# HUMAN DNA SEQUENCES RELATED TO MURINE ENDOGENOUS RETROVIRAL SUPERANTIGENS AND THEIR EXPRESSION IN SJOGREN'S SYNDROME.

**Brian Kenneth Smith Davidson**

Thesis submitted for the degree of M.D.

to the University of Glasgow.

Department of Pathology,

University of Newcastle-upon-Tyne

March 2001



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

## ABSTRACT

The project described in this thesis was undertaken to identify human DNA sequences related to the MMTV 3'LTR gene (vSag) that encodes endogenous superantigens in mice. Using sequence specific PCR primers and murine probes specific for the first and second conserved region of murine vSag, six human clones have been isolated from a DNA library. These human clones have been characterised through a process of hybridisation analysis, subcloning and sequencing. These DNA sequences show limited regions of homology with murine vSag, although none of these clones have extended regions of homology over the entire vSag gene. These vSag-related sequences do not lie in close proximity to other MMTV genes, as indicated by the failure of these clones to hybridise strongly with MMTV *gag, pol* and *env* probes. DNA database searches indicated that three of the six human DNA sequences with limited regions of homology to murine vSag were greater than 98% identical with sequences submitted as part of the Human Genome Project. None of these sequences represented known human genes. Potential open reading frames were identified in five out of the six human DNA sequences, with predicted translated proteins ranging in length from 43 to 156 amino acids.

In the second part of this thesis, evidence of expression of human vSag-related sequences was sought in the autoimmune condition Primary Sjogren's Syndrome (1°SS). This series of experiments was carried out to test the hypothesis that retroviral superantigens are involved in the aetiology of this condition. vSag-related sequences unique to 1°SS were not identified in DNA isolated from the peripheral blood of patients with this condition. Messenger RNA transcripts hybridising with one of the human vSag-related sequences were identified in the minor salivary gland tissue from three out of five patients with 1°SS, with a second vSag-related sequence hybridising to a similar sized mRNA transcript in one of these patients. In experiments to test an

alternative hypothesis that the retroviral gene HTLV-1 *tax* has an aetiological role in 1°SS, sequence specific PCR primers for this gene failed to amplify appropriate sized products from cDNA derived from the salivary glands of five 1°SS patients.

These experiments demonstrate that the human genome does not contain genes that are closely related to murine vSag, and that the DNA sequences sharing short regions of homology with this murine gene do not lie within a proviral configuration. The potential biological significance of these weakly related DNA sequences remains to be determined.

**Acknowledgements**

I would like to thank Dr F. May and Professor B. Westley for their helpful guidance throughout this study. I am also indebted to Dr I. Griffiths for his advice and continued encouragement and to Dr Sarah Bartram for her endless moral support.

# CONTENTS

# List of Figures

# List of Tables

# 1 CHAPTER ONE

# Introduction

## *1.1 Superantigens*

Over the past few years it has become apparent that diverse groups of microbial pathogens produce a novel class of antigens which interact with the immune system in a unique manner (1, 2). These superantigens differ from conventional antigens by their ability to react with a large percentage of the T lymphocyte repertoire, by interacting with the T cell antigen receptor (TCR) outside the conventional antigen-binding site (3, 4). As a result, up to 30% of the T cell repertoire is capable of interacting with a single superantigen. This compares with a responding frequency of less than 1 in 1000 for conventional peptide antigens. Like conventional peptide antigens, superantigens are only recognised by T cells when presented by MHC class II molecules, however they bind to these molecules outside the usual peptide binding groove (5). As a result, polymorphic differences in MHC that affect peptide binding do not usually affect superantigen binding, and responses to superantigens are not 'MHC restricted' to the same extent as responses to conventional antigens (6).

Superantigens are produced by a diverse range of bacteria, mycoplasma and viruses, and have markedly different structures (7). Despite this structural diversity, these different subclasses of superantigens appear to interact with the immune system in similar ways. They produce a variety of pathological effects after infection, which appear to result from over-stimulation of immune system (8). In addition to acute 'toxic syndromes', it has been suggested that superantigens may cause autoimmune diseases (9).

This introductory chapter will focus on two of the main subgroups of superantigens; the bacterial enterotoxins and the superantigens produced by the murine retrovirus, mouse mammary tumour virus (MMTV). It will consider the structure of these superantigens and how they interact with MHC molecules and the T cells receptor. It will describe evidence that the human genome contains DNA sequences that are related to MMTV. Finally evidence suggesting the involvement of retroviruses and superantigens in the development of autoimmune diseases will be discussed. It will first consider in more detail how interactions with superantigens differ from those to conventional peptide antigens.

### 1.1.1 The T lymphocyte-MHC interaction with conventional antigens

The T lymphocyte antigen receptor (TCR) is a heterodimer with alpha and beta chains encoded by gene complexes on chromosomes 14 and 7 respectively (10). The diversity of the T cell repertoire is generated by recombination of the genes encoding the TCR. Each chain has a constant region, encoded by a C gene segment, and a variable region, encoded by J (junctional), D (diversity) and V (variable) gene segments for the beta chain, J and V gene segments for the alpha chain. There are at least 50 different V gene segments, 20 D gene segments and 6 J gene segments. The Vβ genes are divided into approximately 20 families based on amino acid homology. The recombination of these genes, together with junctional diversity generated by imprecise V-J, V-D and D-J joining, allows a potential repertoire in the order of $10^{12}$ different T cells (11).

The TCR recognises antigens, in a processed form of between 9-14 amino acids, in association with molecules encoded by the major histocompatibility gene complex (MHC) of antigen presenting cells (12). The part of the TCR involved in recognition of the antigen-MHC complex is that part encoded at the joining regions of the V and J gene segments of both alpha and beta chains, resulting in a hypervariable region with

maximum diversity at the site of antigen recognition (13). There are two classes of MHC molecules involved in antigen presentation. Class I MHC molecules are found on all nucleated cells and present antigen to $CD8^+$ T cells. Class II MHC molecules are found primarily on B lymphocytes and 'professional' antigen presenting cells such as macrophages and dendritic cells. MHC class II molecules present antigen to $CD4^+$ T lymphocytes (14), often referred to as helper T cells because of their central role in the immune response. Soluble antigens may be internalised by antigen presenting cells by phagocytosis (e.g. macrophages and dendritic cells) or in the cases of B lymphocytes by endocytosis after binding to surface immunoglobulin (the B cell specific antigen receptor). Such exogenous antigens are partially digested in a series of endosomes where they encounter class II MHC molecules (15). Here appropriately sized antigenic peptides (typically 14 amino acids in length) bind to newly exposed 'peptide binding grooves' on the MHC molecule, following cleavage of an invariant chain from the MHC class II assembly (16). The binding of antigenic peptide to the MHC binding groove may signal transport of the complex to the cell surface, thereby exposing the processed peptide to potentially reactive CD4+ T lymphocytes. By contrast, antigens produced internally by a cell, for example following a viral infection, are partially digested by a specialised proteosome complex within the cytoplasm of the cell (17). The processed 'endogenous' peptides are then transported to the cell golgi apparatus aided by the 'transporter-associated proteins' encoded by TAP 1 and TAP 2 genes (18, 19). Here they encounter MHC class I molecules, where appropriately sized antigenic peptides (usually 9 amino acids in length) may bind to the peptide binding groove, signalling the transport of the antigen-MHC complex to the cell surface.

### 1.1.2 The T lymphocyte-MHC interaction with superantigens

The response to superantigens differs from that to conventional antigens in a number of important ways (20). Superantigens are not processed by antigen presenting cells

into short peptides and do not lie within the binding groove of MHC molecules. Instead they bind, in an unprocessed form, to the outside of the MHC (5, 13). During T cell recognition, the hypervariable region of the TCR is not involved, but instead the superantigen binds outside the conventional antigen recognition site, to a region encoded by the variable segment of the TCR beta chain (21-23). As there are a relatively limited number of V gene segments, a large percentage of the T cell repertoire, all sharing the same V beta region, may react to the same superantigen. This results in a rapid and excessive immune response, due to the large number of potentially reactive T cells, and also because processing of the superantigen is not necessary.

This novel mechanism of superantigen recognition appears to have other important functional consequences. Whereas the T cell response to a conventional antigen presented by MHC class II molecules results in proliferation of a limited number of clones of CD4+ cells, the excessive proliferative response to superantigens is typically followed by inactivation, or even deletion, of the reactive cells (24). This resultant anergy is manifested by unresponsiveness following subsequent exposure to the superantigen. Furthermore both CD4+ and CD8+ T cells are able to respond to superantigens presented by MHC class II molecules (25). As mentioned earlier, CD8+ T cell responses to conventional antigens are restricted by MHC class I molecules, while CD4+ T cell responses are restricted by MHC class II. This results from the ability of CD8 and CD4 molecules to stabilise TCR interactions with MHC class I and class II respectively. The observation that CD8+ T cells respond to superantigens presented by class II MHC (25) suggests that superantigens may also stabilise TCR-MHC interactions allowing Vβ-specific T cells to respond to MHC class II bound superantigens irrespective of whether they are CD4 or CD8 positive.

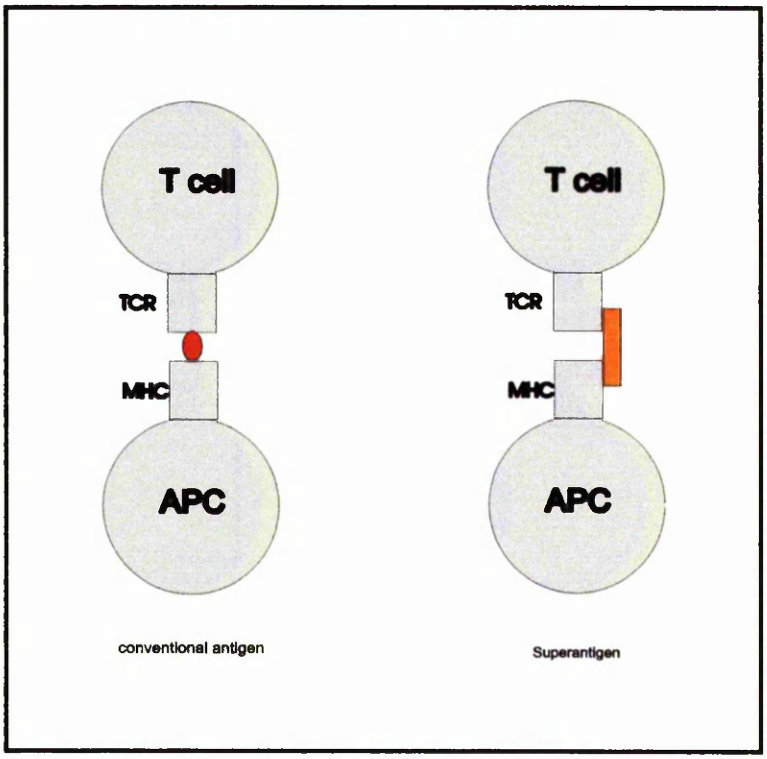



**Figure 1.1 Schematic representation of conventional antigen and superantigen recognition.** The conventional peptide antigen is shown in red, sitting in the peptide-binding grove of the MHC molecule, and recognised by the hypervariable region of the TCR. The superantigen is shown in orange, binding to the MHC outside the peptide-binding groove, and recognised by the Vβ region of the TCR.

Weak superantigen effects have been described where endogenous retroviral superantigens do not cause clear-cut Vβ deletions (26). It has been suggested that superantigens with lower affinity for a TCR Vβ region, may require "stabilising" interactions between MHC and TCR (27). These weak superantigens may require interactions with other regions of the TCR, such as Vα or DJ regions (28, 29). It has been suggested that the affinity for an endogenous superantigen will determine whether individual T cell clones are clonally deleted, anergized (switched off), or unaffected, when the superantigens are encountered in the thymus (6).

### 1.1.3 MHC restriction with conventional antigens versus superantigens

A further important distinction to be made between conventional and superantigen responses concerns their interaction with the MHC molecule. As already mentioned, conventional antigens are presented to the T cell bound to the MHC, resulting in a phenomenon referred to as MHC restriction. MHC molecules are highly polymorphic, and this polymorphism is concentrated at the site of the peptide-binding groove (30). The different MHC genes in man are referred to as HLA (human leukocyte associated) types. An individual within a population will inherit his HLA type from his parents. There are six well characterised HLA class I loci but HLA A, B and C genes encode the principle transplantation antigens in humans and are expressed on all nucleated cells. Each encodes a 3 domain molecule which complexes with an invariant single domain chain, β2-microglobulin to give the functional HLA class I antigen. Individual haplotypes contain up to 14 different class II loci, clustered in three major subregions, HLA DR, DQ and DP. Each subregion contains at least one functional β locus and one functional α locus. These encode the α and β chains of the functional class II HLA antigens. A second expressed DR molecule is encoded on most haplotypes (encoded by DRA and either DRB3, DRB4 or DRB5), resulting in four distinct expressed class II molecules per haplotype. Each MHC molecule has a different peptide-binding groove, capable of presenting a large variety of different processed peptides. However not all appropriately sized peptides can be accommodated by the same type of MHC molecule. Failure to present a processed peptide by any of the types of MHC molecule an individual possessed would result in an absent immune response, hence those HLA types would be associated with unresponsiveness to that antigen. Likewise an HLA type with a particularly high affinity for a processed peptide, might increase the chances of a specific T cell response developing to the antigen from which that peptide originated.

Although superantigens are presented in the context of MHC molecules, they do not bind to the highly polymorphic peptide-binding groove of the MHC molecule (5, 31). Responses to superantigens are therefore not restricted by MHC type to the same extent as conventional antigens, although minor polymorphisms at the superantigen-binding site result in a hierarchy of MHC-dependent responses (32, 33). As many different MHC types can present superantigens, most members of the population will make responses to an individual superantigen. This could have catastrophic consequences if the response was deleterious to the individual. Evolutionary pressures may therefore have been in place to allow the deletion of superantigen responsive T cells, as will be discussed later.

### 1.1.4 Bacterial superantigens

The term 'superantigen' was first applied to the bacterial enterotoxins responsible for food poisoning and toxic shock (34, 35). Although it was known that these toxins were able to stimulate large numbers of T cells, both in humans and in mice, they differed from mitogens. They did not activate *in vitro* as large a proportion of T cells as mitogens such as concanavalin A or phaetohaemaglutinin and some T cell clones could not be stimulated by particular toxins. Furthermore stimulation was dependent on the presence of class II bearing cells.

These toxins comprise a large group of proteins produced by various bacteria including staphylococci and streptococci (1). The family of staphylococcal enterotoxins (SEA, B, C1, C2, C3, D and E) are produced by different strains of staphylococcal aureus and cause food poisoning in man (9). Related toxins also produced by staph aureus cause tampon-related toxic shock (toxic shock syndrome toxin, TSST1) and scalded skin syndrome (exfoliating toxins A and B). Toxins produced by streptococcus

pyogenes (streptococcal pyrogenic enterotoxin; SPE A, B and C) cause a syndrome characterised by rash, fever and shock (36).

#### 1.1.4.1 Structure of enterococcal superantigens

The staphylococcal and streptococcal enterotoxins are globular proteins ranging in size from 24 to 30 kDa. By comparing their amino acid sequences, similarities between the different enterotoxins have been identified. SEA and SEE share 90% amino acid sequence homology, and SEB is closely related the SEC toxin (1). The streptococcal enterotoxins SPE-A and SPE-C are structurally related, and are similar to the staphylococcal enterotoxins SEB and SEC (37). However other enterotoxins share no significant amino acid sequence homology. SPE-B is unrelated to any of the other enterotoxins, and appears to be related to a streptococcal proteinase precursor (38). It is also difficult to align sequences of TSST-1 and the exfoliative toxins with the other enterotoxins (1).

#### 1.1.4.2 TCR Vβ chain specificity for enterococcal superantigens

The specificity of enterotoxin binding with the TCR Vβ chains has been extensively investigated. There is some correlation between the primary amino acid structure of the various enterotoxins and the pattern of Vβ reactivity, although considerable overlap exists (35, 39-43). Table 1.1 shows the staphylococcal enterotoxins separated into three groups according to greatest amino acid sequence homology, giving the mouse TCR Vβ specificity of each. A similar pattern is observed with human TCR Vβ specificity.

| | Enterotoxin | TCR Vβ specificity |
|---|---|---|
| Group 1 | SEA | 1 3 10 11 17 |
| | SED | 3 11 17 |
| | SEE | 11 15 17 |
| Group 2 | SEB | 7 8.1 8.2 8.3 |
| | SEC1 | 3 8.2 8.3 11 |
| | SEC2 | 3 8.2 10 17 |
| | SEC3 | 7 8.2 |
| Group 3 | TSST-1 | 15 16 |
| | ExFT | 10 11 15 |

**Table 1.1 Vβ specificity of bacterial enterotoxins**

### 1.1.4.3 Enterotoxin-MHC interactions

The bacterial enterotoxins were found to have high binding affinities for MHC class II molecules (44, 45). SEA and SEB were shown to cross compete suggesting that they bound to the same site on class II (46). However SEB and TSST1 did not compete so may have different binding sites on class II molecules (47). X-ray crystallography of SEB associated with class II confirmed earlier evidence that the bacterial toxins bind outside the conventional MHC peptide binding groove (31). This had been suspected from experiments showing that the association of SEB with mouse class II molecules did not inhibit presentation of conventional antigenic peptides to T cells (5). Point mutations introduced into the MHC molecule also suggested that the site of binding lay outside the conventional peptide-binding groove (48). The different affinities of TSST-1 for human class II molecules DR and DP suggested that the α domain of the MHC class II molecule HLA-DR1 was essential for high-affinity binding of this

superantigen (49). X-ray crystallography data subsequently showed that SEB bound to the α1 domain of HLA-DR (31). This exclusive interaction with the DR1 α-chain explained the ability of SEB to bind many different DR allotypes, as this chain is conserved in all DR molecules.

The regions of the enterococcal superantigen proteins involved in binding the MHC and TCR Vβ chains have also been identified by mutational analysis. For the SEB enterotoxin, both the binding of MHC and TCR Vβ mapped to the amino-terminal region of the toxin (50). Mutation in a region mapped to residues 41-53 resulted in defective binding of SEB to MHC class II molecules. Residues 60-61 were important in binding TCR Vβ, while a third region (residues 9-23) affected binding to both MHC and TCR Vβ. The three-dimensional structure of a human class II MHC molecule complexed with SEB determined by X-ray crystallography (31) showed the critical residues of SEB involved in binding HLA-DR1 derive from an N-terminal β-barrel domain, with potential hydrogen bonds formed between DR1 and the residues 43-46 of SEB. Three residues from the C-terminal helix of SEB (residues 210-217) were also in contact with the DR1 molecule, together with a salt bridge between the DR1 α-chain residue lysine 39 and the SEB glutamic acid 67.

#### 1.1.4.4 Enterotoxin-TCR Vβ interactions

Some mutations in the SEB molecule affect T cell stimulation but not class II binding, suggesting different residues of SEB specifically interact with the TCR (50). Regions affecting Vβ specificity have been determined by looking at the amino acid sequence differences between SEA and SEE which are 90% homologous but have different patterns to TCR Vβ stimulation (51). Amino acid residues at positions 206 and 207, near the carboxy terminus of the enterotoxins, were found to be primarily responsible

for the differences in Vβ specificity using hybrid molecules with residues exchanged between the two toxins. The crystal structure supports a model in which the MHC interaction with the TCR is distinct form the conventional peptide-mediated interaction. A stretch of residues located in a disuphide loop of SEB lies across a region of the α1 domain of DR1, covering residues that have been implicated in TCR recognition of peptide-MHC complexes (31). This suggests that superantigens may not take advantage of any residual affinity of TCR for MHC derived from positive selection during thymic education. However, the location of the interacting residues on SEB and TCR suggests that the TCR is still positioned in close proximity to the class II peptide-binding site. This may explain how TCR α-chains and MHC polymorphisms can modulate superantigen stimulation.

Several studies have examined the binding site on the TCR Vβ chain for the bacterial superantigens. In view of the conservation of critical amino acids between immunoglobulin and TCR V domains (52), it is generally assumed that the structure of the TCR V region is similar to immunoglobulin. The TCR V domains have therefore been envisaged as forming a series of anti-parallel β strands with the loops corresponding to the three immunoglobulin complementarity-determining regions (CDRs) brought to one face to form the binding site for peptide antigen-MHC (13). Amino acids important for peptide-MHC recognition have been shown to predominate in those regions that correspond to the CDR loops, especially CDR3, which is formed from the junctional Vα-Jα and Vβ-Dβ-Jβ gene segments. In contrast, residues involved in superantigen recognition appear to be in a different region of the TCR, away from the CDR loops.

Amino acid substitution experiments have been used to identify important residues on the TCR Vβ chain, which are involved in binding bacterial enterotoxins. Residues of

murine Vβ3, which binds SEC3, were exchanged with Vβ17, which does not bind with this superantigen (53). This suggested that residues at positions 66, 68, 72 and 74 of the Vβ chain, which mapped to hypervariable region 4 (HV4), were critical for binding of SEC3 to the TCR. These residues were not involved in recognition of a peptide-MHC complex. Similar residues in HV4 (67-77) were critical in the interaction between human Vβ13.2 and the bacterial enterotoxins SEC2 and SEC3 (21).

#### 1.1.4.5 Functional consequence of TCR interaction with enterococcal superantigens

The *in vivo* responses to bacterial enterotoxins have been studied in detail. Injections of SEA in mice gave an initial sharp rise in Vβ3+ and Vβ11+ T cells followed by a dramatic decline in these cell populations (54). This response was observed in both CD4+ and CD8+ T cell subsets. Depletion of T cells was not permanent, and the recovery was in a time scale consistent with the generation of new cells in the thymus in the absence of toxin. This pattern of activation followed by deletion and/or inactivation (anergy) of superantigen responsive T cells appears to be characteristic of superantigen responses. The mechanisms underlying the deletion of the superantigen reactive T cells involves the 'programmed cell death' pathway, apoptosis (55). Mice carrying the lpr/lpr defect have an abnormality of Fas, a molecule involved in signalling apoptosis. In these mice, apoptotic death of superantigen stimulated cells was dramatically reduced (56), although other apoptosis signal pathways in addition to Fas also appear to be involved (57).

This process of switching off the immune response after stimulation with a superantigen may play a role in limiting autoimmune reactions and in the maintenance of immune homeostasis. It may also explain how bacteria benefit from producing superantigens, as clearly they must derive some advantage from the production of

toxins of this type. By initially producing local inflammation, the bacteria may benefit from the increased blood and nutrient supply. The subsequent suppression of the immune response induced by the toxin might reduce the hosts resistance to invasion by the bacteria.

As host-bacterial relationships have evolved concurrently, it might have been expected that the mammalian immune system would have mutated in such a way that the class II failed to bind these enterotoxins, or that the TCR Vβ chain did not bind with the enterotoxin-class II complex. Such mutations might have resulted in loss of function of MHC class II molecules and the enterotoxins may be capitalising on some essential structural feature of the TCR Vβ chain. However as discussed below, large gaps in the T cell receptor repertoire are seen in mice, as a result of deletion of specific Vβ bearing T cells. These deletions remove many of the T cells known to react with enterococcal superantigens. These large gaps in the murine T cell repertoire are associated with the presence of minor lymphocyte stimulating (Mls) genes. The products of these genes are now known to be superantigens, which are encoded by the endogenous retrovirus, mouse mammary tumour virus (MMTV).

It is possible that enterococcal superantigens have caused selection pressures to limit the murine TCR Vβ repertoire and as a result endogenous retroviral superantigens have been retained in the genome. Although murine and human responses to these enterococcal superantigens appear to be similar, there are no equivalent large Vβ specific gaps in the human T cell repertoire. It is not known whether endogenous retroviral superantigens are present in the human genome, and clarification of this was the primary aim of the work described in this thesis.

## *1.2 Retroviruses*

Retroviruses are small RNA viruses that have a unique mode of replication (58). Their RNA genome replicates through a DNA intermediary by a process called reverse transcription, mediated by an RNA-dependent DNA polymerase (reverse transcriptase). The retroviral genome consists of a 60-70S dimer complex of single stranded, positive-sense RNA. The dimer contains two identical subunits, which resemble messenger RNA (mRNA) molecules in having a methylated 5' cap structure and a polyadenylated tract at the 3' end. The genome contains three essential functional genes: *gag*, *pol* and *env*. The *gag* gene encodes the structural core proteins of the virus, the *pol* gene encodes the reverse transcriptase and the *env* gene encodes the envelope proteins. The arrangement of these genes is the same for all known retroviruses: i.e. 5' – *gag* – *pol* – *env* – 3'.

When a retrovirus infects a cell, the retroviral reverse transcriptase makes a double-stranded DNA copy of its genomic RNA. The DNA copy of the retrovirus can then integrate into the host cell's DNA as a provirus. During this process, the terminal sequences present in the RNA genome are duplicated to form repetitive structures called long terminal repeats (LTRs). In the integrated provirus, LTRs consist of three elements, U3, R and U5, which contain transcription initiation and termination signals. Retroviruses with complex genomes also encode accessory proteins to modulate expression, including transactivators of transcription (e.g. *tax* in HTLV-1 (59) and *tat* in HIV (60)) and regulators of expression of virus proteins (e.g. *rev* in HIV (60)).

The integrated retroviral DNA will be duplicated along with normal cellular genes when the infected cell divides. The provirus may remain inactive in a cell, until its expression is induced by an event such as cell activation. If a retrovirus infects a germ cell or early embryo, the retroviral genome can integrate into the host germline DNA. In this

event the retrovirus becomes integrated within the DNA of each cell, and is called an endogenous retrovirus (ERV). These proviruses will be inherited in a stable Mendelian fashion by subsequent generations ('vertical transmission'). Unless mutations occur, the provirus will also retain the capacity to produce infective retroviral particles, potentially resulting in 'horizontal transmission'.

### 1.2.1 Mouse Mammary Tumour Virus encoded superantigens

Certain viruses have been shown to exhibit superantigen activity, which resembles the activity of enterococcal superantigens (61). Of these, the most widely studied is encoded by the retrovirus, mouse mammary tumour virus (MMTV). The gene encoding the superantigen lies within the 3' long terminal repeat (L TR), and has been termed vSag (62). Over 50 strains of MMTV have now been identified, and the vSag genes of these strains show significant polymorphism. It is now recognised that these genes are responsible for the "minor lymphocyte stimulating" (Mls) phenomenon, originally described over 25 years ago (63), where lymphocytes isolated from MHC identical mouse strains were found to proliferate in a mixed lymphocyte reaction, a response normally attributed to differences in the major histocompatibility genes. Unlike the highly polymorphic MHC antigens, which in mice are encoded within the H-2 gene complex on chromosome 17, the Mls antigens were from many unlinked loci, each with two alleles, one stimulatory and the other not. The T cell reactivity was determined by the Vβ domain of the TCR (62, 64-66), and the gene products of different Mls loci were reactive with different Vβ TCR domains. In mouse strains positive for an Mls antigen, the reactive Vβ bearing T cells were found to be deleted in the thymus. The Ect-1 Mls antigen, which deleted Vβ 5 and Vβ 11 bearing T cells, could not be segregated from the MMTV provirus Mtv-9 by genetic crosses (67), suggesting that Mls genes were of retroviral origin. This was confirmed by demonstrating that mice transgenic for the MMTV(GR) deleted Vβ 14+ T cells (68), and that transfecting the 3'L TR ORF of MMTV resulted in stimulation of T cells in a

Vβ specific manner (62, 64-66).

Endogenous MMTV genes are integrated in a stable fashion within the murine genome, and are inherited in a Mendelian fashion. Endogenous superantigens are expressed in the thymus (69), and appear to play a critical role in shaping the murine T cell receptor repertoire. When immature T cells in the thymus encounter endogenous retroviral superantigens, those T cells bearing reactive Vβ regions are deleted (70). In contrast, infective MMTV induces a stimulatory response (71). Exogenous MMTV initially infects B lymphocytes within gut associated lymphoid tissue. Although viral replication does not occur in these cells, expression of the retroviral vSag gene product results in activation of T cells with appropriate TCR Vβ regions (72). Proliferating T cells subsequently cause expansion of B cells harbouring the viral genomes. This proliferation of lymphocytes infected by MMTV ensures replication of the proviral genome.

#### 1.2.1.1 Structure of MMTV encoded superantigens

The MMTV open reading frame that encodes vSag activity produces a protein of about 320 amino acids (73), which is thought to be a type II transmembrane glycoprotein, where the carboxyl terminus is extracellular and the amino terminus is cytoplasmic (74). This is supported by experiments that demonstrate that vSag synthesised *in vitro* in the presence of microsomes adopts an inverted membrane orientation (75). The polymorphism in vSag genes between the different strains of MMTV is concentrated at the carboxyl terminus of the predicted amino acid sequence (66, 76-79). Two conserved regions have been identified within vSag (80). The first contains predominantly hydrophobic residues, and is thought to encode the transmembrane domain. The function of the second conserved region is unknown. Five consensus sites for N-linked glycosylation are found carboxy-terminal to the

hydrophobic region. The polymorphic carboxyl terminus appears to binds to the Vβ region of the TCR during superantigen mediated responses, and as a result determines Vβ specificity.

Evidence of cell surface expression of MMTV encoded superantigens is less clear-cut than for the bacterial enterotoxin, and it has not been possible to isolate the vSag protein in a form which would allow detailed structural analysis of the molecule. A monoclonal antibody raised to a 13 amino acid long peptide derived from the extreme carboxyl terminus of the vSag gene of Mtv-7 (vSag-7), identified cell membrane expressed protein on B cell hybridoma cell lines and was found to block superantigen induced T cell proliferation (81). This monoclonal antibody identified vSag-7 surface expression in LPS-activated B cells from CBA/J (Mtv-7+) mice, but not from any unstimulated cell population or on activated T cells. However, other studies using indirect means of identifying vSag induced responses, such as examining the T cell repertoire for the deletion of vSag reactive T cells, suggest that vSag is also expressed by T cells (82, 83), dendritic cells (84-87), and thymic epithelial cells (88, 89). vSag may therefore be expressed on cells other than B cells, but at levels below the sensitivity of flow cytometric analysis.

Western analysis of the expressed vSag-7 gene identified two products of 18.5 kDa and 45 kDa (81). Translation of the entire vSag-7 gene would give a 321 amino acid product with a predicted core molecular weight of 37.1 kDa. The 45 kDa product could therefore represent this protein in a glycosylated state. The 18.5 kDa protein may result from proteolytic cleavage of the 45 kDa product, or from translational initiation at an internal methionine in the vSag-7 open reading frame. Only the 18.5 kDa protein was identified when monoclonal antibodies were used to precipitate vSag-7 from lysates of $^{125}$I surface-labelled hybridoma cells and LPS-stimulated CBA/J (Mtv-7+)

splenocytes (81). The molecular size of the 18.5 kDa protein was unaffected by treatment with N-glycanase, which removes N-linked oligosaccharide, suggesting that the polypeptide is the result of cleavage of the intact vSag-7 protein at a position carboxy-terminal to the N-linded glycosylation sites (see figure 1.2). Mutations introduced at the putative processing site at position 171 abrogated detectable vSag-7 surface expression in B cells, suggesting that proteolytic processing is required for vSag-7 function (90).

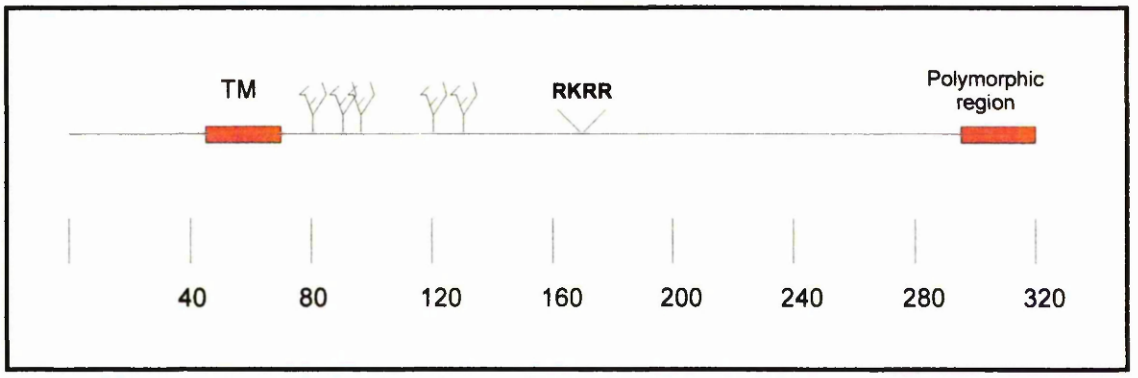



**Figure 1.2 Diagram of the vSag protein encoded by the 3' long terminal repeat (LTR) of MMTV.** The amino acids are numbered. TM denotes the transmembrane region. The polymorphic region lies at the carboxyl terminus of vSag and is predominantly involved in Vβ specificity. RKRR denotes a potential proteolytic cleavage site.

### 1.2.1.2 TCR Vβ chain specificity for MMTV encoded superantigens

Evidence that the carboxyl terminus of vSag binds to the Vβ region of the TCR during superantigen mediated responses comes from the observation that vSag genes with similar carboxy terminal regions stimulate T cells sharing the same Vβ regions (62, 66, 68, 71, 79, 91-94). Transfection experiments using mutant vSag genes also suggest that this region determines TCR Vβ specificity (7). Table 1.2 shows the pattern of TCR Vβ specificity associated with different vSag genes with variable carboxy termini amino acid sequences.

| MMTV | vSag carboxy termini amino acid sequence | TCR Vβ |
|---|---|---|
| GR | DYIYLGTGMHFWGKVFH TKEGTVAGLIEHYSAKTYGMSYTD | 14 |
| C3H | .....I.. .....E | 14,15 |
| | | |
| Mtv-1 | .....IH.....YNSR.EAKRHI...IK.LPLAF | 3,17 |
| Mtv-3 | .....IH.....YNSR.EAKRHI...IK.LPLAF | 3,5,7 |
| Mtv-6 | .....IH.....YNSR.EAKRHI...IK.LPLAF | 3 |
| Mtv-13 | .....IH.....YNSR.EAKRHI...IK.LPLAF | 3 |
| | | |
| Mtv-8 | .....NV...I..Y....A..RQL..I..D.F....NG | 11,12,17 |
| Mtv-9 | .....NV...I..Y....A..R.L..I..D.F....NG | 5,11,12,17 |
| Mtv-11 | .....NV...I..Y....A..RQL..I..D.F....NG | 11,12 |
| | | |
| Mtv-7 | .....N....I.DY.E..AI.KILYNMKYTHG.RVGF.PF | 6,7,8.1,9 |
| Mtv-43 | .....N....I.DY.E..AV.KILYNMKYTHN.RIGF.PF | 6,7,8.1,9 |
| SW | .....N....I.DY.E..AI.KI.YNIKYTHG.RIGF.PF | 6,7,8.1,9 |

**Table 1.2 Vβ specificity of mouse mammary tumour virus superantigens**

### 1.2.1.3 MMTV vSag-MHC interactions

Processed forms of vSag-7 have been shown to bind human MHC class II molecules (HLA-DR1 and HLA-DR4) but not the class I molecule HLA-A2 (95). Both the 28 kDa extracellular domain, and the 18 kDa carboxy-terminal fragment, which had been expressed in Escherichia coli and electrophoresed in an SDS-polyacrylamide gel, bound soluble forms of MHC class II with equal avidity. Although this suggested that the critical MHC binding site on vSag-7 was at a position carboxy-terminal to the putative processing site at position 171, a separate group described a peptide from position 76-119 of vSag-1 which bound to cells expressing the murine MHC class II molecules $I\text{-}A^d$ and $I\text{-}E^d$, and inhibited binding of the bacterial superantigen SEA (96). Taken together, these results suggest that there are two binding sites for MHC class II proteins on the extracellular domain of vSag, one binding to MHC class II in competition with enterotoxin SEA, and a second on the 18 kDa carboxy-terminal fragment binding to distinct site on the class II molecule.

#### 1.2.1.4 MMTV vSag-TCR Vβ interactions

Studies examining the surface of the TCR Vβ region binding to different superantigens showed that vSag engages Vβ on the solvent-exposed face of the molecule (made up of four β strands from residues 1-7, 16-25, 63-68 and 69-76 and several loops). Critical residues for binding vSag-7 were found to be at positions 17, 18, 22, 70 and 71 (22), representing part of the Vβ solvent-exposed face that is relatively distant from the region of Vβ thought to engage MHC. The same solvent exposed face of Vβ is involved in binding the bacterial enterotoxin SEC3 (53), although this involved different critical residues (66, 68, 72, 74) which are relatively close to those involved in binding conventional antigens in association with MHC.

#### 1.2.1.5 Pathological consequences of MMTV

Exogenous MMTV is transmitted to babies through infected milk, infecting B lymphocytes within gut associated lymphoid tissue (97). In the early stages of infection, the retroviral vSag gene is expressed and results in activation of T cells with appropriate TCR Vβ regions (72). The consequent proliferation of lymphocytes infected by MMTV ensures replication of the proviral genome. Continuation of the viral life cycle depends on viral gene expression in mammary tissue. Infection of mammary epithelial cells is dependent on T lymphocytes; nude mice, that lack functional T cells are not susceptible to MMTV-induced tumour development (98). Hormones such as progesterone and prolactin are important in the transcriptional activation of MMTV during pregnancy and lactation (99), and the production of high titres of infectious virus allows the continued transmission of virus to suckling young.
Although mammary tumours are associated with the retrovirus, this depends on where the virus integrates into the murine genome and is not an inevitable consequence of the retroviral infection (100, 101). Where the MMTV provirus is integrated within a

germ line cell, it is inherited in a stable Mendelian fashion. Endogenous MMTV proviruses are present in the germ line of all inbred mice, and most strains contain approximately three to eight proviruses at various genomic locations (102). Most of the proviruses are defective and do not produce infectious virus but have retained a functional superantigen gene and express a vSag protein that results in T cell subset deletion. This is dependent on vSag expression in the developing thymus, which results in the negative selection of reactive Vβ bearing T lymphocytes (103).

The 'vertical transmission' of endogenous MMTV has been shown to confer resistance to infectious MMTV. BALB/c mice congenic for the endogenous retrovirus, Mtv-3, were resistant to experimental MMTV infection, as assessed by measuring viral antigens released in the milk, and by recording the incidence of early mammary tumours (104). This resistance is dependent on the endogenous MMTV causing deletion of the Vβ-bearing T cells specific for the exogenous MMTV (105).

The ability to confer resistance to potentially tumour inducing exogenous MMTV may explain why functional vSag genes have been retained in the murine genome. Although this results in large Vβ-specific gaps in the T cell repertoire, there is no evidence that this causes significant immunodeficiency, even where half the potential T cell repertoire is deleted. Although no other pathological affects appear to result from the presence of these endogenous superantigens, they remain potential targets for the immune system. *In vitro* they are the basis of the Mls reaction described above and *in vivo* they appear to be able to stimulate graft-versus-host disease (GVHD). GVHD typically occurs after transplantation of lymphoid cells incompatible for major histocompatibility antigens (106). A combination of MHC (H-2d) compatible mice that were disparate for vSag was used to study the potential role of host endogenous mouse mammary tumor virus encoded superantigens in the development of lethal

GVHD. The presence of Mtv-7 in the host genome highly increased the rate and severity of GVHD (107). Kinetic analyses of TCR Vβ gene expression in recipient mice indicated a dramatic but transient infiltration of GVHD target organs by vSag-specific T cells. A similar result was observed in another MHC-matched, superantigen disparate, donor-recipient bone marrow transplant combinations, where in the first 2 weeks post-transplant of B10.D2-->BALB/c, approximately 50% of all Thy1.2+ spleen and lymph node cells were found to express T cell receptors utilizing Vβ3 (108).

Chronic GVHD can result in a number of clinical manifestations that resemble autoimmune disease. These include progressive sclerodermatous skin reactions, liver changes resembling primary biliary cirrhosis, feature of Sjogren's syndrome and systemic lupus erythematosus, together with autoantibodies to a variety of tissue antigens (109, 110). MMTV in mice does not appear to be associated with autoimmune pathology. This is presumably because superantigen-reactive Vβ-specific T cells are deleted. However, the possibility remains that autoimmune consequences could result from failure to delete these potentially auto-reactive T cell.

## *1.3 DNA sequences related to MMTV in the human genome*

### 1.3.1 Human endogenous retroviruses

Endogenous retroviruses (ERVs) and retrovirus-like element represent a substantial component of vertebrate genomes and approximately 5-8% of human DNA is retrovirus-related (111, 112). ERVs are thought to have originated from infectious retroviruses because they have a similar structure and share sequence homology. Full length ERVs have a LTR-*gag-pol-env*-LTR structure characteristic of infectious retroviruses. However, many of the ERVs described in the literature are not full length and do not have a complete retroviral structure. Most ERVs have termination codons interrupting open reading frames, and many show partial deletions of structural genes.

In the recently published draft of the human genome, only three full length ERVs were identified where all open reading frames were intact (112).

Human endogenous retroviral sequences appear to belong to two families (113, 114), those related to the type C mammalian retroviruses and those which show a mosaic pattern of homology to mammalian type A, B and D retroviruses (where types A-D refer to the morphological classification of the oncovirinae family of retroviruses). Many of these ERVs were originally identified by screening human genomic libraries with proviral DNA probes under reduced stringency conditions. Such probes included murine leukaemia virus (MuLV), Moloney MuLV (MoMuLV), MMTV, baboon or chimpanzee viruses (115-123).

Human retroviral-like sequences related to C-type retroviruses include the full length provirus ERV3 (116) (chromosome 7) and ERV1 (124), which has a *gag-pol-env*-LTR genomic structure (chromosome 18 q22-23 (125)). Sequence analysis revealed that both clones contain termination codons within the *pol* and *gag* genes, although the *env* gene of ERV3 has a long open reading frame capable of encoding a polypeptide of approximately 650 amino acids (116). These sequences are present as single copies in the human genome, but other human C-type retroviral-related sequences have been identified that have up to 1000 copies per haploid genome. These families include HERV-E (126), HERV-H (127), ERV-9 (126) and HERV-W (128).

B- and D- type retroviruses and A-type particles are thought to originate from a common progenitor on the basis of homologous pol sequences that differ from those of mammalian C-type viruses (129, 130). The human genome contains a large number of retroviral-like sequences with homology to MMTV (a B-type retrovirus) (118-120, 131), as well as to the Syrian hamster intracisternal A-type particle (IAP) (132), and D-type retroviruses (133).

### 1.3.2 Human endogenous retroviral sequences related to MMTV

Human DNA sequences related to MMTV have been isolated by low stringency hybridisation with DNA probes encompassing various regions of the MMTV (117-121). Franklin *et al* (131) analysed 100 recombinant clones, which had been isolated from the DNA of a human breast cancer cell line, by screening with an MMTV *gag-pol* fragment. Cross-hybridisation experiments with subcloned fragments from some of these isolates identified nine distinct subgroups of MMTV-related sequences. Although all subgroups hybridised with the MMTV *gag-pol* probe, they are not closely related to each other. The largest subgroup, comprising 64% of the isolated clones, was found to contain sequences most homologous to MMTV. This group consists of retroviral genomes of 6 to 10 kb and these sequences are members of the HERV-K superfamily of endogenous retroviruses.

HERV-K10 is a full length provirus, and the complete nucleotide sequence was determined almost 15 years ago. (132, 134). It is 9.2 kb in length with LTRs of 968 bp at both ends. The *pol* region of HERV-K10 is closely related to that of A- and D- type retroviruses and especially to the B-type *pol* gene. It contains an open reading frame large enough to allow synthesis of full-length polymerase proteins including reverse transcriptase. The env gene of HERV-K10 shows unusually high similarity with the MMTV env region, even on the level of secondary structure of the predicted amino acid sequences, including potential glycosylation sites (135). However, the LTRs do not show significant sequence homology to the MMTV vSag gene. Several full-length endogenous retroviruses related to HERV-K10 have now been described, including HERV-K(C4) , HERV-K-T47D (isolated from a breast carcinoma cell line (136)) and HTDV/HERV-K (which codes for human teratocarcinoma-derived retrovirus (HTDV) particles (137-140)).

Several human ERVs have been shown to be transcriptionally active in human tissues and cell lines (131, 141-144). Evidence of protein expression has largely been restricted to the presence of antibodies against retroviral gene products (145-148), although a recombinant HERV-K *env* expression vector has recently been described that yielded ENV proteins of the expected molecular mass (139). However, all human endogenous retroviral elements examined to date appear to be replication defective.

Despite the presence of MMTV-related sequences in the human genome, DNA sequences closely related to the MMTV 3'LTR superantigen-encoding gene have not yet been identified. Using sequence specific PCR primers, three human DNA sequences have been identified which hybridised to an MMTV probe (149). One of these human sequences had a short region of homology to murine vSag, although this was only over a region of twenty nucleotides after excluding the PCR primer binding sites. No significant homology to vSag was identified with the other two human sequences, although one sequence was 80% homologous to the *pol* region of an endogenous retroviral-like sequence (RTLV-1) (150).

Evidence that the human immune system can respond to MMTV-encoded superantigens suggests that the human genome may contain related vSag genes. Human T cells respond to murine retroviral superantigens in a Vβ specific manner (151), and human HLA class II molecules can present vSag to both human and mouse T cells (151, 152). In mice, functional vSag genes have been identified that are no longer within a proviral configuration (153). If vSag genes in the human genome were not in close proximity to other retroviral genes, many of the experimental approaches described above would have failed to identify them. The huge number of LTR-related sequences in the human genome, together with the polymorphic nature of the vSag-encoding gene, further complicates the search. As the murine vSag genes

are known to have two highly conserved regions (80), an alternative approach is to screen for human DNA sequences related to these conserved regions. This was the principle strategy of the project detailed in this thesis. Evidence was also sought for the expression of human vSag-related sequences in Sjogren's syndrome, as it has been independently proposed that both endogenous retroviruses and superantigens have an aetiological role in this autoimmune condition (154, 155). The background to these hypotheses will be discussed in the remaining sections of this introduction.

## *1.4 Sjogren's Syndrome*

Primary Sjogren's syndrome (1°SS) is a heterogeneous disease characterised by lymphocytic infiltration of salivary and lacrimal glands and extraglandular structures (156). It is associated with polyclonal B cell activation (157) and the production of organ and non-organ specific autoantibodies (158). The ophthalmologist Henrik Sjogren described the condition in 1933, in an account of 19 patients with dry eyes, of whom 13 had arthritis (159). He introduced the term keratoconjunctivitis sicca to describe the dryness of the eye leading to conjunctival and corneal ulceration. He also reported dryness of the mouth (xerostomia) in some of these patients, and described an inflammatory infiltrate in the salivary glands of one patient at post-mortem.

Sjogren's syndrome is now thought to represent a group of diseases with a common histological feature of chronic lymphocytic infiltration of exocrine glands with subsequent decrease in function. Immunohistological studies of salivary tissue have indicated that the majority of the infiltrating lymphocytes around the salivary glands are CD4+ T cells, and that CD8+ T cells are present in the salivary duct epithelium (160, 161). Although the lacrimal and salivary glands are referred to predominantly, other organs are frequently affected, including the lungs (162-164), kidneys (165, 166), pancreas (167, 168) and skin (169, 170). There is also an increased incidence of

lymphoma in these patients (171). The terms secondary Sjogren's syndrome (2°SS) is used when describing patients who also have a definable connective tissue disorder. 2°SS is observed in 30-50% of patients with rheumatoid arthritis (172), and in 20-30% of patients with systemic lupus erythematosus (SLE) (173). It is also described with other connective tissue disorders such as scleroderma (174), dermatomyositis (175) and is a feature of 'mixed connective tissue disease' (176).

"Primary" Sjogren's syndrome is generally applied to patients without features of other connective tissue disorders, although other autoimmune conditions may co-exist, such as autoimmune thyroid disease (177, 178), or primary biliary cirrhosis (179), where the distinction between 1°SS and 2°SS becomes confusing. There can also problems in classifying 1°SS patients with systemic features resembling those of other connective tissue disorder such as SLE.

Much of the confusion with respect to terminology in Sjogren's syndrome has resulted from a lack of universally accepted diagnostic criteria. In an attempt to clarify this situation, new diagnostic criteria have been developed (180, 181). The European classification for the diagnosis of Sjogren's syndrome is shown in Box 1. In this classification, symptomatic dry eyes and dry mouth are defined by positive responses to a questionnaire. Objective evidence of ocular and salivary involvement, as assessed by tear production and salivary flow rate, together with the histopathological and serological features, make up the six criteria, of which four need to be fulfilled in order to establish a diagnosis of primary Sjogren's syndrome.

**European criteria for primary Sjogren's syndrome**

- Ocular symptoms (positive response to one of three question below)
- Oral symptoms (positive response to one of three questions below)
- Ocular signs (Schirmer's test <5mm wetting in 5 minutes or positive Rose-Bengal staining of conjunctiva)
- Lymphocytic infiltrate on salivary gland biopsy
- Objective evidence of salivary gland involvement (unstimulated salivary flow <1.5mls in 15 minutes)
- Serology (anti-Ro or anti-La antibody positive)*

4 out of 6 criteria required for a diagnosis of 1°SS

*Although the autoantibodies ANA and RF were included as serological markers in both Fox's criteria and the preliminary European criteria, this was subsequently restricted to positive Ro and/or La antibodies following an assessment of the European classification.

**European symptom questionnaire**

Ocular symptoms

- Have you had daily, persistent, troublesome dry eyes for more than 3 months?
- Do you have a recurrent sensation of sand or gravel in the eyes?
- Do you use a tear substitute more than three times a day?

Oral symptoms

- Have you had a daily feeling of dry mouth for more than 3 months?
- Have you had recurrent or persistently swollen salivary glands as an adult?
- Do you frequently drink liquids to aid swallowing dry foods?

**Box 1**

A number of characteristic laboratory features are found in 1°SS, which may reflect an over active immune system in these patients. These include autoantibodies to extractable nuclear antigens (158, 182) (anti-Ro and anti-La antibodies) and increased polyclonal immunoglobulin production (157, 183), which results in often dramatic hypergammaglobulinaemia, with serum immunoglobulin levels significantly higher than the upper limit of the normal range. Severe hypergammaglobulinaemia occasionally results in hyperviscositiy syndromes (184, 185), with arterial or venous occlusive events. The immunochemical properties of the abnormal immunoglobulin sometimes results in the formation of cryoglobulins (186), which typically manifests as a palpable purpuric skin rash. The anti-Ro and anti-La autoantibodies are associated with

systemic features of the condition, such as lung involvement and renal tubular acidosis (164, 187). The increased incidence of lymphoma in these patients also appears to be restricted to those who are anti-Ro antibody positive (188). It is however unclear whether these autoantibodities are pathogenic, or are simply markers of the disease (189).

### 1.4.1 Retroviruses and Sjogren's syndrome

Several strands of evidence implicate retroviruses as potential causal factors in 1°SS, although definitive proof of a viral aetiology has not been established. Clinical features resembling 1°SS develop in some patients infected with the human T cell leukaemia virus type 1 (HTLV-1) (190, 191) and human immunodeficiency virus (HIV) (192), although these patients do not generally develop the Ro and La autoantibodies characteristic of 1°SS (193). In a study by Talal *et al* (194), antibodies to HIV-1 *gag* p24 protein were found in serum samples of 14 of 47 (30%) patients with 1°SS, with two samples also reacting with HIV-1 p17 protein. The patients with positive serology for HIV-1 proteins were anti-Ro and anti-La antibody negative, a negative association that was also found by other investigators (195, 196).

The absence of reactivity against native HIV-1 proteins in these studies ruled out infection with either HIV-1 or a closely related retrovirus. However, a retroviral particle antigenically related to HIV was detected in T-lymphoblastoid cells co-cultured with salivary gland biopsy material from a patient with Sjogren's syndrome (197). Primers to conserved regions of HIV-1 gag, pol and env genes did not amplify nucleic acid sequences from these cells by polymerase chain reaction, even under conditions of low stringency, indicating that they were not infected with a defective form of HIV. The retroviral particles were contained within intracytoplasmic vacuoles, and had a typical A-type morphology, and as a result were termed human intracisternal A-type particles

(hIAP). IAPs in mice are non-infectious endogenous retroviral structures whose transcripts are packaged in A-particles in the cysternae of the endoplasmic reticulum (114). They were first described as inclusion bodies in mouse mammary adenocarcinomas (198). They contain a polyadenylated RNA molecule, a reverse transcriptase and a gag-like 73 kDa protein, and may result from inadequate maturation processing of viral precursor polypeptides or from disturbed virus budding (199).

Evidence of a serological response to HTLV-1 *gag* proteins has also been found in 1°SS patients. Antibodies against synthetic peptides representing the major epitopes on HTLV-1 p19 *gag* and a homologous sequence on the endogenous retrovirus HRES-1 were found in 32% of 1°SS patients (200). Salivary gland biopsies from 31% of 39 patients with 1°SS contained an epithelial cytoplasmic protein reactive with a monoclonal antibody to HTLV-1 p19 *gag*. This antibody also detected antigen in the labial biopsies from 24% of 17 patients with 2°SS, 21% of 14 patients with sicca symptoms and 12.5% of 16 patients with other connective tissue diseases (201). There are also reports that the *tax* gene of HTLV-1 may be expressed in some 1°SS patients. HTLV-1 *tax* sequences, but not *gag*, *pol* or *env* genes, were detected in the minor salivary glands of two out of nine patients with Sjogren's syndrome using in situ hybridisation and PCR (202). In a second study from Japan (203), HTLV1 *tax* mRNA was detected by RT-PCR in the minor salivary glands of 4 out of 14 patients with 1°SS, again in the absence of *gag*, *pol* and *env* retroviral gene expression.

