and side chains of the peptide: GfPN1kpr (VGPLESKPRGPSPLS) which was synthesised and used in an ELISA style assay for evaluation of HERV-K10 reactivity in RA patient sera in chapter 4. (B) A diagram showing the electron and charge distribution across the surface of the peptide.
The above images were produced using a publicly available software package available from the ExPASy server (http://us.expasy.org/spdbv/). This application allows the visualisation of proteins in terms of structure and any interactions they may have wilth other proteins. Deepview application has been developed with the Swiss-MODEL PDB database and modelling server.

A third peptide was also selected for manual synthesis using in-house facilities. PLSK (for sequence see Table 4.2) was derived from the same region as GfPN1kpr, however was derived from the HERV-K provirus, rather than HERV-K10. This was reflected in the 9 amino acids difference between the two and was included in order to identify differences between members of the same sub-family – HERV-K (HML-2).

At present, there appears to be a poor understanding of the rules governing what exactly constitutes a functional epitope (Greenbaum *et al*., 2007) with many studies appearing to have been content with using areas of high hydrophilicity alone as markers for epitopes (Galbraith *et al*., 2001). Furthermore validations of published experimental derived epitopes have shown this approach to be inadequate (*personal observation*). In particular this highlights a more general problem within the field, i.e. that of a lack of standardisation. This is especially evident when confronted with the myriad of different techniques available, both commercially and for public use, with most based around identical principals. Further validation and analysis of larger

datasets of proteins are required, as well as extensive comparison studies between the different approaches.

3.4.3 Conclusions

In this chapter, an *in silico* based methodology for the prediction of B cell epitopes was optimised and applied for the determination of antigenic regions likely to bind antibodies in RA patients. Epitopes in HERVs and RA autoantigens were cross referenced between each other to detect the presence of any sequence homology, and such regions were then short-listed for synthesis as short peptides for incorporation into an ELISA (Chapter 4). During this analysis, the methodology was validated using a small data set of experimentally derived linear epitopes, identified upon 7 randomly selected autoantigens involved in systemic rheumatic diseases.

Currently, one of the largest obstacles for progress with the field of B cell epitope mapping is the lack of standardisation between individual investigations. The introduction of performance scales such as A_{ROC} (Berman *et al*., 2002), may help to solve this problem and act as a bench mark for progress, however techniques such as this must be utilised by all groups researching this area, rather than many groups developing and using their own individual methodologies with no real means for comparison.

From this work the following conclusions can be made:

- The field of B cell epitope mapping requires standardisation, in terms of both tools used as well as controls.
- The level of sequence homology between HERVs and human autoantigens in RA was less than expected.
- Three peptides were identified for synthesis as short peptides and incorporation into a novel ELISA for surveying of patient samples. These epitopes have been identified as epitopes on RA autoantigen proteins that share sequence homology with endogenous retroviral sequences.

Chapter 4

Abbreviations

Optical density (OD)
American Rheumatism Association (ARA)
Bovine serum albumin (BSA)
Chronic fatigue syndrome (CFS)
Enzyme-linked immunosorbent assay (ELISA)
Epstein-Barr virus (EBV)
Human endogenous retrovirus (HERV)
Hydrochloric acid (HCl)
Immunofluoresence (IF)
Inhibition ELISA (iELISA)
Inflammatory bowel disease (IBD)
Matrix assisted laser desorption ionisation - time of flight (MALDI-TOF)
Mixed cryoglobulinaemia (MC)
Non-specific binding (NSB)
Normal healthy donors (NHD)
Osteoarthritis (OA)
Phosphate buffered saline (PBS)
Purified protein derivative (PPD)
Rheumatoid arthritis (RA)
Rheumatoid factors (RF)
Sjogrens syndrome (SS)
Standard error of mean (SEM)
Synovial fluid (SF)
Systemic lupus erythematosus (SLE)
Trifluoroacetic acid (TFA)
Trimethyl-benzene (TMB)
Tween 20 (TW20)

Chapter 4 – Optimisation & development of ELISA for detection of anti-HERV-K antibodies within patient samples

4.1 Introduction

4.1.1 History of the ELISA

Enzyme assays have been in use for several decades, with applications ranging from the detection of antigen in tissues (Nakane and Pierce, 1966) to antibody quantification, using polystyrene centrifuge tubes as the solid phase (Van Weemen and Schuurs, 1971, Engvall and Perlmann, 1972). The 96-well microtitre plate format was first utilised by a group testing for antibodies to the malaria parasite *Plasmodium falciparum* in the mid-1970's (Voller *et al*., 1974). Today, the enzyme immunoassay is a fundamental technique, commonplace in many immunological laboratories worldwide. One of the most common forms of immunological enzymatic tests used is the Enzyme-linked immunosorbent assay (ELISA) (Schmitt and Papisch, 2002), which has widespread use as an assay for specific antibodies and protein antigens in body fluids and tissue extracts (Cantarero *et al*., 1980). As a detection method, ELISA has a number of advantages over the indirect Immunofluoresence (IF) methodologies, including increased specificity and sensitivity (Borg, 1997). A number of previously published studies have already used ELISA in the detection and determination of anti-retroviral antibodies, exploring their potential roles in rheumatic disease (Deas *et al*., 1998, Bengtsson *et al*., 1996, Herve *et al*., 2002) whilst other studies have identified increased levels of HERV transcriptional activity in Rheumatoid Arthritis (RA) patients, compared to disease controls (Ejtehadi *et al*., 2006).

Additionally the HERV-K (HML-2) ELISA was used to correlate levels of antibodies specific for HERV-K (HML-2) Gag and Env proteins, with the presence of Rheumatoid factors (RF). RF were first discovered almost 75 years ago by Waaler (Natvig and Tonder, 1998) and Rose (Christian, 1998) and target the constant region of IgG (Fc region, typically in the γ2-γ3 cleft) (DeLano *et al*., 2000). Initially their activity was thought to be specific for RA. This is now known not to be so (Westwood *et al*., 2006) and despite their inclusion as one of the seven criteria set out by the American rheumatism association (ARA) for the diagnosis of RA (Arnett *et al*., 1988), RF have been identified in other autoimmune [e.g. Sjogrens syndrome (SS) (Shlomchik *et al*., 1987a, Shlomchik *et al*., 1987b), Systemic lupus erythematosus (SLE) (Tan,

1989), Mixed cryoglobulinaemia (MC) (Martin and Pasquali, 1992, Procaccia *et al*., 1987)] and infectious diseases [Chlamydia (Peeling *et al*., 2000), Hepatitis B (Steigmann *et al*., 1974)] and Syphilis (Cerny *et al*., 1985)]. Additionally, their presence has also been detected to a lesser extent in both healthy and elderly individuals and may form part of the normal immune response. Some pathogens have also been reported to be influenced by RF levels, such as Epstein Barr virus (EBV) (Yang *et al*., 2004) thus levels of RF were determined and correlated with levels of responses to HERV-K proteins.

4.1.2 Aims and Objectives

This chapter outlines the serological phase of this investigation addressing whether HERV-K (HML-2) plays a role in the pathogenesis of RA through protein production. Herein, 4 main themes are addressed:

1. To develop and optimize a specific and sensitive ELISA, incorporating epitopes predicted in chapter 3 using bioinformatic analysis.
2. To use the novel ELISA to determine levels of antibodies to predicted epitopes in RA patients and disease controls. This may give an indication as to whether these epitopes play a role in disease pathogenesis.
3. To compare the levels of anti-HERV antibodies in both the synovial fluid and blood, correlating both with levels of RF. This may give an indication as to whether RF can modulate HERV-K activity and also whether the non-specific binding of RF was playing a role in influencing ELISA OD values.
4. To assess the performance of and to validate the output of the bioinformatics methodology, as described in chapter 3, through a number of means including biotinylation of peptides and the use of a negative control peptide.

A novel ELISA was thus developed using the peptide and corresponding polyclonal rabbit bleed in order to optimise assay conditions before using it to test RA patient and control antisera.

4.2 Materials & Methods

All reagents were derived from (Sigma-Aldrich, Gillingham, UK) unless specified otherwise

4.2.1. Patient samples

Blood samples were taken from 48 RA patients who attended the Rheumatology Clinics at New Cross Hospital, Wolverhampton. All RA patients satisfied the criteria set out by the ARA (Arnett *et al*, 1988). Serum samples were also taken from control groups including 27 Osteoarthritis (OA - non-inflammatory joint disease) and 17 inflammatory bowel disease patients (IBD) (7 diagnosed with active Crohn's disease, 7 with inactive Crohn's disease and 3 with Ulcerative colitis). IBD patients acted as non-joint inflammatory disease controls. Furthermore, 23 rheumatic disease controls with Systemic lupus erythematosus (SLE) and 37 Normal healthy donors (NHD) were selected. Specific inclusion criteria stipulated that all patients with RA must be in attendance at the Rheumatology clinic, New Cross Hospital, Wolverhampton, and fulfilled diagnostic criteria set by the American College of Rheumatology (ACR). No exclusion criteria were set for this investigation. All samples were taken by staff working at the clinic and constituted 10ml of venous blood were collected in sterile preservative-free heparin from each patient after informed consent had been given. Blood samples were also obtained from healthy volunteers of known age and sex. In order to maximise accuracy of the investigation all blood samples were processed within 3 hours. All steps were performed under aseptic conditions.

4.2.2 Peptide synthesis

A single peptide 'PLSK' was synthesised manually using F_{moc} solid phase synthesis as described previously (Fields and Noble, 1990). Synthesis was carried out on a 0.2 mM scale. Amino acid side chains were protected as t-butyl esters (Aspartic acid and Glutamine), esters (Serine, Threonine and Tyrosine) and urethanes (Lysine and Histidine). Guanidino and thiol groups of arginine and cysteine were protected by 4-methoxy-2, 3, 6-trimethylbenzenesulphonyl and trityl groups respectively (Hancock and Battersby, 1976). Kaiser tests were employed to verify coupling and deprotection

(Chan and White, 2000). Cleavage and side-chain deprotection was accomplished for appropriate lengths of time with Trifluoroacetic acid (TFA). Biotin was attached to nascent peptides prior to the cleavage step, via the addition of a cysteine residue at the N-terminal. This addition also facilitated the binding of the PPD immunogen via the cysteine's side chain thiol group. Peptide identity was verified by molecular weight using 'Matrix Assisted Laser Desorption Ionisation - Time of Flight' mass spectrometry (MALDI-TOF) (Shimadzu UK Ltd, Milton Keynes, UK). Peptides were lyophilised and resuspended in sterile phosphate buffer to a concentration of 10mM and stored at -20°C. Peptides 'GfPN1kpr' and 'GfPN2eip' were synthesised commercially (Severn Biotech, Kidderminster, UK). 'Negcont1' peptide was synthesised commercially (Severn Biotech, Kidderminster, UK), however like PLSK it lacked specific antiserum.

4.2.3 ELISA micro-titre plate optimisation

Four 96-well microtitre plates [Nunclon Δ surface (Sigma, UK), Falcon Flexiplate (Falcon, CA), HB2 (DYNEX technologies, INC, USA) and HB4 (Immunolon, UK)] were trialed using the GfPN1kpr peptide as described previously (Section 2.6.1). Wells were coated with 400ng/50 µl of peptide and positive serum was diluted in a seven-step two fold dilution series starting at 1/50, and blocked with 2% Bovine serum albumin (BSA)/1 x Phosphate buffered solution (PBS) (0.05% TW20). Absorbance readings were taken at 450 nm after an incubation of 5 mins at room temperature, in the absence of light, using a Trimethyl-benzene (TMB) (Sigma, UK) and a Multiscan platereader (Labsystems, Thermo, CA). All experiments were completed in triplicate and repeated independently three times in order to compensate for intra and inter assay variation.

4.2.4 Determination of optimal antigen concentration

The checkerboard titration technique was used to test both the peptides and their respective terminal bleeds against one another (Crowther, 2000). A second plate was also run with titrated peptide against a second unrelated polyclonal antibody in order to confirm there was no cross-reactivity between the two peptides. Concentrations of peptide started at 80µgml^{-1} and were further diluted two-fold across the plate (column 1-11). Serum was also diluted two-fold over 7 dilutions, ranging from 1/50 to 1/3200.

Diluted antibody only and peptide/ no antibody controls respectively were also included upon the plate. This was repeated using GfPN2eip to coat the wells, in order to assess the potential for cross-reactivity between the polyclonal Kpr-specific bleed and other short peptides incorporating HERV epitopes.

4.2.5 Blocking agent optimisation

Several candidates for blocking agents were chosen from previously published studies and investigations were conducted to determine their blocking efficiency (Table 4.1). Blocking agents were assessed using the 'final rabbit bleeds' for specific peptides - Kpr and Eip (at 1/200 dilution) and an antigen coating of 10μgml^{-1}. The nature of these non-specific interactions was investigated by omitting both the antigen and primary antibody in two separate investigations. The elimination of antigen from the assay allowed the determination of the Non-specific binding (NSB) capacity of the primary antibody within the assay (Figure 4.6). Through omission of the primary antibody, this allowed the determination of the level of non-specific binding occurring as a result of the conjugate (Figure 4.7). All blocking agents were made up in diluent (1 x PBS) to a concentration of 2%, unless otherwise stated. The same diluent was also used as a negative control.

Blocking agent	Reference
Bovine Serum Albumin (BSA)	**(Herrmann *et al.*, 1979)**
Casein	**(Herve *et al.*, 2002)**
Gelatin (fish skin)	**(Gary *et al.*, 1985)**
Milk powder (Marvel)	**(Vogt *et al.*, 1987)**
Ovalbumin	**(Crowther *et al.*, 2001)**

Table 4.1 shows the different blocking agents tested for suitability within the ELISA assay. All blocking buffers were used at 2% concentrations unless otherwise stated.

4.2.6 Incubation of substrate and reaction stopping time

ELISA was performed using HB2 plates and 2% BSA blocking buffer, both of which had been optimised in previous steps. A 1/200 dilution of primary antibody

(GfPN1kpr terminal bleed) was added at 50 µl per well. Absorbance values were read at 450nm using the MS Multiscan plate reader at different time intervals. Readings were taken for reactions immediately after adding TMB substrate, and for periods between 1-10 minutes, at 1 min intervals. All incubations were carried out in the dark and at room temperature. A stopping solution of 50µl of 2N Hydrochloric acid (HCl) was used to stop each enzymatic reaction with each reaction time tested three times in triplicate over three separate plates.

4.2.7 Bleed evaluation

Commercially synthesised peptides and their respective bleeds were tested in an ELISA in order to measure their reactivity at different concentrations and to confirm bleed functionality in addition to determining optimal dilutions for use within the assay. Bleeds for each peptide were taken pre-immunisation, and at 7 day intervals including 7, 14, 21 and 28 days.

4.2.8 Inhibitory studies using GfPN1kpr and GfPN2eip

Inhibition ELISA (iELISA) was used as a method for validating peptide specificity by showing that the bleeds used to develop the assay were specific for the peptide and not due to reactions of a non-specific nature. The iELISA protocol was separated into two phases. An initial phase involved the incubation of the primary antibody with log dilutions of the peptide it had been raised against. Incubated serum was then tested in the assay using the same peptide as antigen. A correlation between an increase in Optical Density (OD) and increasing log dilutions of peptide incubated with the antibody suggested OD were indicative of a specific interaction between epitope and paratope. The inhibition assay was also carried out using an irrelevant peptide (Negcont1) and irrelevant antiserum (GfPN2eip) further emphasising the specificity and lack of cross-reactivity between peptides and bleeds alike. The Negcont1 peptide was included as a negative control in order to eliminate bias.

4.2.9 Investigation of anti-HERV-K Gag & Env antibodies within the synovial fluid

Levels of antibodies to epitopes Kpr and Eip were also measured in the Synovial Fluid (SF). This was investigated by substituting synovial fluid, diluted to 1/200 in 2% blocking buffer (0.05% Tween 20 (TW20), into the ELISA in place of primary serum. This gave an indication of the antibody responses to HERVs at the site of disease, within the joint.

4.2.10 Quantification of rheumatoid factor levels within the synovial fluid & blood

Levels of RF in patient samples (both serum and synovial fluid) were measured using a commercially available quantitative agglutination assay (Plasmatec, Bridport, UK). The assay was carried out according to manufacturers' instructions, with positive and negative controls included on each run.

4.2.11 Biotinylated ELISA

Biotinylated ELISA was carried out using React-bind™ NeutrAvidin™ coated 96-well plates (Pierce, Rockford, IL). 400 ng/50 µl of biotinylated peptides were added to each well and washed three times in 1 x PBS/ 0.1% TW20 before ELISA was carried out as described previously (Section 2.6.1).

4.2.12 Immunofluoresence

2 x 10^3 cells (T47D, Tera1 and human fibroblast-like synoviocytes) were used to seed each well on a BD Falcon culture slide (BD Biosciences, Oxford, UK). Each seeding volume was diluted with 300µl of supplemented medium (for details see Appendix III) and incubated for 12-16 hours at 37°C at 5% CO_2. After incubation, spent media was discarded and cells were washed 3 times for 10 min with 1 X PBS. Cells were then fixed with 2% formaldehyde in PBS/pH 7.4 for 15 min at 20°C, before washing 3 times with 1 x PBS for 10 min. 0.2% Triton X-100 / PBS/ 1% BSA (pH 7.3) was added to permeablise cells, for 5 min on ice. Fixed cells were washed 3 times for 10 min with PBS/ 1% BSA before the slide was incubated in a humidity chamber with

50µl of antibody at 1/200 dilutions in PBS/ 1% BSA/ 0.1% TW20 for 1 hour at room temperature, or overnight at 4°C. After 3 x 10 min washes with PBS/ 1% BSA, slides were incubated for a further hour with the secondary antibody, conjugated to FITC (Dako, Glostrup, Denmark) at appropriate dilutions (manufacturers' recommendations). After a further 3 x 10 min washes with PBS/ 1% BSA, slides were allowed to drain before a single drop of Vectorshield™ mountant (with DAPI) (Vector labs, Burlingame, CA) was added, and sandwiched using a sterile coverslip. Slides were sealed using clear nail varnish before setting and viewing with DAPI fluorescing at 461 nm as blue and FITC fluorescing at 521 nm as green, using the LSM 510 Meta confocal microscope and analysed with LSM 5 Meta software (Carl Zeiss IMT Corp, Thornwood, NY).

4.2.13 Statistical analysis

Statistical analysis of differences between > 2 groups was carried out using a one-way ANOVA test and Tukey-Kramer test in order to compare all groups and individual pairs of groups, not assuming Gaussian distributions amongst samples (Graphpad version 5.1, San Diego, CA). SPSS (SPSS, Chicago, IL) was also used to assess variance between sample results. Analysis of two paired samples was carried out using 2-way paired t-tests, and Mann Whitney tests. P values of <0.05 were taken as significant.

4.3 Results

4.3.1 Peptide/Antigen selection and development

Three synthetic peptides of 15-17aa were designed incorporating epitopes using *in silico* analysis of putative HERV protein sequences involved in potential molecular mimicry with disease autoantigens. One peptide incorporating an epitope derived from a Gypsy/ Ty3 element found in *Arabidopsis thaliana* was also included as a negative control to prevent bias within the analysis (Table 4.2). Peptides 'GfPN1kpr', 'GfPN2eip' and 'Negcont1' were commercially synthesised. Specific antiserum was produced for both Kpr and Eip, but not Negcont1. These peptides contained charged amino acids predicted to be of a hydrophilic nature and thus exposed at the surface of the molecule. High titres of polyclonal antiserum were then harvested after immunisation of rabbits with PPD-coupled Kpr and Eip peptides. A fourth peptide 'PLSK' was synthesised using in-house facilities at the University of Wolverhampton (Section 4.2.2) and its identity confirmed using MALDI-TOF analysis (Figure 4.1). No accompanying bleed was raised to PLSK and its purity was not ascertained by HPLC.

Peptide name	Sequence	Organism derived	Accession no.	Justification
GfPN1kpr (Kpr)	GPSESKPRGTSPLPAG	**HERV-K10 Gag1**	**M14123**	**Homology to predicted epitope on Collagen II and also shared homology with HRES-1 p25**
GfPN2eip (Eip)	PCPKEIPKESKNTEVL	**HERV-K10 Pol/Env**	**M14123**	**Highlighted as superantigen region of HERV-K10 (Herve *et al*, 2001)**
PLSK	VGPLESKPRGPSPLSA	**HERV K Gag**	**Y18890**	**3 alignments with two from Collagen I and one on Calpastatin**
Negcont1	VAVADGRKLNVDGQIK	***Arabidopsis thaliana* Gypsy/ Ty3 element**	**AAG51464/ AC069160**	**Sequence of retroviral origin unrelated to HERVs**

Table 4.2 Table showing the peptides selected as candidates for synthesis and testing against patient antiserum.

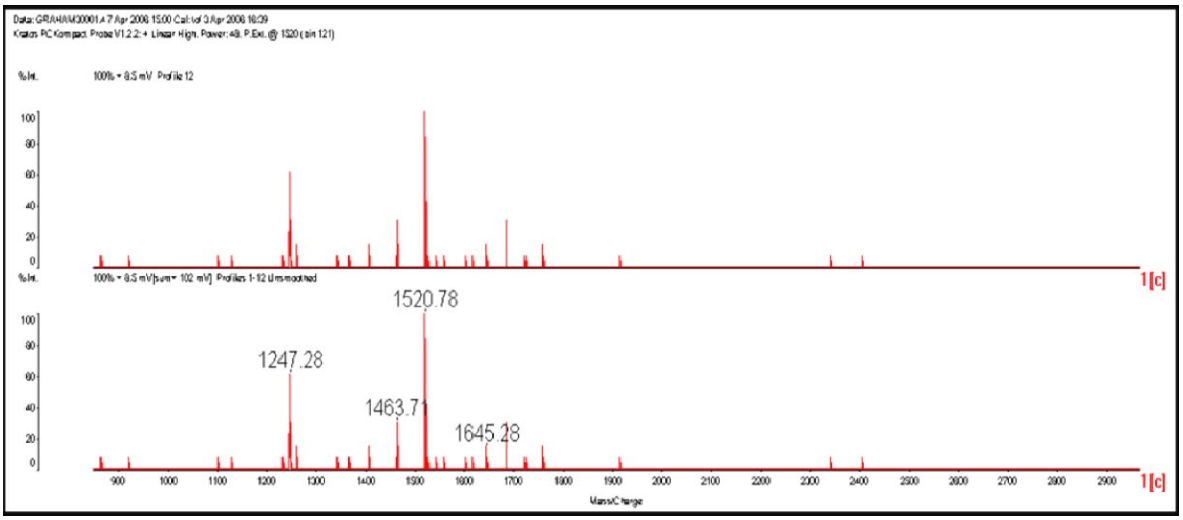


Figure 4.1 MALDI-TOF mass spectrometry analysis of the PLSK peptide, which was manually synthesised using in-house facilities in the Molecular Pharmacology group, University of Wolverhampton. Of particular note is the largest peak at 1520.78 – the expected molecular weight of the synthesised peptide.

During the course of the assay development, the optimisation of a number of variable factors was taken into account. This included optimisation of the type of microtitre plate, the blocking agent, time of reaction between conjugate after addition of substrate before stopping and both the antigen and bleed dilutions to be used.

4.3.2 ELISA plate and Antigen seeding optimisation

The 96-well micro-titre plate is one of the main components of an ELISA assay and one of the more commonly used forms of solid phase (Gonzalez-Buitrago and Gonzalez, 2006). A key feature of the solid-phase ELISA is that antigen can be easily attached to the plate surface by passive adsorption, in a process commonly known as coating. Poor binding of the peptide to the microplate may result in the low performance of the assay and underreporting of actual absorbances, leading to an increase in false-negatives.

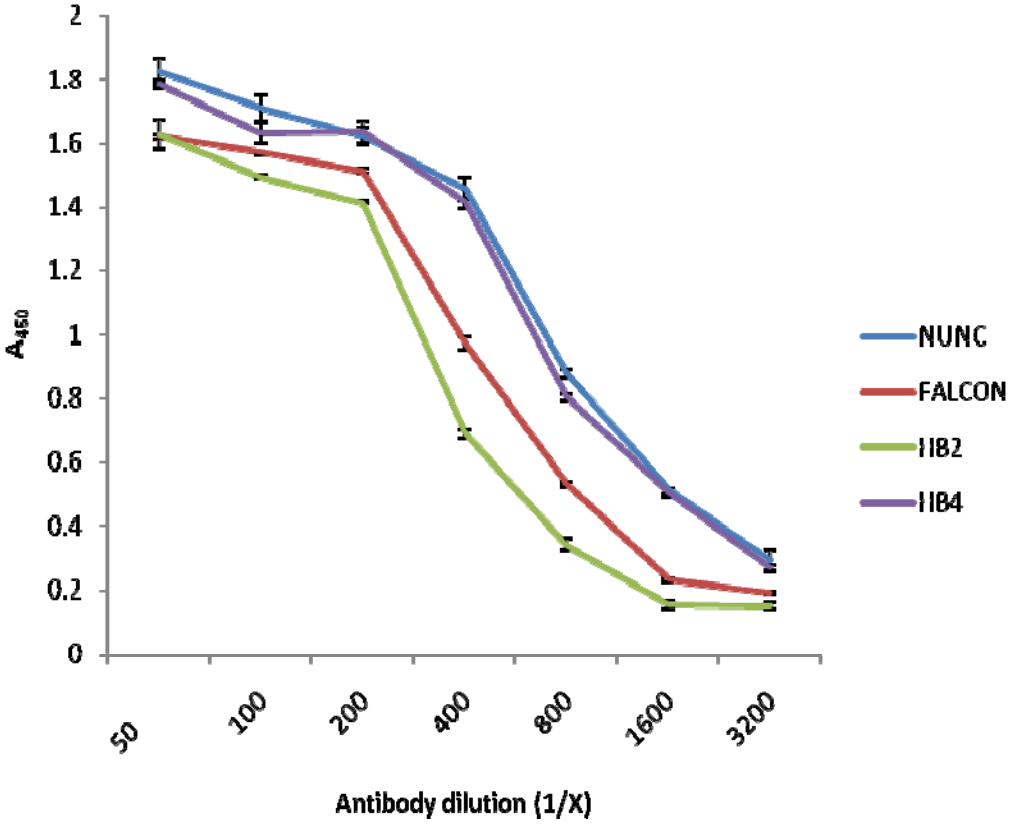


Figure 4.2 Comparison of four separate 96-well microtitre plate types [Nunc (Sigma), Falcon (Falcon), HB2 (Dynex) and HB4 (Immunolon)] using serial dilutions of primary antibodies (terminal bleed to GfPN1kpr). Error bars show Standard error of the mean (SEM).

Serial dilutions of the GfPN1kpr terminal bleed ranging from 1/50, down a two-fold dilution series to 1/3200 were tested against a single concentration of peptide (10 μgml^{-1}) coated on each of the four test plates. HB2 Immunolon plates showed the lowest level of background binding, although binding of all plates was comparable, particularly at lower dilutions (Figure 4.2), i.e. 0.01 - 0.02 at 1/3200. Results, shown in Figure 4.3, indicated a distinct correlation between decreasing antibody dilution and the level of absorbance recorded. This trend was observed at all peptide concentrations. Peaks in OD values were observed at 160 μgml^{-1} of antigen, although levels of standard error were increased compared to other dilutions. For this and economical reasons, the lower coating volume of 10 μgml^{-1} was selected and used for future assays. Additionally, it had been previously reported that adding higher concentrations of peptide to a microtiter well (up to 10 μgml^{-1}), increased the total amount of peptide adsorbed (Butler *et al*., 1987). The reactivity of the Kpr bleed was also tested against the GfPN2eip peptide across the same dilution series (data not shown). This did not

show the same correlation between antibody dilution and OD value, thus validating the specificity of the positive serum, for Kpr and demonstrating its lack of cross-reactivity with other peptides. All coating steps were carried out overnight at 37 ºC. The effect of temperature upon passive adsorption of antigen was also investigated thus confirming 37 ºC as the most optimal temperature for coating. From this investigation it was concluded that HB2 plates presented the lowest background levels. The results outlined above also suggested an acceptable antigen coating concentration was 10 μgml^{-1}.

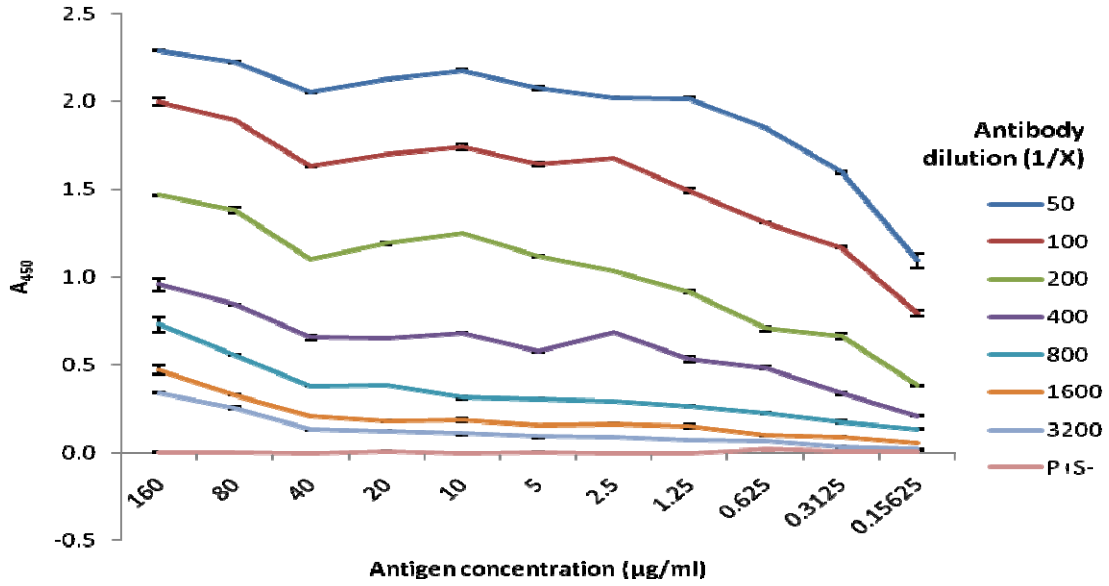


Figure 4.3 Checkerboard titration results in which gradients of antigen were plotted against serum titration. Serum dilutions are labelled in key. Error bars indicate SEM. P+ S- : Peptide, no serum control.

4.3.3 Optimisation of blocking buffers

Measures were also taken to reduce the non-specific adsorption of contaminating proteins present in samples, after coating the bottom of wells with antigen. A literature search was carried out identifying potential candidate blocking agents for use within the assay. Candidate blocking agents assessed (Table 4.1) were tested as described previously (Section 4.2.5). The results for the evaluation of the best blocking agent were shown in Figures 4.4 - 4.5. The results in Figure 4.4 show the influence of blocking buffer upon final OD values when an ELISA was run in the absence of antigen. This gave an overall indication of the level of non-specific binding

of the blocking agent itself to the microtitre plate and its non-specific reactivity concerning the primary antibody. After subtraction of background values 2% BSA showed the lowest level of non-specific binding. Casein and Gelatin both showed high OD values indicating a large degree of non-specific binding was occurring.

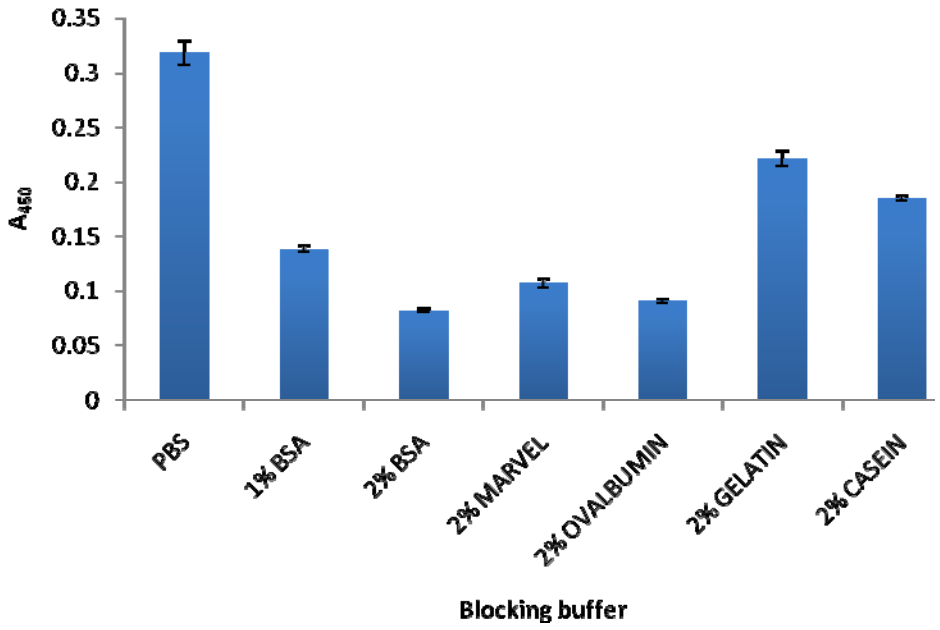


Figure 4.4 Comparison of previously published blocking buffers omitting antigen indicating the effect of non-specific binding due to the primary antibody. The higher the OD value, the less blocking effect and higher non-specific binding is achieved. As shown above, Gelatin and casein showed the highest OD values of those blocking agents tested suggesting a degree of non-specific binding. The blocking agent exhibiting the lowest absorbance was 2% BSA. As expected the highest values were observed using PBS (negative control). These showed a high degree of non-specific binding occurring between the antiserum and the plate. Error bars show SEM.

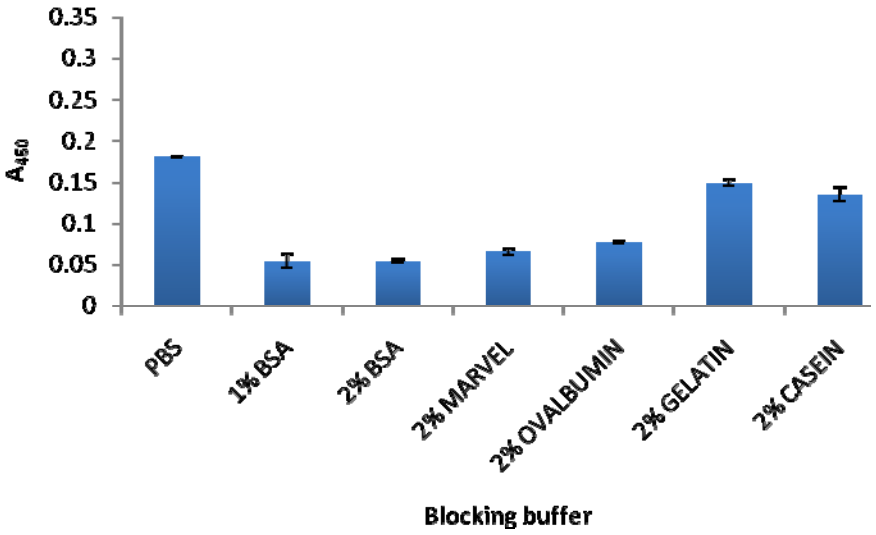


Figure 4.5 The results of blocking buffer optimisation. The results gave an indication of the effect of the omission of the primary antibody, thus allowing determination of the level of non-specific binding due to the conjugate. The higher the OD values, the higher the non-specific binding to the plate. PBS negative control showed the highest degree of non-specific binding as expected. Error bars show SEM.

Results presented in Figure 4.5 showed a further assay carried out in the absence of primary serum. This allowed the level of non-specific binding between the conjugate and the blocking agent to be determined. OD values were lower in comparison to those shown in Figure 4.4. Similar to the first experiment omitting antigen, 2% BSA gave the lowest value, although those using 1% BSA, 2% marvel and ovalbumin were also comparable. 2% BSA was thereby selected for inclusion within further investigations.

4.3.4 Optimisation of reaction incubation time

The optimisation of the incubation of substrate (TMB)/reaction time before stopping was also investigated. The addition of a stopping agent increased the sensitivity of the assay with the effect of incubation time upon final absorbance was shown in Figure 4.6. After 6 minutes, absorbance values increased at a slower rate than that observed prior to 6 minutes suggesting that this point was at the end of the linear phase of the reaction. Absorbance levels still increased after 6 minutes until 10 minutes (data not shown); however the rate of reaction was slower than that observed previously. An incubation period of 5 minutes at room temperature, in the absence of light was chosen for inclusion within the assay as this point was the end of the linear phase of the reaction before enzymatic reaction rates began to slow.

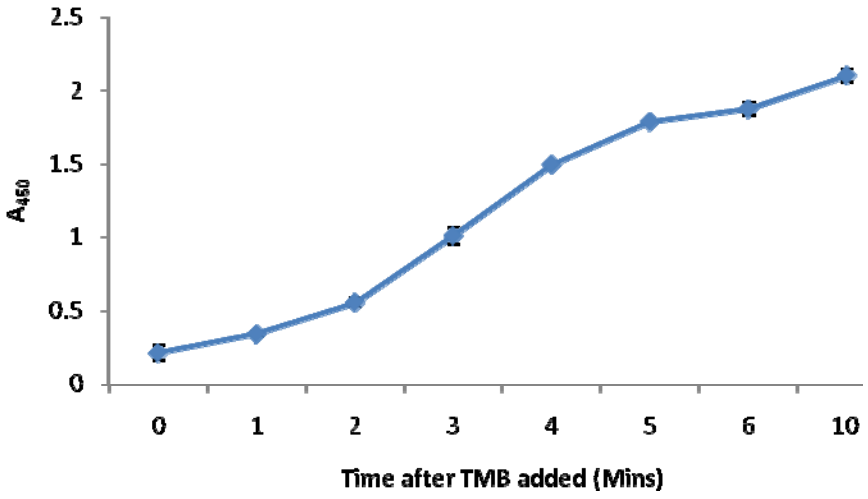


Figure 4.6 The effect of varying substrate incubation upon final absorbance readings. The above graph shows the effect of time upon final absorbance values after TNB substrate has been added to ELISA plate wells. There was a positive correlation between the length of incubation and level of absorbance measured. The rate of reaction is shown to alter after 6 minutes showing a change in reaction kinetics. Error bars show standard error of the mean (SEM). TMB: Trimethyl-Benzene.

4.3.5 Antibody titres of different bleeds raised to peptides

Data presented in Figures 4.7 and 4.8 shows the antibody titres of the different bleeds raised to Kpr and Eip peptides in rabbits. Each bleed taken at 7-day intervals in Figure 4.7 (see legend) showed an increase in strength with time of exposure to the immunogen. This was emphasised in the results as the terminal bleed showed the highest level of absorbance across all dilutions. As expected the pre-bleed provided a base-line of reactivity below all other bleeds. Bleeds raised to Eip however, showed a different trend (Figure 4.8). The terminal bleed appeared to be stronger when undiluted, although it dipped below that of bleed 3 after a 1/200 dilution, and dropped below even bleed 1 at 1/1000 dilution (taken at 7 days post exposure to the immunogen). The difference between bleeds was not considered significant and was attributed to a time delay during delivery of the individual bleeds.

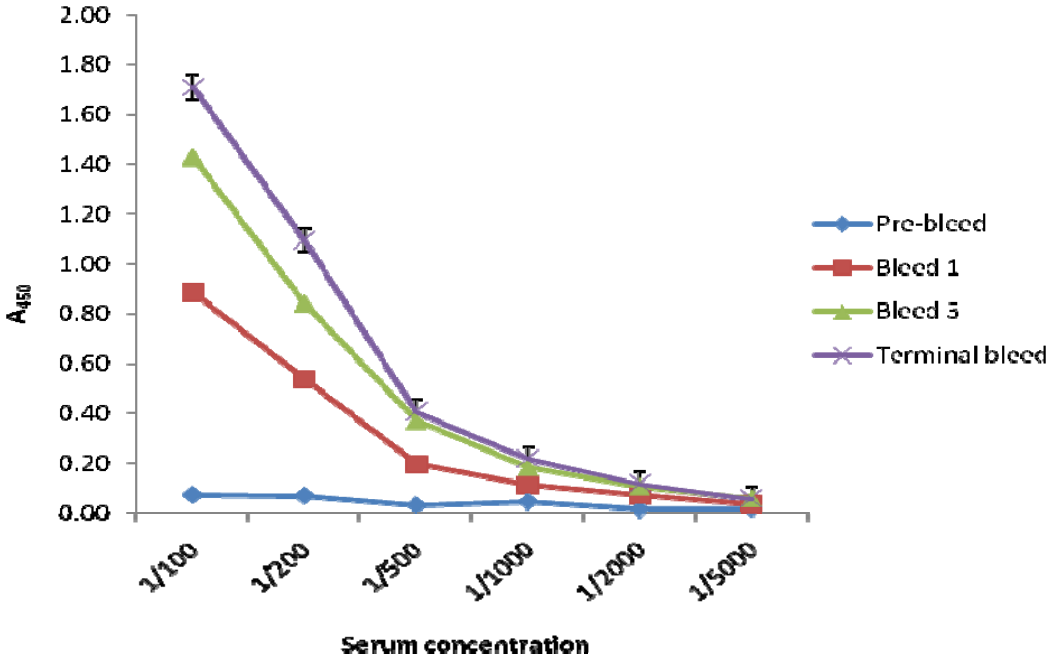


Figure 4.7 The reactivity of pre-bleed, bleed 1, bleed 3 and the terminal bleed with peptide GfPN1kpr. Each bleed raised against Kpr in the above figure represents a single harvest taken from the animal at different time points over a month post immunisation. Pre-bleed represents levels of anti-Kpr prior to immunisation with the peptide. Bleed 1 shows levels of antibodies after 7 days post-immunisation. Bleed 3 shows antibody titres after 21 days and the terminal bleed exhibits levels of anti-Kpr antibodies after the 1 month post immunisation period had expired. In theory, the terminal bleed should show the highest antibody titres, and this is reflected in the above graph. Error bars show SEM.

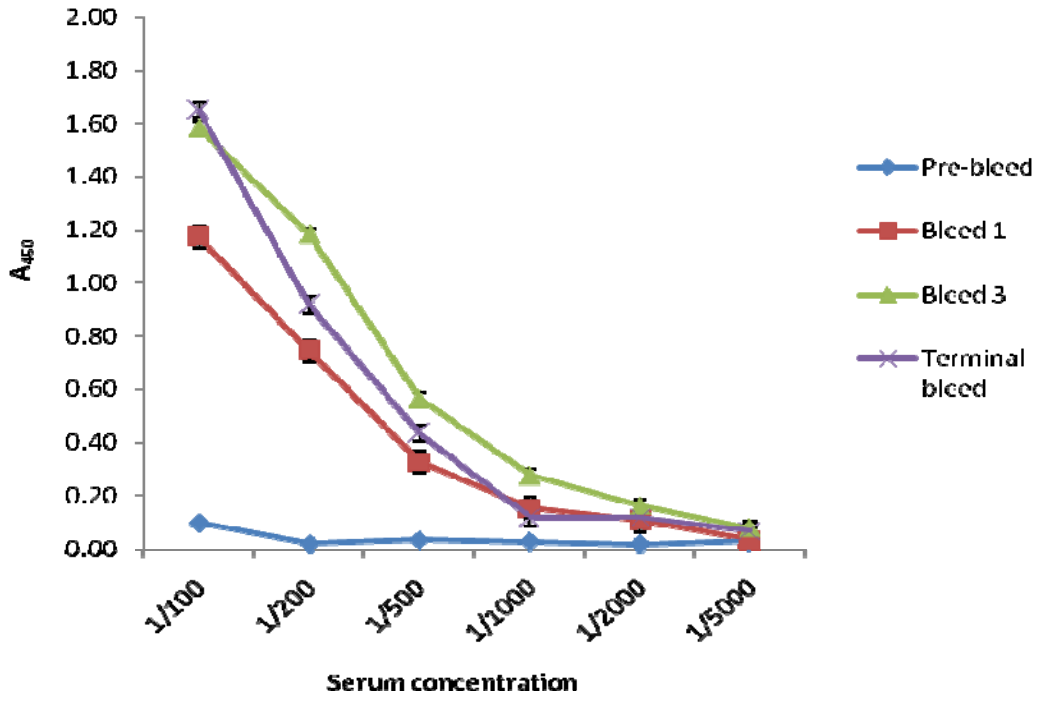
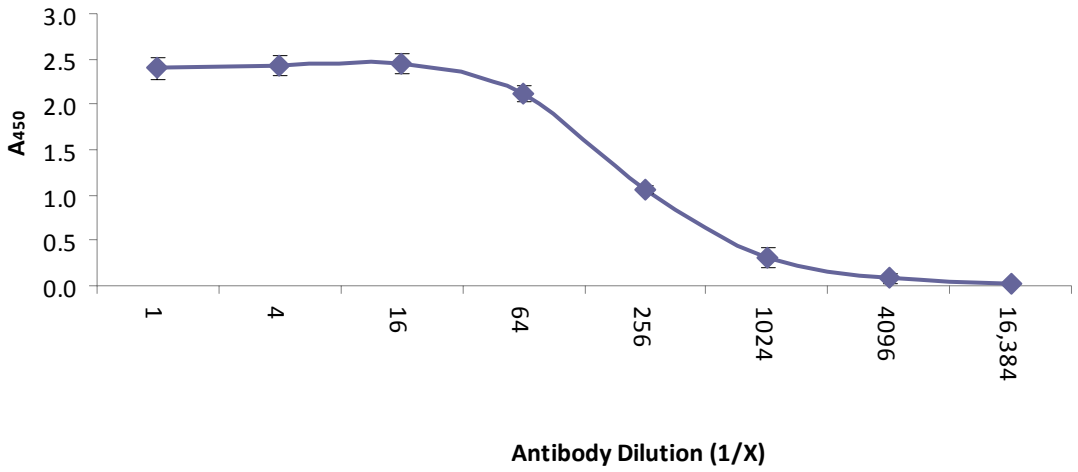


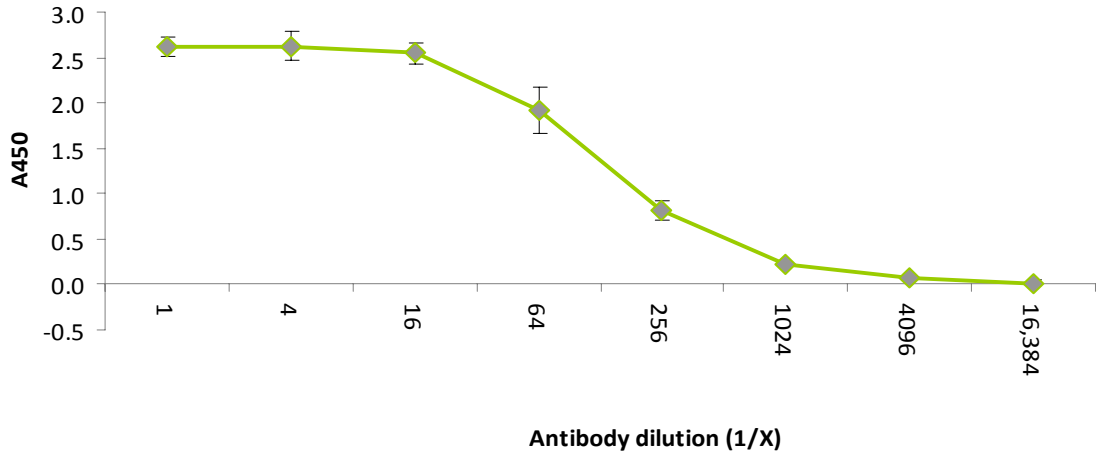
Figure 4.8 The reactivity of the pre-bleed, bleed 1, bleed 3 and the terminal bleed raised to GfPN2eip. Error bars show SEM.

4.3.6 Inhibition assay

Inhibition assays were run in order to confirm the specificity of both bleeds for their respective peptides. This also validated the bioinformatics work showing that the epitopes were able to interact with specific antibodies. Standard curves were constructed for each antiserum (Figures 4.9 - 4.10) and the dilution which gave an absorbance of 0.5 was noted. With both Kpr and Eip antisera, the bleeds were diluted to the appropriate dilution (1/1024) before being incubated with log dilutions of respective peptides taken from a log dilution series.



Figures 4.9 Standard curve generated by titrating out the GfPN1kpr terminal bleed. Error bars show SEM



Figures 4.10 Standard curve generated by titrating out the GfPN2eip terminal bleed. Error bars show SEM

Data presented in Figures 4.11 – 4.12 suggest that antisera raised to both peptides were specific for their respective antigens. Increases in absorbance correlated with increased dilutions of peptides used to incubate with primary antiserum before use in the assay. This increase in OD values was more prominent with the Eip peptide showing a greater specificity between the peptide and antibody. This was shown by a steeper gradient to the curve.

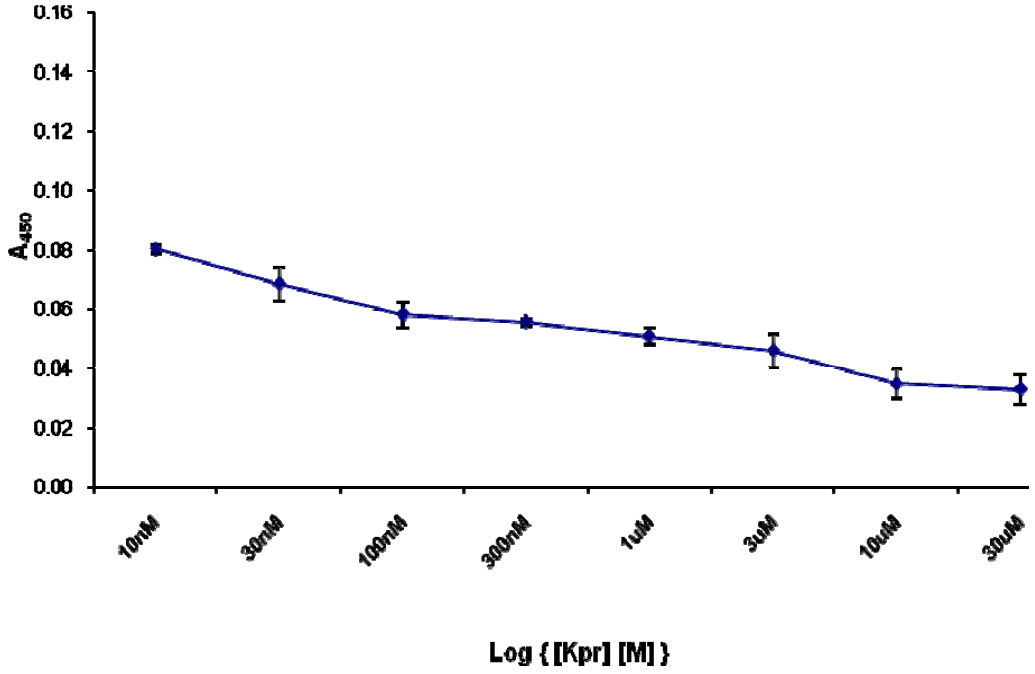


Figure 4.11 The effect of incubating log dilutions of GfPN1kpr peptide with specific polyclonal antiserum. Error bars indicate SEM.

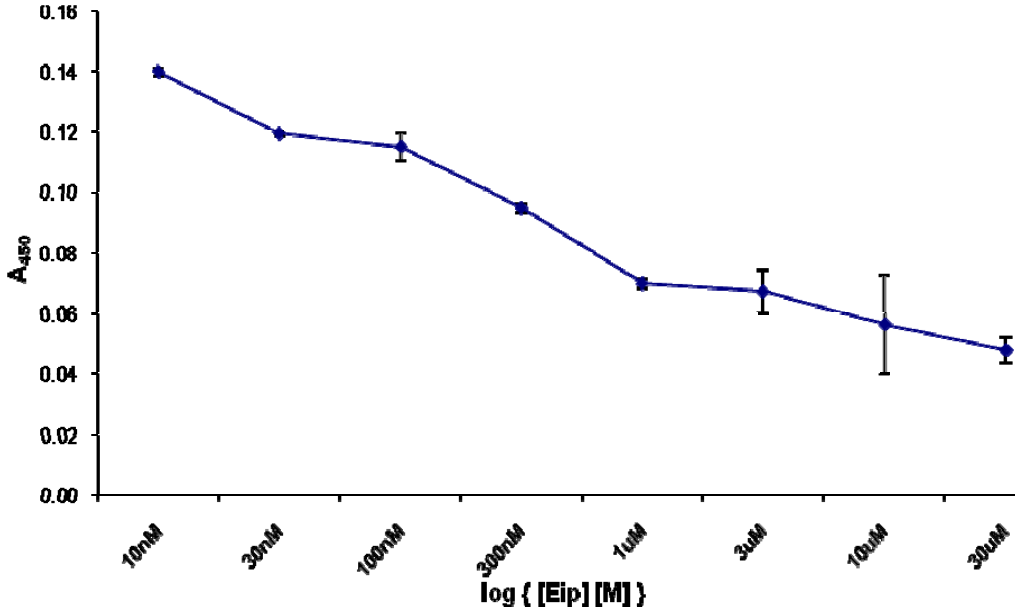


Figure 4.12 The effect of incubating log dilutions of GfPN2eip peptide with specific polyclonal antiserum. Error bars indicate SEM.

In order to demonstrate that the lack of cross-reactivity between peptides was not limited to the two tested, a third unrelated peptide (Negcont1) was also included within the study. 0.5 Log_{10} dilutions were made up as described previously and incubated with both Kpr (terminal bleed) and Eip (bleed 3) at the appropriate dilution (as determined by the standard curves). The reactivity of both bleeds were then investigated, using ELISA to assess any effects upon absorbance values (Figure 4.13). Both bleeds showed no overall change in trend and generally OD values were unaffected by incubation with different dilutions of peptide. This emphasised the overall specificity of the bleeds for their respective peptides, and their lack of cross-reactivity with other unrelated peptides.

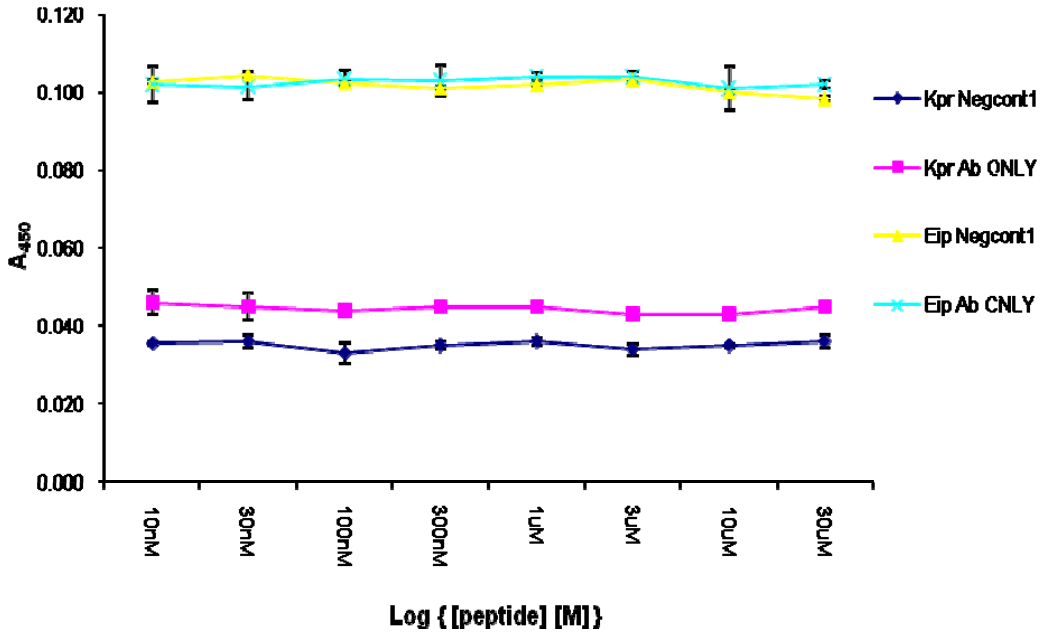


Figure 4.13 The effect of incubating primary serum with unrelated peptide (Negcont1) before use in novel optimised ELISA assay. Error bars indicate SEM.

4.4 Immunofluoresence of HERV-K expression in cultured cells

In order to test antibodies *in vivo,* antibodies raised to peptides GfPN1kpr and GfPN2eip in Immunofluoresence (IF) on a number of cell lines including MCF7, T47D and Human synoviocytes taken from a healthy donor. Figures 4.14 to 4.19 show typical images representative of staining of MCF7 and T47D cell lines. As positive controls, human HLA-ABC (MHC class I) was used to confirm the IF procedure had worked successfully and performed as expected (Figures 4.14 and 4.17). Negative controls included wells with no cells and well with cells but no primary antibody. Primary antibodies were used at 1/200 dilution as used with optimised ELISAs. Differences were apparent in the distribution of staining with Eip in MCF7 and T47Ds showing focal points of intense fluorescence (evident as bright spots as observed in figures 4.15 and 4.18). These tended to correlate with the location of the DAPI stain potentially suggesting this staining may be occurring within the nucleus. Such patterns were not observed with anti-GfPN1kpr, which presented a less intense but more evenly diffuse pattern (Figures 4.16 and 4.19).

Figures 4.20 to 4.23 show images of fibroblast-like synoviocytes (healthy donor) after staining with both Kpr/Eip and DAPI counterstaining. The staining of both HERV-K Gag and Env in these primary cells appears diffuse throughout the cytoplasm of the cell perhaps hinting to the localisation of the retroviral proteins in the cytoplasm where expected before packaging/budding. Cells stained with anti-GfPN2eip appeared to exhibit a higher intensity of staining compared to that observed with GfPN1kpr. This difference appeared lesser in fibroblast-like synoviocytes (derived from a healthy donor) with Kpr showing a similar intensity to Eip. This work validated the bioinformatic research carried out in Chapter 3, showing that antibodies to bind to these HERV epitopes within their *in-vitro* environment. Additionally, this staining shows that HERVs are expressed as proteins to some extent within cells. In those continuous cell lines tested (MCF7 and T47D) the extent to which this may occur may be due to the conditions the cells have been exposed to. The fact that this is evident within primary cells confirms the occurrence within an *in vivo* setting. 'Substrate only' negative controls were not run during immunostaining experiments.

MCF7 cells

Figure 4.14 MCF7 cells stained with positive control antibody (mouse anti-human monoclonal anti-HLA ABC), and labelled with rabbit anti-mouse – FITC conjugate. Green indicates the location of HLA proteins within the cell membrane. The blue stain shows DAPI stain which binds to dsDNA within the nucleus.

Figure 4.15 MCF7 cells stained using rabbit anti-Eip antibodies, and a goat anti-rabbit FITC conjugate antibody. Green staining shows location of HERV-K Env protein. Blue stain shows DAPI stain which indicates the location of the nucleus.

Figure 4.16 MCF7 cells after staining with rabbit anti- Kpr antibodies, with a goat anti-rabbit FITC conjugate. Green strain shows the location of HERV-K Gag protein. Blue stain at the centre of the cell shows DAPI stain bound to dsDNA within the nucleus.

T47D cells

Figure 4.17 Positive control for T47D cells using mouse anti-human HLA-ABC. The green staining indicates the location of MHC-I within the cells membrane. Blue DAPI stain shows location of nucleus in which dsDNA is present.

Figure 4.18 T47D cells stained with rabbit anti-GfPN2eip (Env). These regions show as green after conjugation with a FITC labelled goat anti-rabbit conjugating antibody. Blue regions after DAPI staining show location of DNA and nucleus.

Figure 4.19 T47D cells showing staining with anti-GfPN1kpr (Gag), conjugated with goat anti-rabbit FITC conjugate. Blue regions denote DAPI staining.

Fibroblast-like synoviocytes (healthy donor)

Figure 4.20 Fibroblast-like synoviocytes (donated by healthy individuals) using anti-GfPN2eip (Env) and labelled using goat anti-rabbit FITC conjugating antibody. Green regions denote expression of HERV-Env protein whereas blue staining shows location of location of dsDNA within the cells.

Figure 4.21 Fibroblast-like synoviocytes (donated by healthy individuals) using anti-GfPN2eip (Env) and labelled using goat anti-rabbit FITC conjugating antibody. Green regions denote expression of HERV-Env protein whereas blue staining shows location of location of DNA within the cells.

Figure 4.22 Fibroblast-like synoviocytes (donated by healthy individuals) using anti-GfPN1kpr (Gag) and labelled using goat anti-rabbit FITC conjugating antibody. Green regions denote expression of HERV-Gag protein whereas blue staining shows location of location of DNA within the cells.

Figure 4.23 Fibroblast-like synoviocytes (donated by healthy individuals) using anti-GfPN1kpr (Gag) and labelled using goat anti-rabbit FITC conjugating antibody. Green regions denote expression of HERV-Gag protein whereas blue staining shows location of location of dsDNA within the cells.