These observations have led to the suggestion that endogenous retroviral gene products are expressed in the salivary glands of some patients with 1°SS. There also appear to be antibodies in the serum of some patients that can bind to these retroviral proteins. While it is tempting to speculate that the expression of ERV gene products in salivary gland tissue results in an immune response that subsequently destroys the

gland, there are alternative explanations that need to be considered. Antibodies from patients with 1°SS may cross-react with retroviral gene products, and likewise monoclonal antibodies raised to retroviral gene products may cross-react with non-viral salivary epithelial proteins expressed in 1°SS patients. Activation of endogenous proviruses can be induced *in vitro* by a variety of stimuli, including chemical and physical, or in response to hormones or exogenous viruses (114). Immune responses induced by mitogens or allogeneic cells may also stimulate expression of ERV proteins (154). Expression of retroviral genes in 1°SS salivary glands could therefore result from, rather than cause the lymphocytic infiltration characteristic of the condition. A subsequent immune response to the ERV proteins could perhaps result in a persistent immune response, leading to chronic glandular damage. However, it remains possible that the observed expression of endogenous retroviruses and antibodies to retroviral proteins in 1°SS are simply consequences of the increased immune reactivity seen in this condition, and do not have any pathological consequences.

### 1.4.2 Evidence of a causal relationship between retroviruses and autoimmunity

The observation that some patients infected with HTLV-1 (190, 191) and HIV-1 (192) develop clinical features resembling 1°SS suggests that these exogenous retroviruses can stimulate immune mediated glandular destruction. It has been suggested that infection with an exogenous virus similar to an ERV could result in loss of tolerance to ERV gene products, resulting in a persistent autoimmune response (154). In a transgenic mouse model expressing the cell membrane associated glycoprotein (G) of vesicular stomatitis virus (VSV) as self-antigen, autoantibodies were produced to VSV-G after infection with wild-type VSV (204). In a similar way, HTLV-1 and HIV-1 infection could potentially trigger a response to related ERVs resulting in autoimmune features in some patients.

Retroviral genes could stimulate autoimmune responses through mechanisms other than expression of an autoantigen. An exocrinopathy resembling Sjogren's syndrome develops in mice transgenic for the *tax* gene of HTLV-1 (205). The *tax* gene, situated in a 1.5 kb region flanked by the *env* gene and 3' long terminal repeat (LTR) of HTLV-1, encodes a 40 kDa protein (206). It has been proposed that this protein transactivates the viral long terminal repeat (205) and, possibly a variety of cellular genes, including the interleukin-2 receptor (IL-2R) (207) and interleukin-2 (IL-2) (208), or other lymphokines involved in T helper cell proliferation such as IL-4 (209). Mice transgenic for the HTLV-1 tax gene display histopathological features in the salivary and lacrimal glands that resemble those of 1°SS (205). The extent of histopathological changes correlate directly with the concentration of the tax protein expressed in the nuclei of the glandular epithelial cells. It has been suggested that salivary gland destruction could be a consequence HTLV-1 *tax* transactivation of cytokine genes, which could provide an important 'second signal' for lymphocyte activation.

Another approach to provide evidence of a causal link between retroviruses and autoimmune disease has been to look for 'disease-specific' retroviruses. Although a number of promising retroviruses have recently been described, subsequent studies have shown these retroviral genes to be more widely distributed in the population than first thought. An apparently exogenous retrovirus, HRV-5, was cloned by reverse transcriptase PCR from salivary gland tissue of a patient with Sjogren's syndrome (210). This retrovirus is related to simian D-type retroviruses, rodent intracisternal A-type particles and MMTV, and was not detectable in normal human DNA by Southern blotting or PCR. It was present in a sucrose density gradient fraction corresponding to that of an enveloped retrovirus particle. In subsequent experiments (211) using nested polymerase chain reaction (PCR), HRV-5 proviral DNA was found in salivary glands of only two out of ninety two patients (55 Sjogren's syndrome patients, 37 non-Sjogren's

syndrome patients). One was from a patient who had sicca symptoms but who did not satisfy the criteria for a diagnosis of Sjogren's syndrome. The other was from a patient with secondary Sjogren's syndrome. Using the same method, HRV-5 proviral DNA was detected in synovial tissue from a proportion of patients with a number of rheumatological conditions, including rheumatoid arthritis (12/25 patients), reactive arthritis (3/5), psoriatic arthritis (2/2) and osteoarthritis (3/5) (212).

A novel retrovirus has recently been isolated from patients with multiple sclerosis (MS), and has been termed MS-associated retrovirus (MSRV) (213). Retroviral *pol* fragments were isolated from RNA-purified extracellular viral particles from leptomeningeal and choroid plexus tissue of MS patients using RT-PCR. These now appear to be transcribed from a novel family of endogenous elements, HERV-W (128) which are expressed in normal placental tissue (214). An endogenous retrovirus related to HERV-K has also recently been implicated in the pathogenesis of insulin-dependent diabetes mellitus (IDDM), and has been termed IDDMK1,2-22 (215) although this work has since been discredited. It had been suggested that this ERV encoded a superantigen, based on the observation of Vβ-specific expansion of T cells in the pancreatic glands of two patient with early onset IDDM. Similar Vβ-specific responses are seen in 1°SS as will be discussed below.

### 1.4.3 TCR Vβ-specific T cells in Sjogren's syndrome

Analysis of the TCR variable gene repertoire of T cells in inflammatory sites of autoimmune diseases has provided insights into the nature of pathogenic auto-reactive T cells and the antigens stimulating them. As superantigens stimulate T cells through binding with the TCR Vβ chain, it would be expected that T cells at sites of superantigen-mediated inflammation would share one, or a few TCR Vβ gene products. Furthermore, as the hypervariable region of the TCR (encoded by junctional

(J) and diversity (D) region as well V region genes) are not significantly involved in superantigen-T cell interactions, it would be anticipated that within the expanded Vβ populations, J and D region genes would be randomly represented. Responses to conventional antigens would give a different pattern on examination of T cells at sites of inflammation. A single conventional antigen normally only expands a few clones of T cells, because the frequency of T cells specific for a given antigen is low. As conventional antigens are recognised primarily by the hypervariable region of the TCR, this oligoclonality would be expected in V, J and D region genes.

TCR populations have been studied by quantitative assays of mRNA Vβ, D and J region gene transcripts, or by using monoclonal antibodies specific for expressed variable region gene products. Studies carried out on TCR usage of infiltrating T cells in Sjogren's syndrome minor salivary and lacrimal glands have given conflicting results. Using a semi-quantitative RT-PCR method, Sumida *et al* (216) reported that Vβ2 and Vβ13 transcripts were predominantly expressed in T cells from minor salivary glands of 1°SS patients. In six of seven patients studied, 8.6-16.8% of T cells expressed Vβ2 and in four of seven patients, 13-29.2% T cells expressed Vβ13. The predominance of Vβ2 and Vβ13 genes was apparently specific for salivary glands, because lesser percentages of these two genes were expressed in peripheral blood lymphocytes. The junctional sequences utilised by Vβ2 and Vβ13 bearing T cells from the salivary glands of three of these patients were subsequently analysed (217). RNA was amplified by RT-PCR using Vβ2 or Vβ13 specific primers together with a Cβ primer, and the products cloned. Of 41 Vβ2 transcripts isolated, 34 clones contained different junctional sequences, suggesting that the infiltrating Vβ2-bearing T cells were polyclonal. Of 45 cDNA clones encoding the Vβ13 gene, 35 represented different junctional sequences. However, the junctional sequence gene Jβ2.3 was present on

44% and 46% of Vβ2 positive clones isolated from two of the three patients, whereas Jβ2.1 was found in 23% and 45% of Vβ13 positive clones from the same two patients.

Although the high prevalence of Vβ2 and Vβ13 positive T cells in some 1°SS salivary glands could be explained by a superantigen response, the preferential usage of junctional region genes in some Vβ2 and Vβ13 clones is more in keeping with the response to a conventional antigen. A subsequent study of TCR Vβ usage in the lacrimal glands of five Sjogren's patients (four 1°SS, one 2°SS) did not find a predominance of Vβ2 or Vβ13 positive T cells amongst the infiltrating T cells (218). Although other Vβ families (including Vβ3, Vβ9, Vβ10, Vβ11, Vβ15 and Vβ17) were over-expressed in lacrimal glands compared with peripheral blood lymphocytes, these varied depending on the patient studied, and overall the infiltrating T cells appeared to be polyclonal. In a study using monoclonal antibodies directed against the TCR Vβ region (219), an increase in lymphocytes bearing Vβ2 family gene products was found in peripheral blood, and an increase in both Vβ2 and Vβ8 in the salivary gland infiltrates of eight patients with 1°SS. The TCR Vβ2 gene was also expressed predominantly in non-malignant parotid lymphoproliferative lesions from seven 1°SS patients (220) and in the kidneys of six of seven 1°SS patients with interstitial nephritis (221). Junctional sequences of cDNAs encoding the Vβ2 gene on infiltrating T cells in the kidneys of five 1°SS patients showed that some of the cells expanded clonally, suggesting conventional antigen-driven stimulation rather than superantigen-induced proliferation. However, the Vβ2 clones isolated from the kidney were different from clones isolated from salivary glands of the same patients. Identical clones have however been isolated from both lacrimal and salivary glands in some 1°SS patients (222).

Mechanisms other than the response to superantigens have been suggested to account for the preferential use of TCR Vβ genes in the T cells infiltrating salivary glands in 1°SS patients. Genetic studies have identified polymorphisms in several TCR Vβ gene coding regions that could account for differences in levels of expression (223-225). Allelic differences in the promoter region of Vβ13 have also been identified (226). The prevalence of TCRBV13S2*2 homozygotes is significantly increased in 1°SS patients with hypergammaglobulinaemia (227), suggesting that allelic differences between TCR Vβ genes could account for the high prevalence of Vβ13 positive T cells in some patients. Other studies have shown preferential usage of certain Vβ genes in the T cell repertoire of normal individuals (35, 228, 229). Vβ2 and Vβ13.1 T cells appear to be more prevalent than other Vβ bearing T cells, with approximately 10% of peripheral blood T cells expressing Vβ2 and 4-13% expressing Vβ13.1. The preferential use of Vβ genes in 1°SS could therefore be reflecting more generalised mechanisms, resulting in 'skewing' of the T cell receptor repertoire.

It is clear from analysis of the human T cell receptor repertoire, that large deletions of specific Vβ-bearing T cells do not occur in the way described in mice, where complete Vβ-specific deletions result from T cells interacting with vSag in the thymus. There is, however, evidence that the variations in Vβ expression seen in humans results from a process of selection in the thymus, rather than simply from Vβ gene allelic differences. In a study of the human TCR Vβ gene repertoire in thymus glands of children undergoing cardiovascular surgery (225), significant variations in Vβ gene transcript levels were observed in mature (CD4+ or CD8+ single positive) but not immature (CD4+ CD8+ double positive) thymocytes. As these differences in Vβ expression occurred during the transition from immature to mature T cells, they presumably occurred through a process of selection, possibly by interacting with an endogenous

superantigen. Marked differences in the percentage of peripheral blood T cell bearing Vβ2 have been observed in a normal population (228), where the percentage of Vβ2+ CD4+ T cells was distributed in a bimodal fashion. Individuals were either 'Vβ2 high' or 'Vβ2 low', and family studies indicated that this phenotype was inherited independently from HLA genes or Vβ gene allelic differences. These authors suggested that the 'Vβ2 low' phenotype could result from the expression of an inherited endogenous superantigen, causing incomplete deletion of Vβ2 bearing T cells in the thymus.

The observations described above raise the intriguing possibility that endogenous superantigens are present in humans. The identification and characterisation of human DNA sequences related to vSag will now be described in detail.

# 2 CHAPTER TWO

## Isolation of human DNA sequences related to vSag

### *2.1 Introduction*

Although DNA sequences related to the murine retrovirus MMTV are present in the human genome, human DNA sequences related to the 3'LTR encoded superantigen (vSag) have not been identified. The aim of the work described in this chapter was to identify and isolate human DNA sequences homologous to the two regions of vSag which are most highly conserved between different strain of MMTV. The first conserved region (vSagC1) is 95 bp in length. The predicted amino acid sequence is hydrophobic, and is a putative trans-membrane region. The function of the second conserved region (vSagC2) is unknown, but the 198 bp sequence is predicted to encode a region of the protein lying within an extracellular domain of the superantigen which maybe involved in binding with MHC molecules.

Two strategies were employed to identify human DNA sequences related to vSagC1 and vSagC2. The first used the polymerase chain reaction (PCR) with oligonucleotide primers specific for conserved regions of vSag. PCR primers were selected to amplify 91 and 182 bp products from these two conserved region and also to span the C1 and C2 domains, to give a predicted product of 501 bp (Fig. 2.1). PCR products amplified from normal human placental DNA were subsequently cloned and those with greatest sequence homology to vSag were used to probe a human placental genomic library, thereby isolating the genomic loci from which these PCR products were amplified.

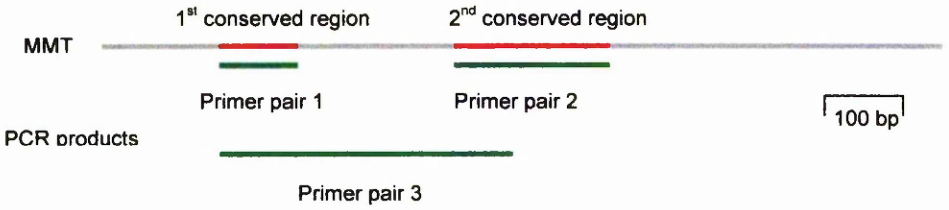



**Figure 2.1 Regions of vSag amplified by PCR primers.** PCR primers to conserved regions of MMTV (vSagC1 and vSagC2) and to the region spanning these conserved regions are indicated.

The second strategy for identifying human sequences related to vSag involved direct screening of the human genomic library with murine DNA from the conserved regions of vSag. These probes were generated from an MMTV-containing plasmid with the PCR primers to vSagC1 and vSagC2. By hybridising under conditions of reduced stringency, any significantly related sequences would be identified, without depending on accurate matching at the regions of the PCR primers.

## *2.2 Materials and Methods*

The methods detailed here are limited to those which were unique, or particularly important to this section of work. Some methods of more widespread use in other sections of this thesis are detailed in the chapter to which they are most relevant. All reagents were of "Analar" quality, and obtained from major commercial UK suppliers unless otherwise stated. All were stored and used according to manufacturers' recommendations.

### 2.2.1 Amplification by Polymerases Chain Reaction with vSag specific primers.

Specific Oligonucleotide primers for PCR were synthesised using a Beckman 1000M DNA synthesiser and their concentration adjusted to 1 mg/ml in TE (10mM Tris pH8, 1 mM EDTA). The primer sequences were derived from the published mouse mammary tumour virus (GR) proviral 3' long terminal repeat (230). Primer pairs specific for vSagC1, vSagC2 and the intervening region were combined as shown below (each primer at 100 μg/ml).

| | | | |
|---|---|---|---|
| Primer pair 1: MMTV 3'LTR (vSagC1) | sense | 210 | TCTGCTGCAAACTTGGCATAGC |
| | antisense | 279 | GGGCTCTCACCCTTGACTCTTT |
| Primer pair 2: MMTV 3'LTR (vSagC2) | sense | 661 | CAAAATAGGAGACAGGTGGTGG |
| | antisense | 821 | CGTGAAAGACTCGCCAGAGCTA |
| Primer pair 3: MMTV 3'LTR (vSagC1-C2) | sense | 210 | TCTGCTGCAAACTTGGCATAGC |
| | antisense | 689 | GGGACTTATAGGGGACCTTACA |

A 1.5 kb Pst 1 fragment of MMTV 3'LTR DNA in the plasmid pBR322 was linearised prior to amplification by digestion with *Pst*I and the concentration adjusted to 1 μg/ml. DNA extracted from murine (exogenous MMTV $GR^+$ or $C3H^+$) and normal human placenta were diluted to 100 μg/ml. Plasmid DNA (1 ng) and murine or human placental DNA (100 ng) were amplified in a thermal cycler in 25 μl volumes under mineral oil with 100 ng of each primer, 0.2 mM dNTPs, 2.5 μg BSA, PCR buffer (10 mM KCl, 20 mM Tris-HCl (pH 8.8 at 25°C), 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100) and Taq polymerase (0.625 units/reaction). Cycles were as follows: 94°C denaturation for 90 seconds, 50°C annealing for 30 seconds and 72°C extension for 30 seconds, followed by 25 cycles of 94°C, 50°C and 72°C each for 30 seconds, with a final extension at 72°C for 120 seconds. PCR products were electrophoresed on 3% agarose minigels (1 g agarose type I (Sigma), 2 g Nusieve (FMC Bioproducts))

in 100 mls TBE (0.9 M Tris-HCl, 0.9 M $H_3BO_3$, 25 mM $Na_2EDTA$)) containing 0.5 μg/ml ethidium bromide.

### 2.2.2 Cloning of PCR products into the pCR vector.

**Ligation and transformation.**

Fresh PCR products were ligated into the pCR vector (Invitrogen) at a 1:1 molar ratio by incubating overnight at 12°C with 50 ng of the vector in 10 μl volumes containing ligation buffer (50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 10mM DTT, 50 μg/ml BSA) and 0.3 units T4 DNA ligase. For each ligation an aliquot of transformation competent cells (OneShot™-Invitrogen) was thawed on ice and gently mixed with 2 μl β mercaptoethanol prior to adding 1 μl of the ligation reaction. After incubating on ice for 30 min, the cells were placed in a 42°C waterbath for 30 seconds and then returned to ice for a further 2 min before adding 450 μl SOC medium (2% w/v bacto-tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_47H_2O$, 20 mM glucose) and placed in a shaker incubator (37°C, 225 rpm) for 1 hour. The transformed cells were incubated overnight at 37°C on 1.5% agar L-agar plates (0.5% w/v yeast extract, 1% w/v bacto-tryptone, 300 mM NaCl, 1.5% w/v L-agar, pH 7.3) containing ampicillin (50 μg/ml) on which 25 μl of X-Gal (40 mg/ml in dimethylformamide) had been spread. The indicator X-Gal results in a blue colour change of transformed cells due to the lacZ gene in the vector. The ligation of a PCR product into the vector insertion site disrupts the lacZ reading frame, preventing the colour change in these colonies. White colonies were therefore selected for DNA minipreparation.

**Plasmid DNA minipreparations.**

The selected colonies were inoculated into 5ml of 2xTY medium (1.6% w/v bacto-tryptone, 1% w/v yeast extract, 300 mM NaCl, pH 7.3) containing 50 μg/ml ampicillin and the cultures grown up overnight, with vigorous shaking at 37°C. Approximately 1.5 ml of the overnight culture was poured into an Eppendorf tube and the cells collected by centrifugation at 5000 rpm for 5 min. The supernatant was then poured off and the bacterial pellet resuspended in 100 μl of a solution containing 50 mM glucose, 10 mM EDTA and 25 mM Tris-HCl (pH 8) and 4 mg/ml lysozyme. The cells were incubated in the lysozyme mixture for 5 min at room temperature, cooled on ice and lysed under alkaline conditions by adding two volumes of a solution of 0.2 M NaOH / 0.1% SDS and incubating on ice for 5 min. A half volume of ice cold 5 M potassium acetate was the added and the tube incubated on ice for 10 minutes before centrifuging at 10000 rpm for 10 min. The supernatant was transferred to a fresh tube and the plasmid DNA extracted once with phenol/chloroform, before precipitating with ethanol at room temperature for 2 min. The DNA pellet was washed with 70% ethanol, dried under vacuum and redissolved in 50 μl TE containing RNase (20 μg/ml). An aliquot of the DNA was then digested with *Eco*RI to cut out the inserted PCR product. The digest was analysed on an agarose gel with *Hinf*I digested pUC 19 as a molecular weight marker.

### 2.2.3 Sequencing of PCR products

Plasmids containing recombinants were sequenced using dye terminator cycle sequencing (Perkin Elmer) with AmpliTaq® DNA Polymerase, FS, which was carried out by the staff of the Molecular Biology Facility at Newcastle University. This was undertaken using an ABI 377 DNA Sequencer (Perkin Elmer), which detects fluorescence from four different dyes, used to identify the A, G, C and T extension reactions. Each dye emits light at a different wavelength when excited by a laser,

allowing all four reactions to be detected in a single gel lane. This strategy improves sequencing accuracy because it eliminated problems caused by variations in electrophoretic mobility from lane to lane.

Polymerase chain reactions were carried out in 20 μl volumes in a reaction mix containing 0.2 μg plasmid DNA, 3.2 pmole universal forward or reverse primers and 8 μl Terminator Premix (comprising A-Dye Terminator, C-Dye Terminator, G-Dye Terminator, T-Dye Terminator, dITP, dATP, dCTP, dTTP, Tris-HCl (pH 9.0), $MgCl_2$, thermal stable pyrophosphatase and AmpliTaq DNA Polymerase, in concentrations determined by the manufacturers). The thermal cycler settings were; 96°C for 30 seconds, 50°C of 15 seconds, 60°C for 4 minutes, repeated for 25 cycles, then held at 4°C prior to purification of the extension products.

Excess dye terminators were remove by ethanol precipitation. The 20 μl contents of each PCR tube was added to 2 μl 3 M sodium acetate (pH 4.6) and 50 μl 95% ethanol in a 1.5 ml microcentrifuge tube, vortexed and incubated on ice for 10 minutes. After centrifuging in a microcentifuge at maximum speed for 15 minutes, the ethanol solution was aspirated with a micropipetter and discarded. The DNA pellet was washed by adding 250 μl 70% ethanol. The ethanol solution was discarded, and the pellet dried in a vacuum centrifuge.

The DNA pellet was re-suspended in 6 μl loading buffer (deionized formamide and 25 mM EDTA (pH8) containing 50 mg/ml Blue dextran in a ratio of 5:1 formamide to EDTA/Blue dextran), denatured by incubating at 90°C for two minute and 1.5 μl of each sample run on a 36 cm well-to-read polyacrylamide gel.

### 2.2.4 Screening a human placental DNA library with vSag-related probes.

A library of human placental DNA, partially digested with *Mbo* I, was constructed using standard techniques in the λ Dash replacement vector (provided by Dr F E May). Large petri-dishes (140mm diameter) were prepared by adding agar (1%) to Phage-broth (0.5% w/v yeast extract, 1% w/v bacto-tryptone, 300 mM NaCl, 0.2% maltose, 10 mM $MgSO_4$) which was autoclaved and cooled to 60°C prior to pouring. Surface moisture was dried off the plates by placing inverted in a 37°C incubator. Top agarose was prepared in a similar manner (but containing only 0.65% agarose) and cooled to 48°C in a waterbath. When the plates were sufficiently dry and the top agarose was at the correct temperature, 0.5 μl of the λ phage placental library was added to 250 μl of $P_2$PLK cells (which had been cultured overnight in L-broth containing 0.2% maltose, centrifuged and resuspended in 10 mM $MgSO_4$ to give an $OD^{600}$ of 0.8) in a sterile 13 ml tube (one per plate). The phage were allowed to adsorb by incubating at 37°C for 15 minutes, after which 7.5 mls of the top agarose was added to the tube and the contents poured onto the agar plates. The bacteria were spread out over the surface of the agar by carefully swirling the petri-dish. The plates were incubated at 37°C overnight to produce lysis. The number of plaques per plate was determined which resulted in the plaques filling the surface of the plate and just touching each other, but without producing confluent lysis.

**Transfer of phage DNA to nitrocellulose**

Recombinant phage were transferred to circular nitrocellulose filters of sufficient size to cover the surface of the plate. The filters were placed onto the agarose surface and left for 10 minutes. The orientation of the filter with respect to the plate was marked during this period, by making a non-symmetrical pattern of four holes in the outer region of the filter using a sterile needle. This pattern was also marked on the base of the plate with a permanent marker pen. The filters were removed from the plate after 10 minutes and placed plaque-side up on Whatman paper to dry, for at least 1 hour.

The dry filters were then laid sequentially onto the surface of three transfer solutions, for 10 seconds each. The transfer solutions (in the order of the treatment sequence) were made up of ; 0.2 M NaOH / 1.5 M NaCl (denaturing solution); 0.4 M Tris-HCl (pH 7.5) / 2 x SSC (neutralising solution) and 2 x SSC alone (wash solution). The treated nitrocellulose was blotted nominally dry and then dried under vacuum for 2 hours at 80°C.

**Preparation of probes for screening the placental DNA library**

The placental DNA library was screened with $^{32}$P-labelled DNA probes synthesised from the human vSag related sequences. PCR products of HRC1and HRC2 were purified prior to use as hybridisation probes. 100 μl of chloroform was added to each PCR product (100 μl + mineral oil). The mixture was vortexed, then centrifuged for 3 minutes at 5000 rpm and the aqueous phase transferred onto 150 μl phenol:chloroform. 50 μl of TE pH 8 was added back to the chloroform, revortexed, recentrifuged and the aqueous phase transferred onto the phenol:chloroform. This was vortexed thoroughly and centrifuged for 5 minutes at 10000 rpm. The supernatant was transferred onto 150 μl chloroform and reextracted. Meanwhile the phenol:chloroform was back extracted with 100 μl TE pH 8, reextracted on the chloroform and the combined aqueous phases collected in a fresh Eppendorf tube. 0.2 μl glycogen, 5 μl 5M NaCl and 600 μl ethanol was added and the DNA precipitated overnight at -20°C. After centrifuging for 10 minutes at 15000 rpm, the DNA pellet was washed in 500 μl 70% ethanol, recentrifuged and dried under vacuum. The DNA was redissolved in 25 μl 1/10 TE and passed over a sephadex G50 minicolumn equilibrated with 1/10 TE. Each sample was eluted with 300 μl 1/10 TE, and 6 x 50 μl fractions collected. 2.5 μl of each fraction was analysed on a 0.8% agarose minigel alongside DNA markers. The fractions containing >10ng DNA were combined and the final concentration estimated.

High specific activity DNA probes were produced by incorporating $^{32}$P-dCTP into the purified PCR products. This method employs DNase I to introduce single strand breaks (or "nicks") into the template, which the allows the enzyme DNA polymerase I to start removing nucleotides at these breaks, replacing them with radiolabelled nucleotides which are introduced into the reaction mixture. The synthesis of all probes was carried out using a commercial nick translation kit (Pharmacia) with 10 ng DNA in a volume of 10 μl (scaled up as necessary). The DNA was first denatured by heating at 98°C for 3 minutes, then a reaction set up containing the DNA, 20 μM (dATP, dGTP and dTTP), 0.5 MBq $^{32}$P-dCTP, 0.5 units DNA polymerase I and 10 pg DNase I. (This mixture, made up from the kit components, also contained Tris-HCl, $MgCl_2$, glycerol, bovine serum albumin and 2-mercaptoethanol; at concentrations not specified by the manufacturers). Following incubation for at least 1 hour at 37°C, the reaction was terminated by the addition of Nick stop buffer. DNA was separated from unincorporated nucleotides by passage over a sephadex G50 minicolumn, eluting with 300 μl TE pH 8. Incorporation of the $^{32}$P label was determined by absorption of a 1 μl aliquot of the probe onto filter paper and measuring counts per minute (cpm) in a scintillation counter.

**Hybridisation**

Hybridisations of the filter bound DNA were carried out in double heat-sealed plastic bags, submerged in a waterbath at 37°C. A 88.5% volume hybridisation mix was used, the volume being made up with salmon sperm DNA (4 mg/ml stock) for the prehybridisations and with salmon sperm plus probe in the case of the hybridisations. The stringency of the hybridisation was determined by the percentage of formamide in the hybridisation solution. For high stringency reactions both prehybridisation and hybridisation solutions contained 50% recrystallised formamide. In addition the prehybridisation and hybridisation solutions contained 4 x SET (100 mM Tris, 0.6 M

NaCl, 10 mM EDTA), 0.1% sodium pyrophosphate, 0.25 mg/ml yeast RNA, 0.2% filtered SDS and 5 x Denhardt's solution (0.1% ficoll, 0.1% polyvinylpyrolidone and 200 μg/ml bovine serum albumin).

The prehybridisation solutions were made up allowing 5 ml per filter and approximately ¾ of this total volume put into a 140 mm petri-dish. Filters were then soaked in the dish, one at a time to form a stack, taking care not to trap air between them and the remaining prehybridisation solution added. Prehybridisations were carried out overnight at room temperature with the petri-dishes placed on a rotary shaker.

Filters were hybridised with a mixture of the probes HRC1 and HRC2 in 50% formamide. Screening was carried with each probe at 10 kBq/ml. The appropriate volume of probe was added to salmon sperm DNA (to give 10% of the hybridisation volume) and denatured by adding 10 M NaOH to a final concentration of 0.7 M. After 10 minutes at room temperature, the probe solution was neutralised by adding 10 N HCl. The resulting solution then had a volume of 11.5% that of the final hybridisation volume. The denatured probe/salmon sperm DNA was then mixed with the required volume of the 88.5% hybridisation solution, such that the final solution had a composition of 0.6 M NaCl, 20 mM EDTA, 200 mM Tris-HCl (pH 8), 0.1% sodium pyrophosphate, 5 x Denhardt's, 250 μg/ml yeast RNA, 0.2% SDS and either 25% or 50% recrystallised formamide. The filters were stacked in the petri-dish as before, but then transferred to a sealing bag (as a stack) and the remaining solution added. The bags were sealed, taking care to exclude air bubbles, and hybridised for 48 hours at 37°C.

The filters were washed in 2 x SSC / 0.1% SDS at room temperature for 30 minutes (2 x 15 minutes), followed by 2 hours (2 x 1 hour) at 65°C. After drying on filter paper, the nitrocellulose filters were exposed against preflashed X-ray film at -70°C.

**Second round screening**

Agarose plugs corresponding to positive hybridisation signals on the autoradiographs were removed from the relevant plate with a sterile loop, after cutting the surface with the "wrong" end of a pasteur pipette. The positive recombinant phage were allowed to diffuse out of the agarose into 2 ml of TMG/NaCl overnight. The positions of the positive plaques were identified by marking the positions of the four holes in the filter onto the film and aligning the plates on top of the film. The phage obtained from the agarose plugs were not pure at this stage (due to the high plaque density) and further rounds of screening were required to isolate the phage containing the hybridising DNA. For the second round of screening, the phage obtained from positive first round plaques were diluted (10 μl phage in 1 ml TMG/NaCl) and plated out on $P_2$PLK cells, as described above (except using 90mm plates). 100μl diluted phage was added to 100 μl $P_2$PLK cells and preadsorbed for 15 minutes prior to adding 2.5 ml top agarose, which was immediately poured onto the prepared 1% agar plates. After overnight incubation at 37°C, the phage DNA was transferred to nitrocellulose filters and hybridised with $^{32}$P-labelled DNA probes, as for the first round.

**Third and forth round screening**

Positive plaques were removed from the second round plates using the tip of a pasteur pipette to cut the top agarose while gently applying suction to lift the gel plug. The plaques were put into 1 ml TMG/NaCl and left to diffuse into the buffer overnight. The phage were diluted (4 μl phage in 50 μl TMG/NaCl) and preadsorbed with 150 μl $P_2$PLK cells, then poured onto prepared agar plates in 2.5 ml top agarose as described above. Filters taken from these plates were hybridised as before. A fourth round of purification, following the third round protocol, gave pure recombinants for isolation.

### 2.2.5 Preparation of DNA from recombinants isolated from the DNA library

**Preparation of high titre lysates**

Pure phage plaques containing hybridising DNA were placed into 500 µl TMG/NaCl in 14 ml bacteriology tubes. After eluting overnight at 4°C, 50 µl $P_2$PLK cells (0.8 $OD^{600}$) were added, preadsorbed for 20 minutes at 37°C and then cultured overnight in 10ml L-broth containing 10 mM $MgSO_4$ (shaking vigorously in 100 ml flasks at 37°C). Following centrifugation at 2500 rpm (Beckman JS 13 rotor), the supernatants were transferred to Wheaton vials and two drops of chloroform added to ensure complete lysis. The titre of the lysates was determined by plating out a series of $P_2$PLK cells which had been preadsorbed with 10 fold dilutions of the phage lysates. By counting the number of plaques per plate, the plaques per ml of each lysate was calculated. The high titre lysates were sealed with parafilm and stored at 4°C.

**Large scale preparation of phage DNA**

PLK cells were cultured in 100 ml L-broth containing 0.2% maltose (shaking vigorously in a 1 litre flask at 37°C) to an $OD^{600}$ of 0.4. Following centrifugation at 2500 rpm (Beckman J 10 rotor), the cells were resuspended in 20 ml 10 mM $MgSO_4$ and aerated for 1 hour at 37°C. 2 ml PLK cells were preadsorbed with phage high titre lysates (2 x $10^6$ phage plaque forming units), then added to 500 ml prewarmed L-broth containing 0.01M $MgSO_4$ and incubated overnight in 2 litre flasks at 37°C with vigorous shaking. 5 ml chloroform was then added to each flask to help lyse the cells and incubated for a further 1 hour at 37°C. After adding 50 µl RNase (10 mg/ml) and 500 µl DNase (1 mg/ml), the flasks were incubated at room temperature for 30 minutes. Then 29.2 g NaCl was dissolved in each solution and incubated on ice for 1 hour. The solutions were transferred to 500 ml centrifuge tubes and spun for 10 minutes at 10000 rpm (Beckman J 10 rotor). The supernatants were poured back into

the flasks (which had been washed) and mixed with 50 g PEG 6000. After incubating on ice overnight, they were centrifuged for 20 minutes at 10000 rpm (Beckman J 10 rotor) and the supernatant discarded. The pellets were resuspended in 10 ml TMG/NaCl and added to 8 ml chloroform in 30 ml Corex tubes. The tubes were sealed with parafilm, vortexed thoroughly and centrifuged for 10 minutes at 8000 rpm (Beckman JS 13 rotor). The supernatants were then transferred to sterile tubes containing 9 g CsCl and mixed by inversion to dissolve the CsCl. Each solutions was transferred to "quick seal" Beckman centrifuge tubes, balanced and sealed excluding all bubbles. They were then centrifuged in a vacuum at 38000 rpm (Beckman 70TC rotor) at 16°C for 36 hours. Following centrifugation, the phage band visible on the CsCl gradient was aspirated by piercing the tube with the needle of a syringe. The contents of each syringe was emptied into dialysing tubing (which was knotted leaving air under pressure in the tubing) and placed in a 1 litre cylinder of dialysing fluid containing 10 mM $MgCl_2$, 10 mM NaCl and 50 mM TRIS-HCl (pH 8). After 150 minutes with 5 changes of dialysing fluid, the contents of the tubing was transferred to Eppendorf tubes, adding 0.5 M EDTA (1:25), Proteinase K (1:40) and 20% SDS (1:40) to the estimated volume in each tube. After incubating for 1 hour at 68°C, the phage DNA was extracted by adding to 2 ml phenol:chloroform in a series of Corex tubes. These were vortexed thoroughly and centrifuged for 5 minutes at 10000 rpm (Beckman JS 13 rotor). The aqueous phase of each tube was transferred to 2 ml phenol:chloroform in a second series of Corex tubes, then vortexed and centrifuged as before. The aqueous phase was then transferred to 2ml 4% isoamylalcohol in chloroform, vortexed, centrifuged and the aqueous phase transferred to a forth series of Corex tubes. The extraction was repeated adding 1 ml TE pH 8 to the first series of Corex tubes giving an estimated final volume of 2 ml in the forth series, to which 40 µl 5 M NaCl and 5 ml ethanol was added to precipitate the DNA. After centrifugation the DNA pellet was washed in 75% ethanol (leaving at room temperature for 30 minutes),

recentrifuged and the pellet dried under vacuum. The DNA was redissolved in 100-300 μl TE pH 8 (depending on the size of the pellet) and the yield estimated by running 1 μl *Eco*RI digested DNA on a minigel alongside a λ *Hind*III marker.

### 2.2.6 Preparation of murine probes to the conserved regions of vSag.

The oligonucleotide primers specific for vSagC1and vSagC2 were combined as shown below (each primer at 100 μg/ml). The 1.5 kb Pst 1 fragment of MMTV 3'LTR DNA in the plasmid pBR322 was linearised prior to PCR amplification by digestion with Pst1 and the concentration adjusted to 1 μg/ml. Two 50 μL PCR reactions were carried out for each primer pair with an MMTV plasmid concentration of 50 ng/ml and each PCR primer at 4 μg/ml. The thermal cycler settings were 94°C denaturation for 90 seconds, 50°C annealing for 30 seconds and 72°C extension for 30 seconds, followed by 30 cycles of 94°C, 50°C and 72°C each for 30 seconds, with a final extension at 72°C for 120 seconds.

| | | | | |
|---|---|---|---|---|
| Primer pair 1: (vSagC1) | MMTV 3'LTR | sense | 210 | TCTGCTGCAAACTTGGCATAGC |
| | | antisense | 279 | GGGCTCTCACCCTTGACTCTTT |
| Primer pair 2: (vSagC2) | MMTV 3'LTR | sense | 661 | CAAAATAGGAGACAGGTGGTGG |
| | | antisense | 821 | CGTGAAAGACTCGCCAGAGCTA |

The PCR products were run on 3% agarose minigels (1 g agarose type 1, 2 g Nusieve in 100 mls TBE) containing 0.5 μg/ml ethidium bromide. The products of primer pair 1 and 2 were of the predicted sizes for vSagC1 (92 bp) and vSagC2 (182 bp). The PCR products were purified prior to use as hybridisation probes as described in 2.2.4.

### 2.2.7 Screening a human placental DNA library with murine probes under reduced stringency conditions.

Screening of a non-amplified human genomic library was undertaken as described 2.2.4, using freshly prepared plates with the same λ phage placental library stock, and the recombinant phage transferred onto nitro-cellulose filters. The placental DNA library was screened with high specific activity DNA probes produced by incorporating $^{32}$P-dCTP into the purified PCR products vSagC1 and vSagC2 as described in chapter two.

Hybridisation of the filter bound DNA was carried out in double heat-sealed plastic bags, submerged in a waterbath at 37°C. A 88.5% volume hybridisation mix was used, the volume being made up with salmon sperm DNA (4 mg/ml stock) for the prehybridisations and with salmon sperm plus probe in the case of the hybridisations. In order to reduce the stringency of the hybridisation reaction, the percentage of formamide in the hybridisation solution was reduced to 25%, and each round of screening was carried out under these conditions. Filters were hybridised for 72 hours at 37°C with a mixture of the probes vSagC1 and vSagC2. To take account of the size difference between the probes, hybridisation was undertaken with vSagC1 at 10 kBq/ml and vSagC2 at 15 kBq/ml.

For filters hybridised under reduced stringency conditions, two wash solutions were prepared. Solution one: 3 x SSC, 0.2% SDS, salmon sperm DNA (0.01 mg/ml) and 0.1% sodium pyrophosphate. Solution two: 25% formamide, 5 x SSC, 0.2% SDS, salmon sperm DNA (0.01 mg/ml) and 0.1% sodium pyrophosphate. The filters were initially rinsed 4 times in solution one, including a 15 minute wash at room temperature. They were subsequently washed at 37°C for 2 hours (2 x 1 hour) in

solution two, followed by 90 minutes (3 x 30 minutes) in solution one. After drying on filter paper, the nitrocellulose filters were exposed against preflashed X-ray film at minus 70°C. Agarose plugs corresponding to positive hybridisation signals on the autoradiographs were removed from the relevant plate and further rounds of screening undertaken to isolate the λ clones containing the hybridising DNA. Recombinant λ phage DNA was prepared as described in 2.2.5.

## *2.3 Results*

### 2.3.1 PCR amplification of vSag and human vSag related products

Specific PCR primers for the two conserved regions of vSag (vSagC1 and vSagC2) and for the intervening region (vSagC1-C2) were synthesised to allow investigation of potentially related vSag sequences in human genomic DNA. The PCR assay conditions were optimised using a 3'LTR MMTV containing plasmid and murine DNA (MMTV $C3H^+$ or $GR^+$). PCR products of the predicted size were generated for each primer pair when the plasmid containing the 1.5 kb *Pst*I fragment of the MMTV 3' LTR was used as the template and identical sized band were obtained with the C3H and GR murine DNA (figure 2.2). The predicted size of the PCR products amplified by each primer pair is shown in Table 2.1.

| Primer pair | Sense (SN) primer | Antisense (ASN) primer | Region of MMTV LTR amplified | PCR product size (bp) |
|---|---|---|---|---|
| 1 | 210 | 279 | C1 | 91 |
| 2 | 661 | 821 | C2 | 182 |
| 3 | 210 | 689 | C1-C2 Intervening region | 501 |

**Table 2.1 Predicted size of vSag PCR products.** Sequence specific PCR primers were selected to allow amplification of the conserved regions of vSag (vSagC1 and vSagC2) and the intervening region. The numbers in columns two and three refer to local primer reference numbers.

These sequence specific primers were used to amplify PCR products from normal human placental DNA. More than one distinct band was visible for each primer pair when human placental DNA was used as the template, some of which were of similar size to the murine PCR product. PCR reactions were carried out at different annealing temperatures (45°, 50° and 55°), the lower temperatures producing less stringent conditions. At an annealing temperature of 45°, at least three additional bands were visible using primer pair 1. Primer pair 2 gave two distinct products, while primer pair 3 generated four bands in addition to the predicted 500bp product of the murine vSag (figure 2.2). With primer pair 1, a faint band was consistently identified in the absence of template DNA. This band was of similar size to the 91 bp product amplified from 3'LTR MMTV, but was still present after new reagents were used, making it unlikely that it resulted from contamination with MMTV DNA.

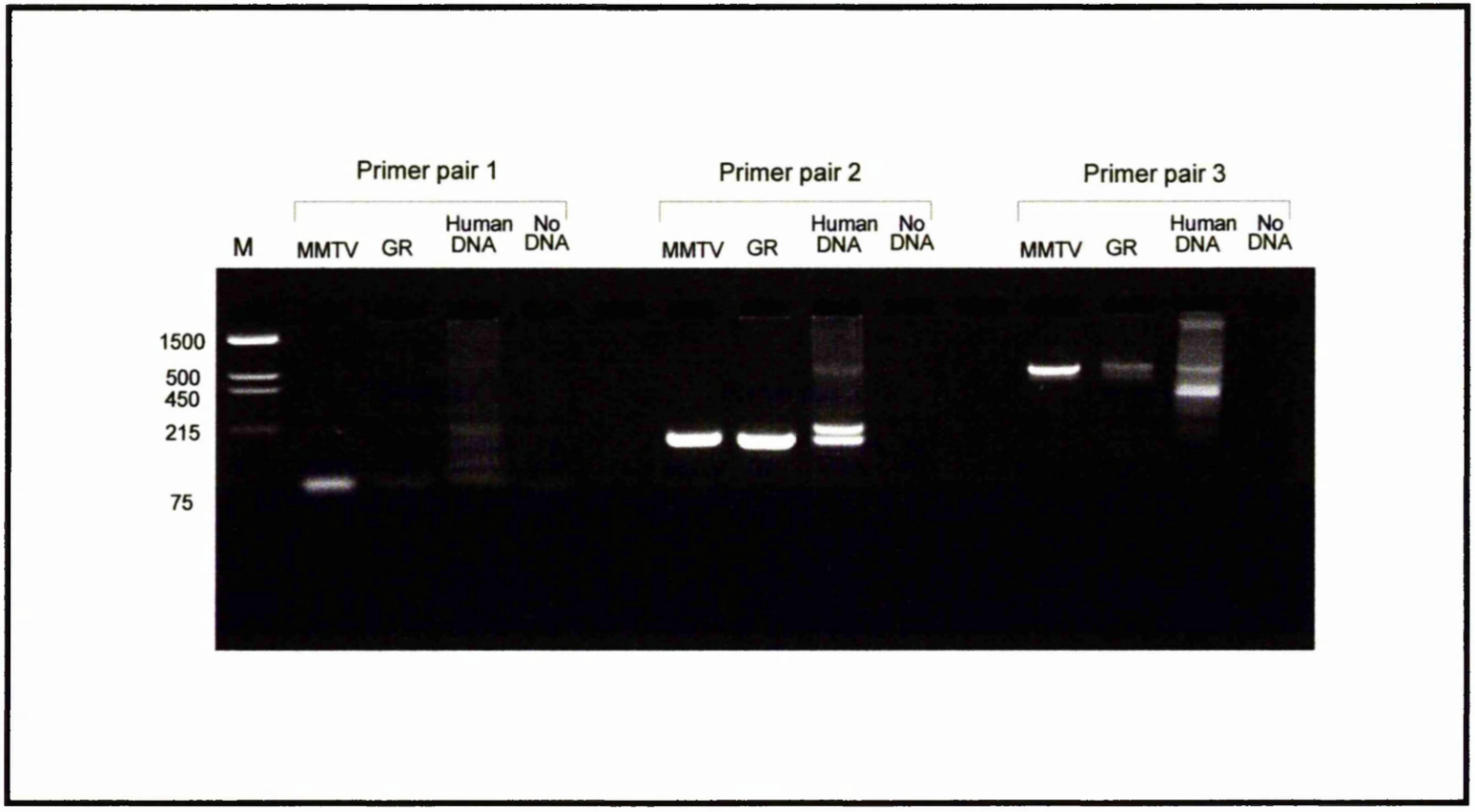



**Figure 2.2 PCR amplification using vSag sequence specific primers.** PCR products were amplified from an MMTV 3'LTR containing plasmid, GR+ murine DNA and human placental DNA with primers for the first and second conserved regions of vSag and the intervening region; vSagC1 (primer pair 1), vSagC2 (primer pair 2) and vSagC1-C2 (primer pair 3). Thermal cycler conditions were as described in section 2.2.1, with 45°C annealing temperature. MMTV: 1.5 kb *Pst*I fragment of MMTV 3'LTR in pBR322 plasmid, GR: murine DNA containing exogenous MMTV, Human DNA: normal human placental DNA, No DNA: negative control, M: molecular weight marker (bp).

As there was more than one PCR product in each reaction, all of which warranted investigation, the PCR products were cloned rather than directly sequenced. A TA cloning vector (pCR, Invitrogen) was used, which takes advantage of the non-template dependent activity of Taq polymerase which adds a single deoxyadenosine to the 3'-end of duplex molecules. The resultant 3' A-overhangs are used to insert the PCR product into the vector which contains single 3' T-overhangs at its insertion site. The different sized PCR products amplified from human DNA at an annealing temperature of 45°C were investigated. The PCR products from reactions using primer pair 1 (SN 210, ASN 279), primer pair 2 (SN 661, ASN 821) and primer pair 3 (SN 210, ASN 689) resulted in the isolation of 6, 2, and 4 different sized recombinants respectively (Table 2.2). The recombinants derived from each PCR product were then sequenced and analysed for homology to murine vSag.

### 2.3.2 DNA sequence analysis of potential human vSag-related PCR products

The purified recombinants isolated from normal human placental DNA were sequenced to identify those with greatest homology to murine vSag. Sequencing was performed using an ABI 377 DNA Sequencer (Perkin Elmer) with forward and reverse universal primers as described in section 2.2.3. The DNA sequence size of each PCR product derived from human DNA using primer pairs 1, 2 and 3 are shown in Table 2.2 and sequence data for the recombinants in Table 2.3.

| Sequence No. | Primer pair | Target DNA | Target DNA sequence size | Human DNA sequence size |
|---|---|---|---|---|
| **1** | **1** | **vSagC1** | **91** | **96** |
| 2 | 1 | vSagC1 | 91 | 128 |
| 3 | 1 | vSagC1 | 91 | 128 |
| 4 | 1 | vSagC1 | 91 | 169 |
| 5 | 1 | vSagC1 | 91 | 192 |
| 6 | 1 | vSagC1 | 91 | 211 |
| **7** | **2** | **vSagC2** | **182** | **176** |
| 8 | 2 | vSagC2 | 182 | 214 |
| 9 | 3 | vSagC1-C2 | 501 | 183 |
| 10 | 3 | vSagC1-C2 | 501 | 351 |
| 11 | 3 | vSagC1-C2 | 501 | 391 |
| 12 | 3 | vSagC1-C2 | 501 | 539 |

**Table 2.2 Sizes of cloned PCR products amplified from human placental DNA with sequence specific primers for vSagC1, vSagC2 and the intervening region (vSagC1-C2).** PCR products were cloned directly into the pCR vector (Invitrogen), without prior agarose purification to separate fragments. PCR products of a size closest to the murine sequences are highlighted in red.

| Seq. No | Target DNA | DNA SEQUENCE |
|---|---|---|
| **1*** | **vSagC1** | TCTGCTGCAA ACTTGGCATA GCTCTGAGTT CAACCATTTC TGCTGCGTTC GCAGCTGGTC<br>TCACAGGAAT TCGAGGGCTC TCACCCTTGA CTCTTT |
| 2 | vSagC1 | TCTGCTGCAA ACTTGGCATA GCCCAAATCA CTCCATGAAA TTGGGGTGTG GCATCTGCCT<br>CAAGGAGCAT TTCTACAACC TCTGCTGCCT CTACCGCAAA TGAAACGGGC TCTCACCCTT<br>GACTCTTT |
| 3 | vSagC1 | TCTGCTGCAA ACTTGGCATA GCTGAAAATA CCAATTAATT TTCAATAGTC CAGCCCATCA<br>TGAGGAAAAG CAGAAATGCC CATCTGGACA CTAAGAGAAT CCCTGTGGGC TCTCACCCTT<br>GACTCTTT |
| 4 | vSagC1 | TCTGCTCCAA ACTTGGCATA GCATTTACCA AAATTTGCCA GAGCAGACAG GCATCTCCAT<br>GTGAAAAGAA TGACTGTAAT GGGCCCCTGA GCCTTGCAGG GAGTATCTGG TCTCCATGAA<br>TGGAAAAGGT GGACATGCTA GGGTGGTGGG CTCTCACCCT TGACTCTTT |
| 5 | vSagC1 | TCTGCTCCAA ACTTGGCATA GCTGCTGTCA GGTGACTGAA GCATTTGGAC TTGAGAAGAA<br>CGCTTTTCCT CCCTGGATCA CGCATTACAA GTACATGCTT TTGAAACAGT GCTTCTGAAC<br>CAGGGTGGTT TTGCTCTCCA AGGACATTTG GCAACTTCTG GAGAAACTTA GGGCTCTCAC<br>CCTTGACTCT TT |
| 6 | vSagC1 | TCTGCTCCAA ACTTGGCATA GCATGAGCTT AGCCAAGTTA ATTAACCTCT CTGGNCCTCC<br>GTATCCTCAT CTGTGAAATG TAGGTAATTA TCTCCACCTC ATACCACTGG GGTAATGGCT<br>GGATGTCATA AAGGCACCTC CCATCATTAA CAGCTGTGTG GCCGGAGAGG CCATGGGCTC<br>TTTCATCCTG GGCTCTCACC CTTGACTCTT T |
| **7*** | **vSagC2** | CAAAATAGGA GACAGGTGGT GGAGGGTGGG ATCTGCCTGC CGAGAACGAG GAAGGGCATC<br>GGTGACAGGT AACAACCTGA GGCACTGGAG AGGGAGAGCA CGCGTGATGT TCCTCAGAAA<br>CGAGTCGCAT TTCCGCATAG CGCCTCCTTC AGACCGTGAA AGACTCGCCA GAGCTA |
| 8 | vSagC2 | CAAAATAGGA GACAGGTGGT GGTAAATGCC TATGTGGACA CAGTCCCACT GCTGCTGCTC<br>AGTGCCGGGA CCCTTGGCAA GCCGCCTGGG CTCTCTGACA CTCAGCCACT CTGCCTGTGA<br>AACAGGATCT TGGTAGTCAA GGGGTTGGAG AGATGGAATG GACGGGGCAG GCCAGGCACT<br>CAGCAAGCGC TCCGTGAAAG ACTCGCCAGA GCTA |
| 9 | vSagC1-C2 | TCTGCTGCAA ACTTGGCATA GCACTGGCAC TCAGTAGACC CTGGACAAGA ACTGCATGGA<br>TGGATGGATG GCCAGATGGG CTAGTAGGCA AGTAGTTACG GAGGACAGGG AAGAATAAGG<br>AATAAGACTA CCTTCAGTTA TTGGTACAAT CCTTTATCAT CATCCAGCTC TCTTTGTTCT<br>GAC |
| 10 | vSagC1-C2 | TCTGCTGCAA ACTTGGCATA GCCCCCAAAG GCCTCTGGGA GTGAGCAGTG GTTCACTTTG<br>GGAAGATCCA ACCTTGAAGA TTACAGAAGA CGTGAGGCAA CTGGTTCTTT TTCTAGACAA<br>CAGCACACAC CTAAGAGAGC AGGACATCTG AAGGCTATAG AAGCCAGTCC ACAATAATGT<br>CAAAACCCAG AATCCCAGGA GACCTAGGAC AGAAGGAACT GAAAAAGTGA TGAAAAATGG<br>CAACCAACTG GCCAAAACAG AGGAGCCAGA GCGAAGCCTA GAAATGTTGA TGATTCGGGA<br>AAATATAGAC TTCCCAAGAG TGAAGACAGG GGACTTATAG GGGACCTTAC A |
| 11 | vSagC1-C2 | TCTGCTGCAA ACTTGGCATA GCTCAGGATG TCATCAGAAA CATCTTCTTT GAAAGTAGTA<br>AGCAAATGGG AGTTCAACCT GCTCATGTGT GAAAATGCTG GTGCGTCTGG GTTTACCCAG<br>CTGGTGACGC TGTGGTCTTC TACTGTCCAA CAGTGGTATC CACAGTACAT CAGCCTCAAT<br>TTCTTTTTTA GGACTGTGAT GCTCTGAGTT GCAGTTCAGG AACATCTCCC CTAGGAATCC<br>AAACTACTAC ACCCTGAAGT TTCATCTTTG TATGCAATTA GTTACTAAGT ACACTGAATC<br>AAATGCTACT TCTAGATTTG GGAGGCTCCC ACTGTCCACA AGGGTGACCA CTACGAGTCT<br>GAATTATAAG GGACTTATAG GGGACCTTAC A |
| 12 | vSagC1-C2 | TCTGCTGCAA ACTTGGCATA GCCCACAGCT CCAATATCTA GTCCAGTGTT GGTGGCTAGT<br>TAAGAAAAGA CAGTGCCTGT TGTGGTCTGT TAACATGATA ATCATTTCAG GGGGTTCTTA<br>AGGACTAAGA GATTCTTGGA ACCTAGTTTT ATCAATATAG CCAGTAGTAC TTGTTGAACA<br>GAGAAGCAGC CTGGTATATA GAAAGATCAT TGAATATGAA GTCAGAATGC CAAATGTGTC<br>CATTTGCTTT GAGACTCCCA CTCTCCATGT AAACAGGAAT ATCTTTGGTG TGCATGCATC<br>CTTTTTCAGT AGGAAACCCC CAGTTTTAAT ATCTGAGATT GTGTTATGAG GCTTAAGTCA<br>GAACAAAGAG AGCTGGATGA TGATAAAGGA TTGTACCAAT AACTGAAGGT AGTCTTATTC<br>CTTATTCTTC CCTGTCCTCC GTAACTACTT GCCTACTAGC CCATCTGGCC ATCCATCCAT<br>CCATGCAGTT CTTGTCCAGG GTCTACTGAG GTCCAGTGCT ATGCCAAGTT TGCAGCAGA |

**Table 2.3 DNA sequences of cloned PCR products amplified from human placental DNA with sequence specific primers for vSagC1, vSagC2 and the intervening region (vSagC1-C2).** PCR products were cloned directly into the pCR vector (Invitrogen), without prior agarose purification to separate fragments. The primer sequences are highlighted (Red: MMTV 3'LTR Sn210, Blue: MMTV 3'LTR Asn279, Pink: MMTV 3'LTR Sn661, Turquoise: MMTV 3'LTR Asn821, Brown: MMTV 3'LTR Asn689). Sequence 1 and 7 shared greatest homology to murine vSagC1 and vSagC2 respectively. Only the 5' sense primer (MMTV 3'LTR Sn210) was present in sequence 9. Sequence 12 had the same primer at each end, with a 3' antisense copy of MMTV 3'LTR Sn210.

### 2.3.3 Sequence analysis for homology to vSag

Homology of human placental DNA derived sequences to murine vSag was assessed using the 'microgenie' computer software (Beckman) of Queen and Korn (231). The human DNA sequences with the greatest homology to murine vSagC1 and vSagC2 were sequence 1 and sequence 7 respectively. No human sequences were identified with significant homology to the region between the two conserved regions (vSagC1-C2). None of the remaining ten PCR products had open reading frames, and were less homologous to murine vSag. Two of these sequences (sequence 9 and 12) were difficult to explain as the sequences of both PCR primers could not be identified. The 3' antisense primer sequence was not identified in the 183 bp product generated using primer pair 3 (sequence 9). The sequence of the vector's multiple cloning site was accurate when sequenced using both forward and reverse primers, indicating a satisfactory sequencing reaction. The PCR product may have been cleaved prior to ligation into the cloning vector, although there was no sequence similarity with the other products identified using this primer pair. The 539 bp product generated with primer pair 3 (sequence 12), although of similar size to the murine vSag product, was found to have the complementary sequence to the sense primer Sn201 at the end of the PCR product instead of the antisense primer Asn689. This PCR product will have resulted from the presence of a DNA sequence the reverse of (but the same sense as) the primer Sn210 downstream from the complementary sequence to which the Sn210 primer had annealed. Extension of the PCR product would produce a sequence complementary to Sn210 to which this primer could then anneal, allowing the polymerase chain reaction to proceed.

The 96 bp product of primer pair 1 (sequence 1) had a similarity index (matches/length) of 71% to the first constant region of murine vSag (vSagC1) with 47% homology when the PCR primer sequences were excluded from the analysis.

The 176 bp product of primer pair 2 (sequence 7) had a similarity index of 52% to the second conserved region of murine vSag (vSagC2) with 38% homology between PCR primers (figure 2.3). These sequences were not found on a DNA database search, and have subsequently been referred to as HRC1 (human-related C1) and HRC2 (human-related C2). Homology to vSag was not found in any of the other sequences and there was no homology between the different sequences amplified using the same primers.

**A**

```
vSagC1  TCTGCTGCAAACTTGGCATAGCTCTGCTTTG..CCTGGGGCTATTGGGG..GAAGTTGCGGTTC.GTGC..T.CGCAG
Seq 1   TCTGCTGCAAACTTGGCATAGCTCTGAGTTCAACCATTT.CTGCTGCGTTCGCAGCTG.GTCTCACAGGAATTCG.AG
        *************************  **   **     **  ** *   * ** ** *  **   *   * ** **

vSagC1  GGCTCTCACCCTTGACTCTTT
Seq 1   GGCTCTCACCCTTGACTCTTT
        *********************

Matches         70      Mismatches      18      Unmatched       11

Length          99      Matches/Length  71%     Matches/Length (excluding primers)  47%
```

**B**

```
vSagC2  CAAAATAGGAGACAGGTGGTGGCAACCAGGGACTTATAGGGGACCTTAC.ATCTACAGACCAACAGATGCCCCCT..T
Seq 7   CAAAATAGGAGACAGGTGGTGGAGGGTGGGATCTGCCTGCCGAGA..ACGAGGAAGGG..CATCGGTGACAGGTAACA
        **********************      **  **    *  **    ** *   *  *  ** * *   *

vSagC2  ACCATATACA..GGA.AGATATGACTTAAATTGGGATAGGTGGGTTACAGTCAA.TGGC.TATAAAGTGTTATATAGAT
Seq 7   ACCTGAGGCACTGGAGAGGGA.GAGC..ACGCGTGAT.GTTC.CTCAGAAACGAGTCGCATTTCC.GCATAGCGCC..T
        ***  *  **  *** **  * **    *   * *** * *   * * *  * * * ** * *   *  *        *

vSagC2  CCCTCCCTTTTCGTGAAAGACTCGCCAGAGCTA
Seq 7   CCTTCAGAC..CGTGAAAGACTCGCCAGAGCTA
        ** **      **********************

Matches         99      Mismatches      69      Unmatched       22

Length          190     Matches/Length  52.1%   Matches/Length (excluding primers)  38%
```

**Figure 2.3 DNA sequence homology between murine vSag and human PCR products.** 'Microgenie' computer software was used to analyse nucleotide sequence homology between murine vSag and the human PCR products amplified with vSagC1 and vSagC2 primers. **A.** murine vSagC1 and a human PCR product (sequence 1) amplified using the sequence specific primers MMTV 3'LTR Sn210 and MMTV 3'LTR Asn 279, and **B.** murine vSagC2 and a human PCR product (sequence 7) amplified using the sequence specific primers MMTV 3'LTR Sn661 and MMTV 3'LTR Asn821. * Matching nucleotides

### 2.3.4 Predicted amino acid sequences

The nucleotide sequences of HRC1 and HRC2 were translated using the 'Microgenie' software. The predicted amino acid sequences are shown together with the homology to the vSag encoded protein in figure 2.4. Both HRC1 and HRC2 sequences had potential open reading frames, with no translation termination codons in two of the reading frames of HRC1 and in one of open reading frame in the HRC2 sequence, multiple stop codons being present in the other reading frames. There were no methionine initiation codons in any of these sequences. The amino acid sequence homology ranged between 29% and 35% (39%-42% if including conservative amino acid substitutions). This homology was significantly influenced by the PCR primer sequences. When these were excluded, the homology between murine vSagC1 and HRC1 was 17%. The amino acid sequence homology between murine vSagC2 and HRC2 was 16% between the PCR primers although this increased to 29% if conservative amino acid substitutions were included.

A.HRC1 1ST open reading frame

```
                                                                --PCR Primer-
vSagC1 I L C C K L G I A L L C L G L L G E V A V R A R R A L T L D S
HRC1   L L Q T W H S S E F N H F C C V R S W S H R N S R A L T L D S

Matches 9      Mismatches     22     Unmatched 0    Conservative substit   3

Length  31     Matches/Length 29%    Similar/Length 39%
```

B. HRC1 2nd open reading frame

```
       --PCR primer-
vSagC1 C C K L G I A L L C L G L L G E V A V R A R * R A L T L D S F
HRC1   C C K L G I A L S S T I S A A F A A G L T G I R G L S P L T L

Matches 11     Mismatches     19     Unmatched 1    Conservative substit   2

Length  31     Matches/Length 35%    Similar/Length 42%
```

C. HRC2 open reading frame

```
       --PCR primer-
vSagC2 K I G D R W W Q P G T Y * R G * P Y I Y R P T D A P L * P Y T
HRC2   K I G D R W W R V G S A C R E R G R A * S V T G N N L R H W R

                                                          --PCR primer-
vSagC2 G R Y D L N F D R W V T V N G Y K V L Y R S L P F R E R L A R A
HRC2   G R A R V M F L R * N E S H * F R I A * P P S D * R E R L A R A

Matches 22     Mismatched     34     Unmatched      7     Conservative substit   6

Lrength 63     Matches/length 35%    Similar/length 41%
```

**Figure 2.4 Predicted amino acid sequence homology between murine vSag and human PCR products.** Homology between the conserved regions of the murine vSag protein and the predicted amino acid sequences of human PCR products HRC1 and HRC2, amplified using sequence specific primers for vSagC1 and vSagC2, was assessed using 'Microgenie' computer software.
**A, B:** potential open reading frames of sequence HRC1. **C:** potential open reading frame of HRC2. Regions representing PCR primer sequences are indicated. Matching amino acids highlighted in red. Conservative amino acid substitutions highlighted in pink.

## Isolation of vSag-related DNA sequences from a genomic library

In order to investigate the human genomic loci containing vSag-related sequences, a human placental DNA library was hybridised with the human sequences HRC1 and HRC2, and the λ Dash clones containing these sequences were isolated. Hybridisation of the genomic library with vSagC1 and vSagC2 DNA under reduced stringency conditions was undertaken to allow the identification of additional human DNA sequences with significant homology to murine vSag.

### 2.3.5 Screening of human placental DNA library with human vSag-related PCR products

Screening of the human placental DNA library with HRC1 and HRC2 probes under high stringency conditions resulted in the identification of 9 λ dash recombinants containing DNA which hybridised to HRC2. No recombinants hybridising to HRC1 were identified. An example of screening is shown in figure 2.5. This figure illustrates the general pattern observed throughout the screening of the library; i.e. a single first round hybridisation signal giving multiple positive signals in the second round stage, with the vast majority of the third round plaques hybridising to the probe.

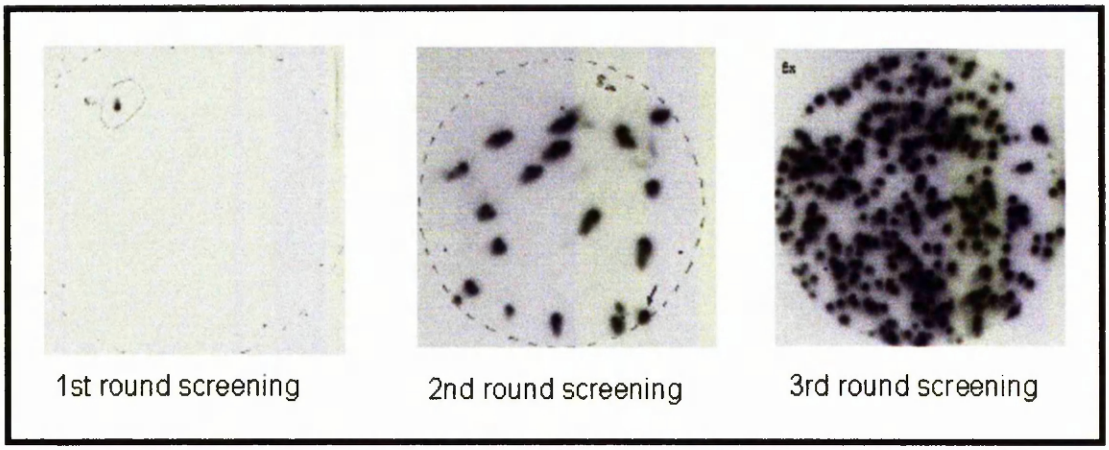
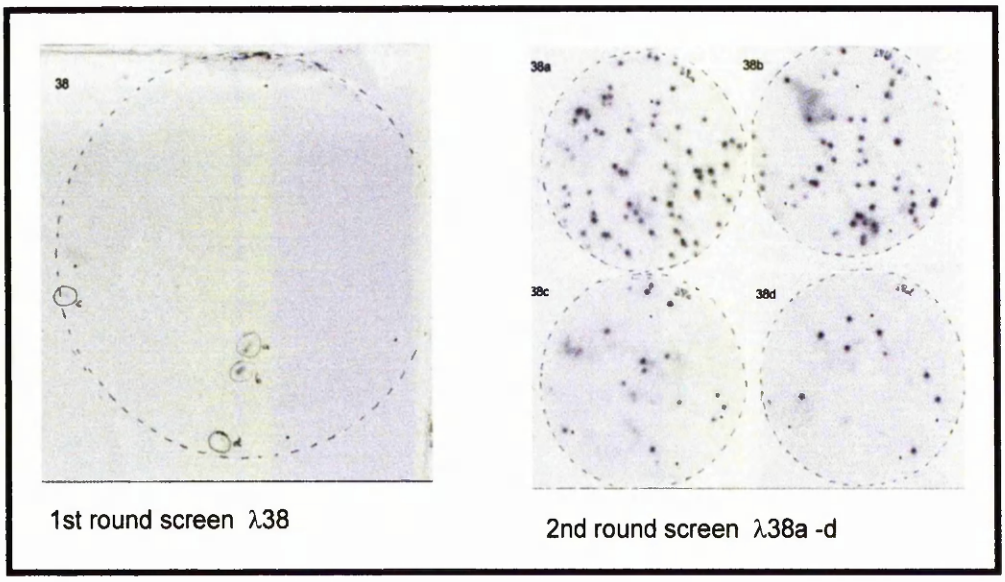



**Figure 2.5 Screening of a placental DNA library with a combination of HRC1 and HRC2 DNA.** Phage DNA was transferred to nitrocellulose as described in Materials and Methods. Filters were hybridised with $^{32}$P-labelled DNA probes at 37°C, in a mixture containing 50% formamide. Plaques which gave hybridisation signals were removed and the recombinant phage were purified by two further rounds of screening (as described in Materials and Methods). The figure shows the autoradiographs corresponding to the three rounds of screening.

### 2.3.6 Screening of human placental DNA library with murine vSag probes

Screening under reduced stringency (25% formamide) with probes amplified from the C1 and C2 regions of murine vSag resulted in the isolation and purification of 32 λ dash recombinants. Examples of positively hybridising DNA recombinants are shown in figure 2.6. Hybridisation with the murine vSag derived probes gave varying strengths of signal. Recombinant 38a is an example of one of the most strongly hybridising recombinants detects, while 39b is one of the weakest.



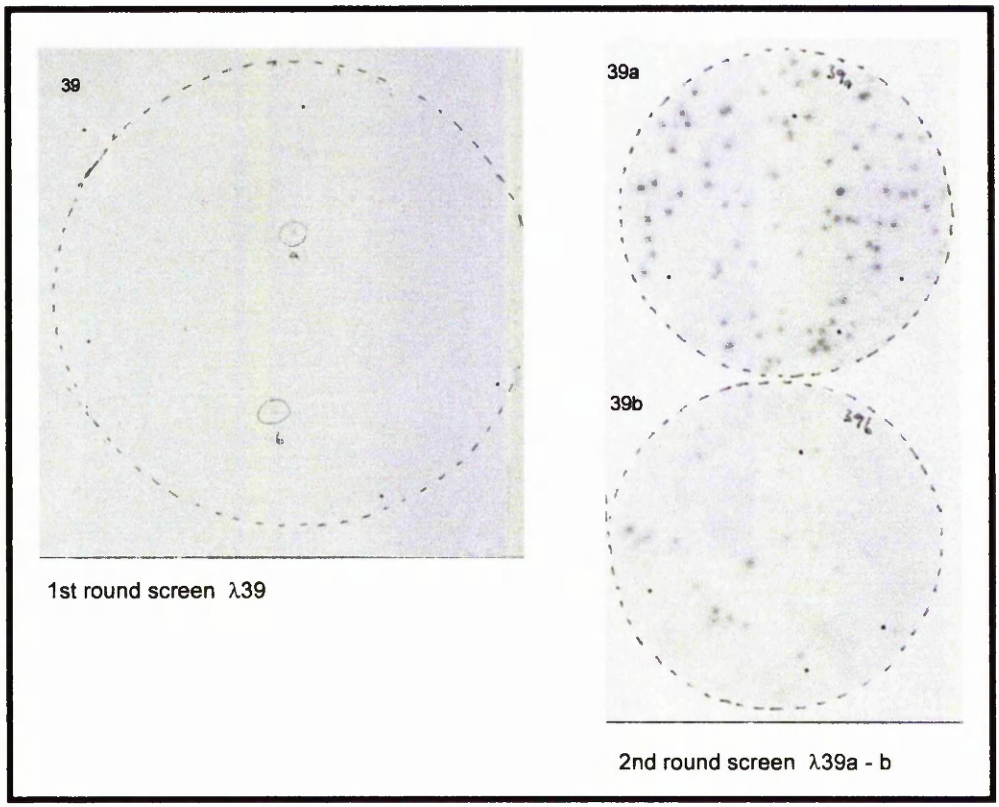




**Figure 2.6 Screening of a placental DNA library with a combination of vSagC1 and vSagC2 DNA under reduced stringency condition.** Phage DNA was transferred to nitrocellulose and filters hybridised with $^{32}$P-labelled DNA probes at 37°C, in a mixture containing 25% formamide. Plaques which gave hybridisation signals were removed and the recombinant phage were purified by two further rounds of screening. The figure shows the autoradiographs corresponding to two rounds of screening giving examples of strong (38a) and weak (39b) hybridisation signals.

### 2.3.7 Preparation of λ dash DNA

The large scale DNA preparation of 41 λ dash recombinants was undertaken; 9 hybridised to the combined HRC1 and HRC2 probes, 32 hybridised to the combined vSagC1 and vSagC2 probes. This gave yields of between 30-100 μg recombinant DNA. The size of the human DNA inserts was estimated after *Eco*RI digestion of the DNA, and ranged from approximately 12 kb to 20 kb.

## *2.4 Discussion*

Using primers derived from conserved regions of murine vSag, 12 PCR products were amplified from human DNA, and subsequently cloned and sequences. Although the human PCR products of the expected size were most promising for further investigation, products of other sizes were also investigated, as these could have resulted from insertions or deletions of otherwise homologous sequences. The most direct method of assessing the significance of the PCR results generated from human placental DNA, was to clone and sequence all of the PCR products. This was achieved by ligating the products of each PCR reaction into a cloning vector. By selecting a large number of clones for each ligation it was possible to isolate a representative of all of the different sized bands identified in the PCR reactions (figure 2.2).

Of the 12 cloned PCR products, two sequences (HRC1 and HRC2) were considered most promising for further investigation. They were selected because they had greatest homology to murine vSag (to the first and second conserved regions) and also had open reading frames, so could potentially encode protein. The sequence similarity indices, calculated by dividing the number of nucleotide matches by the total

length of the sequences being compared, were 71% and 52% for the sequences related to vSagC1 and vSagC2 respectively. However these similarity indices dropped to 47% and 38% respectively when only the region between the PCR primers was compared. (The significance of this degree of homology will be discussed in detail in chapter 4).

Indraccolo *et al* (149) used a similar approach to amplify human DNA sequences using murine vSag PCR primers, although they did not select primers to the conserved region of MMTV LTR, but rather used a random series of primers over the vSag gene. PCR products amplified with MMTV LTR specific primers were subsequently probed with a 1.4 kb *Bg*lII- *Hpa*II MMTV LTR restriction fragment, although this probe extended over the PCR primer binding sites, limiting the significance of any hybridisation. They identified three PCR products that hybridised to the MMTV probe. None of these PCR products were of a size predicted from the murine vSag sequence. A primer pair giving a predicted PCR product of 591 bp resulted in the amplification of two products of 275 bp and 477 bp. A different primer pair giving a predicted PCR product of 355 bp resulted in the amplification of a 410 bp product. Only the 275 bp sequence had a short region of 90% homology to murine vSag over twenty nucleotides. None of these sequences were homologous to the twelve human PCR products described in this chapter.

To investigate the genomic loci from which the two human vSag related sequences HRC1 and HRC2 had been amplified, these PCR products were used to probe a genomic library. The library was screened under high stringency conditions with a mixture of the two PCR products, resulting in the isolation of 9 λ dash recombinants. The characterisation of these clones is described in the following chapters.

The strategy described above depended on the selected PCR primers having very precise sequence similarity to the human related DNA sequences. However, distantly related sequences may have been missed, despite having significant homology, if there were mismatches at critical residues involved in annealing of extension of the PCR primers. Although increasing the number of PCR cycles or using degenerate PCR primers would be expected to increase the number of sequences isolated, many of these would be irrelevant, with no homology to vSag in the DNA sequence between the PCR primer annealing sites. An alternative approach was therefore employed where the human genomic library was screened directly, with murine DNA from the conserved regions of vSag. Probing the placental DNA library under reduced stringency conditions directly with a mixture of the murine probes derived from the conserved regions of vSag (vSagC1 and vSagC2) resulted in the isolation of 32 λ dash clones. It was not possible at this stage to determine whether each probe was hybridising to clones representing unique regions of DNA, or to overlapping stretches of the same sequence. The relative strength of the hybridisation signals varied between clones, but this could have resulted from overlapping clones containing a shorter part to the sequence complementary to the probe.

In summary, this chapter has described the isolation of 41 DNA clones from a human placental DNA library, detected by probes derived from murine vSag and human vSag-related sequences. The next chapter will deal with the preliminary characterisation of these DNA isolates.

# 3 CHAPTER THREE

## Preliminary characterisation of λ clones isolated from a placental DNA library using probes derived from murine vSag and human vSag-related sequences

### *3.1 Introduction*

This chapter describes the initial characterisation of the λ dash DNA recombinants isolated in chapter 2, by hybridisation analysis, restriction mapping and Southern blotting. Clones likely to be of greatest interest were selected for subcloning and sequencing (described in chapter 4) by identifying those which hybridised under higher stringency conditions and by looking for clones which hybridised with other vSag derived probes. Hybridisation analysis of specific DNA sequences was undertaken by "dotting" the cloned DNA of interest onto nitrocellulose filters and hybridising with radioactively labelled DNA probes. The stringency of the hybridisation reaction was controlled by increasing the concentration of formamide, thereby allowing the relative degree of homology of the clone to vSag to be assessed. Hybridisation of the clones to each probe (vSagC1, vSagC2, HRC1 and HRC2) was undertaken to assess whether the homology to vSag extended to adjacent DNA sequences. The λ dash clones were also hybridised to MMTV gag, pol and env probes, to determine whether the isolated DNA sequences were of retroviral origin and remained in a proviral configuration.

Restriction mapping and Southern blotting of digested DNA from λ clones identified those representing overlapping fragments of DNA and determined how many different vSag-related sequences had been isolated. Maps of the restriction enzyme sites within the DNA clones were also needed for subcloning and sequencing procedures (see chapter 4).

## *3.2 Materials and Methods*

### 3.2.1 Characterisation of λ clones by Dot-Blot Hybridisation.

**Dot blotting of λ recombinant DNA**

DNA from the 41 λ dash recombinants was immobilised onto nitrocellulose in the form of matrices of dots, using a "Biodot" manifold (Bio Rad). Using this apparatus, sheets of nitrocellulose are held in a fixed position and the DNA is absorbed in a matrix of dots. In addition to the 41 λ clones, DNA prepared from the following were included as controls: λ dash (negative control), MMTV 3'LTR, vSagC1, vSagC2, HRC1, HRC2.

DNA was prepared for dotting using the following protocol. Phage DNA isolated with murine probes vSagC1 and vSagC2 under reduced stringency (wells 1-32) was "dotted" at 100ng per well, while clones isolated under high stringency conditions with the human derived probe HRC2 (wells 33-41) were "dotted" at 25 ng per well. For control wells: 100ng λ dash (negative control), 5 ng MMTV plasmid (including gag, pol, env and LTR) and 0.4 ng of each PCR product vSagC1, vSagC2, HRC1, HRC2. For each dot required for a given recombinant, the DNA was first diluted with TE, to give a final volume of 41.5 μl. The sample was then denatured by adding 10 μl of 1M NaOH and heating at 37°C for 10 minutes. After cooling on ice, the DNA solution was made up to a final volume of 100 μl by the addition of 15 μl of a solution of 0.3 M Tris-HCl (pH 8), 0.67 M HCl and 1 mg/ml ethidium bromide, followed by 33.5 μl 20 x SSC. After

dotting, the manifold was dismantled and the filters submerged in a solution of 6 x SSC (90 mM $Na_3$Citrate and 0.9 M NaCl) for 2 minutes and dried on Whatman paper for 15 minutes. DNA was finally fixed on the nitrocellulose by baking at 80°C under vacuum for 2 hours. The filters were sealed into airtight bags and stored at 4°C until required.

**Preparation of radiolabelled probes by nick translation**

Hybridisation probes (vSagC1, vSagC2, HRC1, HRC2, MMTV *gag*, MMTV *gag-pol*, MMTV *env*, MMTV LTR) were radiolabelled by nick translation, as described in chapter 2. MMTV *gag* (1.1kb), *gag-pol* (4.0/4.3kb), *env* (2kb) and LTR (1.5kb) probes were derived from subgenomic *Pst*I fragments from MMTV (118).

**Hybridisation**

Hybribisation was carried with each probe at 10 kBq/ml. The stringency of the hybridisation reactions was controlled by varying the percentage of formamide in the hybridisation solution. Filters containing the dot-blotted DNA were prehybridised overnight at room temperature. All hybridisations were at 37°C for 48 hours as described in chapter 2, with between 25-50% formamide. Filters were washed at 37°C (see chapter 2) with formamide at the same concentration used in the hybridisation of the particular filter. The filters were finally dried and exposed against pre-flashed X-ray film.

### 3.2.2 Restriction mapping of λ clones.

The recombinant DNA was cloned into the *Bam*HI site of λ dash. The initial analysis of the clones consisted of digesting the DNA singly with four restriction enzyme (*Eco*RI, *Hind*III, *Bam*HI and *Xba*I). Digests were in 16 µl reaction mixtures, containing 500 ng

of DNA and 100 μg/ml BSA (bovine serum albumin), in a restriction buffer of the composition recommended by the enzyme manufacturer for 3 hours at 37°C. The restriction enzymes were used at 4 fold excess. (N.B. one unit of a restriction enzyme is defined as the amount required to digest 1μg λ DNA to completion in one hour at 37°C). Digested DNA samples were then electrophoresed on 1% agarose gels with *Hind*III digested λ DNA molecular weight markers, for approximately 450 volt-hours. The pattern of bands produced by restriction digest was observed by staining the gel in a solution of 0.5 μg/ml ethidium bromide and illuminating under UV light. The sizes of the restriction fragments produced by each digest were determined by comparison of the migration distance of the bands to the λ DNA markers.

**Double digests of selected λ clones**

Selected λ clones representing six families of overlapping clones were subsequently analysed by double digests with combinations of the four restriction enzymes described above. 500 ng of DNA was digested with an 8 fold excess of the restriction enzymes as shown in Table 3.2 and incubated for 37°C for 5 hours to ensure complete digestion. Digested DNA samples were electrophoresed on 1% agarose gels as described above. Further digests using an additional four restriction enzymes (*Pst*I, *Sac*I, *Sma*I and *Kpn*I) were required for the clones 26a and 44b, due to the absence of *Bam*HI and *Hind*III sites in these clones.

| locus | phage | Restriction enzyme combination | | |
|---|---|---|---|---|
| C1A | 26a+44b | *Eco*RI+*Xba*I | | |
| C1B | 38a | *Hind*III+*Bam*HI | *Hind*III+*Xba*I | *Bam*HI+*Xba*I |
| C1C | 40a | *Eco*RI+*Hind*III | *Eco*RI+*Xba*I | *Hind*III+*Xba*I |
| C2A | 5a | *Eco*RI+*Bam*HI | *Eco*RI+*Xba*I | *Bam*HI+*Xba*I |
| C2B | 37a | *Eco*RI+*Hind*III | *Eco*RI+*Xba*I | *Hind*III+*Xba*I |
| HRC2 | 2x | *Eco*RI+*Hind*III | *Eco*RI+*Bam*HI | *Hind*III+*Bam*HI |

**Table 3.2 Double digests of selected λ clones.**

### 3.2.3 Southern transfer of restriction enzyme digested λ clones.

After imaging the ethidium bromide stained gel on which the digested DNA clones had been electrophoresed, the restriction fragments were transferred to nylon membranes by Southern blotting, to allow hybridisation with the vSag related probes. Prior to transfer, the gels were placed in denaturing solution (0.5 M NaOH and 1.5 M NaCl) for 30 minutes and then neutralised in 1 M Tris-HCl (pH 5) containing 3 M NaCl. The nylon membrane soaked in 20 x SSC (0.3 M $Na_3$Citrate and 3 M NaCl) was laid over the gel on transfer apparatus (comprising a Whatman wick on a glass plate in a tray containing 20 x SSC), and allowed to transfer overnight.

Hybridisation probes, radiolabelled by nick translation, were prepared as described in chapter 2. Hybribisation was carried with each probe at 10 kBq/ml.

## *3.3 Results*

### 3.3.1 Hybridisation analysis of λ clones under varying stringency conditions

The DNA isolated from the human placental library was immobilised onto nitrocellulose filters. As the human placental library had been originally screened with combinations of probes (vSagC1 and vSagC2; HRC1 and HCR2), hybridisation with individual probes identified which vSag-derived probes hybridised most strongly, and assessed whether homology to vSag or other MMTV genes extended to adjacent DNA sequences. The position of the DNA from each λ clone on the nitrocellulose filter is shown in figure 3.1. These 'Hybridot' filters were hybridised with a series of probes: vSagC1, vSagC2, HRC1, HRC2, MMTV *gag*, MMTV *gag-pol*, MMTV *env* and MMTV LTR. The strength of the hybridisation was assessed by varying the percentage of formamide, allowing preliminary grouping of the clones.

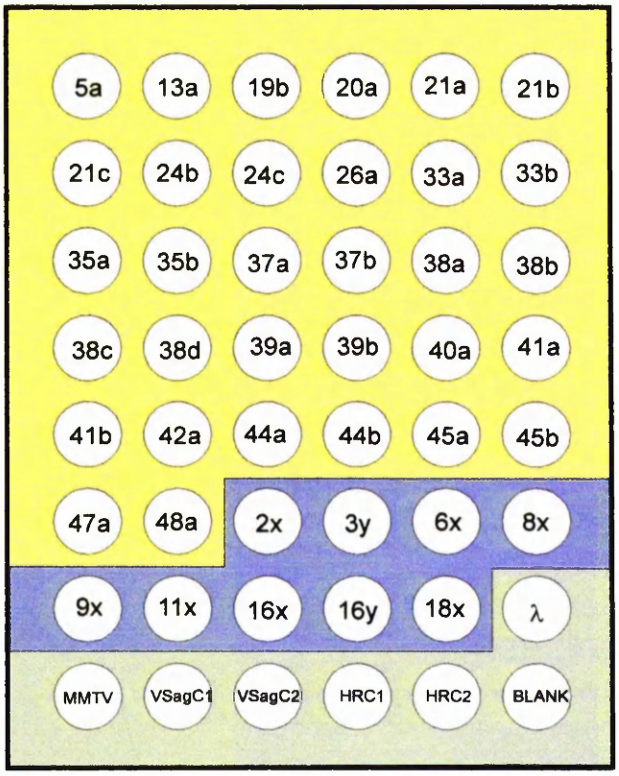
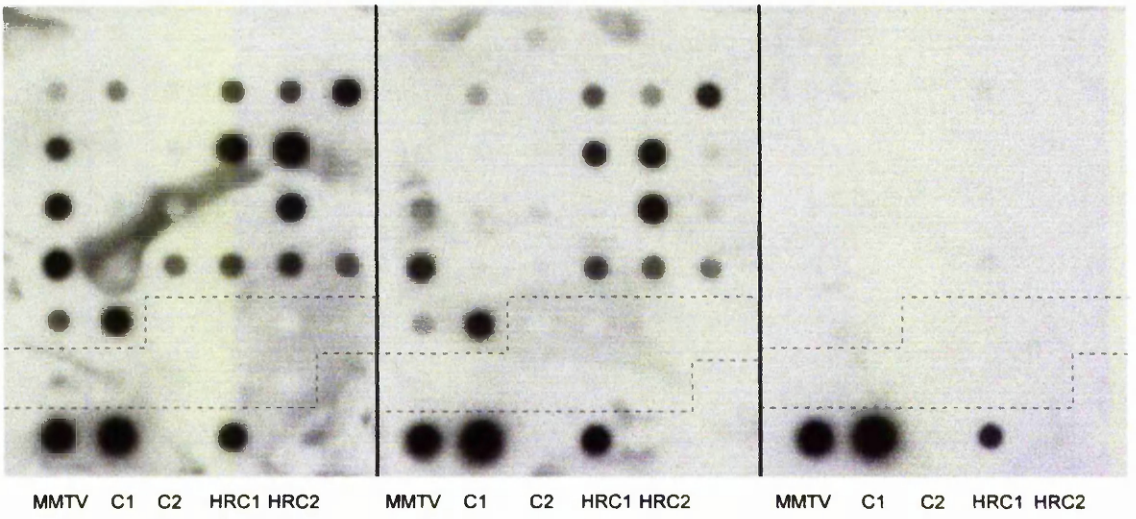



**Figure 3.1 Key for Hybridot filters showing the position of "dotted" DNA from the λ clones.** Yellow: 32 λ clones isolated with the combined vSagC1 and vSagC2 probes; Blue: 9 λ clones isolated with the combined HRC1 and HRC2 probes; Green: control DNA (λ, MMTV, vSagC1, vSagC2, HRC1, HRC2 and No DNA).

For the 32 λ clones isolated from the human placental DNA library with the combination of murine probes vSagC1 and vSagC2, all hybridised preferentially to one of the two conserved region probes (Figure 3.2 and Table 3.3). Under the original screening conditions (25% formamide) all 32 λ clones hybridised to either vSagC1 or vSagC2 probes (Table 3.3). The vSagC1 and vSagC2 probes hybridised strongly with the MMTV plasmid and to their respective positive controls (vSagC1 or vSagC2 PCR product). The vSagC1 probe also hybridised strongly to the HRC1 PCR product control even at increased stringency conditions (35% formamide), while the vSagC2 probe only hybridised weakly to the HRC2 under these conditions.

**A. vSagC1 probe**




**B. vSagC2 probe**

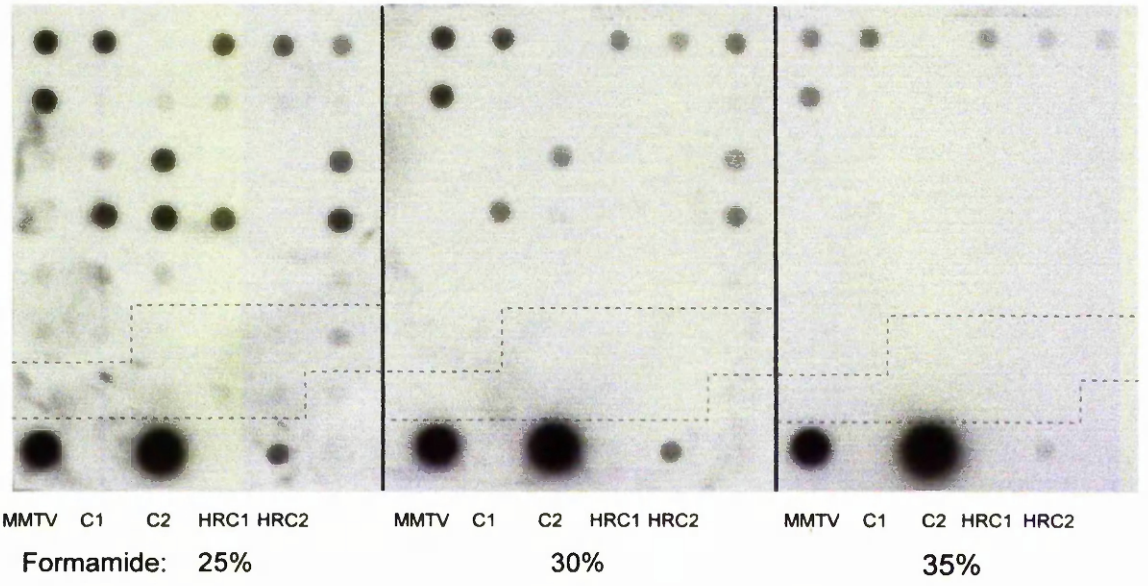



**Figure 3.2 Hybridisation of probes vSagC1 (A) and vSagC2 (B) to the λ DNA recombinants at varying stringency.** λ DNA (clones 1-32: 100 ng per dot, clones 33-41: 25 ng per dot) was immobilised on nitrocellulose filters, using a Biodot manifold. Nitrocellulose strips containing the immobilised DNA were hybridised with the $^{32}$P-α-dCTP-labelled probes (10 kBq/ml) as described in Materials and Methods. The prehybridisation, hybridistion and wash solutions contained 25%, 30% or 35% formamide. The positions of the λ recombinants are shown in figure 3.1.

Of the λ clones preferentially hybridising to the vSagC1 probe, 26a, 33a, 33b, 35a, 38c, 41b, 42a, 44a, 45b, 47a and 48a hybridised weakly to the vSagC2 probe under reduced stringency conditions (25% formamide), but did not hybridise when the formamide concentration was increased to 30%. Of the λ clones preferentially hybridising to the vSagC2 probe, 21c and 38d gave weak hybridisation signals with the vSagC1 probe (25% formamide).

Three of the λ clones isolated with the combined HRC1 and HRC2 probes (8x,16y, 18x) hybridised to the vSagC2 probe, giving a weak hybridisation signal with 25% formamide, but not under higher stringency conditions. None of these clones hybridised to the vSagC1 probe.

Hybridisation of the λ clones with the 1.5 kb *Pst*I fragment of MMTV containing the complete 3'LTR is shown in figure 3.3. All 32 of the clones originally isolated with the combined vSagC1 and vSagC2 probes hybridised to the MMTV 3'LTR when the formamide concentration was 25%, but the strength of hybridisation signal reduced as the fomamide concentration was increased, and at 35% formamide, only 8 clones hybridised weakly (5a, 13a, 20a, 21a, 21b, 21c, 26a, 44b).

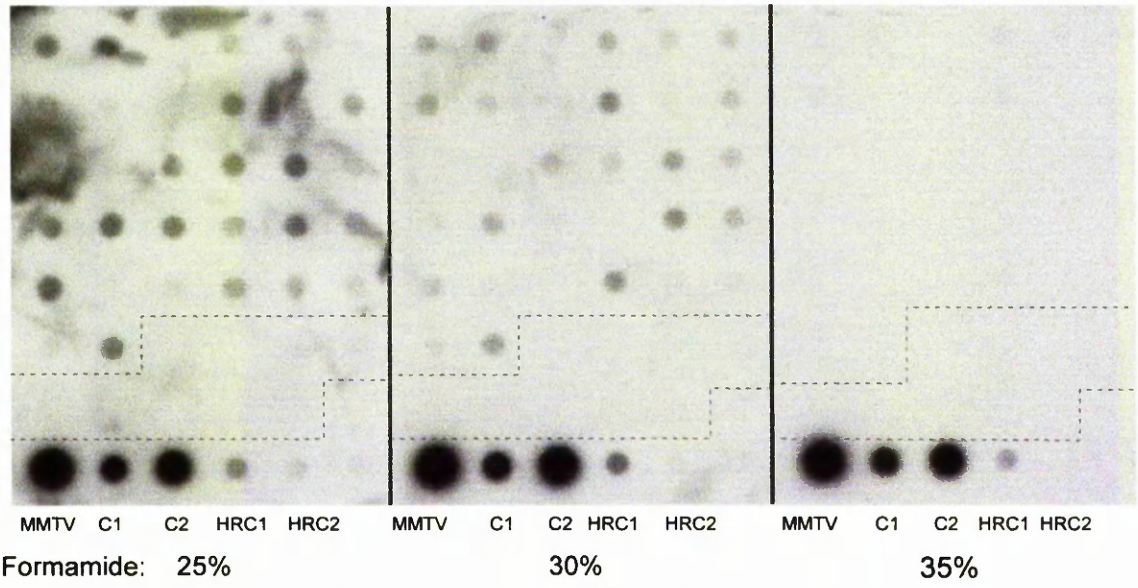



**Figure 3.3 Hybridisation of MMTV 3'LTR to the λ DNA recombinants at varying stringency.** λ DNA (clones 1-32: 100 ng per dot, clones 33-41: 25 ng per dot) was immobilised on nitrocellulose filters, using a Biodot manifold. Nitrocellulose strips containing the immobilised DNA were hybridised with the $^{32}$P-α-dCTP-labelled 1.5 kb *Pst*I fragment of MMTV (10 kBq/ml) as described in Materials and Methods. The prehybridisation, hybridistion and wash solutions contained 25%, 30% or 35% formamide. The positions of the λ recombinants are shown in figure 3.1.

The strength of the hybridisation signal with vSagC1, vSagC2 and MMTV 3'LTR for each of the 32 λ clones originally isolated with the combined vSagC1 and vSagC2 probes is shown in Table 3.3. The λ clones were provisionally grouped based on the strength of the hybridisation signals at varying stringency as shown in Table 3.4. Group 1 clones gave positive hybridising signals under the highest stringency conditions (35% formamide), group 2 clones hybridised strongly at 30% but not at 35% formamide, group 3 clones hybridised strongly at 25% but no more than weakly at 30% formamide, and group 4 clones hybridised only weakly at 25% formamide.

| λ DNA | vSagC1 | | | vSagC2 | | | MMTV 3'LTR | | |
|---|---|---|---|---|---|---|---|---|---|
| | Formamide | | | Formamide | | | Formamide | | |
| | 25% | 30% | 35% | 25% | 30% | 35% | 25% | 30% | 35% |
| 5a | - | - | - | +++ | ++ | + | ++ | + | +/- |
| 13a | - | - | - | +++ | ++ | + | ++ | + | +/- |
| 19b | - | +/- | - | - | - | - | +/- | - | - |
| 20a | - | - | - | +++ | ++ | + | + | + | +/- |
| 21a | - | - | - | +++ | ++ | + | + | + | +/- |
| 21b | - | - | - | +++ | ++ | + | + | + | +/- |
| 21c | +/- | - | - | +++ | ++ | + | ++ | + | +/- |
| 24b | + | + | - | - | - | - | + | + | - |
| 24c | - | - | - | + | +/- | - | +/- | - | - |
| 26a | +++ | ++ | +/- | +/- | - | - | ++ | + | +/- |
| 33a | +++ | + | - | +/- | - | - | + | +/- | - |
| 33b | +++ | +++ | - | +/- | - | - | + | + | - |
| 35a | +++ | - | - | +/- | - | - | + | - | - |
| 35b | - | - | - | + | - | - | +/- | - | - |
| 37a | - | - | - | +++ | ++ | - | ++ | + | - |
| 37b | +++ | +++ | - | - | - | - | ++ | + | - |
| 38a | +++ | +++ | - | - | - | - | ++ | + | - |
| 38b | - | +/- | - | ++ | + | - | + | + | - |
| 38c | +++ | + | - | +/- | - | - | ++ | +/- | - |
| 38d | +/- | +/- | - | +++ | ++ | - | ++ | + | - |
| 39a | - | +/- | - | +++ | +/- | - | ++ | +/- | - |
| 39b | - | - | - | ++ | - | - | + | - | - |
| 40a | +++ | +++ | - | - | - | - | ++ | + | - |
| 41a | - | +/- | - | +++ | + | - | + | + | - |
| 41b | +++ | +++ | - | +/- | - | - | ++ | + | - |
| 42a | + | - | - | +/- | - | - | +/- | - | - |
| 44a | ++ | - | - | +/- | - | - | +/- | - | - |
| 44b | +++ | ++ | +/- | - | - | - | + | + | +/- |
| 45a | ++ | + | - | - | - | - | + | +/- | - |
| 45b | ++ | + | - | +/- | - | - | + | +/- | - |
| 47a | ++ | + | - | +/- | - | - | + | +/- | - |
| 48a | +++ | +++ | - | +/- | - | - | ++ | + | - |
| λ | - | - | - | - | - | - | - | - | - |
| MMTV 3'LTR | +++ | +++ | +++ | +++ | +++ | +++ | +++ | +++ | +++ |
| vSagC1 | +++ | +++ | +++ | - | - | - | +++ | +++ | +++ |
| vSagC2 | - | - | - | +++ | +++ | +++ | +++ | +++ | +++ |
| HRC1 | +++ | +++ | ++ | - | - | - | ++ | ++ | + |
| HRC2 | - | - | - | ++ | + | +/- | + | +/- | - |

**Table 3.3 Hybridisation strength of vSagC1, vSagC2 and MMTV 3'LTR probes to the λ DNA recombinants.** The λ clones isolated from a human placental DNA library with the combined vSagC1 and vSagC2 probes which had been hybridised with $^{32}$P-α-dCTP-labelled vSagC1, vSagC2 and MMTV 3'LTR probes (figure 3.2 and 3.3), were graded for strength of hybridisation at varying stringency conditions (determined by % formamide in prehybridisation, hybridisation and wash solutions):
-negative, +/-very weak, +weak, ++strong, +++very strong.

| | vSagC1 hybridising clones | vSagC2 hybridising clones |
|---|---|---|
| Group 1 | 26a,44b | 5a, 13a, 20a, 21a, 21b, 21c |
| Group 2 | 33b, 37b, 38a, 40a, 41b, 48a | 37a, 38d |
| Group 3 | 33a, 35a, 38c, 44a, 45a, 45b, 47a | 38b, 39a, 41a, 39b |
| Group 4 | 19b, 24b, 42a | 24c, 35b |

**Table 3.4 Preliminary grouping of λ DNA recombinants based on hybridisation strength of vSagC1, vSagC2 and MMTV 3'LTR probes.** λ clones isolated from a human placental DNA library with the combined vSagC1 and vSagC2 probes which had been hybridised with $^{32}$P-α-dCTP-labelled vSagC1, vSagC2 and MMTV 3'LTR probes (figure 3.2 and 3.3), were grouped based on the strength of hybridisation at varying stringency conditions: Group 1= positive hybridisation at 35% formamide, Group 2= strong (++ or +++) hybridisation at 30% but negative at 35% formamide, Group 3= strong (++ or +++) hybridisation at 25% but negative or weak (-, +/- or +) at 30% formamide, Group 4= weak (+/- or +) at 25% formamide only.

As shown in Table 3.4, of the vSagC1 hybridising clones, two clones were categorised as group1, six as group 2, seven as group 3 and three as group 4. Of the vSagC2 hybridising clones, six clones were categorised as group 1, two as group 2, four as group 3 and two as group 4.

The same eight clones hybridising to vSagC1 or vSagC2 at 35% formamide concentration (Group 1) gave weak hybridisation signals with MMTV 3'LTR under these conditions. All of the Group 2 clones hybridised well to the MMTV 3'LTR at 30% but not 35% formamide. The same clones not hybridising to vSagC1or vSagC2 at 30% formamide (24c, 35a, 35b, 39b, 42a, 44a) together with clone19b, failed to hybridise with MMTV 3'LTR at 30% formamide concentration, but hybridisation signals were identified for the remaining 25 clones under this stringency.

### 3.3.2 Hybridisation of λ clones with HRC1 and HRC2 probes

The 'hybridot' filters were hybridised separately with the human PCR products HRC1 and HRC2. Of the 9 λ clones isolated with the combined HRC1 and HRC2 probes (2x, 3y, 6x, 8x, 9x, 11x, 16x, 16y, 18x), all hybridised strongly with the HRC2 probe, with no change in the strength of hybridisation signal as the formamide concentration was increase from 30% to 35% (figure 3.4). The HRC2 probe hybridised strongly with the HRC2 PCR product control, but did not hybridise to the MMTV plasmid or vSagC2 control even at lower stringency (30% formamide). The HRC1 probe did not hybridise with any of the clones, although it hybridised strongly to the HRC1 and vSagC1 PCR product controls and to the MMTV plasmid (data not shown).

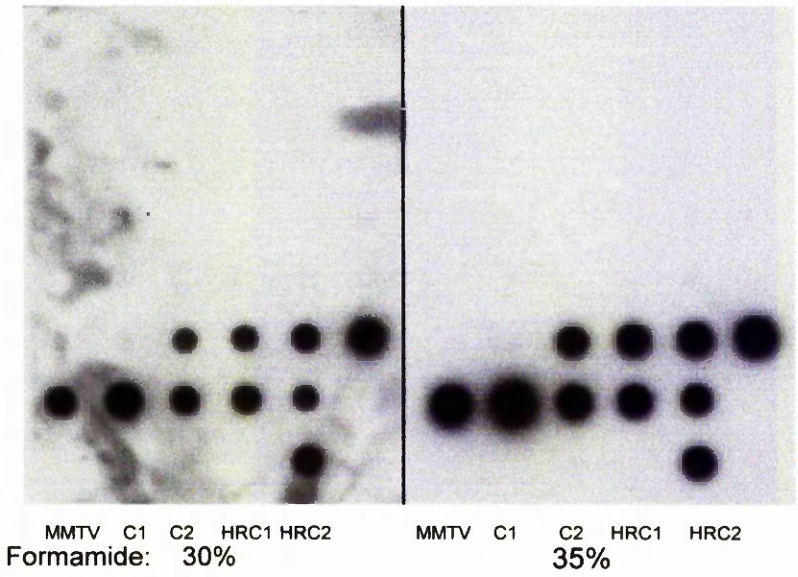



**Figure 3.4 Hybridisation of HRC2 probe to the λ DNA recombinants at varying stringency.** λ DNA (clones 1-32: 100 ng per dot, clones 33-41: 25 ng per dot) was immobilised on nitrocellulose filters, using a Biodot manifold. Nitrocellulose strips containing the immobilised DNA were hybridised with the $^{32}$P-α-dCTP-labelled HRC2 probe (10 kBq/ml) as described in Materials and Methods. The prehybridisation, hybridistion and wash solutions contained 30% or 35% formamide. The positions of the λ recombinants are shown in figure 3.1.

Of the λ clones originally isolated with vSagC1 and vSagC2 probes, 5a and 21c hybridised weakly to the HRC2 probe with a formamide concentration of 30% but not 35%. None of these clones hybridised to the HRC1 probe.

### 3.3.3 Hybridisation of λ clones to MMTV *gag*, *pol*, and *env* derived probes.

Probes derived from the MMTV *gag*, *gag-pol* and *env* regions (subgenomic MMTV *Pst*I digestion fragments 1.1 kb, 4.0/4.3 kb and 2.0 kb respectively (118)) were hybridised with the λ DNA clones at 25% and 30% formamide concentrations. None of the probes hybridised in 30% formamide (data not shown). At lower stringency conditions, the higher background made results difficult to interpret, but the MMTV *gag-pol* probe hybridised to clones 26a, 40a and 45a (Figure 3.5), each of these clones preferentially hybridising to the vSagC1 probe. The MMTV *gag* and *env* probes (MMTV *Pst*I 1.1 kb and 2.0 kb fragments) did not hybridise to any of the human λ recombinants.