4.4.1 MCF7 cells

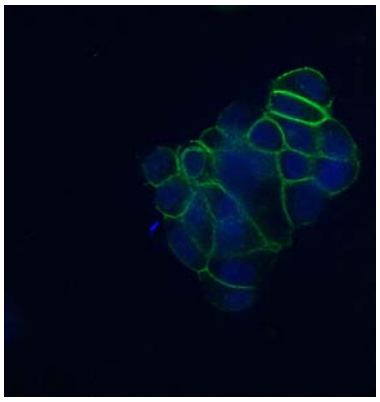


Figure 4.14

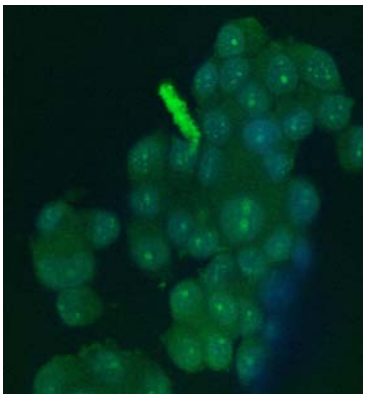


Figure 4.15

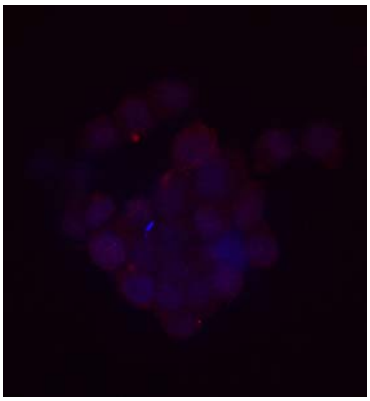


Figure 4.16

4.4.2 T47D cells

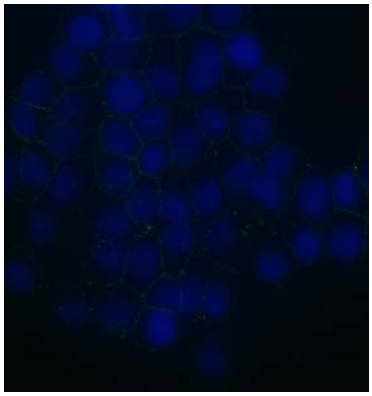


Figure 4.17

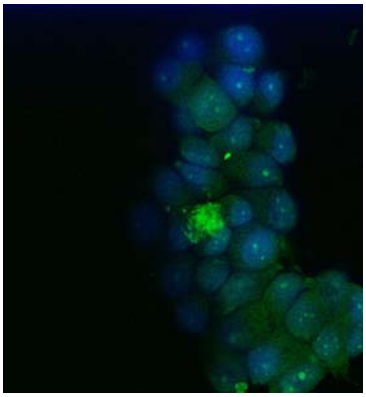


Figure 4.18

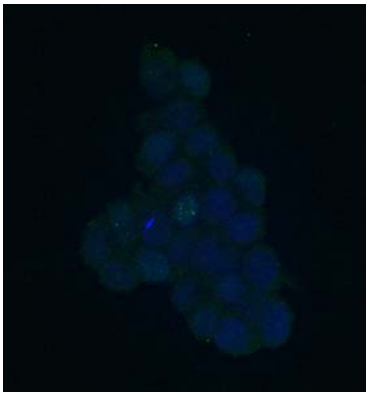


Figure 4.19

4.4.3 Fibroblast-like Synoviocytes derived from Healthy donor

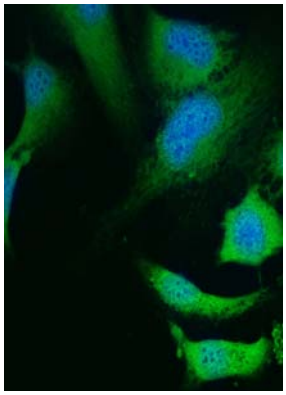


Figure 4.20

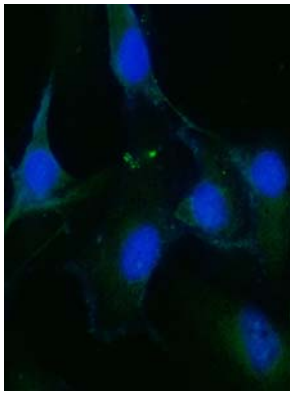


Figure 4.21

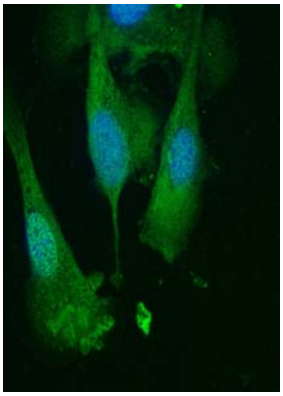


Figure 4.22

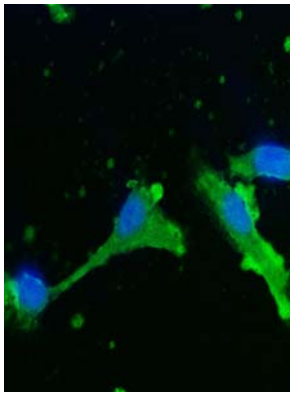


Figure 4.23

4.5 Testing of patient samples

4.5.1 Testing of patient samples using non-biotinylated peptides

Serum was extracted from the whole venous blood of RA patients (n = 48) and the reactivity to six peptides incorporating HERV epitopes (two biotinylated, four non-biotinylated), assessed and compared to control groups including inflammatory non-joint disease (IBD) (n = 20), non-inflammatory joint disease (OA) (n = 27), rheumatic disease (SLE) (n = 23) and healthy samples (NHD) (n =37).

Results in Figure 4.24 show the mean levels of reactivity in patient cohorts, for each peptide. All non-biotinylated peptides, across the patient groups appeared to exhibit a similar level of activity, although this varied somewhat between particular disease groups. Perhaps surprisingly, the reactivity of the patient samples to the negative control peptide (Negcont1) gave comparable means to other non-biotinylated test peptides although this appears to be more specific for particular disease groups, i.e. OA. Overall the level of reactivity to the non-biotinylated peptides was low in general being comparable in some cases to background reactivity, i.e. OD values of < 0.05. As a result, biotinylated peptides performed significantly better than their non-biotinylated counterparts, showing significantly increased OD values ($p < 0.0001$) and possibly hinting that biotinylation is essential for the accurate modelling of some peptides. Additionally, a disease specific significant increase in RA patient reactivity was observed when biotinylation was used. The improvement in orientation of the peptides and its consequent effect upon antibody recognition was also reflected in overall mean ODs per peptide (Table 4.3).

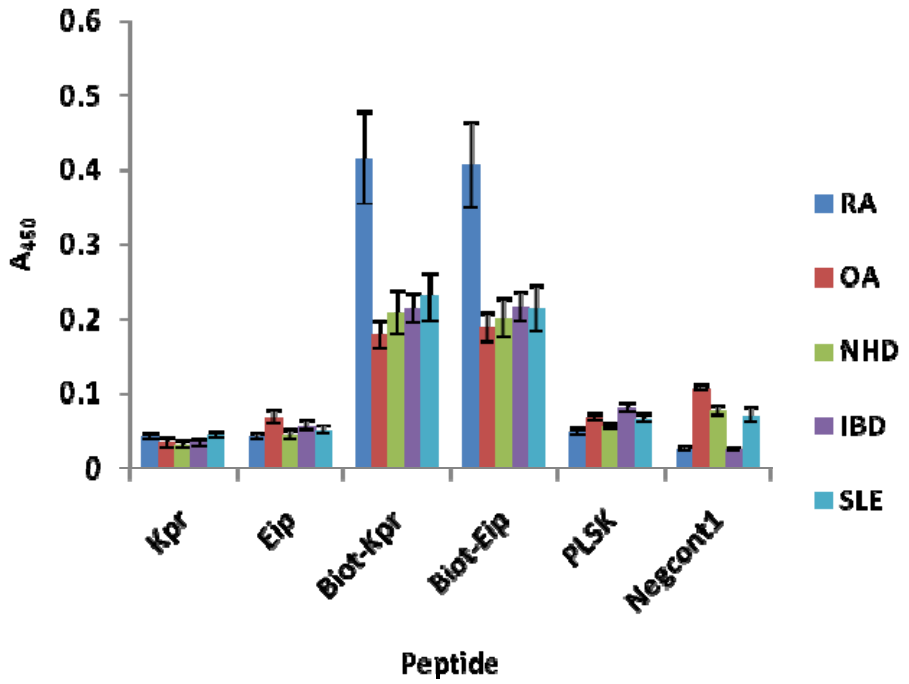


Figure 4.24 Mean OD values for patients in separate disease cohorts against different non-biotinylated peptides (Kpr, Eip, PLSK, Negcont1) and biotinylated (Biot-Kpr and Biot-Eip). Error bars indicate SEM. n = RA: 48, OA: 27, NHD: 37, IBD: 17, SLE: 23.

Peptide	Sample (n)	RA (48)	SLE (23)	OA (27)	IBD (17)	NHD (37)
Kpr	BD	0.042	0.043	0.033	0.034	0.031
	SF	0.034				
Eip	BD	0.042	0.051	0.068	0.057	0.044
	SF	0.037				
Biot - Kpr	BD	0.416	0.230	0.180	0.214	0.209
	SF	0.095				
Biot - Eip	BD	0.406	0.214	0.189	0.216	0.201
	SF	0.097				
PLSK	BD	0.049	0.068	0.069	0.081	0.057
Negcont1	BD	0.026	0.072	0.107	0.025	0.078

Table 4.3 Mean OD values for all disease cohorts against different peptides. The mean OD values per disease cohort per peptide, biotinylation of peptides gave significantly increased reactivity with antibodies. n = RA: 48, OA: 27, NHD: 37, IBD: 17, SLE: 23.

Using the GfPN1kpr (Kpr) peptide (Figure 4.25), it was apparent that SLE and RA patient sera showed the highest mean levels of reactivity (0.043, 0.042 and 0.07 respectively). OA patients could be classified into two groups based upon the OD observed. A cluster of 11 patient samples had OD values greater than the group average, whereas the remaining patients exhibited OD values of less than 0.02. The

group mean value of OA patient reactivity to Kpr was thus reduced due to this second 'lower' cluster. RA patients also showed considerable variation separating into two clusters (upper n = 10), resulting in a group mean value comparable to SLE. These high responders could not be separated upon the basis of clinical details

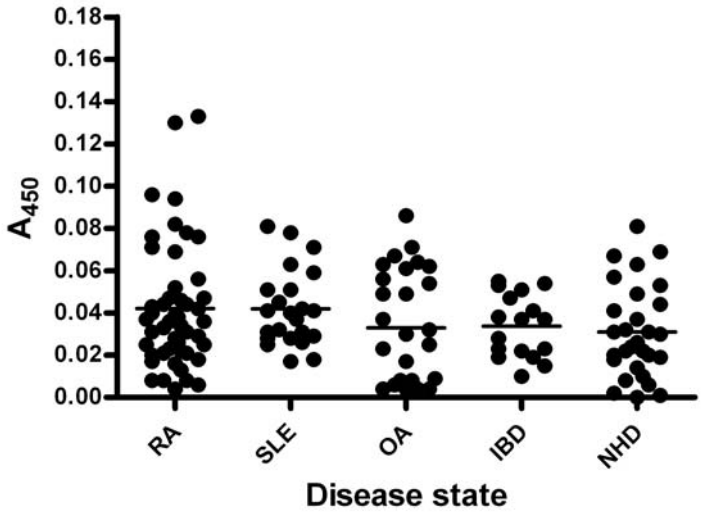


Figure 4.25 OD values for patient sera from different disease groups using non-biotinylated GfPN1kpr peptide.

Generally patients showed a more heterogeneous response to the non-biotinylated GfPN2eip (Eip) peptide, compared to reactivities observed with other peptides tested (Figure 4.26). OA patient cohorts exhibited the highest reactivities of all groups tested (mean OD value of 0.06848). Conversely NHD, SLE and RA patient samples exhibited the lowest levels of anti-Eip antibodies (mean OD values = 0.044, 0.051 and 0.042 respectively). OA patient data exhibited an increased level of variation, with results falling into 3 separate groupings. These comprised of 6 samples showing OD values in excess of the population mean, ten samples clustering around the mean value and 8 samples exhibiting OD values below the disease mean. The IBD patient group showed the third highest levels of reactivity to GfPN2eip, with the three highest OD values taken from three patients with active ulcerative colitis (patient details unknown). The differences between all groups were significant using one-way ANOVA analysis ($p = 0.0105$), with only RA being significantly decreased in comparison to OA ($p < 0.05$) according to Tukey-Kramer test.

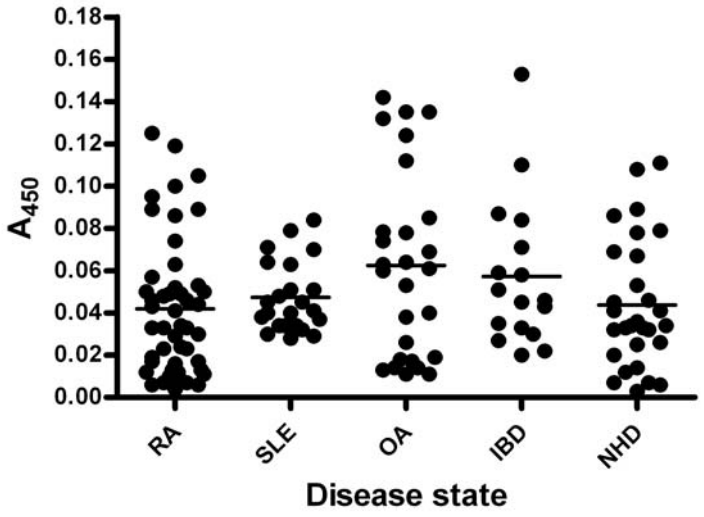


Figure 4.26 OD values for patient sera from different disease groups using non-biotinylated GfPN2eip peptide.

Reactivities to the PLSK peptide were most increased in OA and IBD patient cohorts (Figure 4.27). In the OBD cohort the highest ODs were observed in those patients suffering from inactive Crohn's disease. In the group of OA patients tested, four patients in particular showed high ODs to PLSK. Of these high responders, no correlations in age or sex were apparent (29F, 39M, 52M and 73F). SLE patients showed a decreased reactivity to PLSK in comparison to IBD and OA however this was still elevated above RA and NHD. In both cases this may hit at a role for HERV-K10 (Kpr) as opposed to HERV-K (PLSK) in these diseases. As with OA a small group of three high responding patients were evident (Figure 4.27). High responders in these groups and those in the SLE cohort showed no correlations in age or sex. NHD and RA exhibited comparable reactivities to one another (means of 0.057 and 0.049 respectively). Overall, anti-PLSK responses in RA patients were significantly decreased in comparison to those observed in OA ($p < 0.0001$) and IBD patients ($p < 0.01$).

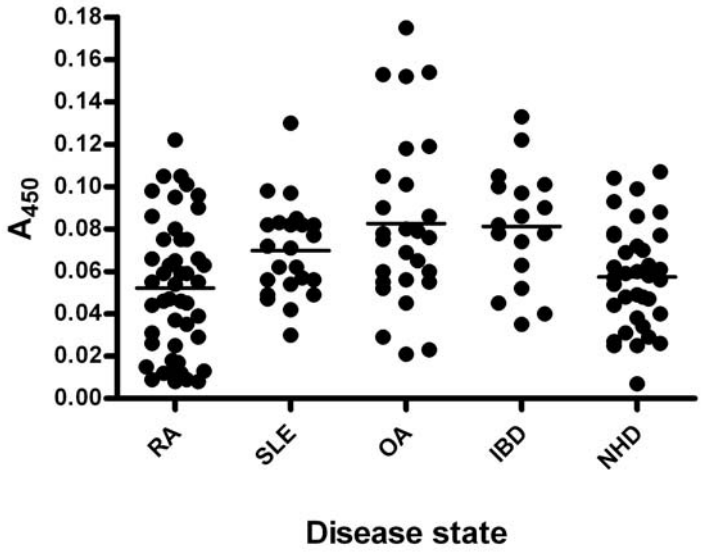


Figure 4.27 The distribution of OD values for reactivity of patient disease cohorts using the PLSK peptide

A negative control peptide 'Negcont1' was also used to test patient sera, in order to determine reactivity to random non-HERV epitopes and to eliminate bias from the results (Figure 4.28). Patterns of reactivity observed with this peptide were intriguing as some disease groups, i.e. RA and IBD, exhibited lower OD values with this peptide whilst other groups such as OA, NHD and SLE, exhibited similar levels of reactivity than expected. SLE and healthy cohorts both showed levels of absorbance and distribution of OD values, comparative to one another (mean OD values = 0.07 and 0.073 respectively) with sample absorbance values varying more from the cohort mean (minimum and maximum values of both 0.02 - 0.14 and 0.01 – 0.14 respectively).

The differences between all groups were significant according to one-way ANOVA analysis ($p < 0.001$) with Tukey-Kramer's test showing the mean OD value from the OA cohort was significantly higher than other sample populations.

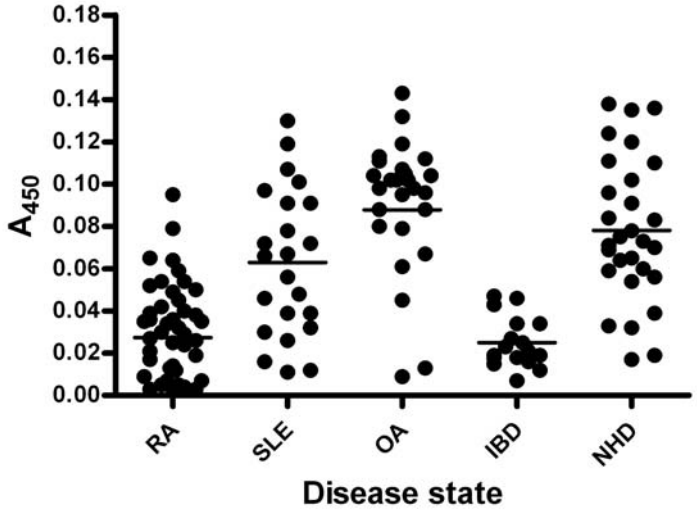


Figure 4.28 Distribution of reactivity in patient sera using the Negcont1 peptide

4.5.2 Biotinylation of GfPN1kpr and GfPN2eip

The reactivity of patient sera with biotinylated forms of the peptides Kpr and Eip was also explored in order to investigate different methods of antigen presentation (Figure 4.29 – 4.30). Overall absorbance values were significantly increased when compared to those produced with non-biotinylated peptides (~0.25 compared to ~0.05 respectively). The effect of biotinylation compared to non-biotinylated peptides was highly significant ($p < 0.0005$) as shown by ANOVA analysis. Of the two peptides, biotinylation of Kpr produced values of increased significance over Eip, despite the fact that Eip showed increased significance over Kpr in its non-biotinylated form. Overall both peptides showed similar levels of reactivity (Figures 4.24 and Table 4.3).

In terms of disease state, biotinylation had the greatest effect upon ODs taken from RA patient samples resulting in an increase in significance of the results for that group.

RA patient serum presented the highest reactivity to the biotinylated Kpr peptide (mean OD value = 0.42) (Figure 4.29) showing the separation of RA patients into two clusters, i.e. those with OD values less than the cohort mean (*n = 30*) and those patients with OD values greater than the cohort mean *(n = 18)*. RA serum overall was significantly increased over all other groups by one way ANOVA analysis [with p values ranging from <0.05 (SLE) to < 0.001(NHD)]. SLE, similarly to non-biotinylated Kpr, also showed a higher degree of reactivity to this peptide in

comparison to other disease cohorts (mean OD value = 0.23). The reactivity of OA patient samples showed reduced activity with the biotinylated peptides, suggesting the opposite of the results exhibited with the non-biotinylated antigen (mean OD value = 0.18). Of all disease groups tested, RA patients showed the greatest increase in significance and levels of reactivity to the biotinylated Kpr peptide, RA patients fell into two distinct subsets - those with OD values higher and those with OD values lower than the group mean. High responders had an average age of 60 years old (range 40-93 years) and comprised of 7 males and 13 females (mean = 0.89). The 7 males were made up of 6 Caucasians and 1 Asian, whereas the 13 females were made up of 12 Caucasians and 1 patient of black African origin. The two highest responders in particular were 79 and 93 years old respectively. Both were male Caucasians with 22 and 26 years disease duration, and both exhibited high levels of RF levels (128 mg/ml).

When healthy samples were tested using biotinylated Kpr, a cluster of four samples gave particularly high readings in comparison to the rest of the population, which showed very little variation from the mean. Further details for these patients - 71, 78, 79 and 81 however were not available, therefore no further conclusions could be drawn as to why they may present increased levels of antibodies to biotinylated Kpr.

As shown in Table 4.3, biotinylated Eip showed a similar distribution of reactivities with levels of absorbance increased by almost 10-fold, compared to those observed using non-biotinylated forms of the same antigen (Figure 4.30). Levels of absorbance were significantly higher than those of non-biotinylated peptides ($p < 0.0001$ by one-way ANOVA) although with the exception of OA patient OD values, all other patient cohorts exhibited reduced reactivity to this peptide compared to biotinylated Kpr. RA patients presented the highest levels of antibodies to the epitope (mean OD value = 0.4062) and this was significantly higher than all other groups [p values ranging from <0.05 (IBD) to <0.001 (OA) by one way ANOVA analysis]. Once more the prevalence in OA patients was lower in comparison to other groups and non-biotinylated peptides (mean OD value = 0.1885).

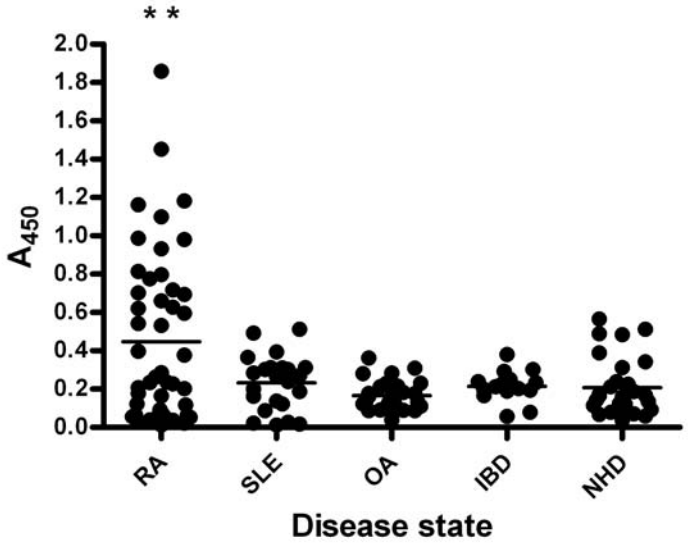


Figure 4.29 OD values for patient serum from different disease groups using biotinylated GfPN1kpr peptide. ** indicates levels of significance of RA over other disease cohorts. Lines show group means.

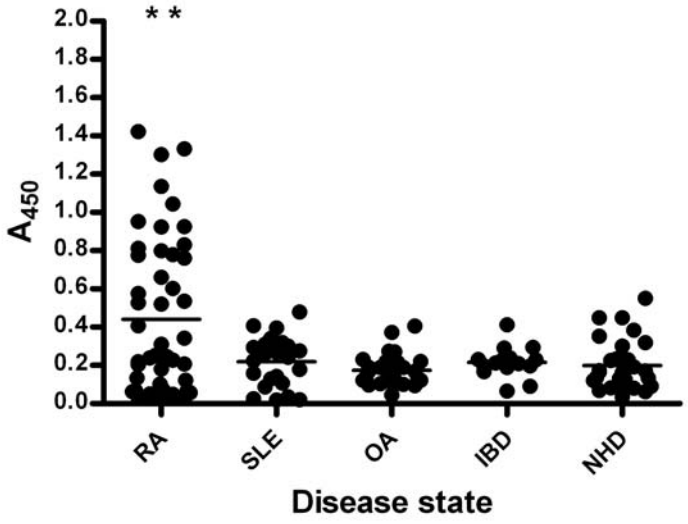


Figure 4.30 OD values for patient serum from different disease groups using biotinylated GfPN2eip peptide. ** indicates levels of significance of RA over other disease cohorts. Lines show group means.

Peptide/Disease	Variance	SLE	OA	IBD	NHD
Kpr	**0.0576**	**NS**	**NS**	**NS**	**NS**
Eip	**0.0365**	**NS**	**NS**	**NS**	**NS**
Biotinylated Kpr	**p <0.0001**	**p <0.05**	**p <0.001**	**p <0.05**	**p <0.01**
Biotinylated Eip	**p <0.0001**	**p <0.01**	**p <0.001**	**p <0.05**	**p <0.01**
PLSK*	**p <0.0001**	**NS**	**p < 0.001**	**p < 0.05**	**NS**
Negcont1*	**p <0.0001**	**p < 0.001**	**p < 0.001**	**NS**	**p < 0.001**

Table 4.4 Levels of significance per peptide of reactivity of RA patients, in comparison to other disease cohorts. P < 0.05 was considered to be significant.

***PLSK and Negcont1 showed RA to be significantly decreased in comparison.**

The level of significance of reactivity in RA patients compared to other disease cohorts is shown in Table 4.4. Table shows the degree of significance by which the reactivity to each peptide of RA patients differ from reactivities observed in other control groups. It is clear from this table the effect of biotinylating the peptides and the difference in patient samples OA patients are shown in the analysis as that group most significantly different from RA patients ($p < 0.001$) with IBD showing the least degree of variance ($p < 0.05$). Both PLSK and Negcont1 both show an increased level of variance over reactivities obtained during the analysis (Figure 4.27 and 4.28 respectively). However RA patients exhibited OD values significantly less than control groups.

4.5.3 Inhibition studies to confirm specificity of patient bleeds for peptides

After serological surveys of patient reactivity to peptides, inhibition studies were carried out on random RA patient sera in order to confirm the reactivity observed in patient samples was reflective of the reactivity to the peptides, i.e. to the HERV-K epitope. Three sera were selected at random and standard curves constructed over a two-fold dilution series (Figure 4.31). Figures 4.32 - 4.34 show the inhibition curves as bleeds after incubation with log dilutions of specific peptide (and irrelevant peptide) prior to application within the ELISA assay. All three patient sera showed antibody responses specific for the Kpr peptide. Incubations with log dilutions of negative control peptide showed no overall change in absorbance suggesting reactivity of patient sera is indicative of an antibody response specific for HERV-K epitopes.

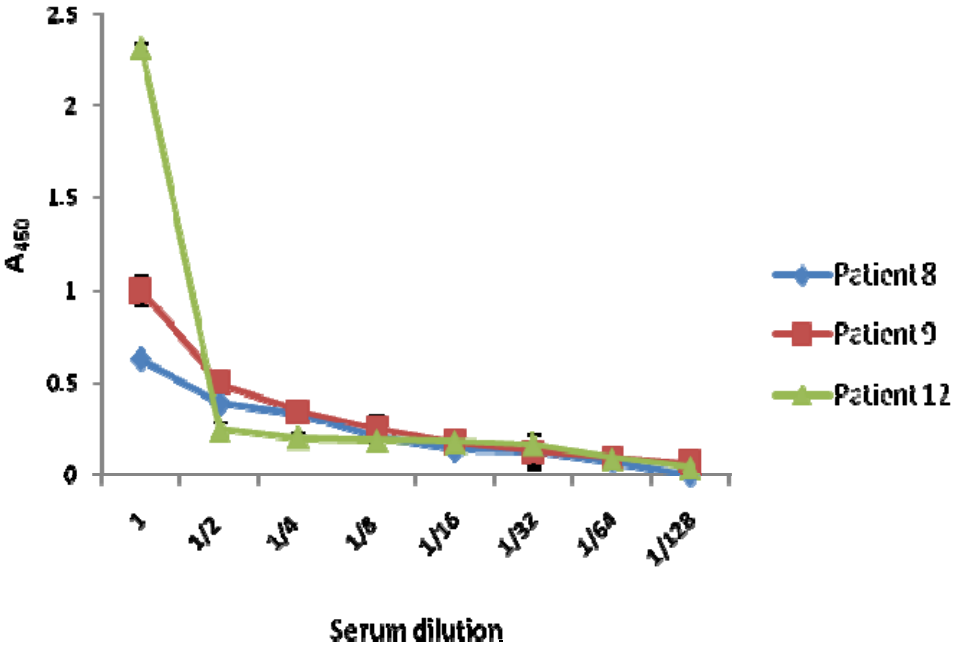


Figure 4.31 Standard curves for patient 8, 9 and 12. Plots show standard curve of antibody for the Kpr peptide alone across 8 dilution series. Error bars show SEM. Blue: Patient 8; Red: Patient 9 ; Green: Patient 12.

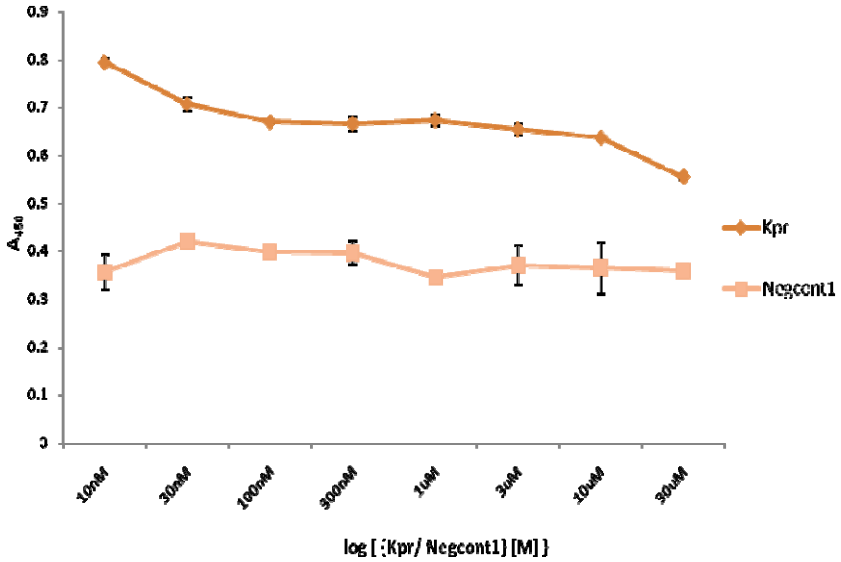


Figure 4.32 Inhibition curve showing degree of specificity for patient bleed GF8 for epitope Kpr over log dilution series of Kpr peptide. As peptide decreases, increased peptide specificity results in increased absorbance. The second plot shows reactivity of the patient serum with a negative control peptide (Negcont1). Overall no change in absorbance is observed over the log dilution series of peptide incubated with the serum. Error bars indicate degree of SEM.
Orange: Kpr ; Beige: Negcont1.

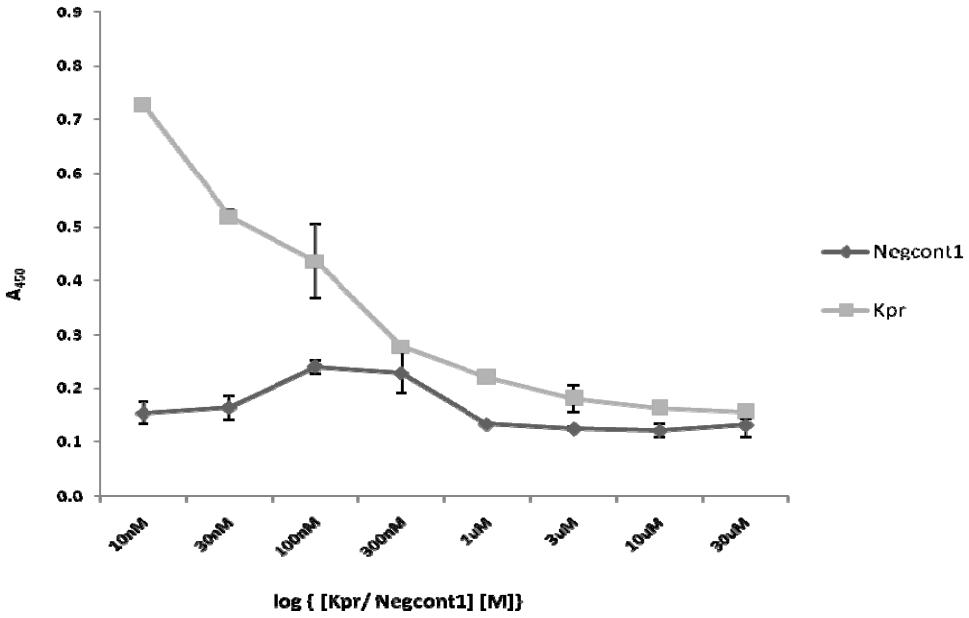


Figure 4.33 Inhibition curve showing degree of specificity for patient bleed GF9 for epitope Kpr over log dilution series of Kpr peptide. As peptide concentrations increase, increased peptide specificity results in decreased absorbance. The second plot shows reactivity of the patient serum with a negative control peptide (Negcont1). Overall no change in absorbance is observed over the log dilution series of peptide incubated with the serum. Error bars indicate degree of SEM. Light grey: Kpr; Dark grey: Negcont1.

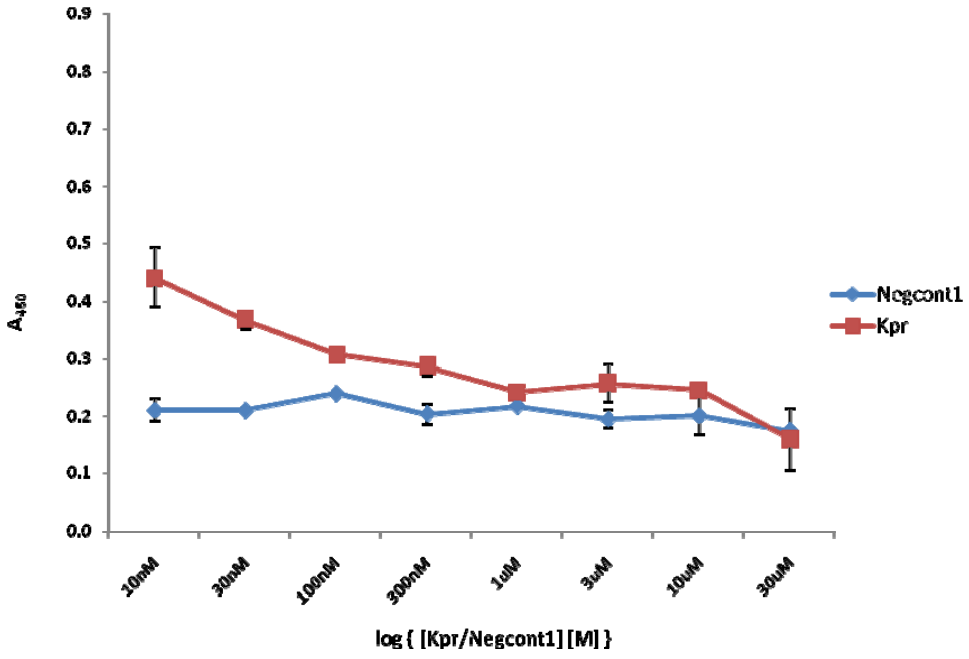


Figure 4.34 Inhibition curve showing degree of specificity for patient bleed GF12 for epitope Kpr over log dilution series of Kpr peptide. As peptide decreases, increased peptide specificity results in increased absorbance. The second plot shows reactivity of the patient serum with a negative control peptide (Negcont1). Overall no change in absorbance is observed over the log dilution series of peptide incubated with the serum. Error bars indicate degree of SEM. Red: Kpr; Blue: Negcont1.

4.5.4 Levels of anti-HERV-K Gag and Env within the Synovial fluid & influence of rheumatoid factor

4.5.4.1 Determination of HERV-K Gag & Env specific antibodies within paired serum and synovial fluid

Paired Synovial fluid (SF) and blood samples taken from RA patients were tested using the novel ELISA assay, in order to identify differences in anti-HERV-K antibody levels within and external to the joint as increased levels within the joint may indicate a locally driven immune process. Non-biotinylated Kpr and Eip both showed increased levels of reactivity in the SF, compared to that in the blood (mean OD values of 0.03 and 0.04 respectively) (Figure 4.35 & Table 4.3). Peaks in paired samples correlated as observed in patients 9 and 16. The exception to this was patient 13 who had elevated anti-HERV-K antibodies to both peptides in their synovial fluid, but not blood. Levels of anti-Kpr and Eip showed increased variation in the SF with more peaks and troughs in the data, in comparison to the blood whose levels appeared more constant. Absorbance values between the Blood (bld) and Synovial fluid (SF) when tested using non-biotinylated peptides were not significantly different from one another using a 2-tailed paired t-test; with Mann Whitney rank test (P values for Kpr and Eip were 0.3507 and 0.7278 respectively).

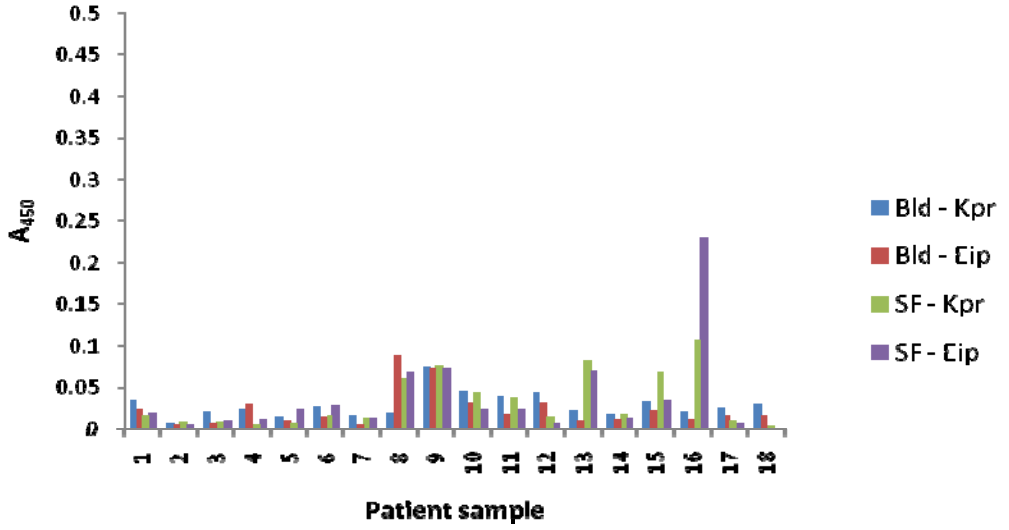


Figure 4.35 OD values from testing RA patient bld sera vs. SF for non-biotinylated Kpr and Eip.

With the non-biotinylated antigen, the synovial fluid showed an increased level of reactivity compared to the blood (means of 0.094 and 0.067 respectively) (Figure

4.36 and Table 4.3). This however may be misrepresentative of the whole dataset and skewed by a small number of patients who showed large increased in the SF, i.e. patients 13, 15 and 16. Biotinylated peptides also showed increased levels of reactivity to HERV peptides in the SF (Figure 4.36). This was particularly evident with patients 2, 4, 7, 10 and 14. Although both biotinylated Kpr and Eip shared similar trends in reactivity, Eip appeared slightly increased in a small number of patients, i.e. sample numbers 4 and 14. Patients 4, 10, 14 and 18 showed the highest levels in both the blood and synovial fluid.

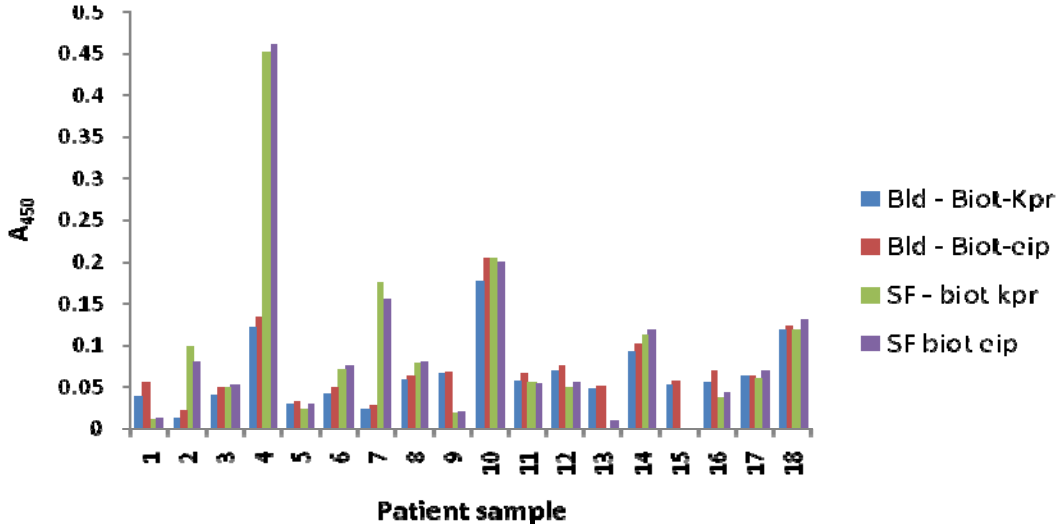


Figure 4.36 OD values from testing RA patient bld sera vs. SF for biotinylated Kpr and Eip.

Presentation	Peptide	Significance (p values)		
		Bld Vs. SF	SF peptide 1 Vs. Peptide 2	Non-biot. Vs. Biot
Non-biotinylated	Kpr	0.5680	0.6481	0.0553
	Eip	0.3328		0.0878
Biotinylated	Kpr	0.2109	0.8195	
	Eip	0.3543		

Table 4.5 Levels of significance between peptides in Bld and SF depending upon their method of presentation.

Using paired t-tests for analysis of paired SF and Bld samples it was concluded that despite reactivity in the SF being elevated to that observed in the blood (Table 4.3), this difference was marginal and not significant (Table 4.5). This lack of significance also extended between peptides and also, surprisingly to the method of antigen presentation, i.e. biotinylation of the peptides.

4.5.4.2 Quantification of RF in serum and its influence upon OD values

RF was shown to correlate with increased reactivities to peptides in peripheral blood (Figure 4.37) with levels particularly elevated in patients 8, 10, 14 and 19. Patients 1 - 18 showed lower OD values to HERV peptides than patients 19 – 41. These patients also exhibited lower levels of rheumatoid factor than patients 19 – 41. It could be concluded that the increased levels of HERV-K antibody responses may be partially masked (or contributed to) by the action of rheumatoid factor.

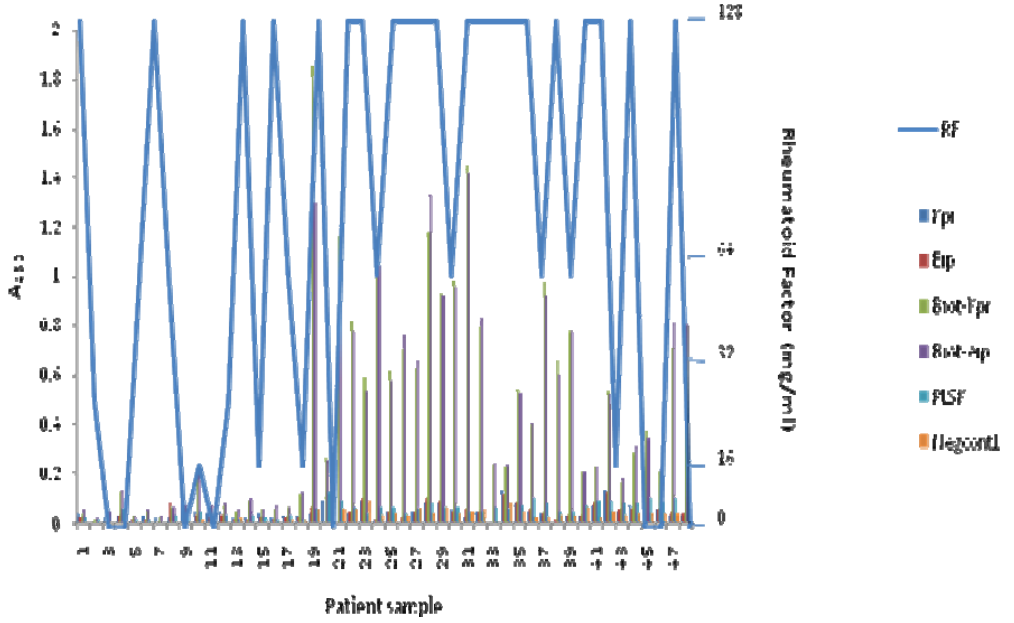


Figure 4.37 Variation in RF levels and its correlation with absorbance values for all 5 peptides tested within the peripheral blood of RA patients. Data on the graph shows levels of RF and correlation with reactivity with each of the five peptides.

4.5.4.3 Presence of RF in paired blood and synovial fluid samples

Figure 4.38 shows levels of RF in paired blood and synovial fluid. Sample sets assayed included 14 RA patients and 3 OA. Of these, 6 RA patients exhibited increased levels of RF in their SF (128 mg/ml). Of these samples 4/6 exhibited equivalent levels in the paired blood. The remaining 2/6 showed the presence of lower levels of RF in their blood. Three RA patients showed increased RF (albeit at lower levels of 16-34 mg/ml) with no detectable levels of RF in the paired SF. The 4 remaining RA patients lacked any detectable levels of rheumatoid factor in both blood and synovial fluid. All 3 OA patients tested for RF exhibited high levels (128 mg/ml) in their SF, with just 1/3 showing equivalent levels within their blood. Overall the RA group exhibits a heterogeneous distribution with RF seropositive and seronegative patients. Additionally, it can be concluded that in the majority of patients (10/17), the levels of rheumatoid factor were equal to or greater within the synovial fluid than they were in the blood irrespective of disease. Of the remaining 7 patients, 4 showed no evidence of RF when tested (patients 3, 9, 13 and 19), whereas the remaining 3 showed levels between 16-32 mg/ml in their peripheral blood (patients 12, 14 and 17).

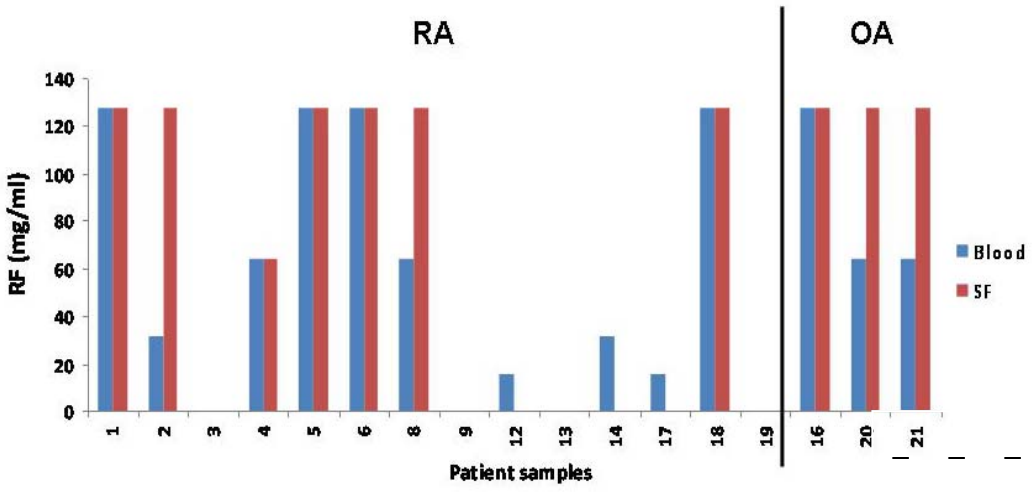


Figure 4.38 Levels of RF in paired samples of blood and synovial fluid in RA and OA patient samples. RA: 1 - 19; OA : 16, 20, 21.

Figure 4.39 showed the levels for rheumatoid factor in paired samples of blood and synovial fluid. Sample sets comprised of 14 RA patients and 3 OA patients. Of 6 RA patients that showed high RF levels of 128 mg/ml, 4 were also accompanied with high levels of RF in the blood. The remaining 2 samples (taken from patients 2 and 8)

showed lower levels of RF, despite the fact that at least one – patient 8 had disease duration of greater than 10 years. Three RA patients showed levels of RF in blood, but not in the SF (patients 12, 14 and 17). Of these patients, levels of RF were low, i.e. 16-32 mg/ml. These patients were male and 2 females. They had an average age of 60.6 years (range of 47 to 68 years). All were white Caucasians, with erosive disease. Disease duration ranged from 3 - >20 years. The remaining 4 RA patients were rheumatoid factor negative. Of these 2 were of an average age of 27.5 years (range 24 and 31 years) with the other patients of an average age of 73.5 years (range 63 and 84). All were female Caucasians with none with any record of erosive disease. Of the OA patients tested (patients 20, 21 and 16) all showed maximum levels of RF in the SF, with 64 – 128 mg/ml in the paired blood.

4.6 Discussion

4.6.1 Experimental optimisation

Immunoinformatic techniques were used to predict linear epitopes of HERV-K and K (HML-2) Gag and Env putative proteins. These epitopes were identified due to their increased antigenicity and thus potential to interact with other immune proteins, such as antibodies and complement. Highlighted epitopes on HERV proteins were cross-referenced with those epitopes identified upon 28 autoantigens, in both RA and rheumatic diseases. By identifying regions of homology between different epitopes, this may hint at the potential for molecular mimicry to induce disease pathogenesis (Oldstone, 2005).

Those candidate epitopes with the best alignments of autoantigen sequences (and thus the greatest potential for molecular mimicry) were short listed and the best three were synthesised as short peptides (Section 4.2.2). A novel indirect ELISA was developed and optimised for the detection of antibodies in patient serum samples, incorporating the novel HERV epitopes. Polyclonal antisera generated to the commercially produced peptide antigens were tested for their strength, sensitivity and specificity and levels of cross reactivity between bleeds (Figure 4.7 – 4.8). Peptide antisera were also investigated as they would be invaluable tools in other applications such as immunofluoresence. Other variable factors such as optimal blocking agents, microtitre plates, antigen/antibody dilutions and time of substrate incubations were also investigated in order to establish the highest level of specificity, sensitivity and accuracy (Figure 4.2 – 4.7). The resultant assays were robust ELISA's which provided invaluable tools for the determination of levels of prevalence of anti-HERV-K(HML-2) antibodies in RA patients and other disease cohorts within this investigation (Figures 4.23 – 4.30). This assay was then used in inhibition studies, thereby validating the work described in Chapter 3 and demonstrating the specificity of the bleeds raised to the peptides. Furthermore, this emphasised the overall specificity of the bleeds for their respective peptides, and the lack of cross-reactivity with other peptides (Figures 4.32 – 4.35).

4.6.2 Patient samples

The novel optimised HERV-K ELISA was used to survey patient sera, identifying different levels of reactivity to HERV-K epitopes across different disease groups. Overall, biotinylated peptides showed significantly increased levels of reactivity with patient sera compared to non-biotinylated peptides, whose reactivites were homogenous albeit at a much lower level of reactivity, i.e. increased slightly above atypical background levels – OD values of ~0.05.

When tested against non-biotinylated Kpr peptide, in a pattern typically observed in all disease groups, OA samples separated into two discrete clusters (Figure 4.25). Of these, the higher responding group comprised of 8 patients. All 8 samples exhibited increased levels of anti-HERV-K Gag with ages ranging from 38-73 years with a mean of 59.1 years. All were rheumatoid factor negative with the exception of one patient from New Cross who had a low RF count of 16 mg/ml, accompanied with erosive damage (as would be expected in OA). Of these 8 patients, 5 were Caucasian. Information was not available for the ethnic origins of the remaining three patients.

Overall there was little variation between mean reactivities of different disease groups, despite the heterogeneity of each group with the non-biotinylated Kpr peptide (Figure 4.25). SLE and RA showed the most reactivity, however a small number of RA patients showed the highest OD values of all samples tested. This may suggest a subset of RA patients develop antibodies to this epitope. This may also be indicative of the nature of RA as an autoimmune disease with many autoantibodies specific for a large number of targets.

When non-biotinylated GfPN2eip peptides were used as the coating antigen; NHD, SLE and RA patients' samples exhibited the lowest levels of anti-Eip antibodies (means of 0.044, 0.051 and 0.042 respectively) (Figure 4.26). The Eip peptide was modelled upon the superantigen region of HERV-K10 Env. Responses in the majority of samples tested showed a wide distribution of reactivities across all disease groups. In many disease groups, the outliers were also able to influence overall group means causing them to appear more reactive as a group than maybe they actually are, e.g. OA and IBD (Figure 4.26). OA patients fell into one of two clusters with absorbance values above and below the cohorts mean value. 6 high responding patients whose OD values were increased over the group mean (OD's varying between 0.01 - 0.015) comprising of 3 male and 3 females, ages 29-73 (mean age 53.8 years). At least 3/6

were Caucasian with 2/3 of those with highest values were under 40 years of age. All 6 exhibited high levels of RF between 64 and 128 mg/ml when tested (Figure 4.37). The lower cluster (Figure 4.26) was made up of 8 patients, ranging between 38 – 82 years (average age of 58.5 years), 5 were female and 3 male, 7/8 were Caucasian with the remaining patient of Asian origin. Rheumatoid factor was recorded in 4/8 samples, with patients having recorded an average disease duration of 9.75 years. Of the three lowest scoring OA samples 2/3 were male, under 55 (55 & 47 years) with greater than 10 years' disease duration. Reactivity to Eip in patient samples showed a higher level of variation than other non-biotinylated antigens tested (Figure 4.24). Additionally, this was the only peptide that did not share sequence homology to an autoantigen such as Collagen II (Kpr) or Collagen I (PLSK) which may have been a factor contributing to the lower values observed. A subset of OA patients and IBD patients however displayed increased reactivities to this peptide, modelled upon the HERV-K superantigen region. Whether this may play a role in any of these diseases may merit further investigation.

The PLSK peptide was synthesised in-house at the University of Wolverhampton and tested upon patient samples alongside commercially synthesised peptides. This short stretch of amino acids was modelled on a similar region of the Gag protein to GfPN1kpr. A 4 amino acid difference between the two, due to derivation from different HERV-K subfamilies, appeared to have significant influence over the reactivity of patient cohorts (Figure 4.18). Overall IBD patients were the disease group that displayed the highest OD values with PLSK. The two highest responding patients were sufferers of active Crohns disease and aged 41 and 43. RA patients also displayed a degree of variation as patients followed a similar distribution as that observed in that observed in non-biotinylated Kpr, i.e. clustering into two discrete groups (Figure 4.27) although overall, RA patients exhibited marginally increased reactivities to PLSK than to Kpr (Table 4.3). Of the 13 patients comprising the higher responding cluster, they exhibited an average age of 54 years (range 50 – 82 years), constituting 2 males and 11 females. Of those, 3 females and 1 male were of Asian origin with the remaining 9 of Caucasian descent. Disease duration of group members showed an average of 15.9 years (range 8-29 years) with 10/13 being RF positive. Although there was incomplete data to make conclusions for all members concerning development of erosive disease, 6/11 had records of erosive disease whilst the remaining 5 showed no records of erosive disease.

Although OD values between Kpr and PLSK were comparable, the main differences observed between the two peptides were in the OA and IBD control groups which showed increases in reactivity, (Figure 4.27). The negative control peptide 'Negcont1' was included to prevent any bias in the resulting data generated from the study (Section 3.3.6) and was modelled upon a retroviral element identified in Arabidopsis thaliana (Haupt *et al*., 2001). It was identified using the same bioinformatics techniques used to identify Kpr, Eip and PLSK and was synthesised commercially at Severn Biotech, Kidderminster. No bleed was generated to the Negcont1 peptide. With the exception of OA, patients from the inflammatory disease cohorts, such as RA and IBD showed a more discrete clustering about their mean OD values, whereas those patient groups with higher means were reflected in the larger degree of heterogeneity within the group's results (Figure 4.28). Overall, both NHD and SLE patients showed wide variation in the antibody responses generated to the control epitope. This may be partially explained by the fact that SLE is a disease whose patients have been frequently been documented to possess large numbers of autoantibodies, although such varying results in healthy patients warrant further investigation. All other disease groups showed distinct patterns of reactivity with lower means in comparison to other peptides. OA patients were the exception to this, by exhibiting a significantly increased level of antibodies reactive to this peptide. Additionally, the apparent reactivity to Negcont1 may be an indication of patient bleed non-specificity as overall levels of reactivity were comparable to those observed with non-biotinylated HERV peptides. It would be of interest to observe those levels of antibodies that bind to a biotinylated form of negative control peptide. On the evidence presented here, this may suggest these act as *bona fide* epitopes involved in OA disease pathogenesis. This however, requires further investigation as to what link exists between this retroviral epitope derived from a plant retroelement and OA disease pathology.

The investigation also examined the impact that antigen presentation had upon final patient reactivities. Kpr and Eip peptides were thus biotinylated and their performance compared to their non-biotinylated counterparts. During the course of this investigation it became clear that levels of reactivity to biotinylated antigen differed significantly to that of non-biotinylated antigen, with OD values differing by almost a factor of ten ($p < 0.0001$)(Table 4.3). Samples exhibited significantly higher levels of reactivity to the biotinylated peptides in comparison to their non-biotinylated

counterparts across all cohorts (Figure 4.24). Such an effect was not unexpected however significant increases in patient responses were observed to both peptides in an effect that was disease specific (Figure 4.39).

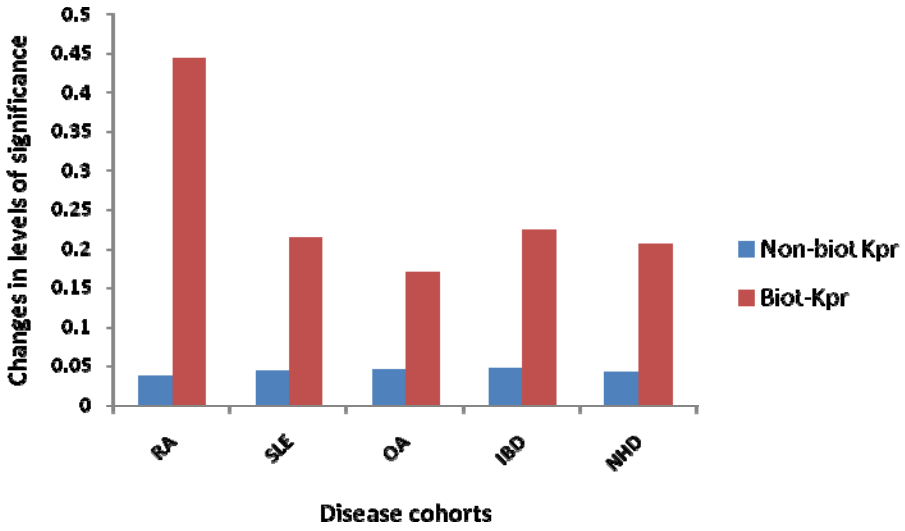


Figure 4.39 Changes in significance in results upon biotinylation of Kpr peptide. Bars represent changes in difference in levels of significance between non-biotinylated and biotinylated forms with patients from different disease cohorts. Analysis was carried out using SPSS software and one-way ANOVA analysis, with bonferroni post test (Section 4.2.13). RA shows the largest increase in significance compared to other disease groups due to biotinylation.

Biotinylated Eip (Figure 4.21) showed a similar pattern of reactivity with RA patients OD values significantly increased above those of all other disease controls. Once again there was an effect of clustering between the samples with one group accumulating below the mean, with the other present above the mean. RA patients showed an increased level of reactivity to both biotinylated peptides, particularly in comparison to disease controls. This was also accompanied by an increase in significance in biotinylated peptides. An effect that was most evident with biotinylated Kpr. Such results may hint at a potential for the epitope to play a role in RA immunopathogenesis.

From the results presented here and in previously published investigations (Shamsuddin and Harris, 1983, Ivanov *et al*., 1992) it was clear that non-biotinylation of peptides resulted in significant modifications in the structural and conformational modelling of the antigens. This low coating efficiency of peptides resulted in alteration of the availability of epitopes on plastic coated peptides leading to a potential masking of those epitopes originally deemed important. Additionally, synthetic peptide

antigens, particularly those of an overly hydrophilic nature containing less than 20 amino acids are notoriously difficult to immobilise upon plates for sufficiently sensitive assays (Fischer and Howden, 1990). It is this process of adsorption, i.e. the hydrophobic interactions between the antigen and the solid phase, which are key to the success of an assay (Kochwa *et al*., 1967, Cash and McGill, 1969). Using the so-called 'region or window of independence', concentrations of antigen within this 'window' exhibited a constant relationship with protein adsorption. It was previously reported that within this region, adsorbed antigen remained stably bound throughout the assay although this window varied with both antigen involved and other external factors such as adsorption temperature. Desorption, however, was also deemed a problem and has been shown to affect the immunogenicity of the antigen, with one investigation showing that surface adsorbed ovalbumin retained only 5% of its antigenic activity when returned to free-solution (Dierks *et al*., 1986). Additionally high concentrations of antigen also led to increased rates of desorption, and thus at all concentrations there exists a dynamic relationship between both adsorption and desorption. Thus it seems likely that the convenient and popular technique of surface adsorption may also appear as a major source of error within the immunoassay.

Several methodologies have been employed to solve this problem including coupling to larger proteins (Oshima and Atassi, 1989), using amino acids as linkers between the peptides and the plates (Loomans *et al*., 1998) and the use of multiple peptide systems (Tam and Zavala, 1989). The most successful of these methods however has been the indirect coating of biotinylated peptides to streptavidin coated plates, which resulted in improved sensitivity (Ivanov *et al*., 1992). For this reason biotinylated peptides were also included within this investigation, in conjunction with strepatavidin-coated plates in order to ascertain the effect of peptide presentation and antigen adsorption and their effects upon epitope presentation (Figure 4.22 – 4.23). According to evidence presented here, biotinylated peptides were more reliable models of *in vivo* epitope conformations, as potential epitopes would not be partially or wholly obstructed by the binding of the peptide to the solid phase (Chames *et al*., 2002).

4.6.3 Anti-HERV-K in SF

The hypothesis behind determination of the prevalence of antibodies to HERV epitopes in the synovial fluid suggested that these epitopes may be driving an B cell

activity within the joints in RA patients (Firestein, 2003). If levels of antigen within the synovial fluid were increased in comparison to those observed within the blood in the same patient, this may be indicative that antibodies producing B cells were present and being stimulated within the joint. The result suggested that there was no significant difference between levels of anti-HERV antibody observed within the blood compared to that within the synovial fluid (Table 4.3). When tested with the non-biotinylated peptides only a minority of patients showed increased levels of anti-HERV-K in the blood and not in their paired SF (Figure 4.35 – 4.36). One exception to this was patient 13 when biotinylated peptides were used (Figure 4.36) who exhibited no significant difference between the two. Interestingly, when reactivity to the non-biotinylated peptide in the SF, was compared to its biotinylated version, the difference between the two was also not significant (although this was borderline in the case of Kpr) (Table 4.5). Biotinylation did result in a more heterogeneous response with specific patients showing increases in anti-HERV antibodies, e.g. patients 2, 4 and 10 (Figure 4.36). Despite both Kpr and Eip peptides exhibiting similar patterns of reactivity in both blood and SF, the absorbances shown in the synovial fluid were slightly elevated over that observed in the blood (Figure 4.35 – 4.36), although overall all OD values were at a low level. The observed increase in levels of antibodies within the SF may be indicative of increased HERV activity within the joint and indicate an antigen-driven reaction within the synovium and synovial fluid, however as shown here those levels shown here were both weak in strength (Table 4.3) and not significantly increased over those observed in the peripheral blood (Table 4.5). Evidence presented here does not support a significant increase in anti-HERV-K antibodies at the site of disease; however despite this marginal increases within the joint have been detected and may represent disease markers within patients.

4.6.4 Levels of RF in serum

One important potential source of non-specific binding in rheumatoid arthritis patients that needed to be taken into account before any conclusions could be made was the presence of Rheumatoid Factor, i.e. antibody specific for the Fc region of IgG. In order to investigate any potential skewing of reactivities by the non-specific binding of RF in the ELISA work, a commercial quantitative latex bead agglutination test was used to accurately determine levels of RF in both patient synovial fluid and blood.

Figure 4.30 shows the relationship between levels of rheumatoid factor present within patient blood for 51 RA patients tested. Overall, the incidence of rheumatoid factor in the blood did appear to correlate with increased OD values in reactivities to peptides (Figure 4.37).

4.7 Conclusions

Six novel indirect ELISA assays incorporating short peptides predicted in Chapter 3 were developed and used to investigate the prevalence of anti-retroviral antibodies in RA patient serum and disease controls.

From this data several conclusions could be drawn:

1. A novel and robust assay has been developed for the detection of anti-HERV antibodies within patient sera. Inhibition studies have shown this assay to be accurate in the determination of anti-HERV antibodies within polyclonal patient sera.

2. Overall reactivities to Eip and Kpr were both increased over other non-biotinylated and biotinylated peptides in the blood ($p < 0.001$).