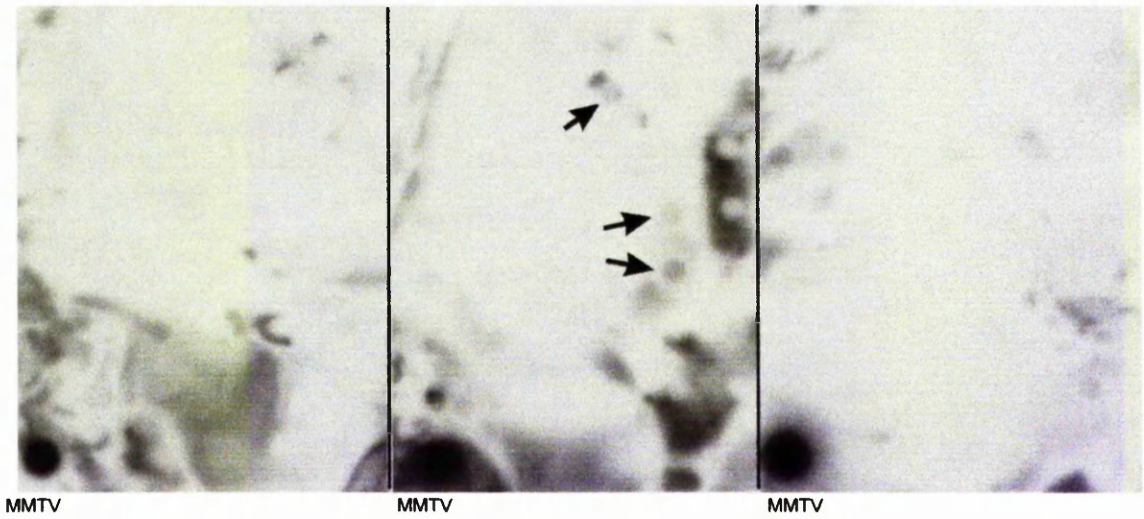



**Figure 3.5 Hybridisation of MMTV *gag*, *gag-pol* and *env* probes to the λ DNA recombinants at varying stringency.** λ DNA (clones 1-32: 100 ng per dot, clones 33-41: 25 ng per dot) was immobilised on nitrocellulose filters, using a Biodot manifold. Nitrocellulose strips containing the immobilised DNA were hybridised with $^{32}$P-α-dCTP-labelled *gag* (1.1 kb), *gag-pol* (4.0/4.3 kb) and *env* (2 kb) subgenomic *Pst*I fragments of MMTV at 10 kBq/ml as described in Materials and Methods. The prehybridisation, hybridistion and wash solutions contained 25%, 30% or 35% formamide. The positions of the λ recombinants are shown in figure 3.1.

### 3.3.4 Restriction analysis of λ clones.

Maps of the restriction endonuclease sites contained within each λ clone were constructed to allow identification of overlapping clones, and to find suitable restriction fragments for subsequent subcloning experiments. The 41 λ clones were each digested singly with four restriction enzymes (*Eco*RI, *Hind*III, *Bam*HI and *Xba*I). A simplified restriction enzyme map of λ Dash is shown in figure 3.6.

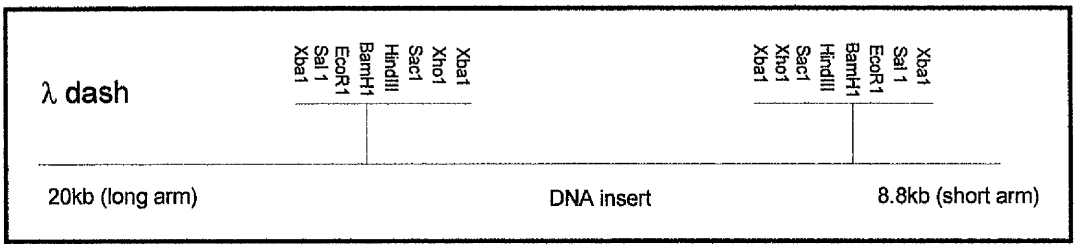



**Figure 3.6 Simplified restriction enzyme map of λ Dash.**

As restriction endonuclease sites for both *Eco*RI and *Xba*I were preserved at each end of the DNA insert, digestion with these enzymes gave 20 kb and 8.8 kb fragments representing the long and short arms of λ dash, the remaining fragments resulting from digestion of the DNA insert. Some *Bam*HI sites were preserved at the cloning site, but the λ Dash *Hind*III sites were deleted at the boundaries of the DNA insert.

Southern blots of the singly digested λ clones (1-32) were hybridised with vSagC1 or vSagC2 depending on which probe preferentially hybridised in the hybridot experiments (figure 3.7). Although 18 λ clones hybridised preferentially to the vSagC1 probe by hybridot analysis, one clone (19b) did not give a hybridisation signal with this probe with Southern blotting but hybridised to the vSagC2 probe instead. Although this clone had apparently hybridised weakly with the vSagC1 probe at 30% formamide, there was no hybridisation at 25% formamide in the Hybridot experiments with either vSagC1 or vSagC2 probes. This may have resulted from the DNA being inadvertently omitted when preparing the Hybridot filters. Therefore of the 32 λ clones isolated from the human placental using probes to the first and second conserved region of the murine vSag gene, 17 clones hybridised to vSagC1 and 15 clones hybridised to vSagC2.

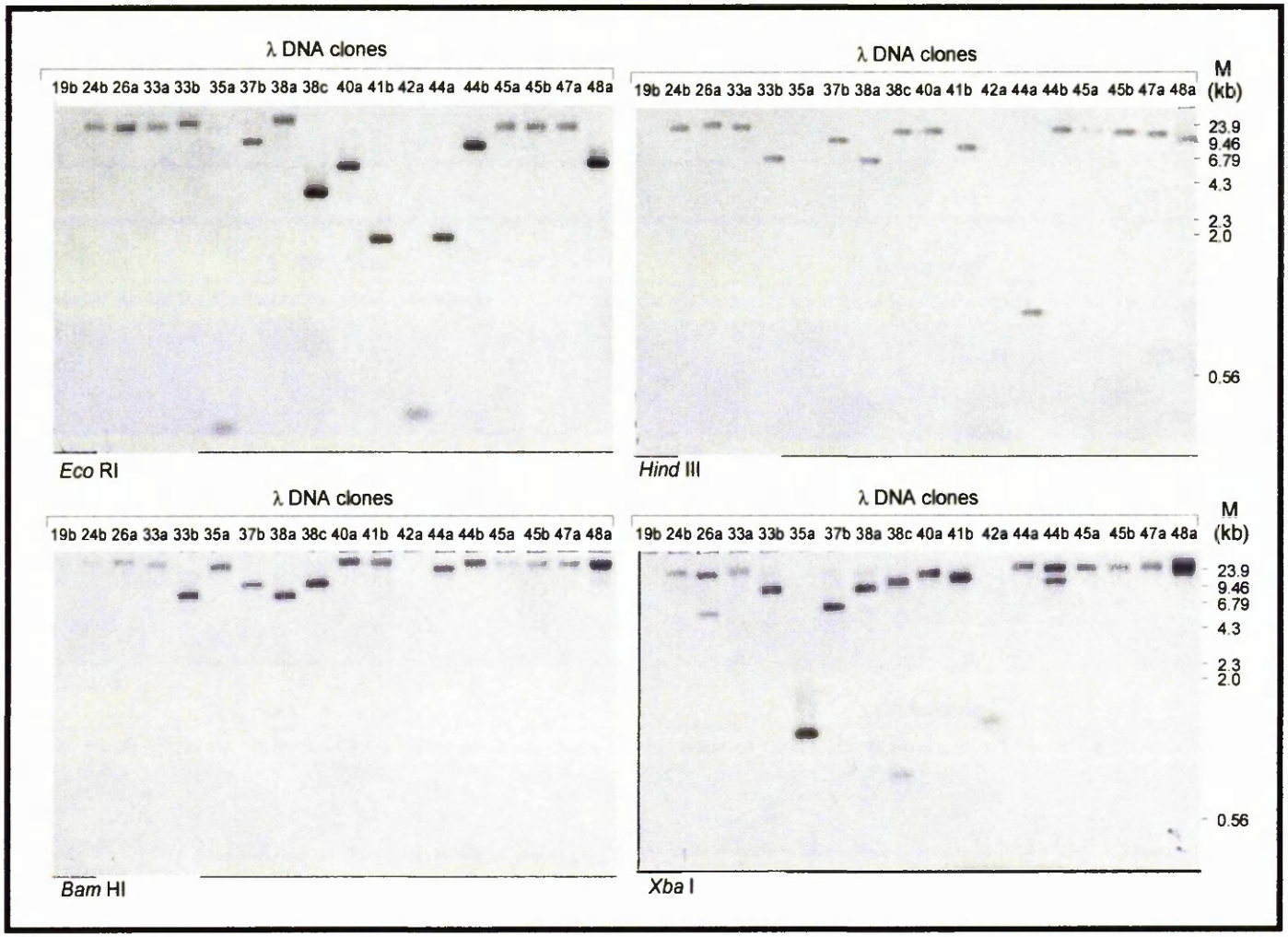



**Figure 3.7a Hybridisation of vSagC1 probe to Southern transfers of λ DNA clones.** DNA from λ clones isolated from a human placental DNA library with combination of vSagC1 and vSagC2 probes was digested singly with *Eco*RI, *Hind*III, *Bam*HI and *Xba*I, separated by gel electrophoresis on a 1% agarose gel, transfered to nylon membrane and hybridised with $^{32}$P-α-dCTP-labelled vSagC1 probe in a solution containing 25% formamide.

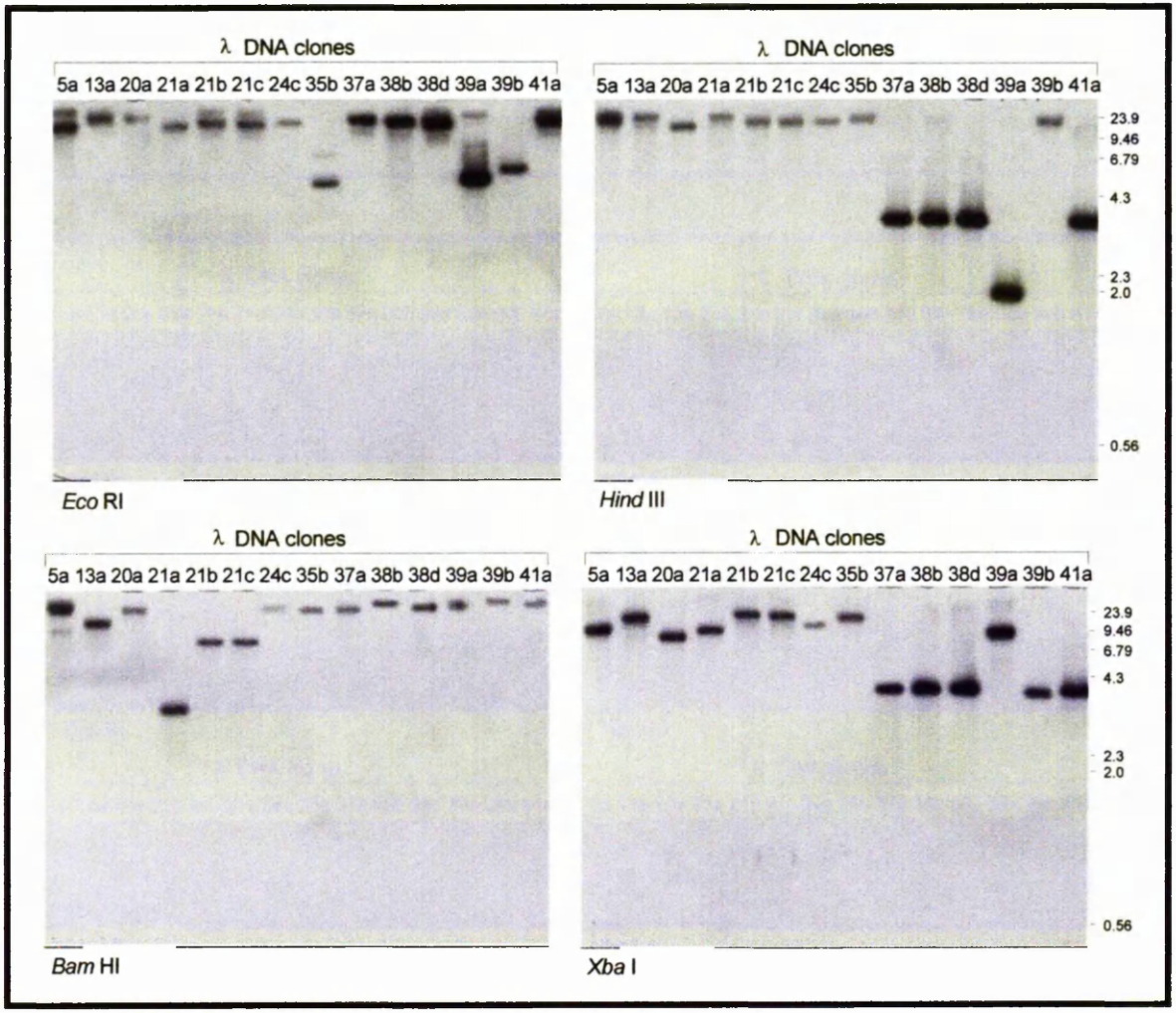
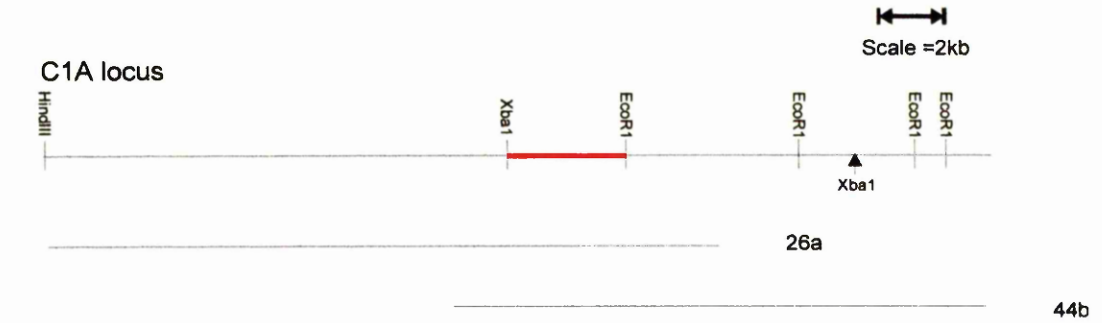



**Figure 3.7b Hybridisation of vSagC2 probe to Southern transfers of λ DNA clones.** DNA from λ clones isolated from a human placental DNA library with combination of vSagC1 and vSagC2 probes was digested singly with *Eco*RI, *Hind*III, *Bam*HI and *Xba*I, separated by gel electrophoresis on a 1% agarose gel, transfered to nylon membrane and hybridised with $^{32}$P-α-dCTP-labelled vSagC2 probe in a solution containing 25% formamide.

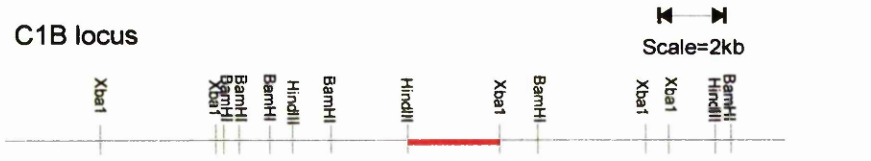
Restriction digests were analysed and restriction maps constructed. These showed five families of overlapping clones which accounted for 18 of the 32 clones isolated with the vSagC1 and vSagC2 probes (figures 3.8a and 3.8b). The remaining 14 λ clones had hybridised relatively weakly in the dot-blot hybridisation experiments (section 3.3.1), and comprised all the group 4 clones and eight of the eleven group 3 hybridising clones.

**vSagC1 hybridising λ clones**

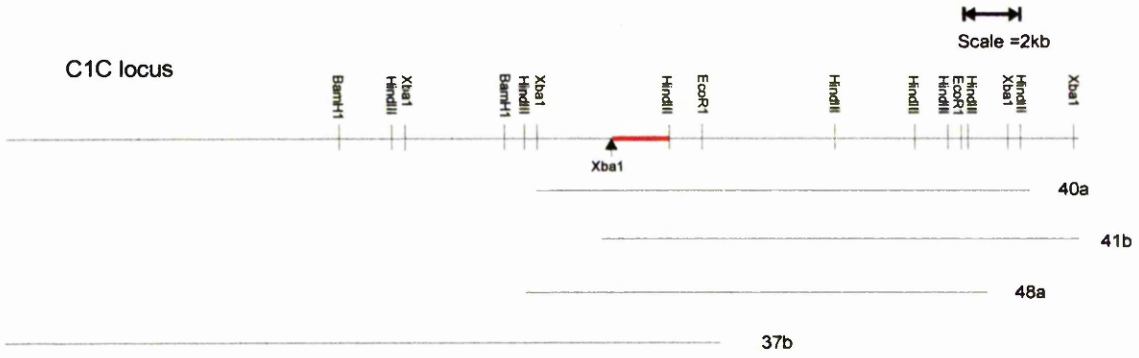
Of the seventeen λ clones isolated using the murine probe vSagC1, eight originated from three genetic loci. Restriction maps of these three loci are shown in figure 3.8a. The region to which the vSagC1 probe hybridises is highlighted. Two overlapping clones originated from locus C1A (26a, 44b), two overlapping clones originated from locus C1B (33b, 38a) and four originated form locus C1C (37b, 40a, 41b, 48a). The remaining nine λ clones hybridising to the vSagC1 probe appeared to be from distinct loci, with no restriction fragment lengths of similar size.




Group A vSagC1 hybribising clones




Group B vSagC1 hybribising clones

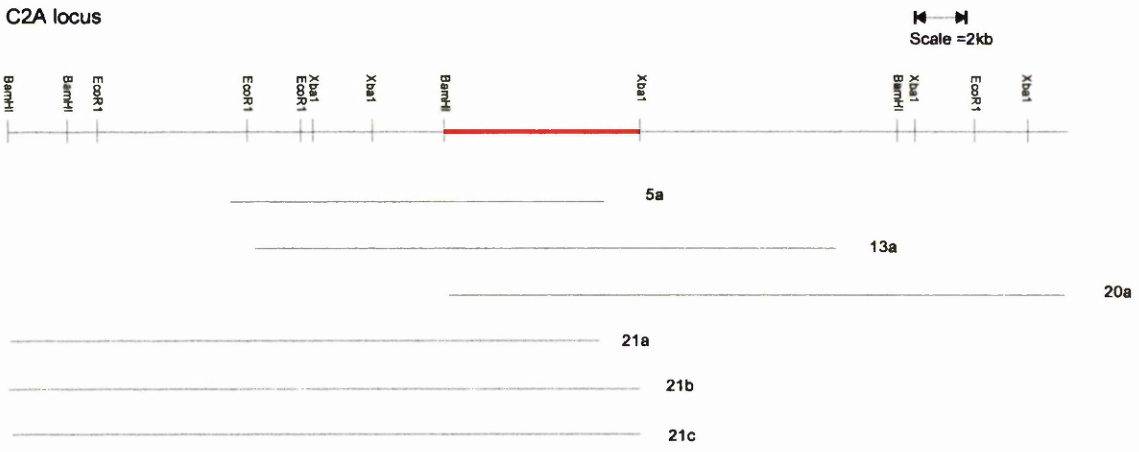



Group C vSagC1 hybribising clones

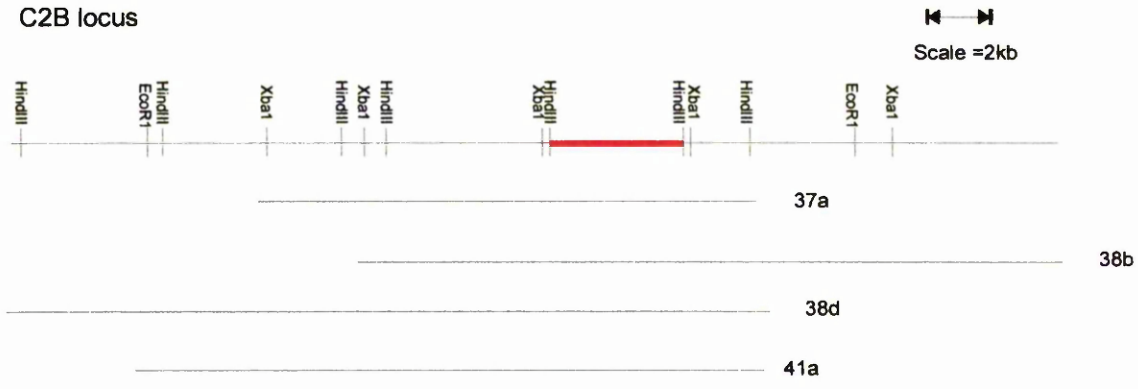
**Figure 3.8a Restriction maps of vSagC1 hybridising λ DNA clones.** The restriction sites within the λ DNA clones were positioned by analysis of single digests with *Eco*RI, *Hind*III, *Bam*HI and *Xba*I. The names of the λ DNA clones are indicated at the right hand side of the appropriate map. The region of the clones to which the vSagC1 probe hybridised is highlighted in red.

**vSagC2 hybridising λ clones**

Of the fifteen λ clones isolated with the murine probe vSagC2, ten originated from two genetic loci. Six overlapping clones were derived from the C2A locus (5a, 13a, 20a, 21a, 21b, 21c) and four overlapping clones from the C2B locus. Restriction maps of these two loci are shown in figure 3.8b. The remaining five λ clones did not share restriction fragments of similar length.




Group A vSagC2 hybridising clones




Group B vSagC2 hybridising clones

**Figure 3.8b Restriction maps of vSagC2 hybridising λ DNA clones.** The restriction sites within the λ DNA clones were positioned by analysis of single digests with *Eco*RI, *Hind*III, *Bam*HI and *Xba*I. The names of the λ DNA clones are indicated at the right hand side of the appropriate map. The region of the clones to which the vSagC2 probe hybridised is highlighted in red.

**HRC2 hybridising λ clones**

The nine λ clones isolated using the HRC2 probe (amplified from human placental DNA with sequence specific PCR primers to the second conserved region of murine vSag) each represented an overlapping clone from the same genomic locus. The restriction map of this locus is shown in figure 3.9.

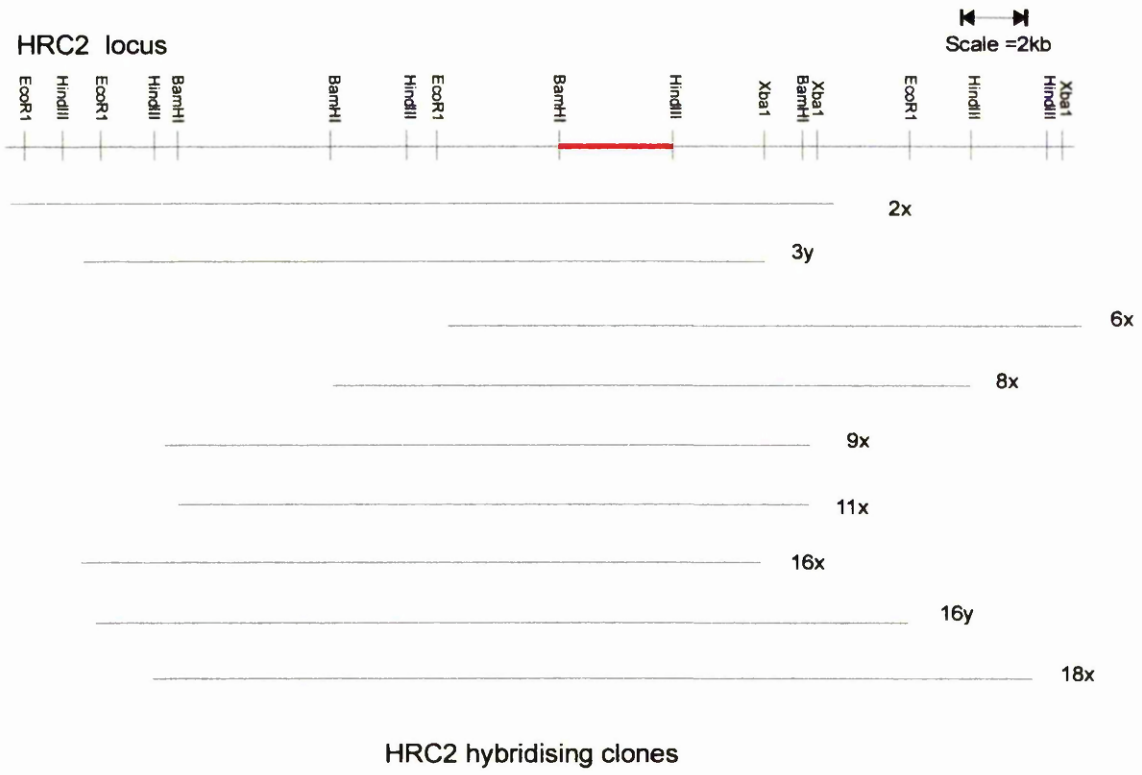



**Figure 3.9 Restriction maps of HRC2 hybridising λ DNA clones.** The restriction sites within the λ DNA clones were positioned by analysis of single digests with *Eco*RI, *Hind*III, *Bam*HI and *Xba*I. The names of the λ DNA clones are indicated at the right hand side of the appropriate map. The region of the clones to which the HRC2 probe hybridised is highlighted in red.

### 3.3.5 Restriction analysis of λ clones by double digests

To confirm the restriction maps for the six loci identified above, and to identify restriction fragments suitable for subsequent subcloning experiments, representative clones were selected and double digests undertaken using pairs of restriction enzymes (Table 3.2). For the clones 38a, 40a, 5a, 37a and 2x, the restriction fragments resulting from these double digests were as predicted from the single digests. Hybridisation of the vSagC1 (38a and 40a), vSagC2 (5a and 37a) and HRC2 (2x) probes to DNA digested with combinations of two restriction enzymes are shown in figure 3.10, giving the restriction fragment sizes in Table 3.5.

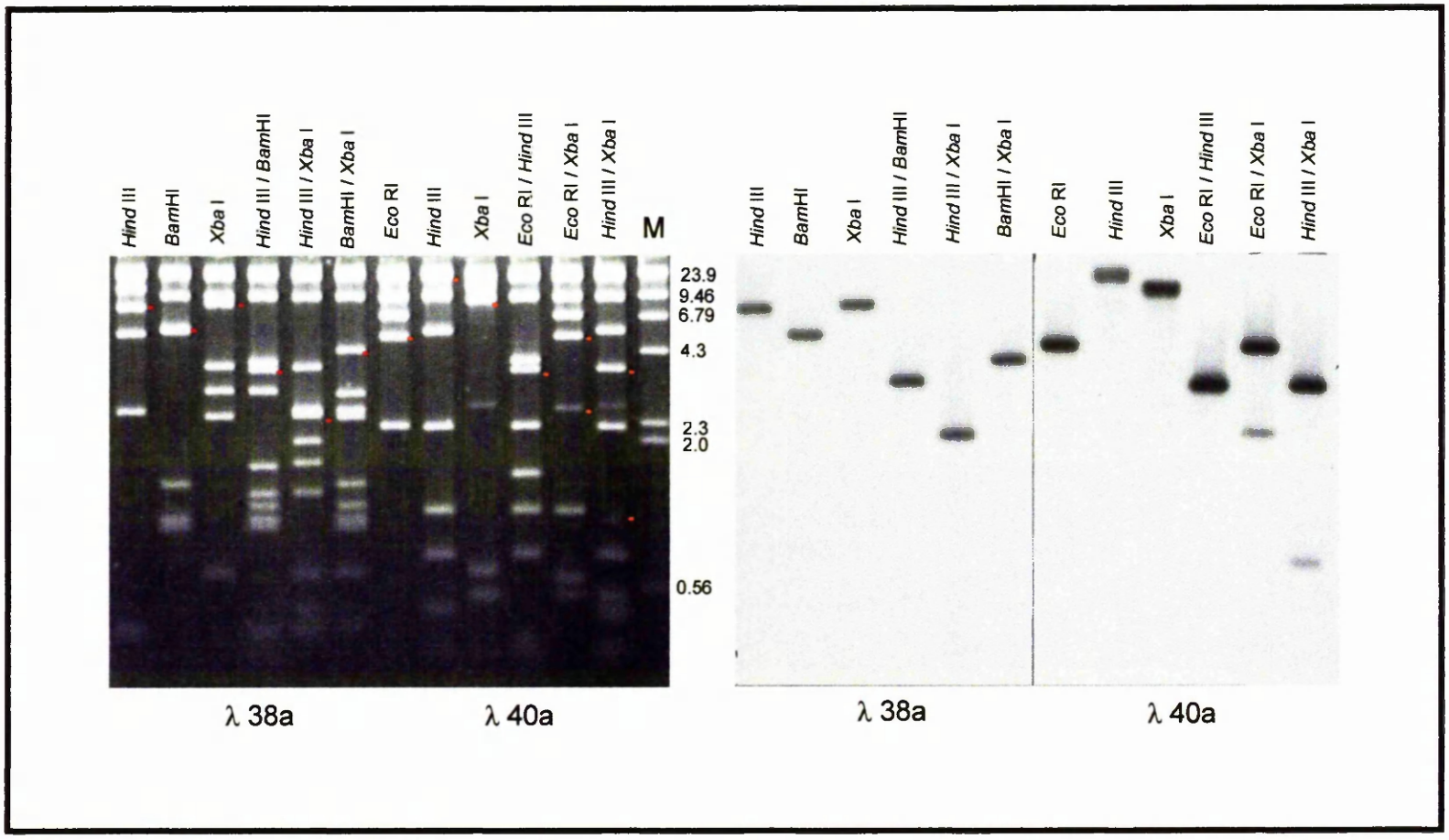



**Figure 3.10a Hybridisation of vSagC1 probe to Southern transfers of λ DNA clones 38a and 40a.** DNA from λ clones 38a and 40a, isolated from a human placental DNA library with combination of vSagC1 and vSagC2 probes was digested singly and doubly with *Hind*III, *Bam*HI and *Xba*I, separated by gel electrophoresis on a 1% agarose gel, transfered to nylon membrane and hybridised with $^{32}$P-α-dCTP-labelled vSagC1 probe in a solution containing 25% formamide. Red marks indicate fragments of λ38a which hybridised to the vSagC1 probe, orange marks indicate fragments of λ40a which hybridised to the vSagC1 probe.
M= λ *Hind*III molecular weight marker (kb).

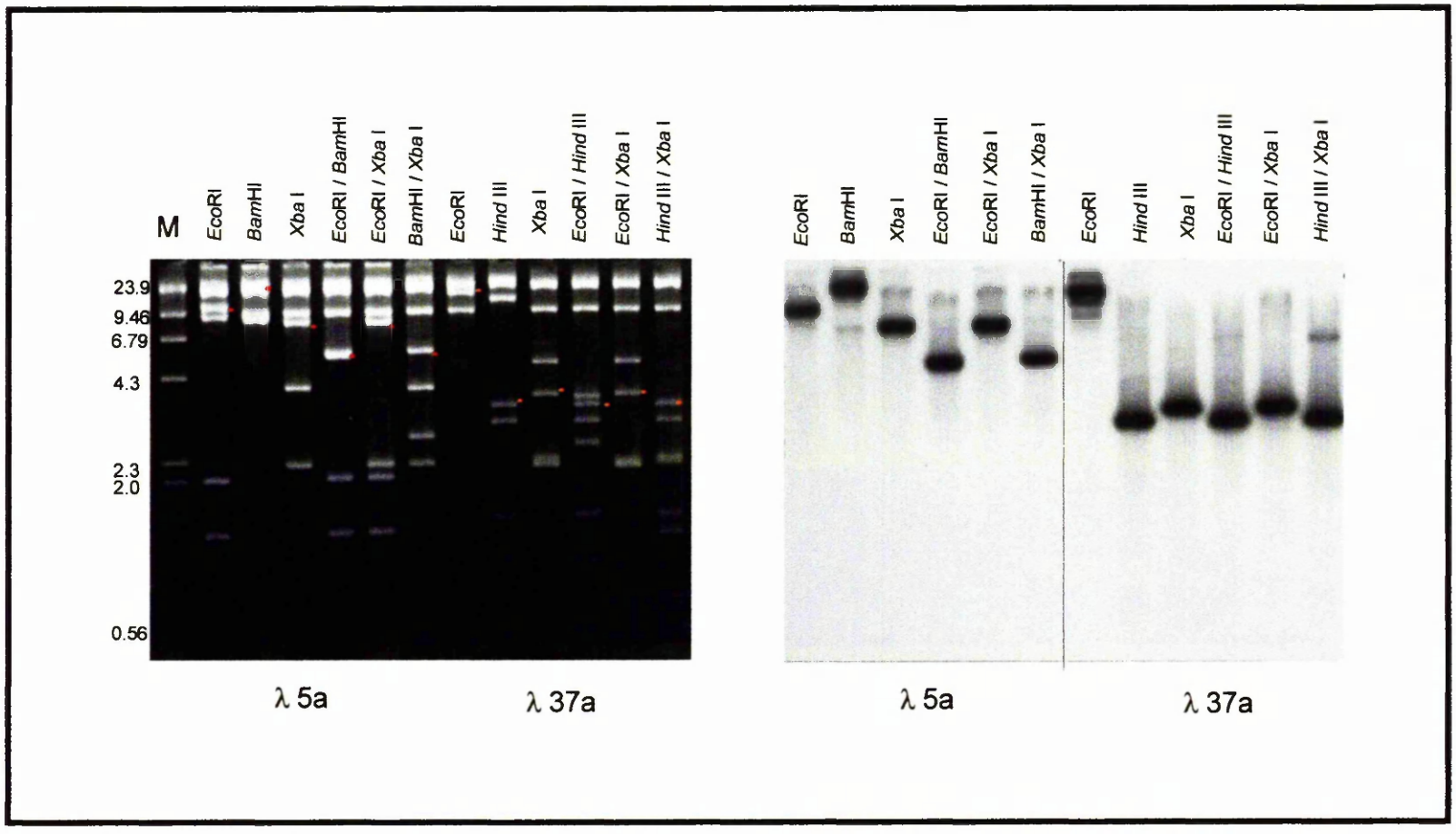



**Figure 3.10b Hybridisation of vSagC2 probe to Southern transfers of λ DNA clones 5a and 37a.** DNA from λ clones 5a and 37a, isolated from a human placental DNA library with combination of vSagC1 and vSagC2 probes was digested singly and doubly with *Eco*RI, *Bam*HI and *Xba*I, separated by gel electrophoresis on a 1% agarose gel, transfered to nylon membrane and hybridised with $^{32}$P-α-dCTP-labelled vSagC2 probe in a solution containing 25% formamide. Red marks indicate fragments of λ5a which hybridised to the vSagC2 probe, orange marks indicate fragments of λ37a which hybridised to the vSagC2 probe.
M= λ *Hind*III molecular weight marker (kb).

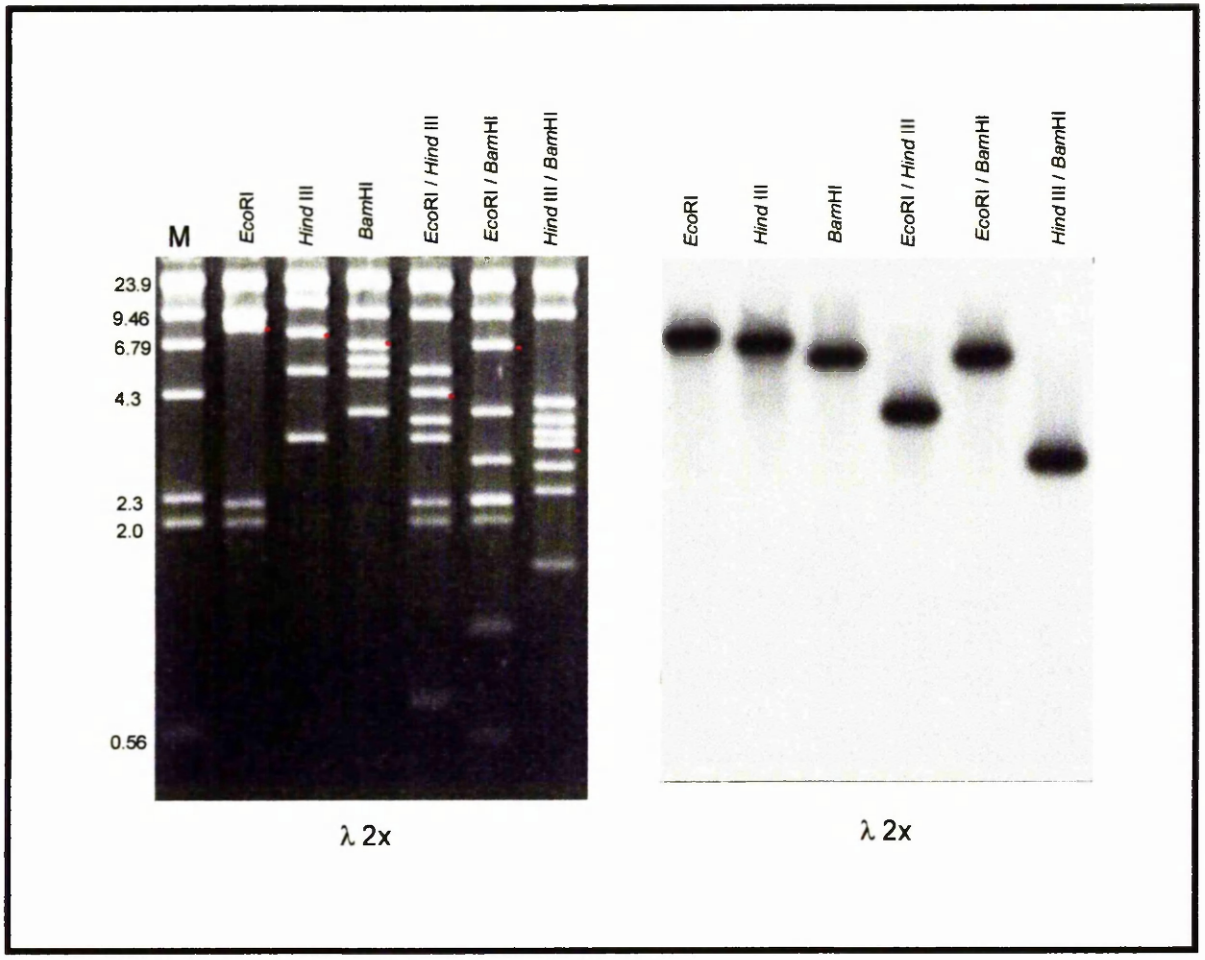



**Figure 3.10c Hybridisation of HRC2 probe to Southern transfers of λ DNA clone 2x.** DNA from λ clone 2x, isolated from a human placental DNA library with combination of HRC1 and HRC2 probes was digested singly and doubly with *Eco*RI, *Hind*III and *Bam*HI, separated by gel electrophoresis on a 1% agarose gel, transfered to nylon membrane and hybridised with $^{32}$P-α-dCTP-labelled HRC2 probe in a solution containing 50% formamide. Red marks indicate fragments which hybridised to the HRC2 probe.
M= λ *Hind*III molecular weight marker (kb).

| loci | probe | phage | Restriction enzymes | Fragment size |
|---|---|---|---|---|
| C1B | vSagC1 | 38a | *Hind*III + *Xba*I | 2.3 kb |
| C1C | vSagC1 | 40a (41b) | *Hind*III + *Xba*I | 1 kb |
| C2A | vSagC2 | 5a (13a) | *Bam*HI + *Xba*I | 5 kb |
| C2B | vSagC2 | 37a | *Hind*III | 3.2 kb |
| HRC2 | HRC2 | 2x | *Bam*HI + *Hind*III | 3.2 kb |

**Table 3.5 Restriction fragments of λ clones hybridising to vSagC1, vSagC2 or HRC2 probes.** The λ clones were digested with a combination of four restriction enzymes (*Eco*RI, *Hind*III, *Bam*HI and *Xba*I). The combination of restriction enzymes giving restriction fragments of a size suitable for subcloning into a plasmid vector are shown. (Note clone 41b used in subsequent cloning instead of 40a as insufficient DNA for detailed subcloning and sequencing experiments. The overlapping clone 13a was used instead of 5a for subcloning, because 5a did not encompassed the entire 7 kb *Bam*HI / *Xba*I fragment hybridising to the vSagC2 probe.)

To confirm that the clones 26a and 44b overlapped, digests with additional restriction enzymes were undertaken. This was necessary as these clones did not have any *Bam*HI sites and there was only one *Hind*III site at the end of the clone 26a, which was not present in the 44b clone. In addition, although *Eco*RI / *Xba*I double digests were consistent with the restriction map shown in figure 3.7, these appeared to be partial digests, resulting in two restriction fragments of varying length which hybridised to the vSagC1 probe. These two clones were singly digested with the restriction enzymes *Pst*I, *Sac*I, *Sma*I and *Kpn*I clearly showing that they represented overlapping clones (figure 3.11). Southern blots gave restriction fragments of 0.6 kb (*Pst*I) and 1.2 kb (*Sac*I) which were subsequently subcloned into the Bluescript vector as described in the following chapter.

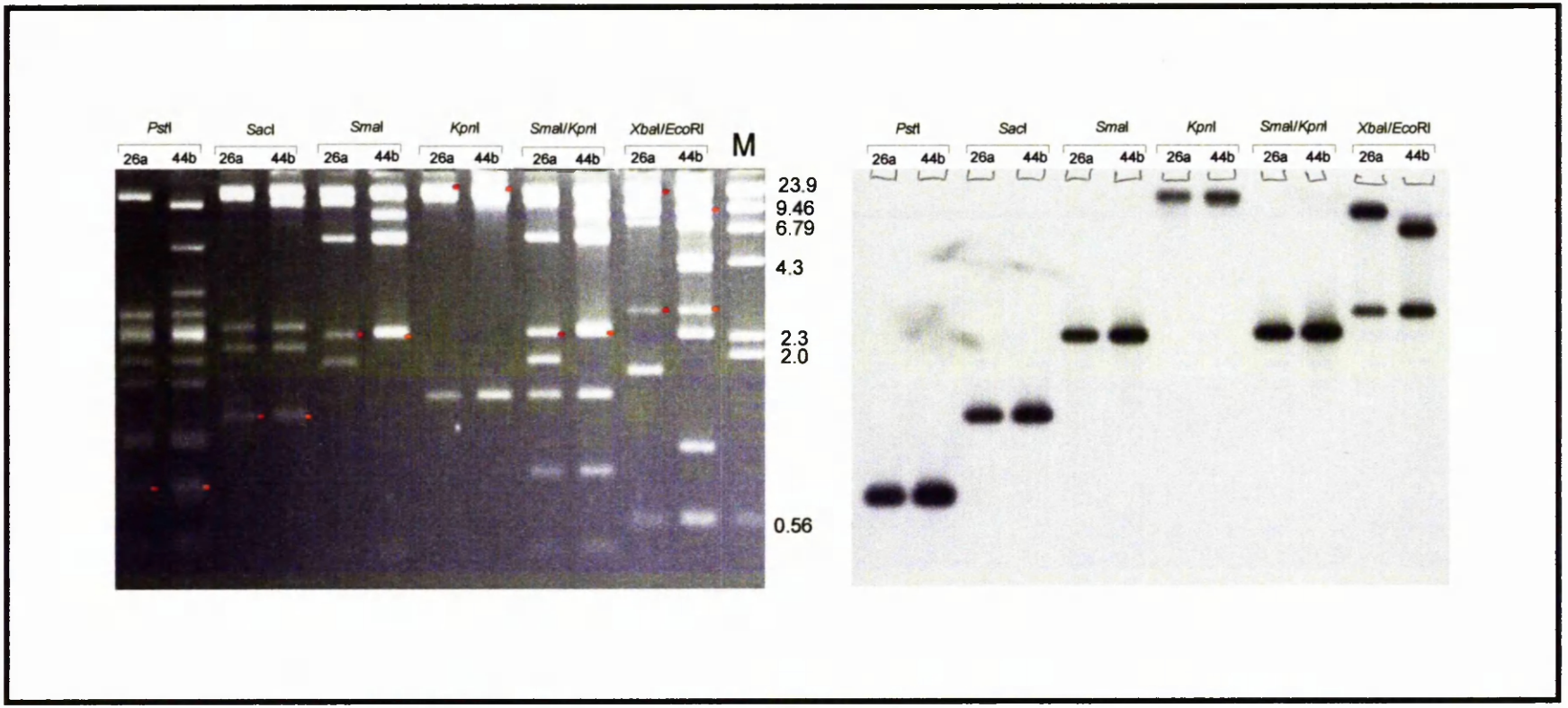



**Figure 3.11 Hybridisation of vSagC1 probe to Southern transfers of λ DNA clones 26a and 44b.** DNA from λ clones 26a and 44b, isoiated from a human placental DNA library with a combination of HRC1 and HRC2 probes was digested singly with *Pst*I, *Sac*I, *Sma*I and *Kpn*I, and doubly with *Sma*I / *Kpn*I and *Xba*I / *Eco*RI, separated by gel electrophoresis on a 1% agarose gel, transferred to a nylon membrane and hybridised with $^{32}$P-α-dCTP-labelled HRC2 probe in a solution containing 25% formamide. Red marks indicate fragments of λ26a which hybridised to the vSagC1 probe, orange marks indicate fragments of λ44b which hybridised to the vSagC1 probe.
M= λ *Hind*III molecular weight marker (kb).

## *3.4 Discussion*

The 41 λ clones isolated in the previous chapter were initially characterised by 'dot-blot' hybridisation experiments, where the purified cloned DNA was immobilised on nitrocellulose filters. These clones had originally been isolated from the human placental library using probes in combination (HRC1 and HRC2, or vSagC1 and SagC2), and by hybridising with individual probes it was possible to identify the region of vSag to which each clone preferentially hybridised. As 32 of these clones were isolated under reduced stringency conditions (25% formamide concentration) using the murine probes vSagC1 and vSagC2, it was of interest to see which clones hybridised at increased stringency. This allowed the clones to be grouped according to their strength of hybridisation under varying stringency, as determined by the concentration of formamide in the hybridisation reaction. This gave an indication of which clones shared greatest homology to the murine vSag constant regions vSagC1 and vSagC2. It also made possible hybridisation experiments looking for clones sharing homology to other regions of MMTV genes beyond the region used to generate the vSagC1 or vSagC2 probes.

Of the 32 clones isolated from the placental library with the combined vSagC1 and vSagC2 probes, eighteen hybridised preferentially to vSagC1 and fourteen to vSagC2. Of the clones hybridising to vSagC1, two hybridised at 35% formamide (group 1), six hybridised strongly at 30% formamide, but not at 35% (group 2), seven hybridised strongly at 25%, but no more than weakly at 30% (group 3) and three hybridised only weakly at 25% formamide (group 4). Using the same scoring to group the fourteen clones preferentially hybridising to the vSagC2 probe, there were six group 1 clones, two group 2 clones, four group 3 clones and two group 4 clones.

None of the clones hybridised strongly to both vSagC1 and vSagC2 probes, although weak 'cross-hybridisations' were found. Eleven of the λ clones preferentially hybridising with the vSagC1 probe also hybridised weakly to vSagC2. Two of the λ clones preferentially hybridising with the vSagC2 probe also hybridised to the vSagC1 probe, again under reduced stringency conditions.

The absence of strong hybridisation with both of the vSag constant region probes suggested that none of the clones had an extended region of homology to vSag spanning the two conserved regions of the retroviral gene. There was also little evidence of homology to other retroviral genes. MMTV *gag*, *pol* and *env* probes failed to hybridise when the formamide concentration was 30%. Weak hybridisation was detected with the gag-pol probe at lower stringency conditions with three of the clones isolated using the vSagC1 probe, although these hybridisation signals were very weak.

Southern blotting of single restriction enzyme digests of the 32 clones isolated with the combined vSagC1 and vSagC2 probes, suggested that 17 clones shared homology with vSagC1 and 15 clones shared homology with vSagC2. One clone (19b) which apparently hybridised preferentially, albeit weakly, to the vSagC1 probe in the hybridot analysis, was found to hybridise only to the vSagC2 probe in the Southern blotting experiments.

The division of the clones into groups based on hybridisation strength was further clarified by restriction mapping. This confirmed that the group 1 clones isolated with the vSagC1 and vSagC2 probes each represented overlapping clones. Two families of overlapping clones made up the group 2 clones isolated with the vSagC1 probe. Four overlapping clones were found within the group 2 and group 3 clones isolated with the

vSagC2 probe. These five families of overlapping clones accounted for 18 of the 32 clones originally isolated with vSagC1 and vSagC2 probes, of which only three clones had been categorised below group 2 based on hybridisation strength.

Of the remaining fourteen clones, all hybridised relatively weakly to their respective probe. Nine were originally isolated using the vSagC1 probe, and represented all of the clones categorised as weakly hybridising group 3 or group 4 clones. The remaining five clones were originally isolated with the vSagC2 probe and again were all either group 3 or group 4 clones according to their hybridisation strength. None of these clones appeared to be overlapping as there were no shared restriction fragment lengths following the single digests. It was not possible to group these fourteen weakly hybridising clones any further, and as they had only weak homology to the murine vSag gene, they were not investigated any further.

All nine of the λ clones isolated from the human placental library with the mixture of vSag-related PCR products HRC1 and HRC2, hybridised to HRC2 alone. No recombinants hybridising to HRC1 were identified. This is in keeping with the result of Southern blotting experiments with normal placental DNA, which also failed to identify a hybridisation signal with the HRC1 probe (data not shown). Only one band was identified on the Southern blots with the HRC2 probe, suggesting that the 9 different λ dash recombinants isolated were most likely to represent overlapping clones from the same genomic loci. This was confirmed by the restriction maps of the nine λ clones (fig. 3.9). Three of these clones (8x, 16y, 18x) hybridised weakly to the vSagC2 probe under reduced stringency (25% formamide), but none hybridised to MMTV *gag*, *pol*, *env* or LTR probes.

The failure to identify the genomic loci from which the HRC1 PCR product was amplified may have resulted from a defective probe. However the same result was obtained with a newly prepared HRC1 probe. In the dot-blot hybridisation experiments the HRC1 probe hybridised to itself at high stringency and to MMTV 3' LTR containing plasmid under reduced stringency conditions, demonstrating that the HRC1 probe hybridised under the conditions employed. Failure of the HRC1 probe to hybridise to genomic DNA under the high stringency conditions employed when the placental library was screened, may have resulted from differences between the human sequence and the murine PCR primer sequences, which account for 44 nucleotide out of a 96 nucleotide sequence. Alternatively, the HRC1 sequence identified by PCR may not be present in normal placental DNA. Complex primer dimers can be generated by PCR reactions and a very faint band of similar size to HRC1 was seen when target DNA was omitted from the reaction (Fig. 2.2). Although the HRC1 sequence does not reveal obvious repetition of the PCR primers, a short stretch of the sense primer is repeated, raising the possibility that this product may be a primer dimer.

In order to validate the restriction maps deduced from the single restriction enzyme digests, representatives from the six families of overlapping clones were selected and digested with combinations of *Eco*RI, *Hind*III, *Xba*I and *Bam*HI. Restriction fragment length analysis and hybridisation confirmed the restriction maps for all but clones 26a and 44b (C1A locus) which required further analysis using additional restriction enzymes. These clones (26a and 44b) did not have any *Bam*HI sites and there was only one *Hind*III site at the end of the clone 26a, which was not present in the 44b clone, so combination digests using these enzymes would have been helpful. In addition, while *Eco*RI / *Xba*I double digests were consistent with the restriction map shown in figure 3.7, these appeared to be partial digests, resulting in two restriction

fragments of varying length which hybridised to the vSagC1 probe. Single digests with the restriction enzymes *Pst*I, *Sac*I, *Sma*I and *Kpn*I confirmed that these two clones overlap and Southern blotting gave restriction fragments of 0.6 kb and 1.5 kb (*Pst*I and *Sac*I digests respectively) which hybridised with the vSagC1 probe and were subsequently used for subcloning.

These experiments have identified the six regions of human DNA with the greatest homology to the murine vSag gene. However they give only limited information on the potential importance of these sequences. The most direct way to investigate further their potential to encode a human superantigen was to sequence these regions of DNA, allowing homology and open reading frame analysis to be carried out. This work is described in the next chapter.

# 4 CHAPTER FOUR

# Analysis of λ clones by nucleotide sequencing of DNA restriction fragments

## *4.1 Introduction*

The initial characterisation of the 41 λ clones isolated in the chapters two and three led to the identification of six human genetic loci that merited further investigation. 27 of the 41 λ clones isolated from the DNA library were from these six loci, as several overlapping clones had been identified. The remaining 14 λ clones hybridised relatively weakly, nine to the vSagC1 probe and five to the vSagC2 probe and were not investigated further. Representative clones from each of the six loci were selected for detailed restriction mapping and subsequent sequencing. Three clones (λ26a, λ38a and λ41b) that hybridised to the vSagC1 probe, and were therefore potentially related to the first constant region of murine vSag, were selected to represent the loci C1A, C1B, and C1C respectively. The clones derived from the other three loci were potentially related to the second constant region of murine vSag. Two clones (λ13a and λ37a) that hybridised to the vSagC2 probe, were selected to represent the loci C2A and C2B respectively. The clone λ2x was selected to represent the group of nine overlapping clones isolated from the DNA library which hybridised to HRC2, the PCR product amplified from human placental DNA with primers to the second constant region of murine vSag. This chapter will describe the subcloning, detailed restriction mapping and sequencing of these λ clones.

As the λ DNA clones were too long to be directly sequenced, the six representative clones (26a, 38a, 41b, 13a, 37a and 2x) were subcloned into a plasmid vector. The

subcloned DNA fragments were then characterised by restriction analysis, allowing detailed maps of the restriction enzyme sites within each clone. These maps were needed for procedures involved in the sequencing process. Restriction fragments less than1 kb in length which hybridising to the relevant probe were required for the automated sequencing techniques employed. This involved further subcloning in some cases.