3. Antiretroviral antibodies were at marginally increased levels within the synovial fluid compared to that in the blood. These increases were not significant.

4. The presence of RF appeared to correlate with higher OD values in the blood in RA patients potentially suggesting these inflated values were potentially contributed to by the presence of RF. This effect was less evident within the synovial fluid, which predominantly showed higher levels of RF than that observed in the blood. RF may play a role in increasing OD values through non-specific binding although such a role would be limited due to antibody: peptide specificity shown in inhibitory studies.

5. When patient sera were tested using biotinylated peptides there was a significant difference between the RA patients and the control populations ($p < 0.001$).

Chapter 5

Abbreviations

18s ribosomal RNA (18s rRNA)
Avian myeloblastosis virus (AMV)
Caprine arthritis encephalitis virus (CAEV)
Coefficient of variation (CV)
Comparative C_T method ($2^{-\Delta\Delta C_T}$)
Comparative threshold (dC_T)
Complementary deoxyribonucleic acid (cDNA)
Deoxy-nucleotides (dNTPs)
Deoxyribonucleic acid (DNA)
Difference between two C_T values (ΔC_T)
Epstein Barr virus negative B-lymphoma (BJAB)
Epstein Barr virus nuclear antigen 1 (EBNA1)
Epstein Barr virus (EBV)
Fibroblast-like synoviocytes (SF)
Fluorescent resonance energy transfer (FRET)
Gene of interest (GOI)
Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)
Housekeeping gene (HKG)
Human endogenous retroviruses (HERVs)
Human endothelial kidney (HEK-293)
Human immunodeficiency virus (HIV)
Human leukocyte antigen (HLA)
Human T cell leukaemia virus-1 (HTLV-1)
Hypoxanthine phosphoribosyltransferase (HPRT)
Late membrane protein 1 (LMP1)
Local regional ethics committees (LRECs)
Luria-Bertani broth (LB broth)
Maedi visna virus (MVV)
Magnesium ions (Mg^{2+})
Messenger riboucleic acid (mRNA)
Normal healthy donors (NHD)
Nuclease-free water (dH_20)
Osteoarthritis (OA)
Peripheral blood mononuclear cells (PBMCs)
Quantitative real-time PCR (QPCR)
Reverse transcriptase (RT)
Reverse transcription polymerase chain reaction (RT-PCR)
Rheumatoid arthritis (RA)
Sjogrens syndrome (SS)
Synovial fluid (SF)
Systemic lupus erythematosus (SLE)
Terrific Broth® super mix (TB)
Threshold cycle (C_T)

Chapter 5:

Development and optimisation of a novel real-time PCR assay for the detection and quantitation of HERV-K *gag* mRNA in RA patient samples

5.1 Introduction

5.1.1 Introduction to real-time Polymerase chain reaction (PCR)

Reverse transcription polymerase chain reaction (RT-PCR) remains the most sensitive technique for detection of mRNA targets and its application within the real-time setting has become the most popular method of quantifying steady state mRNA levels (Bustin and Nolan, 2004). Employing Quantitative real-time PCR (QPCR) offers a novel aspect to this end-point assay, allowing measurement of products generated during each cycle of the reaction, which are directly proportional to the amount of template prior to the start of the PCR process. This enables the determination of starting quantities of DNA or cDNA template within the sample, generated from mRNA. The real-time system is based on the detection and quantitation of a fluorescent reporter (Lee *et al*., 1993) and works by monitoring the fluorescence emitted during the reaction, using this as an indicator of amplicon production during each PCR cycle, (in real-time), as opposed to end point detection (Bustin and Nolan, 2004). In comparison to conventional RT-PCR, QPCR offers a far superior alternative in terms of dynamic range of up to 10^7 fold (compared to 10^3 fold) and for these reasons it has developed diverse applications in a number of fields including functional genomics studies, molecular medicine, forensics, virology, microbiology and biotechnology (http://www.gene-quantification.info/).

Quantitative RT-PCR can be carried out as either a one or two-step reaction. In a one-step reaction both cDNA synthesis and PCR amplification occurs in the same tube in consecutive reactions. Contrastingly, in a two-step reaction the stages take place in different tubes. One-step reactions have been reported to have an increased sensitivity and ability to compare multiple Genes of interest (GOI) (Ginzinger, 2002), however sample cDNA cannot be used for any further downstream applications.

5.1.2 Non-specific or specific detection strategies

There are two types of homogenous fluorescent reporting chemistries presently in use: 'non-specific' and 'specific' detection. Template-specific chemistries require the design and synthesis of one or more custom-made fluorescent probes for each assay (Tyagi and Kramer, 1996). Commonly, these include reporting systems employing Fluorescent resonance energy transfer (FRET) or similar in interactions between donor and quencher molecules as the basis of detection (Gendi *et al*., 1996, Heid *et al*., 1996). The main advantage is the additional level of specificity introduced to the assay, thereby eliminating the requirement for specificity & product confirmation through post PCR processing (Ririe *et al*., 1997).

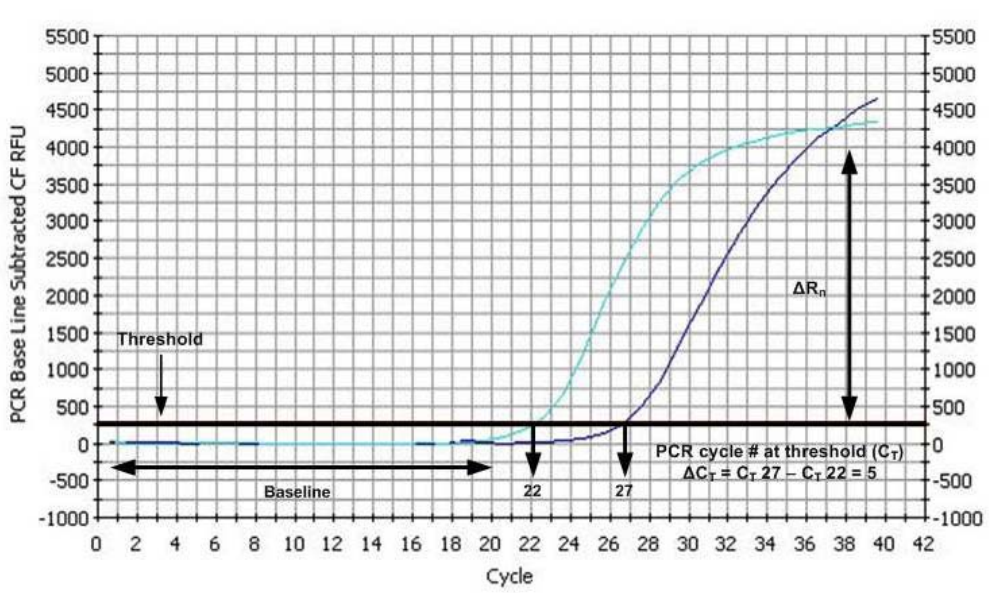


Figure 5.1 Amplification plot illustrating the main nomenclature used in real-time Q-PCR experiments. In this plot a solid line crosses the threshold at the PCR cycle number18, whereas the dotted line crosses at 20. By subtracting the two (20-18) there is a two cycle difference between the two samples or a ΔCt of 2. Due to the exponential nature of PCR, the ΔCt is converted to a linear form. This number indicates the fold-difference between the two samples reflecting variation in starting template abundance. ΔR_n : Increase of fluoresenct signal over threshold per run.

Non-specific detection uses intercalating dyes such as SYBR Green (Yin *et al*., 2001, Varga and James, 2006) that bind to any double-stranded DNA generated during the PCR reaction, emitting enhanced fluorescence (Ishiguro *et al*., 1995). Although intrinsically non-specific in nature, these chemistries could yield quasi-template specific data if used in conjunction with melt curves to aid in the identification of specific amplification products (Ririe *et al*., 1997). One recent investigation even

reported SYBR Green I detection being of higher precision than Taqman (and other 'specific') chemistries (Livak and Schmittgen, 2001). Non-specific assays are advantageous in that they could be used with previously established protocols and primer sets, in addition to being significantly cheaper than the specific detection alternatives (Bustin and Nolan, 2004).

The Threshold cycle (CT) is a parameter most conveniently and most frequently quoted when reporting QPCR results. CT values are commonly used when describing the data quantitatively and form the heart of the assay itself. They are defined as the fractional PCR cycle number at which the fluorescent signal becomes greater than the minimum detection level (baseline or threshold) (Figure 5.1).

5.1.3 Quantitation strategies

QPCR data can be analysed differently depending upon experimental design and the investigative goals. The two methods of analysis available are 'relative quantitation' and 'absolute quantitation' with each having profound impacts upon how each assay is designed and run. 'Relative' is the analytic method of choice for the majority of published assays and involve a comparison made within a sample between its GOI and a control or Housekeeping gene (HKG) whose expression levels should be unchanged by experimental conditions (Bustin and Nolan, 2004). Absolute quantification relies upon the use of a standard curve derived from known quantities of GOI, which is then used to calculate unknowns. The advantage with using absolute quantification comes from its ability to determine the input copy number of transcripts of interest (Gibellini *et al*., 2004). cDNA plasmids containing GOI are common sources of controls in such investigations in which known starting amounts of sample are serial diluted to construct standard curves. This method requires fewer optimisation steps, although is more prone to operator error and sample variation (Bustin and Nolan, 2004) and reported to be less sensitive (Battaglia *et al*., 1998). Relative quantitation of gene expression can be determined using several alternative methods based upon comparison of reference and experimental genes.

5.1.3.1 Relative standard method (relative fold change)

Relative standard quantitation involves one of the experimental samples (1x sample concentration) acting as the calibrator. Experimental values were divided by the calibrator's normalised value in order to give a value of relative expression. As the calibrator is the 1x sample, all other quantities are expressed as the *x*-fold difference relative to the calibrator.

5.1.3.2 Comparative threshold (dC_T)

Through lack of any predetermined standard, this method directly compares relative amount of target sequence to any of the reference values chosen. The result is given as relative to the reference value, i.e. the baseline expression in resting cells, etc. Using this methodology the main criterion for accuracy is that in order for the C_T comparison to hold, both target and reference amplification efficiencies should be equal. This usually requires construction of a standard curve using dilutions of target GOI (Livak and Schmittgen, 2001).

5.1.3.3 Comparative C_T method ($2^{-\Delta\Delta C_T}$)

This method enables the relative quantitation of template without the requirement for a standard curve by calculating relative to an endogenous control (normalizer). Again, for maximum accuracy, both target and reference must have equal amplification efficiencies. It is usual for the reference gene to have a slightly higher expression level to the GOI (and therefore a lower C_T,) and it is the difference between them (ΔC_T) that underpins this methodology (Livak and Schmittgen, 2001). The comparative $\Delta\Delta C_T$ method without correction for PCR efficiency was used for relative quantitation of the GOI against the HKG (Winer *et al*., 1999, Livak and Schmittgen, 2001) thus eliminating the requirement for a standard curve for each sample.

5.1.4. Normalisation strategies

The identification of a valid reference for data normalisation is central to the type of analysis employed in relative quantitation, but also remains one of the more controversial problems concerning real-time analysis (Bustin and Nolan, 2004). Several different strategies have been applied to the normalization of the expression of

RNA of interest (Huggett *et al*., 2005). The most appropriate strategy is to use a reference gene as an internal standard that is assumed to remain constant between experimental groups. It now appears, however, that in most experimental situations the use of conventional reference genes such as Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is inappropriate due to their variability in transcriptional levels (Bustin and Nolan, 2004). Requirements for good HKG's include those that are stably expressed amongst different cell types and are unaffected by experimental conditions (Dheda *et al*., 2005). Table 5.1 shows five candidate housekeeping genes considered for use within the assay framework outlined within this work. Of these, Hypoxanthine phosphoribosyltransferase 1 (*HPRT*) and Beta-Actin (ACTB) were selected for further use on the basis of assay suitability. *GAPDH* could not be used as levels were altered in the presence of Fetal calf serum used in cell culture (Ke *et al*., 2000). 18s ribosomal RNA (18s rRNA) could not be used due to use of oligo-dTs as primers for reverse transcription and the primers for Histone 3.3 (Andersson *et al*., 1996) were not used further as these produced a second peak (indicating a second amplicon) on the melt curves. More recently a number of kits to assess HKGs suitability under experimental conditions have become available commercially, e.g. GeNorm (Primerdesign, Southhampton, UK). These kits assess panels of HKG under the experimental conditions of the investigation in hand as described previously (Vandesompele *et al*., 2002).

For real-time analysis and quantitation, the geometric mean C_T values of two HKGs were combined to increase the accuracy of any resultant quantitations (Vandesompele *et al*., 2002, Bengtsson *et al*., 2005). Hypoxanthine-guanine phosphoribothymidine (HPRT) is an enzyme associated with purine metabolism that functions to salvage purines from degraded DNA in order to renew their synthesis. In this role, it acts as a catalyst in the reaction between guanine and phosphoribosyl pyrophosphate (PRPP) to form GMP. The Beta-actin has been reported to have a role in mRNA localisation and translation, RNA polymerase II transcription (Hu *et al*., 2004) and cell motility (Kislauskis *et al*., 1997).

Relative quantification determines changes in steady state mRNA levels of a gene across multiple samples and expresses it relative to levels of an internal control RNA, this reference often being a housekeeping gene (Pfaffl *et al*., 2002, Wittwer *et al*., 2001). Additionally relative quantification does not require standards with known concentrations and can be any transcript as long as the sequence is known (Bustin,

2002). Accurate quantitation is achieved by comparing GOI vs. reference gene. In the majority of cases this is adequate for investigation of physiological changes.

Housekeeping gene	Primer sequence	Expected product Size	Reference
Hypoxanthine phosphoribosyltransferase 1 (*HPRT*)	For – AAGCTTGCTGGTGAAAGG Rev – AAACATGATTCAAATCCCTGA	134bp	(Ramos-Nino *et al*., 2003)
Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*)	For – CCTGTTCGACAGTCAGCCG Rev – CGACCAAATCCGTTGACTCC	101bp	(Sonveaux *et al*., 2003)
Histone 3.3	For – CTCTACTGGAGGGGTGAAGAA Rev – TGCCTCCTGCAAAGCACCGAT	200bp	(Andersson *et al*., 1996)
Beta-Actin	For – GGCCAACCGCGAGAAGAT Rev – CGTCACCGGAGTCCATCAC	134bp	Wang *et al*., 2002)

Table 5.1 Details of different housekeeping genes considered for use within the investigation. All four were tested three times in triplicate on Tera-1 cellular cDNA (*n*=3). Their suitability was also taken into account in terms of variability under experimental conditions and in disease state.

5.1.5. Molecular investigations of HERVs in RA

As described in Chapter 1, a number of investigations have identified increased levels of HERV genetic material within RA samples compared to disease and healthy controls although, increased HERV transcriptional activity has also been reported in other rheumatic diseases Systemic Lupus Erythematosus (SLE) and Sjogrens Syndrome (SS). The HERV-K10 *gag1* gene has been identified previously to be expressed as encoded protein in certain cancer cell lines (Sauter *et al*., 1995). A number of other studies have focussed upon the expression of *env* which in some HERVs, i.e. HERV-W, appears to be the most intact gene (Antony *et al*., 2006).

In addition to evidence drawn from parallels between human and animal retroviral infections such as Caprine arthritis encephalitis virus (CAEV) and Maedi visna virus (MVV) (Michaels *et al*., 1991), many early investigations focussed upon serological associations between retroviruses and RA. The majority of RA patients were seropositive for Human immunodeficiency virus (HIV) and Human T cell lymphotropic virus-I (HTLV-I) reactive antibodies, despite lacking any evidence of a causative virus (Douvas *et al*., 1996, Nelson *et al*., 1999, Seemayer *et al*., 2002). More recent studies have demonstrated evidence for HERVs at the site of disease, using

molecular techniques with superior specificity and sensitivity such as PCR and *in-situ* hybridisation. Nakagawa *et al*. (1997) showed increases in mRNA levels for HERV-L, ERV-9 and HERV-K in RA patients versus controls. Increases in HERV-K10 mRNA in Peripheral blood mononuclear cells (PBMCs) of RA patients were reported by Nelson *et al* (1999), whose results were also confirmed in a follow-up study (Davari Ejtehadi *et al*., 2005). Other HERV families, for example ERV3, Lambda 4-1, HRV-5 and L1-transposons have also been implicated in RA (Takeuchi *et al*., 1995, Griffiths *et al*., 1999, Brand *et al*., 1999, Neidhart *et al*., 2000). A number of studies however have published contradictory evidence questioning such a role for HERVs (Norval *et al*., 1979, Stoye, 1999, Gaudin *et al*., 2000). Additionally HERV-K (HML-2) Rec protein has recently been shown to be expressed in both normal and rheumatoid synovial cells (Ehlhardt *et al*., 2006).

Interest also surrounded findings that a HERV-K10 LTR (LTR3) was reported within a Human leukocyte antigen (HLA) class II region (Kambhu *et al*., 1990). Almost a decade later, this was identified as a potential disease marker for RA, with its presence significantly increased in patients compared to healthy controls (Seidl *et al*., 1999, Pascual *et al*., 2001, Pascual *et al*., 2003). So far it has become apparent that the relationship between HERVs and RA is complex, and in need of clarification.

5.1.6. Comparison of Reaction efficiencies

In this investigation a real time quantitative PCR assay using SYBR-green detection chemistries was developed to measure levels of HERV-K *gag* expression in mononuclear cells extracted from blood and synovial fluid taken from RA, OA and healthy patient cohorts. ACTB and HPRT expression were used to normalise and calculate relative expression of *gag*. The resultant assay showed greater speed and potential throughput, than conventional PCR whilst being more economical than Taqman alternatives. The Comparative C_T method was used in the assay, as described above, to determine relative quantitation between the GOI (HERV-K *gag*) and the HKG used (ACTB and HPRT). The software 'LinRegPCR' (Ramakers *et al*., 2003) was used to determine the PCR efficiency of all reactions from the cycle fluorescence data generated during each reaction, thereby negating the requirement for a standard curve for each experimental run. Using this methodology it was possible to eliminate all reactions with efficiencies outside the acceptable margin of amplification

efficiencies, i.e. between '1.7 - 2.1' (Ramakers *et al*., 2003), thus ensuring only those samples with comparable efficiencies were used in quantitation. The two-step QPCR approach was preferred, i.e. different tubes for both cDNA synthesis and PCR amplification as this allowed sample cDNA to be used with other downstream applications as opposed to the one off measurement offered by the one step assay. This was also an important choice as the quantity of high quality RNA isolated from patient samples was a major limiting criterion.

During this investigation a novel real time RT-qPCR assay was developed for the accurate quantitation of HERV-K (HML-2) *gag*1 mRNA in RA patient samples compared to healthy and disease controls. This assay was also used to measure levels of HERV activity after treatment of HFLS with pro-inflammatory cytokines that have been reported previously to have significant influences over RA disease pathogenesis, including increased maturation of osteoclasts and fibroblast activation. Additionally, the influence of other viral pathogens over HERV transcriptional activity was explored. The Herpes viruses are one group of viruses that have been associated repeatedly within the literature with modulation of HERV activity (Ruprecht *et al*., 2006). EBV is one viral pathogen which has been associated within the literature with RA disease pathogenesis. EBV genes EBNA1, shared sequence homology with a susceptibility marker for developing RA, in (HLA)-DR1 (Fawcett *et al.,,* 1988) whilst elevated levels of EBNA1 transcripts have been detected within the synovium of RA patients (Saal *et al.,,* 1999) **.** Increased expression of LMP1 was also reported in the cells located at the apex of the villi of the synovial membrane (Takei *et al.,* 1997). LMP-2A was selected as its presence has been reported to transactivate the HERV-K18 SAg (Sutkowski *et al*., 2004).

5.1.7. Aims

This chapter initially set out to describe the design and experimental optimisation of a real-time PCR assay specific for HERV-K *gag*. Once developed this assay was trialed by comparing the levels of HERV-K *gag* transcription in several different cell lines and by investigating the effects of hypomethylation agents upon HERV-K *gag* transcription in testicular carcinoma cells. Application of the real-time assay to the investigation of a potential role of HERV-K in RA immunopathology on a

molecular level was carried out. Such subsequent investigations followed three specific predictions.

- Levels of HERV-K *gag* transcription would be increased in RA patient samples over controls.
- Cell populations responsible for contributions to disease pathogenesis would show the greatest increases in HERV-K *gag* transcription.
- External environmental factors, i.e. other viral pathogens or inflammatory cytokines, play modulatory roles in influencing HERV-K activity in RA patients.

In order to follow these novel lines of investigation in clarifying the exact role of previously reported increases in HERV-K *gag* in RA patients, the highly specific and sensitive real-time PCR assay was used:

1. To quantify levels of HERV-K *gag* mRNA in patient samples (relative to HKG) in comparison to disease controls.
2. To identify cell populations in patient samples in which HERV-K (HML-2) *gag1* activity was increased.
3. To explore the modulatory effect of different viral proteins upon HERV-K (HML-2) *gag1* transcription.
4. To explore the modulatory effects of inflammatory cytokines upon HERV-K (HML-2) *gag1* mRNA levels.

5.2 Materials & Methods

All reagents were obtained from (Sigma-Aldrich, Gillingham, UK) unless specified otherwise

5.2.1 Collection of human blood and synovial fluid

The average age for RA patients was 65 years old *(n* = 29) (range 40–93 years) with a mean duration of disease of 13.5 years (ranging between 4–46 years). Osteoarthritis (OA) patients had an average age of 63 years old (*n* = 12) (range 38-89) with a mean disease duration of 10.5 years (varying between 6–17 years). Details for inflammatory controls were unavailable. Venous blood was collected from were obtained from Normal healthy donors (NHD) comprising departmental staff and normal blood donors *(n* = 6). The average age for healthy donors was 33 years old (range 21-40). Patients with non-RA inflammatory diseases such as gout, oligoarthritis and polyarthritis (*n* = 7) were also included within the analysis. All blood samples consisted of 10mls venous blood and were collected in sterile preservative-free heparin and taken from patients with informed consent. Blood samples were processed within 3 hours. 10 mls of synovial fluid (SF) were collected in empty sterile 20 ml universals with all steps performed under aseptic conditions. All ethical approvals were considered by appropriate Local regional ethics committees (LRECs) and recruitment criteria and methods have been reported in detail elsewhere (Section 2.1) (Davari Ejtehadi *et al*., 2005). Fibroblast-like synoviocyte (SF) primary cells were purchased commercially (Cell applications, San Diego, CA) and were derived from RA, OA and healthy (NHD) donors. All donors were Caucasian, 42, 66 and 58 years of age respectively. OA and NHD were derived from female donors whereas the RA SFs were derived from a male.

5.2.2 Sample processing

Mononuclear cells were separated from 10ml whole blood using a Ficoll gradient. Anti-coagulated venous blood was layered onto 7ml HISTOPAQUE-1077 in a 20ml sterile universal. This was then centrifuged at 400 *x g* for 30 min at room temperature. During centrifugation, erythrocytes and granulocytes were aggregated by polysucrose and rapidly sedimented at the base of the tube whilst lymphocytes and mononuclear cells remained at the plasma-HISTOPAQUE®-1077 interface (Figure

2.1). The upper layer (plasma) ending approximately '0.5cm', above the central opaque interface (containing mononuclear cells) was aspirated and transferred to a sterile universal using a Pasteur pipette. Plasma was stored at -70ºC until required in serological investigations. The opaque interface was also transferred to a sterile centrifuge tube. These cells were resuspended in 1 ml of TRI reagent for RNA extraction (Chomczynski and Sacchi, 1987). Total RNA from patient samples was extracted as described previously by (Davari Ejtehadi *et al*., 2005).

5.2.3 Separation of specific cell populations using paramagnetic beads

For a small number of new samples, different species of lymphocytes were separated from one another to ascertain levels of HERV-K (HML-2) in different cell populations. This was done using a series of washes with paramagnetic beads coated with antibody specific for CD19 (B-cells), CD4 and CD8 (T-cells). The Ficoll interface containing mononuclear cells was separated into three equal aliquots. Each aliquot was then mixed with dynabeads specific for the cell population being extracted i.e. B-cell (CD-19). Dynabeads (25ul/ml^{-1} sample) were added to the interphase and incubated at 4ºC under agitation before incubating at room temperature for 2 min with the MPC Dynal magnet (Dynal technologies, UK). After separation the supernatant was discarded and cells/ dynabeads washed 5 times with PBS/ 0.1% BSA. Beads and cells were separated from one another using commercially available Detachabead® kits (Dynal technologies, UK). Cells and Detachabead® reagents were mixed and incubated together at room temperature for 45 min before placing upon the MPC Dynal magnet, for a further 5 min were decanted into a sterile centrifuge tube and washed three times in RPMI/1% FCS before centrifuging for 8 min at 500 *x g*. Cell pellets were finally resuspended in appropriate volumes of RPMI/1% FCS. These cells were then lysed in RLT buffer for RNA extraction using the QIAAmp kits (Qiagen, UK).

5.2.4 Primer design

Nucleotide reference sequences for HERV-K and HERV-K10 were extracted from the NCBI database (accession numbers M14123 and Y18890). Sequences were copied and pasted from a standard text editor into three different primer design programs: Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi) (Korwin-Kossakowska *et al*., 2006); PrimerSelect (DNAstar Inc, Madison, WI) (Breslauer *et al*.,

1986) and Beacon designer version 5.0 (Premier Biosoft Int., Palo Alto, CA) (Ram and Shanker, 2005). Primers were designed using rules and guidelines reviewed previously (Dieffenbach *et al*., 1995). All primers were subjected to BLAST analysis using the BLASTn database in order to confirm their specificity, and to prevent any cross-reactivity from occurring (Altschul *et al*., 1997). Internal hybridisation probes were designed for all primer sets so that later conversion to specific detection chemistries such as Taqman required minimal optimisation and disruption, if required.

5.2.5 Two - step quantitative real-time RT-PCR

cDNA template was synthesised using 1 µg of total RNA and converted into cDNA using oligo(dT) and Avian Myeloblastosis Virus (AMV) RT enzyme (Promega, Southampton, UK). RT cycles consisted of an initial step of 5 mins at 72°C, followed by addition of mastermix containing 3mM $MgCl_2$ (Promega), 5 x Reverse transcription buffer [250mM Tris-HCl (pH 8.3 at 25ºC), 250mM KCl, 50mM $MgCl_2$, 2.5mM spermidine and 50mM DTT], deoxy-nucleotides (200 µM), oligo (dT) (0.5 µg/µg RNA) and sterile nuclease-free water (dH_20). The reaction mix was incubated at 42 °C for 1 hour before heating to 95°C for 5 mins. For testing in the real-time assay, cDNA was diluted 1:5 in dH_20. Gene specific primers for the GOI and both HKGs were mixed in separate 0.5ml eppendorf tubes. These included 9 µl of sterile nuclease-free dH_20, 0.5 µl HERV-K(HML-2) *gag1* set1 primers (nM), 2.5 µl of template (diluted 1:5 in dH_20) and 12.5 µl Absolute™ QPCR SYBR green master mix (Abgene, Epsom, UK) (Table 5.2). Once mixed, reagents were transferred to a 96-well plastic plate (Abgene, Epsom, UK) and capped with plastic strips (Abgene, Epsom, UK) before being sealed with plate sealers (Abgene, Epsom, UK) in order to prevent evaporation and run using an iCycler real-time thermocycler (Bio-Rad, Hemel Hempstead, UK) using final optimised cycling parameters listed in Table 5.3. All optimisation steps and subsequent assay runs were performed using Absolute™ QPCR SYBR green mix (Abgene, Epsom, UK). This mix had been reported previously within the literature with successful results (Govindarajan *et al*., 2003, Dittmer *et al*., 2006).

Reagent	Volume (µl)	Final concentration
AMV Reverse Transcription 5x reaction buffer	**5**	**1x**
dNTP	**2.5**	**1 mM each dNTP**
Oligo(dT)	**0.5**	**0.5 µg/µg RNA**
RNasin ribonuclear inhibitor	**0.5**	**40 U**
Template RNA	**1-10**	**1 µg**
AMV RT	**0.75**	**30 U**
Nuclease-Free water	**Make up to total of 20 µl**	**-**
Total volume	20 µl	
ABsolute™ QPCR SYBR green mix (2x)	12.5	1x
Forward primer	0.5	0.5 µM
Reverse primer	0.5	0.5 µM
cDNA Template (diluted 1:5)	2.5	< 250ng
dH_20	9	-
Total volume	25 µl	

Table 5.2 Reagent components of the real-time PCR mix making up the novel HERV-K *gag* real-time PCR assay.

Stage	Temp (°C)	Time (mins)	No. of cycles
Initial denaturation	**94**	**15**	**1**
Denaturation	**94**	**15s**	
Annealing	**55.9**	**30s**	**40**
Extension	**72**	**30s**	
Melt curve analysis			
Denaturation	**94**	**30s**	**1**
Start	**60**	**30s**	**1**
Melting step	**60 (+0.5 each cycle)**	**10s**	**80**

Table 5.3 Details of the real-time thermocycling parameters optimised for the *gag*1set1 primer set and both housekeeping genes. S: seconds.

5.2.6 Intra and Inter-assay validation

In order to assess the accuracy and reproducibility of the assay, total RNA was extracted from Tera-1 cells using an RNeasy RNA extraction kit (Qiagen, Crawley, UK). Total RNA was eluted from spin columns in 20 µl volumes of EB buffer and quantified using a spectrophotometer. Total cellular RNA was then applied to a ten-fold serial dilution, and was used to synthesise cDNA using protocols described previously (Section 5.2.5). After first strand synthesis, cDNA was diluted 1:5 in sterile nuclease-free dH_2O and added to reagents (Table 5.2) before being amplified under conditions described previously (Table 5.3). After amplification, 10 µls of reaction mix was run and the 120bp amplification product of *gag* was visualised in the presence of ethidium bromide in an agarose gel under long wave ultra-violet light. Bands were

excised and purified using a MiniElute DNA purification kit (Qiagen, Crawley, UK) and eluted purified amplicons were then quantified using an spectrophotometer (Geneflow, Fradley, UK) before identity was confirmed by sequencing (Section 5.2.8). Intra-assay variation was assessed using five sets of triplicates per plate/run. Inter plate variation was assessed with one sample at 3 specific dilutions, run in triplicate and repeated on five different plates. One master mix for the RT step was made up and divided amongst the seven samples in order to minimise variation. This was repeated three times with each sample, run in triplicate.

5.2.7 Cloning of HERV-K *gag* amplicons into vectors for sequencing

In order to both confirm specificity of the HERV-K10 gag primers, and to determine any disease specific variants, 12 PCR products synthesised from patient samples were cloned into the pGEM-T plasmid (Promega, Southampton, UK). Initially PCR products were run in an agarose gel before being excised and purified using the QIAspin gel purification kit (Qiagen, Crawley, UK). Purified PCR products were quantified in a gel against a low mass DNA ladder (Promega, Southmpton, UK), before being ligated into the plasmid for preparation of DNA and sequencing. For ligating the amplicon into the vector two opposing ratios were set up of 3:1 and 1:3 of template to vector respectively. Ligations were set up with 5 $ng\mu l^{-1}$ and 0.5 $ng\mu l^{-1}$ of template respectively before being added to the vector in ligation buffer according to manufacturer's instructions. Ligation was carried out at room temperature for 1 hour before ligated plasmid and inserts were transformed into 50 µl JM109 competent cells (Stratagene, Stockport, UK) using the heat-shock method. Transformed cells were then incubated for at least 16 hours on LB-agar with ampicillin/IPTG/X-Gal for blue-white screening. White colonies contained plasmids with inserts. Negative controls and uncut pGEM plasmid (0.1 $ngml^{-1}$) were also included for the purpose of calculating transformation efficiency.

5.2.8 PCR product clean up and sequencing of PCR amplicons

In order to confirm the identity of amplified PCR products, and so the specificity of the real-time PCR for HERV-K10 *gag1* reverse transcribed mRNA amplified products were run in a 2% agarose gel. PCR bands were cut from the gel under long wave UV light using a scalpel and then purified using the Minielute PCR

purification kit according to the kit instructions (Qiagen, Crawley, UK). DNA sequencing was performed by the Sequencing service (School of Life Sciences, University of Dundee, Scotland) using Applied Biosystems Big-Dye version 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer (Applied Biosystems, San Diego, CA). Sequencing was carried out using gene specific primers unless otherwise stated. All DNA sequence analysis was performed using Lasergene software suite (version 7.1) (DNAstar, Madison, WI).

Fluorescence data was analysed using the iCycler 'iQ' software (Bio-Rad, Hemel Hempstead, UK) and expressed as a function of the C_T. and exported directly into Excel. PCR reaction efficiencies were calculated using the LinRegPCR software (Ramakers *et al*., 2003) in order to ensure accurate quantification. Validated results showed PCR efficiencies between '1.7 and 2.1', with correlations '>0.999'. Results outside of these margins were discarded and repeated. Final ΔC_t values were calculated by determining average Ct values over repeated runs and applying these normalised values to the relative quantitation formula.

5.2.9 Restriction Digests

Restriction enzymes were used to confirm the identity of sG5-EBV expression plasmids (Section 5.3.4.1). Reactions were performed in 20 µl volumes using 1µg/µl of substrate DNA. Reactions typically consisted of 16.3 µl sterile, deionised waster, 2 µl 10X buffer, 0.2 µl Acetylated BSA (10µgµl), and 1 µl of template DNA (1 µg/µl) in a sterile 0.5 ml eppendorff. Following gentle mixing by pipetting, 0.5 µl restriction enzymes – *Bgl* II and *EcoR*1 (Promega, UK) were added to the reaction mix giving a final volume of 20 µl. Following further gentle mixing, the tubes were incubated at 37°C using a thermocycler (MJ research Inc., Ramsey, US) for four hours. After incubating, digested DNA was visualised using an electrophoretic agarose gel stained with ethidium bromide.

5.2.10 5'-Aza-2'-deoxycytidine treatments

Tera-1 cells were seeded at 5 x 10^4 cells per T75 flask and grown in RPMI medium (Sigma, UK) supplemented with 10% FCS and 20mM L-glutamine as described previously (Section 2.7). Cells were incubated until 70% - 80% confluent at

37ºC in 5% CO_2. A 100 μM solution of 5'-Aza-2'-deoxycytidine was made using dH_20, before being further diluted to 10 μM and 1 μM in supplemented media and added to culture flasks before being incubated for 72 hours at 37ºC in 5% CO_2. Untreated and water only controls were also included within the experimental design to compensate for the effects of diluted media. Each experiment was repeated on three separate occasions in order to give as accurate a result as possible. After 48 hours, cells were trypsinised from the flask surface, and RNA was extracted using RNeasy commercial kits (Qiagen, Crawley, UK). cDNA was synthesised as previously described (Section 5.2.6). All test runs included no reverse transcriptase (RT) and no template negative controls.

5.2.11 Viral protein studies

Fibroblast-like synoviocytes (Cell applications, San Diego, CA), were transfected with expression plasmids containing viral proteins from different viruses that have been associated with RA in the literature. Total cellular RNA was extracted and analysed for changes in HERV-K *gag* expression using quantitative RT-PCR and a HERV - specific retroviral microarray. Human cell lines (both primary and continuous) were transfected with two pSG-5 expression plasmids containing Epstein barr nuclear antigen 1 (EBNA1) and Late membrane protein 1 (LMP1) genes from Epstein Barr virus (EBV) that had been linked previously with the etiology of RA (kindly provided by Dr Paul Murray, University of Birmingham). After transfection of synoviocytes and controls with expression plasmids for EBV using Lipafectamine (Invitrogen, Paisley, UK) according to manufacturers instructions cells, were incubated at 37ºC with 5% CO_2. After 48 hours cells were harvested and total RNA was isolated and purified using RNeasy kits (Qiagen, Crawley, UK). Total RNA was quantified and levels of HERV-K *gag* expression were determined using Quantitative RT-PCR, using iCycler thermocycler (BioRad, UK) and with HERV-specific microarrays.

5.2.13 Transformation of viral gene expression vectors into *E.coli*

Once their identity was confirmed, pSG5-EBNA1 and pSG5-LMP1 were both transformed into JM109 competent cells (Promega) using the heat-shock method (Sambrook J and Russell, 2000). Once transformed into competent cells, plasmid stocks were grown overnight (16 – 20 hours) in Luria-Bertani broth (LB broth) and

bacterial cultures were prepared using QIAprep Minipreps kits (Qiagen, Crawley, UK). Once plasmids had been extracted from bacterial lysates, they were run in a 1% agarose gel in order to confirm identities. pSG5 – LMP1 in particular gave repeatedly low yields of plasmids (<50 ng/μl^{-1}) when grown for 16 – 20 hours in LB media, so media was changed to Terrific Broth, Modified EZmix powder. Individual preparations of plasmids were then combined respectively using ethanol precipitation before being stored at -20ºC. Bacterial clones containing each plasmid were picked from master library plates and stored at -80ºC using Protect bacterial preservers (Fischer Scientific, Leicestershire, UK).

5.2.15 Cytokine treatments

Treatments with pro-inflammatory cytokines were carried out upon human synoviocyte cell lines (RA, OA and healthy) and HEK-293 endothelial cells (used as non-synovial fibroblast control). 5 x 10^3 cells were seeded in vented T75 flasks before incubating overnight until confluent in RPMI supplemented media. Pro-inflammatory cytokines – IL-1β, IL-6 and TNF-α were diluted to required concentrations using medium (un-supplemented). Diluted cytokines were then incubated with cells at 5% $C0_2$ for 48 hours. Cell viability was determined using Celltiter 96 AQ_{ueous} one solution reagent (Promega, UK). All cell viability was in excess of 70% for all treatments. After incubation, cells were harvested and genetic material extracted using QIAspin RNA extraction kits (Qiagen). All treatments were repeated three times.

5.2.17 Statistical analysis

Geometric mean C_T values were calculated using three repeat runs of triplicates. All samples were age/sex matched where possible. Statistical analysis between matched samples involved paired, 2-tailed t-tests with the Mann Whitney rank sum test and variance between different groups was graded using the one-way ANOVA test, with Tukey-Kramer or Bonferroni test to compare specific groups. Statistics and graphical analysis were performed using Prism 5.1 (Graphpad software, San Diego, CA). Unpaired samples were compared using an unpaired, 2-tailed t-test.

5.3 Results

5.3.1 Optimisation and Development of real-time PCR assay for HERV-K gag

5.3.1.1 Primer design and selection of optimal primer pair: A comparison of 3 different programmes for primer design

Table 5.2 list all primers designed and tested for the purposes of inclusion within the novel real-time assay for HERV-K. The location of predicted primer sets and their distribution across the HERV-K sequence, using the prototype HERV-K10 as a model, shows a clustering towards the end of HERV-K10 *gag1* and the beginning of *gag2* genes (Figure 5.2). In HERV-K, such regions would likely correspond to the central capsid sequence that forms the core shell of virus particles (Campillos *et al*., 2006). Primer sets for the envelope region appeared to cluster predominantly at the 5' end of the gene within the *gp36* antigenic region (Sander *et al*., 2005). After primer pairs had been synthesised, they were tested using cDNA from Tera-1 cells, known to harbour increased levels of HERV-K with the potential to form particles.

The levels of Performance attained by each software application were comparable, with all programs predicting at least one functional primer set (Table 5.4). Primer 3 showed the least success with just 33.3% (1/3) primers being functional. Lasergene and Molecular Beacons both showed comparable levels of success (2/3 and 4/8, respectively). Of those primers tested, all resulted in the production of a single expected amplicon. All functional primer pairs tested gave smooth melt curves and signal peaks, indicative of a high level of specificity, with low occurrence of mispriming. In all cases, primer performance was also robust and reproducible; producing consistent C_T values over three runs of triplicates.

The final primer set chosen, on the basis of melt curve analysis and reproducibility for incorporation into the final assay, were 'MB HERV-K10 *gag1* set 1' (Table 5.3). Primers amplified a product of 120 bp from HERV-K *gag*, with a T_m of ~55°C. BLAST analysis of the primer pair revealed no mispriming or sequence similarities with organisms other than HERV-K. Additional analysis showed the absence of any secondary hairpin structures or dimerisation. In order to confirm primer specificity amplification products were sequenced. Figure 5.3 shows the meltcurve

data for a single PCR product of 120 bp from five amplifications, run in a 3% agarose gel (Sigma), stained and visualised under UV with Ethidium bromide (Figure 5.4).

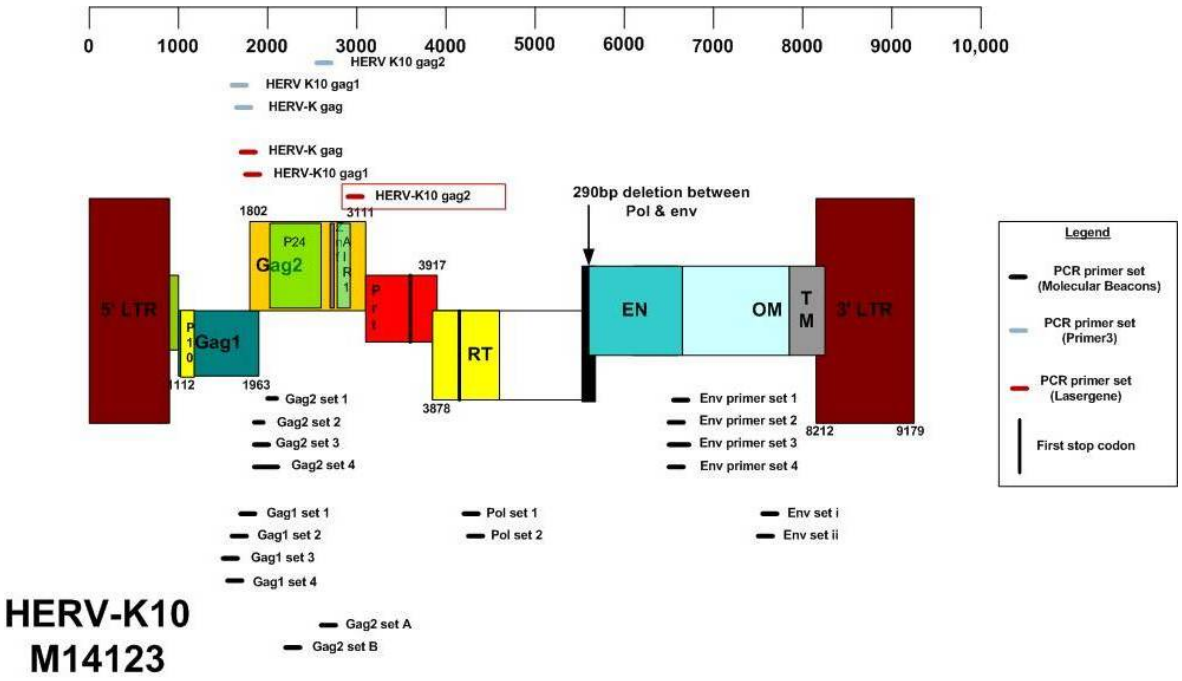


Figure 5.2 Map of the HERV-K family prototype - HERV-K10 sequence showing locations of different primer sets designed for incorporation into a real-time assay. For key see legend. Primer pair selected for use in real-time assay is boxed in red.
TM : Transmembrane, EN: Envelope; RT: Reverse transcriptase; Prt: Protease ; LTR : Long terminal repeat.

Design software	Primer name/ Target gene	Primer sequence	Expected Prod. Size (bp)	PCR Product
Primer3	HERV-K *gag* F (1773-1792)	CCA CCC CCA GAA AGT CAG TA	104	
	HERV-K *gag* R (1863-1877)	TGG TGC CAT AGC ATT GTC TC		
	HERV-K10 *gag*1 F (1618-1637)	TTA ATG GGG CCA TCA *GAG* TC	112	
	HERV-K10 *gag*1 R (1711-1730)	TGC GGT TGG GTC TTA TTT TC		
	HERV-K10 *gag*2 F (2551-2570)	ATG AAA ACG CCA ATC CTG AG	107	X
	HERV-K10 *gag*2 R (2639-2658)	CGA TTC CAT CAC AGG CTT TT		
Lasergene (PrimerSelect)	LSRGENE HERV-K *gag* F (1770-1789)	CGG CCA CCC CCA GAA AGT CA	68	X
	LSRGENE HERV-K *gag* R (1816-1838)	GGT ATG GCT CTG TCC TGT GGT G		
	LSRGENE HERV-K10 *gag*1 F (1771-1796)	GGC CAC CCC CAG AAA GTC AG	184	
	LSRGENE HERV-K10 *gag*1 R (1932-1955)	GAA TTG CCA TGC CTC AGT ATC TCC		
	LSRGENE HERV-K10 *gag*2 F (2952-2972)	CCT CAG GCC CCA CAA CAA ACT	62	X
	LSRGENE HERV-K10 *gag*2 R (2991-3014)	CTG AAA ACC CTG AGG AAC AAA TGG		
Molecular Beacons	MB HERV-K10 *gag*1 set 1 F (1623-1643)	GGG GCC ATC AGA GTC TAA ACC	120	X
	MB HERV-K10 *gag*1 set 1 R (1720-1743)	TGA TAG GCT ACT TGC GGT TGG		
	MB HERV-K10 *gag*1 set 2 F (1621-1642)	ATG GGG CCA TCA *GAG* TCT AAA C	122	
	MB HERV-K10 *gag*1 set 2 R (1722-1743)	TGA TAG GCT ACT TGC GGT TGG		
	MB HERV-K10 *gag*2 set 1 F (1623-1643)	GGG GCC ATC AGA GTC TAA ACC	76	
	MB HERV-K10 *gag*2 set 1 R (1675-1699)	TTG AGG TTG TAA TCT TAC *GAG* CAC		
	MB HERV-K10 *gag*2 set 2 F (1623-1643)	GGG GCC ATC AGA GTC TAA ACC	150	
	MB HERV-K10 *gag*2 set 2 R (1749-1773)	GGT TGT AAT CTT ACG AGC ACC TG		
	MB HERV-K10 *pol* set 1 F (4506-4530)	TGA CTG TTA TAC ATT TCT GCA AGC	87	X
	MB HERV-K10 *pol* set 1 R (4569-4593)	TGA AAA GGA GTA *GAG* GTT TGG ATC		
	MB HERV-K10 *pol* set 2 F (4508-4532)	ACT GTT ATA CAT TTC TGC AAG CAG	109	X
	MB HERV-K10 *pol* set 2 R (4593-4617)	TGA AAA GGA GTA *GAG* GTT TGG ATC		
	MB HERV-K10 *env* set i F (7383-7404)	ATT GCT TAC TTG CAT TGA TTC AAC	125	X
	MB HERV-K10 *env* set i R (7484-7508)	AAA ATA TGG ACG GAT GGT *GAG* G		
	MB HERV-K10 *env* set ii F (7383-7406)	ATT GCT TAC TTG CAT TGA TTC AAC	85	X
	MB HERV-K10 *env* set ii R (7444-7468)	TCA GTC AAA ATA TGGA CGG ATG G		

Table 5.4 All candidate primer pairs that were trialled for inclusion within the real-time QPCR assay. Primers were then tested using the assay and those showing the most promising data were developed further. Primer pair highlighted in yellow indicates that chosen for incorporation into the real-time assay. X – Indicates functional primer pair.

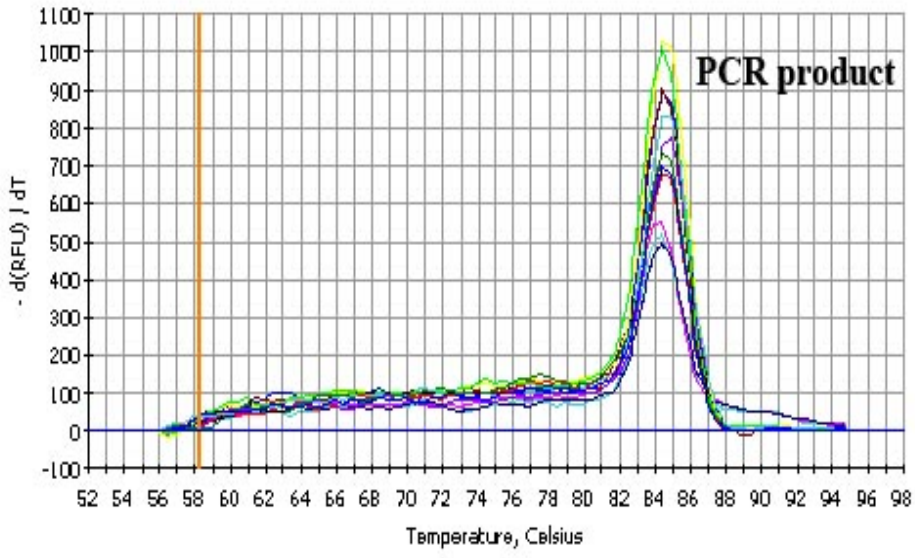


Figure 5.3 Melt curve analysis of HERV-K10 *gag1* set1 primer set designed and incorporated into a HERV-K *gag* specific real-time RT-PCR assay. The single peak represents a single PCR product and the primer's specificity. The different coloured lines indicate reactions carried out in different wells. In this case over a dilution series, each peak is progressively smaller as the starting amount of RNA becomes more diluted.

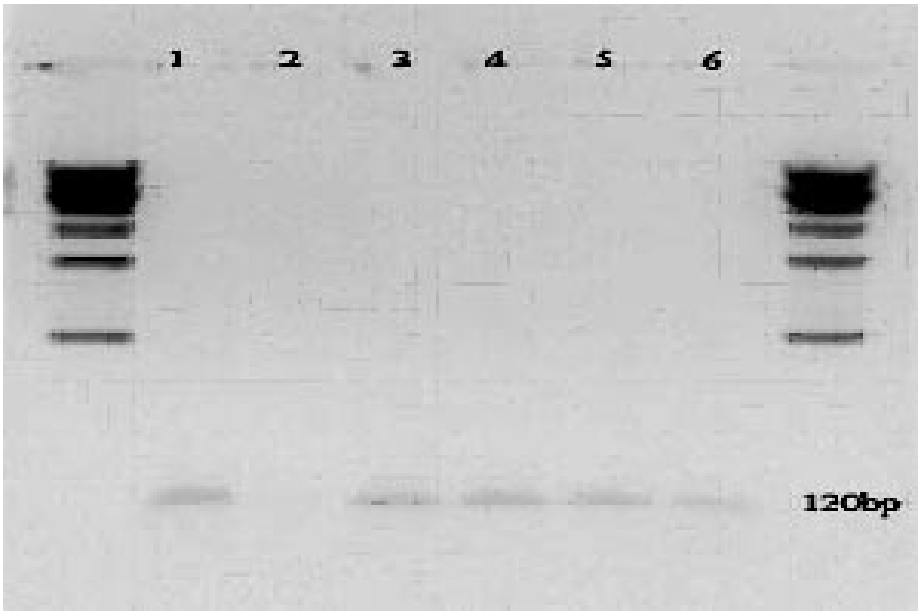


Figure 5.4 A 2% agarose gel shows single PCR product of 120 basepairs (bp) generated using HERV K *gag* set1 primer set with 3 replicates in three samples of total RNA extracted from three different cell lines - Tera-1 cDNA(4, 5, 6) , HEK-293 (3) & Fibroblast-like synovial fibroblasts derived from a healthy donor (1). Run 2 represent dH_20 control (2). Presence of a single band of expected size confirms primer specificity supporting melt-curve data. DNA marker is the 100 bp DNA ladder (Promega).

5.3.1.2 Assay specificity and sensitivity

The sensitivity of the novel real-time quantitative PCR assay was confirmed by the detection of HERV-K *gag* mRNA at dilutions of up to 1 pg. Assay specificity was confirmed by the presence of a single PCR product shown in the melt curve analysis, i.e. exhibiting a single large peak at between 84 - 85°C (Figure 5.3). Additionally, the melt curve showed limited shouldering indicating the absence of primer dimers and other non-specific products. The specificity of PCR products was further confirmed by a single band observed in a 3% agarose gel (Figure 5.4), which was excised, purified and sequenced. Results from DNA sequence analysis (Table 5.5) showed best matches to HERV-K *gag* gene, with HERV-K 10 *gag* in the top 5 most homologous matches in each of the products sequenced thus confirming the primer specificity.

Sample	BLAST results	Accession no.	e value	Gaps	% Similarity
1	*Homo sapiens* endogenous virus Human endogenous retrovirus K HERV-K50A	DQ112146	$2e^{-31}$	0/80	98
1	*Homo sapiens* endogenous virus Human endogenous retrovirus K HERV-K68	DQ112096	$2e^{-31}$	0/80	98
1	*Homo sapiens* endogenous retrovirus HERV-K103	AF164611.1	$1e^{-29}$	0/80	97
1	Human endogenous retrovirus HERV-K10	M14123	$1e^{-29}$	0/80	97
1	*Homo sapiens* endogenous virus Human endogenous retrovirus K HERV-K50D	DQ112152.1	$5e^{-28}$	0/80	96
2	Human endogenous retrovirus K115 complete genome	AY037929.1	$2e^{-31}$	0/80	98
2	*Homo sapiens* human endogenous retrovirus HERV-K(HML-2.HOM)	AF074086.2	$1e^{-29}$	0/80	97
2	*Homo sapiens* endogenous virus Human endogenous retrovirus K HERV-K68	DQ112096.1	$5e^{-33}$	2/90	96
2	Human DNA sequence from clone RP11-29H23 on chromosome 1	AF164610.1	$5e^{-33}$	2/90	96
2	*Homo sapiens* endogenous retrovirus HERV-K102	AF164610.1	$5e^{-33}$	2/90	96
3	*Homo sapiens* endogenous virus Human endogenous retrovirus K HERV-K50A	DQ112146.1	$5e^{-30}$	0/80	97
3	*Homo sapiens* endogenous virus Human endogenous retrovirus K HERV-K68	DQ112096.1	$5e^{-30}$	0/80	97
3	Human endogenous retrovirus K115 complete genome	AY037929.1	$5e^{-30}$	0/80	97
3	*Homo sapiens* chromosome 8, clone RP11-1118M6, complete sequence	AC134684.5	$5e^{-30}$	0/80	97
3	Human endogenous retrovirus HERV-K10	M14123	$2e^{-28}$	0/80	96

Table 5.5 The top 5 matches from BLAST searches to the amplicons sequenced from three PCR products amplified from the Tera-1 cell line and used in experiments to determine sensitivity and specificity. Samples 1, 2 and 3 were the PCR products shown in wells 4, 5 and 6 of Figure 5.4.

5.3.1.3 Primer calibration curve

Initial studies and attempts at optimisation included the determination of the reaction efficiencies of the chosen primer set – HERV-K *gag*1set1. This was calculated using an amplification curve composed of five RNA standards of known quantity (Figure 5.5). These were plotted as a standard curve and the gradient extrapolated to produce the primer efficiencies. The standard curve (Figure 5.6) showed a PCR efficiency of 139.8 %. This was slightly increased over the optimal value of 120%, an effect likely due to the presence of inhibitors and artefacts remaining from the first strand synthesis (Qiagen, 2006). After this stage of optimisation all cDNA was diluted 1:5 to limit this inhibitory effect and the 'Linreg' software programme was employed to calculate the reaction efficiency of each sample run thus eliminating the requirement for a standard curve to be constructed for each run (Section 5.1.4).

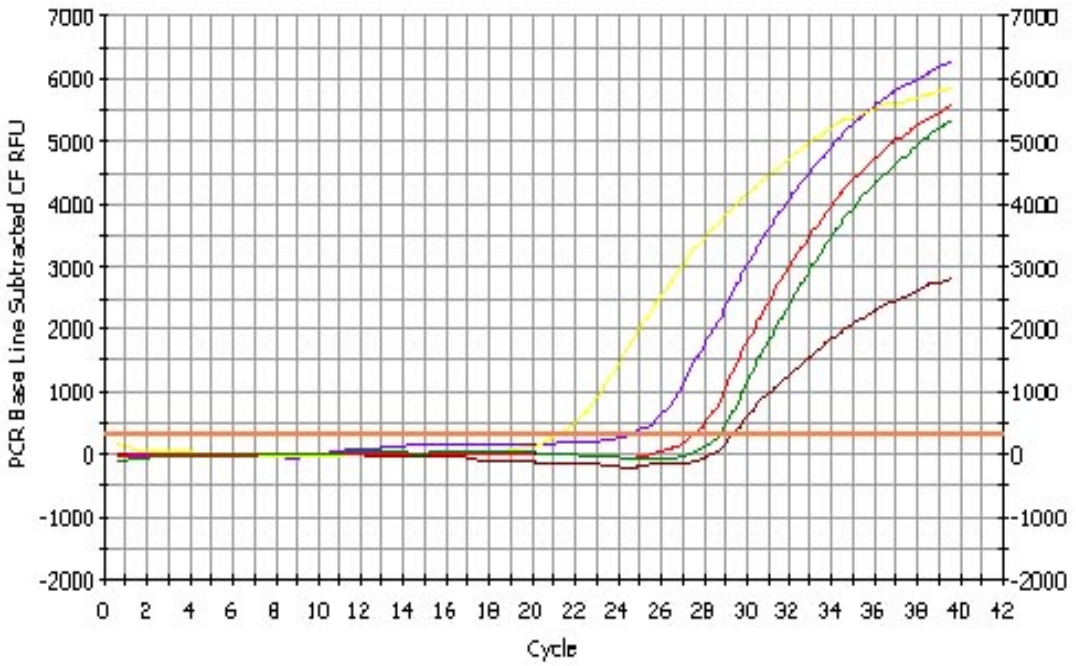


Figure 5.5 Amplification plot of RNA standards derived from Tera-1 total mRNA and used to calculate reaction efficiencies of the primer set HERV-K *gag*1 set1 as shown in Figure 5.6. Plots were constructed using iQ software, Bio Rad).
Key:
Yellow: 100ng. Violet: 10ng. Red: 100pg. Green: 10pg. Black: 1pg.

Orange line represents baseline reaction level as determined using iQ software defaults.

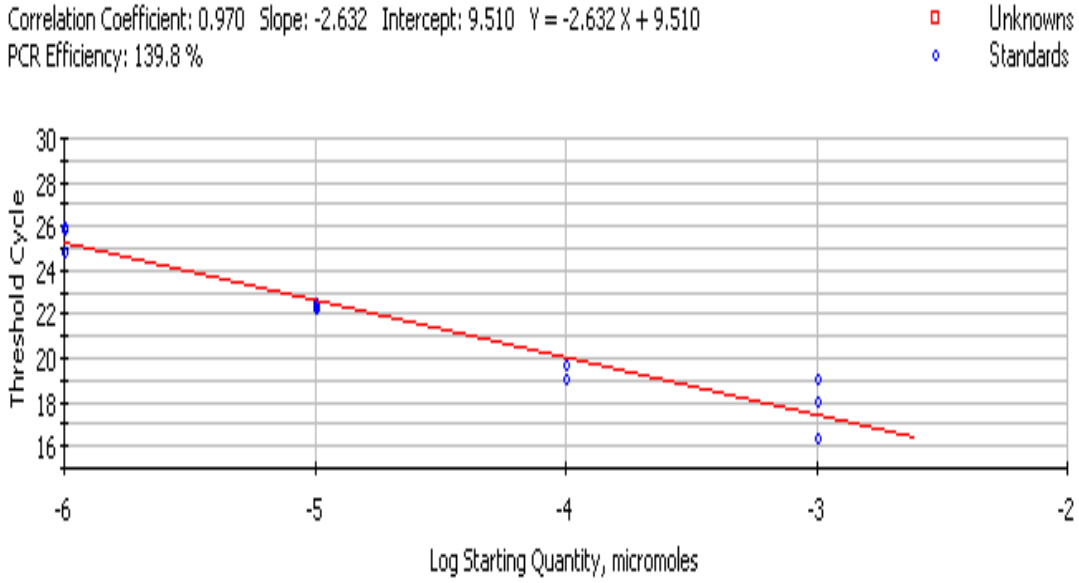


Figure 5.6 Standard curve for HERV-K *gag*1 set 1 primer set. The standard curve was produced using five known dilutions of Tera-1 RNA starting at 1µg. The slope of the line is then used to calculate the primer efficiency. Values over 120% were liklely due to inhibitors or artefacts present from the reverse transcription stage (Bustin *et al.* 2004).

5.3.1.4 Assay optimisation: Primer concentration

In order to maximise the accuracy of the quantitation, primer concentrations were optimised using the primer chess boarding strategy, described by (Gunson *et al*., 2003). Primer concentrations varying between 50 pmol to 1.25 pmol were tested for optimisation purposes. Although only small variations were observed between concentrations of primers tested, data shown in Table 5.5 and Figure 5.7 showed that 25 pmol of forward and reverse primers produced the lowest C_T values. As all reactions were run with the underlying assumption that the initial starting quantity of template was constant, those exhibiting the Lowest C_T values were considered as those with the most optimal and efficient reaction/ primer concentrations.

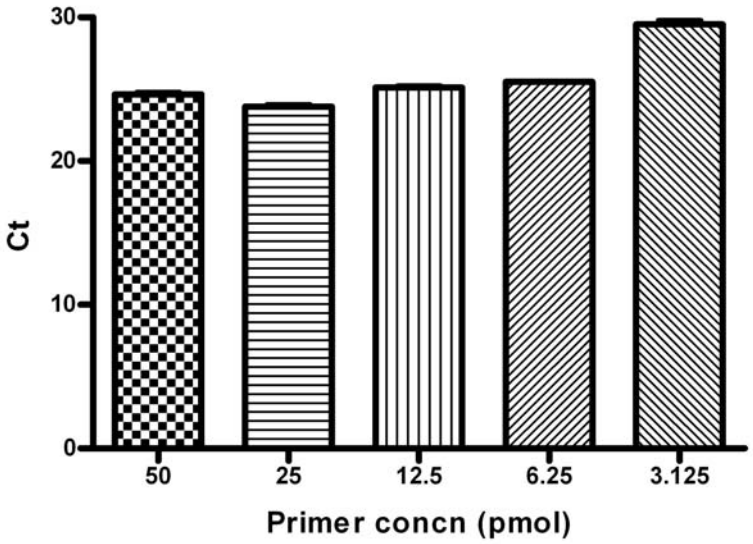


Figure 5.7 The effect of varying primer concentration upon C_T values. Error bars show variation over three runs.

		Forward primer				
	Concn. (pmol)	**50**	**25**	**12.5**	**6.25**	**3.125**
Reverse primer	**50**	24.9	24.4	23	26	27.1
	25	24.6	23.2	23.7	24.6	27.6
	12.5	24.8	25.2	25	24.8	27.9
	6.25	24.6	25	24.3	25.4	28.4
	3.125	25.6	24.5	24.3	25.6	29.4

Table 5.5 Variation in C_T with different primer concentrations using the 'primer chess boarding' technique (Gunson *et al*, 2003).

5.3.1.5 Assay optimisation: Annealing temperature

Samples were tested over a range of annealing temperatures in order to determine the optimal conditions for the '*gag*1 set1' primer set which produced the lowest C_T values. Annealing temperatures tested ranged from 55°C - 57°C at 0.2 °C intervals. Annealing temperature of 55.9 °C gave the lowest C_T values although it should be noted that there was little difference between different annealing temperatures tested (Figure 5.8). An important factor was that this was a very similar temperature to those of the housekeeping genes (whose T_m was 56 ºC; thus both primer cocktails could be run simultaneously under optimal conditions).

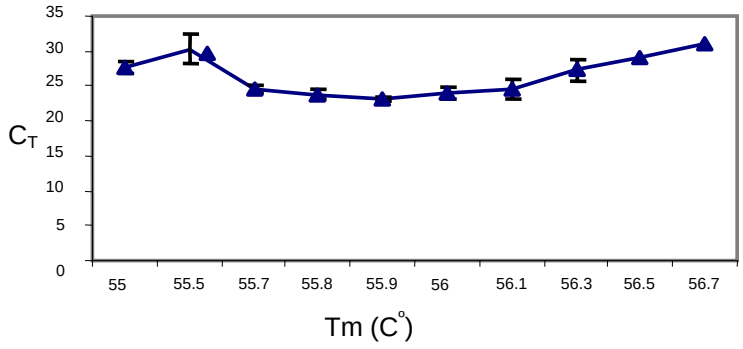


Figure 5.8 The effect of variation of annealing temperature on C_T values for HERV-K10 *gag*1 set1.

5.3.1.6 Intra and Inter-assay validation

In order to assess the accuracy and reproducibility of the assay, C_T values of different serial dilutions of Tera-1 total mRNA were compared (Table 5.6) with a ten-fold dilution series ranging from 1 µg to 1 pg (Figure 5.9). Average C_T values for all intra assay dilutions varied from 14.97 (1 µg) to 29.20 (1 pg). Standard deviation showed little variation between different dilutions until 10 pg, where it increased from ~ 0.3 (1 µg) to over 2.0 (1 pg). For all intra assay runs, Coefficient of variation (CV) was less than 3%, with 1 pg showing both the largest standard error and CV (0.79 and 2.72 respectively). Inter-assay variation observed between runs was increased slightly compared to inter-assay variation. Standard deviation (SD) and CV were comparable across dilutions of 100ng to 100pg although markedly increased at 1pg.

RNA template	Mean C_T	S.D.	CV(%)
	Intra-assay coefficient of variation *(n = 5)*		
1µg	**14.97**	**0.2887**	**1.9288**
100 ng	**23.27**	**0.2887**	**1.2407**
10 ng	**25.93**	**0.1528**	**0.5890**
1 ng	**27.80**	**0.2646**	**0.9517**
100 pg	**28.27**	**0.1155**	**0.4085**
10 pg	**28.73**	**0.2082**	**0.7245**
1pg	**29.20**	**0.7937**	**2.7182**
	Inter-assay coefficient of variation *(n = 5)*		
100 ng	**23.97**	**0.1822**	**0.76**
100 pg	**28.45**	**0.2134**	**0.75**
1 pg	**29.29**	**0.9578**	**3.27**

Table 5.6 Inter assay and intra-assay variation for HERV-K *gag* real-time PCR.. C_T: Threshold cycle, S.D.: Standard deviation, CV: Coefficient of variance.

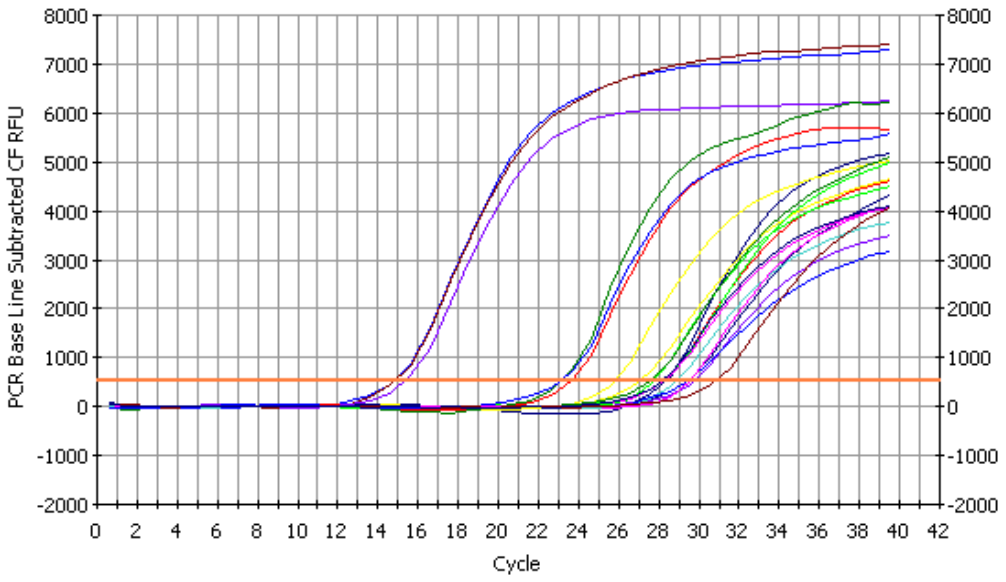


Figure 5.9 Serial dilution series showing amplification of serially diluted RNA standards amplified using the novel HERV-K *gag* real-time PCR assay in order to determine levels of inter and intra-assay variability. Orange line represents baseline as selected as default by machine.

5.3.1.7 Housekeeping genes

Previously published primer sets for four different housekeeping genes (HKGs) were trialled to establish the maximum level of accuracy in quantitation of assay results (Table 5.1). HKG primer pairs were tested using total RNA from two human derived cell lines – Tera-1 and HEK-293. Melt curves for both HKGs showed one single peak in each cell type, indicating the synthesis of a single amplicon. The PCR product of ACTB had a T_m of ~ 87 ºC (Figure 5.10), whereas that of HPRT showed a T_m of 82 ºC (Figure 5.11). PCR products from both HKGs were run in agarose gels, with a single band confirming melt-curve analysis. Amplicons were cut out, purified and sequenced to confirm primer specificity (Figure 5.12 and Table 5.8 for ACTB and HPRT respectively)

Beta actin (ACTB) and Hypoxanthine phosphoribosyltransferase 1 (HPRT) were selected upon the basis of suitability to experimental conditions and primer performance. In order to normalise experimental values for the GOI, the geometric means of both HKGs were combined and used to normalise experimental C_T values, using the ΔC_T method of relative quantitation of HERV-K *gag* gene expression.

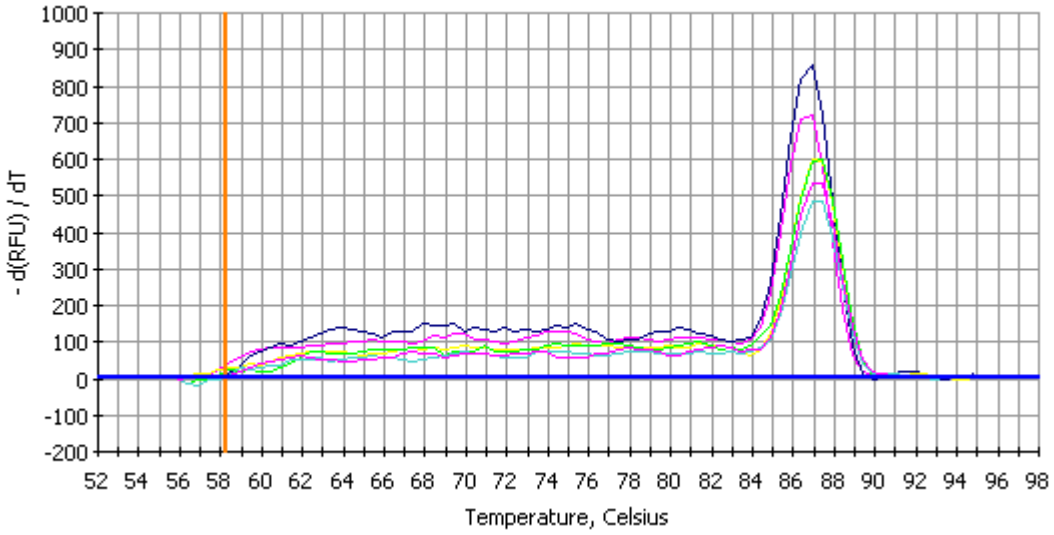


Figure 5.10 Melt curve showing amplification of a single PCR product of Beta-actin using housekeeping gene primers on HEK-293 cells. Different traces represent replicates of same primer set/sample.

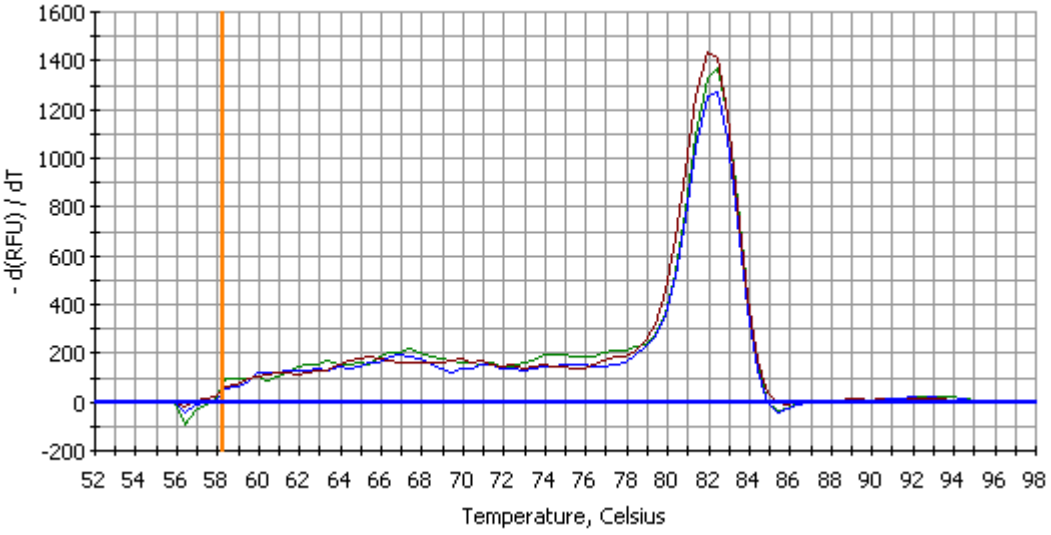


Figure 5.11 Melt curve analysis of HPRT primers used with HEK-293 cells cDNA. The single peak indicates the specificity of the primers and synthesis of a single PCR product. Different traces represent replicate wells of same primer set/sample.

```
                                  ----------XACCXTXTTTGAGACTTC--ACACCCCAGCCATGTACGTTGCTAT Majority
                                          410       420       430       440       450       460       470       480
401  CTGAACCCCAAGGCCAACCGCGAGAAGATGACCCAGATCATGTTTGAGACCTTCAACACCCCAGCCATGTACGTTGCTAT Beta-Actin
1    ----------------------------------CCCACTTTTTGAGACTTC--ACACCCCAGCCATGTACGTTGCTAT 16_001
1    ----------------------------------ACCTGATTTGAGACTTC--ACACCCCAGCCATGTACGTTGCTAT 18_015

                                  CCAGGCTGTGCTATCCCTGTACGCCTCTGGCCGTACTAXCTGGCATCGTGATGGACTCCGGTGACGAXGXCA---XCXXC Majority
                                          490       500       510       520       530       540       550       560
481  CCAGGCTGTGCTATCCCTGTACGCCTCTGGCCGTACCA-CTGGCATCGTGATGGACTCCGGTGACGGGGTCA---CCCAC Beta-Actin
44   CCAGGCTGTGCTGTCCCTGTACGCCTCTGGCCGTACT'ACTGGCATCGTGATGGACTCCGGTGACGAAGCCCTCTACTGC 16_001
43   CCAGGCTGTGCTATCCCTGTACGCCTCTGGCCGTACTA'CTGGCATCGTGATGGACTCCGGTGACGACGACA---GGGTT 18_015

                                  ACTGXXCCCCCATACCAXXGGTATACAAGCAXXXATGGXA----GCXAXTGAAX--GGXTAGXXAAATCCTXAAXAATTA Majority
                                          570       580       590       600       610       620       630       640
557  ACTGTGCCCATCTACGAGGGGTATGCCCTCCCCCATGCCATCCTGCGTCTGGACCTGGCTGGCCGGGACCTGACTGACTA Beta-Actin
124  ATAGCCCCCCCCATACCATCGGAATACAAGTAAGAACGG'G----GCTATTAAA----GATAA--AAATCCTCGAGAATTC 16_001
120  ACTGGTATACCAAACCA--GGTAAACAAGCAGTGATTGAA----GTCA'TGAAT--GGGTAGTTAAATGATTAACAATAA 18_015

                                  TCXXXXXAAXAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX Majority
                                          650       660       670       680       690       700       710       720
637  CCTCATGAAGATCCTCACCGAGCGCGGCTACAGCTTCACCACCACGGCCGAGCGGGAAATCGTGCGTGACATTAAGGAGA Beta-Actin
194  TC'                                                                                                        16_001
192  TAAATAAAAAAAATA'                                                                                           18_015
```

Figure 5.12 Alignment of sequences for amplicon sequence for Beta-actin against reference sequence for gene taken from NCBI (NM_001101). 16_001: Tera-1 cells; 18_015: HEK-293 cells.

Sample	Accession	Description	Max score	Query coverage (%)	E value	Max ident (%)
HPRT 1	NM_000194.1	*Homo sapiens* hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome) (HPRT1), mRNA	161	51%	2^{e}-97	97
	NT_011786.15	*Homo sapiens* chromosome X genomic contig, reference assembly	132	42	1e-28	97
	NW_927721.1	*Homo sapiens* chromosome X genomic contig, alternate assembly (based on Celera assembly)	132	42	1e-28	97
HPRT 2	NM_000194.1	*Homo sapiens* hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome) (HPRT1), mRNA	156	51	8e-36	96
	NT_011786.15	*Homo sapiens* chromosome X genomic contig, reference assembly	128	41	2e-27	97
	NW_927721.1	*Homo sapiens* chromosome X genomic contig, alternate assembly (based on Celera assembly)	128	41	2e-57	97
HPRT 3	NM_000194.1	*Homo sapiens* hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome) (HPRT1), mRNA	93.3	59	6e-17	86
	NT_011786.15	*Homo sapiens* chromosome X genomic contig, reference assembly	94.5	57	1e-05	90
	NW_927721.1	*Homo sapiens* chromosome X genomic contig, alternate assembly (based on Celera assembly)	94.5	57	1e-05	90

Table 5.8 BLAST results from sequenced PCR products from HPRT1 primers shown in Table 5.1. Sequences HPRT1 and HPRT2 were derived from PCR products amplified from HEK-293 cells. HPRT was derived from cDNA synthesised from Tera-1 RNA.

5.3.2 Application of real-time PCR assay specific for HERV-K gag

5.3.2.1 Treatment of cells with 5 – AzacD experimental in Tera-1 cells

Using the novel HERV-K *gag* real-time PCR, the influence of hypomethylating agents upon HERV transcription were investigated in Tera-1 cells. Results showed increasing concentrations of hypomethylating agent resulted in increased HERV-K *gag* transcriptional activity (Figure 5.13). HERV-K *gag* mRNA levels were shown to be increased after treatment with 10 µM of hypomethylation agent, compared to cells treated with lower concentrations (1 µM) and negative controls. The 10 µM hypomethylation treatment was significantly increased over negative controls using one way ANOVA with Tukeys multiple comparison test ($p < 0.001$). 10 µM treatments were not significantly increased over 1 µM treatments ($p > 0.005$).

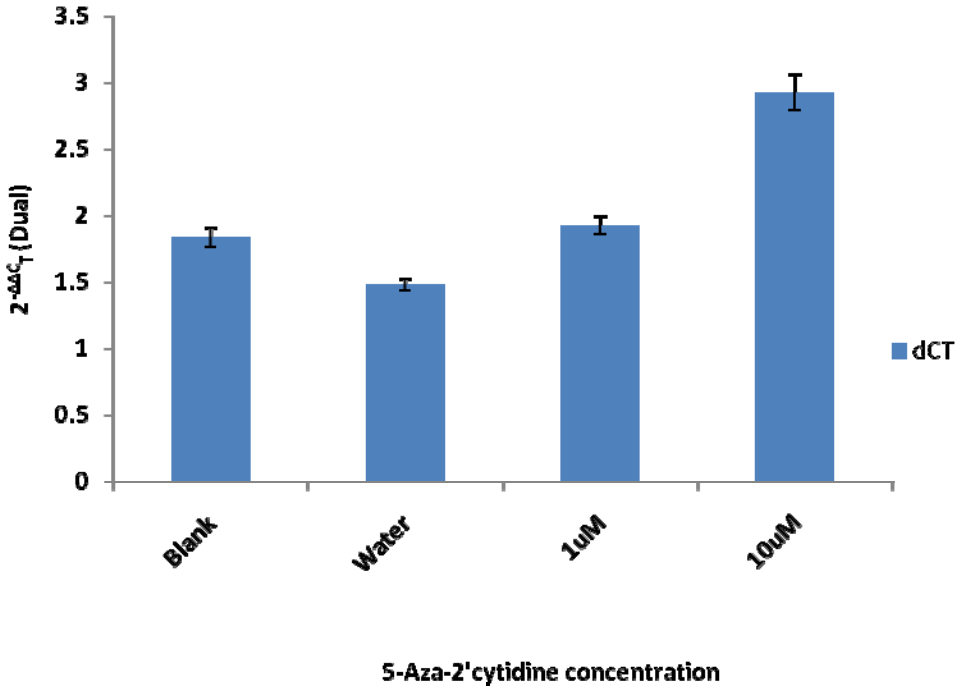


Figure 5.13 Levels of HERV-K10 *gag* mRNA present within Tera-1 cells treated with different concentrations of 5-aza-2'-cytidine relative to two house keeping genes. Error bars indicate degree of standard error in results (*n*=3).

5.3.2.2 Investigation into the levels of HERV-K *gag* RNA in cell lines

The novel real-time PCR assay was used to quantify levels of HERV-K *gag* mRNA in different cell lines. The highest levels of HERV-K mRNA relative to housekeeping genes were observed in the human synoviocytes taken from a donor with RA showing the most elevated level ($2^{-\Delta\Delta C}{}_T$ = ~11) (Figure 5.14). Fibroblast-like synoviocytes (SF) derived from healthy (NHD) and OA donors showed levels comparable to one another ($2^{-\Delta\Delta C}{}_T$ = 3 - 4). Levels in RA SF were significantly increased over OA and NHD by unpaired t-tests (NHD: p = 0.0013; OA: p = 0.0039). All were elevated compared with the non-synovial fibroblast control cell line – HEK-293. Testicular carcinoma (Tera-1) cells, EBV negative B-lymphoma (BJAB) and human leukemic T cell lymphoblast (Jurkat) immortalised cell lines showed levels approximately two-fold that of housekeeping genes (HPRT and ACTB). Levels observed in human breast cancer cell line - MCF7 was reduced in comparison to both lymphoblastic cell lines. In terms of negative controls, the murine fibroblasts showed a low $2^{-\Delta\Delta C}{}_T$ and late C_T with both gene specific and housekeeping primers. A PCR product was identified by melt curve analysis for the beta-actin primer set, giving an indication of sequence conservation between species. No single PCR product was identified by melt-curve analysis, indicating primer specificity for human specific endogenous retroviruses. Human endothelial kidney (HEK-293) cells were also included in the analysis as human fibroblast cell line (of non-synovial origin). They displayed low levels of reactivity only slightly increased over the negative controls (mouse fibroblast).

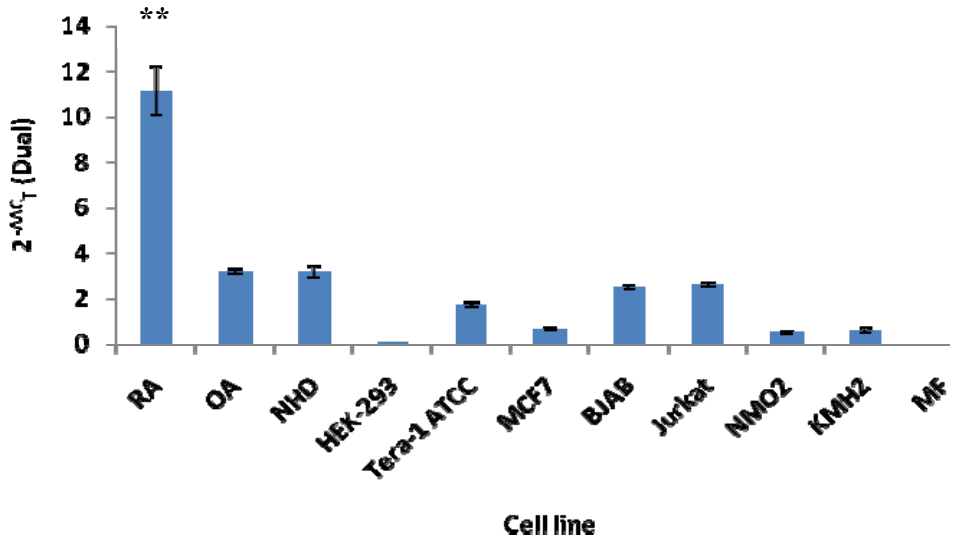


Figure 5.14 $2^{-\Delta\Delta C}{}_T$ values for HERV-K10 *gag* relative to overall HKG expression in different primary and continuous cell lines. RA,OA and NHD are primary cells taken from patients. MF: Mouse fibroblasts. ** indicates the significant increase over both OA and healthy (SF) derived fibroblasts

5.3.3 Use of real-time PCR with RA patient samples and controls

5.3.3.1 Levels of HERV-K *gag* within mononuclear cells of patient venous blood in different disease cohorts

Total cellular mRNA extracted from mononuclear cell populations from venous whole blood of RA patients and control groups were tested using the novel real-time PCR assay to determine levels of HERV-K *gag* mRNA relative to HKG (Figure 5.15)

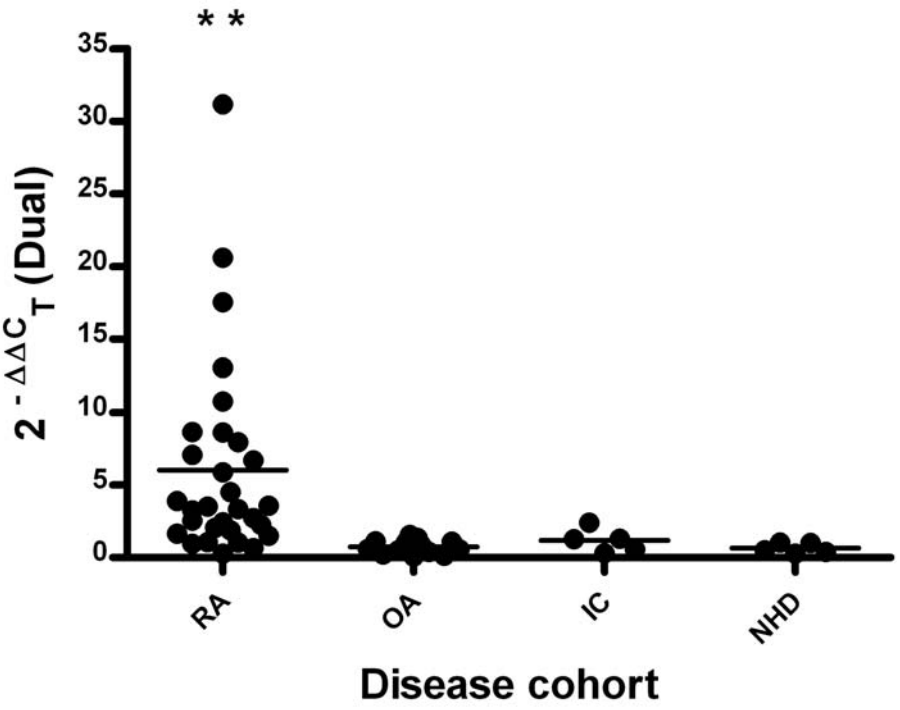


Figure 5.15 The distribution of $2^{-\Delta\Delta C}{}_T$ values for HERV-K10 *gag* $2^{-ddC}{}_T$ relative to overall HKG expression in individual patients. The line indicates the mean of samples.

Figure 5.15 shows that several RA patients exhibited increased levels of *gag* mRNA compared to control groups. Despite variation in $2^{-\Delta\Delta C}{}_T$ values, overall cohort means were elevated in comparison to controls. 11 patients showed $2^{-\Delta\Delta C}{}_T$ values elevated above the group mean. These gave the overall distribution of HERV-K *gag* mRNA a heterogeneous appearance with $2^{-\Delta\Delta C}{}_T$ values (between 5 and 35). All other groups showed limited variation in their results. The second most elevated group was that of the inflammatory disease non-RA controls. Despite the limited size this group

of patients showed a more varied spread of values in comparison to OA and NHDs although the latter represented a small sample set (Figure 5.16). Both healthy and OA cohorts displayed little variation in their $2^{-\Delta\Delta C}{}_T$ results suggesting a possible baseline of activity for HERV-K *gag* in these patient groups. Statistically, the difference between RA and all control groups was significant (OA, NHD and IC) ($p < 0.0001$, < 0.001 and < 0.05 respectively) (Table 5.9). All controls were not significantly different from one another. Using one-way ANOVA analysis to compare all cohorts, all groups were significantly different from one another ($p < 0.0001$).

Disease	RA	OA	IC
OA	< 0.0001 (***)		➢ 0.05
NHD	< 0.001 (**)	➢ 0.05	➢ 0.05
IC	< 0.05 (*)	➢ 0.05	

Table 5.9 Levels of significance of different disease cohorts compared to RA. Analysis was carried out on log-transformed data, using SPSS to carry out one way ANOVA analysis with a Bonferroni post test to compare all columns.

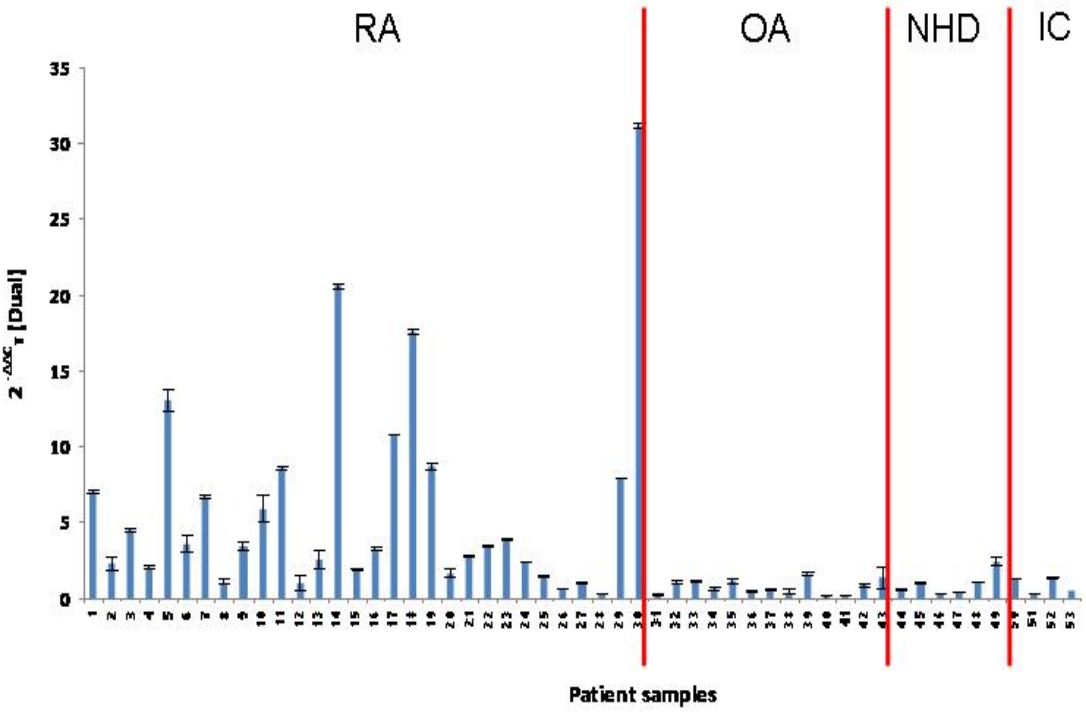


Figure 5.16 $2^{-\Delta\Delta C_T}$ of HERV-K10 *gag* relative to overall HKG. Graph shows disease groups RA – Rheumatoid arthritis (1-30); OA – Osteoarthritis (31-44); NHD - Healthy (43-49); IC - Inflammatory controls (50-54). Error bars show SE for each sample.

5.3.3.2 Age - sex matching of patient samples

Twenty eight patient samples were age-sex matched into 14 pairs, from 14 RA and 14 OA patients. Of those matched pairs, all showed increased *gag* expression in RA patients over their OA counterparts (Figure 5.17). In all pairs, this excess was increased at least ten-fold with mean values of RA patients increased over that observed in OA patients (Table 5.10). The evidence also hinted at a role for gender in HERV activity, as 5 of the top 6 high responders (high responders classified by $2^{-\Delta\Delta C_T}$ levels in excess of 'cohort mean plus 1 Standard deviation') were female. Additionally, female patients exhibited levels of HERV-K gag mRNA almost two-fold that observed in males (Table 5.10) and exhibited an increased heterogeneity in their distribution compared to their male RA counterparts. Of the 6 highest responders (Figure 5.17) (classified as such by exhibiting a dC_T higher than the group mean plus one standard deviation), 5 were female RA patients (aged 51 and 77 years), and 1 male RA patient. The two highest dC_T values were two females of 51 and 53 years of age whilst patient details were incomplete for three, two female were 80 and 83. All high responders had disease durations in excess of 10 years, were rheumatoid factor positive and had histories of erosive disease. Mean ages for both cohorts were comparatively similar (61.3 and 60.96 years for RA and OA patients, and 60.69 and 61.57 years for female and male patients respectively). The highest levels of HERV-K *gag* mRNA observed in an OA patient were from a 38 year old female, of Caucasian descent. The second and third highest responders were female and male, of 78 and 64 years of age, Caucasian descent. All three OA high responders exhibited disease durations of over 10 years.

Overall, RA patients showed levels of HERV-K *gag* mRNA (relative to HKG) that were significantly increased over matched OA patients ($p < 0.0001$ using a two tailed, paired t-test) (Table 5.10). This difference was significant for both genders, i.e. male and female ($p = 0.0067$ and $p = 0.0009$ respectively). Female RA patients were also significantly increased over their male RA paired partners ($p = 0.0392$). This effect was disease specific as in OA the difference between genders (in terms of $2^{-\Delta\Delta C_T}$) was not significant ($p = 0.3182$).

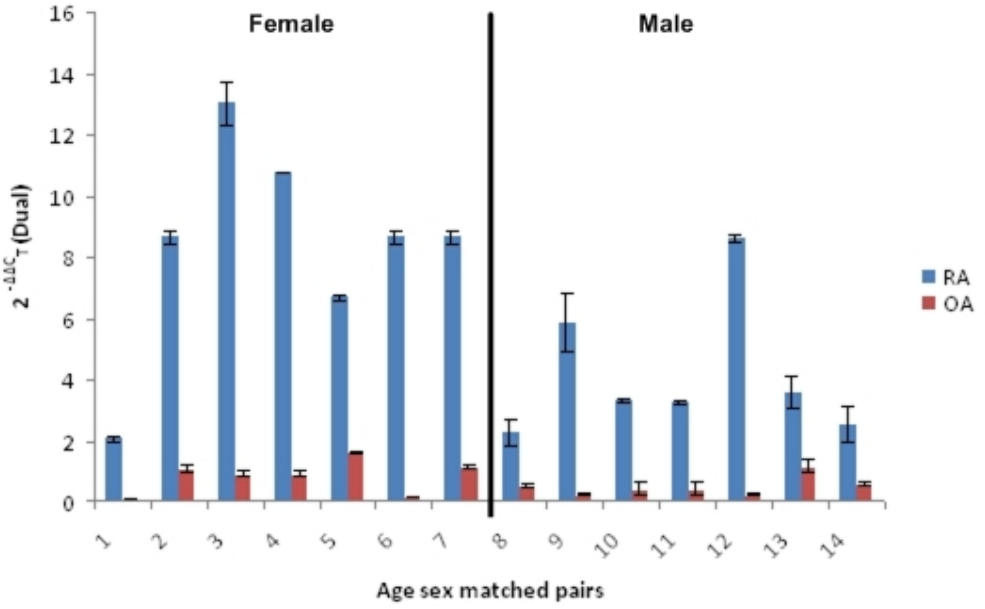


Figure 5.17 $2^{-\Delta\Delta C_T}$ values for HERV-K *gag* levels in OA and RA age: sex matched pairs. Pairs 1-7 represent female: female matched pairs. Pairs 8-14 represent the male matched pairs. Error bars indicate standard experimental error.

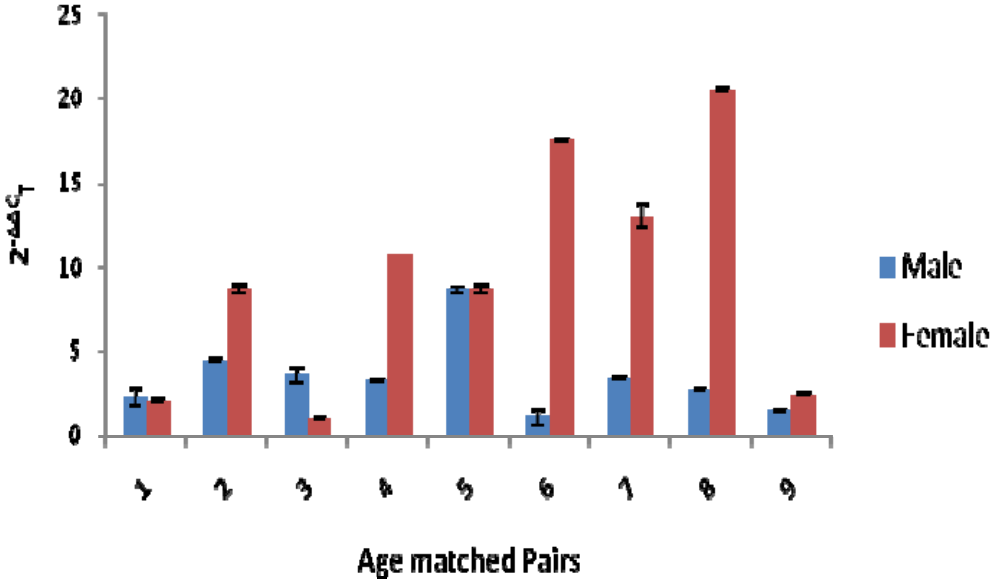


Figure 5.18 Results from the age: sex pairing of RA patients shown in the form of the two opposing genders.

	RA	OA	Significance
Female	8.36	0.81	p = 0.0009 (***)
Male	4.21	0.49	p = 0.0067 (**)
Significance	p = 0.0362 (*)	p = 0.3182 (NS)	P < 0.0001 (***)

Table 5.10 The arithmetic means of the $2^{-\Delta\Delta C_T}$ of age-sex matched pairs in RA and OA cohorts. Statistical data shows the significance of the comparisons of each group using matched pairs and paired (two-way) t-tests. NS: Not Significant. Shaded box represents significance of difference between all matched RA patients vs. all matched OA patients.

These results were also presented as a scatter plot allowing distributions between both cohorts to be examined in greater detail (Figure 5.19). Whilst all OA patients showed a tight clustering about their group mean, RA patients again showed an increasingly heterogeneous response with one cluster of six patients exhibiting levels of HERV-K *gag* mRNA below the cohort mean. All other samples were in excess of the mean. All samples with $2^{-\Delta\Delta C_T}$ levels in excess of the cohort mean were female (with the exception of one). This mirrored distribution observed in the general results (Figures 5.15 – 5.16).

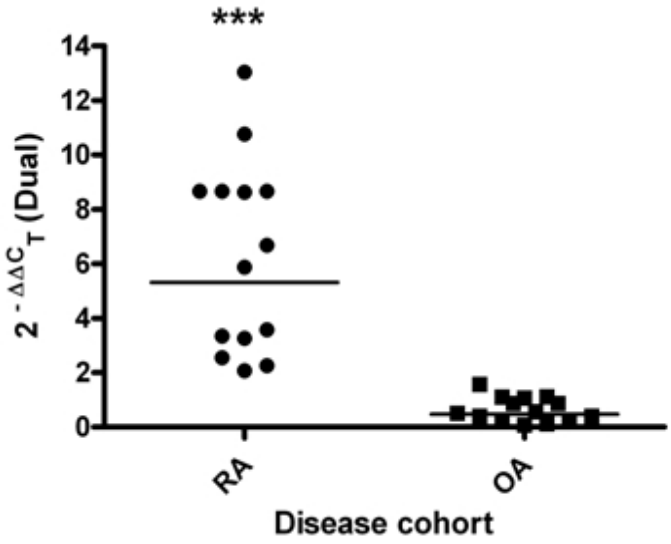


Figure 5.19 The distribution and mean values of both OA and RA age-sex matched samples.

Figures 5.20 – 5.21 show the variation of $2^{-\Delta\Delta C_T}$ with age. Figure 5.20 shows results observed in female RA patients of Caucasian origin. Conversely, Figure 5.21 shows how $2^{-\Delta\Delta C_T}$ levels vary with age in male RA patients of Caucasian origin. A

potential trend in results can be observed in Figure 5.20 with levels of $2^{-\Delta\Delta C}{}_T$ peaking between 40-60 years. This then dips in the sixth decade, before increasing again after 70 years of age. This pattern is not observed in male RA patients who appear to exhibit a baseline of HERV-K *gag* activity that increases in the eighth decade of life (Figure 5.21).

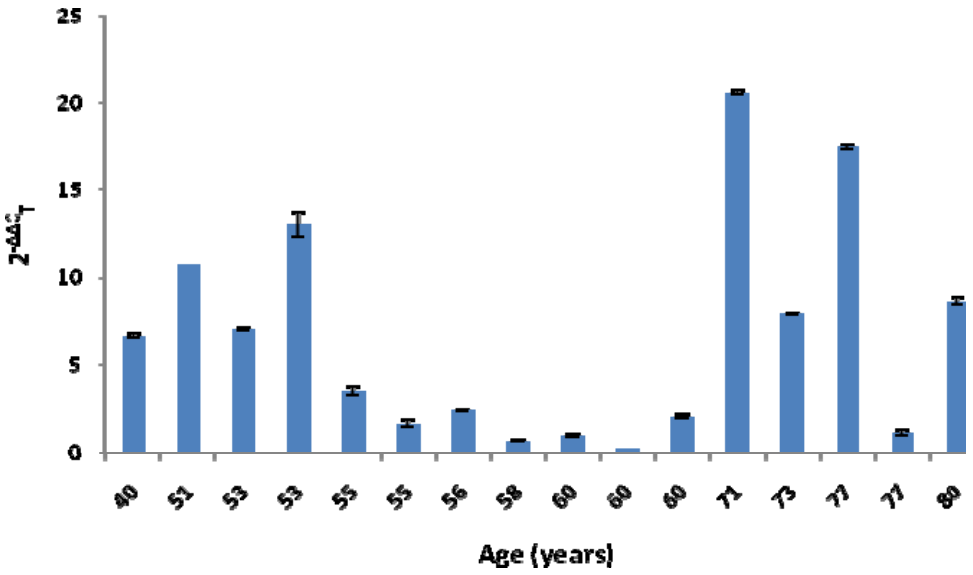


Figure 5.20 Data showing a potential correlation between age and level of HERV-K *gag* activity in female RA patients.

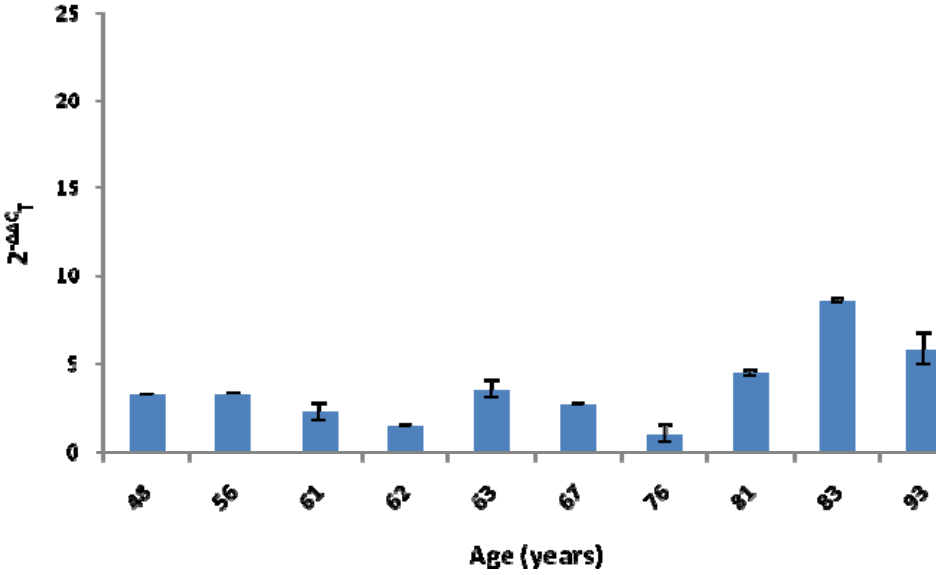


Figure 5.21 Data showing a potential correlation between age and level of HERV-K *gag* activity in male RA patients.

Additionally, disease duration also appeared to contribute to increasing HERV activity. Peaks in $2^{-\Delta\Delta C}{}_T$ correlated with those patients who showed the longest disease duration (Figure 5.22). No trends were observed between $2^{-\Delta\Delta C}{}_T$ and the presence or concentration of Rheumatoid Factor or the development of erosive disease although patient details available at the time of writing were limited on these factors.

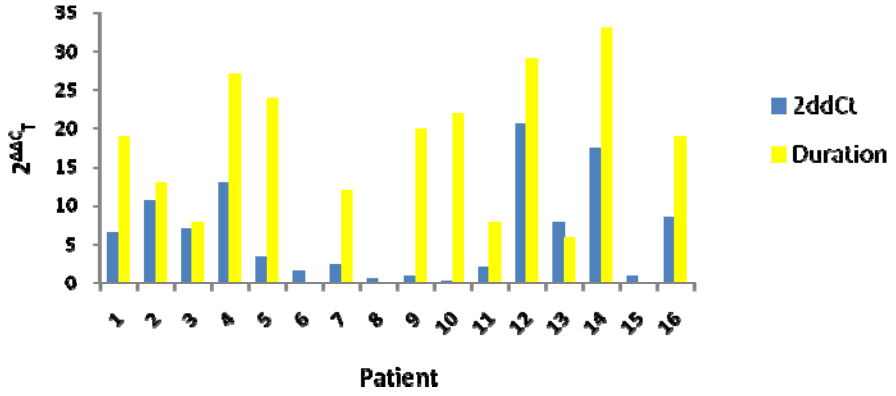


Figure 5.22 The correlation of HERV-K *gag* activity in relation to disease duration in female caucasian RA patients.

5.3.3.3 Cloning and sequencing of HERV-K *gag* PCR products amplified from patient samples

In order to both confirm primer specificity and to identify any disease specific variants in patient samples, HERV-K *gag* PCR amplicons from patient mRNA were amplified via real-time PCR, cloned into the pGEM-T plasmid (Promega, Southampton, UK) to be sequenced (Section 5.2.7). Amplicons were sequenced from 4 RA patients, 2 healthy controls and 2 OA patients were selected at random. In three of the four cases, increasing the insert improved the frequency of white colonies, transformed by vectors with cloned insert (Table 5.11). Problems were encountered during sequencing which prevented the read through of inserted products, therefore purified PCR products were sequenced directly using gene-specific primers. Those sequences retrieved were aligned against both HERV-K *gag*1 and HERV-K *gag* sequences (Figure 5.23).

Sample	Insert : Vector	White	Blue
1	3 : 1	445	58
	1 : 3	318	110
2	3 : 1	214	76
	1 :3	9	23
3	3 : 1	107	76
	1 : 3	10	25
4	3 : 1	0	0
	1 : 3	30	49
+		60	11
Uncut Plasmid only		1	160
Negative control (Media only)		0	0
Transformation efficiency		**1.6 x 10^8 cfu/µg^{-1}**	

Table .5.11 The number of colonies resulting from the each ligation of four RA patient sample PCR products into the pGEM vector. These were used to calculate the overall cloning efficiency for the transformation.

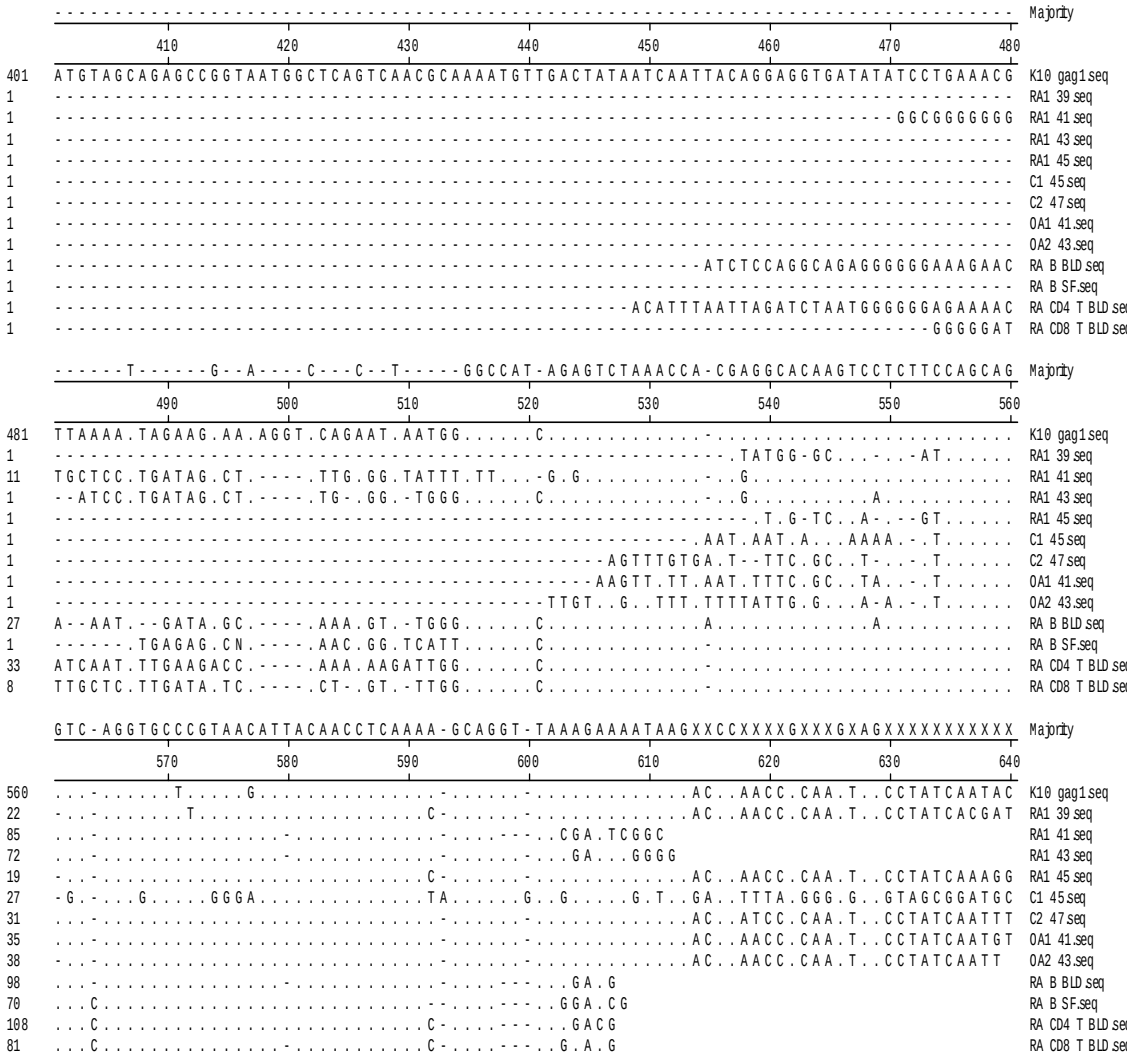


Figure 5.23 Alignment of sequences showing sequencing products amplified from patient samples. Alignment was constructed using MEGAlign version 7 (DNAstar, Madison, Wisconsin).

5.3.3.4 Results for levels of HERV-K *gag* mRNA in different cell types

Figures 5.24 and 5.25 show results of the investigation of levels of HERV-K *gag* in different lymphocyte populations within individual patient samples. Cell populations were separated as described previously (Section 5.2.3) and levels of HERV-K *gag* mRNA measured using the novel real-time PCR.

CD4+ T-cell populations, in both blood and synovial fluid exhibited the highest levels of HERV-K *gag* transcription. Levels within the synovial fluid were elevated in comparison to those observed in the respective cell populations within the blood with the exception of B cells in OA patients. Additionally, both healthy and RA patient lymphocyte populations showed increased levels of HERV-K *gag* transcription, when compared to $2^{-\Delta\Delta C}{}_T$ values exhibited by the OA patient, although levels in RA patients were lower than those observed in healthy lymphocyte populations. This contradicted previous results in PBMCs extracted from in patients samples which demonstrated HERV-K *gag* mRNA to be significantly elevated over both OA and NHDs (Section 5.3.3.1). In OA patient samples, both the T-cell populations showed comparable levels of *gag* activity.

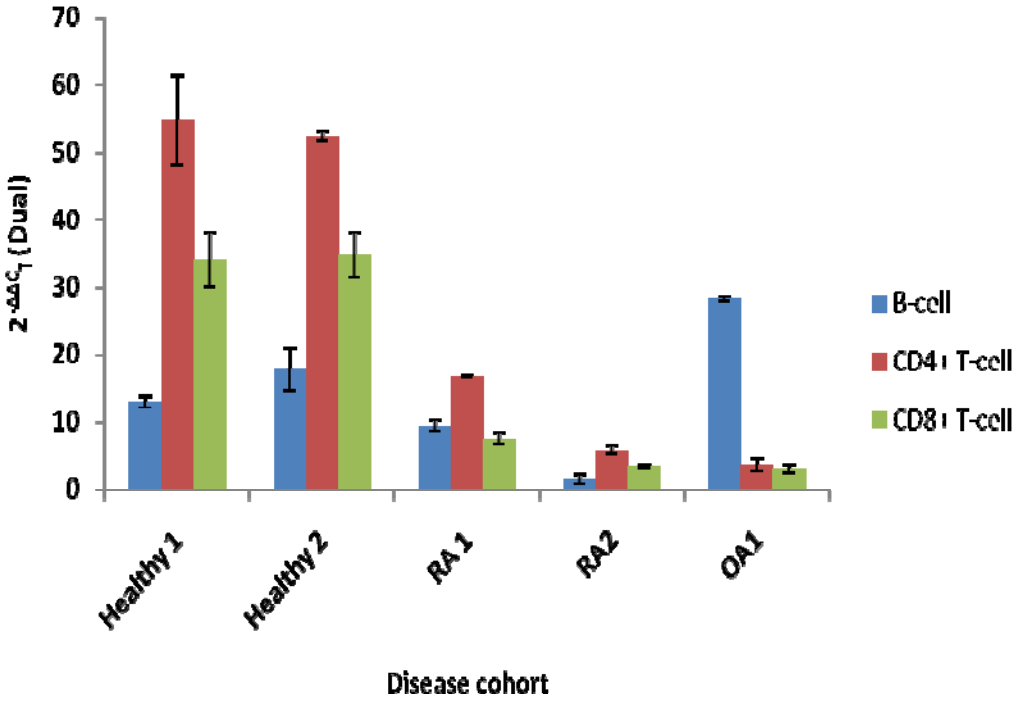


Figure 5.24 Comparison of levels of HERV-K *gag* in different cell types isolated from whole blood across different disease types. Error bars indicate levels of standard error.

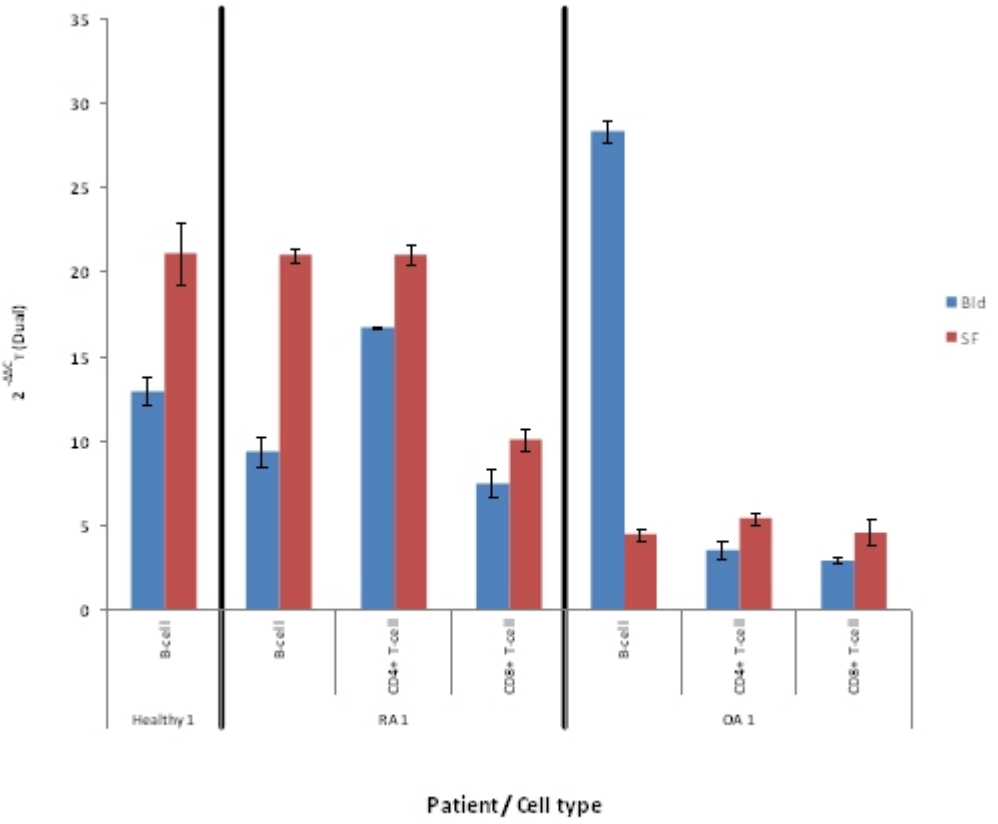


Figure 5.25 Levels of HERV-K *gag* activity relative to HKG expression between cells from blood and synovial fluid. Error bars indicate standard error.

5.3.4 Mechanisms by which HERVs may trigger or influence the immunopathogenesis of Rheumatoid Arthritis

5.3.4.1 The role of viral proteins in influencing HERV-K10 in rheumatoid arthritis pathogenesis

Two expression plasmids were used to investigate the influence of viral genes upon HERV activity. pSG5 plasmids (Stratagene) contained the EBNA1 (pSG5-EBNA1) and LMP1 (pSG5-LMP1) genes from the Epstein Barr virus (EBV). Restriction digests (Section 5.2.9) confirmed the identity of the expression plasmids as shown in Figures 5.26 - 5.27)

Figure 5.26 shows the products of both single and double digests of the pSG5-EBNA1 plasmid. Lane 1 and 2 both show single restriction enzyme digests using *Bgl* II and *EcoR*1 respectively. This had the effect of linearising the plasmid by only partially cutting out the insert, therefore only a single band of just over 6kb is observed i.e. the size of both plasmid and insert. Lane 3 shows the products from a double digest using both restriction enzymes. This completely excises the 2 kb insert from the 4.1 kb pSG5 plasmid.

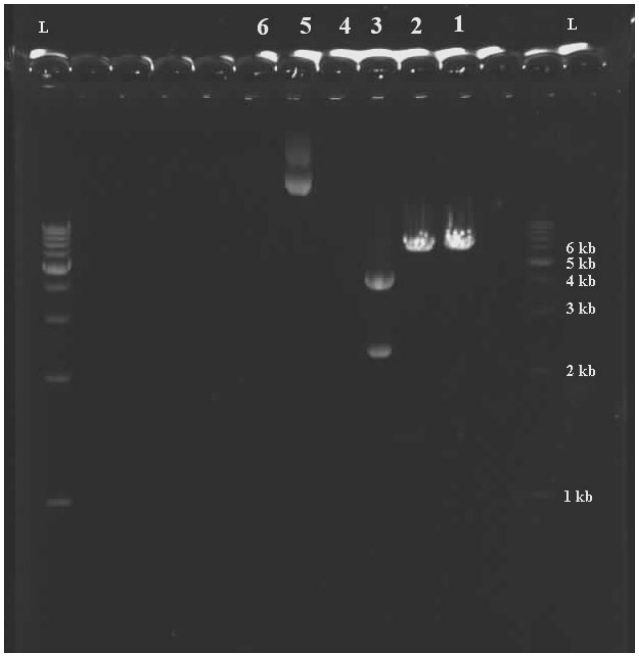


Figure 5.26 Gel confirming the identity of pSG5 – EBNA1 plasmid using a dual restriction digest.
Lanes 1: *EcoR*1. 2: *Bgl* II. 3: Dual digest (*EcoR*1 and *Bgl* II). 4: No template. 5: No enzyme. 6: dH_20.
L : DNA marker Ladder is 1Kb + DNA ladder (Promega, UK).

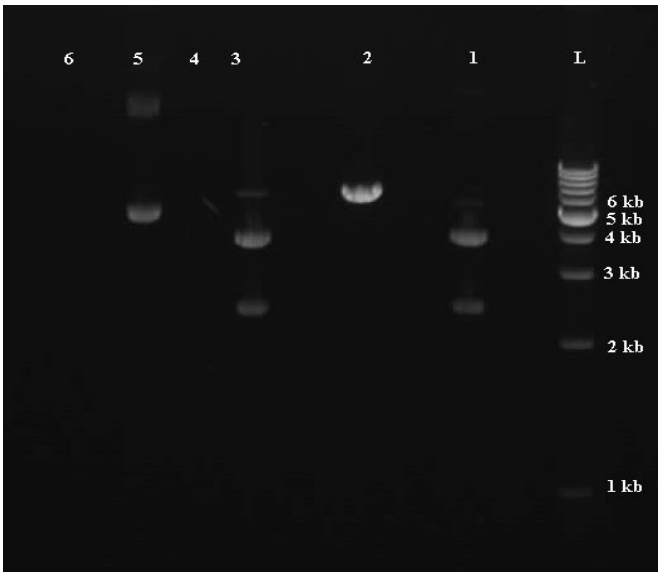


Figure 5.27 Gel confirming the identity of pSG5 – LMP1 plasmid using a dual restriction digest.
Lanes 1: *EcoR*1. 2: *Bgl* II. 3: Dual digest (*EcoR*1 and *Bgl II*). 4: No template. 5: No enzyme. 6: dH_20.
L : DNA marker ladder shown in 1Kb +DNA ladder (Promega, UK)

Additional controls included no template (lane 4), no enzyme (lane 5) and water only (lane 6). A similar pattern as described above was observed for the restriction analysis of pSG5-LMP1 (Figure 5.27).

Real-time PCR was used to determine mRNA levels in human synoviocytes, transfected with expression plasmids containing Epstein barr virus (EBV) genes - EBNA1 and LMP1 (Figure 5.28 and Table 5.12). Cell lines tested included human synoviocytes derived from RA, OA and healthy donors and KMH2 (Hodgkins lymphoma cell line) used as an unrelated control cell line. In addition to the treatments, several negative controls were included within the analysis including untreated, 'vector only' and 'Lipofectamine only' in order to determine the extent of their influence upon HERV-K transcription. All real-time readings were repeated three times in triplicate and normalised with two housekeeping genes – HPRT and ACTB.

The largest increases in HERV-K *gag* mRNA were observed after synoviocytes had been transfected with pSG5-LMP1. Similar effects were observed in all synoviocytes. A similar effect was also observed in KMH2 cells that had been transfected with LMP2a. EBNA1 expression also induced increases in HERV-K *gag* mRNA relative to housekeeping genes, although the increase was not to the extent observed with LMP1. Once again, effects were observed in all cell lines tested (synoviocytes and KMH2).

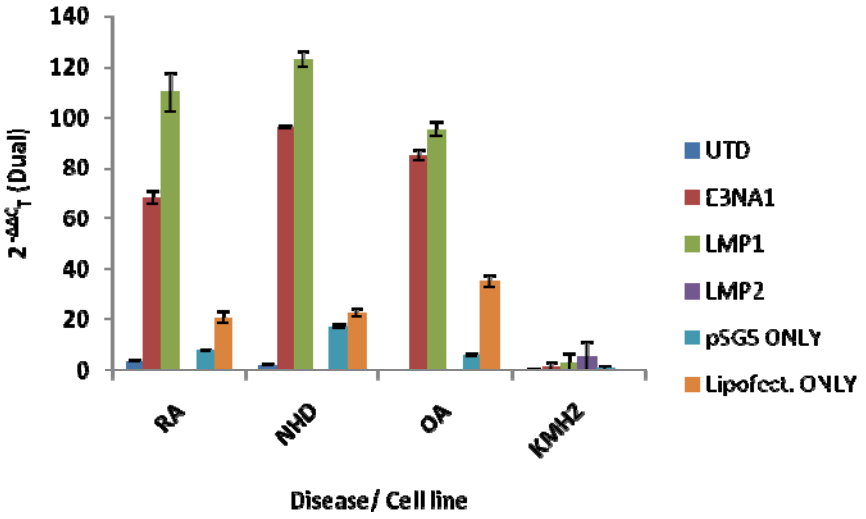


Figure 5.28 Levels of HERV-K *gag* expression in response to the presence of EBV proteins EBNA1, LMP1 and LMP2a in fibroblast-like synoviocytes and KMH2 cells. Expression plasmids for EBV proteins were transfected into cells and HERV-K *gag* measured.