## *4.2 Materials and Methods*

### 4.2.1 Subcloning of λ restriction fragments into Bluescript KSM13⁻

The plasmid vector chosen for the characterisation and nucleotide sequencing of the clones was the Bluescript KSM13- (Stratagene Cloning Systems, San Diego, California). The vector is derived form a pUC parent plasmid, but has been extensively modified to enable its use in sequencing. The plasmid contains an ampicillin resistance gene, which allows the selection of bacteria transformed with KSM13- from non-transformed bacteria, by growing the cells on agar plates treated with ampicillin. The vector also contains the gene for β-galactosidase (Lac Z), which results in the blue coloration of colonies from bacteria transformed with KSM13- when grown of plates containing an inducer and substrate for the Lac Z gene. A series of 21 unique restriction sites, comprising the "multiple cloning site" of the vector are located within the Lac Z gene, resulting in the loss of activity when a DNA fragment is inserted into the multiple cloning site. This system allows the identification of recombinant plasmids by colour testing for β-galactosidase activity on indicator plates.

The multiple cloning site of the Bluescript vector is flanked by RNA polymerase promoter sequences derived from the T7 and T3 phages, one at each end and on opposite strands (to enable the transcription of strand-specific RNA probes from the

cloned insert) and by the universal forward and reverse primer sites to enable sequencing reactions. The composition of the multiple cloning site of KSM13- is shown in figure 4.1.

**T7 promoter**
*Sac*I
*Bst*XI
*Sac*II
*Eag*I
*Not*I
*Xba*I
*Spe*I
*Bam*HI
*Sma*I
*Pst*I
*Eco*RI
*Eco*RV
*Hind*III
*Cla*I
*Sal*I
*Hinc*II
*Acc*I
*Xho*I
*Dra*II
*Apa*I
*Kpn*I
**T3 Promoter**

**Figure 4.1 Multiple cloning site of the KSM13- Vector**

Restriction maps of the six human genetic loci with potential homology to vSag (chapter 3) allowed the production of restriction fragments from the relevant λ clones of between 1 kb -7kb which hybridised with the vSag probes as shown in Table 4.1.

| loci | probe | phage | Restriction enzymes | Fragment size |
|---|---|---|---|---|
| C1A | vSagC1 | λ26a | *Sac*I | 1.2 kb |
| C1B | vSagC1 | λ38a | *Hind*III + *Xba*I | 2.3 kb |
| C1C | vSagC1 | λ41b | *Hind*III + *Xba*I | 1 kb |
| C2A | vSagC2 | λ13a | *Bam*HI + *Xba*I | 7 kb |
| C2B | vSagC2 | λ37a | *Hind*III | 3.2 kb |
| HRC2 | HRC2 | λ2x | *Bam*HI + *Hind*III | 3.2 kb |

**Table 4.1 Restriction fragments of the six representative λ DNA clones which hybridised to vSag derived probes.** These restriction fragments were ligated with the KSM13- plasmid and XL-1 cells transformed by electroporation.

Restriction fragments from these six λ clones were subcloned into the Bluescript KSM13- vector. The bacterial host chosen for the growth of the plasmid was E. coli XL-1. In brief, the subcloning was achieved by digesting both plasmid and λ dash recombinant DNA with the appropriate restriction enzymes, after which the digested DNA was ligated and used to transform XL-1 cells. Bacterial colonies were grown up overnight and replica nitrocellulose filters hybridised with the relevant probe as described in chapter 2. Bacterial colonies transformed with recombinant plasmids that hybridised with the probes were selected and the recombinant DNA was analysed to identify appropriately sized inserts.

**Preparation of KSM13- DNA for subcloning**

In order to accept the restriction fragments of the λ clones, KSM13- DNA was digested with the same restriction enzymes used to produce the fragment and the termini of the linearised plasmid dephosphorylated with calf intestine alkaline phosphatase (CIP) to prevent self-ligation of the vector during subcloning.

10 μg KSM13- DNA was digested to completion with 4 fold excess of the appropriate restriction enzymes (Table 4.1) and then extracted twice with phenol/chloroform, once with chloroform and precipitated with ethanol. The DNA pellet was then dissolved in 50 μl of a buffer containing 50 mM Tris (pH 9), 1 mM $MgCl_2$, 100 μM $ZnCl_2$, and 1 mM

spermidine. The digested plasmid was dephosphorylated by adding 4 μl of alkaline phosphtase (0.09 units CIP which represents a two fold excess of the enzyme) and incubating at 37°C for 1 hour. (One unit of CIP is the amount of enzyme required to dephosphorylate 100 pmol of 5' ends of DNA; which represents 222 μg of a 3 kb linear DNA molecule such as KSM13-). The CIP was inactivated by heating at 68°C for 10 minutes, after first adjusting the reaction mixture to 30mM Tris (pH 8), 50 mM NaCl and 0.5 mM EDTA. The dephosphorylated DNA was then extracted twice with phenol/chloroform, once with chloroform (followed by a TE back extraction) and precipitated with ethanol. The yield of the dephosphorylated plasmid DNA was determined by running an aliquot of each digest on an agarose minigel together with 100 ng undigested KSM13-. The final concentration of the DNA was adjusted to 200 μg/ml with TE and stored at –20°C.

**Subcloning and bacterial transformations**

The λ dash clones shown in Table 4.1 were subcloned by digesting 2.5 μg of DNA for 1 hour at 37°C with a two-fold excess of the appropriate restriction enzymes. Where ligation of blunt ended fragments was required, the restriction fragments ends were 'filled-in' after digestion by incubating with Klenow (0.5 units) and 0.2 mM dNTPs for 5 minutes at room temperature, 5 minutes on ice and 5 minutes at 70°C. After estimating the yield (by electrophoresing an aliquot of the digested DNA on a 0.8% minigel) 500 ng of the digested λ clone was ligated with 50 ng of the dephosphorylated plasmid DNA in a reaction mixture containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 10 mM DTT, 50 μg/ml BSA and T4 DNA ligase (0.2 units in 5 μl reaction mixture). Ligations were incubated overnight at 15°C. The ligation was checked by electorphoresing samples taken before addition of the ligase and after incubation on a minigel.

Transformation of XL-1 cells was achieved by electroporation. For each ligation, an aliquot of XL-1 cells in glycerol was allowed to thaw and 0.5 μl of ligation reaction mixture added. The DNA/cell mixture was transferred to an electroporation cuvette and pulsed with 2.5 kilovolts. Immediately following electoporation, 600 μl of SOC broth was added and the mixture incubated at 37°C for 45 minutes to allow the cells to recover. The cells were then spread on agar plates and incubated overnight at 37°C. Nitrocellulose replica filters were made (see materials and methods chapter 2) and hybridised with the relevant probes, thereby allowing selection of recombinants of digestion fragments containing the region of DNA hybridising to these probes. In subsequent subcloning procedures, cells were spread on indicator plates (see section 2.2.2), incubated overnight at 37°C and colonies containing DNA insert (white) selected for DNA minipreparation as described in chapter 2. Recombinants of the appropriate size were identified by electrophoresing on a 3% agarose gel.

### 4.2.2 Restriction mapping of subcloned DNA.

The initial analysis of the subclones consisted of digestion with a panel of eleven restriction enzymes to determine which sites were contained within each clone. The enzymes used (obtained from a number of suppliers) were; *Apa*I, *Bst*XI, *Eco*RV, *Hinc*II, *Kpn*I, *Pst*I, *Sal*I, *Sma*I, *Xho*I, *Sac*I and *Sac*II. (Further digests of the λ26a subclone were not undertaken, as a detailed restriction map had already been produced using the enzymes *Pst*I, *Sac*I, *Sma*I and *Kpn*I). These restriction enzymes cut only at the multiple cloning site of KSM13- (Figure 4.1). In some cases, these restriction sites would have been excised during the subcloning procedure, if they lay between the two restriction sites into which the recombinant had been ligated. Combinations of these enzymes with others which cut within the vector multiple cloning site only were used to obtain accurate measurements of fragments which would otherwise include the whole plasmid. DNA was digested for 3 hours at 37°C in 5

μl reaction mixtures containing 200 ng of DNA and 50 μg/ml BSA in a restriction buffer of the composition recommended by the enzyme manufacturer. Buffers were selected which were compatible with the combination of enzymes used. The restriction enzymes were used at 4 fold excess. Digested DNA samples were then electrophoresed on 0.8% agarose minigels with *Hind*III-digested λ markers. The pattern of bands produced by restriction digest was observed by staining the gel in a solution of 0.5 μg/ml ethidium bromide and illuminating under ultraviolet light. The sizes of the restriction fragments produced by each digest were determined by comparison of the migration distance of the bands to the λ DNA markers.

### 4.2.3 Automated DNA sequencing of subcloned DNA.

DNA was sequenced by the staff of the Molecular Biology Facility at Newcastle University, using dye terminator cycle sequencing (Perkin Elmer), as described in section 2.2.3. Purified double stranded DNA was used as the template with the universal forward and reverse primers at each end of the cloned insert. This sequencing technique produces readable sequences of up to 500 nucleotides from the primer. Subcloned inserts of up to approximately 1000 nucleotides could therefore be sequenced using both the forward and reverse primers to sequence form each end of the insert (off opposite strands and in opposite directions).

### 4.2.4 BLAST analysis of DNA and predicted amino acid sequences

The statistical significance of the regions of homology identified between the vSag probes and the six genomic loci hybridising to these probes was assessed using the computer facilities provided by the National Center for Biotechnology Information

(National Library of Medicine, National Institute of Health, Bethesda, USA, http:/www.ncbi.hih.gov). BLAST (Basic Local Alignment Search Tool) is the search algorithm used by a number of search tools (see Box 2) to calculate significance of homology using the statistical methods described by Karlin and Altschul (232).

| | |
|---|---|
| blastp | compares an amino acid query sequence against a protein sequence database |
| blastn | compares a nucleotide query sequence against a nucleotide sequence database |
| blastx | compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database |
| tblastn | compares a protein query sequence against a nucleotide sequence database dynamically in all six reading frames (both strands) |
| tblastx | compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database |

**Box 2**

Local alignments are assessed by means of a score, which is computed as the sum of scores for aligned pairs of residues and scores for gaps. The fundamental unit of the BLAST algorithm output is the High-scoring Segment Pair (HSP). An HSP consists of two sequence fragments whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cut-off score. The task of finding HSPs begins with identifying short words of length W in the query sequence that either match or satisfy the threshold score T when aligned with a word of the same length in

a database sequence. These initial word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignments score can be increased. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score drops below zero, or the end of either sequence is reached. The sensitivity of the programme can be adjusted by changing the parameters W, T and X, and by adjusting the reward for matches and penalty for mismatches.

In blastn, the M parameter sets the reward score for a pair of matching residues; the N parameter sets the penalty score for mismatching residues. M and N must be positive and negative integers, respectively. Higher ratios of M:N corresponds to increasing divergence of nucleic acids (PAMs: point accepted mutations per 100 residues). The default values for M and N are 5 and –4, giving a ratio of 1.25. This corresponds to about 47 nucleic acid PAMs, or about 58 amino acid PAMs for protein searches.

In general, the blastn programme is not intended for finding distantly related nucleic acid sequences. The parameters are optimised for speed, not sensitivity. The blastp, blastx, tblastn and tblastx offer more flexibility in scoring systems. The default scoring matrix used by these programmes is the BLOSUM62 matrix (233), but several PAM amino acid scoring matrices are provided in the BLAST software. Each matrix is most sensitive at finding similarities at its particular PAM distance. Searches using a combination of scoring matrices are required particularly when the mutational distance between potential homologs is unknown and the significance of their similarities may be weak.

The principle equation relating the score of an HSP to its expected frequency of chance occurrence is:

$$E = K N \exp(-\text{Lambda } S)$$

where E is the expected frequency of chance occurrence of an HSP having score S (or higher) ; K and lambda are Karlin-Altschul parameters; N is the product of the query and database sequence lengths and exp is the exponentiation function. Lambda may be thought of as the expected increase in reliability of an alignment associated with a unit increase in alignment score. Reliability is expressed in units of information (bits). The expectation E calculated for an alignment between the query sequence and a database sequence can be extrapolated to an expectation over the entire database search, by converting the pairwise expectation to a probability and multiplying the result by the ratio of the entire database size (expressed in residues) to the length of the matching database sequence. Due to inaccuracy in the statistical methods as they are applied in the BLAST programme, whenever E is less than about 0.05, its value can practically be treated as being equal to the probability (p).

Searches were performed using the two "non-redundant" sequence databases maintained by the NCBI (one for proteins and one for nucleic acids). These databases are constructed using sequence data from several sources, including GenBank, the EMBL Data Library, the DNA Database of Japan (DDBJ), as well as data from US and European patents, and are continually updated. Eliminating the large degree of internal "redundancy" between sequences derived from many different sources, allows faster database searches, and avoids large outputs of similar or identical alignments which can obscure novel matches.

## *4.3 Results*

### 4.3.1 Subcloning λ DNA recombinants into the KSM13- vector.

To facilitate DNA sequencing, the restriction fragments of the six representative λ recombinants shown in Table 4.1 were subcloned into KSM13- vector. Subclones containing DNA that hybridised to the relevant probes were identified for each of the six λ clones. The density of the positive colonies produced after transforming XL-1 cells with the KSM13- vector ligated with the 7 kb *Bam*HI/*Xba*I fragment of λ13a was very low. This may have been due to the large size of the recombinant. The sizes of the inserts obtained from digesting minipreps of the purified subclones were as expected based on the restriction maps.

Further subcloning was required prior to DNA sequencing, to produce recombinants no greater than 1 kb in length which hybridised to the relevant probes. More detailed restriction maps were required for all but the λ26a subclone. Additional restriction sites were identified in the other subclones by digesting the subcloned DNA with the panel of restriction enzymes shown in Table 4.2.

| | 38a | 41b | 13a | 37a | 2x |
|---|---|---|---|---|---|
| *Apa*I | - | - | - | + | - |
| *Bst*XI | + | + | + | - | + |
| *Eco*RV | - | - | - | - | - |
| *Hinc*II | - | - | - | - | - |
| *Kpn*I | - | - | + | - | + |
| *Pst*I | - | - | + | + | + |
| *Sal*I | - | - | - | - | - |
| *Sma*I | - | + | + | - | + |
| *Xho*I | - | - | - | - | - |
| *Sac*I | + | - | - | - | - |
| *Sac*II | - | - | - | + | - |

**Table 4.2 Restriction sites within the subcloned DNA restriction fragments.** The restriction fragments of λ38a, λ41b, λ13a, λ37a and λ2x (Table 4.1), which had been subcloned into KSM13-, were singly digested with *Apa*I, *Bst*XI, *Eco*RV, *Hinc*II, *Kpn*I, *Pst*I, *Sal*I, *Sma*I, *Xho*I, *Sac*I and *Sac*II to identify additional restriction sites. (+ restriction site(s) present within recombinant DNA, – no restriction site present within recombinant DNA).

### 4.3.2 Restriction mapping and sequencing of vSagC1 hybridising DNA subclones.

The restriction maps of the three subcloned λ DNA fragments to which the vSagC1 probe hybridised are shown in Figure 4.2.

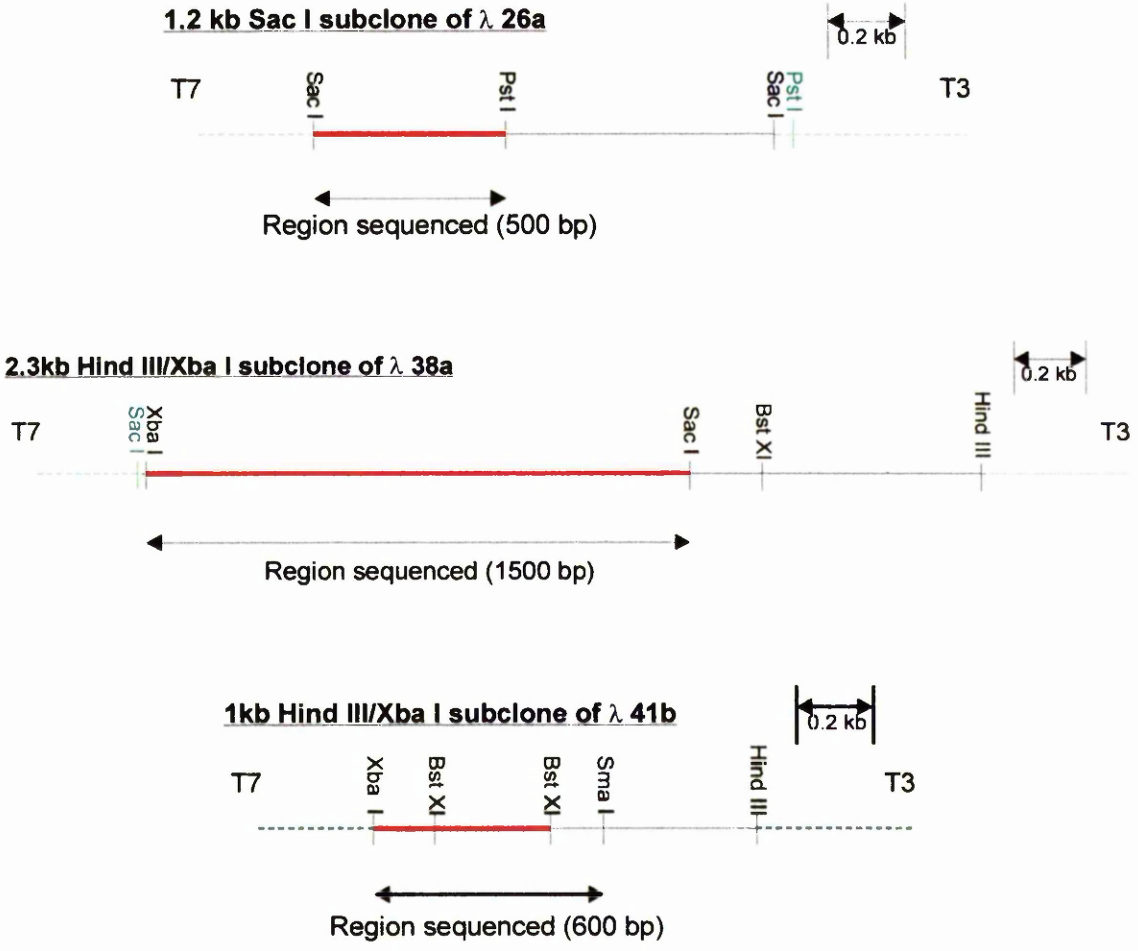



**Figure 4.2 Restriction maps of vSagC1 hybridising DNA subclones.** The restriction sites within the DNA subclones were positioned by analysis of single digests with *Apa*I, *Bst*XI, *Eco*RV, *Hinc*II, *Kpn*I, *Pst*I, *Sal*I, *Sma*I, *Xho*I, *Sac*I and *Sac*II. The region of the subclones to which the vSagC1 probe hybridised is highlighted in red and the region sequenced indicated.

The restriction map for λ26a (deduced from figure 3.11) shows a 0.5 kb region to which the vSagC1 probe hybridised, flanked by *Sac*I and *Pst*I restriction sites, which was an appropriate length for sequencing. The subcloned 1.2kb *Sac*I fragment of λ 26a was digested with *Pst*I to remove the 0.7 fragment downstream from this region and the vector containing the hybridising 0.5 kb fragment re-ligated. Transformed XL-1 cells with this new subclone were then isolated and purified prior to sequencing with both forward and reverse universal primers.

For the 1kb *Hind*III/*Xba*I subclone of λ 41b, the downstream 0.4kb fragment flanked by *Sma*I and *Hind*III was digested out and ligated to give a 0.6 kb recombinant for sequencing. This involved the ligation of blunt ended fragments as the *Sma*I site in the vector's multiple cloning region would have been excised when the original 1 kb *Hind*III/*Xba*I fragment was ligated into this site.

The 1.5 kb *Sac*I restriction fragment of λ38a was subcloned by ligating with the KSM13- vector and transformed XL-1 cells isolated as previously described. The λ38a derived subclone containing the 1.5 kb *Sac*I fragment was still too large to be sequenced, and required further digestion. Restriction enzymes which did not cut the KSM13- vector were used to digest the subclone. The 1.5 *Sac*I fragment was found to have a restriction site for *Bgl*II, as shown in figures 4.2 and 4.3. By digesting out the 0.6 kb fragment upstream and the 0.9 kb fragment downstream from the *Bgl*II site, and ligating the blunt ends of the residual recombinants, two new subclones were isolated. These were sequenced with forward and reverse universal primers giving a sequenced region of 1.5 kb.

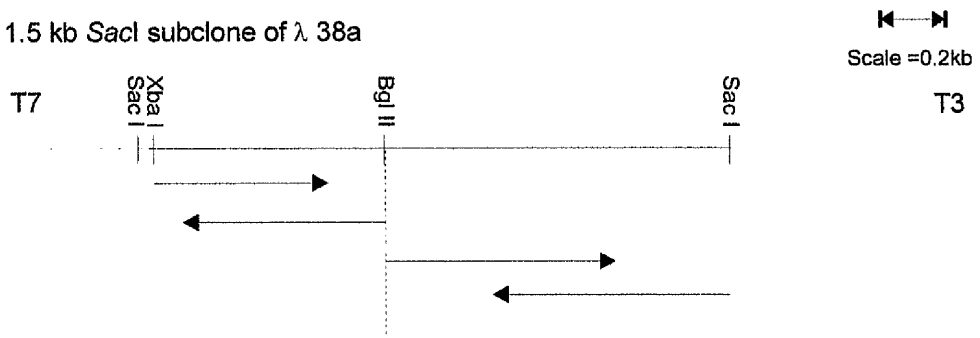



**Figure 4.3 Restriction map of 1.5 kb subclone of λ38a.** Restriction digestion with *Bgl*II divided the 1.5 kb subclone of λ38a into 0.6 kb and 0.9 kb fragments which were subcloned and sequenced separately using the universal forward and reverse primers.

The DNA sequences of the three representative subclones hybridising to vSagC1 are shown in figure 4.4. The lengths of these sequenced regions ranged from 478 – 1422 nucleotides. The reverse (lower) strand is shown for the sequences derived from the λ clones 26a (C1A) as homology to the vSagC1 probe was identified in this orientation. The restriction enzyme sites at each end of the cloned regions are highlighted together with the any restriction sites within the cloned sequences. Alterations of the restriction sites resulting from ligation of blunt ended fragments are also shown.

The graphical representation of a typical sequencing reaction is shown in figure 4.5. This sequencing reaction was performed on the 0.6 kb *Xba*I/*Sma*I subclone of λ 41b using the universal forward primer. This shows the multiple cloning site of the KSM 13-vector starting at position 77 (Sac I- GAGCTC) with the insert cloned into the *Xba*I site (TCTAGA) at positon100. Although the sequence data beyond position 500 was difficult to interpret, sequencing of the reverse strand of the subclone gave a large region of overlap. The two sequences were then merged using DNA analysis software

(Microgenie), giving the complete sequence of the subclone, with a total length of 632 nucleotides.

| Loci | probe | phage | length |
|---|---|---|---|
| C1A | vSagC1 | λ26a | 478 bp |

Reverse strand
*PstI*
CTGCAGCAAT CCCCCNTTCC CAAAGTTANG GGGATTATAA GCACGGAGCC CATCATGCCC 60

NGGCNTANGG ATTTCTTACT GAGAAAACTG GATTCCACAA AAAATCATAG ACTATTACAA 120

AACTGGTTGT TTGAAATTGT ATCAGTAAGA ATAAAAACAT TCCACTCTTG CTTGGAGGAC 180

CTAAGACACA GAGAAGCAGT CTTGACTTCC AAGCCAGGNG AAGGAAAAAG GGGTTTCTGA 240

AGACAACAGT TCCCATCTAT CTTAACTGTA GTATTCACGG AAACAGAACC TGGGGTCCAT 300

GGTGGTCCAG ACCTGATGTT AATGCCNTTA TTTGCAGCTC TGCTTTGCCT GGGGCCTCTC 360

AAAAACACTG CTTCATTTAA ATGTCATCAG AACTCCACAC CTCCCAGAAA GCCCTGGGTT 420

*SacI*
AATTTCATCC ATTCCTTTGT TTGGTGAGAT GTTCCAAGGT TNCCTTGAGC GTGAGCTC 478

| Loci | probe | phage | length |
|---|---|---|---|
| C1B | vSagC1 | λ38a | 1422 bp |

Forward strand

*XbaI*

TCTAGAACTC TCATCTCAAC TCTTGATATC CATCCAGCTT CCTGGAGATG GCTACCCAC 60

CAGAAGTCAG CGACCAGAGA AAACCTCCAC TTCACAAATG CACATTGACA CTGTAAGCAG 120

TGGAATAAAT TGGCCTGTTT TCATTGCAGA AAAATCAGCA TATTCTGAAA TGATAGAAGC 180

TCCCCGTCAA TTAGGGAAAT TAACATCCCT CTCTGTGGTT GGTGCTATTG CTGCACTTCA 240

TTTGCCTGAC CCAAAGCACA GCAGTTTCTT TTTCATGAGC CGTTCCAGTT GGAAAGTAAA 300

TAGGATTTTT GGGGTCTTCT TTCTGTGGGT CCTGCTCCAG ACTGGGCTCT CAATTCTGCC 360

TGTGGTGTCC TCACAGAGTG AGATGGGAGG CAGATGCACA TTGGGGAAAG TTGCGGTTGT 420

GCTATTTGGA GCAAGTAATA ATAGACATAG TAAAGCACAA ATTAAGAATT GTTCTTGCAC 480

*BglII*

AACATGCCCA GCTGTCTCTA GTGAGGAACA GCAGGCGAGT AGATCAGATC TACTGGACTG 540

GATCAATAAG TCCCATTGGT TAAAATGATC GGGGCAAAAT GATTAAGAGA TCCGAGTTTC 600

AGTTTACTTA GTAAGAAAAT ACAGGCATAT CTCAGAAATA TTGCAGGTTC AGTTCTGGGC 660

CACTGCAATA AAGCAAGTAT CATAATAAAG CAGGTCACAC ACATTTTTTT GGTTCCCTAG 720

TACGTATAAA AGTTATATTT GCACTATGCA AGTCTATTAA GTGTACAATA ATATTTTGTC 780

TAAAAAAAGA TGTACATACC TGAATTAAAA AATACTTTAG TGCTAAAAAA TGCTTACCAT 840

GATCTGAGCC TTCAGTGAGT CATAATATTC TTCCTGGCTG GAGGGGTCTG CCTCAGTATT 900

GATGGCTGCT GACGATCGGG TGTGGTTGCT GAAGGTGGGG TGGCTGTGGC AATTTTTCCT 960

TTTTTTTGGG AGAGGGGTTC TGGCTTGTCA TCCTGGCTAG AGGGCAGTGG TGCATTCATA 1020

GCACCCTGTA ATTTCAAATT CCTGGGCTCA AGTAATGTTC TGGCTTCANC CTTCTGCCTC 1080

AGTTTCCCAA ATAGCAAGGA CNACAGGCTA TTCTTGGATA ACTTTTAAAA AGTTTTTTAG 1140

AAACAAGGTC TCACTATGTT GTCCAGACTA TGTGAGCAAT ATTGATTGAT TGATTGATTG 1200

AGTCAGGGTA TCCCTCTGTC GCCCAGGCTG GAGTGCAGTG GTGTGTTCTT GGCTTACTGC 1260

AGCCTCCGCC TCCTGGGCTC AAGTGATCTT CCCGTCTCAG CCTCCCGAGT AGCTGGGACC 1320

ACAGGTGCAT ACCACCATGT CTGGCTGATT TTCTTTTTTG TAATTTTTGA TAGAGATGGG 1380

*SacI*

GTTTTGCCAT GTTGCCGACG TGGGTCTTGA ACTCNTGAGC TC 1422

| loci | probe | phage | length |
|---|---|---|---|
| C1C | vSagC1 | λ41b | 632 bp |

Forward strand
XbaI
TCTAGAGCTG TCGACGCGGC CGCGTAATAC GACTCACTAT AGGGCGAAGA ATTCGGATCC 60

CACATTAAGT CAGGTGCTAC TTGGGGAAGT TGCTTCGTTT TCTTGTGCTC TAAGATCTCT 120
BstXI
CACCCACACG CCTGGTTTTG GGGAGAAGGA TAAGGACGGT GTTTCTAAAC ATGCTGTGTT 180

TCCTATGCAG AGAGGCAGGT GACTGTGTAT GGCTCAACTT GCAGACTCTA AAGCCAACTA 240

CTTTAGAGTT CAGAGTTCAG ACACCAGCTT TGCCTCTTGT CCTTGTAACA TGGCACAGGT 300

TACTTCAACT CTGTAATGCC TCAGTTTCCT CATCTGTGAA ATGGGGGTAA TAACTGTGTT 360

ATACCTCATA AGTGTTGCTT GGAATAATGC CCACTACCAC GTTCAAAATT CANCTGCTGG 420
BstXI
GGCCAAGCAT GGTGTCTGAC CCCTGTAATC CCAGCACTTT GGGAGGCCAA GGCAGGCAGA 480

TGACCGGAGG TTGGGAGTTC GAGACCAGCC TGACCAACAT GACAAAACCC CATCTTTACT 540

AAAAATACAA AAATTAGCAG GGTATGGTGA CAAGCACCTG TAATCCCAGC TACTCAGGAG 600
SmaI/HindIII (blunt)
GCCGAGGCAG GAGAATCACT TGAACCCAGC TT 632

**Figure 4.4 Nucleotide sequences of the representative vSagC1 hybridising subclones.** Plasmids containing recombinants that hybridised to the vSagC1 probe were sequenced using dye terminator cycle sequencing (Perkin Elmer) with AmpliTaq® DNA Polymerase, FS, using universal forward and reverse primers. The positions of restriction endonuclease sites are indicated. Where the sequence data gave an equivocal result for a specific nucleotide, or where there was a discrepancy between the forward and reverse sequencing, the letter N was used. The number on the right indicates the position of the last nucleotide in the line.

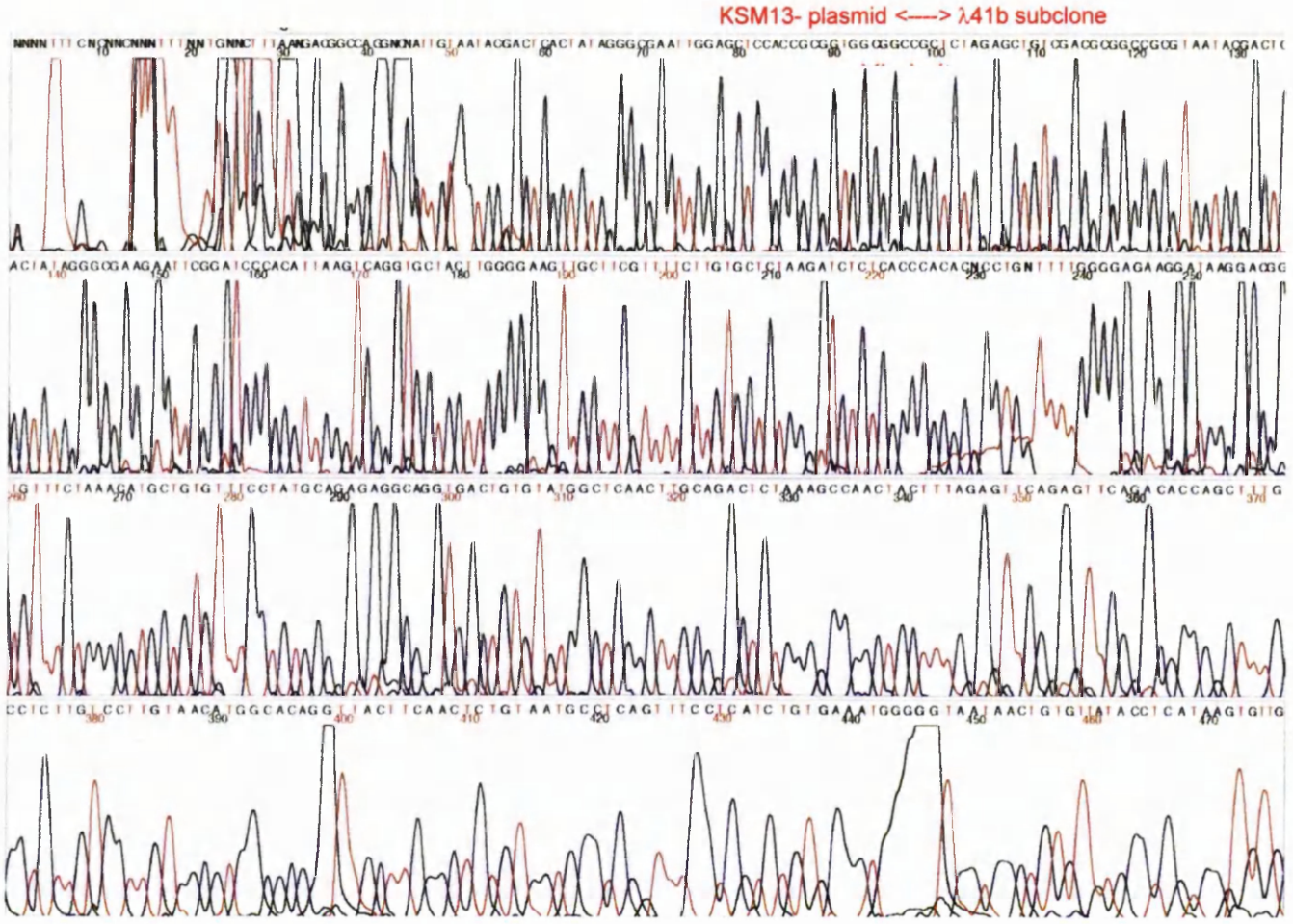



**Figure 4.5 Graphical representation of a sequencing reaction.** The 0.6 kb *Xbal*/*Sma*I subclone of λ 41b was sequenced by dye terminator cycle sequencing (Perkin Elmer) with AmpliTaq® DNA Polymerase, FS, using the universal forward primer. The recombinant is cloned into the Xba I site (TCTAGA) at positon100 as indicated.

### 4.3.3 DNA sequence homology to the murine vSagC1 probe.

Regions of homology between the human DNA sequences and the vSag derived probes were initially identified using the Microgenie computer software which allows simultaneous analysis of both forward and reverse nucleotide strands. The sequences of the three subclones hybridising to the vSagC1 probe are shown in figure 4.6. The similarity indices (number of nucleotide matches divided by the total length of the sequence) ranged from 52.2%-58.2%. For the C1A sequence, there was 100% homology over the 20 nucleotide region corresponding to vSagC1 20-39. For the C1B sequence, a 23/25 (92%) nucleotide match was identified with vSagC1 41-65. Although the C1C sequence had a shorter region of exact homology, over the length of the probe the similarity index was highest (58.2%) with a 39/53 (74%) nucleotide match over the region vSagC1 36-88.

## C1A locus

```
vSagC1  TCTGCTG.CAAACTTGGCATAGCTCTGCTTTGCCTGGGGCTATTGGGGGAAGTTGCGGTTCGTGCTCGCAG
C1A     GTTAATGCCNTTATTTGC..AGCTCTGCTTTGCCTGGGGCCTCTCAAAAACACTGC..TTCAT.TTAAATG
          *  ** *    ** **  ********************   *     *   ***  *** *  *    *

vSagC1  GGCTCTCACCCTTGACTCTTT
C1A     TCATCAGA.ACTCCACACCTC
           **  *  **  ** * *

Matches         48      Mismatches      37      Unmatched       7
Length          92      Matches/Length  52.2%
```

## C1B locus

```
vSagC1  TCTGCTGCAAACTTGGCATA..GCTCTGCTTTGCCTGGGGCTATTGGGGGAAGTTGCGGTTCGTGCTCGCAG
C1B     CCTG.TGGTGTCCTCACAGAGTGAGATGGGAGGCAGATGCACATTGGGGAAAGTTGCGGTT.GTGCTATTTG
         *** **    * *  ** *  *   **    **    *   ******* ********** *****    *

vSagC1  G.GCTC.TCACCCTTGACTCTTT
C1B     GAGCAAGTAATAATAGACATAGT
        * **   * *   * ***    *


Matches         50      Mismatches      44      Unmatched       6
Length          95      Matches/Length  52.6%
```

## C1C locus

```
vSagC1  TCTGCTGCAAACTTGGCATAGCTCTGCTT.TGCCTGGGGCTATTGGGGGAAGTTG...CGG.TTC..GTGCTCGCAG
C1C     TATAGGGCGAAGAATTCGGATCCCACATTAAGTCAGGTGCTACTTGGGGAAGTTGCTTCGTTTTCTTGTGCTCTAAG
        * *   ** **     *  * * *   **  * * ** **** * **********   **  ***  ******  **

vSagC1  GGCTCTCACCCTTGACTCTTT
C1C     ATCTCTCACCC.ACACGCCTG
          *********   ** * *

Matches         57      Mismatches      33      Unmatched       8
Length          98      Matches/Length  58.2%
```

**Figure 4.6 DNA sequence homology between vSagC1 and human subclones hybridising to this probe.** 'Microgenie' computer software was used to align the nucleotide sequences of subcloned λ26a (C1A locus), λ38a (C1B locus) and λ41b (C1C locus) against the sequence of murine vSag (MMTV 3'LTR (230)). Asterisks indicate matching nucleotides. The similarity index is the percentage of matching nucleotides over the length of the sequence.

### 4.3.4 Restriction mapping and sequencing of vSagC2 hybridising DNA subclones.

The restriction maps of the two subcloned λ DNA fragments to which the vSagC2 probe hybridised are shown in Figure 4.7.

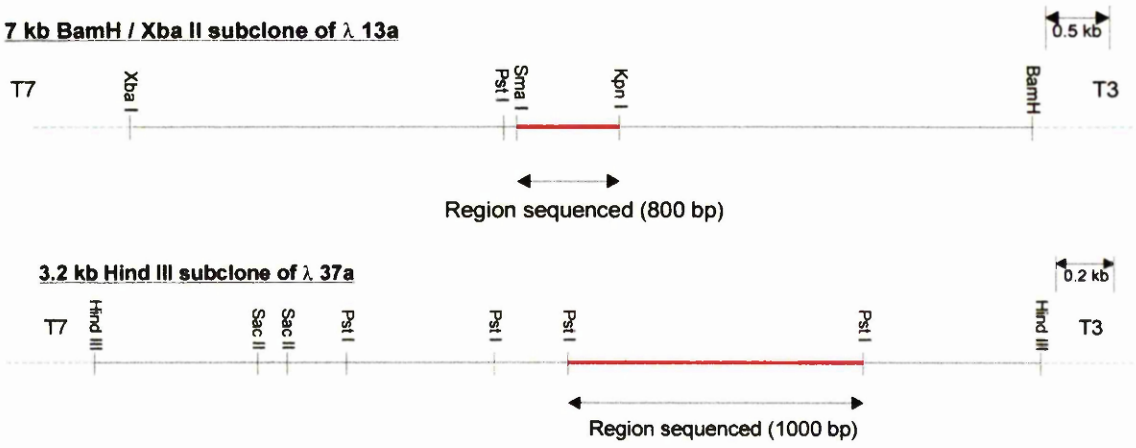



**Figure 4.7 Restriction maps of vSagC2 hybridising DNA subclones.** The restriction sites within the DNA subclones were positioned by analysis of single digests with *Apa*I, *Bst*XI, *Eco*RV, *Hinc*II, *Kpn*I, *Pst*I, *Sal*I, *Sma*I, *Xho*I, *Sac*I and *Sac*II. The region of the subclones to which the vSagC2 probe hybridised is highlighted in red and the region sequenced indicated.

To produce recombinants under 1 kb in length which hybridised to the vSagC2 probe, further subcloning was required prior to sequencing. The 0.8 kb restriction fragment of λ13a flanked by *Sma*I and *Kpn*I, and the 1 kb *Pst*I restriction fragment of λ37a (Figure 4.7) were ligated with the KSM 13- vector. Transformed XL-1 cells were isolated as described previously. New subclones containing the 0.8 kb *Sma*I/*Kpn*I fragment of λ13a and 1 kb *Pst*I fragment of λ37a were identified by digesting minipreps with the appropriate enzymes and electrophoresing on minigels.

The DNA sequences of the two representative subclones hybridising to the vSagC2 are shown in figure 4.8. The lengths of the sequenced regions were 819 bp and 952 bp for the C2A and C2B loci respectively. The restriction enzyme sites at each end of the cloned regions are highlighted.

| loci | probe | phage | length |
|---|---|---|---|
| C2A | vSagC2 | λ13a | 819 bp |

Forward strand
SmaI
CCCGGGTTCA TGCCATTCTC CTGCCTCAGC CTCCTGAGTA GCTGGGACTA CAGGCGCCCG 60

CCACCATGCC CAACTAATTT TTTTGTATTT TTAGTAGAGA TGGGGTTTTA CTGTGTTAGC 120

CAGGATGGTC TTGATCTCCT GACCTCGTGA TCTGCCCACA TCGGCCTCCT AAAGTGTTGG 180

GATTATAGGC GTGAGCCACC GCACCTGGCC GCCACATACT TTTAAACAAC CAGATCTCAC 240

AAGAACTCAC TCACTATCAG CAGGACAGTA CCAAGCCATT CATGAAGGAT CTGCCCCATA 300

ACCCAAACAC CTCCCACTAN GGCCCACTTC CAACACTGGG GAATTATATT TCAACATAAG 360

ANTTGGGGGG ACAAATATCC AAACTATATC ACTTCGGTTT AATTGTCTGT TAATCAGAAC 420

CTTGAAATCA CTGCCTTCTC CTTTGCCTCT TTCAATCAAC CCAAGGCCCA CCCCAGAATT 480

TCTGTACATA TGTCACTTAA CTTGNTAGCA ACCCTTGGGG GAAGGGAGTA GACTACACTC 540

AAGTTCTCTG NACATCCCTA GTCATTGAGT TGCCACTTTC ATGTCTGCAA GTGGCCGGCA 600

GCCGTGCAAT GAGGGAGACA GGTGGTGGCG CCCAGGTCCT GCCCTAGGTA GAAAAGTGAT 660

GGAAGATCCC TGCTTGATGC TGAGGCCCCA GCCGCGCCCG CCCTGGAAAG ATAAAAGAGG 720

AACTGGCCTA TCCTATCCTG ATGCCGAAGG GAAGGGAGAG AAAGGAAAAT NTCCTNTAAG 780

KpnI
AGCAAAGAGC CACCTGGCCC TCCTACAGTG TGTGGTACC 819

| loci | probe | phage | length |
|---|---|---|---|
| C2B | vSagC2 | λ37a | 952 bp |

Forward strand
PstI
CTGCAGTAAG GCTGGACTGG ACTCTTTCAG CCAGGTTTTA GGTGCTCCGG GCACACAGAC 60

GGGAGGATGC GATTTCCCCG TTTACCCTTC CCAGGCCTGA GGGCCTCCAC AGCTGGCCCT 120

CGGCTGTGGC CAGTTATTCC CACTTCTGCC TGCTTTTTTC TTCAGACATC TTATGGTAGC 180

GTCTTGCTAA TGAGGAATTT GGTTTGAGAT TTTGTGTAGT TTTTTACTTG ATGAAATGTC 240

CCAGTTTTAG TTTAACATTG TCCATCACCA NCCTTAAGAT ATAACTTGCT GTTTAAATAC 300

GTCTGTCATT ATTGTTGAAT ATTTACTAAA GTTTCATCAG ACAGGTGGGG CAACCAGGGA 360

TTTCTATTAA GTGANAACAA AATACAGCGT TAATAAGTGA CTCCAAAACC NCCAAAACAT 420

TATGCNAAGT TAAANAAGTA ATATATGAAA AGTTACATAT GGTATGATCC CNTTATATAA 480

ANATCTAAAA TAAANACTCC AGTNGGNGAC ATGGGANACA TGGNTCACAA TTTGGGGAAG 540

AGAGAGTGAA GAAAGAGTTA TGGACCACAG TGTTGGCAAT AACAGAGTTC CAGTAGGATT 600

GGGATTGAAA GGAACCATTT GATTCAGAAG GTGGGAGTTC TTGGCAAAAA CATTCTTGGG 660

GATGGCTGGG CGTGNTGGTT CATGCCCGTA ATCCCAGCAA TTTTGGGAGC CCAAGGCGGG 720

TGGATCACCT GAGGTTCAGG AGTTCAAGAC CAGCTTTGCC AACATGGTGA AACCCTGTNT 780

NTACAAAATA TACAAAAGTT AGCCAGGCGT GGTGGCAGGC GCCTGTAATC CCGGCTGCTC 840

GGGAGGCTGA GGCAGGAGAA TCGCTTGAGC CCAGGAGTCA GTGGTTGCAG TGAGCCGAGG 900

PstI
TTGTGCCATT GCACTCCAGC CTGGGCGGCA GAGTGAGACT CTGTTTCTGC AG 952

**Figure 4.8 Nucleotide sequences of the representative vSagC2 hybridising subclones.** Plasmids containing recombinants that hybridised to the vSagC2 probe were sequenced using dye terminator cycle sequencing (Perkin Elmer) with AmpliTaq® DNA Polymerase, FS, using universal forward and reverse primers. The positions of restriction endonuclease sites are indicated. Where the sequence data gave an equivocal result for a specific nucleotide, or where there was a discrepancy between the forward and reverse sequencing, the letter N was used. The number on the right indicates the position of the last nucleotide in the line.

### 4.3.5 DNA sequence homology to the murine vSagC2 probe.

Regions of homology between the human DNA sequences and vSagC2 are shown in figure 4.9. The similarity indices were lower (48% and 43.7% for the C2A and C2B nucleotide sequences respectively) than with the subclones hybridising to the shorter C1 probe. For the C2A sequence, there was a 21/23 (91%) nucleotide match over the region vSagC2 8-30. A similar region of the C2 probe (vSagC2 10-32) had greatest homology to the C2B sequence with 22/23 (95%) nucleotides matching.

**C2A locus**

```
vSagC2  CAAAATAGGAGACAGGTGGTGGCAACCAGGGACTTA..TAGGG.GACCTTACATCT.ACAGACCAAC.AGATGCC...
C2A     CAATGAGGGAGACAGGTGGTGGCGCCCAGGTCCTGCCCTAGGTAGAAAAGTGATGGAAGATCCCTGCTTGATGCTGAG
        ***    ****************  *****  **    ****  **      **   * *  **  *  *****

vSagC2  CCCTTACCATATACAGGAAGATATGACTTAAATTGGGATAGGTGGGTTACAGTC..AATGGCTA..TAAAGTGTTAT
C2A     GCCCCAGCCGCGCCCGCCCTGGAAAGATAAAAGAGGAACTGGCCTA.TCCTATCCTGATGCCGAAGGGAAGGGAGAG
         **  * *     * *      *    * ***  ** *  **     * *  **   *** * *    *** *  *

vSagC2  ATA.GATCCCTCCCTTTTCGTG.AAAGACTCGCCAGAGCTA
C2A     AAAGGAAAATNTCCTNTAAGAGCAAAGAGCCACCTGGCCCT
        * * **      *** *  * * *****  * ** *  *

Matches         94      Mismatches      87      Unmatched       15

Length          196     Matches/Length 48.0%
```

**C2B locus**

```
vSagC2  CAAAAT.AGGAGACAGGTGGTGGCAACCAGGGACTTATAGGGGACCTTACATCTACAGACCAACAGATGCCCCCTT
C2B     AAGTTTCATCAGACAGGTGG.GGCAACCAGGGATTTCTA..TTAAGTGANAACAAAATACAGCGTTAATAAGTGAC
         *   * *  ********** ************ ** **    *  * * * * * * **      *

vSagC2  ACCATATACAGGAAGATATGACTTAAATTGGGATAGGTGGGTTACAGTCAATGGCTATAAAGTGTTATATAGATC
C2B     TCCAAAACCNCCAAAACATTATGCNAAGTTAAANAAGTAA.TATATGAAAAGTTACATATGGTATGATCCCNTTA
         *** *  *   ** * ** *    ** *   * * **  *    *  **     ***  ** * **     *

vSagC2  CCTCCCTTTTCGTGAAAGACTCGCCAGAGCTA
C2B     TATAAANATCTAAAATAAANACTCCAGTNGGN
          *     *     * * *  * ****

Matches         80      Mismatches      98      Unmatched       5

Length          183     Matches/Length 43.7%
```

**Figure 4.9 DNA sequence homology between vSagC2 and human subclones hybridising to this probe.** 'Microgenie' computer software was used to align the nucleotide sequences of subcloned λ13a (C2A locus) and λ37a (C2B locus) against the sequence of murine vSag (MMTV 3'LTR (228)). Asterisks indicate matching nucleotides. The similarity index is the percentage of matching nucleotides over the length of the sequence.

### 4.3.6 Restriction mapping and sequencing of HRC2 hybridising DNA subclone.

The restriction map of the subcloned λ DNA fragment to which the HRC2 probe hybridised is shown in Figure 4.10.

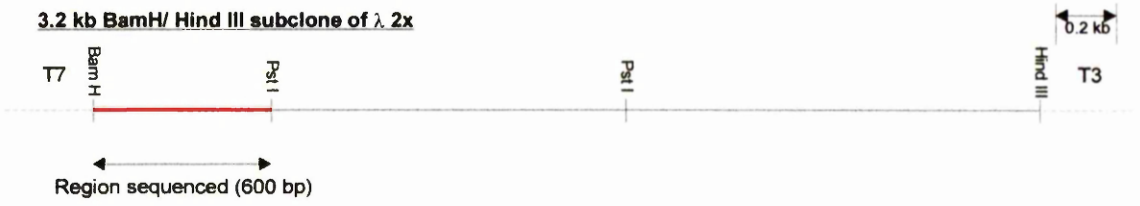



**Figure 4.10 Restriction maps of the HRC2 hybridising DNA subclone.** The restriction sites within the DNA subclone were positioned by analysis of single digests with *Apa*I, *Bst*XI, *Eco*RV, *Hinc*II, *Kpn*I, *Pst*I, *Sal*I, *Sma*I, *Xho*I, *Sac*I and *Sac*II. The region of the subclone to which the HRC2 probe hybridised is highlighted in red and the region sequenced indicated.

The restriction map for λ2x shows a 0.6 kb region to which the HRC2 probe hybridised, flanked by *Bam*HI and *Pst*I restriction sites. The subcloned 3.2 kb *Bam*HI/*Hind*III fragment of λ2x was digested with *Pst*I and *Xho*I to remove the 2.6 kb fragment downstream from this region and the vector containing the hybridising 0.6 kb fragment re-ligated. Transformed XL-1 cells with this new subclone were then isolated and purified prior to sequencing with both forward and reverse universal primers.

The DNA sequence of the subclone hybridising to the HRC2 probe is shown in figure 4.11. The length of the sequenced region was 685 nucleotides. The reverse (lower) strand is shown as homology to the HRC2 probe was identified in this orientation

| loci | probe | phage | length |
|---|---|---|---|
| HRC2 | HRC2 | λ2x | 685 bp |

Reverse strand
*XhoI/PstI (blunt)*

```
  CTCGGGCC ACACTCCCTC CAGAGGCTTT GGGGAGAATC TGTTCCCTGC CTCGCCCAGC 58

TCCTGGGGCT GCTGGCGTTC CTTGGCTCAC GGGCGCTTCC CTCCCATCTC TGCCTCCCTG 118

GTCACATCGC CGCCTCCTCT TCTGGGGTGT GTGTCTGGGT CAAACCTCCC TCTGCCTCCT 178

CCTCATAAGG ANATTTATCC AATTTAGCAC CCACCTAGAT AATCCAGGTA TAATCTTCTC 238

CTCTCAAAAT CATCCATCAA ATCAAATCTT CAAAGACACC TTTTCCAAGG TGAGGCAACC 298

TGTACACCTT TGGCAGACCA TCATTCCACA GCACGCCCAG AAACAGGGGA GCCAGGTGGT 358

GGAGGGTGGG ATCTGCCTGC CGAGAACGAG GAAGGGCATC GGTGACAGGT AACAACCTGA 418

GGCACTGGAG AGGGAGAGCA CGCGTGATGT TCCTCAGAAA CGAGTCGCAT TTCCGCATAG 478

CGCCTCCTTC AGACCGTGAG AGATCGCCAG AGGGCGTGGG CTTTCCGTTC TGTTCTCTGA 538

ACCTACAGTT AAGCCCAAGC CAGGGGCCGC TTTTCCCTAA TGAACATGCC TTTCTTCGGT 598

GGACCTGACC CTGTTACAGG AAAGCCGCGC TTTGATGAAA CCAGTTCTTT AAAATGAAAC 658
                         BamH
ATGGTATTAA AAGCCATAAC AGGATCC 685
```

**Figure 4.11 Nucleotide sequences of the representative HRC2 hybridising subclone.** Plasmid containing the recombinant that hybridised to the HRC2 probe was sequenced using dye terminator cycle sequencing (Perkin Elmer) with AmpliTaq® DNA Polymerase, FS, using universal forward and reverse primers. The positions of restriction endonuclease sites are indicated. Where the sequence data gave an equivocal result for a specific nucleotide, or where there was a discrepancy between the forward and reverse sequencing, the letter N was used. The number on the right indicates the position of the last nucleotide in the line.