Sample	Cell line/ Disease	Treatment	$2^{-\Delta\Delta C}_T$	SE	p value (to Utd)	Significance
1	RA	UTD	3.508	0.160	-	-
2	RA	EBNA1	68.392	2.506	< 0.001	**
3	RA	LMP1	109.926	7.712	< 0.001	**
4	RA	pSG5 ONLY	7.666	0.210	-	NS
5	RA	Lipofect. ONLY	20.727	2.322	< 0.01	*
6	NHD	UTD	2.138	0.056	-	-
7	NHD	EBNA1	96.244	0.334	< 0.001	**
8	NHD	LMP1	122.880	2.924	< 0.001	**
9	NHD	pSG5 ONLY	17.245	0.711	-	NS
10	NHD	Lipofect. ONLY	22.636	1.580	-	NS
11	OA	UTD	0.824	0.047	-	-
12	OA	EBNA1	85.031	2.178	< 0.001	**
13	OA	LMP1	95.391	2.879	< 0.001	**
14	OA	pSG5 ONLY	6.047	0.533	-	NS
15	OA	Lipofect. ONLY	35.151	2.356	< 0.01	*
16	KMH2	UTD	0.565	0.100	-	-
17	KMH2	EBNA1	68.562	1.383	< 0.001	**
18	KMH2	LMP1	79.809	3.074	< 0.001	**
19	KMH2	LMP2	100.500	5.424	< 0.001	**
20	KMH2	pSG5 ONLY	5.470	0.815	-	NS

Table 5.12 2^{-ddCT} values for treatments of viral proteins using synoviocyte cell lines and KMH2 control cell line. Statistical significance of each treatment was relative to that observed in untreated cells using a paired t-test. NS = not significant.

Of the control groups, both pSG5 (no insert) and lipofectamine only controls were elevated above the untreated cells. This may suggest some of the observed increases in HERV-K *gag* may have been attributable to the experimental conditions. However these increases were only small compared to the increases when viral genes were present. As described in Section 5.3.2.2, untreated RA synoviocytes exhibited significantly increased levels of HERV-K *gag* mRNA when compared to disease controls. However, after treatment with viral genes, synoviocytes taken from healthy donors showed the greatest increases in HERV-K *gag* transcriptional activity Figure 5.24). All treatments showed significance over untreated and treatment controls. The level of significance remained consistent for all treatments irrespective of the cell line investigated, i.e. $p < 0.001$ (Table 5.12). For all treatments, the lipofectamine showed an increased significance above untreated and plasmid only controls. This suggests that the actual process of transfection itself may lead to an increase in HERV-K *gag*. Despite this, treatments were of increased significance to controls suggesting these were representative of true effects rather than artefacts of the experimental process.

Microarray analysis of the influence of different viral genes upon different members of the HERV-K superfamily was also investigated (Figure 6.2).

5.3.4.2 Modulative properties of inflammatory cytokines

After treatment with varying concentrations of inflammatory cytokines it was apparent that cells treated with higher concentrations of pro-inflammatory cytokines exhibited reduced levels of HERV-K *gag* transcriptional activity (Figure 5.29). This was particularly evident in synoviocytes derived from diseased donors, i.e. RA and OA. Normal healthy synoviocytes showed the least differences between high and low concentrations of cytokine treatments. The non-synovial fibroblast controls (HEK-293) exhibited a base level of activity far lower in comparison to those levels observed in synoviocytes. Levels in HEK-293 cells also showed decreased levels of HERV-K *gag* transcription in response to being treated with higher concentrations of IL-1 Beta and IL-6.

Differences in the potential for modulating HERV transcription between the cytokines themselves were also evident. IL-1 β appeared to show the largest increase in healthy synoviocytes, however was untested in any of the other cell lines. Of those cytokines used in all cell lines, IL-6 appeared to show the greatest degree of modulation

with the $2^{-\Delta\Delta C}{}_T$ values for 1 ngml^{-1} being almost five-fold higher in both RA and OA than the 1 ngml^{-1}. TNF-α treatments showed almost a two-three fold difference between the two concentrations. However, 1 ngml^{-1} treatments showed overall peaks in $2^{-\Delta\Delta C}{}_T$ values for HERV-K *gag* mRNA (relative to HKG). All placebo controls containing water were comparable to untreated negative controls.

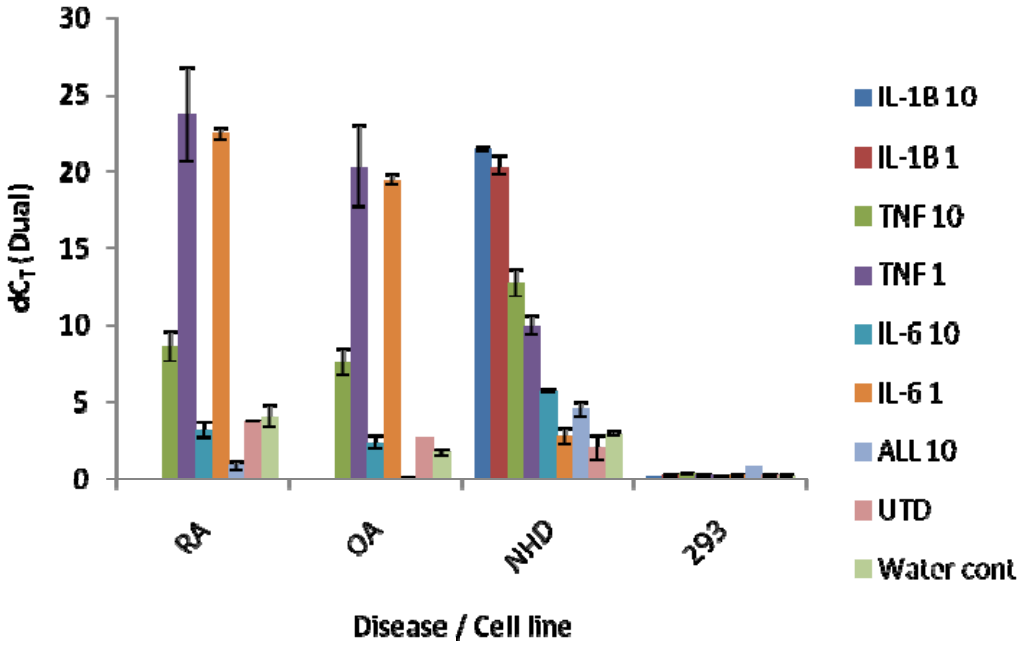


Figure 5.29 The effect of inflammatory cytokines upon HERV-K *gag*. Cytokines at concentrations of 1 ngml^{-1} had a greater influence upon *gag* expression compared to 10 ngml^{-1}, that appeared to show an inhibitory effect. Mouse fibroblasts showed basal levels across all treatments as expected. This graph shows data normalised to ATCB and HPRT genes. Error bars indicate degree of standard error.

5.5 Discussion

5.5.1 Assay optimisation

In this study, an accurate, novel and economical SYBR green real-time RT-PCR was designed to quantify expression of HERV-K *gag* in Rheumatoid Arthritis patients relative to disease controls. Quantitative PCR was a sufficiently sensitive technique to detect <1pg of template (Figure 5.9) and whose specificity was confirmed by both melt curve analysis (Figure 5.3) and sequencing of purified amplicons (Figure 5.4 and Table 5.5).

The optimised assay was then used to investigate the influence of hypomethylation agents upon HERV activity (Section 5.3.2.1). Findings in this investigation confirmed that previously reported in studies that hypomethylation of DNA can lead to increases in transcriptional activity (Ogasawara *et al*., 2003). A number of in vitro cell lines were also surveyed for the presence of HERV-K *gag* (Figure 5.14). Results of HERV-K *gag* expression within two of these cell lines, i.e. MCF7 and human fibroblast-like fibroblasts (from a healthy donor) were backed up by immunohistochemical staining analysis (Figure 4.16, 4.19 and 4.22 – 4.23) using antibodies raised to HERV *gag* epitopes predicted in Chapter 3. Real-time analysis relative to the combined geometric means of two HKG (Section 5.3.1.7) showed significant increases in RA fibroblast-like synoviocytes, compared to disease and healthy controls (Figure 5.14). Additionally, EBV and HTLV immortalised lymphoblastic cell lines – BJAB and Jurkat cell lines showed increased levels of HERVs compared to other cell lines tested. Whether this has any relevance to the viral immortalisation of these cell lines or to any effects of long-term *in-vitro* culture may warrant further investigation.

5.5.2 Patient Samples

Once validated and optimised, the real-time RT-PCR was used to investigate the levels of HERV-K *gag* mRNA in RA patient samples versus disease controls (Section 5.3.3). From this analysis several trends became evident within the results. The level of HERV-K *gag* expression in RA patients was significantly increased in comparison to that of disease controls, particularly that over OA ($p < 0.0001$), NHD ($p < 0.001$) and

inflammatory controls ($p < 0.05$) (Figure 5.15). As inflammatory controls were also increased slightly, this may suggest part of an inflammatory pathway or element of the inflammatory environment may influence HERV-K *gag* expression. This has been suggested previously in work by Serra *et al* who investigated the role of HERVs in the autoimmune response of MS (Serra *et al*., 2003)

Gender may play a significant role in the influence of HERV activity. This supports previous reports which have linked the action of sex hormones to modulation of HERV-K activity (Wang-Johanning *et al*., 2003). Such an effect was particularly notable in the age-sex matching analysis of patient samples (Section 5.3.3.2). Statistical analysis of results (Table 5.10) showed both male and female RA patients were significantly increased over OA matched patients ($p < 0.0001$). This was also the case when the gender specific analysis was completed (Female RA vs. OA: $p = 0.0009$; Male RA vs. OA: $p = 0.0067$) although significance of the difference in male patients was smaller than that observed with females. RA females also exhibited significantly increased levels of HERV-K *gag* mRNA to their paired male counterparts ($p = 0.0362$). When OA females were compared with paired male OA patients this was not the case ($p = 0.3182$) and the effect was disease specific for RA. This observation was disease specific and not observed in disease controls ($p = 0.3182$). Additionally, RA, like many autoimmune diseases has been associated with the actions of sex hormones. Such associations coupled with the fact that disease remissions are often observed during pregnancy (Olsen and Kovacs, 2002) and like most autoimmune diseases, patients appear to be predominantly female (Goronzy and Weyand, 2005) such evidence may contribute to a case for a role for HERVs in Rheumatoid Arthritis pathogenesis.

The real-time data also shows an increased level of $2^{-\Delta\Delta C_T}$ observed in older patients, although this appeared to be reduced within the sixth decade (Figure 5.20). Interestingly, this trend was observed only in female RA patients, and not males (Figure 5.21). RA is a disease that is most commonly triggered in the sixth decade of life (Goronzy and Weyand, 2005). Furthermore results suggest an association between $2^{-\Delta\Delta C_T}$ and disease duration, with increasing disease duration showing a positive correlation with HERV-K *gag* activity (Figure 5.22). It may be that these patients who develop RA before the sixth decade may represent those subsets that may be susceptible to the disease genetically, or come into contact with environmental triggers. These patients in the sixth decade who show a decrease in HERV-K mRNA may represent a cohort that develop disease at this age, and at this point, $2^{-\Delta\Delta C_T}$ remains low

before it increases as the disease progresses. Unfortunately the data available at the time of writing is incomplete and further research is required before making further conclusions (Figure 5.30). It is unlikely however those HERVs are the only factor in a disease such as RA and those developing the disease earlier may be due to a myriad of genetic and environmental variables.

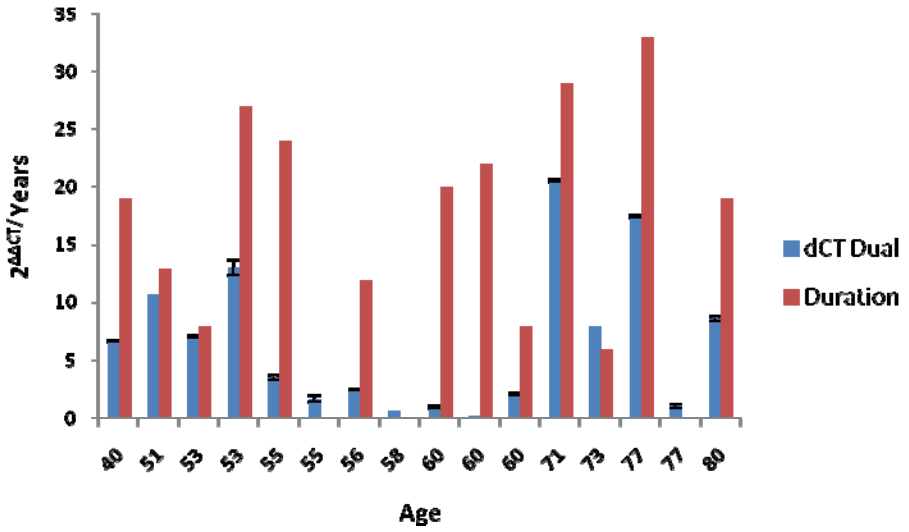


Figure 5.30 The correlation between age, disease duration and HERV-K *gag* activity in RA patients. As shown in the graph levels of HERV-K mRNA decrease within the 6th decade, before it increases in subsequent years. Figure 5.22 shows a correlation with HERV-K mRNA levels and disease duration.
Red bars show disease duration and its correlation with age and HERV-K activity.
Gaps in the line indicate where patient information is unavailable.

5.5.3 HERVs in lymphocyte subpopulations

Part of the investigation also involved the analysis of HERV-K *gag* activity in individual lymphocyte cell populations extracted and purified from patient blood (Section 5.3.3). From the results of this investigation it was evident that HERV expression was not uniform across different cell types. This supported previous studies looking at HERV expression in different tissue and cell types, showing their level of HERV expression to be influenced by the tissues role or metabolic purpose (Seifarth, 2005). Results in this investigation, despite using a limited sample set, showed that HERV-K *gag* activity appeared increased in CD4+ T cells in both healthy (NHDs) and RA patients (Figure 5.24). CD4+ lymphocytes form an important population of cells in RA pathology (Feldmann *et al*., 1996) which may have further implications in disease

pathogenesis. The one exception to this was in OA patients where B-cells showed a large increase over other cell types however due to the limited sample set further investigation is required before any conclusions can be made. Despite its preliminary nature however, they do present an intriguing picture of HERV activity within the cellular microenvironment. The vast majority of these cells are likely to be activated; therefore it would be useful to quantify the level of HERV-K *gag* in quiescent and activated cellular status in order to gain a clearer picture of HERV status.

Further investigation would also identify whether the significant increase of HERV-K *gag* mRNA in B-cells of the OA patient was an artefact or contamination of the sample, or a true representation of HERV-K in OA patients B cells. Furthermore Figure 5.24 shows levels of HERV expression to be decreased in RA patients compared to healthy controls, whereas the quantification of HERV-K *gag* levels in fibroblast-like synoviocytes and other PBMCs of other patients (Figure 5.14 and 5.15) to be increased compared to disease and healthy controls. Clearly these results are contradictory to one another and require further clarification.

Synovial fibroblasts have previously been suggested as key players in the process of synovitis (Huber *et al*., 2006) as they carry out a number of important functions in RA maintaining disease pathogenesis (Section 1.2.6.3). This may suggest that HERVs are in fact present not in activated lymphocytes entering the joint, but in the fibroblast-like synoviocytes, influencing the propagation of disease pathology. This not only places them at the scene of the crime, but shows their upregulation in a cell type, pivotal to the maintenance of disease pathology. Further evidence supporting this observation was previously reported identifying increases in HERV-K expression within the synovial tissue (Ehlhardt *et al*., 2006). This finding was also found on a protein level with increased levels of antibodies being detected within the synovial fluid in comparison to the blood (Section 4.5.4) although increases were not significant (Table 4.5). Additionally our data suggested that HERV-K (HML-2) *gag* mRNA was increased in individual cell populations within the synovial fluid (Figure 4.25). This has been reported previously within the literature (Ejtehadi *et al*., 2006) suggesting that HERVs may form an important part of the disease process, by acting within local cell populations within the joint.

5.5.4 Other factors influencing HERVs

The heterogeneous presentation of HERV-K *gag* in RA patients suggests the lack of a straight-forward association to a disease, already renowned for its highly variable presentation. It appears from this data that a number of factors in both the patients' disease state, i.e. gender, age, disease duration and their genetics, e.g. MHC haplotype, play roles in contributing to the overall disease development. The real-time PCR assay was used to determine how HERVs may be upregulated by endogenous factors within the joint as well as by external ones. Two other factors explored by this investigation included the role of exogenous viral pathogens (Section 5.3.4.1) and the cytokine component of ongoing immune responses within the joint in RA (Section 5.3.4.2).

The presence of exogenous viruses was modulated through the transfection of expression plasmids containing genes cloned from EBV. Those genes investigated were EBNA1, LMP1 and LMP2, with all three having been previously linked with RA disease pathogenesis (Takei *et al*., 1997, Saal *et al*., 1999, Blaschke *et al*., 2000). All three viral genes were shown to have a positive effect upon HERV transcription when present within the expression vector within the synoviocytes (Figure 5.28). The largest increases were observed compared to EBV LMP. LMP2 was only trialled in the KMH2 cell lines; therefore its effects upon HERV-K *gag* require further verification (Table 5.12). In all cell lines tested, i.e. RA, OA and healthy synoviocytes and KMH2 cells both EBNA1 and LMP1 transfection led to significant increases in HERV-K *gag* mRNA levels ($p < 0.001$). These increases were greatest in NHD derived synoviocytes with levels of $2^{\Delta\Delta dC}{}_{T}$ being increased almost 122 fold. HERV-K activity was also increased by transfection reagent controls (lipofectamine) and to a lesser extent in empty vector controls, however these levels were well below those observed with viral gene transfection.

Additionally, cDNA extracted from these transfected synoviocytes were tested using a HERV gene-chip (Seifarth *et al*, 2001) in order to determine activity across the HERV-K superfamily (Appendix ii). From this analysis it showed the HERV-K (HML-2) sub-family (including HERV-K10) was not expressed to a detectable level. This is contradictory to the real-time data which suggests that HERV-K gag is significantly upregulated by the presence of the viral genes (Figure 5.28). The most likely reason for this is that whilst the real-time is specific for HERV-K *gag,* the HERV

microarray is based upon *pol* specific primers. For this reason it may be concluded that this may be the reason the two sets of results do not agree, but also that a full particle is not being produced. The *gag* gene is showing increases in transcripts that are not being reflected in the *env* transcripts. This data is supported by several previously published reports of members of the Herpes family interacting with HERVs (Sutkowski *et al*., 2004, Nellaker *et al*., 2006, Mameli *et al*., 2007).

The second part of the analysis looked at the effect of pro-inflammatory cytokines – IL-1β, IL-6 and TNF-α upon HERV-K (HML-2) *gag*1 transcription (Section 5.3.4.2). Each cytokine was treated at three concentrations – untreated, 1ng/ml and 10ng/ml. These concentrations were chosen in order to simulate those concentrations in-vivo. Concentrations listed in Table 5.13 are those reported in the literature to be typical of circulatory concentrations found in RA patients (Danis *et al*., 1992). Cell lines were treated with cytokines for 48 hours before cells were trypsinised and RNA extracted (Section 5.2.5). After being incubated with pro-inflammatory cytokines, at levels typically observed in RA patients (~1ng/ml) HML-2 *gag*1 was increased in both RA and OA derived cells. At concentrations ten-times those observed in RA patients however, HML-2 *gag*1` transcription appeared to be reduced by over a half (Figure 5.29). In fibroblast-like synoviocytes derived from a healthy donor, however higher concentrations appeared to result in increased transcription as expected. Levels of HERV-K (HML-2) *gag*1 mRNA were similar in all three cell lines in water only and untreated controls. These levels were similar to those observed in Figure 5.14. Additionally, when the non-synovial fibroblast control cell line was treated, levels of HML-2 *gag*1 remained significantly lower than that observed in synoviocytes.

Cytokine	Average level in RA patient blood	Reference
IL-1β	1ng/ml	
IL-6	0.44ng/ml	(Danis *et al*., 1992)
TNF-α	1.2ng/ml	

Table 5.13 Typical circulating oncentrations of pro-inflammatory cytokines under investigation in RA patients.

5.6 Conclusions

In this investigation a sensitive, specific and robust real-time PCR assay was developed, optimised and applied to determine levels of HERV-K *gag* mRNA levels relative to housekeeping genes. This assay detected that HERV-K *gag* transcriptional activity was significantly increased in RA patients compared to disease and healthy controls. Additionally, this real-time PCR assay was used to investigate endogenous factors within the microenvironment of the joint that may be capable of influencing HERV-K activity. Two factors addressed by this investigation were those of exogenous viral pathogens and components of the immune response that play prominent roles in RA disease pathogenesis. Investigations outlined here suggest that both of these factors present within the synovial microenvironment have the potential to influence HERV-K activity, whether on a transcriptional or translational level, i.e. independently of disease state. Further study is however required in order to eliminate HERV-K activity from disease pathogenesis.

A number of conclusions can be drawn from this work.

1. Levels of HERV-K *gag* (relative to HKG) were significantly increased in RA patients, compared to disease controls ($p < 0.0001$).

2. Factors such as age, gender and disease duration play important roles in influencing transcriptional activity of HERV-K *gag*. No correlations could be determined between HERV-K *gag* $2^{-\Delta\Delta C}{}_T$ and the occurrence of Rheumatoid factor, or erosive disease.

3. In mononuclear cell populations in peripheral blood, CD4+ T cells appeared to harbour increased levels of HERV-K *gag* compared to other populations of lymphocytes. These levels were found to be significantly increased in cells taken from the synovial fluid. The relevance of these findings requires further investigation due to the limited sample set available.

4. The presence and expression of Herpesvirus viral proteins, such as those derived from Epstein Barr virus has a positive effect upon HERV-K *gag* transcription. Investigations

showed increased levels of *gag* transcription compared to controls, after transfection of expression plasmids containing EBV genes EBNA1, LMP1 and LMP2. This effect may be gene-specific for *gag.*

5. Proinflammatory cytokines may exert a modulatory effect upon HERV transcriptional activity. At the levels of cytokines present within an RA patient, HERV-K *gag* shows increased transcriptional activity. Ten times these elevated levels show a negative association with HERV-K *gag* transcription.

Chapter 6:

Abbreviations

B-Cell Epitope Prediction (BCEPRED)
Encyclopedia of DNA Elements (ENCODE)
Epstein - Barr virus (EBV)
Human Endogenous Retroviruses (HERVs)
Micro RNAs (miRNA)
Normal Healthy Donor (NHD)
Oral Contraceptive pill (OCP)
Osteoarthritis (OA)
Premature Termination Codon (PTC)
Rheumatoid Arthritis (RA)
RNA Interference (RNAi)
Small interfering RNAs (siRNA)
Sperm Adhesion Molecule-1 (SPAM1)

6.1 Final Discussion

6.1.1 The Central Dogma

'Find the gene, cure the disease', was the optimistic mantra cited by geneticists throughout the 1980's and 90's which embodied the driving force behind the genome cloning revolution (Salmon and Akbar, 2004). Although this phrase embodied a period of paradigms reshaping the field of molecular biology, the reality was less optimistic as the journey from identifying a gene to proving its role and curing the resultant disease was generally a long and arduous one. Over two decades ago the link between Rheumatoid Arthritis (RA) and retroviruses was first reported (Norval *et al*., 1979, Hart *et al*., 1979). Since then this link has been the subject of numerous investigations and reviews (Perl, 2003, Ejtehadi *et al*., 2006) involving several families of Human Endogenous retroviruses (HERVs) in particular (Nakagawa *et al*., 1997). However, no conclusive evidence had been identified that links HERVs to RA disease pathogenesis.

The main purpose of this project was therefore to test the hypothesis that HERVs may contribute to RA disease pathogenesis. The research also involved the development and optimisation of several novel molecular and serological tools and reagents, for use in the detection of HERV gene products in patient samples and cell lines. The three approaches taken were based upon the central dogma (Figure 6.1) and used to determine at which level of activity HERVs may contribute towards disease pathology.

The first approach assessed similarities in the primary structure of HERVs and identified shared regions in autoantigens, implicated in RA disease pathogenesis. Both sets of sequences were mapped antigenically by an optimised and validated method of B-cell epitope prediction and epitopes from both sets were aligned with one another using bioinformatic analysis (Section 3.3.5). From this analysis, a short list of epitopes was drawn up that were shared by both HERVs and autoantigens (Appendix II). These suggested that HERVs may trigger the process of molecular mimicry, resulting in the ensuing cross-reactive autoimmune response.

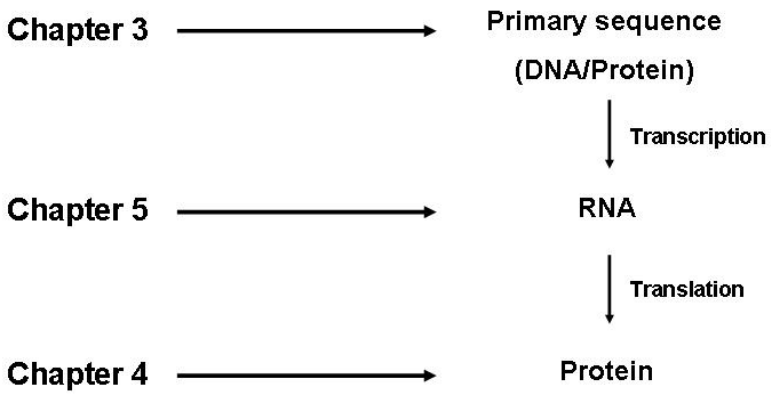


Figure 6.1 The central dogma of molecular biology (Watson and Crick, 1953) described the one-way direction of flow of genetic information in most cellular activities. The first step is the synthesis of RNA using a DNA-dependent RNA polymerase (transcription). The second step is the conversion of RNA to polypeptides / proteins (translation).

Once candidate epitopes had been selected (Section 3.3.7), they were synthesised as short peptides and incorporated into an ELISA for surveying of patient sera (Chapter 4). Reactivity of patient sera to the peptides gave an indication as to whether HERV proteins were contributing to disease, via the induction of antibodies that may also cross-react with self-proteins (Section 4.5). All serological results were confirmed by inhibition studies using patient sera (Section 4.5.3). Such results confirmed the presence of a complete *gag* open reading frame and HERV protein production, thus a potential contribution to disease on a translational level. On a transcriptional level, a novel real-time RT-PCR assay was developed for the detection of HERV-K (HML-2) *gag* mRNA in both patient samples and *in-vitro* models (Chapter 5). This was used to determine levels of HERV-K mRNA in patient samples and disease controls (Section 5.3.3). A novel quantitative PCR was then used to investigate exact mechanisms whereby HERVs were upregulated in RA. Increased levels of HERV-K *gag* were detected upon induced expression of viral proteins in an *in-vitro* model using primary cells taken from RA patients, Osteoarthritis (OA) patients and Normal Healthy donors (NHD) (Section 5.3.4.1). Transfected cells were significantly

increased over controls ($p < 0.001$) showing the expression of Epstein-Barr virus (EBV) proteins, notably LMP1 led to increases in HERV-K *gag* transcription (Figure 5.28). Additionally pro-inflammatory cytokines were also shown to have a modulatory effect upon HERV-K *gag* transcription (Section 5.3.4.2). These results showed that although pro-inflammatory cytokines were capable of modulating increases in HERV-K activity, at increased concentrations this appeared inhibitory, resulting in reductions in HERV-K *gag* expression (Figure 5.29).

6.1.2 Bioinformatic analysis

A final prediction strategy was devised which attained a maximal level of accuracy through the incorporation of dynamic threshold levels and secondary structure, i.e. the location of beta-turns (Section 3.3.3). Using 21 experimentally derived epitopes as 'gold standards', the performance of this methodology was assessed and compared to that of a publicly available database – B-Cell epitope Prediction (BCEPRED) (Section 3.3.4). BCEPRED is an online server that made predictions of B-cell epitopes using the information comprising of 1029 known B-cell epitopes. From the data presented in Chapter 3, the optimal ExPASy method (incorporating input from beta-turns) attained a prediction accuracy of 71%. This was compared to that of BCEPRED which predicted 85% of 'gold standard' epitopes. Despite this increased overall accuracy, the BCEPRED protocol exhibited reduced sensitivity (22% to 41%) and specificity (17% to 25.4%) compared to the ExPASy methodology. Thereby, overall the ExPASy strategy presented a more accurate method of prediction. Despite this, the limitations of the sliding windows method inherently restricted the level of prediction accuracy. Typically, the sliding windows procedure could attain an average of 60% accuracy (Claverie and Notredame, 2006). Other limitations of this analysis included the limited data set of 'gold standard' epitopes used to assess performance, which would require expanding for any future analysis (Greenbaum *et al*., 2007). Additionally, the methodology was limited to the identification of continuous epitopes (Blythe and Flower, 2005), although many discontinuous epitopes may consist of stretches of several consecutive amino acids, thereby making them continuous in their own right (Van Regenmortel, 2006).

B-cell epitope prediction has been notoriously difficult to predict due to the lack of knowledge surrounding the precise interactions between antibodies and their

antigens (Section 3.1.2). Additionally, in comparison to its sister field of T-cell epitope prediction, B-cell epitope prediction is in urgent need of development and standardisation (Petrovsky and Brusic, 2006). For the present line of investigation there are very few clearly set out rules or standards for both successful prediction and validation. Questions should also be raised concerning the acceptable level of sequence homology that makes such a match significant. Is mere sequence homology enough in order to generate a level of cross-reactivity capable of causing disease (Wucherpfennig, 2001).

Initially, bioinformatic analysis was used to determine the degree of sequence homology that exists between HERV derived proteins and host proteins. Such shared regions located in regions most likely to be identified by antibodies (epitopes) were identified and short-listed before validating by serological analysis in Chapter 4. One clear conclusion arising from this work was that perhaps surprisingly there was not as great level of homology between autoantigen and HERV epitopes or primary sequences as previously expected perhaps suggesting that HERVs may not be triggering RA through molecular mimicry (Section 3.3.5).

The investigation of the nature of HERVs and their association with RA employed both serological and molecular approaches, described in Chapters 4 and 5. Once candidate epitopes from Chapter 3 had been identified, these were synthesised as short peptides and incorporated as part of an optimised ELISA to survey patient sera (Section 4.5). Secondly the development and optimisation of a HERV-K *gag* specific real-time RT-PCR assay was used to detect and quantify *gag* mRNA. Both approaches showed significant increases in HERV reactivity in RA patient samples over controls.

Several important factors came to light during the molecular and serological testing of patient samples, although these trends were more prominent in the real-time data. Gender appeared to play a central influence over the level of HERV activity within an individual (Section 5.3.3.2). The real-time data showed females exhibited an almost two-fold increase on levels observed in age-matched males (Table 5.10). More interesting was the fact that although female levels of HERV-K *gag* mRNA were significantly increased compared to matched males in RA ($p = 0.0362$), this effect was disease specific and the two genders were not significantly different in OA ($p = 0.3182$). It has been reported previously within the literature that low androgen levels act as a risk factor for RA onset (James, 1993). This had led to suggestions that the Oral contraceptive pill (OCP) exerted a protective factor for development of RA,

although many of these reports have been found to be contradictory (Silman, 1992, Masi *et al*., 1995). This is further supported as during pregnancy the disease goes into remission (Olsen and Kovacs, 2002). These observations reaffirm the immunological/pro-inflammatory nature of the disease as one of a T_H1 nature. The reasons for the sex bias in RA and other autoimmune diseases are unclear but may include such factors as sex-related differences in immune responsiveness, response to infection, sex steroid effects, and sex-linked genetic factors (Whitacre *et al*., 1999). Additionally, increased levels of HERV-K *gag* in female samples may be due to the females having increased numbers of CD4+ lymphocytes in their peripheral blood. Additionally, this cell population was identified as harbouring the highest levels of HERV-K *gag* in RA patients and healthy controls (Section 5.3.3.4) although these findings should be considered preliminary due to small patient sample size. Increases in HERV-K *gag* in female patients therefore may be due to an increased pro-inflammatory environment established, due to increased levels of estrogens (Cutolo *et al*., 1998). Such a difference may have been identified as a result the action of a gender-specific ERV-derived promoter such as Sperm Adhesion molecule-1 (SPAM1) genes, identified in the testes of both humans and mice (Dunn and Mager, 2005).

6.1.3 Transcription and Translation

Intriguingly gender and disease duration showed weak if no association with absorbance values and serum reactivities (both markers of HERV protein synthesis) (Appendix i). Age and sex matched patients exhibited no trends in their results suggesting that translational and transcriptional mechanisms may be influenced by different host factors. It has been observed previously that differences in levels of antibodies in general, between the sexes are increasingly difficult to identify, thereby supporting findings within this work (Vranckx *et al*., 1986). Real-time PCR data showed increases in mRNA present in RA patients ($p < 0.0001$) compared to disease controls, before translation. In comparison the levels of immune responses to HERV proteins were not as significant as may be expected from the increases observed in the transcriptional data. Although this may only serve as an indirect comparison between the two, further investigation would be required to directly quantify levels of HERV-K gag protein, thus enabling a direct comparison between the two. Potentially such a comparison may highlight important differences in HERV activity upon both

translational and transcriptional levels which may also be reflected within the clinical histories. The significant level of mRNA present in comparison to controls could be explained by several factors. Premature termination codons (PTC) in mRNAs may be encountered by ribosomes, during translation, thereby preventing a complete read through of the open reading frame and synthesis of a functional Gag protein. The process of nonsense mediated decay may also ensure the premature degradation of mRNAs containing PTCs or miRNAs operating at a post transcriptional stage may all account for such increases in mRNAs. This would account for a decreased level of translated protein, in comparison to mRNA. This would be plausible due to the well-documented accumulation of mutations and PTCs in HERV sequences (Nelson *et al*., 2003). However, anti-HERV antibodies were still significantly increased in RA patients compared to controls, suggesting that translation upon some scale was occurring, albeit potentially to a lesser extent, thus making this unlikely.

The increased levels of HERV mRNA observed in patients may be due to external or exogenous factors, affecting mRNA stability or blocking translational processes, thereby preventing or limiting the translational potential of HERV-K (HML-2) *gag*1. SiRNA and miRNAs pathways are two likely candidates for such host mechanisms that may serve such roles (Pillai *et al*., 2007). miRNA in particular is known to repress translation at a post-initiation stage (Seggerson *et al*., 2002, Petersen *et al*., 2006) in addition to being a reversible process (Pillai *et al*., 2007). miRNA mechanisms have also been previously reported to target retroviral activity within human cells (Lecellier *et al*., 2005, Smalheiser and Torvik, 2005, Smalheiser and Torvik, 2006, Triboulet *et al*., 2007). In addition to miRNAs evolution has also led to the development of a number of other ancient, anti-retroviral genome defence mechanisms which may fulfil such a role. These include piRNA (Aravin *et al*., 2007), APOBEC3G (Holmes *et al*., 2007) and paramutation/transvection (Ashe and Whitelaw, 2007) with all three reported to suppress retroviral activity within the host genome.

An intriguing question from this data also asks if HERVs have evolved to equilibrate with their host, as expected, why would HERV-K synthesise any detectable levels of mRNA? Additionally, why are these levels significantly increased in RA patients compared to controls? And does this mRNA play other roles within the host genome?

With advances in technology and completion of several genome sequencing projects, (e.g. Yeast Genome Directory, 1997, *C. elegans* genome sequencing

consortium, 1998, International Human Genome Sequencing Consortium, 2004) the focus of investigations is now addressing the roles of those genes identified as active (Mansia *et al*., 2007, Wellcome Trust Case Control Consortium, 2007). With attention turning to the transcriptome, it has become unexpectedly apparent that the genome displays a greater degree of plasticity than thought previously (Check, 2007). Preliminary data from the first wave of such projects, including that of the Encyclopedia of DNA Elements (ENCODE) consortium, now suggest that from analysis of a fraction of the genome that a larger role than previously considered is regulated by the transcription of vast stretches of non-protein coding DNA (Birney *et al*., 2007). This suggests that genes are intricately regulated by a network of promoters provided by the so-called "junk-DNA", made up of retroviral insertions, repeat elements, etc (ENCODE project consortium, 2007). Is it plausible to suggest a potential role for HERVs in this capacity? Much of the 'junk DNA' is composed of repeats, transposable elements and non-coding DNA largely from introns (Gibbs, 2003). It is in these sequences that retroviral integration clusters have already been shown to play roles in miRNA regulation (Smalheiser and Torvik, 2006, Makunin *et al*., 2007). Additionally, recent reports from studies in mice have implicated an association between retroviral integrations and development of cancers through the alteration of miRNA profiles (He *et al*., 2005, Lum *et al*., 2007, Makunin *et al*., 2007). The existence of an entire endogenous proviral reading frame in the anti-sense orientation capable of producing transcripts in the Complement C4 open reading frame (Mack *et al*., 2004) suggests the potential to adopt such a role and the resulting implications. This however is still a developing field and the significance of these findings in humans is still yet to be revealed.

Such mRNA species may also interact directly with lymphocyte species resulting in their activation. Recent studies have established the capacity for DNA and DNA-associated autoantigens to activate autoimmune B cells via the sequential engagement of BCR and Toll like receptor-9 (TLR-9) (Leadbetter *et al*., 2002, Viglianti *et al*., 2003). Such interactions have also now been established in the activation of autoreactive B cells and RNA/RNA associated autoantigens through interactions with BCR and TLR-7 (Lau, 2005). This supports the observations observed in clinical studies of Rituximab with amelioration of disease symptoms following elimination of CD-19 B cells from patients (Edwards *et al*., 2006). It is therefore possible that autoantibody production may well reflect the presence of increased levels of HERV

mRNA rather than protein. This may emphasise the importance of TLR signalling pathways in autoimmune disease pathogenesis (Lin and Peng, 2005).

6.1.4 Heterogeneity of disease and HERV polymorphisms

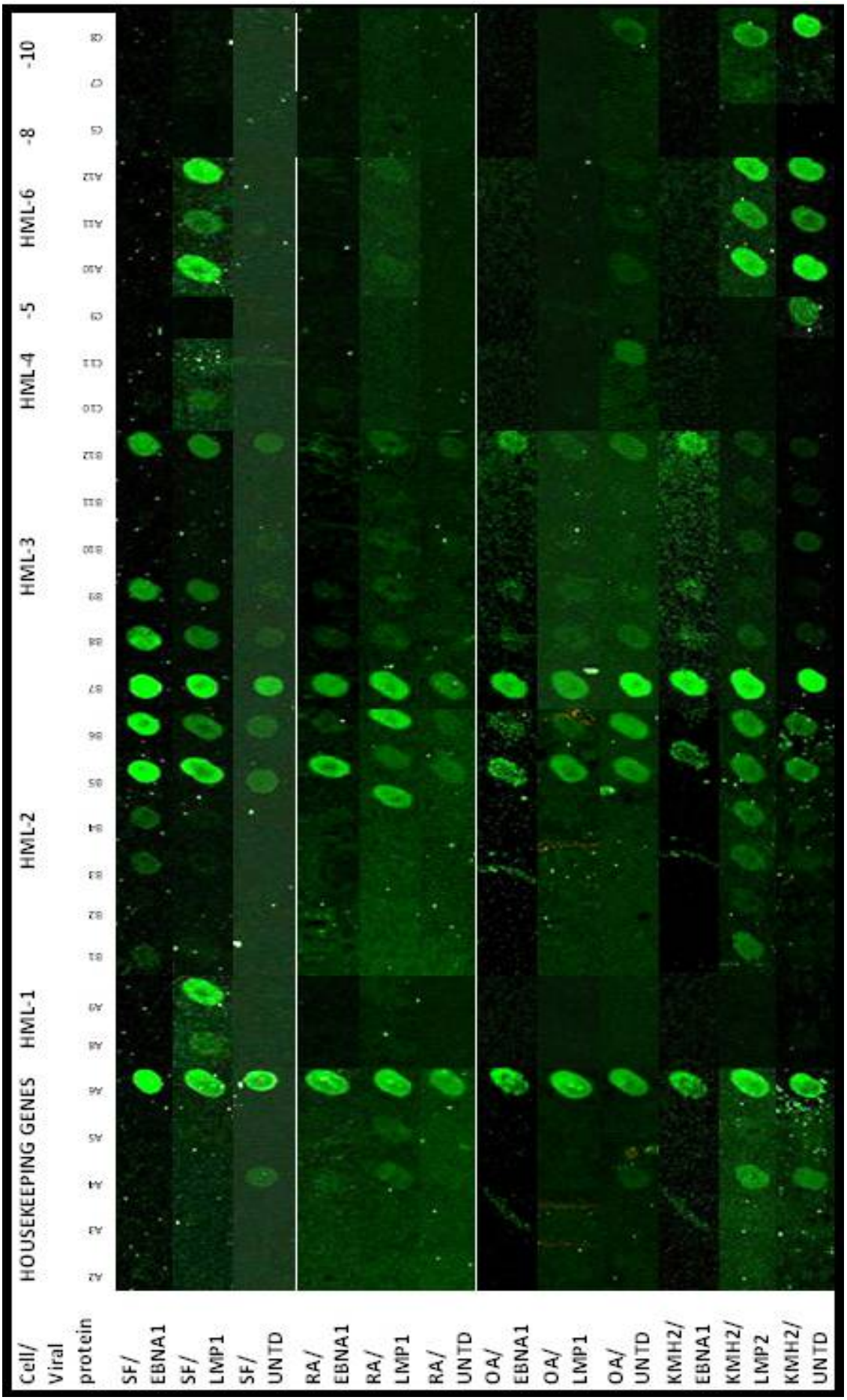
The literature suggests that there is significant activity of HERVs on a transcriptional level. It appears that HERV expression is heterogeneous, varying depending upon individual host factors, i.e. ethnicity (Turner *et al*., 2001, Moyes *et al*., 2005), cell populations (Krieg *et al*., 1992), gender (Chapter 5) and tissue types (Seifarth *et al*., 2005). This is reflected in results presented within this investigation (Figures 4.24 – 4.31 and Figure 5.15). Other factors that may play a role behind this heterogeneous presentation may be due to host repressive mechanisms. This includes the Piwi - piRNA pathway which constitutes an ancient defence mechanism targeting retroviral elements and transposons. Piwis and piRNAs form a system distinct from RNAi and miRNA pathways, although they do share several similarities with miRNAs (Lau *et al*., 2006). Importantly, they are substantially longer than their miRNA counterparts (24 to 30 nt) and poorly conserved amongst closely related species (Aravin *et al*., 2006, Girard *et al*., 2006). Due to this ill conserved nature, individual polymorphisms may develop in individuals resulting in the variation in HERV activities observed in patients. This level of heterogeneity however may not simply be a reflection of host factors as the provirus itself may also constitute an important variable. The issue of proviral polymorphism has recently become topical, with the discovery of several insertional polymorphic members of the HERV-K (HML-2) family – HERV-K113 and HERV-K115 (Turner *et al*., 2001). The existence of polymorphisms within HERV families may help to explain the occurrence of disease affecting only a small proportion of the population (Moyes *et al*., 2007). Such polymorphisms can be either sequence or insertional in nature, although with one exception, disease associations have yet to be clearly established (Marguerat *et al*., 2004). Insertional polymorphisms present an intriguing set of proviruses, particularly when recent studies have shown the potential for reinfection demonstrated by the HERV-K (HML-2) subfamily (Belshaw *et al*., 2004). Additionally the examples of HERV-K(C4), HERV-K113 and K115 show the potential roles for such polymorphisms (Mack *et al*., 2004, Turner *et al*., 2001).

Additionally although polymorphisms may pay a role in differential expression levels in diseased patients, expression between proviral genes within these

polymorphisms may also be a factor. This is illustrated by comparison of the HERV microarray results using samples transfected with EBV genes in expression plasmids (Section 5.3.4.1). Although the *gag* based real-time PCR showed significant differences when EBNA1 and LMP1 were transfected into synoviocytes, results from an *env* based gene chip showed little change in samples among members of the HERV-K superfamily, most notably in HERV-K10 (Figure 6.2 – row B1). This may be explained by the fact that some HERV polymorphisms may have a complete ORF for a single gene, rather than a whole provirus being expressed. Additionally, this may also account for the decrease in significance observed in the ELISA data when the reactivity of Eip peptide is compared to Kpr. Expression of single HERV proteins have been implicated in pathological changes previously within the literature in both germ tumours (Sauter et al., 1996) and in serving important functions within the host, e.g. HERV-W *env* and expression within the trophoblasts of the primate and human placenta (Muir *et al*., 2004)

The possibility that other families of HERVs may also play a role in rheumatoid arthritis should also not be overlooked, although HERV-K constitutes the most plausible candidate. Other HERVs previously linked to RA include ERV3, ERV9, HERV-L and HERV-E (CD5) (Nakagawa *et al*., 1997, Renaudineau *et al*., 2005).

Figure 6.2 Alignment of false colour chip sets corresponding to HERV class II transcriptional activity, observed in healthy and disease derived synoviocytes (RA, OA and healthy) and KMH2 cells transfected with expression plasmids containing Epstein barr virus EBNA1, LMP1 and LMP2a.



Key to Figure 6.2:
HKG **A2**: Ubiquitin (U49869), **A3**: GAPDH (NM_02046.1), **A4**: Alpha-tubulin (NM_006082), **A5**: ATCB (E01094), **A6**: HPRT (NM_000194).
HML-1 **A8**: HML-1 (U35102), **A9**: HML-1 seq29 (S77579),
HML-2 **B1**: HERV-K10 (M14123), **B2**: HERV clone M3.5 (U87592), **B3**: HERV-K (D1.2)(U87595), **B4**: HERV-K2.HOM (U87592), **B5**: HERV-K(HP1)(U87588), **B6**: HERV-K (P1.10)(U87594),
HML-3 **B7**: HERV-K HML-3 (U35236), **B8**: HERV1 (S66676), **B9**: RT244(S77583), **B10**: Seq26 (AC073115), **B11**: Seq34 (AL592449), **B12**: Seq42 (AF047595).
HML-4 **C10**: HERV-K HML-4 Seq10 (AF047591), **C11**: HERV-K-T47D (AF020092).
HML-5 **C9**: HERV-K HML-5 (U35161).
HML-6 **A10**: HERV-K HML-6 (U60269), **A11**: Seq38 (AC010328), **A12**: Seq56 (AC018558).
HML-8 **C5**: HERV-K (P3)(AF030047).
HML-10 **C7**: Seq31 (AL162734), **C8**: HERV-K C4 (U07856).

Overall polymorphic HERVs make better candidates than their ancient counterparts for roles in disease. Therefore, under the right environmental conditions, in genetically predisposed hosts, certain HERV polymorphisms may play a role in disease activation. Another factor requiring clarification was the fact that it had been long known that RA presents itself under the auspices of varying clinical manifestations, disease courses and outcomes. This may be attributed to the possibility that it is in fact still an umbrella term for several separate clinical entities (Weyand *et al*., 1998). A similar situation is observed in SLE with classic lupus, drug-induced lupus and subcutaneous lupus all of which are now recognised as separate disorders (Routsias *et al*., 2004, Conrad and Schlosler, 2002). The heterogeneous nature of the results may indicate particular subsets of patients within a disease. This may be influenced by multiple polymorphism subtypes and observed within clinical symptoms as a varied presentation of seroconversion, erosive disease and extra-articular complications (Weyand *et al*., 1992, Goronzy *et al*., 2004) and germinal centre development (Goronzy and Weyand, 2005).

The second factor important in RA is the disease duration. Once again in the real-time PCR results outlined in chapter 5, longer disease duration showed a positive correlation with HERV-K *gag* transcriptional activity (Figure 5.22). This also seemed to be confirmed in the serological results using the ELISA.

Cell separation strategies were also used to show distribution of HERV activity in different cell types. CD4+ T cells were identified as those with the highest levels of HERV-K *gag* activity in RA and healthy (NHD) patients. This would appear to fit with the fact that T cells (notably CD4+) appear to be those most permissive for retroviral infection, e.g. HIV and HTLV-I. The activation status of these cells and whether this increased HERV activity is due to cellular activation is unknown and requires further investigation. One curious finding was that levels of HERV-K *gag* were shown to be increased in PBMCs of patient samples in RA patients' to controls groups (Figure

5.15). In Figure 5.24, which shows a breakdown of PBMCs into individual lymphocyte populations, the opposite is shown (although the numbers of samples are limited). Here, different cells within the peripheral blood were separated out using paramagnetic beads and levels of HERV-K gag in each population were quantified. Results showed the opposite of that observed in 30 patient samples. What else is present in PBMCs that could account for such a contradiction arising? Monocytes are the other main cell type present within peripheral blood, which did not feature in cell separation strategies (Section 5.3.3.4). Monocytes are phagocytic cells that circulate in the blood for a short period before maturing into macrophage. Macrophages play important roles in RA disease pathogenesis. Macrophage-like synoviocytes are activated in disease and play pivotal roles in synthesising large quantities of pro-inflammatory cytokines such as IL-1, IL-6 and TNF-α in addition to upregulating MMPs, proteinases and free-radicals (Firestein, 2003). Additionally, the macrophage-like synoviocytes are the most predominant cell type at the cartilage-pannus junction (Chung *et al*, 2006). It may be that macrophages from blood or macrophage-like synoviocytes harbour the highest level of HERVs of all PBMC cell populations. Whether this allows HERVs to play a role in disease pathology, or again, they play a role of bystander activation requires further investigation. Levels of HERV-K *gag* were also shown to be significantly increased in cell populations isolated from the synovial fluid, e.g. Table 4.5. This was supported by increased levels of anti-HERV antibodies being identified within the synovial fluid. This was an important finding as it suggested HERVs were increasingly active at the site of disease, supporting their 'triggering of disease' role. The fact that HERVs are upregulated at the site of disease however, may be circumstantial.

RA patients' lymphocytes are characterised by significantly shortened telomeres (Goronzy and Weyand, 2005). This may be explained via increased levels of HERVs contributing to a HERV-specific immune response. The fact that patients mount an antibody response to HERVs on a protein levels, and increased transcriptional activity may hint a chronic host response targeting HERVs from an early age. Such a response may lead to the repeated clonal expansion of immune lymphocytes potentially resulting in the senescence of T-cells (Goronzy *et al*., 2006). This repeated exposure and targeting by the immune system may also be reflected in the significantly shortened telomeres, potentially caused by a persistent viral antigen such as HERV encoded proteins (Pawelec *et al*., 2005). This shortening of telomeres is also associated with the

major genetic risk factor in RA – the HLA-DRB1 allele further substantiating such a link (Schonland *et al*., 2003).

6.1.5 Future work

Hypothesis	Synopsis
T-cell epitope mapping	RA disease pathology is T cell dependent, as well as B cell dependent (Firestein, 2003). Analysis using the TEPITOPE programme (Hammer et al., 1997) may show alternative anti-HERV responses via T cell pathways which may play important roles in RA pathology. This may also identify those antigenic regions that overlapping with both T and B cell epitopes, helping to clarify what constitutes an antigenic region.
Biotinylation of negative control peptide (Negcont1)	How would patient sera react to the biotinylated Negcont1. Would this also result in significant increases in OD values or was this effect specific for HERVK peptides. Such an investigation would validate the use of biotinylation in future ELISA work.
Identification of discrete epitopes in peptides	Kpr is a peptide of 28 aa's in length. Kpr could be synthesised as several shorter overlapping peptides in order to identify the antigenic epitope on the peptide. Observing the effect of this upon patient sera reactivity may provide hints to both peptide antibody recognition mechansms, but also about the location of the discrete epitopes themselves.
Role of HERVs in diseases other than Rheumatoid Arthritis	Appendix IV shows scatter plots of data generated from chronic fatigue syndrome patients and their reactivity to the HERV-K peptides identified within this research. Despite a small sample size (n = 3) this data shows the potential for varying levels of expression within this cohort to different HERV-K peptides. Further research is necessary before any conclusions can be drawn.
Investigation of other HERV families in RA immunopathogenesis.	Other HERVs of interest include HERV-L, ERV9 (Nakagawa *et al*., 1997) and HERV-E (CD5) (Renaudineau *et al*., 2005). Such investigations may include the generation of CD5+ rheumatoid factor producing B cell line (Burastero et al., 1990) in order to further investigate the effects of rheumatoid factor upon HERV activity.

The innate immune response and HERVs	An investigation of interactions between HERVs and innate response. Such an investigation would identify the role of TLRs in anti-HERV responses. Such investigations may involve the transfection of retroviral RNA and proteins into cells and the use of real-time PCR and gene chip analysis to quantitate levels of approopriate TLRs
HERV-K K/O experiments in synovial tissue/ fibroblast-like synoviocytes	What would be effects on cells if HERV-K mRNA expression were knocked out or suppressed? RNA interference could be used to reduce HERV-K activity as described previously (Oricchio *et al*., 2007). Additional approaches including the use of miRNAs to suppress/reduce HERV-K gag levels. Such work could be carried out in transgenic mouse models lacking background HERV activity (Galli et al., 2005). Changes in HERV gene expression would be detected using molecular techniques such as genome analysis and quantitative PCR.
Cell populations harbouring HERV-K	Which immune cells harbour the highest level of HERV-K *gag* activity? In order to further investigate preliminary findings described within this report (Section 5.3.3.4) a larger cohort of samples should be tested and levels of HERV-K *gag* quantified using real-time PCR. Identification of the cell type harbouring increases in HERV-K may help to identify contributions to disease pathology by HERVs.
The roles of viral proteins and their modulation of HERVs	The effect of expressing transactivating proteins such as HTLV Tax and HIV Tat upon HERV-K *gag* expression would be explored in synoviocytes advancing work described in Section 5.3.4.1, using real-time PCR.
HERVs and NF-KB	NF-kB is activated by pro-inflammatory cytokines IL-1β, IL-6 and TNF-α. The effect of modulating NF-**K**B upon HERV-K *gag* activity could be explored. Investigation should be broadened to include other pro and non-inflammatory cytokines.

6.2. Conclusions

As the optimistic geneticists of the early 1990's stated, 'find the gene, cure the disease' (Salmon and Akbar, 2004), at least one candidate gene is described here.

The fact that HERV-K *gag* is significantly upregulated in RA patients, compared to disease controls suggests that it may play a role in disease pathology, even potentially fulfilling the elusive role of disease trigger. Until now upregulation of HERV-K has only been established in RA patients compared to controls within peripheral blood cells (Ejtehadi *et al*., 2006) and at the site of disease (Ehlhardt *et al*., 2006). In recent years however, the field of molecular biology has been turned upon its head with gene expression now known to be far more intricately regulated than previously thought (Greally, 2007). As science delves deeper into the regulatory workings of the genome, more repression mechanisms become apparent, with the majority operating by complex pathways and mechanisms (Stefani, 2007). HERV-K *gag* is significantly upregulated in RA patients, particularly in females, on both a transcriptional and (although to a lesser extent) on a protein level. This however does not necessarily mean it plays an active role in the disease propagation. The results presented here suggest that HERVs are upregulated (on a transcriptional level) although whether or not this is due to other external factors within the joint microenvironment, such as expression of other viral pathogens, or components of the immune response present within the joint requires further investigation. Evidence presented here also hints that the high levels of HERV-K mRNA may in fact be down-regulated or suppressed at the translational level by host or unknown factors.

Chapter 7:

References

ADAMS, B. S., CHA, H. C., CLEARY, J., HAIYING, T., WANG, H., SITWALA, K. & MARKOVITZ, D. M. (2003) DEK binding to class II MHC Y-box sequences is gene- and allele-specific. *Arthritis Res Ther,* 5**,** R226-33.

AGRAWAL, A., EASTMAN, Q. M. & SCHATZ, D. G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature,* 394**,** 744-51.

ALBERT, L. J. & INMAN, R. D. (1999) Molecular mimicry and autoimmunity. *N Engl J Med,* 341**,** 2068-74.

ALIX, A. J. (1999) Predictive estimation of protein linear epitopes by using the program PEOPLE. *Vaccine,* 18**,** 311-4.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-402.

AMFT, N., CURNOW, S. J., SCHEEL-TOELLNER, D., DEVADAS, A., OATES, J., CROCKER, J., HAMBURGER, J., AINSWORTH, J., MATHEWS, J., SALMON, M., BOWMAN, S. J. & BUCKLEY, C. D. (2001) Ectopic expression of the B cell-attracting chemokine BCA-1 (CXCL13) on endothelial cells and within lymphoid follicles contributes to the establishment of germinal center-like structures in Sjogren's syndrome. *Arthritis Rheum,* 44**,** 2633-41.

ANDERSSEN, S., SJOTTEM, E., SVINENG, G. & JOHANSEN, T. (1997) Comparative analyses of LTRs of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset, African green monkey, and man. *Virology,* 234**,** 14-30.

ANDERSSON, A. C., VENABLES, P. J., TONJES, R. R., SCHERER, J., ERIKSSON, L. & LARSSON, E. (2002) Developmental expression of HERV-R (ERV3) and HERV-K in human tissue. *Virology,* 297**,** 220-5.

ANDERSSON, G., SVENSSON, A. C., SETTERBLAD, N. & RASK, L. (1998) Retroelements in the human MHC class II region. *Trends Genet,* 14**,** 109-14.

ANDERSSON, M. L., MEDSTRAND, P., YIN, H. & BLOMBERG, J. (1996) Differential expression of human endogenous retroviral sequences similar to mouse mammary tumor virus in normal peripheral blood mononuclear cells. *AIDS Res Hum Retroviruses,* 12**,** 833-40.

ANTONY, J. M., IZAD, M., BAR-OR, A., WARREN, K. G., VODJGANI, M., MALLET, F. & POWER, C. (2006) Quantitative analysis of human endogenous retrovirus-W env in neuroinflammatory diseases. *AIDS Res Hum Retroviruses,* 22**,** 1253-9.

ARAVIN, A., GAIDATZIS, D., PFEFFER, S., LAGOS-QUINTANA, M., LANDGRAF, P., IOVINO, N., MORRIS, P., BROWNSTEIN, M. J., KURAMOCHI-MIYAGAWA, S., NAKANO, T., CHIEN, M., RUSSO, J. J., JU, J., SHERIDAN, R., SANDER, C., ZAVOLAN, M. & TUSCHL, T. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature,* 442**,** 203-7.

ARAVIN, A. A., HANNON, G. J. & BRENNECKE, J. (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science,* 318**,** 761-4.

AREND, W. P., MALYAK, M., GUTHRIDGE, C. J. & GABAY, C. (1998) Interleukin-1 receptor antagonist: role in biology. *Annu Rev Immunol,* 16**,** 27-55.

ARMBRUESTER, V., SAUTER, M., KRAUTKRAEMER, E., MEESE, E., KLEIMAN, A., BEST, B., ROEMER, K. & MUELLER-LANTZSCH, N. (2002) A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res,* 8**,** 1800-7.

ARNETT, F. C., EDWORTHY, S. M., BLOCH, D. A., MCSHANE, D. J., FRIES, J. F., COOPER, N. S., HEALEY, L. A., KAPLAN, S. R., LIANG, M. H., LUTHRA, H. S. & ET AL. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum,* 31**,** 315-24.

ARNETT, F. C., REVEILLE, J. D., MOUTSOPOULOS, H. M., GEORGESCU, L. & ELKON, K. B. (1996) Ribosomal P autoantibodies in systemic lupus erythematosus. Frequencies in different ethnic groups and clinical and immunogenetic associations. *Arthritis Rheum,* 39**,** 1833-9.

ASHE, A. & WHITELAW, E. (2007) Another role for RNA: a messenger across generations. *Trends Genet,* 23**,** 8-10.

AUGER, I., TOUSSIROT, E. & ROUDIER, J. (1998) HLA-DRB1 motifs and heat shock proteins in rheumatoid arthritis. *Int Rev Immunol,* 17**,** 263-71.

AUPPERLE, K. R., BOYLE, D. L., HENDRIX, M., SEFTOR, E. A., ZVAIFLER, N. J., BARBOSA, M. AND FIRESTEIN, G. S. (1998) Regulation of synoviocyte proliferation, apoptosis, and invasion by the p53 tumor suppressor gene. *Am J Pathol,* **152,** 1091-8.

ARAVIND, L. & KOONIN, E. V. (1999) G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem Sci,* 24, 342-4.

BALDI, P., BRUNAK, S., CHAUVIN, Y., ANDERSEN, C. A. & NIELSEN, H. (2000) Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics,* 16**,** 412-24.

BALTIMORE, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature,* 226**,** 1209-11.

BALTIMORE, D. (1975) Tumor viruses: 1974. *Cold Spring Harb Symp Quant Biol,* 39 Pt 2**,** 1187-200.

BANDZIULIS, R. J., SWANSON, M. S. & DREYFUSS, G. (1989) RNA-binding proteins as developmental regulators. *Genes Dev,* 3**,** 431-7.

BATTAGLIA, M., PEDRAZZOLI, P., PALERMO, B., LANZA, A., BERTOLINI, F., GIBELLI, N., DA PRADA, G. A., ZAMBELLI, A., PEROTTI, C. & ROBUSTELLI DELLA CUNA, G. (1998) Epithelial tumour cell detection and the unsolved problems of nested RT-PCR: a new sensitive one step method without false positive results. *Bone Marrow Transplant,* 22**,** 693-8.

BAUST, C., SEIFARTH, W., SCHON, U., HEHLMANN, R. & LEIB-MOSCH, C. (2001) Functional activity of HERV-K-T47D-related long terminal repeats. *Virology,* 283**,** 262-72.

BEARDSLEY, D. J., TANG, C., CHEN, B. G., LAMBORN, C., GOMES, E. & SRIMATKANDADA, V. (2003) The disulfide-rich region of platelet glycoprotein (GP) IIIa contains hydrophilic peptide sequences that bind anti-GPIIIa autoantibodies from patients with immune thrombocytopenic purpura (ITP). *Biophys Chem,* 105**,** 503-15.

BEASLEY, B. E. & HU, W. S. (2002) cis-Acting elements important for retroviral RNA packaging specificity. *J Virol,* 76**,** 4950-60.

BELSHAW, R., PEREIRA, V., KATZOURAKIS, A., TALBOT, G., PACES, J., BURT, A. & TRISTEM, M. (2004) Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci U S A,* 101**,** 4894-9.
BELSHAW, R., DAWSON, A. L., WOOLVEN-ALLEN, J., REDDING, J., BURT, A. & TRISTEM, M. (2005) Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J Virol,* 79, 12507-14.
BENGTSSON, A., BLOMBERG, J., NIVED, O., PIPKORN, R., TOTH, L. & STURFELT, G. (1996) Selective antibody reactivity with peptides from human endogenous retroviruses and nonviral poly(amino acids) in patients with systemic lupus erythematosus. *Arthritis Rheum,* 39**,** 1654-63.
BENGTSSON, M., STAHLBERG, A., RORSMAN, P. & KUBISTA, M. (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res,* 15**,** 1388-92.
BENIT, L., DESSEN, P. & HEIDMANN, T. (2001) Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *J Virol,* 75**,** 11709-19.
BERGER, C., WEBER-BORNHAUSER, S., EGGENBERGER, J., HANES, J., PLUCKTHUN, A. & BOSSHARD, H. R. (1999) Antigen recognition by conformational selection. *FEBS Lett,* 450**,** 149-53.
BERGLUND, P., STIGHALL, M., JIRSTROM, K., BORGQUIST, S., SJOLANDER, A., HEDENFALK, I. & LANDBERG, G. (2005) Cyclin E overexpression obstructs infiltrative behavior in breast cancer: a novel role reflected in the growth pattern of medullary breast cancers. *Cancer Res,* 65, 9727-34.
BERMAN, H. M., BATTISTUZ, T., BHAT, T. N., BLUHM, W. F., BOURNE, P. E., BURKHARDT, K., FENG, Z., GILLILAND, G. L., IYPE, L., JAIN, S., FAGAN, P., MARVIN, J., PADILLA, D., RAVICHANDRAN, V., SCHNEIDER, B., THANKI, N., WEISSIG, H., WESTBROOK, J. D. & ZARDECKI, C. (2002) The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr,* 58**,** 899-907.
BERNER, B., AKCA, D., JUNG, T., MULLER, G. A. & REUSS-BORST, M. A. (2000) Analysis of Th1 and Th2 cytokines expressing CD4+ and CD8+ T cells in rheumatoid arthritis by flow cytometry. *J Rheumatol,* 27**,** 1128-35.
BHASKARAN, R. & PONNUSWAMY, P. K. (1988) Positional Flexibilities of Amino-Acid Residues in Globular-Proteins. *International Journal of Peptide and Protein Research,* 32**,** 241-255.
BIECHE, I., LAURENT, A., LAURENDEAU, I., DURET, L., GIOVANGRANDI, Y., FRENDO, J. L., OLIVI, M., FAUSSER, J. L., EVAIN-BRION, D. & VIDAUD, M. (2003) Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element. *Biol Reprod,* 68**,** 1422-9.
BIRNEY, E., STAMATOYANNOPOULOS, J. A., DUTTA, A., GUIGO, R., GINGERAS, T. R., MARGULIES, E. H., WENG, Z., SNYDER, M., DERMITZAKIS, E. T., THURMAN, R. E., KUEHN, M. S., TAYLOR, C. M., NEPH, S., KOCH, C. M., ASTHANA, S., MALHOTRA, A., ADZHUBEI, I., GREENBAUM, J. A., ANDREWS, R. M., FLICEK, P., BOYLE, P. J., CAO, H., CARTER, N. P., CLELLAND, G. K., DAVIS, S., DAY, N., DHAMI, P., DILLON, S. C., DORSCHNER, M. O., FIEGLER, H., GIRESI, P. G., GOLDY, J., HAWRYLYCZ, M., HAYDOCK, A., HUMBERT, R., JAMES, K. D., JOHNSON, B. E., JOHNSON, E. M., FRUM, T. T., ROSENZWEIG, E. R., KARNANI, N., LEE, K., LEFEBVRE, G. C., NAVAS, P. A., NERI, F.,

PARKER, S. C., SABO, P. J., SANDSTROM, R., SHAFER, A., VETRIE, D., WEAVER, M., WILCOX, S., YU, M., COLLINS, F. S., DEKKER, J., LIEB, J. D., TULLIUS, T. D., CRAWFORD, G. E., SUNYAEV, S., NOBLE, W. S., DUNHAM, I., DENOEUD, F., REYMOND, A., KAPRANOV, P., ROZOWSKY, J., ZHENG, D., CASTELO, R., FRANKISH, A., HARROW, J., GHOSH, S., SANDELIN, A., HOFACKER, I. L., BAERTSCH, R., KEEFE, D., DIKE, S., CHENG, J., HIRSCH, H. A., SEKINGER, E. A., LAGARDE, J., ABRIL, J. F., SHAHAB, A., FLAMM, C., FRIED, C., HACKERMULLER, J., HERTEL, J., LINDEMEYER, M., MISSAL, K., TANZER, A., WASHIETL, S., KORBEL, J., EMANUELSSON, O., PEDERSEN, J. S., HOLROYD, N., TAYLOR, R., SWARBRECK, D., MATTHEWS, N., DICKSON, M. C., THOMAS, D. J., WEIRAUCH, M. T., GILBERT, J., et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447**,** 799-816.

BLASCHKE, S., SCHWARZ, G., MONEKE, D., BINDER, L., MULLER, G. & REUSS-BORST, M. (2000) Epstein-Barr virus infection in peripheral blood mononuclear cells, synovial fluid cells, and synovial membranes of patients with rheumatoid arthritis. *J Rheumatol,* 27**,** 866-73.

BLOND, J. L., BESEME, F., DURET, L., BOUTON, O., BEDIN, F., PERRON, H., MANDRAND, B. & MALLET, F. (1999) Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. *J Virol,* 73**,** 1175-85.

BLOND, J. L., LAVILLETTE, D., CHEYNET, V., BOUTON, O., ORIOL, G., CHAPEL-FERNANDES, S., MANDRAND, B., MALLET, F. & COSSET, F. L. (2000) An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J Virol,* 74**,** 3321-9.

BLUTHNER, M., MAHLER, M., MULLER, D. B., DUNZL, H. & BAUTZ, F. A. (2000) Identification of an alpha-helical epitope region on the PM/Scl-100 autoantigen with structural homology to a region on the heterochromatin p25beta autoantigen using immobilized overlapping synthetic peptides. *J Mol Med,* 78**,** 47-54.

BLYTHE, M. J. & FLOWER, D. R. (2005) Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci,* 14**,** 246-8.

BODMAN-SMITH, M. D., CORRIGALL, V. M., KEMENY, D. M. & PANAYI, G. S. (2003) BiP, a putative autoantigen in rheumatoid arthritis, stimulates IL-10-producing CD8-positive T cells from normal individuals. *Rheumatology (Oxford),* 42**,** 637-44.

BOGAN, A. A. & THORN, K. S. (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol,* 280**,** 1-9.

BOLLER, K., JANSSEN, O., SCHULDES, H., TONJES, R. R. & KURTH, R. (1997) Characterization of the antibody response specific for the human endogenous retrovirus HTDV/HERV-K. *J Virol,* 71**,** 4581-8.

BONNER, T. I., O'CONNELL, C. & COHEN, M. (1982) Cloned endogenous retroviral sequences from human DNA. *Proc Natl Acad Sci U S A,* 79**,** 4709-13.

BORG, A. A. (1997) Antibodies to cytokeratins in inflammatory arthropathies. *Semin Arthritis Rheum,* 27**,** 186-95.

BORMAN, A. M., QUILLENT, C., CHARNEAU, P., DAGUET, C., CLAVEL, F. (1995) Human immunodeficiency virus type 1 Vif-mutant particles from

restrictive cells: role of Vif in correct particle assembly and infectivity. *J Virol*, 69, 2058-67.
BRAND, A., GRIFFITHS, D. J., HERVE, C., MALLON, E. & VENABLES, P. J. (1999) Human retrovirus-5 in rheumatic disease. *J Autoimmun,* 13**,** 149-54.
BRAND, D. D., KANG, A. H. & ROSLONIEC, E. F. (2004) The mouse model of collagen-induced arthritis. *Methods Mol Med,* 102**,** 295-312.
BRANDEN C. & TOOZE., J. (1999) *Introduction to Protein structure,* London, Routledge.
BRENDEL, V., DOHLMAN, J., BLAISDELL, B. E. & KARLIN, S. (1991) Very long charge runs in systemic lupus erythematosus-associated autoantigens. *Proc Natl Acad Sci U S A,* 88**,** 1536-40.
BRODSKY, I., FOLEY, B., HAINES, D., JOHNSTON, J., CUDDY, K. & GILLESPIE, D. (1993) Expression of HERV-K proviruses in human leukocytes. *Blood,* 81**,** 2369-74.
BRODY, J. R. & KERN, S. E. (2004) Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *Biotechniques,* 36**,** 214-6.
BROWNING, M. T., SCHMIDT, R. D., LEW, K. A. & RIZVI, T. A. (2001) Primate and feline lentivirus vector RNA packaging and propagation by heterologous lentivirus virions. *J Virol,* 75**,** 5129-40.
BUCKLEY, C. D. (2003) Michael Mason prize essay 2003. Why do leucocytes accumulate within chronically inflamed joints? *Rheumatology (Oxford),* 42**,** 1433-44.
BUCKLEY, C. D., AMFT, N., BRADFIELD, P. F., PILLING, D., ROSS, E., ARENZANA-SEISDEDOS, F., AMARA, A., CURNOW, S. J., LORD, J. M., SCHEEL-TOELLNER, D. & SALMON, M. (2000) Persistent induction of the chemokine receptor CXCR4 by TGF-beta 1 on synovial T cells contributes to their accumulation within the rheumatoid synovium. *J Immunol,* 165**,** 3423-9.
BUCKLEY, C. D., FILER, A., HAWORTH, O., PARSONAGE, G. & SALMON, M. (2004) Defining a role for fibroblasts in the persistence of chronic inflammatory joint disease. *Ann Rheum Dis,* 63 Suppl 2**,** ii92-ii95.
BUSHMAN, F. D. (2003) Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell,* 115**,** 135-8.
BUSTIN, S. A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol,* 25**,** 169-93.
BUSTIN, S. A. & NOLAN, T. (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech,* 15**,** 155-66.
BUTLER, J. E., PETERMAN, J. H., SUTER, M. & DIERKS, S. E. (1987) The immunochemistry of solid-phase sandwich enzyme-linked immunosorbent assays. *Fed Proc,* 46**,** 2548-56.
C. ELEGANS GENOME SEQUENCING CONSORTIUM (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science,* 282**,** 2012-8.
CAMPILLOS, M., DOERKS, T., SHAH, P. K. & BORK, P. (2006) Computational characterization of multiple Gag-like human proteins. *Trends Genet,* 22**,** 585-9.
CANTARERO, L. A., BUTLER, J. E. & OSBORNE, J. W. (1980) The adsorptive characteristics of proteins for polystyrene and their significance in solid-phase imunoassays. *Anal Biochem,* 105**,** 375-82.
CANTRELL, M. A., EDERER, M. M., ERICKSON, I. K., SWIER, V. J., BAKER, R. J. & WICHMAN, H. A. (2005) MysTR: an endogenous retrovirus family in

mammals that is undergoing recent amplifications to unprecedented copy numbers. *J Virol,* 79**,** 14698-707.
CARASSO, R., SANDBANK, U. & COHN, D. (1976) Electron microscopy findings in familial hypokalemic periodic paralysis. *Harefuah,* 90**,** 115-7.
CASH, J. D. & MCGILL, R. C. (1969) Fibrinolytic response to moderate exercise in young male diabetics and non-diabetics. *J Clin Pathol,* 22**,** 32-5.
CATRINA, A. I., ULFGREN, A. K., LINDBLAD, S., GRONDAL, L. & KLARESKOG, L. (2002) Low levels of apoptosis and high FLIP expression in early rheumatoid arthritis synovium. *Ann Rheum Dis,* 61**,** 934-6.
CERNY, E. H., FARSHY, C. E., HUNTER, E. F. & LARSEN, S. A. (1985) Rheumatoid factor in syphilis. *J Clin Microbiol,* 22**,** 89-94.
CERTO, J. L., KABDULOV, T. O., PAULSON, M. L., ANDERSON, J. A. & HU, W. S. (1999) The nucleocapsid domain is responsible for the ability of spleen necrosis virus (SNV) Gag polyprotein to package both SNV and murine leukemia virus RNA. *J Virol,* 73**,** 9170-7.
CHACON, N., LOSADA, S., BERMUDEZ, H., CESARI, I. M., HOEBEKE, J. & NOYA, O. (2003) Immunogenicity of polymerizable synthetic peptides derived from a vaccine candidate against schistosomiasis: the asparaginyl endopeptidase (Sm32). *Immunol Lett,* 88**,** 199-210.
CHAMES, P., HOOGENBOOM, H. R. & HENDERIKX, P. (2002) Selection of antibodies against biotinylated antigens. *Methods Mol Biol,* 178**,** 147-57.
CHAN W.C. & WHITE, P. D. (Eds.) (2000) *Fmoc Solid Phase Peptide Synthesis: A Practical Approach,* New York, Oxford University Press.
CHANG, X., YAMADA, R. & YAMAMOTO, K. (2005) Inhibition of antithrombin by hyaluronic acid may be involved in the pathogenesis of rheumatoid arthritis. *Arthritis Res Ther,* 7**,** R268-73.
CHECK, E. (2007) Genome project turns up evolutionary surprises. *Nature,* 447**,** 760-1.
CHEN, H. J., CARR, K., JEROME, R. E. & EDENBERG, H. J. (2002) A retroviral repetitive element confers tissue-specificity to the human alcohol dehydrogenase 1C (ADH1C) gene. *DNA Cell Biol,* 21**,** 793-801.
CHIU, Y. L. AND GREENE, W. C. (2008) The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol,* **26,** 317-53.
CHOU, P. Y. & FASMAN, G. D. (1974) Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry,* 13**,** 211-22.
CHOY, E. H. & PANAYI, G. S. (2001) Cytokine pathways and joint inflammation in rheumatoid arthritis. *N Engl J Med,* 344**,** 907-16.
CHRISTIAN, C. L. (1998) The discovery of the rheumatoid factor. II. Rose, Ragan, Pearce & Lipman. 1948. *Clin Exp Rheumatol,* 16**,** 345-9.
CHUNG, E. Y., KIM, S. J. & MA, X. J. (2006) Regulation of cytokine production during phagocytosis of apoptotic cells. *Cell Res,* 16**,** 154-61.
CIANCIOLO, G. J., COPELAND, T. D., OROSZLAN, S. & SNYDERMAN, R. (1985) Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science,* 230**,** 453-5.
CLAUSEN, B. E., BRIDGES, S. L., JR., LAVELLE, J. C., FOWLER, P. G., GAY, S., KOOPMAN, W. J. & SCHROEDER, H. W., JR. (1998) Clonally-related immunoglobulin VH domains and nonrandom use of DH gene segments in rheumatoid arthritis synovium. *Mol Med,* 4**,** 240-57.

CLAVERIE, J. & NOTREDAME, C. (2006) *Bioinformatics for Dummies,* London, John Wiley & Sons.
COFFIN J. M., HUGHES S. H & VARMUS, H. E. (Eds.) (1999) *Retroviruses,* Woodbury, Cold Spring Harbor Laboratory Press.
COHEN, M., POWERS, M., O'CONNELL, C. & KATO, N. (1985) The nucleotide sequence of the env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. *Virology,* 147**,** 449-58.
COLE, G. A., HOGG, T. L. & WOODLAND, D. L. (1995) T cell recognition of the immunodominant Sendai virus NP324-332/Kb epitope is focused on the center of the peptide. *J Immunol,* 155**,** 2841-8.
COLMEGNA, I. & GARRY, R. F. (2006) Role of endogenous retroviruses in autoimmune diseases. *Infect Dis Clin North Am,* 20**,** 913-29.
CONRAD, B., WEIDMANN, E., TRUCCO, G., RUDERT, W. A., BEHBOO, R., RICORDI, C., RODRIQUEZ-RILO, H., FINEGOLD, D. & TRUCCO, M. (1994) Evidence for superantigen involvement in insulin-dependent diabetes mellitus aetiology. *Nature,* 371**,** 351-5.
CONRAD, B., WEISSMAHR, R. N., BONI, J., ARCARI, R., SCHUPBACH, J. & MACH, B. (1997) A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell,* 90**,** 303-13.
CONRAD, K. & SCHLOSLER, W. (2002) *Autoantibodies in systemic autoimmune disease,* Berlin, Pabst Science Publishers.
CORDONNIER, A., CASELLA, J. F. & HEIDMANN, T. (1995) Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence. *J Virol,* 69**,** 5890-7.
CORRIGALL, V. M. & PANAYI, G. S. (2002) Autoantigens and immune pathways in rheumatoid arthritis. *Crit Rev Immunol,* 22**,** 281-93.
CREIGHTON, T. E. (1992) Protein folding. Up the kinetic pathway. *Nature,* 356**,** 194-5.
CROWTHER J.R. (Ed.) (2000) *The ELISA Guidebook,* Vienna, Humana Press.
CUFF, J. A. & BARTON, G. J. (1999) Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins,* 34**,** 508-19.
CUTOLO, M., SULLI, A., VILLAGGIO, B., SERIOLO, B. & ACCARDO, S. (1998) Relations between steroid hormones and cytokines in rheumatoid arthritis and systemic lupus erythematosus. *Ann Rheum Dis,* 57**,** 573-7.
CUTOLO, M., VILLAGGIO, B., CRAVIOTTO, C., PIZZORNI, C., SERIOLO, B. & SULLI, A. (2002) Sex hormones and rheumatoid arthritis. *Autoimmun Rev,* 1**,** 284-9.
CYRANOWSKI, J. M., FRANK, E., WINTER, E., RUCCI, P., NOVICK, D., PILKONIS, P., FAGIOLINI, A., SWARTZ, H. A., HOUCK, P. & KUPFER, D. J. (2004) Personality pathology and outcome in recurrently depressed women over 2 years of maintenance interpersonal psychotherapy. *Psychol Med,* 34**,** 659-69.
D'ARPA, P., MACHLIN, P. S., RATRIE, H., 3RD, ROTHFIELD, N. F., CLEVELAND, D. W. & EARNSHAW, W. C. (1988) cDNA cloning of human DNA topoisomerase I: catalytic activity of a 67.7-kDa carboxyl-terminal fragment. *Proc Natl Acad Sci U S A,* 85**,** 2543-7.
DANIS, V. A., FRANIC, G. M., RATHJEN, D. A., LAURENT, R. M. & BROOKS, P. M. (1992) Circulating cytokine levels in patients with rheumatoid arthritis: results of a double blind trial with sulphasalazine. *Ann Rheum Dis,* 51**,** 946-50.

DAVARI EJTEHADI, H., FREIMANIS, G. L., ALI, H. A., BOWMAN, S., ALAVI, A., AXFORD, J., PERERA, S. A., CALLAGHAN, R. & NELSON, P. N. (2005) The potential role of human endogenous retrovirus K10 in pathogenesis of rheumatoid arthritis: a preliminary study. *Ann Rheum Dis*.

DAVIS, L. S., CUSH, J. J., SCHULZE-KOOPS, H. & LIPSKY, P. E. (2001) Rheumatoid synovial CD4+ T cells exhibit a reduced capacity to differentiate into IL-4-producing T-helper-2 effector cells. *Arthritis Res,* 3**,** 54-64.

DAY, R. (2002) Adverse reactions to TNF-alpha inhibitors in rheumatoid arthritis. *Lancet,* 359**,** 540-1.

DE CASTRO, J. M. & BALAGURA, S. (1976) Insulin pretreatment facilitates recovery after dorsal hippocampal lesions. *Physiol Behav,* 16**,** 517-20.

DE PARSEVAL, N., ALKABBANI, H. & HEIDMANN, T. (1999) The long terminal repeats of the HERV-H human endogenous retrovirus contain binding sites for transcriptional regulation by the Myb protein. *J Gen Virol,* 80 (Pt 4)**,** 841-5.

DEAS, J. E., LIU, L. G., THOMPSON, J. J., SANDER, D. M., SOBLE, S. S., GARRY, R. F. & GALLAHER, W. R. (1998) Reactivity of sera from systemic lupus erythematosus and Sjogren's syndrome patients with peptides derived from human immunodeficiency virus p24 capsid antigen. *Clin Diagn Lab Immunol,* 5**,** 181-5.

DECANNIERE, K., DESMYTER, A., LAUWEREYS, M., GHAHROUDI, M. A., MUYLDERMANS, S. & WYNS, L. (1999) A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. *Structure,* 7**,** 361-70.

DEEN, K. C. & SWEET, R. W. (1986) Murine mammary tumor virus pol-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus pol gene. *J Virol,* 57**,** 422-32.

DELANO, W. L., ULTSCH, M. H., DE VOS, A. M. & WELLS, J. A. (2000) Convergent solutions to binding at a protein-protein interface. *Science,* 287**,** 1279-83.

DELVES, P. J. & ROITT, I. M. (2000) The Immune System- First of Two Parts.

DHEDA, K., HUGGETT, J. F., CHANG, J. S., KIM, L. U., BUSTIN, S. A., JOHNSON, M. A., ROOK, G. A. & ZUMLA, A. (2005) The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal Biochem,* 344**,** 141-3.

DIEFFENBACH, C. W., T.M.J. LOWE & G.S.DVEKSLER. (1995) General concepts of primer design. IN C.W.DIEFFENBACH & DVEKSLER, G. S. (Eds.) *PCR Primer: A laboratory manual.* 1 ed., CSHL Press.

DIERKS, S. E., BUTLER, J. E. & RICHERSON, H. B. (1986) Altered recognition of surface-adsorbed compared to antigen-bound antibodies in the ELISA. *Mol Immunol,* 23**,** 403-11.

DIKE, L. E. & FARMER, S. R. (1988) Cell adhesion induces expression of growth-associated genes in suspension-arrested fibroblasts. *Proc Natl Acad Sci U S A,* 85**,** 6792-6.

DINARELLO, C. A. (1996) Biologic basis for interleukin-1 in disease. *Blood,* 87**,** 2095-147.

DITTMER, A., VETTER, M., SCHUNKE, D., SPAN, P. N., SWEEP, F., THOMSSEN, C. & DITTMER, J. (2006) Parathyroid hormone-related protein regulates tumor-relevant genes in breast cancer cells. *J Biol Chem,* 281**,** 14563-72.

DONG, Y. B., CHENG, J. Y., HUANG, R. Q., SMITH, M. D. & ZUR LOYE, H. C. (2003) Self-assembly of coordination polymers from AgX (X = SbF(6)(-), PF(6)(-), and CF(3)SO(3)(-)) and oxadiazole-containing ligands. *Inorg Chem,* 42**,** 5699-706.

DONNES, P. & ELOFSSON, A. (2002) Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics,* 3**,** 25.

DORAN, M. F., POND, G. R., CROWSON, C. S., O'FALLON, W. M. & GABRIEL, S. E. (2002) Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period. *Arthritis Rheum,* 46**,** 625-31.

DORNER, T. (2006) Crossroads of B cell activation in autoimmunity: rationale of targeting B cells. *J Rheumatol Suppl,* 77**,** 3-11.

DOUVAS, A., TAKEHANA, Y., EHRESMANN, G., CHERNYOVSKIY, T. & DAAR, E. S. (1996) Neutralization of HIV type 1 infectivity by serum antibodies from a subset of autoimmune patients with mixed connective tissue disease. *AIDS Res Hum Retroviruses,* 12**,** 1509-17.

DUKE, O., PANAYI, G. S., JANOSSY, G. & POULTER, L. W. (1982) An immunohistological analysis of lymphocyte subpopulations and their microenvironment in the synovial membranes of patients with rheumatoid arthritis using monoclonal antibodies. *Clin Exp Immunol,* 49**,** 22-30.

DUNN, C. A. & MAGER, D. L. (2005) Transcription of the human and rodent SPAM1 / PH-20 genes initiates within an ancient endogenous retrovirus. *BMC Genomics,* 6**,** 47.

DUNN, C. A., MEDSTRAND, P. & MAGER, D. L. (2003) An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon. *Proc Natl Acad Sci U S A,* 100**,** 12841-6.

EDWARDS, J. C., CAMBRIDGE, G. & ABRAHAMS, V. M. (1999) Do self-perpetuating B lymphocytes drive human autoimmune disease? *Immunology,* 97**,** 188-96.

EDWARDS, J. C., CAMBRIDGE, G. & LEANDRO, M. J. (2006) B cell depletion therapy in rheumatic disease. *Best Pract Res Clin Rheumatol,* 20**,** 915-28.

EGGLETON, P., WARD, F. J., JOHNSON, S., KHAMASHTA, M. A., HUGHES, G. R., HAJELA, V. A., MICHALAK, M., CORBETT, E. F., STAINES, N. A. & REID, K. B. (2000) Fine specificity of autoantibodies to calreticulin: epitope mapping and characterization. *Clin Exp Immunol,* 120**,** 384-91.

EHLHARDT, S., SEIFERT, M., SCHNEIDER, J., OJAK, A., ZANG, K. D. & MEHRAEIN, Y. (2006) Human endogenous retrovirus HERV-K(HML-2) Rec expression and transcriptional activities in normal and rheumatoid arthritis synovia. *J Rheumatol,* 33**,** 16-23.

EHRENSTEIN, M. R., EVANS, J. G., SINGH, A., MOORE, S., WARNES, G., ISENBERG, D. A. & MAURI, C. (2004) Compromised function of regulatory T cells in rheumatoid arthritis and reversal by anti-TNFalpha therapy. *J Exp Med,* 200**,** 277-85.

EJTEHADI, H. D., FREIMANIS, G. L., ALI, H. A., BOWMAN, S., ALAVI, A., AXFORD, J., CALLAGHAN, R. & NELSON, P. N. (2006) The potential role of human endogenous retrovirus K10 in the pathogenesis of rheumatoid arthritis: a preliminary study. *Ann Rheum Dis,* 65**,** 612-6.

ELKON, K., SKELLY, S., PARNASSA, A., MOLLER, W., DANHO, W., WEISSBACH, H. & BROT, N. (1986) Identification and chemical synthesis of a ribosomal protein antigenic determinant in systemic lupus erythematosus. *Proc Natl Acad Sci U S A,* 83**,** 7419-23.

EMINI, E. A., HUGHES, J. V., PERLOW, D. S. & BOGER, J. (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol,* 55**,** 836-9.
ENCODE PROJECT CONSORTIUM (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447**,** 799-816.
ENGVALL, E. & PERLMANN, P. (1972) Enzyme-linked immunosorbent assay, Elisa. 3. Quantitation of specific antibodies by enzyme-labeled anti-immunoglobulin in antigen-coated tubes. *J Immunol,* 109**,** 129-35.
EVANS, C. F., HORWITZ, M. S., HOBBS, M. V. & OLDSTONE, M. B. (1996) Viral infection of transgenic mice expressing a viral protein in oligodendrocytes leads to chronic central nervous system autoimmune disease. *J Exp Med,* 184**,** 2371-84.
FARMERIE, W. G., LOEB, D. D., CASAVANT, N. C., HUTCHISON, C. A., 3RD, EDGELL, M. H. AND SWANSTROM, R. (1987) Expression and processing of the AIDS virus reverse transcriptase in Escherichia coli. *Science,* **236,** 305-8.
FASSBENDER, H. G. (1983) Histomorphological basis of articular cartilage destruction in rheumatoid arthritis. *Coll Relat Res,* 3**,** 141-55.
FELDMANN, M., BRENNAN, F. M. & MAINI, R. N. (1996) Rheumatoid arthritis. *Cell,* 85**,** 307-10.
FELSON, D. T., ANDERSON, J. J., BOERS, M., BOMBARDIER, C., FURST, D., GOLDSMITH, C., KATZ, L. M., LIGHTFOOT, R., JR., PAULUS, H., STRAND, V. & ET AL. (1995) American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum,* 38**,** 727-35.
FIELDS, G. B. & NOBLE, R. L. (1990) Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int J Pept Protein Res,* 35**,** 161-214.
FIRESTEIN, G. S. (1999) Starving the synovium: angiogenesis and inflammation in rheumatoid arthritis. *J Clin Invest,* 103**,** 3-4.
FIRESTEIN, G. S. (2003) Evolving concepts of rheumatoid arthritis. *Nature,* 423**,** 356-61.
FIRESTEIN, G. S. & ZVAIFLER, N. J. (2002) How important are T cells in chronic rheumatoid synovitis?: II. T cell-independent mechanisms from beginning to end. *Arthritis Rheum,* 46**,** 298-308.
FISCHER, P. M. & HOWDEN, M. E. (1990) Direct enzyme-linked immunosorbent assay of anti-peptide antibodies using capture of biotinylated peptides by immobilized avidin. *J Immunoassay,* 11**,** 311-27.
FLEISCHER, B. (1991) Stimulation of human T cells by microbial 'superantigens'. *Immunol Res,* 10**,** 349-55.
FLOWER, D. R. (2003) Towards *in silico* prediction of immunogenic epitopes. *Trends Immunol,* 24**,** 667-74.
FLUGEL, R. M., RETHWILM, A., MAURER, B. & DARAI, G. (1987) Nucleotide sequence analysis of the env gene and its flanking regions of the human spumaretrovirus reveals two novel genes. *Embo J,* 6**,** 2077-84.
FOURNEL, S. & MULLER, S. (2003) Synthetic peptides in the diagnosis of systemic autoimmune diseases. *Curr Protein Pept Sci,* 4**,** 261-74.
FRANKLIN, G. C., CHRETIEN, S., HANSON, I. M., ROCHEFORT, H., MAY, F. E. & WESTLEY, B. R. (1988) Expression of human sequences related to those of mouse mammary tumor virus. *J Virol,* 62**,** 1203-10.

FRITSCH, R., ESELBOCK, D., SKRINER, K., JAHN-SCHMID, B., SCHEINECKER, C., BOHLE, B., TOHIDAST-AKRAD, M., HAYER, S., NEUMULLER, J., PINOL-ROMA, S., SMOLEN, J. S. & STEINER, G. (2002) Characterization of autoreactive T cells to the autoantigens heterogeneous nuclear ribonucleoprotein A2 (RA33) and filaggrin in patients with rheumatoid arthritis. *J Immunol,* 169**,** 1068-76.

FU, W., GORELICK, R. J. & REIN, A. (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J Virol,* 68**,** 5013-8.

FU, Y. X., HUANG, G., WANG, Y. & CHAPLIN, D. D. (1998) B lymphocytes induce the formation of follicular dendritic cell clusters in a lymphotoxin alpha-dependent fashion. *J Exp Med,* 187**,** 1009-18.

FUJIWARA, R., KUTSUMI, Y., HAYASHI, T., NISHIO, H., KOSHINO, Y., SHIMADA, Y., NAKAI, T. & MIYABO, S. (1995) Relation of angiographically defined coronary artery disease and plasma concentrations of insulin, lipid, and apolipoprotein in normolipidemic subjects with varying degrees of glucose tolerance. *Am J Cardiol,* 75**,** 122-6.

GALBRAITH, D. N., KELLY, H. T., DYKE, A., REID, G., HAWORTH, C., BEEKMAN, J., SHEPHERD, A. & SMITH, K. T. (2000) Design and validation of immunological tests for the detection of Porcine endogenous retrovirus in biological materials. *J Virol Methods,* 90**,** 115-24.

GALLI, U. M., SAUTER, M., LECHER, B., MAURER, S., HERBST, H., ROEMER, K. & MUELLER-LANTZSCH, N. (2005) Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors. *Oncogene,* 24**,** 3223-8.

GARCIA, K. C., RONCO, P. M., VERROUST, P. J., BRUNGER, A. T. & AMZEL, L. M. (1992) Three-dimensional structure of an angiotensin II-Fab complex at 3 A: hormone recognition by an anti-idiotypic antibody. *Science,* 257**,** 502-7.

GAUDIN, P., IJAZ, S., TUKE, P. W., MARCEL, F., PARAZ, A., SEIGNEURIN, J. M., MANDRAND, B., PERRON, H. & GARSON, J. A. (2000) Infrequency of detection of particle-associated MSRV/HERV-W RNA in the synovial fluid of patients with rheumatoid arthritis. *Rheumatology (Oxford),* 39**,** 950-4.

GEILER, T., KRIEGSMANN, J., KEYSZER, G. M., GAY, R. E. & GAY, S. (1994) A new model for rheumatoid arthritis generated by engraftment of rheumatoid synovial tissue and normal human cartilage into SCID mice. *Arthritis Rheum,* 37**,** 1664-71.

GENDI, N. S., GIBSON, K. & WORDSWORTH, B. P. (1996) Effect of HLA type and hypocomplementaemia on the expression of parvovirus arthritis: one year follow up of an outbreak. *Ann Rheum Dis,* 55**,** 63-5.

GENOVESE, M. C., BECKER, J. C., SCHIFF, M., LUGGEN, M., SHERRER, Y., KREMER, J., BIRBARA, C., BOX, J., NATARAJAN, K., NUAMAH, I., LI, T., ARANDA, R., HAGERTY, D. T. & DOUGADOS, M. (2005) Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition. *N Engl J Med,* 353**,** 1114-23.

GIBBS, W. W. (2003) The unseen genome: gems among the junk. *Sci Am,* 289**,** 26-33.

GIBELLINI, D., VITONE, F., SCHIAVONE, P., PONTI, C., LA PLACA, M. & RE, M. C. (2004) Quantitative detection of human immunodeficiency virus type 1 (HIV-1) proviral DNA in peripheral blood mononuclear cells by SYBR green real-time PCR technique. *J Clin Virol,* 29**,** 282-9.

GIFFORD, R., KABAT, P., MARTIN, J., LYNCH, C. & TRISTEM, M. (2005) Evolution and distribution of class II-related endogenous retroviruses. *J Virol,* 79**,** 6478-86.
GIFFORD, R. & TRISTEM, M. (2003) The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes,* 26**,** 291-315.
GINZINGER, D. G. (2002) Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol,* 30**,** 503-12.
GIRARD, A., SACHIDANANDAM, R., HANNON, G. J. & CARMELL, M. A. (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature,* 442**,** 199-202.
GOBERDHAN, D. C. & WILSON, C. (2003) PTEN: tumour suppressor, multifunctional growth regulator and more. *Hum Mol Genet,* 12 Spec No 2**,** R239-48.
GOLDBACH-MANSKY, R., LEE, J., MCCOY, A., HOXWORTH, J., YARBORO, C., SMOLEN, J. S., STEINER, G., ROSEN, A., ZHANG, C., MENARD, H. A., ZHOU, Z. J., PALOSUO, T., VAN VENROOIJ, W. J., WILDER, R. L., KLIPPEL, J. H., SCHUMACHER, H. R., JR. & EL-GABALAWY, H. S. (2000) Rheumatoid arthritis associated autoantibodies in patients with synovitis of recent onset. *Arthritis Res,* 2**,** 236-43.
GOLDBLATT, F. & ISENBERG, D. A. (2005) New therapies for rheumatoid arthritis. *Clin Exp Immunol,* 140**,** 195-204.
GOLDSMITH, M. A., WARMERDAM, M. T., ATCHISON, R. E., MILLER, M. D., AND GREENE, W. C. (1995). Dissociation of the CD4 downregulation and viral infectivity enhancement functions of human immunodeficiency virus type 1 Nef. *J Virol, 69*, 4112-4121.
GONZALEZ-BUITRAGO, J. M. & GONZALEZ, C. (2006) Present and future of the autoimmunity laboratory. *Clin Chim Acta,* 365**,** 50-7.
GORONZY, J. J., FUJII, H. & WEYAND, C. M. (2006) Telomeres, immune aging and autoimmunity. *Exp Gerontol,* 41**,** 246-51.
GORONZY, J. J., MATTESON, E. L., FULBRIGHT, J. W., WARRINGTON, K. J., CHANG-MILLER, A., HUNDER, G. G., MASON, T. G., NELSON, A. M., VALENTE, R. M., CROWSON, C. S., ERLICH, H. A., REYNOLDS, R. L., SWEE, R. G., O'FALLON, W. M. & WEYAND, C. M. (2004) Prognostic markers of radiographic progression in early rheumatoid arthritis. *Arthritis Rheum,* 50**,** 43-54.
GORONZY, J. J. & WEYAND, C. M. (2005) Rheumatoid arthritis. *Immunol Rev,* 204**,** 55-73.
GORONZY, J. J., ZETTL, A. & WEYAND, C. M. (1998) T cell receptor repertoire in rheumatoid arthritis. *Int Rev Immunol,* 17**,** 339-63.
GOVINDARAJAN, K. R., KANGUEANE, P., TAN, T. W. & RANGANATHAN, S. (2003) MPID: MHC-Peptide Interaction Database for sequence-structure-function information on peptides binding to MHC molecules. *Bioinformatics,* 19**,** 309-10.
GRANTHAM, R. (1974) Amino acid difference formula to help explain protein evolution. *Science,* 185**,** 862-4.
GRASSI, W., DE ANGELIS, R., LAMANNA, G. & CERVINI, C. (1998) The clinical features of rheumatoid arthritis. *Eur J Radiol,* 27 Suppl 1**,** S18-24.
GRAVALLESE, E. M. (2002) Bone destruction in arthritis. *Ann Rheum Dis,* 61 Suppl 2**,** ii84-6.

GREALLY, J. M. (2007) Genomics: Encyclopaedia of humble DNA. *Nature,* 447**,** 782-3.
GREENBAUM, J. A., ANDERSEN, P. H., BLYTHE, M., BUI, H. H., CACHAU, R. E., CROWE, J., DAVIES, M., KOLASKAR, A. S., LUND, O., MORRISON, S., MUMEY, B., OFRAN, Y., PELLEQUER, J. L., PINILLA, C., PONOMARENKO, J. V., RAGHAVA, G. P., VAN REGENMORTEL, M. H., ROGGEN, E. L., SETTE, A., SCHLESSINGER, A., SOLLNER, J., ZAND, M. & PETERS, B. (2007) Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J Mol Recognit,* 20**,** 75-82.
GRIESMACHER, A. & PEICHL, P. (2001) Autoantibodies associated with rheumatic diseases. *Clin Chem Lab Med,* 39**,** 189-208.
GRIFFITHS, D. J., COOKE, S. P., HERVE, C., RIGBY, S. P., MALLON, E., HAJEER, A., LOCK, M., EMERY, V., TAYLOR, P., PANTELIDIS, P., BUNKER, C. B., DU BOIS, R., WEISS, R. A. & VENABLES, P. J. (1999) Detection of human retrovirus 5 in patients with arthritis and systemic lupus erythematosus. *Arthritis Rheum,* 42**,** 448-54.
GUNSON, R., GILLESPIE, G. & W, F. C. (2003) Optimisation of PCR reactions using primer chessboarding. *J Clin Virol,* 26**,** 369-73.
HAIN, N. A., STUHLMULLER, B., HAHN, G. R., KALDEN, J. R., DEUTZMANN, R. & BURMESTER, G. R. (1996) Biochemical characterization and microsequencing of a 205-kDa synovial protein stimulatory for T cells and reactive with rheumatoid factor containing sera. *J Immunol,* 157**,** 1773-80.
HAMMER, J., STURNIOLO, T. & SINIGAGLIA, F. (1997) HLA class II peptide binding specificity and autoimmunity. *Adv Immunol,* 66, 67-100.
HANAOKA, R., KASAMA, T., MURAMATSU, M., YAJIMA, N., SHIOZAWA, F., MIWA, Y., NEGISHI, M., IDE, H., MIYAOKA, H., UCHIDA, H. & ADACHI, M. (2003) A novel mechanism for the regulation of IFN-gamma inducible protein-10 expression in rheumatoid arthritis. *Arthritis Res Ther,* 5**,** R74-81.
HANCOCK, W. S. & BATTERSBY, J. E. (1976) A new micro-test for the detection of incomplete coupling reactions in solid-phase peptide synthesis using 2,4,6-trinitrobenzenesulphonic acid. *Anal Biochem,* 71**,** 260-4.
HARAGUCHI, S., GOOD, R. A. & DAY, N. K. (1995) Immunosuppressive retroviral peptides: cAMP and cytokine patterns. *Immunol Today,* 16**,** 595-603.
HART, H., MCCORMICK, J. N. & MARMION, B. P. (1979) Viruses and lymphocytes in rheumatoid arthritis. II. Examination of lymphocytes and sera from patients with rheumatoid arthritis for evidence of retrovirus infection. *Ann Rheum Dis,* 38**,** 514-25.
HATTORI, T., TAKAHASH, K., YUTANI, Y., FUJISAWA, T., NAKANISHI, T. & TAKIGAWA, M. (2000) Rheumatoid arthritis-related antigen 47kDa (RA-A47) is a product of colligin-2 and acts as a human HSP47. *J Bone Miner Metab,* 18**,** 328-34.
HAUPT, W., FISCHER, T. C., WINDERL, S., FRANSZ, P. & TORRES-RUIZ, R. A. (2001) The centromere1 (CEN1) region of *Arabidopsis thaliana*: architecture and functional impact of chromatin. *Plant J,* 27**,** 285-96.
HE, L., THOMSON, J. M., HEMANN, M. T., HERNANDO-MONGE, E., MU, D., GOODSON, S., POWERS, S., CORDON-CARDO, C., LOWE, S. W., HANNON, G. J. & HAMMOND, S. M. (2005) A microRNA polycistron as a potential human oncogene. *Nature,* 435**,** 828-33.
HEID, C. A., STEVENS, J., LIVAK, K. J. & WILLIAMS, P. M. (1996) Real time quantitative PCR. *Genome Res,* 6**,** 986-94.

HERNANDEZ AVILA, M., LIANG, M. H., WILLETT, W. C., STAMPFER, M. J., COLDITZ, G. A., ROSNER, B., ROBERTS, W. N., HENNEKENS, C. H. & SPEIZER, F. E. (1990) Reproductive factors, smoking, and the risk for rheumatoid arthritis. *Epidemiology,* 1**,** 285-91.

HERNIOU, E., MARTIN, J., MILLER, K., COOK, J., WILKINSON, M. & TRISTEM, M. (1998) Retroviral diversity and distribution in vertebrates. *J Virol,* 72**,** 5955-66.

HERVE, C. A., LUGLI, E. B., BRAND, A., GRIFFITHS, D. J. & VENABLES, P. J. (2002) Autoantibodies to human endogenous retrovirus-K are frequently detected in health and disease and react with multiple epitopes. *Clin Exp Immunol,* 128**,** 75-82.

HERVE, C. A., WAIT, R. & VENABLES, P. J. (2003) Glucose-6-phosphate isomerase is not a specific autoantigen in rheumatoid arthritis. *Rheumatology (Oxford),* 42**,** 986-8.

HIDA, S., MIURA, N. N., ADACHI, Y. & OHNO, N. (2007) Cell wall beta-glucan derived from Candida albicans acts as a trigger for autoimmune arthritis in SKG mice. *Biol Pharm Bull,* 30**,** 1589-92.

HISCOX, J. A. (2002) The nucleolus--a gateway to viral infection? *Arch Virol,* 147**,** 1077-89.

HOLLSTEIN, M., SIDRANSKY, D., VOGELSTEIN, B. AND HARRIS, C. C. (1991) p53 mutations in human cancers. *Science,* **253,** 49-53.

HOLMES, R. K., MALIM, M. H. & BISHOP, K. N. (2007) APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci,* 32**,** 118-28.

HOPP, T. P. (1993) Retrospective: 12 years of antigenic determinant predictions, and more. *Pept Res,* 6**,** 183-90.

HOPP, T. P. (1994) Different views of protein antigenicity. *Pept Res,* 7**,** 229-31.

HOPP, T. P. & WOODS, K. R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U S A,* 78**,** 3824-8.

HORAI, R., SAIJO, S., TANIOKA, H., NAKAE, S., SUDO, K., OKAHARA, A., IKUSE, T., ASANO, M. & IWAKURA, Y. (2000) Development of chronic inflammatory arthropathy resembling rheumatoid arthritis in interleukin 1 receptor antagonist-deficient mice. *J Exp Med,* 191**,** 313-20.

HU, P., WU, S. & HERNANDEZ, N. (2004) A role for beta-actin in RNA polymerase III transcription. *Genes Dev,* 18**,** 3010-5.

HUBER, B. T., HSU, P. N. & SUTKOWSKI, N. (1996) Virus-encoded superantigens. *Microbiol Rev,* 60**,** 473-82.

HUBER, L. C., DISTLER, O., TARNER, I., GAY, R. E., GAY, S. & PAP, T. (2006) Synovial fibroblasts: key players in rheumatoid arthritis. *Rheumatology (Oxford),* 45**,** 669-75.

HUGGETT, J., DHEDA, K., BUSTIN, S. & ZUMLA, A. (2005) Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun,* 6**,** 279-84.

HYRICH, K. L. & INMAN, R. D. (2001) Infectious agents in chronic rheumatic diseases. *Curr Opin Rheumatol,* 13**,** 300-4.

IANNONE, F., CORRIGAL, V. M. & PANAYI, G. S. (1996) CD69 on synovial T cells in rheumatoid arthritis correlates with disease activity. *Br J Rheumatol,* 35**,** 397.

IKEUCHI, H., KUROIWA, T., HIRAMATSU, N., KANEKO, Y., HIROMURA, K., UEKI, K. & NOJIMA, Y. (2005) Expression of interleukin-22 in rheumatoid arthritis: potential role as a proinflammatory cytokine. *Arthritis Rheum,* 52**,** 1037-46.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM (2004) Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-45.
ISHIGURO, T., SAITOH, J., YAWATA, H., YAMAGISHI, H., IWASAKI, S. & MITOMA, Y. (1995) Homogeneous quantitative assay of hepatitis C virus RNA by polymerase chain reaction in the presence of a fluorescent intercalater. *Anal Biochem,* 229**,** 207-13.
ISHIKAWA, H., HIRATA, S., ANDOH, Y., KUBO, H., NAKAGAWA, N., NISHIBAYASHI, Y. & MIZUNO, K. (1996) An immunohistochemical and immunoelectron microscopic study of adhesion molecules in synovial pannus formation in rheumatoid arthritis. *Rheumatol Int,* 16**,** 53-60.
ITOH, Y. & REICHLIN, M. (1992) Autoantibodies to the Ro/SSA antigen are conformation dependent. I: Anti-60 kD antibodies are mainly directed to the native protein; anti-52 kD antibodies are mainly directed to the denatured protein. *Autoimmunity,* 14**,** 57-65.
IVANOV, V. S., SUVOROVA, Z. K., TCHIKIN, L. D., KOZHICH, A. T. & IVANOV, V. T. (1992) Effective method for synthetic peptide immobilization that increases the sensitivity and specificity of ELISA procedures. *J Immunol Methods,* 153**,** 229-33.
IVARS, F. (2007) Superantigen-induced regulatory T cells in vivo. *Chem Immunol Allergy,* 93**,** 137-60.
IWAKURA, Y., SAIJO, S., KIOKA, Y., NAKAYAMA-YAMADA, J., ITAGAKI, K., TOSU, M., ASANO, M., KANAI, Y. & KAKIMOTO, K. (1995) Autoimmunity induction by human T cell leukemia virus type 1 in transgenic mice that develop chronic inflammatory arthropathy resembling rheumatoid arthritis in humans. *J Immunol,* 155**,** 1588-98.
IWAKURA, Y., TOSU, M., YOSHIDA, E., TAKIGUCHI, M., SATO, K., KITAJIMA, I., NISHIOKA, K., YAMAMOTO, K., TAKEDA, T., HATANAKA, M. & ET AL. (1991) Induction of inflammatory arthropathy resembling rheumatoid arthritis in mice transgenic for HTLV-I. *Science,* 253**,** 1026-8.
JACKS, T., POWER, M. D., MASIARZ, F. R., LUCIW, P. A., BARR, P. J. AND VARMUS, H. E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression *Nature,* **331,** 280-3.
JAMES, J. A., GROSS, T., SCOFIELD, R. H. & HARLEY, J. B. (1995) Immunoglobulin epitope spreading and autoimmune disease after peptide immunization: Sm B/B'-derived PPPGMRPP and PPPGIRGP induce spliceosome autoimmunity. *J Exp Med,* 181**,** 453-61.
JAMES, J. A. & HARLEY, J. B. (1992) Linear epitope mapping of an Sm B/B' polypeptide. *J Immunol,* 148**,** 2074-9.
JAMES, J. A. & HARLEY, J. B. (1998) B-cell epitope spreading in autoimmunity. *Immunol Rev,* 164**,** 185-200.
JAMES, W. H. (1993) Rheumatoid arthritis, the contraceptive pill, and androgens. *Ann Rheum Dis,* 52**,** 470-4.
JANIN, J. (1979) Surface and inside volumes in globular proteins. *Nature,* 277**,** 491-2.
JENSEN, S., GASSAMA, M. P. & HEIDMANN, T. (1999) Taming of transposable elements by homology-dependent gene silencing. *Nat Genet,* 21**,** 209-12.
JONES, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol,* 292**,** 195-202.
JONES, S. & THORNTON, J. M. (1996) Principles of protein-protein interactions. *Proc Natl Acad Sci U S A,* 93**,** 13-20.

JORDAN, I. K., ROGOZIN, I. B., GLAZKO, G. V. & KOONIN, E. V. (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet,* 19**,** 68-72.

JUNAKOVIC, N., TERRINONI, A., DI FRANCO, C., VIEIRA, C. & LOEVENBRUCK, C. (1998) Accumulation of transposable elements in the heterochromatin and on the Y chromosome of Drosophila simulans and Drosophila melanogaster. *J Mol Evol,* 46**,** 661-8.

JUNGEL, A., DISTLER, J. H., KUROWSKA-STOLARSKA, M., SEEMAYER, C. A., SEIBL, R., FORSTER, A., MICHEL, B. A., GAY, R. E., EMMRICH, F., GAY, S. & DISTLER, O. (2004) Expression of interleukin-21 receptor, but not interleukin-21, in synovial fibroblasts and synovial macrophages of patients with rheumatoid arthritis. *Arthritis Rheum,* 50**,** 1468-76.

KAMBHU, S., FALLDORF, P. & LEE, J. S. (1990) Endogenous retroviral long terminal repeats within the HLA-DQ locus. *Proc Natl Acad Sci U S A,* 87**,** 4927-31.

KAPLAN, M. (2002) Eculizumab (Alexion). *Curr Opin Investig Drugs,* 3**,** 1017-23.

KARPLUS, P. A. & SCHULZ, G. E. (1985) Prediction of Chain Flexibility in Proteins - a Tool for the Selection of Peptide Antigens. *Naturwissenschaften,* 72**,** 212-213.

KATZOURAKIS, A., TRISTEM, M., PYBUS, O. G. & GIFFORD, R. J. (2007) Discovery and analysis of the first endogenous lentivirus. *Proc Natl Acad Sci U S A,* 104**,** 6261-5.

KAUR, H. & RAGHAVA, G. P. (2003) A neural-network based method for prediction of gamma-turns in proteins from multiple sequence alignment. *Protein Sci,* 12**,** 923-9.

KAUR, H. & RAGHAVA, G. P. (2004) A neural network method for prediction of beta-turn types in proteins using evolutionary information. *Bioinformatics,* 20**,** 2751-8.

KE, L. D., CHEN, Z. & YUNG, W. K. (2000) A reliability test of standard-based quantitative PCR: exogenous vs endogenous standards. *Mol Cell Probes,* 14**,** 127-35.

KEFFER, J., PROBERT, L., CAZLARIS, H., GEORGOPOULOS, S., KASLARIS, E., KIOUSSIS, D. & KOLLIAS, G. (1991) Transgenic mice expressing human tumour necrosis factor: a predictive genetic model of arthritis. *Embo J,* 10**,** 4025-31.

KEYSTONE, E. (2002) Treatments no longer in development for rheumatoid arthritis. *Ann Rheum Dis,* 61 Suppl 2**,** ii43-5.

KIM, H. J. & BEREK, C. (2000) B cells in rheumatoid arthritis. *Arthritis Res,* 2**,** 126-31.

KINOMOTO, M., KANNO, T., SHIMURA, M., ISHIZAKA, Y., KOJIMA, A., KURATA, T., SATA, T. & TOKUNAGA, K. (2007) All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res,* 35**,** 2955-64.

KISLAUSKIS, E. H., ZHU, X. & SINGER, R. H. (1997) beta-Actin messenger RNA localization and protein synthesis augment cell motility. *J Cell Biol,* 136**,** 1263-70.

KJELLMAN, C., SJOGREN, H. O. & WIDEGREN, B. (1995) The Y chromosome: a graveyard for endogenous retroviruses. *Gene,* 161**,** 163-70.

KJELLMAN, C., SJOGREN, H. O. & WIDEGREN, B. (1999) HERV-F, a new group of human endogenous retrovirus sequences. *J Gen Virol,* 80 (Pt 9)**,** 2383-92.

KLAVINSKIS, L. S., WHITTON, J. L. & OLDSTONE, M. B. (1989) Molecularly engineered vaccine which expresses an immunodominant T-cell epitope induces cytotoxic T lymphocytes that confer protection from lethal virus infection. *J Virol,* 63**,** 4311-6.

KLIMIUK, P. A., YANG, H., GORONZY, J. J. & WEYAND, C. M. (1999) Production of cytokines and metalloproteinases in rheumatoid synovitis is T cell dependent. *Clin Immunol,* 90**,** 65-78.

KLIPPEL K. H. & DIEPPE, P. A. (Eds.) (1997) *Rheumatology,* St. Louis ; London : Mosby, Mosby.

KNOSSL, M., LOWER, R. & LOWER, J. (1999) Expression of the human endogenous retrovirus HTDV/HERV-K is enhanced by cellular transcription factor YY1. *J Virol,* 73**,** 1254-61.

KOCHWA, S., BROWNELL, M., ROSENFIELD, R. E. & WASSERMAN, L. R. (1967) Adsorption of proteins by polystyrene particles. I. Molecular unfolding and acquired immunogenicity of IgG. *J Immunol,* 99**,** 981-6.

KOHSAKA, H., YAMAMOTO, K., FUJII, H., MIURA, H., MIYASAKA, N., NISHIOKA, K. & MIYAMOTO, T. (1990) Fine epitope mapping of the human SS-B/La protein. Identification of a distinct autoepitope homologous to a viral gag polyprotein. *J Clin Invest,* 85**,** 1566-74.

KOLASKAR, A. S. & TONGAONKAR, P. C. (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett,* 276**,** 172-4.

KORTHAUER, U., GRAF, D., MAGES, H. W., BRIERE, F., PADAYACHEE, M., MALCOLM, S., UGAZIO, A. G., NOTARANGELO, L. D., LEVINSKY, R. J. & KROCZEK, R. A. (1993) Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM. *Nature,* 361**,** 539-41.

KOSHLAND, D. E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A,* 44**,** 98-104.

KOUSKOFF, V., KORGANOW, A. S., DUCHATELLE, V., DEGOTT, C., BENOIST, C. & MATHIS, D. (1996) Organ-specific disease provoked by systemic autoimmunity. *Cell,* 87**,** 811-22.

KRANE, D. E. & RAYMER., M. L. (2002) *Fundamental Concepts of Bioinformatics* New York, Benjamin Cummings.

KREMER, J. M., WESTHOVENS, R., LEON, M., DI GIORGIO, E., ALTEN, R., STEINFELD, S., RUSSELL, A., DOUGADOS, M., EMERY, P., NUAMAH, I. F., WILLIAMS, G. R., BECKER, J. C., HAGERTY, D. T. & MORELAND, L. W. (2003) Treatment of rheumatoid arthritis by selective inhibition of T-cell activation with fusion protein CTLA4Ig. *N Engl J Med,* 349**,** 1907-15.

KRIEG, A. M., GOURLEY, M. F., KLINMAN, D. M., PERL, A. & STEINBERG, A. D. (1992) Heterogeneous expression and coordinate regulation of endogenous retroviral sequences in human peripheral blood mononuclear cells. *AIDS Res Hum Retroviruses,* 8**,** 1991-8.

KROESEN, S., WIDMER, A. F., TYNDALL, A. & HASLER, P. (2003) Serious bacterial infections in patients with rheumatoid arthritis under anti-TNF-alpha therapy. *Rheumatology (Oxford),* 42**,** 617-21.

KROOT, E. J., DE JONG, B. A., VAN LEEUWEN, M. A., SWINKELS, H., VAN DEN HOOGEN, F. H., VAN'T HOF, M., VAN DE PUTTE, L. B., VAN RIJSWIJK, M. H., VAN VENROOIJ, W. J. & VAN RIEL, P. L. (2000) The prognostic value of anti-cyclic citrullinated peptide antibody in patients with recent-onset rheumatoid arthritis. *Arthritis Rheum,* 43**,** 1831-5.

KULSKI, J. K., GAUDIERI, S., INOKO, H. & DAWKINS, R. L. (1999) Comparison between two human endogenous retrovirus (HERV)-rich regions within the major histocompatibility complex. *J Mol Evol,* 48**,** 675-83.

KUWANA, M., KABURAKI, J., MIMORI, T., TOJO, T. & HOMMA, M. (1993) Autoantigenic epitopes on DNA topoisomerase I. Clinical and immunogenetic associations in systemic sclerosis. *Arthritis Rheum,* 36**,** 1406-13.

LA MANTIA, G., MAGLIONE, D., PENGUE, G., DI CRISTOFANO, A., SIMEONE, A., LANFRANCONE, L. & LANIA, L. (1991) Identification and characterization of novel human endogenous retroviral sequences prefentially expressed in undifferentiated embryonal carcinoma cells. *Nucleic Acids Res,* 19**,** 1513-20.

LA MANTIA, G., MAJELLO, B., DI CRISTOFANO, A., STRAZZULLO, M., MINCHIOTTI, G. & LANIA, L. (1992) Identification of regulatory elements within the minimal promoter region of the human endogenous ERV9 proviruses: accurate transcription initiation is controlled by an Inr-like element. *Nucleic Acids Res,* 20**,** 4129-36.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. (2001) Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LANDRY, J. R., ROUHI, A., MEDSTRAND, P. & MAGER, D. L. (2002) The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter. *Mol Biol Evol,* 19**,** 1934-42.

LARSSON, E. & ANDERSSON, G. (1998) Beneficial role of human endogenous retroviruses: facts and hypotheses. *Scand J Immunol,* 48**,** 329-38.

LAU, C. M., BROUGHTON, C., TABOR, A. S., AKIRA, S., FLAVELL, R. A., MAMULA, M. J., CHRISTENSEN, S. R., SHLOMCHIK, M. J., VIGLIANTI, G. A., RIFKIN, I. R. & MARSHAK-ROTHSTEIN, A. (2005) RNA-associated autoantigens activate B cells by combined B cell antigen receptor/Toll-like receptor 7 engagement. *J Exp Med,* 202**,** 1171-7.

LAU, N. C., SETO, A. G., KIM, J., KURAMOCHI-MIYAGAWA, S., NAKANO, T., BARTEL, D. P. & KINGSTON, R. E. (2006) Characterization of the piRNA complex from rat testes. *Science,* 313**,** 363-7.

LEADBETTER, E. A., RIFKIN, I. R., HOHLBAUM, A. M., BEAUDETTE, B. C., SHLOMCHIK, M. J. & MARSHAK-ROTHSTEIN, A. (2002) Chromatin-IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. *Nature,* 416**,** 603-7.

LEANDRO, M. J., EDWARDS, J. C. & CAMBRIDGE, G. (2002) Clinical outcome in 22 patients with rheumatoid arthritis treated with B lymphocyte depletion. *Ann Rheum Dis,* 61**,** 883-8.

LECELLIER, C. H., DUNOYER, P., ARAR, K., LEHMANN-CHE, J., EYQUEM, S., HIMBER, C., SAIB, A. & VOINNET, O. (2005) A cellular microRNA mediates antiviral defense in human cells. *Science,* 308**,** 557-60.

LEE, B. & RICHARDS, F. M. (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol,* 55**,** 379-400.

LEE, D. M., FRIEND, D. S., GURISH, M. F., BENOIST, C., MATHIS, D. & BRENNER, M. B. (2002) Mast cells: a cellular link between autoantibodies and inflammatory arthritis. *Science,* 297**,** 1689-92.

LEE, L. G., CONNELL, C. R. & BLOCH, W. (1993) Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res,* 21**,** 3761-6.

LEE, W. J., KWUN, H. J. & JANG, K. L. (2003) Analysis of transcriptional regulatory sequences in the human endogenous retrovirus W long terminal repeat. *J Gen Virol,* 84**,** 2229-35.

LEFKOWITH, J. B., KIEHL, M., RUBENSTEIN, J., DIVALERIO, R., BERNSTEIN, K., KAHL, L., RUBIN, R. L. & GOURLEY, M. (1996) Heterogeneity and clinical significance of glomerular-binding antibodies in systemic lupus erythematosus. *J Clin Invest,* 98**,** 1373-80.

LEHMANN, P. V., FORSTHUBER, T., MILLER, A. & SERCARZ, E. E. (1992) Spreading of T-cell autoimmunity to cryptic determinants of an autoantigen. *Nature,* 358**,** 155-7.

LEVELY, M. E., BANNOW, C. A., SMITH, C. W. & NICHOLAS, J. A. (1991) Immunodominant T-cell epitope on the F protein of respiratory syncytial virus recognized by human lymphocytes. *J Virol,* 65**,** 3789-96.

LEVITT, M. (1976) A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol,* 104**,** 59-107.

LEWIS, S. M. & WU, G. E. (1997) The origins of V(D)J recombination. *Cell,* 88**,** 159-62.

LIAO, D., PAVELITZ, T., KIDD, J. R., KIDD, K. K. & WEINER, A. M. (1997) Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *Embo J,* 16**,** 588-98.

LIN, L. & PENG, S. L. (2005) Interleukin-18 receptor signaling is not required for autoantibody production and end-organ disease in murine lupus. *Arthritis Rheum,* 52**,** 984-6.

LINDESKOG, M., MAGER, D. L. & BLOMBERG, J. (1999) Isolation of a human endogenous retroviral HERV-H element with an open env reading frame. *Virology,* 258**,** 441-50.

LINDESKOG, M., MEDSTRAND, P., CUNNINGHAM, A. A. & BLOMBERG, J. (1998) Coamplification and dispersion of adjacent human endogenous retroviral

HERV-H and HERV-E elements; presence of spliced hybrid transcripts in normal leukocytes. *Virology,* 244**,** 219-29.
LIPSKY, P. E. (2001) Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. *Nat Immunol,* 2**,** 764-6.
LIU, A. Y. & ABRAHAM, B. A. (1991) Subtractive cloning of a hybrid human endogenous retrovirus and calbindin gene in the prostate cell line PC3. *Cancer Res,* 51**,** 4107-10.
LIVAK, K. J. & SCHMITTGEN, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.
LO CONTE, L., CHOTHIA, C. & JANIN, J. (1999) The atomic structure of protein-protein recognition sites. *J Mol Biol,* 285**,** 2177-98.
LOOMANS, E. E., GRIBNAU, T. C., BLOEMERS, H. P. & SCHIELEN, W. J. (1998) Adsorption studies of tritium-labeled peptides on polystyrene surfaces. *J Immunol Methods,* 221**,** 131-9.
LUM, A. M., WANG, B. B., LI, L., CHANNA, N., BARTHA, G. & WABL, M. (2007) Retroviral activation of the mir-106a microRNA cistron in T lymphoma. *Retrovirology,* 4**,** 5.
LUO, K., WANG, T., BINDONG, L., CHUNJUAN, T., ZUOXIANG, X., KAPPES, J. AND YU, X. F. (2007) Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J Virol*, 81, 7238-7248.
MACDONALD, K. P., NISHIOKA, Y., LIPSKY, P. E. & THOMAS, R. (1997) Functional CD40 ligand is expressed by T cells in rheumatoid arthritis. *J Clin Invest,* 100**,** 2404-14.
MACK, M., BENDER, K. & SCHNEIDER, P. M. (2004) Detection of retroviral antisense transcripts and promoter activity of the HERV-K(C4) insertion in the MHC class III region. *Immunogenetics,* 56**,** 321-32.
MADANI, N. AND KABAT, D. (1998) An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein. *J Virol,* 72, 10251-5.
MAHLER, M., MIERAU, R. & BLUTHNER, M. (2000) Fine-specificity of the anti-CENP-A B-cell autoimmune response. *J Mol Med,* 78**,** 460-7.
MAKUNIN, I. V., PHEASANT, M., SIMONS, C. & MATTICK, J. S. (2007) Orthologous MicroRNA Genes Are Located in Cancer-Associated Genomic Regions in Human and Mouse. *PLoS ONE,* 2**,** e1133.
MALE, D., BROSTOFF J. & I ROITT (2006) *Immunology,* London, Mosby.
MALIM, M. H., TILEY, L. S., MCCARN, D. F., RUSCHE, J. R., HAUBER, J. AND CULLEN, B. R. (1990) HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell,* 60, 675-83.
MALONEY, D. G., GRILLO-LOPEZ, A. J., WHITE, C. A., BODKIN, D., SCHILDER, R. J., NEIDHART, J. A., JANAKIRAMAN, N., FOON, K. A., LILES, T. M., DALLAIRE, B. K., WEY, K., ROYSTON, I., DAVIS, T. & LEVY, R. (1997) IDEC-C2B8 (Rituximab) anti-CD20 monoclonal antibody therapy in patients with relapsed low-grade non-Hodgkin's lymphoma. *Blood,* 90**,** 2188-95.
MAMELI, G., ASTONE, V., ARRU, G., MARCONI, S., LOVATO, L., SERRA, C., SOTGIU, S., BONETTI, B. & DOLEI, A. (2007) Brains and peripheral blood mononuclear cells of multiple sclerosis (MS) patients hyperexpress MS-

associated retrovirus/HERV-W endogenous retrovirus, but not Human herpesvirus 6. *J Gen Virol,* 88**,** 264-74.
MANGASARIAN, A., AND TRONO, D. (1997). The multifaceted role of HIV Nef. *Res Virol, 148*, 30-33.