### 4.3.7 DNA sequence homology to the murine vSagC2 and HRC2 probes.

The DNA sequence of the subclone of λ2x is shown in figure 4.12. This clone was isolated with the probe HRC2, the PCR product amplified from human DNA with primers to the C2 region of vSag. The figure compares the cloned sequence with both the PCR product used as the probe and the second conserved region of murine vSag. 100% homology was found between the cloned sequence and the probe for the region

between the PCR primers (132 nucleotides). The PCR primer binding regions were similar but not identical to the primers. The 5' PCR binding site had 20/25 (80%) matching nucleotides with 2 mismatches and 3 unmatched nucleotides. The 3' PCR binding site had 17/22 (77%) matched nucleotides, with 4 mismatched and one unmatched. Over the length of the region binding to the HRC2 probe, the homology to the murine sequence vSagC2 was 47.7%, with localised regions of greater homology at each end of the probe at the PCR primer binding sites.

```
vSagC2  CA.AAATA..GGAGACAGGTGGTGGCAACCAGGGACTTATAGGGGACCTTAC.ATCTACAGACCAACAGATGCCCCCT.
Probe   CA.AAATA..GGAGACAGGTGGTGGAGGGTGGGATCTGCCTGCCGAGA..ACGAGGAAGGG..CATCGGTGACAGGTAA
2x      CAGAAACAGGGGAGCCAGGTGGTGGAGGGTGGGATCTGCCTGCCGAGA..ACGAGGAAGGG..CATCGGTGACAGGTAA
        ** *** *  **** **********      **  **    *  **    ** *   *  *  ** * *   *

vSagC2  .TACCATATACA..GGA.AGATATGACTTAAATTGGGATAGGTGGGTTACAGTCAA.TGGC.TATAAAGTGTTATATAG
Probe   CAACCTGAGGCACTGGAGAGGGA.GAGC..ACGCGTGAT.GTTC.CTCAGAAACGAGTCGCATTTCC.GCATAGCGCC.
2x      CAACCTGAGGCACTGGAGAGGGA.GAGC..ACGCGTGAT.GTTC.CTCAGAAACGAGTCGCATTTCC.GCATAGCGCC.
          ***  *  **  *** **  * **    *   * *** * *   * * *  * * * ** * *   *  *

vSagC2  ATCCCTCCCTTTTCGTGAAAGACTCGCCAGAGCTA
Probe   .TCCTTCAGAC..CGTGAAAGACTCGCCAGAGCTA
2x      .TCCTTCAGAC..CGTGAGAGA.TCGCCAGAGGGC
         *** **       ***** *** *********

Matches         92      Mismatches      75      Unmatched       26

Length          193     Matches/Length 47.7%
```

**Figure 4.12 Sequence homology between the λ2x subclone, probe (HRC2) and the second conserved region of vSag (vSagC2).** The HRC2 probe was amplified from human genomic DNA by PCR using primers to the second conserved region of vSag. The PCR primer sequences are highlighted blue. Nucleotide homology between the sequence derived from the λ2x and vSagC2 is indicated by asterisk.

### 4.3.8 BLAST analysis of human DNA sequences hybridising to vSag

The nucleotide sequences of the six human sub-clones isolated using vSag derived probes were subjected to BLAST analysis, thereby determining whether they were homologous to any known human genes. This analysis was also useful in assessing how significant the regions of homology to vSag were likely to be. These search programmes generate a list of matching database sequences together with an alignment score and Expect (E) valve. The Expect value (E) is a parameter that

describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences. In the new versions of the BLAST programmes, the E value is used instead of the P value (probability) to report the significance of matches. For example, an E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see 1 match with a similar score simply by chance.

The nucleotide sequences shown in figures 4.4, 4.8 and 4.11 were submitted as query sequences via the GenBank World Wide Web server (http;/www.ncbi.nih.gov) and were matched to sequences in the NCBI non-redundant nucleotide database (see 4.2.4). Contaminating human Alu subfamily sequences were obtained with these searches for C1B, C1C, C2A and C2B when the sequences were submitted in unedited form. In order to avoid this contamination, the regions of the nucleotide sequences representing the contaminating Alu subfamily sequences were excluded prior to submission. The output derived from these searches is shown in figure 4.13 (top twenty matches). The C1A, C1C and C2A sequences gave almost exact matches with previously identified human DNA sequences, each the result of large sequencing projects and did not represent known genes. The C1A sequence was 98% identical to a DNA sequence mapped to chromosome 10, giving an E value of 2e-27. The C1C sequence was 98% identical to a DNA sequence mapped to chromosome 6 (E value 1e-36), and the C2A sequence shared 96% homology with a DNA sequence mapped to chromosome 9, giving an E value of 6e-86. Relatively weak alignments to murine vSag genes were identified for C1A (E value 0.084) and C1B (E value 0.36), but not for the other sequences. There were no significant alignments to any known retroviral genes.

As functionally related genes can share relatively short regions of homology at the nucleotide level, where the homologous sequences encode critical regions of the protein, the DNA database searches were repeated restricting the query sequences to the short regions with greatest homology to vSag. Nucleotide sequences ranging from 20-50 bp in length were submitted for BLAST analysis using the NCBI non-redundant DNA database. Significant matches to vSag sequences were obtained for C1A and C1B (E values 0.006 and 0.05 respectively, data not shown). A weak match was identified between vSag and C2A (E value 2.9). The remaining sequences (C1C, C2B and HRC2) failed to matched with vSag sequences in the DNA database.

Query sequence C1A

| Sequences producing significant alignments: | Score (bits) | E Value |
|---|---|---|
| gb\|AC006101.3\|AC006101 citb_338_f_24, complete sequence (Ho... | 125 | 2e-27 |
| emb\|AL357095.2\|CNS05TDU Human chromosome 14 DNA sequence **... | 42 | 0.021 |
| ref\|NC_001503.1\| Mouse mammary tumor virus, complete genome | 40 | 0.084 |
| gb\|AF263910.1\|AF263910 Mus musculus endogenous mouse mammar... | 40 | 0.084 |
| gb\|AF009658.1\|AF009658 Mus musculus host/virus junction fra... | 40 | 0.084 |
| gb\|AF228551.1\|AF228551 Exogenous mouse mammary tumor virus ... | 40 | 0.084 |
| gb\|AF228550.1\|AF228550 Endogenous mouse mammary tumor virus... | 40 | 0.084 |
| gb\|AF119342.1\|MTV3LTR2 Mus musculus mammary tumor virus-3 s... | 40 | 0.084 |
| gb\|AF119341.1\|MTV3LTR1 Mus musculus mammary tumor virus-3 5... | 40 | 0.084 |
| gb\|AF136900.1\|AF136900 Mouse mammary tumor virus strain RII... | 40 | 0.084 |
| gb\|AF033807.1\|AF033807 Mouse mammary tumor virus complete p... | 40 | 0.084 |
| gb\|U71271.1\|MMU71271 Mouse mammary tumor virus superantigen... | 40 | 0.084 |
| gb\|U13861.1\|XXU13861 pMSG-CAT cloning vector, complete sequ... | 40 | 0.084 |
| gb\|U13860.1\|XXU13860 pMSG cloning vector, complete sequence | 40 | 0.084 |
| gb\|AF043690.1\|AF043690 Mus musculus Mouse mammary tumor vir... | 40 | 0.084 |
| gb\|AF043689.1\|AF043689 Mus musculus Mouse mammary tumor vir... | 40 | 0.084 |
| gb\|AF043688.1\|AF043688 Mus musculus Mouse mammary tumor vir... | 40 | 0.084 |
| gb\|U02432.1\|XXU02432 Cloning vector pMAMneo, complete sequence | 40 | 0.084 |
| gb\|U02431.1\|XXU02431 Cloning vector pMAMneo-CAT, complete s... | 40 | 0.084 |
| gb\|U02430.1\|XXU02430 Cloning vector pMAMneoBlue, complete s... | 40 | 0.084 |

## Query sequence C1B

| Sequences producing significant alignments: | Score (bits) | E Value |
|---|---|---|
| gb\|AF196968.1\|AF196968 Homo sapiens PAC 704H1777 chromosome... | 38 | 0.36 |
| gb\|AC002109.1\|AC002109 Genomic sequence from Mouse 9, compl... | 38 | 0.36 |
| gb\|M11024.1\|MUSERMMTR Mouse endogenous mammary tumor virus ... | 38 | 0.36 |
| gb\|AC021049.12\|AC021049 Homo sapiens 12p12-21.8-27.2 BAC RP... | 36 | 1.4 |
| emb\|Z85996.1\|HS431A14 Human DNA sequence from PAC 431A14 on... | 36 | 1.4 |
| gb\|M36801.1\|HUMHXMA05 Human hemopexin gene, exon 7 | 36 | 1.4 |
| gb\|AC019014.1\|AC019014 Homo sapiens clone RP11-221G19 from ... | 34 | 5.6 |
| gb\|AC009229.3\|AC009229 Homo sapiens clone RP11-314C9, compl... | 34 | 5.6 |
| gb\|AC000064.1\|HSAC000064 Human BAC clone RG083M05 from 7q21... | 34 | 5.6 |
| gb\|AC003010.1\|HUAC003010 Homo sapiens Chromosome 16 BAC clo... | 34 | 5.6 |
| gb\|AC007566.1\|AC007566 Homo sapiens clone RG010G05, complet... | 34 | 5.6 |
| emb\|AL031627.1\|CEY102A5C Caenorhabditis elegans cosmid Y102... | 34 | 5.6 |
| emb\|AL133398.2\|AL133398 Human DNA sequence from clone RP1-2... | 34 | 5.6 |
| emb\|Z82058.1\|CET27C5 Caenorhabditis elegans cosmid T27C5, c... | 34 | 5.6 |
| emb\|Z93241.11\|HS222E13 Human DNA sequence from clone RP1-22... | 34 | 5.6 |
| emb\|Z86090.10\|HS229A8 Human DNA sequence from clone 229A8 o... | 34 | 5.6 |
| emb\|AL022316.2\|HS126B4 Human DNA sequence from clone CTA-12... | 34 | 5.6 |
| emb\|AL035071.17\|HS1085F17 Human DNA sequence from clone 108... | 34 | 5.6 |
| emb\|AL117257.1\|HS213J1PB Human DNA sequence from clone 213J... | 34 | 5.6 |
| dbj\|AB009048.1\|AB009048 Arabidopsis thaliana genomic DNA, c... | 34 | 5.6 |

## Query sequence C1C

| Sequences producing significant alignments: | Score (bits) | E Value |
|---|---|---|
| emb\|AL357057.19\|AL357057 Human DNA sequence from clone RP11... | 157 | 1e-36 |
| emb\|X97191.1\|RNMAFAEX1 R.norvegicus MAFA gene, exon1 (join acros... | 50 | 2e-04 |
| dbj\|AB023310.1\|AB023310 Cyanidioschyzon merolae gene for FtsH2, ... | 48 | 6e-04 |
| gb\|AF119676.1\|AF119676 Mus musculus small GTP-binding protein RA... | 46 | 0.002 |
| gb\|AF055898.1\|AF055898 Zea mays alanine aminotransferase (alt) g... | 46 | 0.002 |
| gb\|AF030385.1\|AF030385 Zea mays nitrate-induced NOI protein gene... | 46 | 0.002 |
| gb\|AF240002.1\|AF240002 Mus musculus adenine nucleotide transloca... | 44 | 0.010 |
| gb\|AF043433.1\|AF043433 Anopheles gambiae putative pupal-specific... | 44 | 0.010 |
| gb\|AF135125.1\|AF135125 Mus musculus nuclear factor kappa B subun... | 44 | 0.010 |
| gb\|AF078900.1\|AF078900 Oryctolagus cuniculus hensin (DMBT1) gene... | 44 | 0.010 |
| gb\|AF105143.1\|AF105143 Brassica napus chromosome N3 disease resi... | 44 | 0.010 |
| gb\|AF078925.1\|AF078925 Homo sapiens P2X1 receptor gene, partial cds | 44 | 0.010 |
| emb\|AX027357.1\|AX027357 Sequence 5 from Patent WO0037488 | 44 | 0.010 |
| emb\|Y14838.1\|HSCHEMR23 Homo sapiens ChemR23 gene | 44 | 0.010 |
| dbj\|AB015670.1\|AB015670 Bacillus sp. genes for CDase, CGTase, MB... | 44 | 0.010 |
| emb\|AJ010735.1\|ATAJ10735 Arabidopsis thaliana gr1 gene | 44 | 0.010 |
| ref\|NC_001141.1\| Saccharomyces cerevisiae chromosome IX, complet... | 40 | 0.15 |
| emb\|Z46833.1\|SC8277 S.cerevisiae chromosome IX cosmid 8277 | 40 | 0.15 |
| dbj\|D10595.1\|YSCNPS1 Saccharomyces cerevisiae NPS1 gene for nucl... | 40 | 0.15 |
| gb\|M83755.1\|YSCSTH1A Saccharomyces cerevisiae STH1 gene, complet... | 40 | 0.15 |

## Query sequence C2A

| Sequences producing significant alignments: | Score (bits) | E Value |
|---|---|---|
| emb\|AL158828.14\|AL158828 Human DNA sequence from clone RP11-359J... | 321 | 6e-86 |
| gb\|AF227510.1\|AF227510 Homo sapiens chromosome 21 map 21q22.1 cl... | 40 | 0.34 |
| emb\|AL163248.2\|HS21C048 Homo sapiens chromosome 21 segment HS21C048 | 40 | 0.34 |
| emb\|AL161572.2\|ATCHRIV68 Arabidopsis thaliana DNA chromosome 4, ... | 40 | 0.34 |
| emb\|AL035524.1\|ATT13J8 Arabidopsis thaliana DNA chromosome 4, BA... | 40 | 0.34 |
| gb\|AE003465.1\|AE003465 Drosophila melanogaster genomic scaffold ... | 38 | 1.4 |
| gb\|AC006346.1\|AC006346 Homo sapiens PAC clone RP5-820G22 from 7,... | 38 | 1.4 |
| gb\|AC068130.3\|AC068130 Homo sapiens clone RP11-640F22, complete ... | 38 | 1.4 |
| gb\|AC020908.6\|AC020908 Homo sapiens chromosome 19 clone CTD-2528... | 38 | 1.4 |
| ref\|NM_008741.1\| Mus musculus neuron specific gene family member... | 38 | 1.4 |
| gb\|AC004096.1\|AC004096 Mouse Cosmid ma66a100 from 14D1-D2, compl... | 38 | 1.4 |
| gb\|AC003119.1\|HUAC003119 Human Chromosome 16 BAC clone CIT987SK-... | 38 | 1.4 |
| gb\|AC007055.3\|AC007055 Homo sapiens chromosome 14 clone BAC 201F... | 38 | 1.4 |
| emb\|AL121983.13\|HSA160H22 Human DNA sequence from clone RP11-160... | 38 | 1.4 |
| gb\|AC003995.1\|AC003995 Mouse Cosmid ma66a094 from 14D1-D2 (T-Cel... | 38 | 1.4 |
| emb\|AL031660.16\|HS387E22 Human DNA sequence from clone RP3-387E2... | 38 | 1.4 |
| emb\|AL031286.1\|HS681J21 Human DNA sequence from clone 681J21 on ... | 38 | 1.4 |
| gb\|AC000373.1\|HSAC000373 Human cosmid g1862d218, complete sequen... | 38 | 1.4 |
| emb\|AL116514.1\|CNS01D8Q Botrytis cinerea strain T4 cDNA library ... | 38 | 1.4 |
| emb\|AL116450.1\|CNS01D6Y Botrytis cinerea strain T4 cDNA library ... | 38 | 1.4 |

## Query sequence C2B

| Sequences producing significant alignments: | Score (bits) | E Value |
|---|---|---|
| emb\|AL354773.8\|AL354773 Human DNA sequence from clone RP5-855L24... | 40 | 0.31 |
| gb\|AF131215.1\|AF131215 Homo sapiens chromosome 8 map 8p23-p22 cl... | 38 | 1.2 |
| gb\|AC016396.5\|AC016396 Homo sapiens clone RP11-179B15, complete ... | 38 | 1.2 |
| gb\|AE000058.1\|MPAE000058 Mycoplasma pneumoniae section 58 of 63 ... | 38 | 1.2 |
| emb\|AL033392.5\|HS403M6 Human DNA sequence from clone 403M6 on ch... | 38 | 1.2 |
| gb\|AC010163.7\|AC010163 Homo sapiens chromosome 10 clone RP11-90J... | 36 | 4.9 |
| gb\|AC008046.4\|AC008046 Arabidopsis thaliana chromosome I BAC F5A... | 36 | 4.9 |
| gb\|AC016925.15\|AC016925 Homo sapiens X BAC RP11-434J4 (Roswell P... | 36 | 4.9 |
| gb\|AC035150.1\|AC035150 Homo sapiens chromosome 19, BAC CIT978SKB... | 36 | 4.9 |
| gb\|AC007126.6\|AC007126 Homo sapiens chromosome 4 clone C0496D08,... | 36 | 4.9 |
| gb\|AE001811.1\|AE001811 Thermotoga maritima section 123 of 136 of... | 36 | 4.9 |
| gb\|AC006213.1\|AC006213 Homo sapiens, clone hRPK.15_A_1, complete... | 36 | 4.9 |
| emb\|AL159152.11\|AL159152 Human DNA sequence from clone RP11-165N... | 36 | 4.9 |
| gb\|AC004764.1\|AC004764 Homo sapiens chromosome 5, P1 clone 255g5... | 36 | 4.9 |
| emb\|AL121781.38\|HSJ1164C1 Human DNA sequence from clone RP5-1164... | 36 | 4.9 |
| emb\|AL049778.3\|CNS00004 Human chromosome 14 DNA sequence *** IN ... | 36 | 4.9 |
| emb\|AL354592.1\|LMFL8138 Leishmania major Friedlin cosmid L8138, ... | 36 | 4.9 |
| emb\|AL080274.21\|HS1012F16 Human DNA sequence from clone RP5-1012... | 36 | 4.9 |
| emb\|AL031965.7\|HS355N11 Human DNA sequence from clone 355N11 on ... | 36 | 4.9 |
| gb\|U71155.1\|TMU71155 Thermotoga maritima DNA mismatch repair pro... | 36 | 4.9 |

## Query sequence HRC2

| Sequences producing significant alignments: | Score (bits) | E Value |
|---|---|---|
| emb\|Y13901.1\|HSFGFR4G Homo sapiens FGFR-4 gene | 44 | 0.020 |
| gb\|AC004900.2\|AC004900 Homo sapiens PAC clone RP4-816G1 from 14q... | 36 | 4.8 |
| gb\|AC007380.3\|AC007380 Homo sapiens BAC clone RP11-150O13 from 2... | 36 | 4.8 |
| gb\|AE001274.1\|AE001274 Leishmania major chromosome 1, complete s... | 36 | 4.8 |
| emb\|AL359204.10\|AL359204 Human DNA sequence from clone RP11-147H... | 36 | 4.8 |
| emb\|AL138716.6\|AL138716 Human DNA sequence from clone CTA-KB521C... | 36 | 4.8 |
| emb\|AL137881.12\|AL137881 Human DNA sequence from clone RP11-40A8... | 36 | 4.8 |
| emb\|Z82245.1\|HS799F10 Human DNA sequence from clone CTA-799F10 o... | 36 | 4.8 |
| gb\|M57766.1\|MACIN11 M.fascicularis interleukin 11 mRNA, complete... | 36 | 4.8 |
| gb\|AC011811.42\|AC011811 Homo sapiens chromosome unknown clone b2... | 34 | 19 |
| ref\|NC_001726.1\| Carrot mottle mimic virus, complete genome | 34 | 19 |
| gb\|AC011555.5\|AC011555 Homo sapiens chromosome 19 clone LLNLR-28... | 34 | 19 |
| ref\|NM_016151.1\| Homo sapiens thousand and one amino acid protei... | 34 | 19 |
| gb\|AC009145.4\|AC009145 Homo sapiens chromosome 16 clone RP11-556... | 34 | 19 |
| gb\|AC007056.4\|AC007056 Homo sapiens clone from human chromosome ... | 34 | 19 |
| gb\|AC002523.1\|AC002523 Homo sapiens Xq28 BACs 360 F12, GSHB-555C... | 34 | 19 |
| gb\|AC000057.1\|HSAC000057 Human BAC clone RG067M09 from 7q21-7q22... | 34 | 19 |
| gb\|AF061943.1\|AF061943 Homo sapiens prostate derived STE20-like ... | 34 | 19 |
| gb\|AC005588.1\|AC005588 Homo sapiens PAC clone RP5-1161G23 from 7... | 34 | 19 |
| gb\|AC005034.1\|AC005034 Homo sapiens BAC clone RP11-342K6 from 2,... | 34 | 19 |

**Figure 4.13 BLAST search of a non-redundant nucleotide database with DNA sequences isolated from a human genomic library with vSag-derived probes.** The Basic Local Alignment Search Tool (BLAST) software provided on the National Center for Biological Information (NCBI) web site was used to search a non-redundant nucleotide database (658,628 sequences, 2,198,227,069 total letters, search date 26/9/2000) for DNA sequences homologous to the subclones λ26a (locus C1A), λ38a (locus C1B), λ41b (locus C1C), λ13a (locus C2A), λ37a (locus C2B) and λ2x (locus HRC2). Matches to murine vSag sequences are highlighted in red.

### 4.3.9 Open reading frame analysis of predicted amino acid sequences.

Potential open reading frames within the six sequenced subclones were identified using the NCBI "ORF Finder" programme. This tool identifies all possible open reading frames in a DNA sequence by locating the standard and alternative stop and start codons. Open reading frames with potential encoding regions ranging between 43-156 amino acids, which contained limited regions of homology with vSag, were identified for the sequences C1A, C1B, C1C, C2A and HRC2 (figure 4.14). No open reading frames were found within C2B.

```
C1A open reading frame (Length: 43 aa)

    316 atgttaatgccnttatttgcagctctgctttgcctggggcctctc
        M  L  M  P  L  F  A  A  L  L  C  L  G  P  L
    361 aaaaacactgcttcatttaaatgtcatcagaactccacacctccc
        K  N  T  A  S  F  K  C  H  Q  N  S  T  P  P
    406 agaaagccctgggttaatttcatccattcctttgtttggtga 447
        R  K  P  W  V  N  F  I  H  S  F  V  W  *
```

```
C1B open reading frame (Length: 156 aa)

     98 atgcacattgacactgtaagcagtggaataaattggcctgttttc
        M  H  I  D  T  V  S  S  G  I  N  W  P  V  F
    143 attgcagaaaaatcagcatattctgaaatgatagaagctccccgt
        I  A  E  K  S  A  Y  S  E  M  I  E  A  P  R
    188 caattagggaaattaacatccctctctgtggttggtgctattgct
        Q  L  G  K  L  T  S  L  S  V  V  G  A  I  A
    233 gcacttcatttgcctgacccaaagcacagcagtttctttttcatg
        A  L  H  L  P  D  P  K  H  S  S  F  F  F  M
    278 agccgttccagttggaaagtaaataggatttttggggtcttcttt
        S  R  S  S  W  K  V  N  R  I  F  G  V  F  F
    323 ctgtgggtcctgctccagactgggctctcaattctgcctgtggtg
        L  W  V  L  L  Q  T  G  L  S  I  L  P  V  V
    368 tcctcacagagtgagatgggaggcagatgcacattggggaaagtt
        S  S  Q  S  E  M  G  G  R  C  T  L  G  K  V
    413 gcggttgtgctatttggagcaagtaataatagacatagtaaagca
        A  V  V  L  F  G  A  S  N  N  R  H  S  K  A
    458 caaattaagaattgttcttgcacaacatgcccagctgtctctagt
        Q  I  K  N  C  S  C  T  T  C  P  A  V  S  S
    503 gaggaacagcaggcgagtagatcagatctactggactggatcaat
        E  E  Q  Q  A  S  R  S  D  L  L  D  W  I  N
    548 aagtcccattggttaaaatga 568
        K  S  H  W  L  K  *
```

C1C open reading frame (Length: 64 aa)

```
  11 ctgtcgacgcggccgcgtaatacgactcactatagggcgaagaat
     L  S  T  R  P  R  N  T  T  H  Y  R  A  K  N
  56 tcggatcccacattaagtcaggtgctacttggggaagttgcttcg
     S  D  P  T  L  S  Q  V  L  L  G  E  V  A  S
 101 ttttcttgtgctctaagatctctcacccacacgcctggttttggg
     F  S  C  A  L  R  S  L  T  H  T  P  G  F  G
 146 gagaaggataaggacggtgtttctaaacatgctgtgtttcctatg
     E  K  D  K  D  G  V  S  K  H  A  V  F  P  M
 191 cagagaggcaggtga 205
     Q  R  G  R  *
```

(Note alternative initiation codon- ctg)

C2A open reading frame (Length: 79 aa)

```
 581 atgtctgcaagtggccggcagccgtgcaatgagggagacaggtgg
     M  S  A  S  G  R  Q  P  C  N  E  G  D  R  W
 626 tggcgcccaggtcctgccctaggtagaaaagtgatggaagatccc
     W  R  P  G  P  A  L  G  R  K  V  M  E  D  P
 671 tgcttgatgctgaggccccagccgcgcccgccctggaaagataaa
     C  L  M  L  R  P  Q  P  R  P  P  W  K  D  K
 716 agaggaactggcctatcctatcctgatgccgaagggaagggagag
     R  G  T  G  L  S  Y  P  D  A  E  G  K  G  E
 761 aaaggaaaatntcctntaagagcaaagagccacctggccctccta
     K  G  K  X  P  X  R  A  K  S  H  L  A  L  L
 806 cagtgtgtggtacc 819
     Q  C  V  V
```

HRC2 open reading frame (Length: 102 aa)

```
 301 ctgtacacctttggcagaccatcattccacagcacgcccagaaac
     L  Y  T  F  G  R  P  S  F  H  S  T  P  R  N
 346 aggggagccaggtggtggagggtgggatctgcctgccgagaacga
     R  G  A  R  W  W  R  V  G  S  A  C  R  E  R
 391 ggaagggcatcggtgacaggtaacaacctgaggcactggagaggg
     G  R  A  S  V  T  G  N  N  L  R  H  W  R  G
 436 agagcacgcgtgatgttcctcagaaacgagtcgcatttccgcata
     R  A  R  V  M  F  L  R  N  E  S  H  F  R  I
 481 gcgcctccttcagaccgtgagagatcgccagagggcgtgggcttt
     A  P  P  S  D  R  E  R  S  P  E  G  V  G  F
 526 ccgttctgttctctgaacctacagttaagcccaagccaggggccg
     P  F  C  S  L  N  L  Q  L  S  P  S  Q  G  P
 571 cttttccctaatgaacatgcctttcttcggtggacctga 609
     L  F  P  N  E  H  A  F  L  R  W  T  *
```

(Note alternative initiation codon- ctg)

**Figure 4.14 Potential open reading frames predicted form the DNA sequences isolated from a human genomic library with vSag-derived probes.** Open reading frames were identified with the NCBI programme "ORF finder" with standard (atg-C1A, C1B, C2A) and alternative (ctg-C1C, HRC2) start codons. All sequences terminate with a tga stop codon except the subclone C1A which potentially extends beyond the sequenced region.

The predicted amino acid sequences from these potential encoding regions were submitted as query sequences for analysis using the blastp programme with the NCBI non-redundant protein database (234). Protein database matches with vSag amino acid sequences were found for the subclone C2A (figure 4.15), resulting from a short region of homology over 12 amino acids (figure 4.16). Weak alignments were also identified for the C1B and C1C subclones (figure 4.16). No other regions of significant homology were found between the human sequences and murine vSag.

Query sequence C2A (predicted amino acid sequence)

| Sequences producing significant alignments: | Score (bits) | E Value |
|---|---|---|
| ref\|NP_055611.1\| KIAA0748 gene product >gi\|3882217\|dbj\|BAA3... | 34 | 0.14 |
| emb\|CAA46436.1\| (X65339) ORF of MTV-7; within long terminal... | 29 | 4.2 |
| pir\|\|S23059 superantigen Mtv43 - mouse mammary tumor virus ... | 29 | 4.2 |
| sp\|P10261\|PR7R_MMTVG PROTEIN PR73 (3'ENDOGENOUS) >gi\|61628\|... | 29 | 4.2 |
| emb\|CAA44757.1\| (X63025) mammary tumour virus 3, superantig... | 29 | 4.2 |
| sp\|P03319\|PR73_MMTVC PROTEIN PR73 | 29 | 4.2 |
| emb\|CAA44755.1\| (X63024) mammary tumour virus 1, superantig... | 29 | 4.2 |
| prf\|\|0803245A L-long term repeat (Mus musculus) | 29 | 4.2 |
| gb\|AAD24771.1\| (AF119342) superantigen 3 (Mus musculus) | 29 | 4.2 |
| dbj\|BAA07620.1\| (D38639) retroviridae protein (Mouse mammar... | 29 | 4.2 |
| gb\|AAC16281.1\| (AF043689) vSAG protein (Mus musculus) | 29 | 4.2 |
| gb\|AAC57063.1\| (U71271) superantigen (Mouse mammary tumor v... | 29 | 4.2 |
| gb\|AAB21468.1\| exoMTV vSAG=viral superantigen type II membr... | 29 | 4.2 |
| sp\|P03321\|PR7L_MMTVG PROTEIN PR73 (5'ENDOGENOUS) >gi\|74536\|... | 29 | 4.2 |
| ref\|NP_056884.1\| Pr48 >gi\|130841\|sp\|P10260\|PR73_MMTVB PROTE... | 29 | 4.2 |
| gb\|AAB33886.1\| MTV protein (mouse mammary tumor virus MMT, ... | 29 | 4.2 |
| gb\|AAB29186.1\| (S67365) vSAG23=T cell receptor V beta 7-spe... | 29 | 4.2 |
| gb\|AAD26464.1\|AF136899_1 (AF136899) superantigen (Mouse mam... | 29 | 4.2 |
| gb\|AAD26463.1\|AF136898_1 (AF136898) superantigen (Mouse mam... | 29 | 4.2 |
| dbj\|BAA05390.1\| (D26359) superantigen (Exogenous mouse mamm... | 29 | 4.2 |

**Figure 4.15 BLAST search of a non-redundant protein database with the predicted amino acid sequences of clones isolated from a human genomic library with vSag-derived probes.** The Basic Local Alignment Search Tool (BLAST) software (234) provided on the National Center for Biological Information (NCBI) web site was used to search a non-redundant protein database (536,932 sequences; 168,632,586 total letters: search date 26/9/2000) for sequences homologous to the predicted amino acid sequences derived from subclones λ26a (locus C1A), λ38a (locus C1B), λ41b (locus C1C), λ13a (locus C2A), λ37a (locus C2B) and λ2x (locus HRC2). Matches to murine vSag sequences are highlighted in red.

```
C1B  : 32 LGKVAV 37
       LG+VAV
vSag : 24 LGEVAV 29

Score = 13.0 bits (23), Expect = 323715
Identities = 5/6 (83%), Positives = 6/6 (99%)


C1C  : 24 LLGEVASFSCALRSLT 39
       LLGEVA    A R+LT
vSag : 23 LLGEVA--VRARRALT 36

Score = 14.8 bits (28), Expect = 96050
Identities = 10/16 (62%), Positives = 11/16 (68%), Gaps = 2/16 (12%)


C2A  : 4   GDRWWRPGPALG 15
       GDRWW+PG   G
vSag : 201 GDRWWQPGTYRG 212

Score = 29.2 bits (60), Expect = 4.2
Identities = 8/12 (66%), Positives = 9/12 (74%)
```

**Figure 4.16 BLAST alignments between vSag and predicted amino acid sequences.** BLAST alignments were generated from searching a non-redundant protein database (234) with the predicted amino acid sequences of clones λ38a (locus C1B), λ41b (locus C1C) and λ13a (locus C2A). Significant alignment were not identified with λ26a (locus C1A), λ37a (locus C2B) and λ2x (locus HRC2).

Figure 4.17 shows the positions of each of the open reading frames against the 320 amino acid vSag gene product. (Note the C1B sequence contains four methionine residues at position 1, 25, 60 and 96, the $Met^{60}$ being taken as the start codon in the alignment shown in figure 4.17)

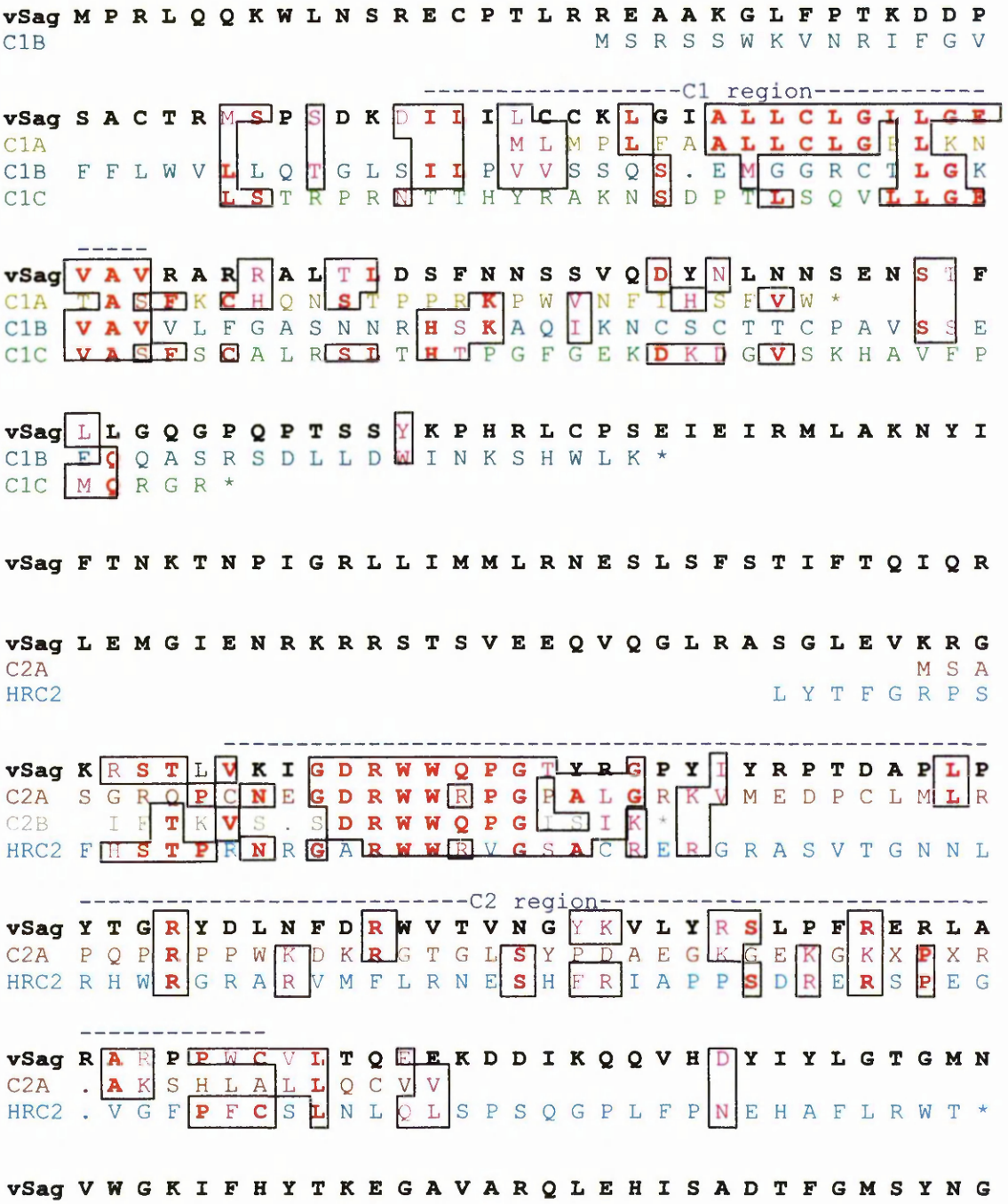



**Figure 4.17 Alignment of open reading frames of sequences C1A, C1B, C1C, C2A and HRC2 against the complete vSag gene product.** Identical and similar residues shared by more than one sequence are boxed, where identical residues are highlighted in red and similar residues highlighted in pink. A short region of sequence alignment is shown for C2B (18 AA), although this did not represent an open reading frame. The conserved regions of vSag (C1 and C2) are indicated, and represent the regions to which probes were targeted. Similar residues were members of the set; L=I=V=M, R=K=H, F=Y=W, S=T, E=Q, D=N.

## *4.4 Discussion*

Previous chapters have described the identification and isolation of the human genomic clones hybridising with greatest affinity to the murine vSag derived probes. The most direct way to determine the significance of these regions of DNA was to sequence them. This allowed the accurate assessment of homology to vSag and by identifying open reading frames, gave an indication as to whether the DNA sequences had the potential to encode proteins. As the sequencing reactions could give accurate data for lengths up to 500 bp, it was necessary to digest the original λ clones (with inserts up to 10 kb), identify the fragments hybridising to the probes and re-ligate these smaller fragments into another vector (KSM13-).

Sequence data was obtained from recombinant inserts ranging from 478 -952 bp. Accuracy of the sequence data was improved by sequencing forward and reverse strands of DNA, merging the complementary strands using computer software (Microgenie). For longer sequences (>800 bp) the overlap between forward and reverse strands was relatively short, resulting in potential sequence errors. To minimise such errors, the graphical output of the sequence data was analysed visually, and the letter "N" inserted where doubt existed. The number of undetermined nucleotides per sequence ranged from 1 to 15 (0.15%-1.67%), with the largest number being present in the longest uninterrupted sequence (952 bp *Pst*I subclone of λ37a).

The initial assessment of the sequence data was undertaken using computer software (Microgenie) which identifies regions of homology between two sequences. The DNA sequences of the subclones were compared with the complete murine vSag sequence. Homology to murine vSag was identified only within the region of the subclones binding to the probes. The degree of homology, expressed as a 'similarity

index' (number of matches divided by the length of the sequence) was relatively low. For the subclones isolated using the vSagC1 probe, the similarity indices over the length of the probe (91bp) ranged from 52.2-58.2%, while the similarity indices for the subclones isolated using the 182bp vSagC2 probe, were 43.7- 48%. For each subclone there was a short region of high homology, with 20/20 nucleotides matching (C1A), or 22/23 matching (C2B). Only sequence C1C had a longer region with 39/53 nucleotide matches.

For some of the subclones, the short regions of strong homology to the vSag probes had high C:G contents which would have increased their hybridisation melting temperature ($T_m$). This is a consequence of G-C interaction involving formation of three hydrogen bonds rather than the two involved in A-T interactions. The C:G content of the 91bp C1 probe was 56%. This increased to 65% over the 20/20 nucleotide stretch with homology to C1A. The C:G content of the 182 bp C2 probe was 45%, which increased to 65% over the region of the probe sharing 21/23 matching nucleotides with C2A. The C:G content of the short hybridising regions of the other subclones ranged from 50-58%.

The sequence data from the subclone hybridising to the HRC2 probe revealed the extent of homology that the PCR primers had with the genomic DNA from which the probe was originally amplified. The region to which the 22 nucleotide long 5' primer annealed had 20 matching nucleotides with 3 unmatched (gaps) and 2 mismatched nucleotides. The 3' primer (22 nucleotides long) annealed to a region with 17 matching nucleotides, with 1 unmatched and 4 mismatched nucleotides. These primer binding sites had the greatest homology to the vSag sequence. Over the length of the probe (176 bp) the similarity index was 47.7% which was similar to that of the C2A

and C2B sequences isolated directly with the probe to murine vSagC2. There were again no significant regions of homology out with the region binding to the probe.

While experiment is the best method of determining biological significance, mathematical analysis by sequence similarity search programmes can indicate which similarities are unlikely to have arisen by chance and can be considered statistically significant. The NCBI (National Center for Biotechnology Information, National Library of Medicine, National Institute of Health, Bethesda, USA, http:/www.ncbi.hih.gov) was used as this provided easy access via the internet to a range of up-to-date DNA and protein databases. When the complete sequences of the subclones were submitted for database searching, in four of the six sequences (C1B, C1C, C2A and C2B), human Alu subfamily sequences with high E (Expect) values were present throughout the output data. (The E value is number of hits one can "expect" to see just by chance when searching a database of a particular size). The Alu family (named due to the presence of the recognition sequence for the restriction endonuclease *Alu*I within the sequence) are short interspersed repetitive nuclear elements (SINEs) which are scattered throughout the genome. Their function is unknown. SINEs vary in length form 130-300 bp and are repeated some 7000 times. These repetitive sequence can confound database searches, as sequences with lower, but significant, alignment scores may not be identified. To get round this problem, it is recommended that database searching should be performed in phases, with a query first compared to a small database containing domains representative of large sequence families. Regions of the query that match one or more of these domains can then be masked (or deleted) prior to full scale searching, thereby eliminating most of the redundant output. By doing this, matches to vSag sequences in the NCBI (non-redundant) DNA database were identified for the sequences C1A (E value 0.084) and C1B (E value 0.36). Almost exact sequence matches were identified for three of the sequences,

C1A matching with a region of chromosome 10, C1C with a region of chromosome 6 and C2A with a region of chromosome 9. In each case the database sequences were from large sequencing projects, and did not represent known genes. Matches to vSag sequences were not identified in database searches with the other sequences.

The database searches with the sequence data from the 6 representative subclones indicated that they were only weakly related to vSag genes with no statistically significant matches being identified with vSag sequences. Functionally related genes can however share relatively short regions of homology at the nucleotide level, where the homologous sequences encode critical regions of the protein. When the DNA database searches were repeated restricting the query sequences to the short regions with greatest homology to vSag (range 20-50 nucleotides), significant matches to vSag sequences were obtained for C1A and C1B (E values 0.006 and 0.05 respectively).

The presence of open reading frames (ORFs) within a DNA sequence can give an indication to the location of protein-encoding regions (i.e. exons). The 'ORF finder' programme (NCBI) was used to identify standard and alternative initiation and termination codons and deduce amino acid sequences in the 6 subclones. Open reading frames containing short regions of homology to vSag in the predicted amino acid sequence were identified in 5 of the 6 subclones (C1A, C1B, C1C, C2A, HRC2). These open reading frames ranged in size from 43 to 156 amino acids. Alignment of these predicted amino acid sequences against that of vSag demonstrates localised areas of homology, shared by each of group sequences, C1A, C1B and C1C sharing homology in the first conserved region of vSag, and C2A and HRC2 sharing homology in the second conserved region of vSag. None of the open reading frames extended to include both first and second conserved regions.

The potential biological significance of the human vSag-related sequences described above will be discussed in chapter 7. We will first turn our attention to the presence and expression of vSag related sequences in the human autoimmune condition, Sjogren's syndrome.

# 5 CHAPTER FIVE

# Search for vSag related sequences in DNA and mRNA from patients with primary Sjogren's syndrome

## *5.1 Introduction*

The preceding chapters have described the search for endogenous retroviral superantigens in normal human genomic DNA. There are several pathological conditions in which such genes could play an important role. As described in chapter one, both superantigens and retroviruses have been implicated in many of the human autoimmune diseases, but to date, there have been no published studies looking for the presence of murine vSag related sequences in these conditions.

Sjogren's syndrome was chosen as a model human autoimmune condition for these studies. This disease is characterised by lymphocytic infiltration of lacrimal and salivary glands, causing destruction of the glands giving rise to the classical features of 'dry eyes' and 'dry mouth'. The condition can be secondary to other autoimmune conditions, such as rheumatoid arthritis or primary biliary cirrhosis, or can overlap with disorders which share similar systemic features (eg systemic lupus erythematosus (SLE), scleroderma or mixed connective tissue disease). Diagnostic criteria for the 'primary' condition (1°SS) are shown in Chapter 1 (section 1.4).

A number of clinical features are found in 1°SS, which may reflect an over active immune system in these patients. The characteristics of an overactive immune system could theoretically result from a response to a superantigen. This chapter describes

the search for potential retroviral superantigen sequences in 1°SS. Eight patients were selected who fulfilled the European diagnostic criteria for 1°SS and were anti-Ro antibody positive. DNA was extracted from peripheral blood samples of these patients and investigated for the presence of genomic vSag-related sequences. mRNA was extracted from minor salivary gland biopsies from five of these patients to look for evidence of gene expression. The PCR primers directed against the conserved regions of murine vSag, which had been used to identify potential vSag related sequences in normal human DNA in chapter two, were used to look for evidence of vSag related sequences unique to 1°SS patients. The murine vSag constant region probes (vSagC1 and vSagC2), which had been used to isolate vSag related sequences from the human placental DNA library (chapter 2), were hybridised with 1°SS peripheral blood DNA to look for evidence of other vSag sequences not identified in the human genomic library. Evidence of gene expression was sought by hybridising mRNA from minor salivary gland biopsies with the human vSag related sequences C1A, C1B, C1C, C2A, C2B and HRC2, shown in previous chapters to be present in normal genomic DNA.

## *5.2 Patients and Methods*

### 5.2.1 Patients

Eight patients, who were under regular review in a specialist clinic (Department of Rheumatology, Freeman Hospital), were recruited. Written consent was obtained for isolation of DNA for the genetic studies. Patients fulfilling the diagnostic criteria for rheumatoid arthritis, SLE or scleroderma were excluded from the study. Any patients with sarcoidosis, graft-versus-host disease, acquired immune deficiency syndrome or pre-existing lymphoma were also excluded. All were white Caucasian females with a median age at diagnosis of 55 years (range 17-65). The median disease duration was

10 years (range 6-20 years). They were asked to complete the 6 item symptom questionnaire described by Vitali *et al* (180, 181). A full clinical assessment was made looking for evidence of extraglandular manifestations (skin rashes, Raynaud's, lymphadenopathy, arthritis, lung involvement, serositis and neuropathy) and glandular manifestations including a Schirmer's test and unstimulated salivary flow rate. The Schirmer's test was considered abnormal if there was less than 5 mm wetting after 5 minutes as defined in the European criteria. The inability to produce more than 0.5 ml of saliva in 5 minutes was also considered abnormal. Blood was taken for extraction of DNA. Minor salivary gland biopsies were available in some patients (stored at –70°C) where they had been performed for diagnostic purposes. Other blood investigations included full blood count, ESR, urea and electrolytes, CRP, liver function tests, thyroid function tests, immunoglobulins, serum electrophoresis. An extended autoantibody profile was carried out by the Immunology Department at Royal Victoria Infirmary. ANA and salivary gland antibodies were assessed by immunofluorescence with titres >1/40 reported as positive. Antibodies to DNA, Ro, La, RNP and Sm, were assayed by a commercial ELISA (Sigma) while RF was determined by a particle agglutination assay (positive titre >1/40).

### 5.2.2 Extraction of DNA from peripheral blood of 1°SS patients.

5 ml of peripheral blood in EDTA was mixed with 15 ml SDW and centrifuged at 4,000 rpm (Beckman JS 13 rotor) for 15 min at 4°C. After draining off the supernatant, 10 ml of sucrose buffer containing 1.09 g sucrose, 10 mM Tris (pH 7.5) and 5 mM $MgCl_2$ was added, and vortexed thoroughly to resuspend the nuclei. This was re-centrifuged at 4,000 rpm (Beckman JS 13 rotor) for 10 minutes at 4°C, and the supernatant drained off thoroughly. The pellet was then resuspended in 0.5 ml of TE buffer and vortexed to obtain a uniform suspension of nuclei. 5 ml of extraction buffer containing

50 μg RNase, 10 mM Tris (pH 8), 50 mM EDTA and 0.5% SDS was added to the suspension of nuclei while vortexing continuously. The suspension was incubated for 3 hours at 50°C with 500 μg proteinase K and then extracted overnight with 5 ml of phenol at room temperature. After centrifuging at 10,000 rpm (Beckman JS 13 rotor) for 10 minutes at 4°C, the aqueous phase was transferred to a clean Corex tube to which 100 mM NaCl and 2 volumes of 100% ethanol was added, mixing well to precipitate the high molecular weight DNA. This was recovered by centrifuging at 10,000 rpm (Beckman JS 13 rotor) for 10 minutes at 4°C. After pouring off the ethanol, the DNA pellet was re-suspended in 75% ethanol and left overnight at 4°C, then centrifuged at 10,000 rpm (Beckman JS 13 rotor) for 10 minutes at 4°C, and the ethanol drained off again. The DNA pellet was dissolved in TE buffer, and after measuring the DNA concentration on a spectrophotometer, was diluted to give a final concentration of 500 μg/ml.

### 5.2.3 PCR amplification using vSag specific primers

DNA extracted from the peripheral blood of 1°SS patients was used as template DNA for polymerase chain reactions using the specific PCR primers to the first and second conserved regions of murine vSag (see chapter 2). PCR reactions were under the conditions describe in section 2.2.1, with 100 ng of patient DNA in each reaction. Control samples were DNA extracted from peripheral blood of a normal human blood donor and DNA from the GR mouse strain. Thermal cycler setting were as described in section 2.2.1, except that reactions were continued for 30 cycles.

PCR products were electrophoresed on 3% agarose minigels containing 0.5 μg/ml ethidium bromide.

### 5.2.4 Southern analysis of 1°SS DNA using vSag related probes.

DNA extracted from the peripheral blood of the 8 1°SS patients was digested with the restriction enzymes *Hind*III and *Xba*I. DNA extracted from peripheral blood of a normal blood donor was used as a control. Digestions were in 16 μl reaction mixtures, containing 5 μg of DNA and 100 μg/ml BSA , in a restriction buffer of the composition recommended by the enzyme manufacturer. The restriction enzymes were used at 4 fold excess and digestions carried out at 37°C for three hours. Digested DNA samples were then run on 1% agarose gels with *Hind*III digested λ DNA molecular weight markers, for approximately 450 volt-hours. After staining with ethidium bromide to determine the position of the molecular weight markers, the digested DNA was transferred to nylon membranes as described in chapter three.

Hybridisation probes, radiolabelled by nick translation, were prepared as described in chapter 2. Four probes were prepared; the murine probes vSagC1 and vSagC2, and the human PCR products, HRC1 and HRC2, amplified with primers to the first and second conserved region of vSag. Hybribisations were carried out under reduced stringency conditions (25% formamide) with the vSagC1 and vSagC2 probes, and under higher stringency condition (50% formamide) with the HRC1 and HRC2. Hybribisations were carried out with each probe at 40kBq/ml. In order to differentiate the predicted restriction fragments based on the known restriction maps (chapter 3), *Hind*III digested DNA was hybridised with the vSagC1 and HRC2 probes, while *Xba*I digested DNA hybridised with vSagC2 and HRC1 probes. (Note a restriction map was not available for HRC1, as λ clones had not been isolated from the genomic library with this probe).

### 5.2.5 Extraction of mRNA from minor salivary gland biopsies.

Messenger RNA was extracted from minor salivary gland biopsies in five of the eight 1°SS patients (clinical features shown in Table 5.2). Each lip biopsy showed characteristic histological features of 1°SS with a focus score ≥ 1 (focus defines as an agglomeration of at least 50 mononuclear cells; focus score defined as the number of foci per 4 $mm^2$ of glandualr tissue). There were between 4 and 6 minor salivary glands for each patient, which were stored at -70°C prior to extraction of mRNA. A salivary gland biopsy from an otherwise healthy patient with chronic sialadenitis secondary to parotid duct obstruction was used as a normal control.

**Isolation of mRNA with Oligo$(dT)_{25}$ magnetic beads**

Messenger RNA was isolated using magnetic beads covalently bound to an oligo dT sequence (Dynabeads, Dynal, Oslo, Norway). This method relies on the base pairing between the oligo dT sequence and the poly (A)+ residues of most messenger RNA. The oligo$(dT)_{25}$ beads are superparamagnetic polystyrene beads, 2.8 μm in diameter, with 25 nucleotide long chains of deoxy-thymidylate attached to the bead surface via a 5' linker group. By placing an Eppendorf tube containing the beads against a magnet, the beads and attached mRNA can be isolated.

For each isolation, 250 μl of the stock bead suspension in PBS was added to an RNase-free Eppendorf tube and placed against a magnet. After the suspension cleared (30 seconds), the supernatant was discarded, and the beads washed once with 0.2 mls lysing/binding solution containing 100mM Tris-HCl (pH 8), 500 mM LiCl, 10 mM EDTA (pH 8), 1% SDS and 5 mM dithiothreitol (DTT). For each patient, frozen tissue was transferred to a glass homogenizer containing 1 ml lysis/binding buffer and homogenised manually for 2 minutes. The viscosity was reduced by passing the lysate three times through a 21 gauge needle with a 1 ml syringe. The lysate was centrifuged

for 1 minute in an Eppendorf centrifuge to remove debris and the supernatant added to the prewashed oligo $dT_{25}$ magnetic beads, mixed well, and left to anneal for 5 minutes at room temperature. The tube was placed against the magnet for 5 minutes to isolated the annealed mRNA. The beads were washed three times in 1 ml washing buffer containing 10 mM Tris-HCl (pH 8), 0.15 M LiCl, 1 mM EDTA and 0.1% SDS, mixing thoroughly in the washing buffer and removing the supernatant completely between each washing step. The mRNA was eluted by incubating at 65°C for 2 minutes in15 μl of 2mM EDTA (pH 8). The tube was placed against the magnet, and the supernatant containing the mRNA transferred to new Rnase-free tubes. The concentration of mRNA was measured by spectrophotometry and adjusted to 100μg/ml, giving an approximatie yield of 3 μg mRNA per patient sample.