MANGENEY, M., DE PARSEVAL, N., THOMAS, G. & HEIDMANN, T. (2001) The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties. *J Gen Virol,* 82**,** 2515-8.
MANIATIS, T., GOODBOURN, S. & FISCHER, J. A. (1987) Regulation of inducible and tissue-specific gene expression. *Science,* 236**,** 1237-45.
MANSIA, G., DE BACKER, G., DOMINICZAK, A., CIFKOVA, R., FAGARD, R., GERMANO, G., GRASSI, G., HEAGERTY, A. M., KJELDSEN, S. E., LAURENT, S., NARKIEWICZ, K., RUILOPE, L., RYNKIEWICZ, A., SCHMIEDER, R. E., STRUIJKER BOUDIER, H. A. & ZANCHETTI, A. (2007) 2007 ESH-ESC Guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Blood Press,* 16**,** 135-232.
MARE, L. & TRINCHERA, M. (2007) Comparative analysis of retroviral and native promoters driving expression of beta1,3-galactosyltransferase beta3Gal-T5 in human and mouse tissues. *J Biol Chem,* 282**,** 49-57.
MARGUERAT, S., WANG, W. Y., TODD, J. A. & CONRAD, B. (2004) Association of human endogenous retrovirus K-18 polymorphisms with type 1 diabetes. *Diabetes,* 53**,** 852-4.
MARTIN, A. C., CHEETHAM, J. C. & REES, A. R. (1989) Modeling antibody hypervariable loops: a combined algorithm. *Proc Natl Acad Sci U S A,* 86**,** 9268-72.
MARTIN, M. A., BRYAN, T., RASHEED, S. & KHAN, A. S. (1981) Identification and cloning of endogenous retroviral sequences present in human DNA. *Proc Natl Acad Sci U S A,* 78**,** 4892-6.
MARTIN, T. & PASQUALI, J. L. (1992) CD5 negative IGM rheumatoid factor B cells in B-chronic lymphocytic leukemia and benign mixed cryoglobulinemia. *Leuk Lymphoma,* 7**,** 55-62.
MASI, A. T., FEIGENBAUM, S. L. & CHATTERTON, R. T. (1995) Hormonal and pregnancy relationships to rheumatoid arthritis: convergent effects with immunologic and microvascular systems. *Semin Arthritis Rheum,* 25**,** 1-27.
MASUKO-HONGO, K., KATO, T. & NISHIOKA, K. (2003) Virus-associated arthritis. *Best Pract Res Clin Rheumatol,* 17**,** 309-18.
MAYER, J., EHLHARDT, S., SEIFERT, M., SAUTER, M., MULLER-LANTZSCH, N., MEHRAEIN, Y., ZANG, K. D. & MEESE, E. (2004) Human endogenous retrovirus HERV-K(HML-2) proviruses with Rec protein coding capacity and transcriptional activity. *Virology,* 322**,** 190-8.
MAZZUCCHELLI, L., BLASER, A., KAPPELER, A., SCHARLI, P., LAISSUE, J. A., BAGGIOLINI, M. & UGUCCIONI, M. (1999) BCA-1 is highly expressed in Helicobacter pylori-induced mucosa-associated lymphoid tissue and gastric lymphoma. *J Clin Invest,* 104**,** R49-54.
MCCLAIN, M. T., RAMSLAND, P. A., KAUFMAN, K. M. & JAMES, J. A. (2002) Anti-sm autoantibodies in systemic lupus target highly basic surface structures of complexed spliceosomal autoantigens. *J Immunol,* 168**,** 2054-62.

MCGUFFIN, L. J., BRYSON, K. & JONES, D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics,* 16**,** 404-5.
MCNEILAGE, L. J., MACMILLAN, E. M. & WHITTINGHAM, S. F. (1990) Mapping of epitopes on the La(SS-B) autoantigen of primary Sjogren's syndrome: identification of a cross-reactive epitope. *J Immunol,* 145**,** 3829-35.
MEDSTRAND, P. & BLOMBERG, J. (1993) Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. *J Virol,* 67**,** 6778-87.
MEDSTRAND, P., LINDESKOG, M. & BLOMBERG, J. (1992) Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. *J Gen Virol,* 73 (Pt 9)**,** 2463-6.
MEDSTRAND, P. & MAGER, D. L. (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. *J Virol,* 72**,** 9782-7.
MEDSTRAND, P., VAN DE LAGEMAAT, L. N. & MAGER, D. L. (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res,* 12**,** 1483-95.
MEESE, E., GOTTERT, E., ZANG, K. D., SAUTER, M., SCHOMMER, S. & MUELLER-LANTZSCH, N. (1996) Human endogenous retroviral element k10 (HERV-K10): chromosomal localization by somatic hybrid mapping and fluorescence in situ hybridization. *Cytogenet Cell Genet,* 72**,** 40-2.
MEISLER, M. H. & TING, C. N. (1993) The remarkable evolutionary history of the human amylase genes. *Crit Rev Oral Biol Med,* 4**,** 503-9.
MENARD, H. A., LAPOINTE, E., ROCHDI, M. D. & ZHOU, Z. J. (2000) Insights into rheumatoid arthritis derived from the Sa immune system. *Arthritis Res,* 2**,** 429-32.
MEYLAN, F., DE SMEDT, M., LECLERCQ, G., PLUM, J., LEUPIN, O., MARGUERAT, S. & CONRAD, B. (2005) Negative thymocyte selection to HERV-K18 superantigens in humans. *Blood,* 105**,** 4377-82.
MI, S., LEE, X., LI, X., VELDMAN, G. M., FINNERTY, H., RACIE, L., LAVALLIE, E., TANG, X. Y., EDOUARD, P., HOWES, S., KEITH, J. C., JR. & MCCOY, J. M. (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature,* 403**,** 785-9.
MICHAELS, F. H., BANKS, K. L. & REITZ, M. S., JR. (1991) Lessons from caprine and ovine retrovirus infections. *Rheum Dis Clin North Am,* 17**,** 5-23.
MIKECZ, K., BRENNAN, F. R., KIM, J. H. & GLANT, T. T. (1995) Anti-CD44 treatment abrogates tissue oedema and leukocyte infiltration in murine arthritis. *Nat Med,* 1**,** 558-63.
MIKSZTA, J. A., JANG, Y. S. & KIM, B. S. (1997) Role of a C-terminal residue of an immunodominant epitope in T cell activation and repertoire diversity. *J Immunol,* 158**,** 127-35.
MILICH, D. R., MCLACHLAN, A., THORNTON, G. B. & HUGHES, J. L. (1987) Antibody production to the nucleocapsid and envelope of the hepatitis B virus primed by a single synthetic T cell site. *Nature,* 329**,** 547-9.
MILLS, R. E., BENNETT, E. A., ISKOW, R. C. & DEVINE, S. E. (2007) Which transposable elements are active in the human genome? *Trends Genet,* 23**,** 183-91.
MIMORI, T. & TANAKA, M. (2002) [Novel autoantibodies and their target antigens in rheumatoid arthritis]. *Nippon Rinsho,* 60**,** 2263-8.

MINOTA, S., KOYASU, S., YAHARA, I. & WINFIELD, J. (1988) Autoantibodies to the heat-shock protein hsp90 in systemic lupus erythematosus. *J Clin Invest,* 81**,** 106-9.

MITCHELL, R. S., BEITZEL, B. F., SCHRODER, A. R., SHINN, P., CHEN, H., BERRY, C. C., ECKER, J. R. & BUSHMAN, F. D. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol,* 2**,** E234.

MOHAN, N., EDWARDS, E. T., CUPPS, T. R., OLIVERIO, P. J., SANDBERG, G., CRAYTON, H., RICHERT, J. R. & SIEGEL, J. N. (2001) Demyelination occurring during anti-tumor necrosis factor alpha therapy for inflammatory arthritides. *Arthritis Rheum,* 44**,** 2862-9.

MOUDGIL, K. D., CHANG, T. T., ERADAT, H., CHEN, A. M., GUPTA, R. S., BRAHN, E. & SERCARZ, E. E. (1997) Diversification of T cell responses to carboxy-terminal determinants within the 65-kD heat-shock protein is involved in regulation of autoimmune arthritis. *J Exp Med,* 185**,** 1307-16.

MOUDGIL, K. D. (1998) Diversification of response to hsp65 during the course of autoimmune arthritis is regulatory rather than pathogenic. *Immunol Rev,* 164, 175-84.

MOULT J., HUBBARD T., FIDELIS K. & ZEMLA., A. (2000) Critical Assessment of Techniques for Protein Structure Prediction (CASP4). *Fourth Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction.* Asilomar Conference Center

MOYES, D., GRIFFITHS, D. J. & VENABLES, P. J. (2007) Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. *Trends Genet,* 23**,** 326-33.

MOYES, D. L., MARTIN, A., SAWCER, S., TEMPERTON, N., WORTHINGTON, J., GRIFFITHS, D. J. & VENABLES, P. J. (2005) The distribution of the endogenous retroviruses HERV-K113 and HERV-K115 in health and disease. *Genomics,* 86**,** 337-41.

MU, J., KANZAKI, T., SI, X., TOMIMATSU, T., FUKUDA, H., SHIOJI, M., MURATA, Y., SUGIMOTO, Y. & ICHIKAWA, A. (2003) Apoptosis and related proteins in placenta of intrauterine fetal death in prostaglandin f receptor-deficient mice. *Biol Reprod,* 68**,** 1968-74.

MUIR, A., LEVER, A. & MOFFETT, A. (2004) Expression and functions of human endogenous retroviruses in the placenta: an update. *Placenta,* 25 Suppl A**,** S16-25.

MULLER-LADNER, U., KRIEGSMANN, J., FRANKLIN, B. N., MATSUMOTO, S., GEILER, T., GAY, R. E. & GAY, S. (1996) Synovial fibroblasts of patients with rheumatoid arthritis attach to and invade normal human cartilage when engrafted into SCID mice. *Am J Pathol,* 149**,** 1607-15.

MULLER-LADNER, U., PAP, T., GAY, R. E. & GAY, S. (2003) Gene transfer as a future therapy for rheumatoid arthritis. *Expert Opin Biol Ther,* 3**,** 587-98.

MULLER, Y. A., CHEN, Y., CHRISTINGER, H. W., LI, B., CUNNINGHAM, B. C., LOWMAN, H. B. & DE VOS, A. M. (1998) VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 A resolution and mutational analysis of the interface. *Structure,* 6**,** 1153-67.

MURO, Y., IWAI, T. & OHASHI, M. (1996) A charged segment mainly composed of basic amino acids forms an autoepitope of CENP-A. *Clin Immunol Immunopathol,* 78**,** 86-9.

MURPHY, E. L., WANG, B., SACHER, R. A., FRIDEY, J., SMITH, J. W., NASS, C. C., NEWMAN, B., OWNBY, H. E., GARRATTY, G., HUTCHING, S. T. & SCHREIBER, G. B. (2004) Respiratory and urinary tract infections, arthritis, and asthma associated with HTLV-I and HTLV-II infection. *Emerg Infect Dis,* 10**,** 109-16.

NAKAGAWA, K., BRUSIC, V., MCCOLL, G. & HARRISON, L. C. (1997) Direct evidence for the expression of multiple endogenous retroviruses in the synovial compartment in rheumatoid arthritis. *Arthritis Rheum,* 40**,** 627-38.

NAKAJIMA, T., AONO, H., HASUNUMA, T., YAMAMOTO, K., MARUYAMA, I., NOSAKA, T., HATANAKA, M. & NISHIOKA, K. (1993) Overgrowth of human synovial cells driven by the human T cell leukemia virus type I tax gene. *J Clin Invest,* 92**,** 186-93.

NAKANE, P. K. & PIERCE, G. B., JR. (1966) Enzyme-labeled antibodies: preparation and application for the localization of antigens. *J Histochem Cytochem,* 14**,** 929-31.

NAKANO, K., OKADA, Y., SAITO, K. & TANAKA, Y. (2004) Induction of RANKL expression and osteoclast maturation by the binding of fibroblast growth factor 2 to heparan sulfate proteoglycan on rheumatoid synovial fibroblasts. *Arthritis Rheum,* 50**,** 2450-8.

NAKELCHIK, M. & MANGINO, J. E. (2002) Reactivation of histoplasmosis after treatment with infliximab. *Am J Med,* 112**,** 78.

NATVIG, J. B. & TONDER, O. (1998) The discovery of the rheumatoid factor. I. Erik Waaler. 1940. *Clin Exp Rheumatol,* 16**,** 340-4.

NEIDHART, M., RETHAGE, J., KUCHEN, S., KUNZLER, P., CROWL, R. M., BILLINGHAM, M. E., GAY, R. E. & GAY, S. (2000) Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. *Arthritis Rheum,* 43**,** 2634-47.

NELLAKER, C., YAO, Y., JONES-BRANDO, L., MALLET, F., YOLKEN, R. H. & KARLSSON, H. (2006) Transactivation of elements in the human endogenous retrovirus W family by viral infection. *Retrovirology,* 3**,** 44.

NELSON, P. N., BOWMAN, S. J., HAY, F. C., LANCHBURY, J. S., PANAYI, G. S. & LEVER, A. M. (1995) Absence of exogenous retroviruses in Felty's syndrome. *Br J Rheumatol,* 34**,** 185-7.

NELSON, P. N., CARNEGIE, P. R., MARTIN, J., DAVARI EJTEHADI, H., HOOLEY, P., RODEN, D., ROWLAND-JONES, S., WARREN, P., ASTLEY, J. & MURRAY, P. G. (2003) Demystified. Human endogenous retroviruses. *Mol Pathol,* 56**,** 11-8.

NELSON, P. N., GOODALL, M. & JEFFERIS, R. (1994) Characterisation of putative monoclonal anti-G3m(u) and anti-G3m(g) reagents and their antigenic determinants. *Immunol Invest,* 23**,** 39-45.

NELSON, P. N., HOOLEY, P., RODEN, D., DAVARI EJTEHADI, H., RYLANCE, P., WARREN, P., MARTIN, J. & MURRAY, P. G. (2004) Human endogenous retroviruses: transposable elements with potential? *Clin Exp Immunol,* 138**,** 1-9.

NELSON, P. N., LEVER, A. M., SMITH, S., PITMAN, R., MURRAY, P., PERERA, S. A., WESTWOOD, O. M., HAY, F. C., EJTEHADI, H. D. & BOOTH, J. C. (1999) Molecular investigations implicate human endogenous retroviruses as mediators of anti-retroviral antibodies in autoimmune rheumatic disease. *Immunol Invest,* 28**,** 277-89.

NEMETHY, G. & SCHERAGA, H. A. (1980) Stereochemical requirements for the existence of hydrogen bonds in beta-bends. *Biochem Biophys Res Commun,* 95**,** 320-7.
NEWMAN, W. (1982) Selective blockade of human natural killer cells by a monoclonal antibody. *Proc Natl Acad Sci U S A,* 79**,** 3858-62.
NISHIMOTO, N. & KISHIMOTO, T. (2004) Inhibition of IL-6 for the treatment of inflammatory diseases. *Curr Opin Pharmacol,* 4**,** 386-91.
NISHIMOTO, N. & KISHIMOTO, T. (2006) Interleukin 6: from bench to bedside. *Nat Clin Pract Rheumatol,* 2**,** 619-26.
NORVAL, M., HART, H. & MARMION, B. P. (1979) Viruses and lymphocytes in rheumatoid arthritis. I. Studies on cultured rheumatoid lymphocytes. *Ann Rheum Dis,* 38**,** 507-13.
OGASAWARA, H., OKADA, M., KANEKO, H., HISHIKAWA, T., SEKIGAWA, I. & HASHIMOTO, H. (2003) Possible role of DNA hypomethylation in the induction of SLE: relationship to the transcription of human endogenous retroviruses. *Clin Exp Rheumatol,* 21**,** 733-8.
OKULICZ, W. C. & ACE, C. I. (2003) Temporal regulation of gene expression during the expected window of receptivity in the rhesus monkey endometrium. *Biol Reprod,* 69**,** 1593-9.
OLDSTONE, M. B. (2005) Molecular mimicry, microbial infection, and autoimmune disease: evolution of the concept. *Curr Top Microbiol Immunol,* 296**,** 1-17.
OLSEN, N. J., BROOKS, R. H., CUSH, J. J., LIPSKY, P. E., ST CLAIR, E. W., MATTESON, E. L., GOLD, K. N., CANNON, G. W., JACKSON, C. G., MCCUNE, W. J., FOX, D. A., NELSON, B., LORENZ, T. & STRAND, V. (1996) A double-blind, placebo-controlled study of anti-CD5 immunoconjugate in patients with rheumatoid arthritis. The Xoma RA Investigator Group. *Arthritis Rheum,* 39**,** 1102-8.
OLSEN, N. J. & KOVACS, W. J. (2002) Hormones, pregnancy, and rheumatoid arthritis. *J Gend Specif Med,* 5**,** 28-37.
ONO, M., YASUNAGA, T., MIYATA, T. & USHIKUBO, H. (1986) Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *J Virol,* 60**,** 589-98.
OOKA, S., MATSUI, T., NISHIOKA, K. & KATO, T. (2003) Autoantibodies to low-density-lipoprotein-receptor-related protein 2 (LRP2) in systemic autoimmune diseases. *Arthritis Res Ther,* 5**,** R174-80.
ORICCHIO, E., SCIAMANNA, I., BERALDI, R., TOLSTONOG, G. V., SCHUMANN, G. G. & SPADAFORA, C. (2007) Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene,* 26**,** 4226-33.
ORTENDAHL, M., HOLMES, T., SCHETTLER, J. D. & FRIES, J. F. (2002) The methotrexate therapeutic response in rheumatoid arthritis. *J Rheumatol,* 29**,** 2084-91.
OSHIMA, M. & ATASSI, M. Z. (1989) Comparison of peptide-coating conditions in solid phase plate assays for detection of anti-peptide antibodies. *Immunol Invest,* 18**,** 841-51.
PADYUKOV, L., SILVA, C., STOLT, P., ALFREDSSON, L. & KLARESKOG, L. (2004) A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum,* 50**,** 3085-92.

PALIARD, X., WEST, S. G., LAFFERTY, J. A., CLEMENTS, J. R., KAPPLER, J. W., MARRACK, P. & KOTZIN, B. L. (1991) Evidence for the effects of a superantigen in rheumatoid arthritis. *Science,* 253**,** 325-9.

PALMARINI, M., MURA, M. & SPENCER, T. E. (2004) Endogenous betaretroviruses of sheep: teaching new lessons in retroviral interference and adaptation. *J Gen Virol,* 85**,** 1-13.

PALMER K. T. & COOPER, C. (2006) Work-Related Disorders of the Upper Limb. *Reports on the Rheumatic Diseases,* 5**,** 1-7.

PAP, T., FRANZ, J. K., HUMMEL, K. M., JEISY, E., GAY, R. & GAY, S. (2000a) Activation of synovial fibroblasts in rheumatoid arthritis: lack of Expression of the tumour suppressor PTEN at sites of invasive growth and destruction. *Arthritis Res,* 2**,** 59-64.

PAP, T., MULLER-LADNER, U., GAY, R. E. & GAY, S. (2000b) Fibroblast biology. Role of synovial fibroblasts in the pathogenesis of rheumatoid arthritis. *Arthritis Res,* 2**,** 361-7.

PAP, T., AUPPERLE, K. R., GAY, S., FIRESTEIN, G. S. AND GAY, R. E. (2001) Invasiveness of synovial fibroblasts is regulated by p53 in the SCID mouse in vivo model of cartilage invasion. *Arthritis Rheum,* **44,** 676-81.

PARKER, J. M., GUO, D. & HODGES, R. S. (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry,* 25**,** 5425-32.

PAROLINI, S., BOTTINO, C., FALCO, M., AUGUGLIARO, R., GILIANI, S., FRANCESCHINI, R., OCHS, H. D., WOLF, H., BONNEFOY, J. Y., BIASSONI, R., MORETTA, L., NOTARANGELO, L. D. & MORETTA, A. (2000) X-linked lymphoproliferative disease. 2B4 molecules displaying inhibitory rather than activating function are responsible for the inability of natural killer cells to kill Epstein-Barr virus-infected cells. *J Exp Med,* 192**,** 337-46.

PARSONAGE, G., FALCIANI, F., BURMAN, A., FILER, A., ROSS, E., BOFILL, M., MARTIN, S., SALMON, M. & BUCKLEY, C. D. (2003) Global gene expression profiles in fibroblasts from synovial, skin and lymphoid tissue reveals distinct cytokine and chemokine expression patterns. *Thromb Haemost,* 90**,** 688-97.

PASCUAL, M., KOELEMAN, B. P., EERLIGH, P., ROEP, B. O. & MARTIN, J. (2003) Distribution of HERV-LTR elements in HLA-DQB1 alleles: comments on the article by Bieda et al. *Diabetologia,* 46**,** 869-70; author reply 870-1.

PASCUAL, M., MARTIN, J., NIETO, A., GIPHART, M. J., VAN DER SLIK, A. R., DE VRIES, R. R. & ZANELLI, E. (2001) Distribution of HERV-LTR elements in the 5'-flanking region of HLA-DQB1 and association with autoimmunity. *Immunogenetics,* 53**,** 114-8.

PASPARAKIS, M., ALEXOPOULOU, L., EPISKOPOU, V. & KOLLIAS, G. (1996) Immune and inflammatory responses in TNF alpha-deficient mice: a critical requirement for TNF alpha in the formation of primary B cell follicles, follicular dendritic cell networks and germinal centers, and in the maturation of the humoral immune response. *J Exp Med,* 184**,** 1397-411.

PATTISON, D. J., SILMAN, A. J., GOODSON, N. J., LUNT, M., BUNN, D., LUBEN, R., WELCH, A., BINGHAM, S., KHAW, K. T., DAY, N. & SYMMONS, D. P. (2004a) Vitamin C and the risk of developing inflammatory polyarthritis: prospective nested case-control study. *Ann Rheum Dis,* 63**,** 843-7.

PATTISON, D. J., SYMMONS, D. P., LUNT, M., WELCH, A., LUBEN, R., BINGHAM, S. A., KHAW, K. T., DAY, N. E. & SILMAN, A. J. (2004b) Dietary risk factors for the development of inflammatory polyarthritis: evidence for a role of high level of red meat consumption. *Arthritis Rheum,* 50**,** 3804-12.
PAVLICEK, A., PACES, J., ELLEDER, D. & HEJNAR, J. (2002) Processed pseudogenes of human endogenous retroviruses generated by LINEs: their integration, stability, and distribution. *Genome Res,* 12**,** 391-9.
PAWELEC, G., AKBAR, A., CARUSO, C., SOLANA, R., GRUBECK-LOEBENSTEIN, B. & WIKBY, A. (2005) Human immunosenescence: is it infectious? *Immunol Rev,* 205**,** 257-68.
PEELING, R. W., WANG, S. P., GRAYSTON, J. T., BLASI, F., BOMAN, J., CLAD, A., FREIDANK, H., GAYDOS, C. A., GNARPE, J., HAGIWARA, T., JONES, R. B., ORFILA, J., PERSSON, K., PUOLAKKAINEN, M., SAIKKU, P. & SCHACHTER, J. (2000) *Chlamydia pneumoniae* serology: interlaboratory variation in microimmunofluorescence assay results. *J Infect Dis,* 181 Suppl 3**,** S426-9.
PELLEQUER, J. L. & WESTHOF, E. (1993) PREDITOP: a program for antigenicity prediction. *J Mol Graph,* 11**,** 204-10, 191-2.
PELTON, B. K., NORTH, M., PALMER, R. G., HYLTON, W., SMITH-BURCHNELL, C., SINCLAIR, A. L., MALKOVSKY, M., DALGLEISH, A. G. & DENMAN, A. M. (1988) A search for retrovirus infection in systemic lupus erythematosus and rheumatoid arthritis. *Ann Rheum Dis,* 47**,** 206-9.
PERL, A. (2001) Endogenous retroviruses in pathogenesis of autoimmunity. *J Rheumatol,* 28**,** 461-4.
PERL, A. (2003) Role of endogenous retroviruses in autoimmune diseases. *Rheum Dis Clin North Am,* 29**,** 123-43, vii.
PERRON, H., JOUVIN-MARCHE, E., MICHEL, M., OUNANIAN-PARAZ, A., CAMELO, S., DUMON, A., JOLIVET-REYNAUD, C., MARCEL, F., SOUILLET, Y., BOREL, E., GEBUHRER, L., SANTORO, L., MARCEL, S., SEIGNEURIN, J. M., MARCHE, P. N. & LAFON, M. (2001) Multiple sclerosis retrovirus particles and recombinant envelope trigger an abnormal immune response in vitro, by inducing polyclonal Vbeta16 T-lymphocyte activation. *Virology,* 287**,** 321-32.
PETERHANS, E., ZANONI, R. & BERTONI, G. (1999) How to succeed as a virus: strategies for dealing with the immune system. *Vet Immunol Immunopathol,* 72**,** 111-7.
PETERSEN, C. P., BORDELEAU, M. E., PELLETIER, J. & SHARP, P. A. (2006) Short RNAs repress translation after initiation in mammalian cells. *Mol Cell,* 21**,** 533-42.
PETROPOULUS, C. (1997) Retroviral taxonomy, protein structures, sequences and genetic maps. IN COFFIN, J. M., HUGHES, S. H., AND VARMUS, H. E., (Ed.) *Retroviruses.* New York, Cold Spring Harbour Laboratory press.
PETROVSKY, N. & BRUSIC, V. (2006) Bioinformatics for study of autoimmunity. *Autoimmunity,* 39**,** 635-43.
PFAFFL, M. W., GEORGIEVA, T. M., GEORGIEV, I. P., ONTSOUKA, E., HAGELEIT, M. & BLUM, J. W. (2002) Real-time RT-PCR quantification of insulin-like growth factor (IGF)-1, IGF-1 receptor, IGF-2, IGF-2 receptor, insulin receptor, growth hormone receptor, IGF-binding proteins 1, 2 and 3 in the bovine species. *Domest Anim Endocrinol,* 22**,** 91-102.

PIERER, M., RETHAGE, J., SEIBL, R., LAUENER, R., BRENTANO, F., WAGNER, U., HANTZSCHEL, H., MICHEL, B. A., GAY, R. E., GAY, S. & KYBURZ, D. (2004) Chemokine secretion of rheumatoid arthritis synovial fibroblasts stimulated by Toll-like receptor 2 ligands. *J Immunol,* 172**,** 1256-65.

PILLAI, R. S., BHATTACHARYYA, S. N. & FILIPOWICZ, W. (2007) Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol,* 17**,** 118-26.

PILLINGER, M. H. & ABRAMSON, S. B. (1995) The neutrophil in rheumatoid arthritis. *Rheum Dis Clin North Am,* 21**,** 691-714.

POGUE, G. P., HOFMANN, J., DUNCAN, R., BEST, J. M., ETHERINGTON, J., SONTHEIMER, R. D. & NAKHASI, H. L. (1996) Autoantigens interact with cis-acting elements of rubella virus RNA. *J Virol,* 70**,** 6269-77.

POLGAR, A., FALUS, A., KOO, E., UJFALUSSY, I., SESZTAK, M., SZUTS, I., KONRAD, K., HODINKA, L., BENE, E., MESZAROS, G., ORTUTAY, Z., FARKAS, E., PAKSY, A. & BUZAS, E. I. (2003) Elevated levels of synovial fluid antibodies reactive with the small proteoglycans biglycan and decorin in patients with rheumatoid arthritis or other joint diseases. *Rheumatology (Oxford),* 42**,** 522-7.

PONFERRADA, V. G., MAUCK, B. S. & WOOLEY, D. P. (2003) The envelope glycoprotein of human endogenous retrovirus HERV-W induces cellular resistance to spleen necrosis virus. *Arch Virol,* 148**,** 659-75.

PONNUSWAMY, P. K., PRABHAKARAN, M. & MANAVALAN, P. (1980) Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. *Biochim Biophys Acta,* 623**,** 301-16.

POSTEPSKI, J., OPOKA-WINIARSKA, V., KOZIOL-MONTEWKA, M., KOROBOWICZ, A. & TUSZKIEWICZ-MISZTAL, E. (2003) Role of mycoplasma pneumoniae infection in aetiopathogenesis of juvenile idiopatic arthritis. *Med Wieku Rozwoj,* 7**,** 271-7.

POWELL, A. M. & BLACK, M. M. (2001) Epitope spreading: protection from pathogens, but propagation of autoimmunity? *Clin Exp Dermatol,* 26**,** 427-33.

PROCACCIA, S., GASPARINI, A., COLUCCI, A., LANZANOVA, D., BIANCHI, M., FORCELLINI, P., VILLA, P., BLASIO, R. & ZANUSSI, C. (1987) ELISA determined IgM, IgG and IgA rheumatoid factors in rheumatoid arthritis and in other connective tissue diseases. *Clin Exp Rheumatol,* 5**,** 335-42.

QIAGEN (2006) Critical Factors for Successful Real-Time PCR Nanogene.

RAHMAN, A. & HIEPE, F. (2002) Anti-DNA antibodies--overview of assays and clinical correlations. *Lupus,* 11**,** 770-3.

RAM, S. & SHANKER, R. (2005) Computing TaqMan probes for multiplex PCR detection of E. coli O157 serotypes in water. *In Silico Biol,* 5**,** 499-504.

RAMAKERS, C., RUIJTER, J. M., DEPREZ, R. H. & MOORMAN, A. F. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett,* 339**,** 62-6.

RAMAN, C. (2002) CD5, an important regulator of lymphocyte selection and immune tolerance. *Immunol Res,* 26**,** 255-63.

RAMOS-NINO, M. E., SCAPOLI, L., MARTINELLI, M., LAND, S. & MOSSMAN, B. T. (2003) Microarray analysis and RNA silencing link fra-1 to cd44 and c-met expression in mesothelioma. *Cancer Res,* 63**,** 3539-45.

RANDEN, I., BROWN, D., THOMPSON, K. M., HUGHES-JONES, N., PASCUAL, V., VICTOR, K., CAPRA, J. D., FORRE, O. & NATVIG, J. B. (1992) Clonally

related IgM rheumatoid factors undergo affinity maturation in the rheumatoid synovial tissue. *J Immunol,* 148**,** 3296-301.
RANTAPAA-DAHLQVIST, S., DE JONG, B. A., BERGLIN, E., HALLMANS, G., WADELL, G., STENLUND, H., SUNDIN, U. & VAN VENROOIJ, W. J. (2003) Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum,* 48**,** 2741-9.
RAU, R., HERBORN, G., MENNINGER, H. & BLECHSCHMIDT, J. (1997) Comparison of intramuscular methotrexate and gold sodium thiomalate in the treatment of early erosive rheumatoid arthritis: 12 month data of a double-blind parallel study of 174 patients. *Br J Rheumatol,* 36**,** 345-52.
REES, A. R., STAUNTON, D., WEBSTER, D. M., SEARLE, S. J., HENRY, A. H. & PEDERSEN, J. T. (1994) Antibody design: beyond the natural limits. *Trends Biotechnol,* 12**,** 199-206.
REFF, M. E., CARNER, K., CHAMBERS, K. S., CHINN, P. C., LEONARD, J. E., RAAB, R., NEWMAN, R. A., HANNA, N. & ANDERSON, D. R. (1994) Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20. *Blood,* 83**,** 435-45.
RENAUDINEAU, Y., VALLET, S., LE DANTEC, C., HILLION, S., SARAUX, A. & YOUINOU, P. (2005) Characterization of the human CD5 endogenous retrovirus-E in B lymphocytes. *Genes Immun,* 6**,** 663-71.
REPASKE, R., STEELE, P. E., O'NEILL, R. R., RABSON, A. B. & MARTIN, M. A. (1985) Nucleotide sequence of a full-length human endogenous retroviral segment. *J Virol,* 54**,** 764-72.
REUS, K., MAYER, J., SAUTER, M., ZISCHLER, H., MULLER-LANTZSCH, N. & MEESE, E. (2001) HERV-K(OLD): ancestor sequences of the human endogenous retrovirus family HERV-K(HML-2). *J Virol,* 75**,** 8917-26.
REVIRON, D., PERDRIGER, A., TOUSSIROT, E., WENDLING, D., BALANDRAUD, N., GUIS, S., SEMANA, G., TIBERGHIEN, P., MERCIER, P. & ROUDIER, J. (2001) Influence of shared epitope-negative HLA-DRB1 alleles on genetic susceptibility to rheumatoid arthritis. *Arthritis Rheum,* 44**,** 535-40.
RINI, J. M., SCHULZE-GAHMEN, U. & WILSON, I. A. (1992) Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science,* 255**,** 959-65.
RIRIE, K. M., RASMUSSEN, R. P. & WITTWER, C. T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem,* 245**,** 154-60.
RIZOU, C., IOANNIDIS, J. P., PANOU-POMONIS, E., SAKARELLOS-DAITSIOTIS, M., SAKARELLOS, C., MOUTSOPOULOS, H. M. & VLACHOYIANNOPOULOS, P. G. (2000) B-Cell epitope mapping of DNA topoisomerase I defines epitopes strongly associated with pulmonary fibrosis in systemic sclerosis. *Am J Respir Cell Mol Biol,* 22**,** 344-51.
ROBERTS, C. R., HERRMANN, M., AICHER, W. K., MULLERLADNER, U., MATSUMOTO, S., GAY, R. E., KALDEN, J. R. & GAY, S. (1995) Sequences Derived from Pcr Amplification of Virus-Like Particles (Vlp) Present in Synovial-Fluid of Patients with Rheumatoid Arthritis(Ra). *Arthritis and Rheumatism,* 38**,** 388-388.
RONAGHY, A., PRAKKEN, B. J., TAKABAYASHI, K., FIRESTEIN, G. S., BOYLE, D., ZVAILFLER, N. J., ROORD, S. T., ALBANI, S., CARSON, D.

A. & RAZ, E. (2002) Immunostimulatory DNA sequences influence the course of adjuvant arthritis. *J Immunol,* 168**,** 51-6.
ROSE, G. D., GIERASCH, L. M. & SMITH, J. A. (1985) Turns in peptides and proteins. *Adv Protein Chem,* 37**,** 1-109.
ROUTSIAS, J. G., TZIOUFAS, A. G. & MOUTSOPOULOS, H. M. (2004) The clinical value of intracellular autoantigens B-cell epitopes in systemic rheumatic diseases. *Clin Chim Acta,* 340**,** 1-25.
ROUTSIAS, J. G., TZIOUFAS, A. G., SAKARELLOS-DAITSIOTIS, M., SAKARELLOS, C. & MOUTSOPOULOS, H. M. (1996) Epitope mapping of the Ro/SSA60KD autoantigen reveals disease-specific antibody-binding profiles. *Eur J Clin Invest,* 26**,** 514-21.
ROY-BURMAN, P. (1995) Endogenous env elements: partners in generation of pathogenic feline leukemia viruses. *Virus Genes,* 11**,** 147-61.
RUBIN, H. (1965) Genetic Control of Cellular Susceptibility to Pseudotypes of Rous Sarcoma Virus. *Virology,* 26**,** 270-6.
RUPRECHT, K., OBOJES, K., WENGEL, V., GRONEN, F., KIM, K. S., PERRON, H., SCHNEIDER-SCHAULIES, J. AND RIECKMANN, P. (2006) Regulation of human endogenous retrovirus W protein expression by herpes simplex virus type 1: implications for multiple sclerosis. *J Neurovirol,* 12, 65-71.
RYAN, F. P. (2004) Human endogenous retroviruses in health and disease: a symbiotic perspective. *J R Soc Med,* 97**,** 560-5.
SAAG, K. G., CRISWELL, L. A., SEMS, K. M., NETTLEMAN, M. D. & KOLLURI, S. (1996) Low-dose corticosteroids in rheumatoid arthritis. A meta-analysis of their moderate-term effectiveness. *Arthritis Rheum,* 39**,** 1818-25.
SAAL, J. G., KRIMMEL, M., STEIDLE, M., GERNETH, F., WAGNER, S., FRITZ, P., KOCH, S., ZACHER, J., SELL, S., EINSELE, H. & MULLER, C. A. (1999) Synovial Epstein-Barr virus infection increases the risk of rheumatoid arthritis in individuals with the shared HLA-DR4 epitope. *Arthritis Rheum,* 42**,** 1485-96.
SAHA, S. & RAGHAVA, G. P. S. (2004) BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. *Artificial Immune Systems, Proceedings.*
SALAFFI, F., FERRACCIOLI, G., PERONI, M., CAROTTI, M., BARTOLI, E. & CERVINI, C. (1994) Progression of erosion and joint space narrowing scores in rheumatoid arthritis assessed by nonlinear models. *J Rheumatol,* 21**,** 1626-30.
SALGHETTI, S., MARIANI, R., AND SKOWRONSKI, J. (1995). Human immunodeficiency virus type 1 Nef and p56lck protein-tyrosine kinase interact with a common element in CD4 cytoplasmic tail. *Proc Natl Acad Sci U S A 92*, 349-353.
SALMON, M. & AKBAR, A. N. (2004) Telomere erosion: a new link between HLA DR4 and rheumatoid arthritis? *Trends Immunol,* 25**,** 339-41.
SALMON, M., SCHEEL-TOELLNER, D., HUISSOON, A. P., PILLING, D., SHAMSADEEN, N., HYDE, H., D'ANGEAC, A. D., BACON, P. A., EMERY, P. & AKBAR, A. N. (1997) Inhibition of T cell apoptosis in the rheumatoid synovium. *J Clin Invest,* 99**,** 439-46.
SALMONS, B. & GUNZBURG, W. H. (1987) Current perspectives in the biology of mouse mammary tumour virus. *Virus Res,* 8**,** 81-102.
SALTARELLI, M., QUERAT, G., KONINGS, D. A., VIGNE, R. & CLEMENTS, J. E. (1990) Nucleotide sequence and transcriptional analysis of molecular clones of CAEV which generate infectious virus. *Virology,* 179**,** 347-64.

SALZER, U., CHAPEL, H. M., WEBSTER, A. D., PAN-HAMMARSTROM, Q., SCHMITT-GRAEFF, A., SCHLESIER, M., PETER, H. H., ROCKSTROH, J. K., SCHNEIDER, P., SCHAFFER, A. A., HAMMARSTROM, L. & GRIMBACHER, B. (2005) Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet,* 37**,** 820-8.

SAMBROOK J & RUSSELL, D. (2000) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

SANDER, D. M., SZABO, S., GALLAHER, W. R., DEAS, J. E., THOMPSON, J. J., CAO, Y., LUO-ZHANG, H., LIU, L. G., COLMEGNA, I., KOEHLER, J., ESPINOZA, L. R., ALEXANDER, S. S., HART, D. J., TOM, D. M., FERMIN, C. D., JASPAN, J. J., KULAKOSKY, P. C., TENENBAUM, S. A., WILSON, R. B. & GARRY, R. F. (2005) Involvement of human intracisternal A-type retroviral particles in autoimmunity. *Microsc Res Tech,* 68**,** 222-34.

SAUTER, M., SCHOMMER, S., KREMMER, E., REMBERGER, K., DOLKEN, G., LEMM, I., BUCK, M., BEST, B., NEUMANN-HAEFELIN, D. & MUELLER-LANTZSCH, N. (1995) Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas. *J Virol,* 69**,** 414-21.

SCHETT, G., HAYER, S., ZWERINA, J., REDLICH, K. & SMOLEN, J. S. (2005) Mechanisms of Disease: the link between RANKL and arthritic bone disease. *Nat Clin Pract Rheumatol,* 1**,** 47-54.

SCHMITT, J. & PAPISCH, W. (2002) Recombinant autoantigens. *Autoimmun Rev,* 1**,** 79-88.

SCHON, U., SEIFARTH, W., BAUST, C., HOHENADL, C., ERFLE, V. & LEIB-MOSCH, C. (2001) Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. *Virology,* 279**,** 280-91.

SCHONLAND, S. O., LOPEZ, C., WIDMANN, T., ZIMMER, J., BRYL, E., GORONZY, J. J. & WEYAND, C. M. (2003) Premature telomeric loss in rheumatoid arthritis is genetically determined and involves both myeloid and lymphoid cell lineages. *Proc Natl Acad Sci U S A,* 100**,** 13471-6.

SCHUELER-FURMAN, O., ELBER, R. & MARGALIT, H. (1998) Knowledge-based structure prediction of MHC class I bound peptides: a study of 23 complexes. *Fold Des,* 3**,** 549-64.

SCOFIELD, R. H. (2004) Autoantibodies as predictors of disease. *Lancet,* 363**,** 1544-6.

SCOFIELD, R. H. & HARLEY, J. B. (1991) Autoantigenicity of Ro/SSA antigen is related to a nucleocapsid protein of vesicular stomatitis virus. *Proc Natl Acad Sci U S A,* 88**,** 3343-7.

SEEMAYER, C. A., KOLB, S. A., NEIDHART, M., OHSHIMA, S., GAY, R. E., MICHEL, B. A., GAY, S., BONI, J. & SCHUPBACH, J. (2002) Absence of inducible retroviruses from synovial fibroblasts and synovial fluid cells of patients with rheumatoid arthritis. *Arthritis Rheum,* 46**,** 2811-3.

SEEMAYER, C. A., KUCHEN, S., NEIDHART, M., KUENZLER, P., RIHOSKOVA, V., NEUMANN, E., PRUSCHY, M., AICHER, W. K., MULLER-LADNER, U., GAY, R. E., MICHEL, B. A., FIRESTEIN, G. S. & GAY, S. (2003) p53 in rheumatoid arthritis synovial fibroblasts at sites of invasion. *Ann Rheum Dis,* 62**,** 1139-44.

SEGGERSON, K., TANG, L. & MOSS, E. G. (2002) Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene lin-28 after translation initiation. *Dev Biol,* 243**,** 215-25.

SEIDL, C., DONNER, H., PETERSHOFEN, E., USADEL, K. H., SEIFRIED, E., KALTWASSER, J. P. & BADENHOOP, K. (1999) An endogenous retroviral long terminal repeat at the HLA-DQB1 gene locus confers susceptibility to rheumatoid arthritis. *Hum Immunol,* 60**,** 63-8.

SEIFARTH, W., BAUST, C., MURR, A., SKLADNY, H., KRIEG-SCHNEIDER, F., BLUSCH, J., WERNER, T., HEHLMANN, R. & LEIB-MOSCH, C. (1998) Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. *J Virol,* 72**,** 8384-91.

SEIFARTH, W., FRANK, O., ZEILFELDER, U., SPIESS, B., GREENWOOD, A. D., HEHLMANN, R. & LEIB-MOSCH, C. (2005) Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *J Virol,* 79**,** 341-52.

SEKINE, T., MASUKO-HONGO, K., MATSUI, T., ASAHARA, H., TAKIGAWA, M., NISHIOKA, K. & KATO, T. (2001) Recognition of YKL-39, a human cartilage related protein, as a target antigen in patients with rheumatoid arthritis. *Ann Rheum Dis,* 60**,** 49-54.

SERRA, C., MAMELI, G., ARRU, G., SOTGIU, S., ROSATI, G. & DOLEI, A. (2003) In vitro modulation of the multiple sclerosis (MS)-associated retrovirus by cytokines: implications for MS pathogenesis. *J Neurovirol,* 9**,** 637-43.

SHAHINIAN, A., PFEFFER, K., LEE, K. P., KUNDIG, T. M., KISHIHARA, K., WAKEHAM, A., KAWAI, K., OHASHI, P. S., THOMPSON, C. B. & MAK, T. W. (1993) Differential T cell costimulatory requirements in CD28-deficient mice. *Science,* 261**,** 609-12.

SHAMSUDDIN, A. M. & HARRIS, C. C. (1983) Improved enzyme immunoassays using biotin-avidin-enzyme complex. *Arch Pathol Lab Med,* 107**,** 514-7.

SHEEHY, A. M., GADDIS, N. C., CHOI, J. D. & MALIM, M. H. (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature,* 418**,** 646-50.

SHEN, G. Q., SHOENFELD, Y. & PETER, J. B. (1998) Anti-DNA, antihistone, and antinucleosome antibodies in systemic lupus erythematosus and drug-induced lupus. *Clin Rev Allergy Immunol,* 16**,** 321-34.

SHLOMCHIK, M., NEMAZEE, D., VAN SNICK, J. & WEIGERT, M. (1987a) Variable region sequences of murine IgM anti-IgG monoclonal autoantibodies (rheumatoid factors). II. Comparison of hybridomas derived by lipopolysaccharide stimulation and secondary protein immunization. *J Exp Med,* 165**,** 970-87.

SHLOMCHIK, M. J., MARSHAK-ROTHSTEIN, A., WOLFOWICZ, C. B., ROTHSTEIN, T. L. & WEIGERT, M. G. (1987b) The role of clonal selection and somatic mutation in autoimmunity. *Nature,* 328**,** 805-11.

SHRAKE, A. & RUPLEY, J. A. (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol,* 79**,** 351-71.

SICAT, J., SUTKOWSKI, N. & HUBER, B. T. (2005) Expression of human endogenous retrovirus HERV-K18 superantigen is elevated in juvenile rheumatoid arthritis. *J Rheumatol,* 32**,** 1821-31.

SILMAN, A. J. (1992) Pregnancy and scleroderma. *Am J Reprod Immunol,* 28**,** 238-40.

SIMON, J. H., GADDIS, N. C., FOUCHIER, R. A. AND MALIM, M. H. (1998) Evidence for a newly discovered cellular anti-HIV-1 phenotype. *Nat Med,* 4, 1397-400.

SIMON, M., GIRBAL, E., SEBBAG, M., GOMES-DAUDRIX, V., VINCENT, C., SALAMA, G. & SERRE, G. (1993) The cytokeratin filament-aggregating protein filaggrin is the target of the so-called "antikeratin antibodies," autoantibodies specific for rheumatoid arthritis. *J Clin Invest,* 92**,** 1387-93.
SJOTTEM, E., ANDERSSEN, S. & JOHANSEN, T. (1996) The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3' to the TATA box. *J Virol,* 70**,** 188-98.
SMALHEISER, N. R. & TORVIK, V. I. (2005) Mammalian microRNAs derived from genomic repeats. *Trends Genet,* 21**,** 322-6.
SMALHEISER, N. R. & TORVIK, V. I. (2006) Alu elements within human mRNAs are probable microRNA targets. *Trends Genet,* 22**,** 532-6.
SMETS, F. & SOKAL, E. M. (2002) Lymphoproliferation in children after liver transplantation. *J Pediatr Gastroenterol Nutr,* 34**,** 499-505.
SMITH, J. M. & HAIGH, J. (1974) The hitch-hiking effect of a favourable gene. *Genet Res,* 23**,** 23-35.
SMITH, K. A., NELSON, P. N., WARREN, P., ASTLEY, S. J., MURRAY, P. G. & GREENMAN, J. (2004) Demystified...recombinant antibodies. *J Clin Pathol,* 57**,** 912-7.
SMOLEN, J. S. & STEINER, G. (2003) Therapeutic strategies for rheumatoid arthritis. *Nat Rev Drug Discov,* 2**,** 473-88.
SONVEAUX, P., BROUET, A., HAVAUX, X., GREGOIRE, V., DESSY, C., BALLIGAND, J. L. & FERON, O. (2003) Irradiation-induced angiogenesis through the up-regulation of the nitric oxide pathway: implications for tumor radiotherapy. *Cancer Res,* 63**,** 1012-9.
SPECTOR, T. D. & HOCHBERG, M. C. (1990) The protective effect of the oral contraceptive pill on rheumatoid arthritis: an overview of the analytic epidemiological studies using meta-analysis. *J Clin Epidemiol,* 43**,** 1221-30.
SPENCER, T. E., JOHNSON, G. A., BAZER, F. W., BURGHARDT, R. C. & PALMARINI, M. (2007) Pregnancy recognition and conceptus implantation in domestic ruminants: roles of progesterone, interferons and endogenous retroviruses. *Reprod Fertil Dev,* 19**,** 65-78.
STANFIELD, R. L., FIESER, T. M., LERNER, R. A. & WILSON, I. A. (1990) Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 A. *Science,* 248**,** 712-9.
STAUFFER, Y., THEILER, G., SPERISEN, P., LEBEDEV, Y. & JONGENEEL, C. V. (2004) Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun,* 4**,** 2.
STEENBAKKERS, P. G., BAETEN, D., ROVERS, E., VEYS, E. M., RIJNDERS, A. W., MEIJERINK, J., DE KEYSER, F. & BOOTS, A. M. (2003) Localization of MHC class II/human cartilage glycoprotein-39 complexes in synovia of rheumatoid arthritis patients using complex-specific monoclonal antibodies. *J Immunol,* 170**,** 5719-27.
STEFANI, G. (2007) Roles of microRNAs and their targets in cancer. *Expert Opin Biol Ther,* 7**,** 1833-40.
STEIGMANN, F., DOURDOUREKAS, D., SHOBASSAY, N., VITTAL, S. B., SZANTO, P. B., AINIS, H., KHIN, U. & TELISCHI, M. (1974) Humoral response in patients with liver disease. *Am J Gastroenterol,* 61**,** 349-55.
STOLT, P., BENGTSSON, C., NORDMARK, B., LINDBLAD, S., LUNDBERG, I., KLARESKOG, L. & ALFREDSSON, L. (2003) Quantification of the influence

of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis,* 62**,** 835-41.
STOYE, J. P. (1999) The pathogenic potential of endogenous retroviruses: a sceptical view. *Trends Microbiol,* 7**,** 430; author reply 431-2.
STOYE, J. P. (2002) An intracellular block to primate lentivirus replication. *Proc Natl Acad Sci U S A,* 99**,** 11549-51.
STRANSKY, G., VERNON, J., AICHER, W. K., MORELAND, L. W., GAY, R. E. & GAY, S. (1993) Virus-like particles in synovial fluids from patients with rheumatoid arthritis. *Br J Rheumatol,* 32**,** 1044-8.
STREMLAU, M., OWENS, C. M., PERRON, M. J., KIESSLING, M., AUTISSIER, P. & SODROSKI, J. (2004) The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature,* 427**,** 848-53.
STREMLAU, M., PERRON, M., LEE, M., LI, Y., SONG, B., JAVANBAKHT, H., DIAZ-GRIFFERO, F., ANDERSON, D. J., SUNDQUIST, W. I. & SODROSKI, J. (2006) Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc Natl Acad Sci U S A,* 103**,** 5514-9.
STRYNADKA, N. C., REDMOND, M. J., PARKER, J. M., SCRABA, D. G. & HODGES, R. S. (1988) Use of synthetic peptides to map the antigenic determinants of glycoprotein D of herpes simplex virus. *J Virol,* 62**,** 3474-83.
SUN, Y., ZENG, X. R., WENGER, L., FIRESTEIN, G. S. AND CHEUNG, H. S. (2004) P53 down-regulates matrix metalloproteinase-1 by targeting the communications between AP-1 and the basal transcription complex. *J Cell Biochem,* **92,** 258-69.
SUNDBERG, E. J., LI, Y. & MARIUZZA, R. A. (2002) So many ways of getting in the way: diversity in the molecular architecture of superantigen-dependent T-cell signaling complexes. *Curr Opin Immunol,* 14**,** 36-44.
SUNDBERG, E. J. & MARIUZZA, R. A. (2002) Molecular recognition in antibody-antigen complexes. *Adv Protein Chem,* 61**,** 119-60.
SUTKOWSKI, N., CHEN, G., CALDERON, G. & HUBER, B. T. (2004) Epstein-Barr virus latent membrane protein LMP-2A is sufficient for transactivation of the human endogenous retrovirus HERV-K18 superantigen. *J Virol,* 78**,** 7852-60.
SUTKOWSKI, N., CONRAD, B., THORLEY-LAWSON, D. A. & HUBER, B. T. (2001) Epstein-Barr virus transactivates the human endogenous retrovirus HERV-K18 that encodes a superantigen. *Immunity,* 15**,** 579-89.
SUZUKI, A., YAMADA, R., CHANG, X., TOKUHIRO, S., SAWADA, T., SUZUKI, M., NAGASAKI, M., NAKAYAMA-HAMADA, M., KAWAIDA, R., ONO, M., OHTSUKI, M., FURUKAWA, H., YOSHINO, S., YUKIOKA, M., TOHMA, S., MATSUBARA, T., WAKITANI, S., TESHIMA, R., NISHIOKA, Y., SEKINE, A., IIDA, A., TAKAHASHI, A., TSUNODA, T., NAKAMURA, Y. & YAMAMOTO, K. (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet,* 34**,** 395-402.
SUZUKI, M., TETSUKA, T., YOSHIDA, S., WATANABE, N., KOBAYASHI, M., MATSUI, N. & OKAMOTO, T. (2000) The role of p38 mitogen-activated protein kinase in IL-6 and IL-8 production from the TNF-alpha- or IL-1beta-stimulated rheumatoid synovial fibroblasts. *FEBS Lett,* 465**,** 23-7.
SVENSSON, A. C. & ANDERSSON, G. (1997) Presence of retroelements reveal the evolutionary history of the human DR haplotypes. *Hereditas,* 127**,** 113-24.

SWEENEY, S. E. & FIRESTEIN, G. S. (2004) Rheumatoid arthritis: regulation of synovial inflammation. *Int J Biochem Cell Biol,* 36**,** 372-8.
SZEKANECZ, Z. & KOCH, A. E. (2000) Cell-cell interactions in synovitis. Endothelial cells and immune cell migration. *Arthritis Res,* 2**,** 368-73.
SZEKANECZ, Z. & KOCH, A. E. (2001) Chemokines and angiogenesis. *Curr Opin Rheumatol,* 13**,** 202-8.
TAKAYANAGI, H., IIZUKA, H., JUJI, T., NAKAGAWA, T., YAMAMOTO, A., MIYAZAKI, T., KOSHIHARA, Y., ODA, H., NAKAMURA, K. & TANAKA, S. (2000) Involvement of receptor activator of nuclear factor kappaB ligand/osteoclast differentiation factor in osteoclastogenesis from synoviocytes in rheumatoid arthritis. *Arthritis Rheum,* 43**,** 259-69.
TAKEI, M., MITAMURA, K., FUJIWARA, S., HORIE, T., RYU, J., OSAKA, S., YOSHINO, S. & SAWADA, S. (1997) Detection of Epstein-Barr virus-encoded small RNA 1 and latent membrane protein 1 in synovial lining cells from rheumatoid arthritis patients. *Int Immunol,* 9**,** 739-43.
TAKEMURA, S., KLIMIUK, P. A., BRAUN, A., GORONZY, J. J. & WEYAND, C. M. (2001) T cell activation in rheumatoid synovium is B cell dependent. *J Immunol,* 167**,** 4710-8.
TAKEUCHI, K., KATSUMATA, K., IKEDA, H., MINAMI, M., WAKISAKA, A. & YOSHIKI, T. (1995) Expression of endogenous retroviruses, ERV3 and lambda 4-1, in synovial tissues from patients with rheumatoid arthritis. *Clin Exp Immunol,* 99**,** 338-44.
TALAL, N., GARRY, R. F., SCHUR, P. H., ALEXANDER, S., DAUPHINEE, M. J., LIVAS, I. H., BALLESTER, A., TAKEI, M. & DANG, H. (1990) A conserved idiotype and antibodies to retroviral proteins in systemic lupus erythematosus. *J Clin Invest,* 85**,** 1866-71.
TAM, J. P. & ZAVALA, F. (1989) Multiple antigen peptide. A novel approach to increase detection sensitivity of synthetic peptides in solid-phase immunoassays. *J Immunol Methods,* 124**,** 53-61.
TAN, E. M. (1989) Do autoantibodies inhibit function of their cognate antigens in vivo? *Arthritis Rheum,* 32**,** 924-5.
TANI, Y., TIWANA, H., HUKUDA, S., NISHIOKA, J., FIELDER, M., WILSON, C., BANSAL, S. & EBRINGER, A. (1997) Antibodies to Klebsiella, Proteus, and HLA-B27 peptides in Japanese patients with ankylosing spondylitis and rheumatoid arthritis. *J Rheumatol,* 24**,** 109-14.
TARLINTON, R. E., MEERS, J. & YOUNG, P. R. (2006) Retroviral invasion of the koala genome. *Nature,* 442**,** 79-81.
TARUSCIO, D. & MANTOVANI, A. (1998) Human endogenous retroviral sequences: possible roles in reproductive physiopathology. *Biol Reprod,* 59**,** 713-24.
TARUSCIO, D. & MANUELIDIS, L. (1991) Integration site preferences of endogenous retroviruses. *Chromosoma,* 101**,** 141-56.
TASSABEHJI, M., STRACHAN, T., ANDERSON, M., CAMPBELL, R. D., COLLIER, S. & LAKO, M. (1994) Identification of a novel family of human endogenous retroviruses and characterization of one family member, HERV-K(C4), located in the complement C4 gene cluster. *Nucleic Acids Res,* 22**,** 5211-7.
TEMIN, H. M. & MIZUTANI, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature,* 226**,** 1211-3.
TER STEEGE, J., VIANEN, M., VAN BILSEN, J., BIJLSMA, J., LAFEBER, F. & WAUBEN, M. (2003) Identification of self-epitopes recognized by T cells in

rheumatoid arthritis demonstrates matrix metalloproteinases as a novel T cell target. *J Rheumatol,* 30**,** 1147-56.
TERATO, K., HARPER, D. S., GRIFFITHS, M. M., HASTY, D. L., YE, X. J., CREMER, M. A. & SEYER, J. M. (1995) Collagen-induced arthritis in mice: synergistic effect of E. coli lipopolysaccharide bypasses epitope specificity in the induction of arthritis with monoclonal antibodies to type II collagen. *Autoimmunity,* 22**,** 137-47.
TIAN, J., ATKINSON, M. A., CLARE-SALZLER, M., HERSCHENFELD, A., FORSTHUBER, T., LEHMANN, P. V. AND KAUFMAN, D. L. (1996) Nasal administration of glutamate decarboxylase (GAD65) peptides induces Th2 responses and prevents murine insulin-dependent diabetes. *J Exp Med,* 183, 1561-7.
TING, C. N., ROSENBERG, M. P., SNOW, C. M., SAMUELSON, L. C. & MEISLER, M. H. (1992) Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev,* 6, 1457-65.
TISCH, R., YANG, X. D., SINGER, S. M., LIBLAU, R. S., FUGGER, L. AND MCDEVITT, H. O. (1993) Immune response to glutamic acid decarboxylase correlates with insulitis in non-obese diabetic mice. *Nature,* 366, 72-5.
TORMO, J., BLAAS, D., PARRY, N. R., ROWLANDS, D., STUART, D. & FITA, I. (1994) Crystal structure of a human rhinovirus neutralizing antibody complexed with a peptide derived from viral capsid protein VP2. *Embo J,* 13**,** 2247-56.
TOWERS, G. J. (2007) The control of viral infection by tripartite motif proteins and cyclophilin A. *Retrovirology*, 4, 40.
TRABANDT, A., GAY, R. E., FASSBENDER, H. G. & GAY, S. (1991) Cathepsin B in synovial cells at the site of joint destruction in rheumatoid arthritis. *Arthritis Rheum,* 34**,** 1444-51.
TRIBOULET, R., MARI, B., LIN, Y. L., CHABLE-BESSIA, C., BENNASSER, Y., LEBRIGAND, K., CARDINAUD, B., MAURIN, T., BARBRY, P., BAILLAT, V., REYNES, J., CORBEAU, P., JEANG, K. T. & BENKIRANE, M. (2007) Suppression of microRNA-silencing pathway by HIV-1 during virus replication. *Science,* 315**,** 1579-82.
TRISTEM, M. (2000) Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J Virol,* 74, 3715-30.
TURNER, B. G. AND SUMMERS, M. F. (1999) Structural biology of HIV. *J Mol Biol,* 285, 1-32.
TURNER, G., BARBULESCU, M., SU, M., JENSEN-SEAMAN, M. I., KIDD, K. K. & LENZ, J. (2001) Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr Biol,* 11**,** 1531-5.
TYAGI, S. & KRAMER, F. R. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol,* 14**,** 303-8.
TZIOUFAS, A. G., TZORTZAKIS, N. G., PANOU-POMONIS, E., BOKI, K. A., SAKARELLOS-DAITSIOTIS, M., SAKARELLOS, C. & MOUTSOPOULOS, H. M. (2000) The clinical relevance of antibodies to ribosomal-P common epitope in two targeted systemic lupus erythematosus populations: a large cohort of consecutive patients and patients with active central nervous system disease. *Ann Rheum Dis,* 59**,** 99-104.
TZIOUFAS, A. G., WASSMUTH, R., DAFNI, U. G., GUIALIS, A., HAGA, H. J., ISENBERG, D. A., JONSSON, R., KALDEN, J. R., KIENER, H.,

SAKARELLOS, C., SMOLEN, J. S., SUTCLIFFE, N., VITALI, C., YIANNAKI, E. & MOUTSOPOULOS, H. M. (2002) Clinical, immunological, and immunogenetic aspects of autoantibody production against Ro/SSA, La/SSB and their linear epitopes in primary Sjogren's syndrome (pSS): a European multicentre study. *Ann Rheum Dis,* 61**,** 398-404.