### 5.2.6 Northern analysis of 1°SS salivary gland mRNA using vSag related probes.

RNA samples were electrophoresed on a denaturing gel containing 1.2% type II agarose and 17% formaldehyde in PB buffer. Prior to loading, for each lane on the gel, 0.5 μg mRNA was denatured by incubating for 20 minutes at 65°C in a solution containing 17.9% formaldehyde, 0.05 mM EDTA, 0.05 mM DMSO and 2% bromophenol blue in PB buffer. 10 μg total RNA from the MCF-7 cell line was denatured in the same way to run as a molecular weight marker. Samples were run on the gel for approximately 450 volt-hours. After imaging the ethidium bromide stained gel to determine the position of the marker ribosomal RNA, the denatured mRNA was transferred to nylon membranes by Northern blotting, to allow hybridisation with the vSag related probes. Prior to transfer, the gel was washed in 20x SSC (0.3 M $Na_3$Citrate and 3 M NaCl) for 40 minutes. The nylon membranes, which had been soaked in 20 x SSC, were laid over the gel on transfer apparatus (described in chapter 3), and allowed to transfer for 24 hours. The nylon membranes were then

baked at 80°C for 2 hours and irradiated for 90 seconds on a transilluminator prior to hybridisation.

**Preparation of vSag related probes for hybridisation**

Probes were prepared from the vSag related subclones described in chapter five. In order to isolate the insert DNA for use as probes, the subclones were used as templates for PCR reactions, using T3 and T7 primers. The size of the product for each subclone is shown in Table 5.1. The original HRC2 probe was used, as this shared 95% homology with the subclone of λ 2x (see chapter 4).

| Sequence | Subclone | PCR product size |
|---|---|---|
| C1A | 26a *SacI* | 1200 bp |
| C1B | 38a *XbaI* / *BglII* | 600 bp |
| C1C | 41b *XbaI* / *SmaI* | 600 bp |
| C2A | 13a *SmaI* / *KpnI* | 800 bp |
| C2B | 37a *PstI* | 1000 bp |

**Table 5.1 Sizes of PCR products amplified from subclones for use as hybridisation probes.** T3 and T7 primers were used to amplify the DNA inserts from the subclones of λ26a (locus C1A), λ38a (locus C1B), λ41b (locus C1C), λ13a (locus C2A) and λ37a (locus C2B). These PCR products were purified and radiolabelled by nick translation as described in chapter two.

The plasmids were linearised by digesting with the restriction enzyme *ScaI* which does not cut within the multiple cloning site of KSM13-. The digested plasmids were run on a 0.8% agarose minigel, to estimate the yield and to ensure the restriction enzyme had not cut within the DNA insert. Two restriction fragment were present after digesting the subclone 37a *PstI*, indicating that a second restriction site for *ScaI* was present within the recombinant DNA. The subclone 37a was therefore not linearised when used as a template for PCR, but the denaturation temperature of the first four

cycles of the PCR reaction was increased to 96°C to account for this. *Sca*I restriction sites were not present within the other recombinants.

PCR reactions were with 2 ng plasmid DNA in 50 μl volumes under mineral oil, with 200 ng of each primer (T3 and T7), 0.2 mM dNTPs, 5 μg BSA, PCR buffer and Taq polymerase (1.25 units/reaction). The thermal cycler settings were 25 cycles of 94°C denaturation for 30 seconds, 55°C annealing for 30 seconds and 72°C extension for 60 seconds (with 4 cycles of 96°C denaturation temperature where non-linearised plasmid was used as template DNA). PCR products were run on 3% agarose minigels (1g agarose type 1, 2g Nusieve in 100mls TBE) containing 0.5μg/ml ethidium bromide.

Prior to use as hybridisation probes, the PCR products were purified and radiolabelled by nick translation as described in chapter two. Hybridisations were carried out in 50% formamide at 42°C for 3 days with each probe at 160 kB/ml. Filters were washed at 37°C (see chapter two) with 50% formamide and were finally dried and exposed against pre-flashed X-ray film.

## *5.3 Results*

### 5.3.1 Clinical features of 1°SS patients

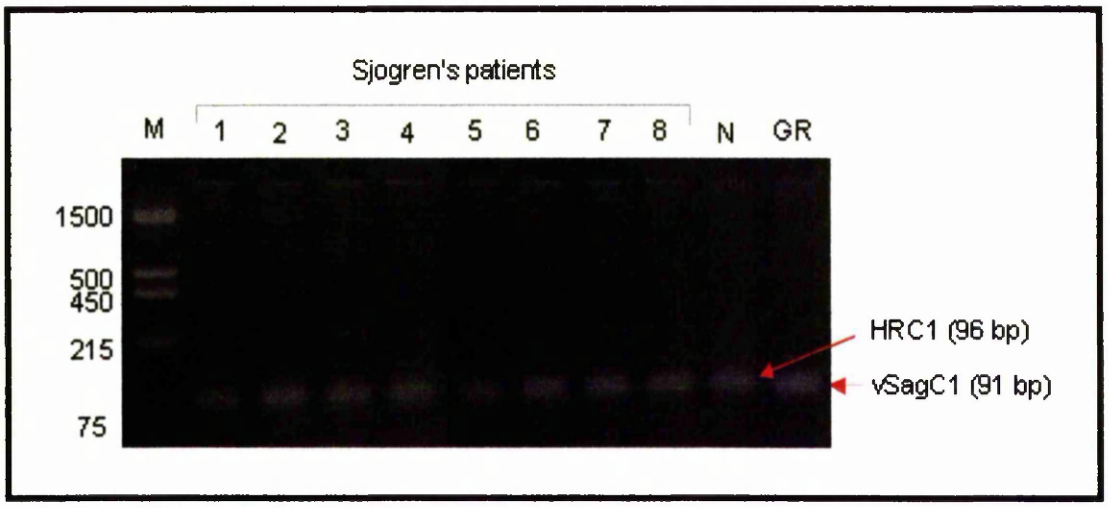
The clinical details of the eight patients are shown in Table 5.2. Each patient fulfilled 6/6 of the European criteria for the diagnosis of 1°SS. All were anti-Ro and anti-La antibody positive. The mean IgG level was 21.3±9.4 g/L (normal range 6-15g/L). Four patients had lung involvement (3/8 pleural effusions, 1/8 fibrosing alveolitis). Lymphadenopathy was present in 3/8 patients, parotid swelling in 4/8 and interstitial nephritis associatied with renal tubular acidosis in 1/8. Three patients were on oral corticosteroids but none of the patients were on other types of immunosuppressive medication.

| Patient | Age at diagnosis | Disease duration | Lung disease | Parotid/LN enlarged | Renal disease | Steroid dose | IgG (g/L) |
|---|---|---|---|---|---|---|---|
| 1 | 51 | 20 | P.Effusion | | | 7.5 mg | 23.5 |
| 2* | 58 | 6 | | Parotid | | | 17.9 |
| 3* | 65 | 11 | | Parotid/LN | | | 13.6 |
| 4* | 32 | 16 | P.Effusion | Parotid | | | 39 |
| 5 | 62 | 8 | FA | | | 7.5 mg | 13.4 |
| 6* | 17 | 6 | | Parotid/LN | | | 23.8 |
| 7* | 45 | 10 | P.Effusion | | RTA | 5 mg | 10.5 |
| 8 | 55 | 7 | | LN | | | 28.5 |

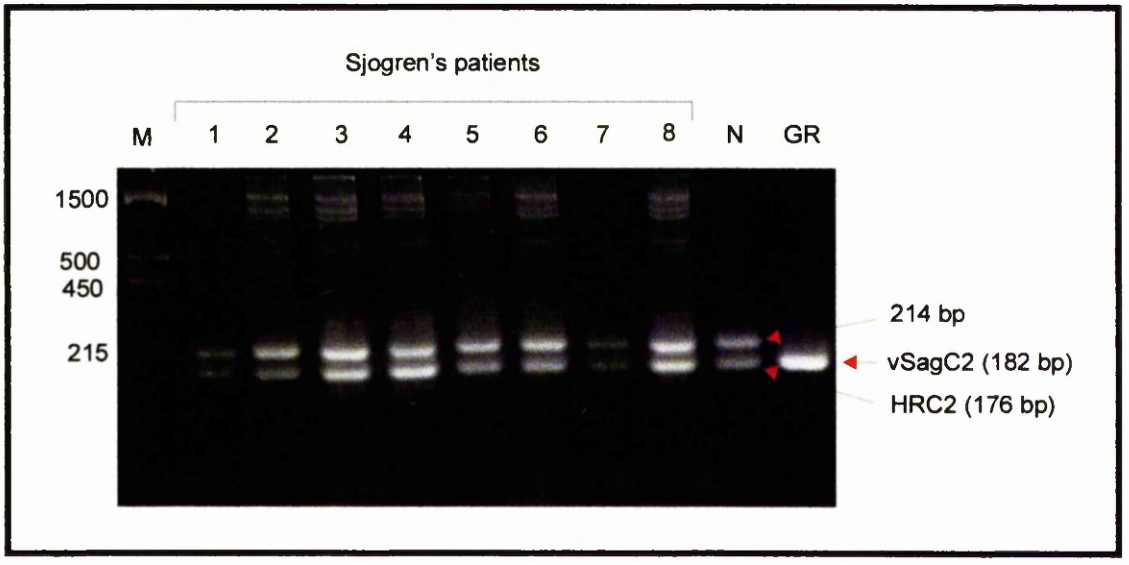
**Table 5.2 Clinical details of patients with 1°SS.** DNA was extracted from peripheral blood as described in Materials and Methods. All patients fulfilled 6/6 European criteria and were anti-Ro and anti-La antibody positive. P.Effusion= pleural effusion; FA= fibrosing alveolitis; LN= lymph node enlargement; RTA= renal tubular acidosis.
* RNA extracted from minor salivary glands in these patients

### 5.3.2 PCR amplification using vSag specific primers with DNA extracted from 1°SS patients

Sequence specific PCR primers to the first and second conserved regions of murine vSag (vSagC1 and vSagC2) were used to amplify products from DNA isolated from the peripheral blood of patients with 1°SS. The sizes of these PCR products were compared with products amplified from normal human DNA by electorphoresing on agarose gels, to investigate whether any unique products were found in the patient group. The PCR products amplified from eight 1°SS patients are shown in figure 5.1. There were no PCR products unique to 1°SS patients. PCR amplification with the primer pair to vSagC1 gave a single product of similar size to the 96 bp product cloned and sequenced from normal human DNA (see chapter 2), the equivalent murine vSag sequence being 91 bp. As with normal human DNA, two main PCR products (176 and 214 bp) were identified using the primers to vSagC2. Under the conditions employed ($50^0$C annealing temperature with an increase in PCR cycles from 25 to 30), a number of larger PCR products were identified with this primer pair, both in the 1°SS patients and in normal DNA. These were not considered significant in view of the large discrepancy in size between these products and the 182 bp product amplified from the vSag gene of the GR mouse DNA.




**a. PCR primers to vSagC1**




**b. PCR primers to vSagC2**

**Figure 5.1 PCR products amplified from 1°SS DNA with primers to vSagC1 and vSagC2.** DNA was extracted from peripheral blood of eight patients with1°SS and one normal control (N). (a) Products amplified with sequence specific PCR primer to the first conserved regions of murine vSag (vSagC1). (b) Products amplified with sequence specific PCR primer to the second conserved regions of murine vSag (vSagC2). The products amplified from GR+ murine DNA were of the sizes predicted for vSagC1 (91 bp) and vSagC2 (182 bp). M= molecular weight marker (bp).

### 5.3.3 Southern analysis of 1°SS DNA using vSagC1 and vSagC2 probes.

The murine probes vSagC1 and vSagC2 were hybridised under reduced stringency conditions to DNA extracted from the peripheral blood of 1°SS patients which had been digested with the restriction enzymes *Hind*III and *Xba*I. Autoradiographs were analysed for the presence of restriction fragments corresponding to the sizes predicted for the C1A, C1B, C1C, C2A and C2B loci (chapter 3). Only very weak hybridisation signals were identified when the *Hind*III digested DNA was hybridised with the vSagC1 probe (data not shown). These weakly hybridising restriction fragments did not correspond to the lengths to the HindIII restriction fragments of the λ clones isolated in chapter two.

The autoradiograph of a Southern blot hybridised under reduced stringency conditions (25% formamide) with the murine probe vSagC2 is shown in figure 5.2. A 4 kb *Xba*I fragment was identified after hybridising the vSagC2 probe with the *Xba*I digested DNA, which corresponded to the restriction fragment of the λ37a clone (locus C2B). This restriction fragment was present in all eight patients and normal DNA. Hybridising restriction fragments otherwise gave very weak signals. A 9 kb *Xba*I restriction fragment, corresponding to the *Xba*I restriction fragment of λ13a (locus C2A), was weakly visible in the DNA of patients 3 and 4. The vSagC2 probe also hybridised very weakly to a 7 kb *Xba*I restriction fragment of the DNA extracted from patient 3, a size corresponding to the *Xba*I restriction fragment of λ39a (Figure 3.7b). This clone had not been sub-cloned or sequenced due to the relative weakness of its hybridisation in the dot-blot experiments (see chapter 3).

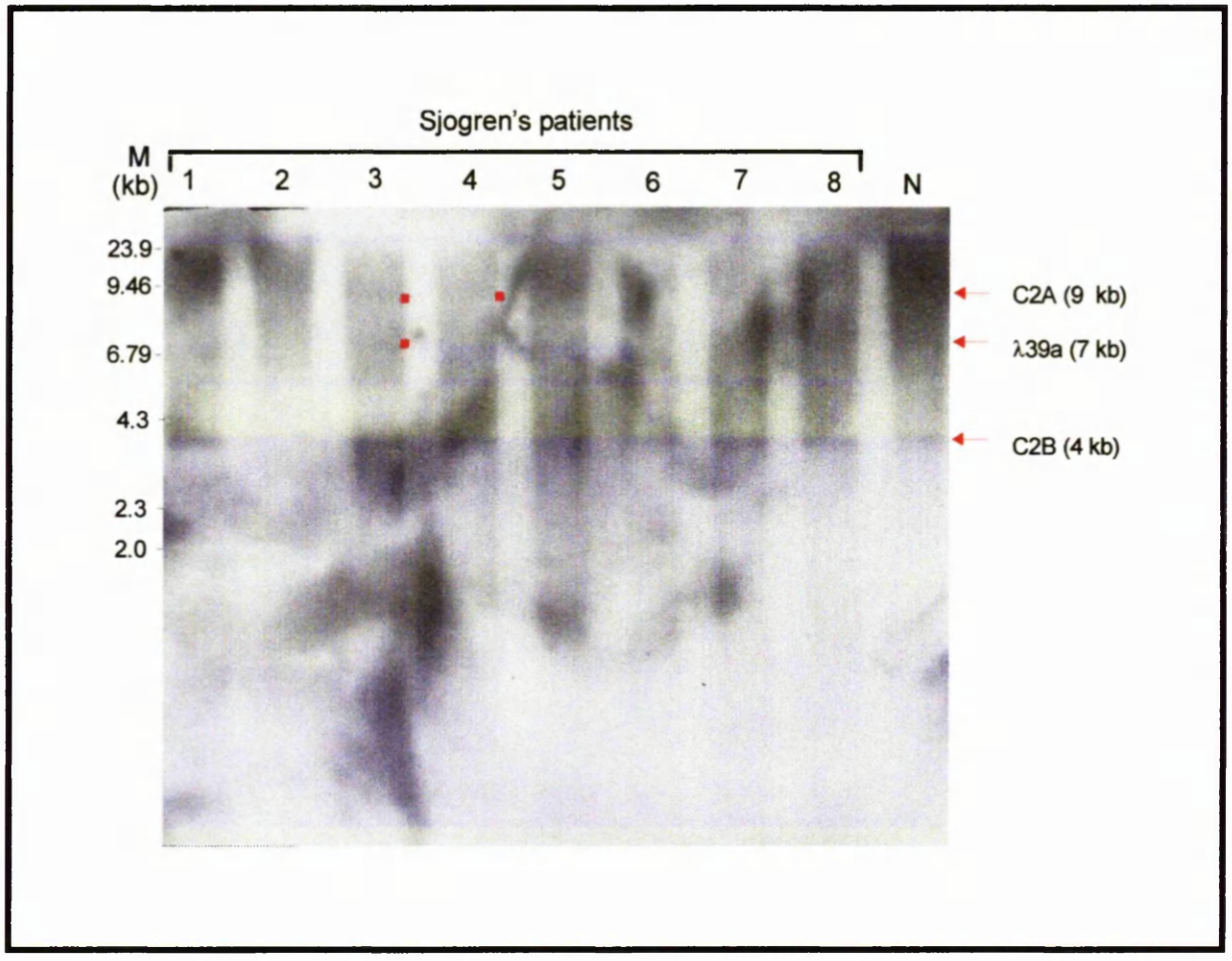



**Figure 5.2 Hybridisation of peripheral blood DNA from 1°SS patients with the vSagC2 probe.** DNA was extracted from the peripheral blood of eight 1°SS patients (lanes 1-8) and one normal control (N). The DNA was digested with *Xba*I, electrophoresed on a 1% agarose gel, and transferred to a nylon membrane. The DNA was hybridised with the vSagC2 probe under reduced stringency conditions (25% formamide in the prehybridisation, hybridisation and wash solutions). Red marks indicate weakly hybridising 7 kb and 9 kb restriction fragments.
M= molecular weight marker (kb).

### 5.3.4 Southern analysis of 1°SS DNA using HRC1 and HRC2 probes.

DNA from the 1°SS patients was hybridised with probes prepared from the 96 bp (HRC1) and 176 bp (HRC2) PCR products amplified from normal human DNA with primers to the first and second conserved region of murine vSag respectively. Hybridisations were carried out under high stringency conditions (50% formamide). No restriction fragments were identified which hybridised to the HRC1 probe, even when the stringency of the hybridisation was lowered, by reducing the formamide concentration from 50% to 45%. Figure 5.3 shows an autoradiograph of *Hind*III digested peripheral blood DNA from patients with 1°SS, which had been hybridised with the HRC2 probe. This probe hybridised to a 7.5 kb restriction fragment (patients 3, 4, 5, 7, 8 and normal DNA) which corresponds to the restriction fragment length predicted from the restriction map of the HRC2 locus (chapter 3). The probe hybridised to an apparently longer restriction fragment (7.6kb) in patients 1, 2 and 6, suggesting the presence of a small insert of approximately 100 nucleotides in these patients.

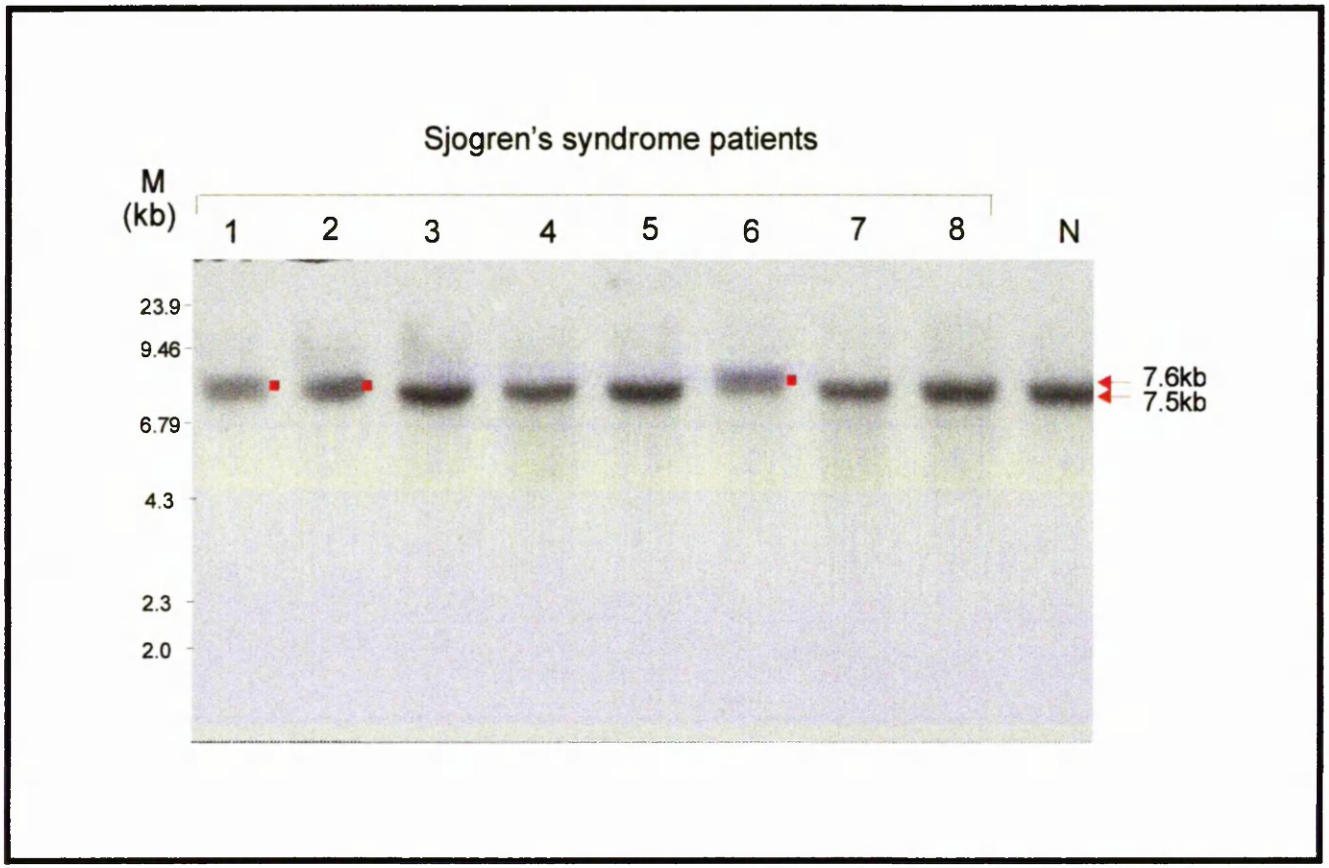



**Figure 5.3 Hybridisation of peripheral blood DNA from 1°SS patients with the HRC2 probe.** DNA was extracted from the peripheral biood of eight 1°SS patients (lanes 1-8) and one normal control (N). The DNA was digested with *Hind*III, electrophoresed on a 1% agarose gel, and transferred to a nylon membrane. The DNA was hybridised with the HRC2 probe under high stringency conditions (50% formamide in hybridisation solution). Red marks indicate 7.6 kb restriction fragments.
M= molecular weight marker (kb).

### 5.3.5 Northern analysis of 1°SS salivary gland mRNA using vSag related probes.

To look for evidence of expression of the human genomic DNA sequences sharing regions of homology with murine vSag sequences, mRNA extracted from the salivary glands of 1°SS patients was hybridised with the probes C1A, C1B, C1C, C2A, C2B and HRC2. Only weak hybridisation signals were identified. The C2A probe hybridised to a 1.87 kb mRNA transcript from patient 6 (figure 5.4). The C2B probe hybridised to similar sized mRNA transcripts in patients 3 and 6 and to a 2.2 kb mRNA transcript in patient 2. Hybridisation to discrete mRNA bands was not identified with the probes C1A, C1B, C1C and HRC2 (data not shown). The 1.87 mRNA transcript was of a similar size to 18S ribosomal RNA (corresponding length = 1874 nucleotides). To confirm that the probes were not hybridising with contaminating ribosomal RNA, they were hybridised with total RNA extracted from the MCF-7 cell line, which gave an S18 ribosomal RNA band (which had been used as a molecular weight marker). None of the probes hybridised with S18 ribosomal RNA (data not shown), suggesting true mRNA species had been identified.

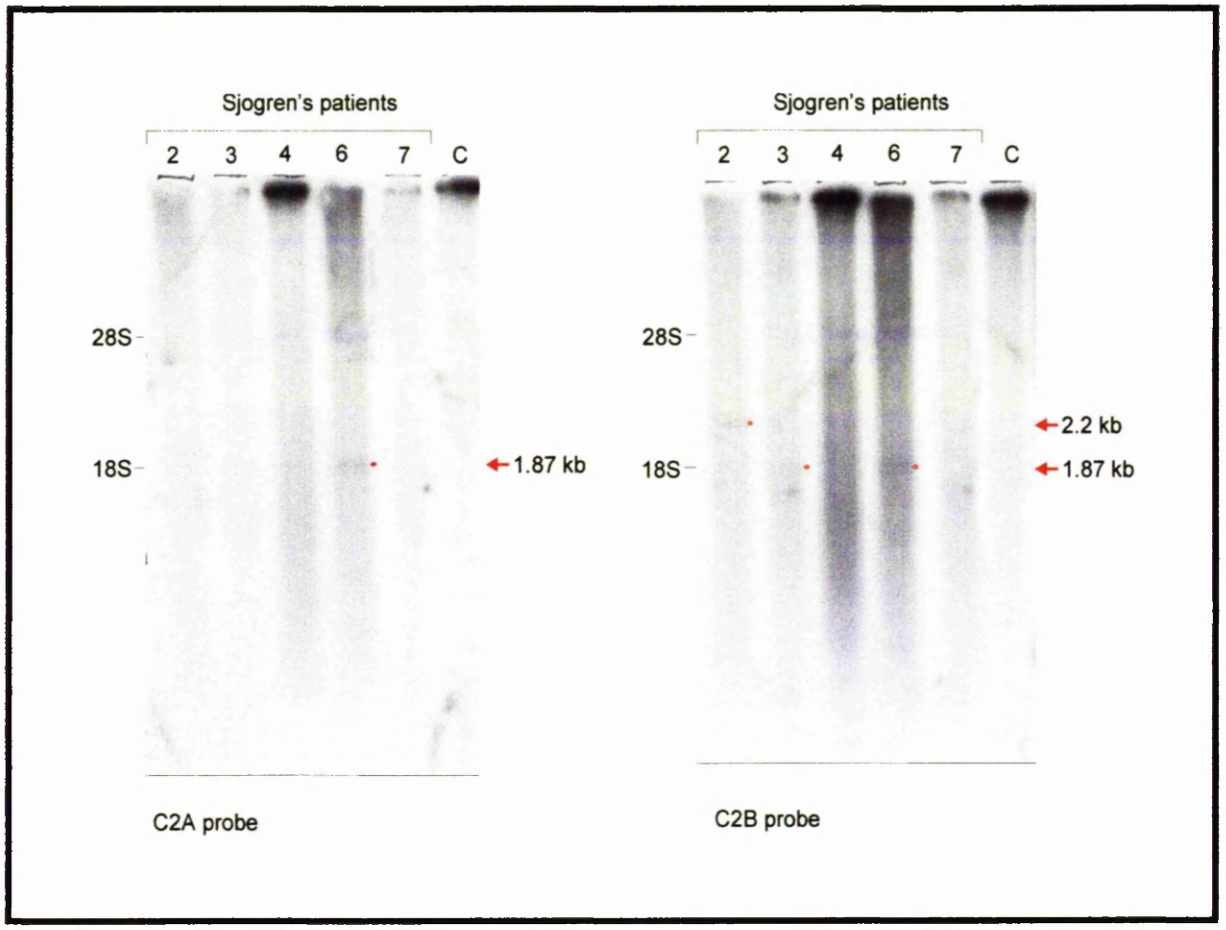



**Figure 5.4 Hybridisation of mRNA extracted from minor salivary glands of 1°SS patients with C2A and C2B probes.** mRNA was extracted from the minor salivary glands of five 1°SS patients and one control patient (C) with chronic sialadenitis using oligo$(dT)_{25}$ magnetic beads. The mRNA was electrophoresed on a denaturing gel containing 1.2% agarose and 17% formaldehyde, and transferred to a nylon membrane. The mRNA was hybridised with C2A and C2B probes under high stringency conditions (for 3 days at 42°C with 50% formamide in hybridisation solution). Red marks indicate mRNA transcripts hybridising to the probes. The position of 18S and 28S ribosomal RNA markers are shown on the left.

## *5.4 Discussion*

In this chapter, evidence was sought for the presence and expression of vSag-related DNA sequences in patients with 1°SS. As vSag genes are highly polymorphic in mice, with individual strains having several different vSag genes at different sites within their genomes, it was relevant to look for similar polymorphisms in human genomic DNA. This was done using both PCR and Southern blotting techniques. PCR had the advantage of being extremely sensitive, but could potentially fail to identify unique products if two or more were of the same length. Southern blotting of digested peripheral blood DNA with vSag probes was therefore also used, as this would allow the identification of multiple vSag-related sequences in the genome, showing any differences between patient and normal DNA as restriction fragment length polymorphisms (RFLPs).

PCR primers specific for the first conserved region of murine vSag amplified the same sized product from the DNA of all eight 1°SS patients which was of comparable size to the product amplified from normal human DNA. This 96 bp product compared to the 91 bp product amplified from the vSag gene of the GR mouse DNA. When sequenced from normal DNA, this PCR product had a similarity index of 71% to the murine vSag sequence (chapter two), however the identities of the PCR products amplified from 1°SS patients were not confirmed either by sequencing or restriction enzyme digestion. The PCR products amplified with primers to the second conserved region of murine vSag were also of similar size in 1°SS and normal DNA. These primers gave two products of 176 bp and 214 bp, compared to the 182 bp product of murine vSag, with a similarity index of 52.1% between vSagC2 and the 176 bp product of normal DNA. Again the PCR products amplified from the 1°SS patients were not sequenced. Although a number of larger products were amplified from both 1°SS patient and normal DNA using the primers to vSagC2, these were not likely to be significant, as

they were relatively weak bands, and were not of comparable size to the murine vSag product.

The Southern blots of 1°SS peripheral blood DNA were probed under reduced stringency conditions with the murine probes vSagC1 and vSagC2. Hybridisation with the murine vSag probes failed to identify hybridising restriction fragments unique to 1°SS patients and there were no obvious restriction fragment length polymorphisms between normal and 1°SS DNA. Only very weak hybridisation signals were identified with vSagC1. None of these weakly hyridising *Hind*III restriction fragments corresponded to the lengths of the restriction fragments identified in the λ clones isolated from the human placental DNA library (chapter 2). This may simply reflect the relative insensitivity of Southern blotting, however reducing the stringency of the hybridisation further, by decreasing the formamide concentration to 20% resulted in very high background signal. The vSagC2 probe gave a clearer hybridisation signal with a 4 kb restriction fragment, a length corresponding to the *Xba*I restriction fragment of the clone λ 37a (locus C2B) isolated from normal human placental DNA library. Other hybridisation signals were very weak. 9 kb and 7 kb *Xba*I restriction fragments were just visible in the 1°SS patients 3 and 4, as well as in normal DNA. These restriction fragments corresponded to the 9 kb *Xba*I fragment of λ13a (locus C2A), and the 7 kb *Xba*I fragment of λ39a. This latter clone had not been investigated beyond the dot-blot hybridisation analysis (chapter 3) as it had been shown to have relatively weak hybridisation strength.

The probes HRC1 and HRC2, prepared from the PCR products of normal human DNA with primers to the first and second conserved regions of vSag respectively (see chapter two), were hybridised with Southern blots of 1°SS and normal DNA. This showed that the 176 bp PCR product amplified with vSagC2 specific primers was

present as a single copy in the genome of all eight of the 1°SS patients (figure 5.3). The *Hind*III restriction fragment of DNA from the 1°SS patients was approximately the same length as that identified in the λ clones isolated from the normal human placental DNA library, however there was an apparent minor size difference between the restriction fragments in three of the patients. This could have resulted from the insertion of a small DNA fragment within the HRC2 locus in these patients.

Probing the Southern blots with the 96 bp human PCR product amplifed with primers to the first conserved region of vSag failed to identify any hybridising restriction fragments, either in normal or 1°SS DNA. This probe had also failed to hybridise to any of the λ clones from the normal human placental DNA library (chapter two). The failure of this probe to hybridise under high stringency conditions could have resulted from its relatively low C:G content. At 50% between the PCR primers, this was significantly below that of the HRC2 probe, which had a C:G content of 56.8%, and would have resulted in a lower $T_m$ for hybridisation. Hybridisations were therefore repeated under less stringent condition (45% formamide) with a freshly prepared PCR probe, but again no hybridisation signals were identified. It was therefore concluded that the 96 bp PCR product amplified from human DNA with primers specific for the first conserved region of vSag was a spurious product, that could have resulted from complex primer dimer formation.

Northern blotting experiments were performed to look for evidence of expression of the human DNA sequences C1A, C1B, C1C, C2A, C2B and HRC2 in the minor salivary glands of 1°SS patients. These glands had multiple foci of lymphocytic infiltrates, and if vSag related genes were involved in driving a sustained immune reactions in these patients, it would be expected that mRNA would be detectable at this site. mRNA was extracted directly from the minor salivary glands of five patients

with 1°SS using oligo$(dT)_{25}$ magnetic beads. A salivary gland biopsy from a patient with chronic sialadenitis resulting from parotid duct obstruction was used as a normal control. The oligo$(dT)_{25}$ magnetic beads were chosen for mRNA extraction, as this gave reasonable yields of intact mRNA from the minor salivary gland biopsies. Although some mRNA species might have been lost with this technique, for example, those lacking poly (A)+ residues as a result of partial degradation, this method eliminated the need for an intermediate total RNA step, giving high purity mRNA, with minimal contamination from DNA or ribosomal RNA.

Although less sensitive, Northern blotting was used rather than RT-PCR, as even minute levels of DNA contamination would give positive results with RT-PCR, and in the absence of known introns, the PCR products would be of the same length in DNA as in mRNA. Only weak hybridisation signals were identified, suggesting that the mRNA species were rare. The C2A and C2B probes hybridised to a 1.87 kb mRNA from one 1°SS patient who had both parotid gland swelling and lymphadenopathy. A similar sized mRNA species, which hybridised only with the C2B probe, was identified in a second patient with lymphadenopathy. The C2B probe also hybridised with a 2.2 kb mRNA species from a third 1°SS patient with parotid gland enlargement.

The 1.87 kb mRNA was of similar size to S18 ribosomal RNA, raising the possibility that the probes might be hybridising with contaminating ribosomal RNA. However, none of the probes hybridised with S18 ribosomal RNA from total RNA extracted from the MCF-7 cell line. These transcripts would be comparable in length to the murine vSag transcripts, although the open reading frame in the sequence C2A was only 240 nucleotides (79 amino acids), and no significant open reading frames were identified in the sequence C2B. There were no hybridising transcripts in the remaining two 1°SS

patients or in the control. No transcripts were identified which hybridised to the probes C1A, C1B, C1C or HRC2.

The significance of these putative mRNA transcripts is unclear. The salivary gland samples from which they were extracted came from patients with parotid swelling (Talbe 5.2). Although it is unlikely that the human vSag-related sequences describe in this thesis have an aetiological role in 1°SS, their expression could reflect increase cell turnover within the salivary tissue. The potential relevance of these findings will be discussed further in the general discussion in chapter seven.

# 6 CHAPTER SIX

# Search for HTLV-1 *tax* in DNA and mRNA from patients with primary Sjogren's syndrome

## *6.1 Introduction*

The preceding chapters have described the identification and characterisation of human DNA sequences sharing regions of homology with the murine retroviral superantigen gene, vSag, encoded in the 3'LTR of the mouse mammary tumour virus. Evidence of expression of these DNA sequences was sought in the autoimmune condition 1°SS, by looking for mRNA transcripts in the minor salivary glands of these patients. Putative transcripts were identified in three of the five patients examined. The C2A probe hybridised with a distinct mRNA band from a minor salivary gland extract of one patient, while the C2B probe hybridised to distinct mRNA bands in three patients. Both these probes shared limited regions of homology with the second conserved region of murine vSag.

Other retroviral gene transcripts have been described in the minor salivary glands of 1°SS patients (For reviews see (154, 235, 236)). As discussed in chapter one, an exocrinopathy resembling Sjogren's syndrome develops in mice transgenic for the *tax* gene (205), encoded by the retrovirus human T cell leukaemia virus type 1 (HTLV-1). HTLV-1, which is endemic in the West Indies and in South West Japan, is aetiologically associated with adult T cell leukaemia /lymphoma (ATL) (237-239), tropical spastic paraparesis (TSP) (240-242) and possibly multiple sclerosis (243). HTLV-1 *tax* sequences, but not *gag*, *pol* or *env* genes, have been detected in the

minor salivary glands of some patients with Sjogren's syndrome in two separate studies (202, 203), although it was not found in peripheral blood DNA in British patients with rheumatological conditions (244).

The disparity between these reports could have resulted from the geographical origin of the patients studied. The report by Sumida *et al* (203)originated from Japan, where HTLV-1 is endemic, with a 10-25% prevalence of serum antibodies to retroviral proteins in adults over 40 year (245). It was therefore of interest to look for evidence of HTLV-1 *tax* in the white Caucasian patients with 1°SS from the North East of England.

## *6.2 Materials and Methods*

### 6.2.1 Patients

The DNA samples were extracted from the peripheral blood of the eight 1°SS patients describes in section 5.2.1. mRNA was extracted from minor salivary gland biopsies of five of these patients (see Table 5.2, chapter five).

### 6.2.2 PCR amplification with HTLV-1 *tax* specific primers.

Specific Oligonucleotide primers for PCR were synthesised by a Beckman 1000M DNA synthesiser and their concentration adjusted to 1mg/ml in TE pH 8. The primer sequences were the same as those used by Mariette *et al* (202, 246) (SK 43/44) and were combined as shown below (each primer at 100 μg/ml). The position of the PCR primer is shown in figure 6.1.

| | | | |
|---|---|---|---|
| SK43 | sense | 638 | CGGATACCCAGTCTACGTGT |
| SK44 | antisense | 776 | CGATGGACGCGTTATCGGCTC |

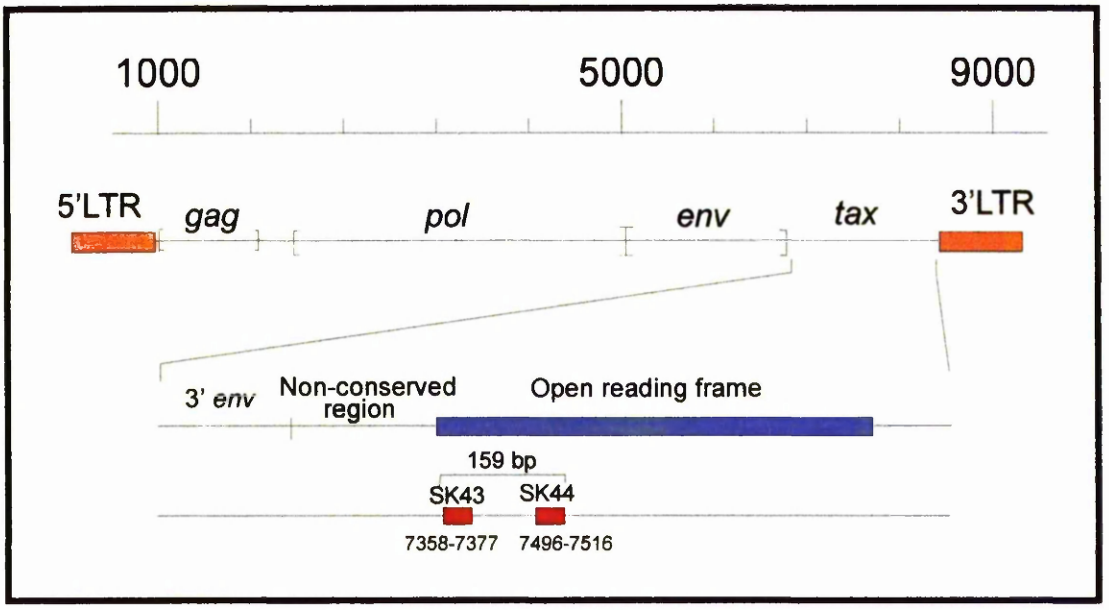



**Figure 6.1 Open reading frame in the 3'LTR of HTLV-1 encoding *tax*.** The position of the PCR primer SK43 and SK44 are shown (red boxes).

A cDNA clone of HTLV-1 *tax* derived from the Hut 102 cell line in the plasmid pSP65 (supplied by Dr Ruth Jarrett, Leukaemia Research Fund Centre, University of Glasgow) was linearised prior to PCR amplification by digestion with *Hind*III and the concentration adjusted to 1 μg/ml. DNA extracted from normal human placenta (negative control) and peripheral blood of 1°SS patients was diluted to 100 μg/ml. PCR reactions were carried out in 25 μl volumes under mineral oil with 100 ng of each primer, 0.2 mM dNTPs, 2.5 μg BSA, PCR buffer and Taq polymerase (0.625 units/reaction). 1 ng plasmid DNA and 100 ng human placental or 1°SS peripheral blood DNA were amplified in a thermal cycler (96°C denaturation for 90 seconds, 45°C annealing for 30 seconds and 72°C extension for 30 seconds, followed by 25 cycles of 96°C, 50°C and 72°C each for 30 seconds, with a final extension at 72°C for 120 seconds). The degree of stringency of the reactions was adjusted by altering the annealing temperatures (45°C,50°C and 55°C) in the thermal cycle (the lower temperatures producing less stringent conditions). PCR products were run on 3% agarose minigels (1 g type I agarose, 2 g Nusieve in 100 ml TBE) containing 0.5 μg/ml ethidium bromide.

### 6.2.3 Reverse transcription of minor salivary gland mRNA

mRNA extracted from the minor salivary glands of five patients with 1°SS and a control with chronic sialadenitis (see section 5.2.5) was reverse transcribed in 10 μl volumes by incubating 100 ng mRNA for 1 hour at 37°C in a reaction mix containing 1.25 μg random primers, 0.5 mM dNTPs, 10 mM DTT, RT buffer and 5 units RNA guard (incubating at 37°C for 3 minutes prior to adding the Rtase MoMuLV). 1 μl of cDNA (corresponding to 10 ng mRNA) was amplified by PCR with the HTLV-1 *tax* specific primers as described in section 6.2.1. The *Hind*III linearised plamid containing the cDNA clone of HTLV-1 tax (1 ng) was used as a positive control for the PCR assay. As a positive control for the reverse transcription, ribosomal primers were used to amplify cDNA reverse transcribed from 20 ng total RNA from the MCF-7 cell line. Samples were electrophoresed on 3% agarose minigels containing 0.5 μg/ml ethidium bromide.

## *6.3 Results*

### 6.3.1 Optimisation of the HTLV-1 *tax* PCR

The PCR conditions were optimised using the *Hind*III linearised plasmid containing the cDNA clone of HTLV-1 *tax* as a template. No PCR products were amplified with the initial thermal cycler settings; 30 cycles of denaturation 94°C for 30 seconds, annealing 50°C for 30 seconds, extension 72°C for 30 seconds. As the C:G content of the expected 159 bp product was 60%, giving a high predicted $T_m$, the denaturation temperature was increased to 96°C. PCR reactions were also carried out with 5% formamide or 5% DMSO added to the reaction mixture, again to facilitate

denaturation. Under each of these conditions, the predicted 159 bp product was amplified from the HTLV-1 *tax* containing plasmid.

### 6.3.2 PCR amplification of peripheral blood DNA from 1°SS patients using HTLV-1 *tax* specific primers.

PCR reactions were carried out under the conditions described above using the HTLV-1 *tax* specific primers with DNA extracted from the peripheral blood of eight patients with 1°SS together with normal placental DNA. The 159 bp product predicted from the HTLV-1 *tax* sequence was generated when the *Hind*III linearised plasmid containing the cDNA clone of HTLV-1 *tax* was used as the template (figure 6.2). No PCR products of this size were identified in the DNA extracted from peripheral blood of 1°SS patients or in the normal placental DNA control when PCR annealing temperature was 50°C (data not shown). Lowering the stringency of the PCR reaction, by reducing the annealing temperature to 45°C, resulted in the formation of a 1.3 kb product in the eight 1°SS patients and in the normal DNA control. A very weak band of similar size to the HTLV-1 *tax* product was also present in the patient samples and normal placental DNA under these PCR conditions. These products were not sequenced to confirm their identity.


**Figure 6.2 PCR products amplified from 1°SS DNA with primers to HTLV-1 *tax*.** DNA was extracted from peripheral blood of eight patients with1°SS and one normal control (N). Products were amplified with sequence specific primer for HTLV-1 *tax* at an annealing temperature of 45°C. A 159 bp product was amplified from *Hind*III linearised plasmid containing the tax gene. In addition to a 1.3 kb product, a very weak band of similar size to the HTLV-1 *tax* product is present in the DNA extracted from the peripheral blood of 1°SS patients and normal placental DNA .
M= molecular weight marker (bp).

### 6.3.3 RT-PCR of mRNA extracted from minor salivary biopsies of 1°SS patients using HTLV-1 *tax* specific primers

mRNA extracted from the minor salivary glands of five 1°SS patients was reverse transcribed and assayed for the presence of HTLV-1 *tax* using specific PCR primers. Ribosomal primers amplified an appropriate sized product from the total RNA, confirming successful reverse transcription. No PCR products were identified of an equivalent size to that obtained using the HTLV-1 *tax* containing plasmid as template DNA (figure 6.3). There was therefore no evidence of HTLV-1 *tax* mRNA in the 1°SS patients' salivary gland biopsies.

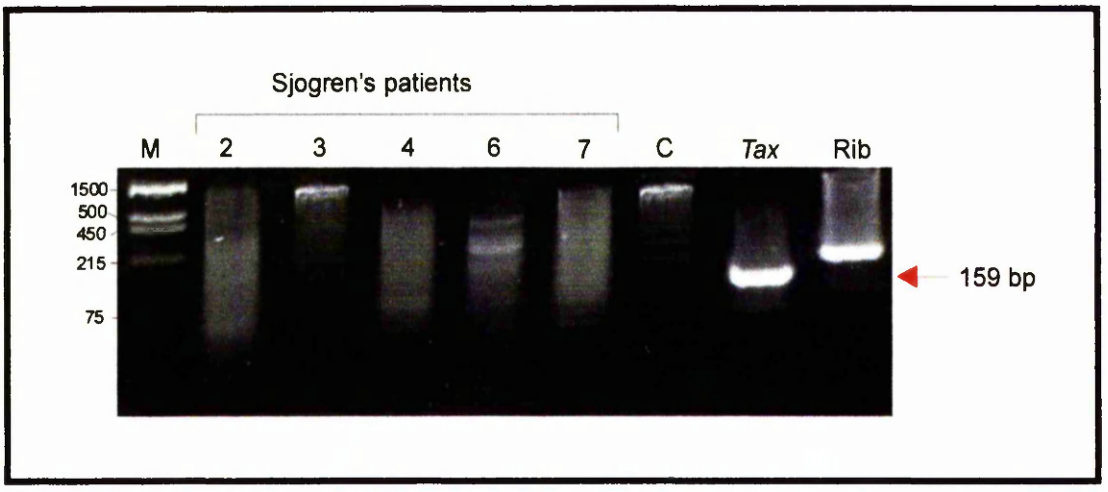



**Figure 6.3 RT-PCR of 1°SS mRNA from minor salivary gland biopsies with primers to HTLV-1 *tax*.** mRNA was extracted from the minor salivary glands of five 1°SS patients and one control patient (C) with chronic sialadenitis using oligo$(dT)_{25}$ magnetic beads. The mRNA was reverse transcribed prior to PCR amplification with sequence specific primer for HTLV-1 *tax* at an annealing temperature of 45°C. A 159 bp product was amplified from *Hind*III linearised plasmid containing the tax gene but not from mRNA extracted from minor salivary glands of 1°SS patients. Rib= RT-PCR of total RNA from MCF-7 cell line with primers to ribosomal RNA. M= molecular weight marker (bp).

## *6.4 Discussion*

In this chapter, the peripheral blood DNA and salivary gland mRNA extracted from the 1°SS patients described in Table 5.2 was investigated for the presence of HTLV-1 *tax*. The sequence specific primers described by Mariette *et al* (202, 246) (figure 6.1) were used in PCR and RT-PCR assays. These authors had found evidence of HTLV-1 *tax* DNA by PCR and in situ hybridisation in minor salivary glands from two of nine 1°SS patients (202). In our study, PCR products were only identified when the stringency of the PCR reaction was reduced, by lowering the annealing temperature. In addition to a 1.3 kb product, a very weak band of similar size to the 159 bp HTLV-1 tax product was identified (figure 6.2). This product was present in the peripheral blood DNA of all eight 1°SS patients, but also in the normal human placental DNA. These PCR products were not sequenced, and so conclusions about their identity cannot be

made, however if weakly related to HTLV-1 tax, these amplified products do not appear by be unique to 1°SS syndrome patients.

There was no evidence of HTLV-1 *tax* in the mRNA extracted from the minor salivary glands of the five 1°SS patients studied using RT-PCR. This contrasts with a report from Japan (203), where HTLV-1 *tax* mRNA was identified in the minor salivary glands of four out of fourteen 1°SS patients, again using RT-PCR. HTLV-1 *tax* was also found in the DNA extracted from the minor salivary glands of these patients. The identity of these PCR products was confirmed by sequencing, giving 100% homology to HTLV-1 *tax* over the sequenced region. In both this study and the report by Mariette *et al* (202), there was no evidence of HTLV-1 *gag*, *pol* or *env* sequences and HTLV-1 *tax* was not identified in peripheral blood DNA. There was also no evidence of serum antibodies to HTLV-1 proteins in any of the patients.

The absence of antibodies to HTLV-1 in these two reports is at odds with the results of other investigators, where antibodies to HTLV-1 were present in around 30% of 1°SS patients (247). In the Japanese study by Sumida *et al*, because HTLV-1 was endemic in the area, patients with positive serology were excluded. As up to 30% of the 1°SS patients might be expected to have cross-reactive antibodies, the patients studied would be an unrepresentative group. The study by Mariette *et al* (202) was undertaken in France. Although the ethnic origin of their nine patients was not reported, none had previously lived in areas endemic for HTLV-1. Only four of these nine patients had primary Sjogren's syndrome, the other five having Sjogren's syndrome associated with rheumatoid arthritis (RA). Of the two patients in which HTLV-1 tax DNA was identified, one had primary and the other secondary Sjogren's syndrome. These conditions are clinically and immunogenetically distinct, making it is difficult to interpret these results.

In a study from the UK (244), PCR was used to look for evidence of HTLV-1 *tax* and HIV-1 *gag* DNA sequences in patients with rheumatological conditions who had serum antibodies cross-reacting with these retroviruses. Antibodies to HTLV-1 were identified in nine of sixty-two patients with rheumatological conditions (4/30 with RA, 3/13 with polymyositis, 2/5 with systemic lupus erythematosus and 0/14 with 1°SS) compared to two out of a control population of eighty, comprising osteoarthritis, Crohn's disease and bacterial endocarditis, and none out of 30 healthy controls. Antibodies to HIV-1 p24 were detected in one RA and one 1°SS patient and in two osteoarthritic patients. PCR using HTLV-1 *tax* specific primers (SK43 and SK44) failed to amplify the expected 159 bp product from peripheral blood DNA extracted from those patients with positive serology. However PCR for HTLV-1 *tax* was not undertaken in any of the patients with 1°SS.

Our study of 1°SS patients from the North East of England has also failed to identify HTLV-1 *tax* in the peripheral blood DNA or mRNA extracted from minor salivary gland biopsies. Taken together with the above report (244), these results do not lend support to the hypothesis that HTLV-1 *tax* is involved in the aetiology of 1°SS.

# 7 CHAPTER SEVEN

## General Discussion

The principle reason for undertaking this study was to determine whether endogenous superantigens are present in the human genome. All known endogenous superantigens in mice are encoded by vSag, a retroviral gene that lies within an open reading frame in the 3'LTR of MMTV. As the human genome contains numerous MMTV-related sequences, it was appropriated to conduct a search for human DNA sequences related to vSag. It was decided to look for these sequences in DNA rather than for mRNA transcripts, as it was not known which tissues might express human endogenous superantigens, and although searching within genomic DNA may have resulted in the identification of non-coding sequences, any potentially expressible MMTV-related superantigen would be identified. To avoid missing endogenous superantigens that did not lie within a proviral configuration, it was decided to screen human genomic DNA directly with probes and PCR primers derived from the MMTV superantigen-encoding gene, rather than screening MMTV related clones isolated with *gag-pol* probes.

In this study, six genomic loci have been identified in human DNA that contain DNA sequences with short regions of homology to vSag. There were no human DNA sequences that were closely related to vSag over the entire length of this gene. Furthermore, the six weakly vSag-related sequences did not appear to lie within a proviral configuration. These sequences did, however, have open reading frames, giving predicted proteins ranging between 43 and 156 amino acids in length.

Although smaller than the predicted 320 amino acid sequence of murine vSag, could these human sequences with short regions of homology to vSag be biologically relevant? As discussed in chapter one, there is evidence that the biologically active form of vSag is an 18 kDa molecule, significantly smaller than the predicted 320 amino acid protein. This may result from either proteolytic digestion of the full-length vSag, or from the use of an alternative translation initiation codon, but both would result in a predicted active protein of approximately 150 amino acids. This truncated protein encompasses the second conserved region (vSagC2) together with the highly polymorphic carboxyl terminal that is predicted to bind to the Vβ region of the T cell receptor. It is therefore theoretically possible that the three clones (C2A, C2B and HRC2) sharing short regions of homology with vSagC2 could possess some of the functions of the biologically active 18 kDa murine vSag protein. In particular, a seven amino acid sequence motif (DRWWQPG) appears to be conserved between murine vSag and the three vSagC2-related clones. As discussed in chapter one, this region of murine vSag may be involved in the interaction with the MHC class II molecule, which could explain why it is so highly conserved.