UEMATSU, S. & AKIRA, S. (2006) Toll-like receptors and innate immunity. *J Mol Med,* 84**,** 712-25.

UEMATSU, Y., WEGE, H., STRAUS, A., OTT, M., BANNWARTH, W., LANCHBURY, J., PANAYI, G. & STEINMETZ, M. (1991) The T-cell-receptor repertoire in the synovial fluid of a patient with rheumatoid arthritis is polyclonal. *Proc Natl Acad Sci U S A,* 88**,** 8534-8.

URNOVITZ, H. B. & MURPHY, W. H. (1996) Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease. *Clin Microbiol Rev,* 9**,** 72-99.

VAN BOEKEL, M. A., VOSSENAAR, E. R., VAN DEN HOOGEN, F. H. & VAN VENROOIJ, W. J. (2002) Autoantibody systems in rheumatoid arthritis: specificity, sensitivity and diagnostic value. *Arthritis Res,* 4**,** 87-93.

VAN DE LAGEMAAT, L. N., MEDSTRAND, P. & MAGER, D. L. (2006) Multiple effects govern endogenous retrovirus survival patterns in human gene introns. *Genome Biol,* 7**,** R86.

VAN DER GRAAFF, W. L., PRINS, A. P., NIERS, T. M., DIJKMANS, B. A. & VAN LIER, R. A. (1999) Quantitation of interferon gamma- and interleukin-4-producing T cells in synovial fluid and peripheral blood of arthritis patients. *Rheumatology (Oxford),* 38**,** 214-20.

VAN DER HEIJDE, D. M. (1995) Joint erosions and patients with early rheumatoid arthritis. *Br J Rheumatol,* 34 Suppl 2**,** 74-8.

VAN GENT, D. C., MIZUUCHI, K. & GELLERT, M. (1996) Similarities between initiation of V(D)J recombination and retroviral integration. *Science,* 271**,** 1592-4.

VAN REGENMORTEL, M. H. (2006) Immunoinformatics may lead to a reappraisal of the nature of B cell epitopes and of the feasibility of synthetic peptide vaccines. *J Mol Recognit,* 19**,** 183-7.

VAN WEEMEN, B. K. & SCHUURS, A. H. (1971) Immunoassay using antigen-enzyme conjugates. *FEBS Lett,* 15**,** 232-236.

VANDESOMPELE, J., DE PRETER, K., PATTYN, F., POPPE, B., VAN ROY, N., DE PAEPE, A. & SPELEMAN, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol,* 3**,** RESEARCH0034.

VARGA, A. & JAMES, D. (2006) Real-time RT-PCR and SYBR Green I melting curve analysis for the identification of Plum pox virus strains C, EA, and W: effect of amplicon size, melt rate, and dye translocation. *J Virol Methods,* 132**,** 146-53.

VAUGHAN, J. H., FOX, R. I., ABRESCH, R. J., TSOUKAS, C. D., CURD, J. G. & CARSON, D. A. (1984) Thoracic duct drainage in rheumatoid arthritis. *Clin Exp Immunol,* 58**,** 645-53.

VENKATACHALAM, C. M. (1968) Stereochemical criteria for polypeptides and proteins. VI. Non-bonded energy of polyglycine and poly-L-alanine in the crystalline beta-form. *Biochim Biophys Acta,* 168**,** 411-6.

VERHOEF, C. M., VAN ROON, J. A., VIANEN, M. E., BRUIJNZEEL-KOOMEN, C. A., LAFEBER, F. P. & BIJLSMA, J. W. (1998) Mutual antagonism of

rheumatoid arthritis and hay fever; a role for type 1/type 2 T cell balance. *Ann Rheum Dis,* 57**,** 275-80.
VIGLIANTI, G. A., LAU, C. M., HANLEY, T. M., MIKO, B. A., SHLOMCHIK, M. J. & MARSHAK-ROTHSTEIN, A. (2003) Activation of autoreactive B cells by CpG dsDNA. *Immunity,* 19**,** 837-47.
VIHINEN, M., TORKKILA, E. & RIIKONEN, P. (1994) Accuracy of protein flexibility predictions. *Proteins,* 19**,** 141-9.
VILLARREAL, L. P. (1997) On viruses, sex, and motherhood. *J Virol,* 71**,** 859-65.
VILLESEN, P., AAGAARD, L., WIUF, C. & PEDERSEN, F. S. (2004) Identification of endogenous retroviral reading frames in the human genome. *Retrovirology,* 1**,** 32.
VINEN, C. S. & OLIVEIRA, D. B. (2003) Acute glomerulonephritis. *Postgrad Med J,* 79**,** 206-13; quiz 212-3.
VINOGRADOVA, T., VOLIK, S., LEBEDEV, Y., SHEVCHENKO, Y., LAVRENTYEVA, I., KHIL, P., GRZESCHIK, K. H., ASHWORTH, L. K. & SVERDLOV, E. (1997) Positioning of 72 potentially full size LTRs of human endogenous retroviruses HERV-K on the human chromosome 19 map. Occurrences of the LTRs in human gene sites. *Gene,* 199**,** 255-64.
VOIGT, L. F., KOEPSELL, T. D., NELSON, J. L., DUGOWSON, C. E. & DALING, J. R. (1994) Smoking, obesity, alcohol consumption, and the risk of rheumatoid arthritis. *Epidemiology,* 5**,** 525-32.
VOLLER, A., BIDWELL, D., HULDT, G. & ENGVALL, E. (1974) A microplate method of enzyme-linked immunosorbent assay and its application to malaria. *Bull World Health Organ,* 51**,** 209-11.
VRANCKX, R., MUYLLE, L., COLE, J., MOLDENHASER, R. & PEETERMANS, M. E. (1986) HBV vaccinations in medical and paramedical staff: the impact of age on immunization results. *Vox Sang,* 50**,** 220-2.
WANG-JOHANNING, F., FROST, A. R., JIAN, B., EPP, L., LU, D. W. & JOHANNING, G. L. (2003) Quantitation of HERV-K env gene expression and splicing in human breast cancer. *Oncogene,* 22**,** 1528-35.
WANG, C. L., WU, Y. T., LEE, C. J., LIU, H. C., HUANG, L. T. & YANG, K. D. (2002a) Decreased nitric oxide production after intravenous immunoglobulin treatment in patients with Kawasaki disease. *J Pediatr,* 141**,** 560-5.
WANG, J., LIU, X., LI, F. & YAO, L. (2002b) Rheumatoid arthritis is auto-immunoreaction to collagen II in cartilage happened in synovial tissue. *Med Hypotheses,* 59**,** 411-5.
WATSON, J. D. & CRICK, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature,* 171**,** 737-8.
WEIDHAAS, J. B., ANGELICHIO, E. L., FENNER, S. & COFFIN, J. M. (2000) Relationship between retroviral DNA integration and gene expression. *J Virol,* 74**,** 8382-9.
WEISS, R. A. (2006) The discovery of endogenous retroviruses. *Retrovirology,* 3**,** 67.
WEISS, R. A., FRIIS, R. R., KATZ, E. & VOGT, P. K. (1971) Induction of avian tumor viruses in normal cells by physical and chemical carcinogens. *Virology,* 46**,** 920-38.
WELLCOME TRUST CASE CONTROL CONSORTIUM (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* 447**,** 661-78.

WELLING, G. W., WEIJER, W. J., VAN DER ZEE, R. & WELLING-WESTER, S. (1985) Prediction of sequential antigenic regions in proteins. *FEBS Lett,* 188**,** 215-8.
WESTHOF, E., ALTSCHUH, D., MORAS, D., BLOOMER, A. C., MONDRAGON, A., KLUG, A. & VAN REGENMORTEL, M. H. (1984) Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature,* 311**,** 123-6.
WESTWOOD, O. M., NELSON, P. N. & HAY, F. C. (2006) Rheumatoid factors: what's new? *Rheumatology (Oxford),* 45**,** 379-85.
WEYAND, C. M. (2000) New insights into the pathogenesis of rheumatoid arthritis. *Rheumatology (Oxford),* 39 Suppl 1**,** 3-8.
WEYAND, C. M., HICOK, K. C., CONN, D. L. & GORONZY, J. J. (1992) The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis. *Ann Intern Med,* 117**,** 801-6.
WEYAND, C. M., KLIMIUK, P. A. & GORONZY, J. J. (1998) Heterogeneity of rheumatoid arthritis: from phenotypes to genotypes. *Springer Semin Immunopathol,* 20**,** 5-22.
WHITACRE, C. C., REINGOLD, S. C. & O'LOONEY, P. A. (1999) A gender gap in autoimmunity. *Science,* 283**,** 1277-8.
WHITTON, J. L., TISHON, A., LEWICKI, H., GEBHARD, J., COOK, T., SALVATO, M., JOLY, E. & OLDSTONE, M. B. (1989) Molecular analyses of a five-amino-acid cytotoxic T-lymphocyte (CTL) epitope: an immunodominant region which induces nonreciprocal CTL cross-reactivity. *J Virol,* 63**,** 4303-10.
WILDER, R. L. & ELENKOV, I. J. (1999) Hormonal regulation of tumor necrosis factor-alpha, interleukin-12 and interleukin-10 production by activated macrophages. A disease-modifying mechanism in rheumatoid arthritis and systemic lupus erythematosus? *Ann N Y Acad Sci,* 876, 14-31.
WILLIAMS, D. G., SHARPE, N. G., WALLACE, G. & LATCHMAN, D. S. (1990) A repeated proline-rich sequence in Sm B/B' and N is a dominant epitope recognized by human and murine autoantibodies. *J Autoimmun,* 3**,** 715-25.
WILLS, J. W. & CRAVEN, R. C. (1991) Form, function, and use of retroviral gag proteins. *Aids,* 5**,** 639-54.
WINCHESTER, R. J. & GREGERSEN, P. K. (1988) The molecular basis of susceptibility to rheumatoid arthritis: the conformational equivalence hypothesis. *Springer Semin Immunopathol,* 10**,** 119-39.
WINER, J., JUNG, C. K., SHACKEL, I. & WILLIAMS, P. M. (1999) Development and validation of real-time quantitative reverse transcriptase-polymerase chain reaction for monitoring gene expression in cardiac myocytes in vitro. *Anal Biochem,* 270**,** 41-9.
WITTWER, C. T., HERRMANN, M. G., GUNDRY, C. N. & ELENITOBA-JOHNSON, K. S. (2001) Real-time multiplex PCR assays. *Methods,* 25**,** 430-42.
WOOLEY, P. H., LUTHRA, H. S., SINGH, S. K., HUSE, A. R., STUART, J. M. & DAVID, C. S. (1984) Passive transfer of arthritis to mice by injection of human anti-type II collagen antibody. *Mayo Clin Proc,* 59**,** 737-43.
WRIGHT, D. A. & VOYTAS, D. F. (2002) Athila4 of Arabidopsis and Calypso of soybean define a lineage of endogenous plant retroviruses. *Genome Res,* 12**,** 122-31.
WUCHERPFENNIG, K. W. (2001) Structural basis of molecular mimicry. *J Autoimmun,* 16**,** 293-302.

XIONG, Y. & EICKBUSH, T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *Embo J,* 9**,** 3353-62.
YANG, L., HAKODA, M., IWABUCHI, K., TAKEDA, T., KOIKE, T., KAMATANI, N. & TAKADA, K. (2004) Rheumatoid factors induce signaling from B cells, leading to Epstein-Barr virus and B-cell activation. *J Virol,* 78**,** 9918-23.
YEAST GENOME DIRECTORY (1997) The yeast genome directory. *Nature,* 387**,** 5.
YIANNAKI, E. E., TZIOUFAS, A. G., BACHMANN, M., HANTOUMI, J., TSIKARIS, V., SAKARELLOS-DAITSIOTIS, M., SAKARELLOS, C. & MOUTSOPOULOS, H. M. (1998) The value of synthetic linear epitope analogues of La/SSB for the detection of autoantibodies to La/SSB; specificity, sensitivity and comparison of methods. *Clin Exp Immunol,* 112**,** 152-8.
YIN, J. L., SHACKEL, N. A., ZEKRY, A., MCGUINNESS, P. H., RICHARDS, C., PUTTEN, K. V., MCCAUGHAN, G. W., ERIS, J. M. & BISHOP, G. A. (2001) Real-time reverse transcriptase-polymerase chain reaction (RT-PCR) for measurement of cytokine and growth factor mRNA expression with fluorogenic probes or SYBR Green I. *Immunol Cell Biol,* 79**,** 213-21.
YODER, J. A., WALSH, C. P. & BESTOR, T. H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet,* 13**,** 335-40.
YU, M., JOHNSON, J. M. & TUOHY, V. K. (1996) A predictable sequential determinant spreading cascade invariably accompanies progression of experimental autoimmune encephalomyelitis: a basis for peptide-specific therapy after onset of clinical disease. *J Exp Med,* 183**,** 1777-88.
YU, X., YU, Y., LIU, B., LOU, K., KONG, W., MAO, P. AND YU, X. F. (2003) Induction of APOBEC3G ubiquitination and degraded by an HIV-1-Cul5-SCF complex. *Science*, 302, 1056-60.
ZANDMAN-GODDARD, G. & SHOENFELD, Y. (2002) HIV and autoimmunity. *Autoimmun Rev,* 1**,** 329-37.
ZARE, F., BOKAREWA, M., NENONEN, N., BERGSTROM, T., ALEXOPOULOU, L., FLAVELL, R. A. & TARKOWSKI, A. (2004) Arthritogenic properties of double-stranded (viral) RNA. *J Immunol,* 172**,** 5656-63.
ZHOU, Z. & MENARD, H. A. (2002) Autoantigenic posttranslational modifications of proteins: does it apply to rheumatoid arthritis? *Curr Opin Rheumatol,* 14**,** 250-3.
ZIEGLER, B., GAY, R. E., HUANG, G. Q., FASSBENDER, H. G. & GAY, S. (1989) Immunohistochemical localization of HTLV-I p19- and p24-related antigens in synovial joints of patients with rheumatoid arthritis. *Am J Pathol,* 135**,** 1-5.
ZOU, J., ZHANG, Y., THIEL, A., RUDWALEIT, M., SHI, S. L., RADBRUCH, A., POOLE, R., BRAUN, J. & SIEPER, J. (2003) Predominant cellular immune response to the cartilage autoantigenic G1 aggrecan in ankylosing spondylitis and rheumatoid arthritis. *Rheumatology (Oxford),* 42**,** 846-55.
ZSIROS, J., JEBBINK, M. F., LUKASHOV, V. V., VOUTE, P. A. & BERKHOUT, B. (1999) Biased nucleotide composition of the genome of HERV-K related endogenous retroviruses and its evolutionary implications. *J Mol Evol,* 48**,** 102-11.
ZUURBIER, C. J. & HUIJING, P. A. (1993) Changes in geometry of actively shortening unipennate rat gastrocnemius muscle. *J Morphol,* 218**,** 167-80.
ZVAIFLER, N. J. (1973) The immunopathology of joint inflammation in rheumatoid arthritis. *Adv Immunol,* 16**,** 265-336.
ZWERINA, J., HAYER, S., TOHIDAST-AKRAD, M., BERGMEISTER, H., REDLICH, K., FEIGE, U., DUNSTAN, C., KOLLIAS, G., STEINER, G., SMOLEN, J. & SCHETT, G. (2004) Single and combined inhibition of tumor

necrosis factor, interleukin-1, and RANKL pathways in tumor necrosis factor-induced arthritis: effects on synovial inflammation, bone erosion, and cartilage destruction. *Arthritis Rheum,* 50**,** 277-90.

Appendix I:
HERV/Autoantigen Epitope maps

HERV Epitope maps

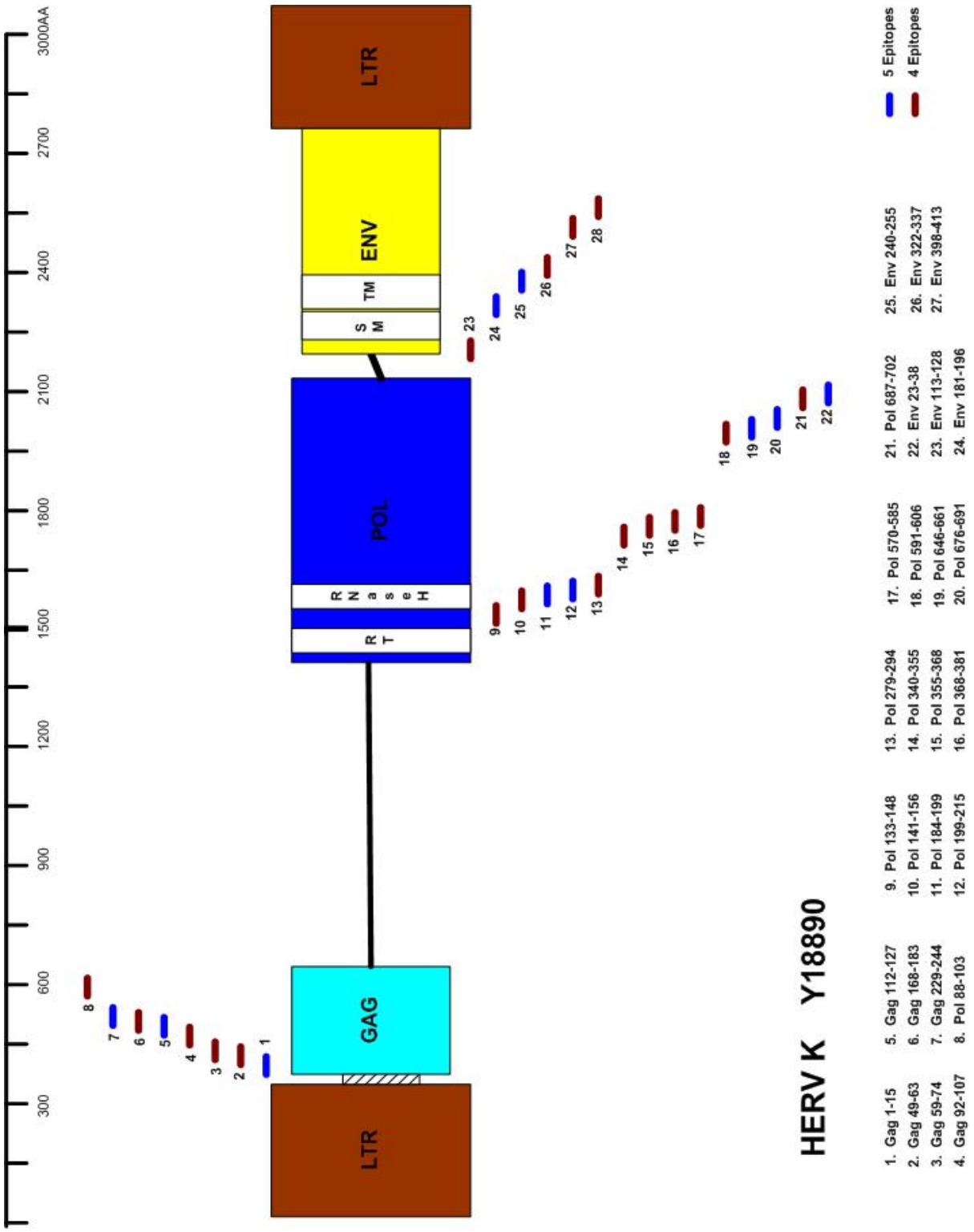


Figure APP1.1. Epitope map for HERV-K (Y168890)

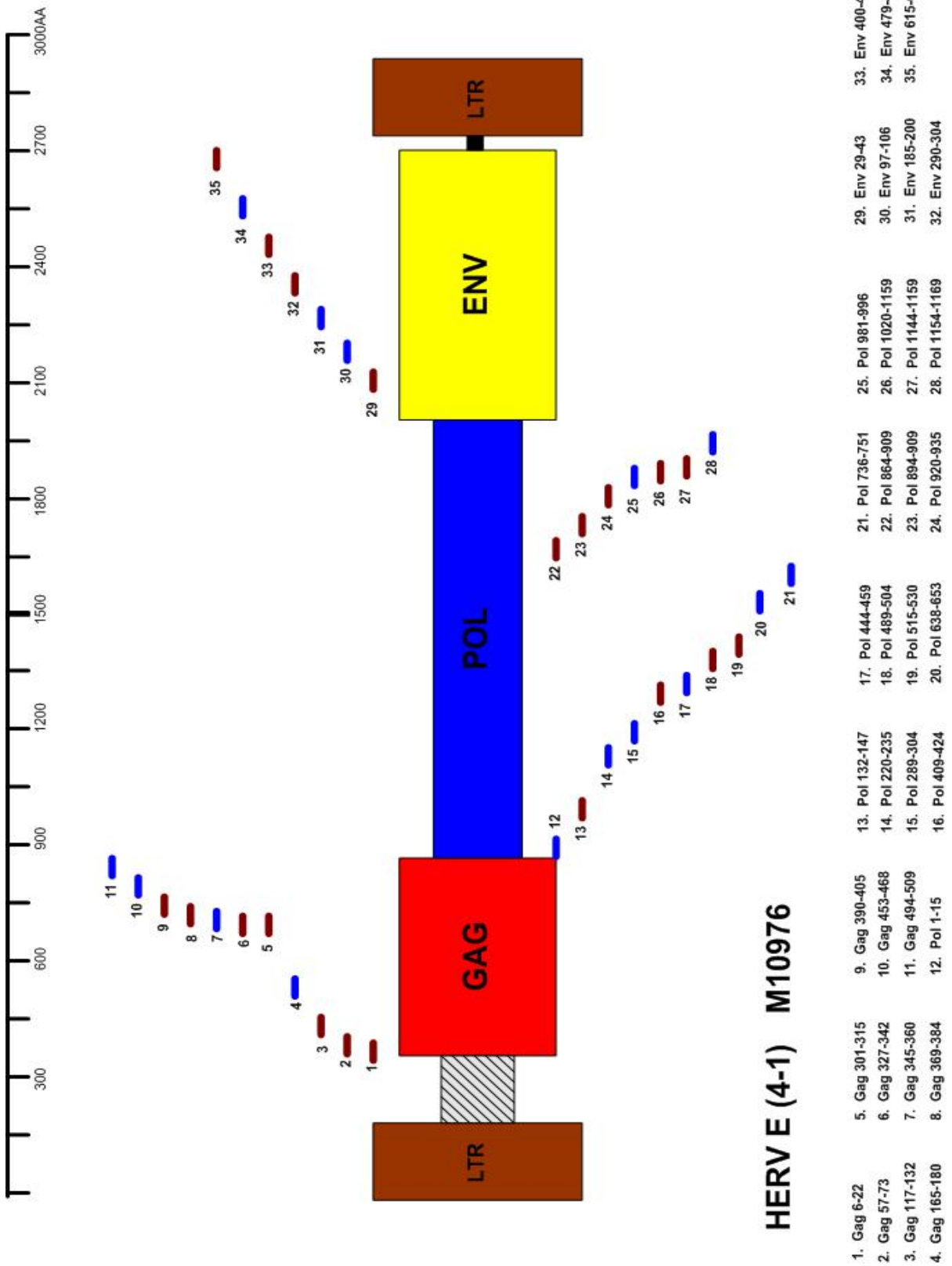


Figure APP1.2. Epitope map for HERV-E (4-1) (M10976)

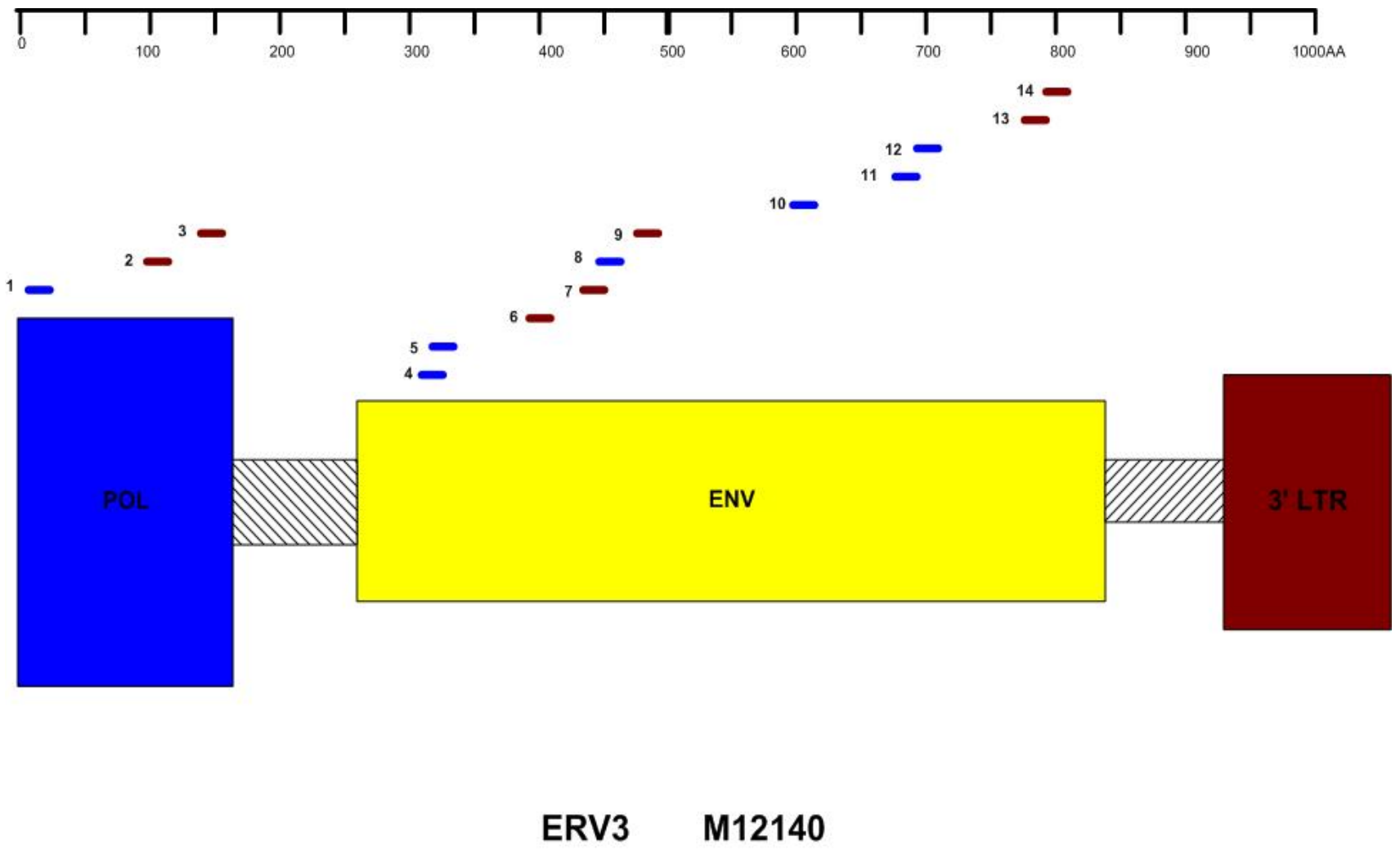


Figure APP1.3. Epitope map for ERV3 (M12140)

Autoantigen epitope maps

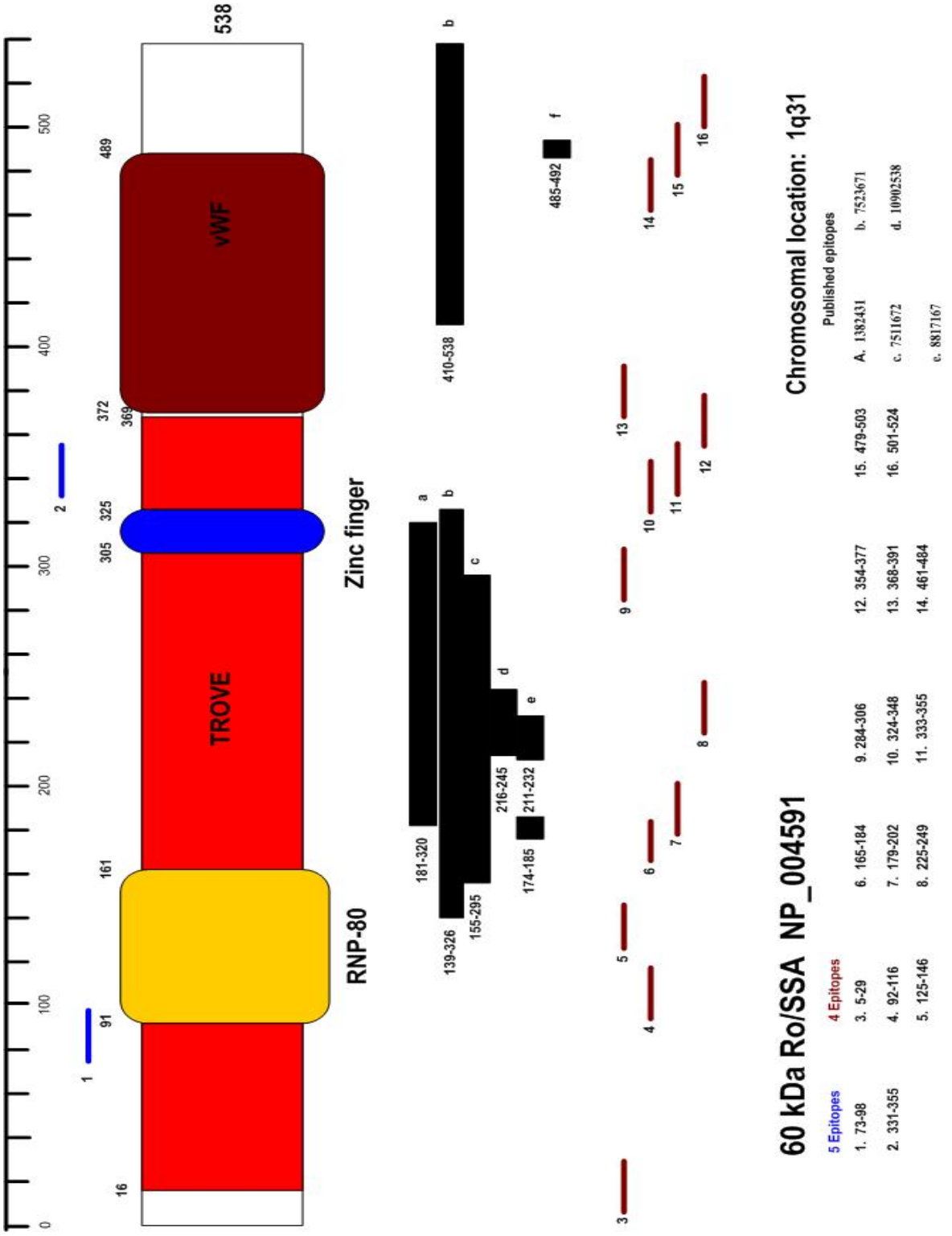


Figure APP1.4. Epitope map for Ro/SSA 60 (NP_004591)

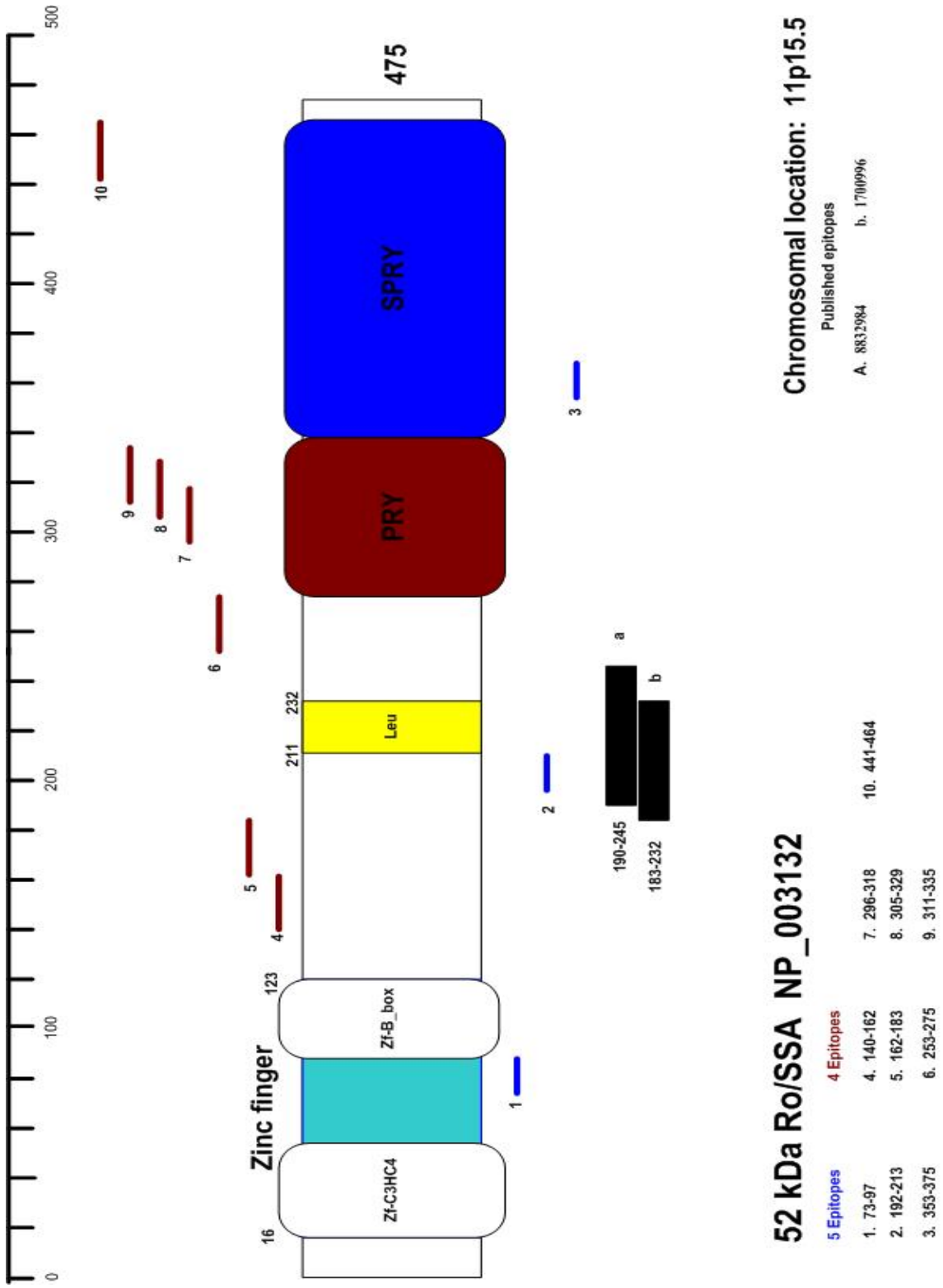


Figure APP1.5. Epitope map for Ro/SSA 52 (NP_003132)

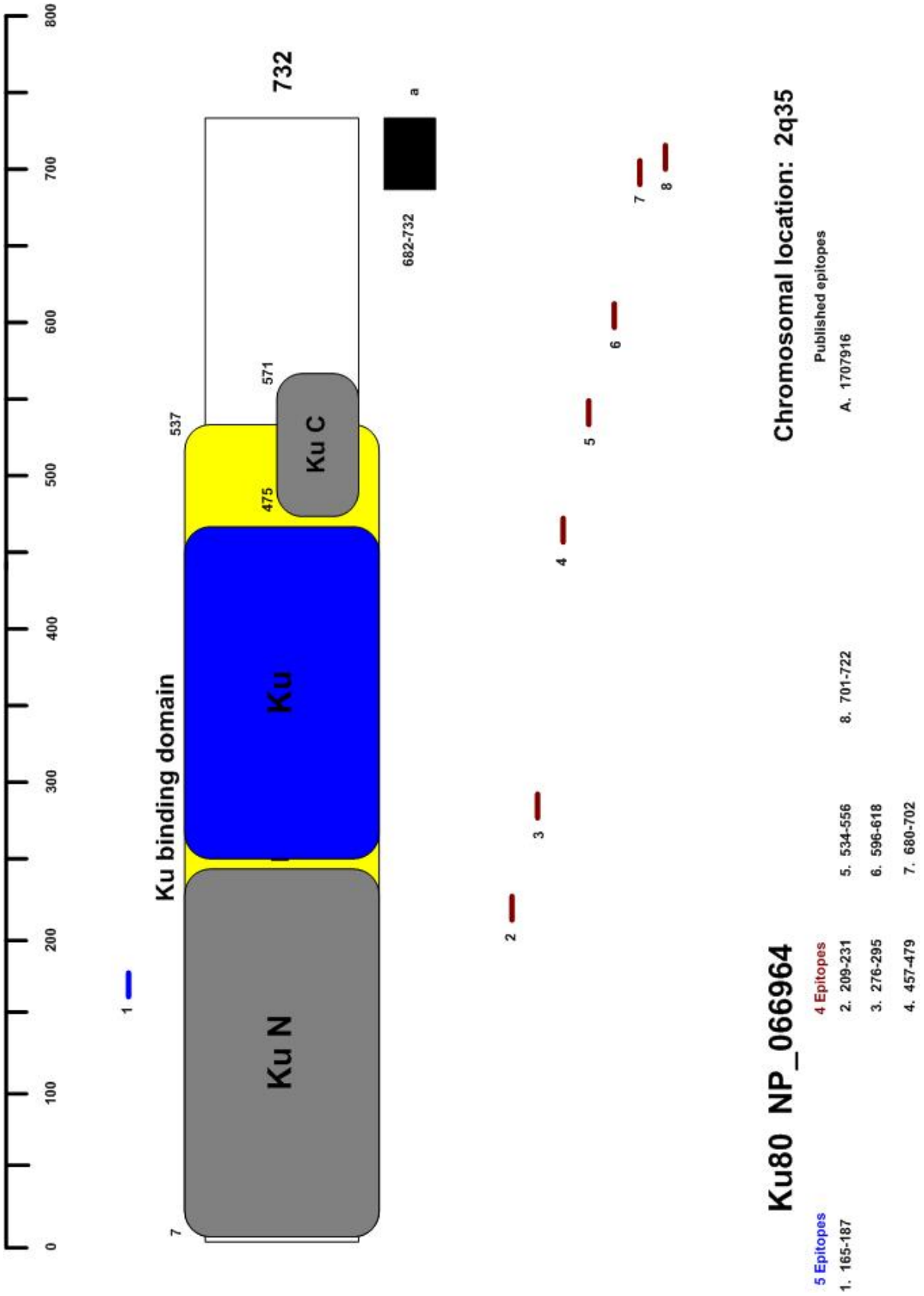


Figure APP1.6. Epitope map for Ku80 (NP_066964)

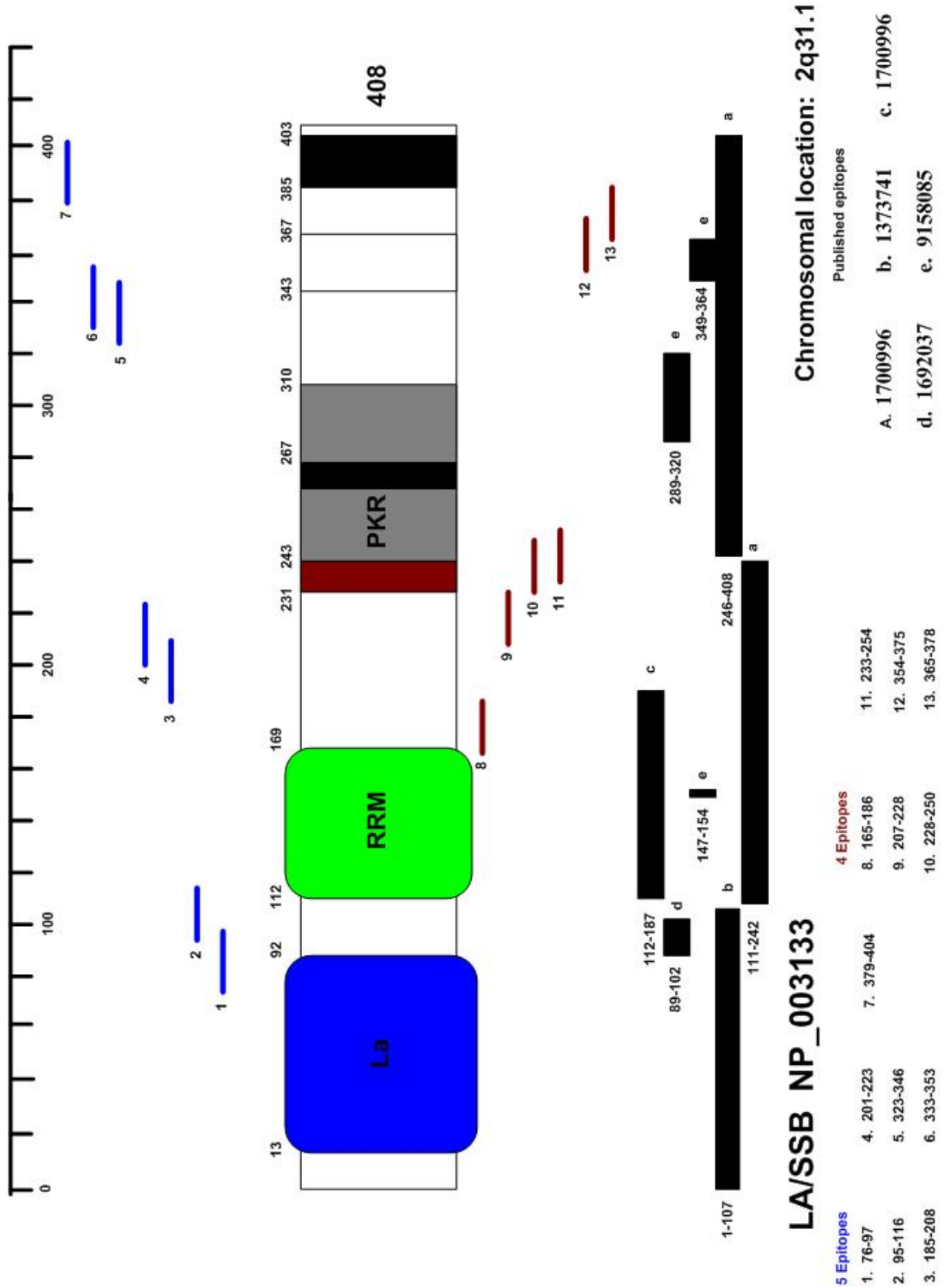


Figure APP1.7. Epitope map for La/SSB (NP_003133)

Appendix *II:*

Full ELISA results for all diseases tested - CFS.

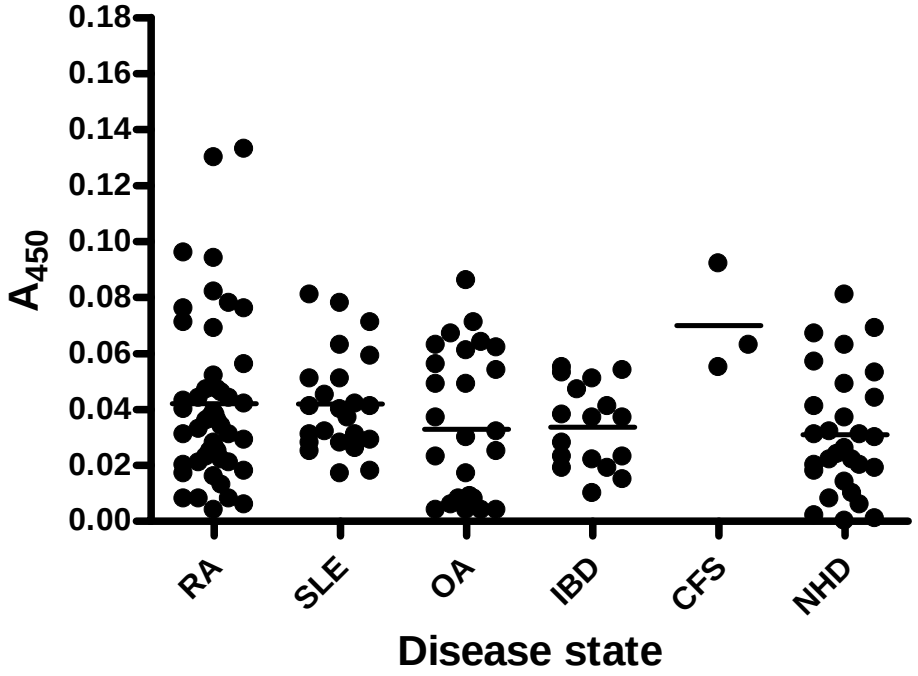


Figure APP2.1 OD values for patient sera from different disease groups using non-biotinylated GfPN1kpr peptide.

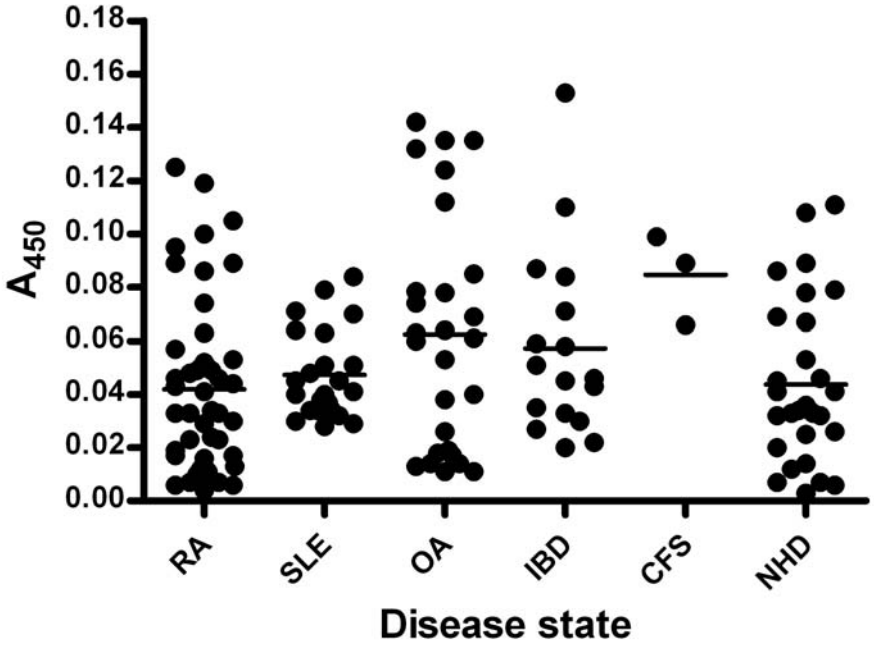


Figure APP2.2 OD values for patient sera from different disease groups using non-biotinylated GfPN2eip peptide.

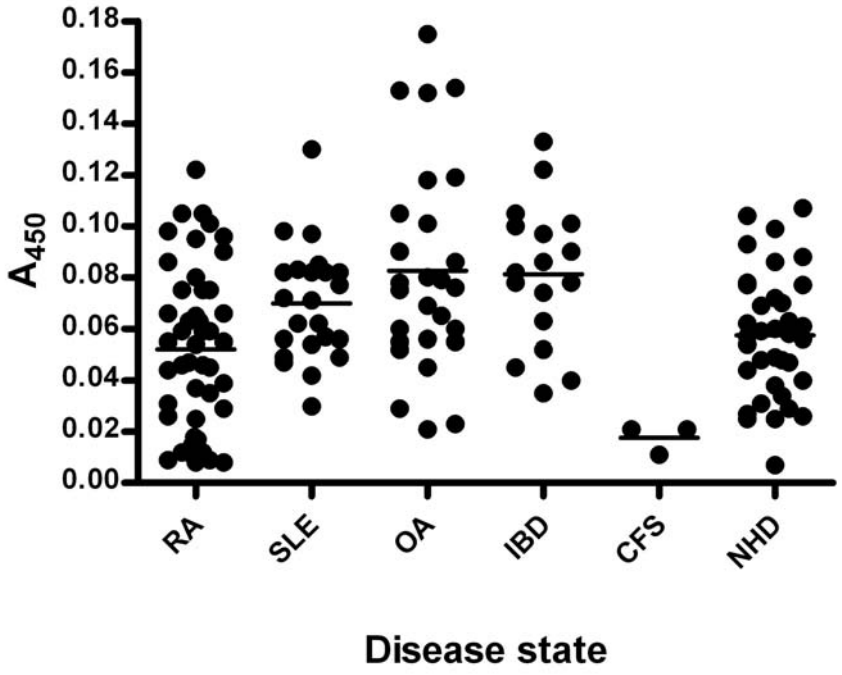


Figure APP2.3 The distribution of OD values for reactivity of patient disease cohorts using the PLSK peptide

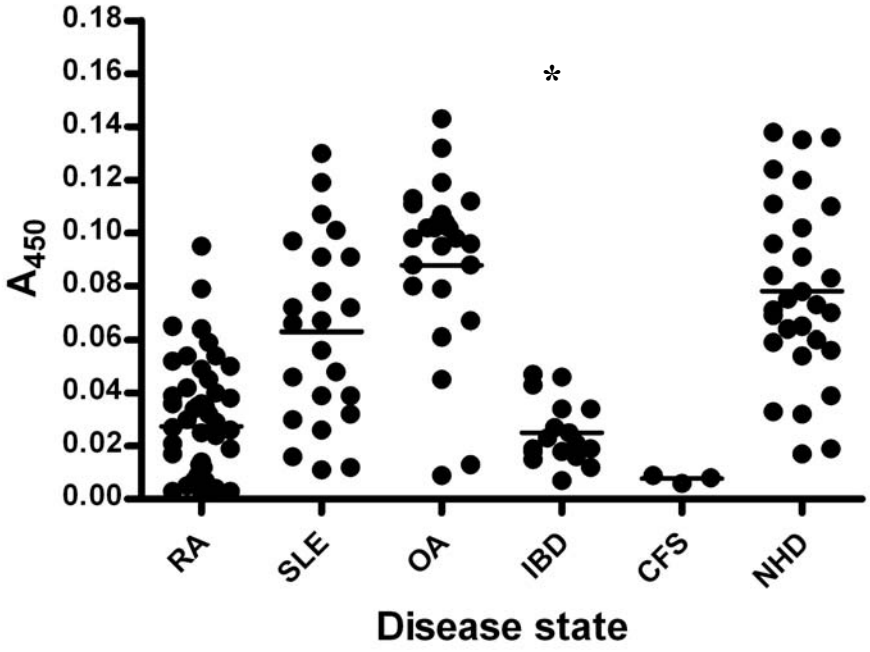


Figure APP2.4 Distribution of reactivity in patient sera using the Negcont1 peptide

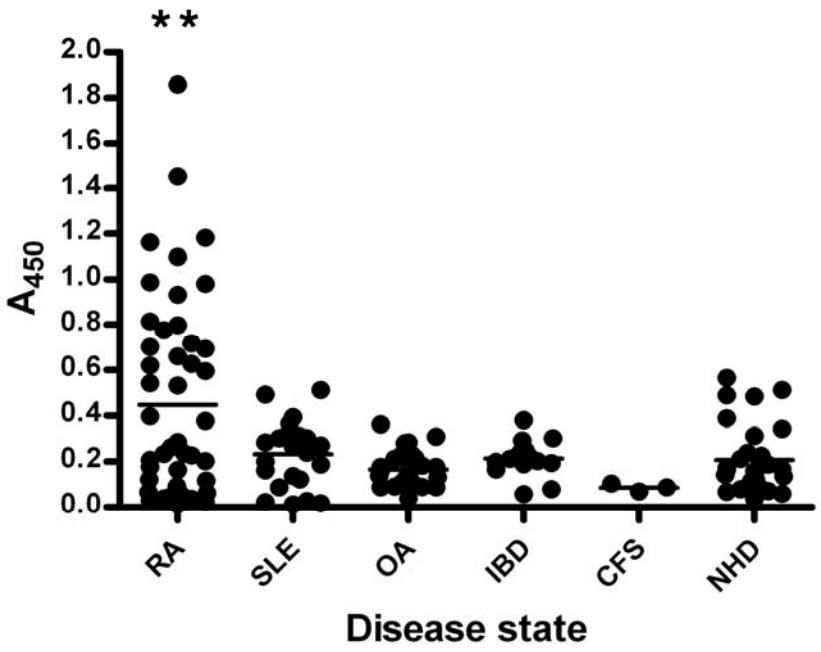


Figure APP2.5 OD values for patient serum from different disease groups using biotinylated GfPN1kpr peptide. ** indicates levels of significance of RA over other disease cohorts. Lines show group means.

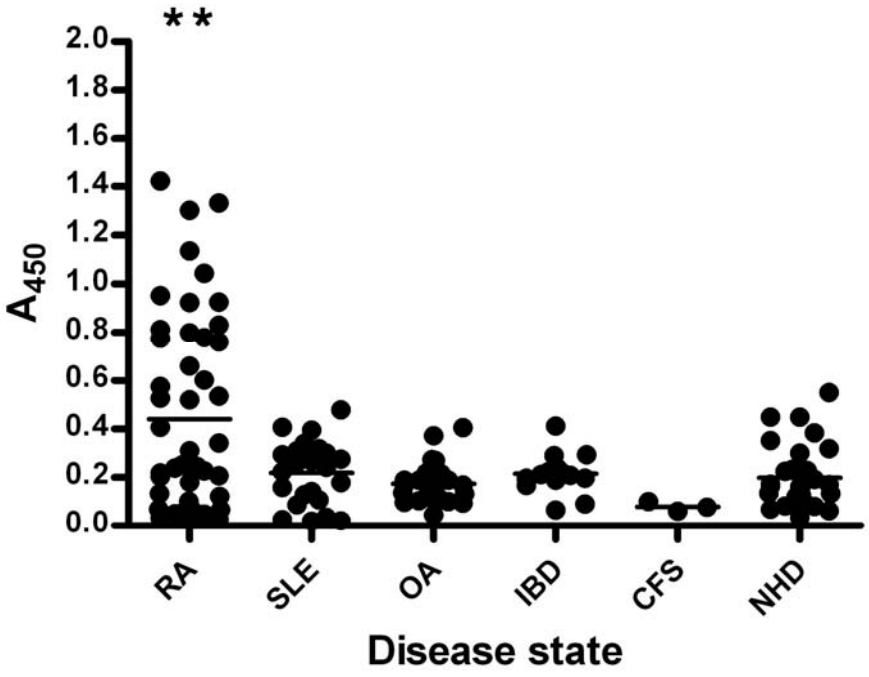


Figure APP2.6 OD values for patient serum from different disease groups using biotinylated GfPN2eip peptide. ** indicates levels of significance of RA over other disease cohorts. Lines show group means.

Appendix *III:*

Cell culture reagents and conditions

Cell culture reagents

Growth media and conditions

Equipment

Hanks' Buffered salt solution (Sigma, H6648)
1xTrypsin-EDTA solution (Sigma, T4415)
FCS (PAA, A14-144)
DMSO – Dimethyl sulphoxide (Sigma, D8418)
500ml filter system (0.2mm nylon) (Corning)
T25/T75/T175 Tissue culture flasks (Sarstedt)
10ml/25ml pipettes (Sarstedt)
1000µl pipette tips (Starstedt)
Sterile aspirator tips (Starstedt)
20ml Centrifuge tubes
Cryovials (Nalgene)

Methodology

To one bottle of RPMI 1640 (Tera-1/GH/BJAB/Jurkat/B95.8) add:

- Fetal bovine serum (10%)
- L-Glutamine (0.3g/l)
- **Either**: Penicillin (50000IU/l) or Streptomycin (50mg/l)

To one bottle of MEM (HEK-293) add:

- Fetal bovine serum (10%)
- 1M HEPES
- 0.1mM non-essential amino acids

To one bottle of MEM (MCF7) add:

- Fetal bovine serum (10%)
- 1.5g/l sodium bicarbonate
- 0.1mM non essential amino acids
- 1mM sodium pyruvate

To one bottle of human synoviocyte basal medium (HFLS) add:

- One bottle of human fibroblast growth supplement (Cell applications, San Diego).

1. Pour all of the above ingredients into the upper unit of a 500ml filter system, and apply vacuum pressure.
2. Label bottle with cell line used for, media and additives/supplements, name and date.

Feeding and maintenance of cell lines (subculture): Tera-1/GH/BJAB

Feeding

1. Incubate medium in water bath at 37˚C.
2. Aspirate off/ take off the contents of the flask – 7ml for T25/28ml for T175.
3. Incubate flask at 37˚C under appropriate conditions.

Passaging

1. Aspirate flask and wash twice with 2ml Hanks' buffered salt solution.
2. Add 2ml Trypsin-EDTA to each and incubate for ~3 minutes (or until cells detach from the flask surface).
3. Add 15ml of media into 2xT75 flasks (30ml of T175 flasks).
4. Resuspend trypsinised cells in 8ml of media to inhibit Trypsin and add 2.5ml of media/ trypsinised cells to each flask (assuming 1 in 4 split between equal sized flasks). For higher seedings add more cell suspension, etc.

Nb. If transferring to different sized flasks

T75 $\Rightarrow$ T175 – **Double**

T175 $\Rightarrow$ T75 – **Halve**

5. Label flasks and incubate at 37˚C.

Feeding and maintenance of cell lines (subculture): B95.8

Feeding

1. Aspirate and discard 5 ml spent media from flask taking care not to take up suspended cells if possible.
2. Replace spent medium with fresh supplemented media.

Passaging

3. Aspirate cells off into a 50ml universal flask.
4. Wash flask twice with 2ml Hanks' buffered salt solution.

5. Add 2ml 0.5M Trypsin-EDTA to the flask and incubate for ~3 minutes at 37˚C, 5% CO_2 until cells detach from the flask surface. Detached cells are diluted in fresh medium and added to universal containing cells.
6. Detached and suspended cells are centrifuged at 1000 rpm for 5 minutes to pellet cells.
7. Pelleted cells are resuspended in fresh medium before reseeding into sterile flasks.

Maintenance of cell lines: Fibroblasts/ HFLS

1. Aspirate flask and wash twice with 2ml Hanks' buffered salt solution.
2. Add 2ml Trypsin-EDTA to each flask and incubate for a minimum of 3 minutes (or until cells detach from the flask surface).
3. Add 15ml of media into 2xT75 flasks (30ml of T175 flasks).
4. Resuspend trypsinised cells in 5ml of neutralising solution to inhibit Trypsin and add 2.5ml of neutralised cells to each flask (assuming 1 in 4 split between equal sized flasks). For higher seedings add more cell suspension, etc.
5. Take 5ml of the media and cells and transfer to new flask.
6. To the 5ml in the new flask add 15ml media.
7. Label and incubate new flasks at 37˚C, 5% CO_2.

Freezing down of cells: All cell lines

1. Pour the old cell culture media into a beaker of virkon.
2. Add 5ml of medium into each flask.
3. Scrape the cells in the flask using a cell scraper.
4. Pour the cells & media into universal 25ml tubes, and label each tube accordingly.
5. Spin down cells at 1500xg for 5 minutes.
6. Some of the medium supernatant should be eluted off the cell pellet, leaving about 0.5ml (for larger pellets, leave more media supernatant).
7. The cell pellet should be resuspended in the media supernatant.
8. Make up 'freezing mix' using FCS and DMSO.

The two reagents should be present at a 9:1 ratio respectively.

9. 0.5ml of freezing mix should be added to the resuspended cell pellet, and this should be transferred to the relevantly labelled cryovials.
10. The freezing media should be added slowly to the resuspended cells.
11. The end concentration of the mixture should be:

7 :3 :1

Medium : FCS: DMSO

12. The cryovials are then transferred to the fridge (+4˚C) for 30 minutes, -20˚C overnight and then onto liquid nitrogen after a days storage.

Resuscitation of cell lines from frozen

1. Add 6ml of supplemented media to a T25 flask, and warm at 37˚C in an incubator.
2. Retrieve the cryovial from liquid nitrogen storage and transfer to a small vessel containing water for 1 minute (do not submerge cryovial).
3. Using a pipette, transfer the thawed cells to a T25 flask and incubate at 37˚C under appropriate conditions until cells adhere (4-24 hours). Label flask with contents, passage number, name and date.
4. Aspirate media from the flask and replace with 8ml fresh media, before incubating.
5. Once cells approach confluency, transfer contents to T75 flask as follows.
6. Aspirate off contents, and wash twice in 2ml of HBSS.
7. Aspirate flask and add 1 ml of Trypsin-EDTA before incubating for ~3 minutes, or until cells detach.
8. Add 10ml of medium to T75 flask.
9. Add 4 ml of media to T25 flask to inhibit trypsin and add 5ml of contents to T75 flask.
10. Label flask and incubate at 37˚C until cells require feeding and subculturing.

Appendix *IV:*

Publications and Presentations

Published papers

EJTEHADI, H. D., FREIMANIS, G. L., ALI, H. A., BOWMAN, S., ALAVI, A., AXFORD, J., CALLAGHAN, R. & NELSON, P. N. (2006) The potential role of human endogenous retrovirus K10 in the pathogenesis of rheumatoid arthritis: a preliminary study. *Ann Rheum Dis,* 65**,** 612-6.

NELSON, P.N., FREIMANIS, G.L. & RODEN, D. (2008) Human endogenous retroviruses: evolutionary dynamics and chromosomal location. *Encyclopedia of Life Sci, - in press.*

NELSON, P.N., WESTWOOD, O.M.R., FREIMANIS, G., RODEN, D., RYLANCE, P.B., SASSOURI, S., HAY FC. (2008) Antigenic Regions on IgG1Fc Identified using Bioinformatics and Monoclonal Rheumatoid Factors: A preliminary study. *Clinical Medicine: Arthritis and Musculoskeletal Disorders* (accepted Dec 07 - *in press*).

Abstracts and presentations

Oral presentations.

Retrovirus workshop; P1

Freimanis, G. & Nelson, P. N. (2007) Cytokine modulation of endogenous retroviruses in rheumatoid arthritis. SGM 160th Annual meeting, Manchester.

Workshop 2: Human endogenous retroviruses:

Immunology; W2.1

Freimanis, G., Roden, D. & Nelson, P.N. (2004) Retroviral molecular mimicry in rheumatic disease: a bioinformatic approach. Immuunology, 113, supplement 1, 96-97.

Poster presentations

Autoimmunity,

Freimanis, G., Shaw, M., Roden, D., Etjhadi, E., Ali, H., Veitch, A & Nelson, P.N. (2007) A role for HERV-K10 in rheumatoid arthritis? *British Society for Immunology Annual congress*, Glasgow.

Autoimmunity,

Freimanis, G. & Nelson, P. N. (2005) B cell epitope mapping – A bioinformatic approach. *British Society for Immunology Annual congress*, Harrogate.

1st place, Carillion Prize 2006, University of Wolverhampton & Carillion plc. May, 2006. Cause or effect? Investigating the role of retroviruses in Rheumatoid arthritis.