The observation that the human vSag-related clones do not lie within a proviral configuration raises questions regarding the origin of the retroviral vSag gene. In mice, fully functional vSag genes have been identified that do not lie within a complete proviral configuration. Hybridisation data indicates that the Mtv-6 provirus is missing most of the *gag* and *env* genes and the entire *pol* gene (248). It was originally thought that this isolated vSag gene could be an integrated copy of a vSag transcript, however sequence data suggests it is a deletion mutant (153). This large deletion (6.2 kb) could have occurred at the stage of viral replication, or as a result of recombination after integration of a complete provirus.

An alternative interpretation of the existence of human vSag-related sequences that are not in proviral configuration is that they could represent cellular sequences, which during evolution have been acquired by infectious retroviruses. Directly transforming retroviruses carry within their genome a transduced cellular gene encoding a protein involved in normal cell cycle regulation (58). When a cell is infected by a directly transforming retrovirus, RNA encoding the transduced cell growth-controlling protein is expressed in a dysregulated fashion leading to cellular proliferation and ultimate oncogenic transformation. These genes that affect host cell behaviour do not arise out of retroviral RNA, but rather are incorporated into the retroviral genome from RNA transcribed from cellular homologues (249). It is therefore possible that the open reading frame in the MMTV 3'LTR that encodes vSag is derived from a host protein of similar function. A Sag gene of cellular origin could have been maintained in the retroviral genome through positive selection pressures, resulting from the stimulatory effects of vSag inducing proliferation of cells in which the provirus was integrated.

It has been suggested that the possible function of a cellular Sag gene would be to serve as a co-ligand for peptide antigen recognition (250). The co-ligand would participate in normal T cell receptor function, helping to align the TCR and co-receptor (CD4 or CD8) over the peptide-binding site in the MHC class II molecule. The major function of a peptide antigen would be to induce a conformational change in this already partially cross-linked TCR, the conformational change of such bound receptors leading to signals that generate T cell activation. If this hypothesis were true, the retroviral superantigens could be transduced copies of these cellular superantigens, and through mutation, may only have retained limited homology.

It is clear from the bacterial enterococcal toxin, that superantigens can have remarkably diverse primary structures and yet have closely related biological functions

(1). Further research will be necessary to investigate the potential biological activity of the human clones sharing short regions of homology with murine vSag. Further study of cDNA clones is required to prove that vSag-related sequences are transcribed. The next step would be to produce proteins encoded by the cloned vSag-related sequences, using an expression vector system. The potential function of these proteins could then be assess in a number of ways.

In order to prove that expressed human vSag-related sequences have the potential to function as superantigens, it will be necessary demonstrate two fundamental properties; that these proteins can bind to MHC class II molecules, and that the protein-class II complex can stimulate T cells in a Vβ specific manner. A strategy which has been used to demonstrate the superantigen function of an endogenous retroviral protein derived from the pancreas of patients with type I diabetes mellitus (215), could be adapted to serve this purpose. In these experiments, retroviral transfectants of monocyte and B lymphocyte cell lines were generated and tested for their ability to stimulated MHC-compatible and incompatible T cell lines in a Vβ-specific manner. This system allows the demonstration of MHC class II dependence, by controlling the expression of class II molecules, through the action of interferon-γ (IFN-γ). Direct evidence of MHC class II binding could be obtained by experiments similar to those undertaken by Mottershead *et al* (95), who demonstrated binding of recombinant forms of murine vSag proteins to HLA-DR molecules. Monoclonal antibodies could be raised against predicted human vSag-related sequences, in an attempt to demonstrated inhibition of T cell Vβ specific proliferation, as has been reported with monoclonal antibodies against the carboxyl terminal region of murine vSag (81). These monoclonal antibodies could also be used to look for evidence of human vSag-related protein expression on different cell types.

If the experimental approaches described above were to demonstrate that the human vSag-related sequences do not have superantigen activity, it raises intriguing questions as to why such genes have been conserved in the murine genome, but are not present in man, where other MMTV-related sequences clearly exist. It is possible that the human genome never contained retroviral derived Sag genes, in which case chance homology has resulted in the isolation of the DNA sequences described in this thesis. Alternatively, human vSag genes may have existed in the past, but have not been conserved in the human genome. As discussed in chapter one, endogenous retroviral superantigens appear to protect against exogenous MMTV infection and subsequent mammary tumour formation in susceptible individuals (104, 105). This protection against retroviral induced tumours would provide a selective pressure to maintain functional endogenous vSag genes. Endogenous superantigens may also protect against the effect of bacterial enterotoxin superantigens, which presumably act as a virulence factor favouring the bacteria. If human endogenous retroviral superantigens do not exist, it suggests that similar selective pressures are not present.

There are a number of possible explanations why evolutionary pressures in humans may not have resulted in the retention of endogenous vSag genes. Such genes may not protect against the development of retroviral induced breast carcinoma. Indeed, despite extensive research in this area, a causal retroviral agent in human breast cancer has not been identified. In any case, human breast cancer typically affects women beyond there reproductive years, limiting the selective advantage of a protective gene. Bacterial enterotoxins do however have pathological effects in humans, causing food poisoning and toxic shock. Protection against these toxins would presumably be advantageous to the human host.

The protective effect of vSag genes in mice results from the deletion Vβ specific T cells, causing large gaps in the T cell repertoire (105). Although this has not been associated with immune deficiency syndromes in mice, it is conceivable that such deletions could result in subtle immunological defects, which could perhaps result in impaired response to some human specific pathogens, or a failure of immunological surveillance against tumour antigens. The price paid for protection against superantigen-encoding exogenous retroviruses and bacterial enterotoxins may simply have been too high in humans to constitute a positive selective pressure. However, as new human infectious retroviruses evolve, this situation might change. There is evidence that HIV-1 encodes a superantigen (9), and if stable integration in germ line cells were to occur, it is conceivable that this gene might provide protection in subsequent generations.

In the final sections of this thesis, evidence was sought for the involvement of endogenous retroviral genes in the human autoimmune condition primary Sjogren's syndrome. In addition to human vSag-related sequences, the presence and expression of HTLV-1 *tax* was studied, as this putative transactivator gene has recently been implicated in the pathogenesis of 1°SS (202, 203, 205). Although disease specific endogenous retroviral genes were not identified, a number of interesting observations were made which merit further investigation.

The peripheral blood DNA of 1°SS patients was initially examined for the presence of unique vSag-related sequences. The strength of hybridisation signals with vSag probes was no greater with 1°SS DNA than with normal DNA, excluding the possibility that a gene closely related to murine vSag is present in the genome of patients with this condition. There was however an apparent restriction fragment length polymorphism in three of eight 1°SS patients, on hybridising the HRC2 probe with

*Hind*III digested DNA. Further analysis of this possible polymorphism in 1°SS DNA is required before any significance can be attached to these findings, including analysis of a normal control population and other disease groups, to determine whether the polymorphism is disease specific.

The next step in assessing the potential involvement of vSag-related sequences in 1°SS was to look for evidence of gene expression in salivary gland biopsies, the commonest site of inflammation in these patients. Rare mRNA transcripts were detected in salivary gland biopsies from three of five 1°SS patients that hybridised with the vSag-related sequence C2B, the mRNA from one of these three patient also hybridising with the C2A probe. Cloning and sequencing of these transcripts is required to confirm that they have been transcribed from the genetic loci from which the probes were derived. Again analysis of a larger patient and control population is required before an association with 1°SS can be made.

Finally, PCR amplification with specific primer for HTLV-1 *tax* did not find evidence of this retroviral gene in peripheral blood DNA or in cDNA from the salivary glands of the 1°SS patients studied. However, under reduced stringency conditions, a faint band of the predicted size was identified in the DNA of 1°SS patients and normal control samples. Cloning and sequencing of this PCR product is required to determine whether it may represent an endogenous retroviral sequence related to HTLV-1 *tax*.

As discussed in chapter one, gene expression at a site of pathology does not necessarily imply a pathogenic role for the expressed gene. If the human vSag-related DNA sequences described in this thesis were to encode superantigens that could be demonstrated to stimulate T cells in a Vβ-specific manner, a causal role could be inferred if the infiltrating T cells in the salivary glands shared that Vβ region. Potential

therapeutic strategies that have been proposed for the treatment of superantigen-mediated autoimmune disease include TCR Vβ blocking monoclonal antibodies, or treatment directed against vSag genes. However, the evidence that superantigens cause autoimmune disease remains highly controversial. Overall the observations in this thesis do not support the hypothesis that superantigens have a causal role in 1°SS, and such therapeutic goals remain a long way in the future.

Much of the work described in this thesis has involved the isolation, cloning, subcloning and sequencing of human DNA. Such information will be readily available in the future. The first draft of the human genome project has recently been published and covers an estimated 88% of the human genome (112, 251). It is interesting that only three of the six genomic loci sequenced in this study were present in the searches of the human genome database, and highlights the need for further work to complete this international project. Open access to sequence databases will allow research of the nature described in this thesis to progress at a rapid pace. However, the interpretation of the mass of sequence data will remain a challenge for decades to come.

## References


1. Marrack P, Kappler J. The staphylococcal enterotoxins and their relatives. Science 1990;248(4959):1066.

2. Fleischer B. Superantigens produced by infectious pathogens: molecular mechanism of action and biological significance. International Journal of Clinical & Laboratory Research 1994;24(4):193-7.

3. Irwin MJ, Hudson KR, Ames KT, Fraser JD, Gascoigne NR. T-cell receptor beta-chain binding to enterotoxin superantigens. Immunological Reviews 1993;131:61-78.

4. Herman A, Kappler JW, Marrack P, Pullen AM. Superantigens: mechanism of T-cell stimulation and role in immune responses. Annual Review of Immunology 1991;9:745-72.

5. Dellabona P, Peccoud J, Kappler J, Marrack P, Benoist C, Mathis D. Superantigens interact with MHC class II molecules outside of the antigen groove. Cell 1990;62(6):1115-21.

6. Woodland DL, Blackman MA. How do T-cell receptors, MHC molecules and superantigens get together? Immunology Today 1993;14(5):208-12.

7. Marrack P, Winslow GM, Choi Y, Scherer M, Pullen A, White J, et al. The bacterial and mouse mammary tumor virus superantigens; two different families of proteins with the same functions. Immunological Reviews 1993;131:79-92.

8. Marrack P, Blackman M, Kushnir E, Kappler J. The toxicity of staphylococcal enterotoxin B in mice is mediated by T cells. Journal of Experimental Medicine 1990;171(2):455-64.

9. Kotzin BL, Leung DY, Kappler J, Marrack P. Superantigens and their potential role in human disease. Advances in Immunology 1993;54:99-166.

10. Acuto O, Reinherz EL. The human T-cell receptor. Structure and function. New England Journal of Medicine 1985;312(17):1100-11.

11. Marrack P, Kappler J. The T-cell repertoire for antigen and MHC. Immunology Today 1988;9(10):308-15.

12. Germain RN, Margulies DH. The biochemistry and cell biology of antigen processing and presentation. Annual Review of Immunology 1993;11:403-50.

13. Jorgensen JL, Reay PA, Ehrich EW, Davis MM. Molecular components of T-cell recognition. Annual Review of Immunology 1992;10:835-73.

14. Weenink SM, Gautam AM. Antigen presentation by MHC class II molecules. Immunology & Cell Biology 1997;75(1):69-81.

15. Watts C. Capture and processing of exogenous antigens for presentation on MHC molecules. Annual Review of Immunology 1997;15:821-50.

16. Castellino F, Zhong G, Germain RN. Antigen presentation by MHC class II molecules: invariant chain function, protein trafficking, and the molecular basis of diverse determinant capture. Human Immunology 1997;54(2):159-69.

17. Solheim JC. Class I MHC molecules: assembly and antigen presentation. Immunological Reviews 1999;172:11-9.

18. Momburg F, Hammerling GJ. Generation and TAP-mediated transport of peptides for major histocompatibility complex class I molecules. Advances in Immunology 1998;68:191-256.

19. Elliott T. How does TAP associate with MHC class I molecules? Immunology Today 1997;18(8):375-9.

20. Shoukry NH, Lavoie PM, Thibodeau J, D'Souza S, Sekaly RP. MHC class II-dependent peptide antigen versus superantigen presentation to T cells. Human Immunology 1997;54(2):194-201.

21. Choi YW, Herman A, DiGiusto D, Wade T, Marrack P, Kappler J. Residues of the variable region of the T-cell-receptor beta-chain that interact with S. aureus toxin superantigens. Nature 1990;346(6283):471-3.

22. Pullen AM, Wade T, Marrack P, Kappler JW. Identification of the region of T cell receptor beta chain that interacts with the self-superantigen Mls-1a. Cell 1990;61(7):1365-74.

23. Pullen AM, Bill J, Kubo RT, Marrack P, Kappler JW. Analysis of the interaction site for the self superantigen Mls-1a on T cell receptor V beta. Journal of Experimental Medicine 1991;173(5):1183-92.

24. MacDonald HR, Lees RK, Baschieri S, Herrmann T, Lussow AR. Peripheral T-cell reactivity to bacterial superantigens in vivo: the response/anergy paradox. Immunological Reviews 1993;133:105-17.

25. Chvatchko Y, MacDonald HR. CD8+ T cell response to Mls-1a determinants involves major histocompatibility complex class II molecules. Journal of Experimental Medicine 1991;173(3):779-82.

26. Woodland DL, Blackman MA. The fringes of superantigen reactivity: what do they tell us? Research in Immunology 1993;144(3):202-5.

27. Deckhut AM, Chien Y, Blackman MA, Woodland DL. Evidence for a functional interaction between the beta chain of major histocompatibility complex class II and the T cell receptor alpha chain during recognition of a bacterial superantigen. Journal of Experimental Medicine 1994;180(5):1931-5.

28. Daly K, Nguyen P, Hankley D, Zhang WJ, Woodland DL, Blackman MA. Contribution of the TCR alpha-chain to the differential recognition of bacterial and retroviral superantigens. Journal of Immunology 1995;155(1):27-34.

29. Blackman MA, Woodland DL. Role of the T cell receptor alpha-chain in superantigen recognition. Immunologic Research 1996;15(2):98-113.

30. Klein J, Sato A. The HLA system. First of two parts. New England Journal of Medicine 2000;343(10):702-9.

31. Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Chi YI, et al. Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. Nature 1994;368(6473):711-8.

32. Wen R, Blackman MA, Woodland DL. Variable influence of MHC polymorphism on the recognition of bacterial superantigens by T cells. Journal of Immunology 1995;155(4):1884-92.

33. Blackman MA, Lund FE, Surman S, Corley RB, Woodland DL. Major histocompatibility complex-restricted recognition of retroviral superantigens by V beta 17+ T cells. Journal of Experimental Medicine 1992;176(1):275-80.

34. White J, Herman A, Pullen AM, Kubo R, Kappler JW, Marrack P. The V beta-specific superantigen staphylococcal enterotoxin B: stimulation of mature T cells and clonal deletion in neonatal mice. Cell 1989;56(1):27-35.

35. Choi YW, Kotzin B, Herron L, Callahan J, Marrack P, Kappler J. Interaction of Staphylococcus aureus toxin "superantigens" with human T cells. Proceedings of the National Academy of Sciences of the United States of America 1989;86(22):8941-5.

36. Bohach GA, Fast DJ, Nelson RD, Schlievert PM. Staphylococcal and streptococcal pyrogenic toxins involved in toxic shock syndrome and related illnesses. Critical Reviews in Microbiology 1990;17(4):251-72.

37. Bohach GA, Handley JP, Schlievert PM. Biological and immunological properties of the carboxyl terminus of staphylococcal enterotoxin C1. Infection & Immunity 1989;57(1):23-8.

38. Hauser AR, Schlievert PM. Nucleotide sequence of the streptococcal pyrogenic exotoxin type B gene and relationship between the toxin and the streptococcal proteinase precursor. Journal of Bacteriology 1990;172(8):4536-42.

39. Callahan JE, Herman A, Kappler JW, Marrack P. Stimulation of B10.BR T cells with superantigenic staphylococcal toxins. Journal of Immunology 1990;144(7):2473-9.

40. Kappler J, Kotzin B, Herron L, Gelfand EW, Bigler RD, Boylston A, et al. V beta-specific stimulation of human T cells by staphylococcal toxins. Science 1989;244(4906):811-3.

41. Mollick JA, McMasters RL, Grossman D, Rich RR. Localization of a site on bacterial superantigens that determines T cell receptor beta chain specificity. Journal of Experimental Medicine 1993;177(2):283-93.

42. Hudson KR, Robinson H, Fraser JD. Two adjacent residues in staphylococcal enterotoxins A and E determine T cell receptor V beta specificity. Journal of Experimental Medicine 1993;177(1):175-84.

43. Abe J, Forrester J, Nakahara T, Lafferty JA, Kotzin BL, Leung DY. Selective stimulation of human T cells with streptococcal erythrogenic toxins A and B. Journal of Immunology 1991;146(11):3747-50.

44. Fischer H, Dohlsten M, Lindvall M, Sjogren HO, Carlsson R. Binding of staphylococcal enterotoxin A to HLA-DR on B cell lines. Journal of Immunology 1989;142(9):3151-7.

45. Scholl P, Diez A, Mourad W, Parsonnet J, Geha RS, Chatila T. Toxic shock syndrome toxin 1 binds to major histocompatibility complex class II molecules. Proceedings of the National Academy of Sciences of the United States of America 1989;86(11):4210-4.

46. Fraser JD. High-affinity binding of staphylococcal enterotoxins A and B to HLA-DR. Nature 1989;339(6221):221-3.

47. Scholl PR, Diez A, Geha RS. Staphylococcal enterotoxin B and toxic shock syndrome toxin-1 bind to distinct sites on HLA-DR and HLA-DQ molecules. Journal of Immunology 1989;143(8):2583-8.

48. Herman A, Labrecque N, Thibodeau J, Marrack P, Kappler JW, Sekaly RP. Identification of the staphylococcal enterotoxin A superantigen binding site in the beta 1 domain of the human histocompatibility antigen HLA-DR. Proceedings of the National Academy of Sciences of the United States of America 1991;88(22):9954-8.

49. Karp DR, Teletski CL, Scholl P, Geha R, Long EO. The alpha 1 domain of the HLA-DR molecule is essential for high-affinity binding of the toxic shock syndrome toxin-1. Nature 1990;346(6283):474-6.

50. Kappler JW, Herman A, Clements J, Marrack P. Mutations defining functional regions of the superantigen staphylococcal enterotoxin B. Journal of Experimental Medicine 1992;175(2):387-96.

51. Irwin MJ, Hudson KR, Fraser JD, Gascoigne NR. Enterotoxin residues determining T-cell receptor V beta binding specificity. Nature 1992;359(6398):841-3.

52. Chothia C, Boswell DR, Lesk AM. The outline structure of the T-cell alpha beta receptor. EMBO Journal 1988;7(12):3745-55.

53. White J, Pullen A, Choi K, Marrack P, Kappler JW. Antigen recognition properties of mutant V beta 3+ T cell receptors are consistent with an immunoglobulin-like structure for the receptor. Journal of Experimental Medicine 1993;177(1):119-25.

54. McCormack JE, Callahan JE, Kappler J, Marrack PC. Profound deletion of mature T cells in vivo by chronic exposure to exogenous superantigen. Journal of Immunology 1993;150(9):3785-92.

55. Nagata S. Apoptosis by death factor. Cell 1997;88(3):355-65.

56. Mogil RJ, Radvanyi L, Gonzalez-Quintial R, Miller R, Mills G, Theofilopoulos AN, et al. Fas (CD95) participates in peripheral T cell deletion and associated apoptosis in vivo. International Immunology 1995;7(9):1451-8.

57. Miethke T, Vabulas R, Bittlingmaier R, Heeg K, Wagner H. Mechanisms of peripheral T cell deletion: anergized T cells are Fas resistant but undergo proliferation-associated apoptosis. European Journal of Immunology 1996;26(7):1459-67.

58. Baltimore D. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. Cell 1985;40(3):481-2.

59. Felber BK, Paskalis H, Kleinman-Ewing C, Wong-Staal F, Pavlakis GN. The pX protein of HTLV-I is a transcriptional activator of its long terminal repeats. Science 1985;229(4714):675-9.

60. Cullen BR. HIV-1 auxiliary proteins: making connections in a dying cell. Cell 1998;93(5):685-92.

61. Acha-Orbea H. Retroviral superantigens. Chemical Immunology 1992;55:65-86.

62. Choi Y, Kappler JW, Marrack P. A superantigen encoded in the open reading frame of the 3' long terminal repeat of mouse mammary tumour virus. Nature 1991;350(6315):203-7.

63. Festenstein H. Immunogenetic and biological aspects of in vitro lymphocyte allotransformation (MLR) in the mouse. Transplantation Reviews 1973;15:62-88.

64. Woodland DL, Lund FE, Happ MP, Blackman MA, Palmer E, Corley RB. Endogenous superantigen expression is controlled by mouse mammary tumor proviral loci. Journal of Experimental Medicine 1991;174(5):1255-8.

65. Beutner U, Frankel WN, Cote MS, Coffin JM, Huber BT. Mls-1 is encoded by the long terminal repeat open reading frame of the mouse mammary tumor provirus Mtv-7. Proceedings of the National Academy of Sciences of the United States of America 1992;89(12):5432-6.

66. Pullen AM, Choi Y, Kushnir E, Kappler J, Marrack P. The open reading frames in the 3' long terminal repeats of several mouse mammary tumor virus integrants encode V beta 3-specific superantigens. Journal of Experimental Medicine 1992;175(1):41-7.

67. Woodland D, Happ MP, Bill J, Palmer E. Requirement for cotolerogenic gene products in the clonal deletion of I-E reactive T cells. Science 1990;247(4945):964-7.

68. Acha-Orbea H, Shakhov AN, Scarpellino L, Kolb E, Muller V, Vessaz-Shaw A, et al. Clonal deletion of V beta 14-bearing T cells in mice transgenic for mammary tumour virus. Nature 1991;350(6315):207-11.

69. Henrard D, Ross SR. Endogenous mouse mammary tumor virus is expressed in several organs in addition to the lactating mammary gland. Journal of Virology 1988;62(8):3046-9.

70. Kappler JW, Staerz U, White J, Marrack PC. Self-tolerance eliminates T cells specific for Mls-modified products of the major histocompatibility complex. Nature 1988;332(6159):35-40.

71. Acha-Orbea H, Palmer E. Mls--a retrovirus exploits the immune system. Immunology Today 1991;12(10):356-61.

72. Held W, Shakhov AN, Izui S, Waanders GA, Scarpellino L, MacDonald HR, et al. Superantigen-reactive CD4+ T cells are required to stimulate B cells after infection with mouse mammary tumor virus. Journal of Experimental Medicine 1993;177(2):359-66.

73. Choi Y, Marrack P, Kappler JW. Structural analysis of a mouse mammary tumor virus superantigen. Journal of Experimental Medicine 1992;175(3):847-52.

74. Korman AJ, Bourgarel P, Meo T, Rieckhof GE. The mouse mammary tumour virus long terminal repeat encodes a type II transmembrane glycoprotein. EMBO Journal 1992;11(5):1901-5.

75. Knight AM, Harrison GB, Pease RJ, Robinson PJ, Dyson PJ. Biochemical analysis of the mouse mammary tumor virus long terminal repeat product. Evidence for the molecular structure of an endogenous superantigen. European Journal of Immunology 1992;22(3):879-82.

76. Donehower LA, Fleurdelys B, Hager GL. Further evidence for the protein coding potential of the mouse mammary tumor virus long terminal repeat: nucleotide sequence of an endogenous proviral long terminal repeat. Journal of Virology 1983;45(3):941-9.

77. Crouse CA, Pauley RJ. Molecular cloning and sequencing of the MTV-1 LTR: evidence for a LTR sequence alteration. Virus Research 1989;12(2):123-37.

78. King LB, Lund FE, White DA, Sharma S, Corley RB. Molecular events in B lymphocyte differentiation. Inducible expression of the endogenous mouse mammary tumor proviral gene, Mtv-9. Journal of Immunology 1990;144(8):3218-27.

79. Rudy CK, Kraus E, Palmer E, Huber BT. Mls-1-like superantigen in the MA/MyJ mouse is encoded by a new mammary tumor provirus that is distinct from Mtv-7. Journal of Experimental Medicine 1992;175(6):1613-21.

80. Brandt-Carlson C, Butel JS, Wheeler D. Phylogenetic and structural analyses of MMTV LTR ORF sequences of exogenous and endogenous origins. Virology 1993;193(1):171-85.

81. Winslow GM, Scherer MT, Kappler JW, Marrack P. Detection and biochemical characterization of the mouse mammary tumor virus 7 superantigen (Mls-1a). Cell 1992;71(5):719-30.

82. Waite DJ, Sunshine GH. Neonatal tolerance induction to Mlsa. II. T cells induce tolerance to Mlsa. Cellular Immunology 1988;117(1):78-88.

83. Webb SR, Sprent J. Induction of neonatal tolerance to Mlsa antigens by CD8+ T cells. Science 1990;248(4963):1643-6.

84. Mazda O, Watanabe Y, Gyotoku J, Katsura Y. Requirement of dendritic cells and B cells in the clonal deletion of Mls-reactive T cells in the thymus. Journal of Experimental Medicine 1991;173(3):539-47.

85. Inaba M, Inaba K, Hosono M, Kumamoto T, Ishida T, Muramatsu S, et al. Distinct mechanisms of neonatal tolerance induced by dendritic cells and thymic B cells. Journal of Experimental Medicine 1991;173(3):549-59.

86. Metlay JP, Pure E, Steinman RM. Distinct features of dendritic cells and anti-Ig activated B cells as stimulators of the primary mixed leukocyte reaction. Journal of Experimental Medicine 1989;169(1):239-54.

87. Hamilos DL, Mascali JJ, Chesnut RW, Young RM, Ishioka G, Grey HM. The role of dendritic cells as stimulators of minor lymphocyte-stimulating locus-specific T cell responses in the mouse. I. Differential capacity of dendritic cells to stimulate minor lymphocyte-stimulating locus-reactive T cell hybridomas and the primary anti-minor lymphocyte-stimulating locus mixed lymphocyte reaction. Journal of Immunology 1989;142(4):1069-78.

88. Ramsdell F, Lantz T, Fowlkes BJ. A nondeletional mechanism of thymic self tolerance. Science 1989;246(4933):1038-41.

89. Roberts JL, Sharrow SO, Singer A. Clonal deletion and clonal anergy in the thymus induced by cellular elements with different radiation sensitivities. Journal of Experimental Medicine 1990;171(3):935-40.

90. Park CG, Jung MY, Choi Y, Winslow GM. Proteolytic processing is required for viral superantigen activity. Journal of Experimental Medicine 1995;181(5):1899-904.

91. Held W, Shakhov AN, Waanders G, Scarpellino L, Luethy R, Kraehenbuhl JP, et al. An exogenous mouse mammary tumor virus with properties of Mls-1a (Mtv-7). Journal of Experimental Medicine 1992;175(6):1623-33.

92. Huber BT. Mls superantigens: how retroviruses influence the expressed T cell receptor repertoire. Seminars in Immunology 1992;4(5):313-8.

93. Jouvin-Marche E, Cazenave PA, Voegtle D, Marche PN. V beta 17 T-cell deletion by endogenous mammary tumor virus in wild-type-derived mouse strain. Proceedings of the National Academy of Sciences of the United States of America 1992;89(8):3232-5.

94. Jouvin-Marche E, Marche PN, Six A, Liebe-Gris C, Voegtle D, Cazenave PA. Identification of an endogenous mammary tumor virus involved in the clonal deletion of V beta 2 T cells. European Journal of Immunology 1993;23(11):2758-64.

95. Mottershead DG, Hsu PN, Urban RG, Strominger JL, Huber BT. Direct binding of the Mtv7 superantigen (Mls-1) to soluble MHC class II molecules. Immunity 1995;2(2):149-54.

96. Torres BA, Griggs ND, Johnson HM. Bacterial and retroviral superantigens share a common binding region on class II MHC antigens. Nature 1993;364(6433):152-4.

97. Hainaut P, Francois C, Calberg-Bacq CM, Vaira D, Osterrieth PM. Peroral infection of suckling mice with milk-borne mouse mammary tumour virus: uptake of the main viral antigens by the gut. Journal of General Virology 1983;64(Pt 12):2535-48.

98. Tsubura A, Inaba M, Imai S, Murakami A, Oyaizu N, Yasumizu R, et al. Intervention of T-cells in transportation of mouse mammary tumor virus (milk factor) to mammary gland cells in vivo. Cancer Research 1988;48(22):6555-9.

99. Gunzburg WH, Salmons B. Factors controlling the expression of mouse mammary tumour virus. Biochemical Journal 1992;283(Pt 3):625-32.

100. Nusse R, Varmus HE. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell 1982;31(1):99-109.

101. Dickson C, Peters G. Proteins encoded by mouse mammary tumour virus. Current Topics in Microbiology & Immunology 1983;106:1-34.

102. Kozak C, Peters G, Pauley R, Morris V, Michalides R, Dudley J, et al. A standardized nomenclature for endogenous mouse mammary tumor viruses. Journal of Virology 1987;61(5):1651-4.

103. Tomonari K, Fairchild S, Rosenwasser OA. Influence of viral superantigens on V beta- and V alpha-specific positive and negative selection. Immunological Reviews 1993;131:131-68.

104. Hainaut P, Castellazzi M, Gonzales D, Clausse N, Hilgers J, Crepin M. A congenic line of the BALB/c mouse strain with the endogenous mouse mammary tumor virus proviral gene Mtv-3: tissue-specific expression and correlation with resistance to mouse mammary tumor virus infection and tumorigenesis. Cancer Research 1990;50(12):3754-60.

105. Golovkina TV, Chervonsky A, Dudley JP, Ross SR. Transgenic mouse mammary tumor virus superantigen expression prevents viral infection. Cell 1992;69(4):637-45.

106. Ferrara JL, Deeg HJ. Graft-versus-host disease. New England Journal of Medicine 1991;324(10):667-74.

107. Miconnet I, Roger T, Seman M, Bruley-Rosset M. Critical role of endogenous Mtv in acute lethal graft-versus-host disease. European Journal of Immunology 1995;25(2):364-8.

108. Jones MS, Riley R, Hamilton BL, Paupe J, Perez D, Levy RB. Endogenous superantigens in allogeneic bone marrow transplant recipients rapidly and selectively expand donor T cells which can produce IFN-gamma. Bone Marrow Transplantation 1994;14(5):725-35.

109. Sherer Y, Shoenfeld Y. Autoimmune diseases and autoimmunity post-bone marrow transplantation. Bone Marrow Transplantation 1998;22(9):873-81.

110. Parkman R. Is chronic graft versus host disease an autoimmune disease? Current Opinion in Immunology 1993;5(5):800-3.

111. Jurka J. Repeats in genomic DNA: mining and meaning. Current Opinion in Structural Biology 1998;8(3):333-7.

112. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.

113. Leib-Mosch C, Brack-Werner R, Werner T, Bachmann M, Faff O, Erfle V, et al. Endogenous retroviral elements in human DNA. Cancer Research 1990;50(17 Suppl):5636S-5642S.

114. Friedlander A, Patarca R. Endogenous proviruses. Critical Reviews in Oncogenesis 1999;10(1-2):129-59.

115. Noda M, Kurihara M, Takano T. Retrovirus-related sequences in human DNA: detection and cloning of sequences which hybridize with the long terminal repeat of baboon endogenous virus. Nucleic Acids Research 1982;10(9):2865-78.

116. O'Connell C, O'Brien S, Nash WG, Cohen M. ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. Virology 1984;138(2):225-35.

117. Callahan R, Drohan W, Tronick S, Schlom J. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proceedings of the National Academy of Sciences of the United States of America 1982;79(18):5503-7.

118. May FE, Westley BR, Rochefort H, Buetti E, Diggelmann H. Mouse mammary tumour virus related sequences are present in human DNA. Nucleic Acids Research 1983;11(12):4127-39.

119. Westley B, May FE. The human genome contains multiple sequences of varying homology to mouse mammary tumour virus DNA. Gene 1984;28(2):221-7.

120. May FE, Westley BR. Structure of a human retroviral sequence related to mouse mammary tumor virus. Journal of Virology 1986;60(2):743-9.

121. Deen KC, Sweet RW. Murine mammary tumor virus pol-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus pol gene. Journal of Virology 1986;57(2):422-32.

122. Martin MA, Bryan T, Rasheed S, Khan AS. Identification and cloning of endogenous retroviral sequences present in human DNA. Proceedings of the National Academy of Sciences of the United States of America 1981;78(8):4892-6.

123. Repaske R, O'Neill RR, Steele PE, Martin MA. Characterization and partial nucleotide sequence of endogenous type C retrovirus segments in human chromosomal DNA. Proceedings of the National Academy of Sciences of the United States of America 1983;80(3):678-82.

124. Bonner TI, O'Connell C, Cohen M. Cloned endogenous retroviral sequences from human DNA. Proceedings of the National Academy of Sciences of the United States of America 1982;79(15):4709-13.

125. O'Brien SJ, Bonner TI, Cohen M, O'Connell C, Nash WG. Mapping of an endogenous retroviral sequence to human chromosome 18. Nature 1983;303(5912):74-7.

126. Lower R, Lower J, Kurth R. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proceedings of the National Academy of Sciences of the United States of America 1996;93(11):5177-84.

127. Lindeskog M, Mager DL, Blomberg J. Isolation of a human endogenous retroviral HERV-H element with an open env reading frame. Virology 1999;258(2):441-50.

128. Blond JL, Beseme F, Duret L, Bouton O, Bedin F, Perron H, et al. Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. Journal of Virology 1999;73(2):1175-85.

129. Chiu IM, Callahan R, Tronick SR, Schlom J, Aaronson SA. Major pol gene progenitors in the evolution of oncoviruses. Science 1984;223(4634):364-70.

130. Ono M, Toh H, Miyata T, Awaya T. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. Journal of Virology 1985;55(2):387-94.

131. Franklin GC, Chretien S, Hanson IM, Rochefort H, May FE, Westley BR. Expression of human sequences related to those of mouse mammary tumor virus. Journal of Virology 1988;62(4):1203-10.

132. Ono M. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. Journal of Virology 1986;58(3):937-44.

133. Simpson GR, Patience C, Lower R, Tonjes RR, Moore HD, Weiss RA, et al. Endogenous D-type (HERV-K) related sequences are packaged into retroviral particles in the placenta and possess open reading frames for reverse transcriptase. Virology 1996;222(2):451-6.

134. Ono M, Yasunaga T, Miyata T, Ushikubo H. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. Journal of Virology 1986;60(2):589-98.

135. Ono M, Kawakami M, Ushikubo H. Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. Journal of Virology 1987;61(6):2059-62.

136. Seifarth W, Baust C, Murr A, Skladny H, Krieg-Schneider F, Blusch J, et al. Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. Journal of Virology 1998;72(10):8384-91.

137. Herbst H, Sauter M, Kuhler-Obbarius C, Loning T, Mueller-Lantzsch N. Human endogenous retrovirus (HERV)-K transcripts in germ cell and trophoblastic tumours. Apmis 1998;106(1):216-20.

138. Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N, Meese E. An almost-intact human endogenous retrovirus K on human chromosome 7. Nature Genetics 1999;21(3):257-8.

139. Tonjes RR, Limbach C, Lower R, Kurth R. Expression of human endogenous retrovirus type K envelope glycoprotein in insect and mammalian cells. Journal of Virology 1997;71(4):2747-56.

140. Tonjes RR, Boller K, Limbach C, Lugert R, Kurth R. Characterization of human endogenous retrovirus type K virus-like particles generated from recombinant baculoviruses. Virology 1997;233(2):280-91.

141. Rabson AB, Steele PE, Garon CF, Martin MA. mRNA transcripts related to full-length endogenous retroviral DNA in human cells. Nature 1983;306(5943):604-7.

142. Rabson AB, Hamagishi Y, Steele PE, Tykocinski M, Martin MA. Characterization of human endogenous retroviral envelope RNA transcripts. Journal of Virology 1985;56(1):176-82.

143. Gattoni-Celli S, Kirsch K, Kalled S, Isselbacher KJ. Expression of type C-related endogenous retroviral sequences in human colon tumors and colon cancer cell lines. Proceedings of the National Academy of Sciences of the United States of America 1986;83(16):6127-31.

144. Medstrand P, Lindeskog M, Blomberg J. Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. Journal of General Virology 1992;73(Pt 9):2463-6.

145. Witkin SS, Sarkar NH, Kinne DW, Breed CN, Good RA, Day NK. Antigens and antibodies cross-reactive to the murine mammary tumor virus in human breast cyst fluids. Journal of Clinical Investigation 1981;67(1):216-22.

146. Tomana M, Kajdos AH, Niedermeier W, Durkin WJ, Mestecky J. Antibodies to mouse mammary tumor virus-related antigen in sera of patients with breast carcinoma. Cancer 1981;47(11):2696-703.

147. Levine PH, Mesa-Tejada R, Keydar I, Tabbane F, Spiegelman S, Mourali N. Increased incidence of mouse mammary tumor virus-related antigen in Tunisian patients with breast cancer. International Journal of Cancer 1984;33(3):305-8.

148. Litvinov S, Kryukova I, Hint E, Purde M. Antibodies to MuMTV proteins in the sera of mammary carcinoma patients' healthy daughters. Archiv fur Geschwulstforschung 1986;56(6):407-12.

149. Indraccolo S, Gunzburg WH, Leib-Mosch C, Erfle V, Salmons B. Identification of three human sequences with viral superantigen-specific primers. Mammalian Genome 1995;6(5):339-44.

150. Maeda N, Kim HS. Three independent insertions of retrovirus-like sequences in the haptoglobin gene cluster of primates. Genomics 1990;8(4):671-83.

151. Labrecque N, McGrath H, Subramanyam M, Huber BT, Sekaly RP. Human T cells respond to mouse mammary tumor virus-encoded superantigen: V beta restriction and conserved evolutionary features. Journal of Experimental Medicine 1993;177(6):1735-43.

152. Subramanyam M, McLellan B, Labrecque N, Sekaly RP, Huber BT. Presentation of the Mls-1 superantigen by human HLA class II molecules to murine T cells. Journal of Immunology 1993;151(5):2538-45.

153. Cho K, Ferrick DA, Morris DW. Structure and biological activity of the subgenomic Mtv-6 endogenous provirus. Virology 1995;206(1):395-402.

154. Krieg AM, Gourley MF, Perl A. Endogenous retroviruses: potential etiologic agents in autoimmunity. FASEB Journal 1992;6(8):2537-44.

155. Youinou P, Saraux A, Le Goff P, Le Corre R. Superantigens in autoimmune disease. Revue Du Rhumatisme, English Edition 1995;62(9):591-7.

156. Talal N. Overview of Sjogren's syndrome. Journal of Dental Research 1987;66 Spec No:672-4.

157. Oxholm P, Manthorpe R, Prause JU. Polyclonal B-cell activation is related to exocrine manifestations of primary Sjogren's syndrome. Journal of Autoimmunity 1989;2(4):515-20.

158. Harley JB. Autoantibodies in Sjogren's syndrome. Journal of Autoimmunity 1989;2(4):383-94.

159. Sjogren H. Zur kenntnis der keratoconjunctivitis sicca (The subject of keratoconjunctivitis sicca). Archives of Ophthlology (Kobnhaven) 1933;11(2):1-151.

160. Fox RI, Adamson TCd, Fong S, Young C, Howell FV. Characterization of the phenotype and function of lymphocytes infiltrating the salivary gland in patients with primary Sjogren syndrome. Diagnostic Immunology 1983;1(3):233-9.

161. Lindahl G, Hedfors E, Klareskog L, Forsum U. Epithelial HLA-DR expression and T lymphocyte subsets in salivary glands in Sjogren's syndrome. Clinical & Experimental Immunology 1985;61(3):475-82.

162. Bariffi F, Pesci A, Bertorelli G, Manganelli P, Ambanelli U. Pulmonary involvement in Sjogren's syndrome. Respiration 1984;46(1):82-7.

163. Papiris SA, Saetta M, Turato G, La Corte R, Trevisani L, Mapp CE, et al. CD4-positive T-lymphocytes infiltrate the bronchial mucosa of patients with Sjogren's syndrome. American Journal of Respiratory & Critical Care Medicine 1997;156(2 Pt 1):637-41.

164. Davidson BK, Kelly CA, Griffiths ID. Ten year follow up of pulmonary function in patients with primary Sjogren's syndrome. Annals of the Rheumatic Diseases 2000;59(9):709-12.

165. Moutsopoulos HM, Cledes J, Skopouli FN, Elisaf M, Youinou P. Nephrocalcinosis in Sjogren's syndrome: a late sequela of renal tubular acidosis. Journal of Internal Medicine 1991;230(2):187-91.

166. Aasarod K, Haga HJ, Berg KJ, Hammerstrom J, Jorstad S. Renal involvement in primary Sjogren's syndrome. Qjm 2000;93(5):297-304.

167. Coll J, Navarro S, Tomas R, Elena M, Martinez E. Exocrine pancreatic function in Sjogren's syndrome. Archives of Internal Medicine 1989;149(4):848-52.

168. Ostuni PA, Gazzetto G, Chieco-Bianchi F, Riga B, Plebani M, Betterle C, et al. Pancreatic exocrine involvement in primary Sjogren's syndrome. Scandinavian Journal of Rheumatology 1996;25(1):47-51.

169. Oxholm P, Oxholm A, Manthorpe R. Diagnostic significance of immunohistological skin examination in patients with primary Sjogren's syndrome and other chronic inflammatory connective tissue diseases. Scandinavian Journal of Rheumatology - Supplement 1986;61:173-6.

170. Velthuis PJ, Nieboer C, Kater L, Hene RJ. A prospective immunofluorescence study of immune deposits in the skin of primary Sjogren's syndrome. Acta Dermato-Venereologica 1989;69(6):487-91.

171. Jordan RC, Speight PM. Lymphoma in Sjogren's syndrome. From histopathology to molecular pathology. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, & Endodontics 1996;81(3):308-20.

172. Andonopoulos AP, Drosos AA, Skopouli FN, Acritidis NC, Moutsopoulos HM. Secondary Sjogren's syndrome in rheumatoid arthritis. Journal of Rheumatology 1987;14(6):1098-103.

173. Nossent JC, Swaak AJ. Systemic lupus erythematosus VII: frequency and impact of secondary Sjogren's syndrome. Lupus 1998;7(4):231-4.

174. Alarcon-Segovia D, Ibanez G, Hernandez-Ortiz J, Velazquez-Forero F, Gonzalez-Jimenez Y. Sjogren's syndrome in progressive systemic sclerosis (scleroderma). American Journal of Medicine 1974;57(1):78-85.

175. Ringel SP, Forstot JZ, Tan EM, Wehling C, Griggs RC, Butcher D. Sjogren's syndrome and polymyositis or dermatomyositis. Archives of Neurology 1982;39(3):157-63.

176. Ohtsuka E, Nonaka S, Shingu M, Yasuda M, Nobunaga M. Sjogren's syndrome and mixed connective tissue disease. Clinical & Experimental Rheumatology 1992;10(4):339-44.

177. Foster H, Fay A, Kelly C, Charles P, Walker D, Griffiths I. Thyroid disease and other autoimmune phenomena in a family study of primary Sjogren's syndrome. British Journal of Rheumatology 1993;32(1):36-40.

178. Karsh J, Pavlidis N, Weintraub BD, Moutsopoulos HM. Thyroid disease in Sjogren's syndrome. Arthritis & Rheumatism 1980;23(11):1326-9.

179. Tsianos EV, Hoofnagle JH, Fox PC, Alspaugh M, Jones EA, Schafer DF, et al. Sjogren's syndrome in patients with primary biliary cirrhosis. Hepatology 1990;11(5):730-4.

180. Vitali C, Bombardieri S, Moutsopoulos HM, Balestrieri G, Bencivelli W, Bernstein RM, et al. Preliminary criteria for the classification of Sjogren's syndrome. Results of a prospective concerted action supported by the European Community. Arthritis & Rheumatism 1993;36(3):340-7.

181. Vitali C, Bombardieri S, Moutsopoulos HM, Coll J, Gerli R, Hatron PY, et al. Assessment of the European classification criteria for Sjogren's syndrome in a series of clinically defined cases: results of a prospective multicentre study. The European Study Group on Diagnostic Criteria for Sjogren's Syndrome. Annals of the Rheumatic Diseases 1996;55(2):116-21.

182. Tan EM. Autoantibodies to nuclear antigens (ANA): their immunobiology and medicine. Advances in Immunology 1982;33:167-240.

183. Hay EM, Freemont AJ, Kay RA, Bernstein RM, Holt PJ, Pumphrey RS. Selective polyclonal increase of immunoglobulin G1 subclass: a link with Sjogren's syndrome. Annals of the Rheumatic Diseases 1990;49(6):373-7.

184. Blaylock WM, Waller M, Normansell DE. Sjogren's syndrome: hyperviscosity and intermediate complexes. Annals of Internal Medicine 1974;80(1):27-34.

185. Alarcon-Segovia D, Fishbein E, Abruzzo JL, Heimer R. Serum hyperviscosity in Sjogren's syndrome interaction between IgG and IgG rheumatoid factor. Annals of Internal Medicine 1974;80(1):35-42.

186. Ramos-Casals M, Cervera R, Yague J, Garcia-Carrasco M, Trejo O, Jimenez S, et al. Cryoglobulinemia in primary Sjogren's syndrome: prevalence and clinical characteristics in a series of 115 patients. Seminars in Arthritis & Rheumatism 1998;28(3):200-5.

187. Kelly CA, Foster H, Pal B, Gardiner P, Malcolm AJ, Charles P, et al. Primary Sjogren's syndrome in north east England--a longitudinal study. British Journal of Rheumatology 1991;30(6):437-42.

188. Davidson BK, Kelly CA, Griffiths ID. Primary Sjogren's syndrome in the North East of England: a long-term follow-up study. Rheumatology (Oxford) 1999;38(3):245-53.

189. Isenberg DA. Autoantibodies: markers of disease or pathogenic? Annals of the New York Academy of Sciences 1997;823:256-62.

190. Cartier L, Castillo JL, Cea JG, Villagra R. Chronic dacryosialadenitis in HTLV I associated myelopathy. Journal of Neurology, Neurosurgery & Psychiatry 1995;58(2):244-6.

191. Nakamura H, Eguchi K, Nakamura T, Mizokami A, Shirabe S, Kawakami A, et al. High prevalence of Sjogren's syndrome in patients with HTLV-I associated myelopathy. Annals of the Rheumatic Diseases 1997;56(3):167-72.

192. Atkinson JC, Schiodt M, Robataille S, Greenspan D, Greenspan JS, Fox PC. Salivary autoantibodies in HIV-associated salivary gland disease. Journal of Oral Pathology & Medicine 1993;22(5):203-6.

193. Kordossis T, Paikos S, Aroni K, Kitsanta P, Dimitrakopoulos A, Kavouklis E, et al. Prevalence of Sjogren's-like syndrome in a cohort of HIV-1-positive patients: descriptive pathology and immunopathology. British Journal of Rheumatology 1998;37(6):691-5.

194. Talal N, Dauphinee MJ, Dang H, Alexander SS, Hart DJ, Garry RF. Detection of serum antibodies to retroviral proteins in patients with primary Sjogren's syndrome (autoimmune exocrinopathy). Arthritis & Rheumatism 1990;33(6):774-81.

195. Cattaneo R, Tosoni C, Franzini M, Manca N, Gelmi M, Turano A, et al. Antibodies to retroviral proteins in Sjogren's syndrome. Annals of the Rheumatic Diseases 1993;52(8):621.

196. Coll J, Palazon J, Yazbeck H, Gutierrez J, Aubo C, Benito P, et al. Antibodies to human immunodeficiency virus (HIV-1) in autoimmune diseases: primary Sjogren's syndrome, systemic lupus erythematosus, rheumatoid arthritis and autoimmune thyroid diseases. Clinical Rheumatology 1995;14(4):451-7.

197. Garry RF, Fermin CD, Hart DJ, Alexander SS, Donehower LA, Luo-Zhang H. Detection of a human intracisternal A-type retroviral particle antigenically related to HIV. Science 1990;250(4984):1127-9.

198. Guerin M. Corps d'inclusion dans les adenocarcinomes mammaires de la souris. Bullitin of Cancer 1955;42:14-28.

199. Nusse R, van der Ploeg L, van Duijn L, Michalides R, Hilgers J. Impaired maturation of mouse mammary tumor virus precursor polypeptides in lymphoid leukemia cells, producing intracytoplasmic A particles and no extracellular B-type virions. Journal of Virology 1979;32(1):251-8.

200. Brookes SM, Pandolfino YA, Mitchell TJ, Venables PJ, Shattles WG, Clark DA, et al. The immune response to and expression of cross-reactive retroviral gag sequences in autoimmune disease. British Journal of Rheumatology 1992;31:735-42.

201. Shattles WG, Brookes SM, Venables PJ, Clark DA, Maini RN. Expression of antigen reactive with a monoclonal antibody to HTLV-1 P19 in salivary glands in Sjogren's syndrome. Clinical & Experimental Immunology 1992;89(1):46-51.

202. Mariette X, Agbalika F, Daniel MT, Bisson M, Lagrange P, Brouet JC, et al. Detection of human T lymphotropic virus type I tax gene in salivary gland epithelium from two patients with Sjogren's syndrome. Arthritis & Rheumatism 1993;36:1423-8.

203. Sumida T, Yonaha F, Maeda T, Kita Y, Iwamoto I, Koike T, et al. Expression of sequences homologous to HTLV-I tax gene in the labial salivary glands of Japanese patients with Sjogren's syndrome. Arthritis & Rheumatism 1994;37(4):545-50.

204. Zinkernagel RM, Cooper S, Chambers J, Lazzarini RA, Hengartner H, Arnheiter H. Virus-induced autoantibody response to a transgenic viral antigen. Nature 1990;345(6270):68-71.

205. Green JE, Hinrichs SH, Vogel J, Jay G. Exocrinopathy resembling Sjogren's syndrome in HTLV-1 tax transgenic mice. Nature 1989;341(6237):72-4.

206. Sodroski J. The human T-cell leukemia virus (HTLV) transactivator (Tax) protein. Biochimica et Biophysica Acta 1992;1114(1):19-29.

207. Leung K, Nabel GJ. HTLV-1 transactivator induces interleukin-2 receptor expression through an NF-kappa B-like factor. Nature 1988;333(6175):776-8.

208. Maruyama M, Shibuya H, Harada H, Hatakeyama M, Seiki M, Fujita T, et al. Evidence for aberrant activation of the interleukin-2 autocrine loop by HTLV-1-encoded p40x and T3/Ti complex triggering. Cell 1987;48(2):343-50.

209. Arai N, Nomura D, Villaret D, DeWaal Malefijt R, Seiki M, Yoshida M, et al. Complete nucleotide sequence of the chromosomal gene for human IL-4 and its expression. Journal of Immunology 1989;142(1):274-82.

210. Griffiths DJ, Venables PJ, Weiss RA, Boyd MT. A novel exogenous retrovirus sequence identified in humans. Journal of Virology 1997;71(4):2866-72.

211. Rigby SP, Griffiths DJ, Weiss RA, Venables PJ. Human retrovirus-5 proviral DNA is rarely detected in salivary gland biopsy tissues from patients with Sjogren's syndrome. Arthritis & Rheumatism 1997;40(11):2016-21.

212. Griffiths DJ, Cooke SP, Herve C, Rigby SP, Mallon E, Hajeer A, et al. Detection of human retrovirus 5 in patients with arthritis and systemic lupus erythematosus. Arthritis & Rheumatism 1999;42(3):448-54.

213. Perron H, Garson JA, Bedin F, Beseme F, Paranhos-Baccala G, Komurian-Pradel F, et al. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. The Collaborative Research Group on Multiple Sclerosis. Proceedings of the National Academy of Sciences of the United States of America 1997;94(14):7583-8.

214. Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, et al. 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. Journal of Virology 2000;74(7):3321-9.

215. Conrad B, Weissmahr RN, Boni J, Arcari R, Schupbach J, Mach B. A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. Cell 1997;90(2):303-13.

216. Sumida T, Yonaha F, Maeda T, Tanabe E, Koike T, Tomioka H, et al. T cell receptor repertoire of infiltrating T cells in lips of Sjogren's syndrome patients. Journal of Clinical Investigation 1992;89(2):681-5.

217. Yonaha F, Sumida T, Maeda T, Tomioka H, Koike T, Yoshida S. Restricted junctional usage of T cell receptor V beta 2 and V beta 13 genes, which are overrepresented on infiltrating T cells in the lips of patients with Sjogren's syndrome. Arthritis & Rheumatism 1992;35(11):1362-7.

218. Mizushima N, Kohsaka H, Tsubota K, Saito I, Miyasaka N. Diverse T cell receptor beta gene usage by infiltrating T cells in the lacrimal glands of Sjogren's syndrome. Clinical & Experimental Immunology 1995;101(1):33-8.

219. Smith MD, Lamour A, Boylston A, Lancaster FC, Pennec YL, van Agthoven A, et al. Selective expression of V beta families by T cells in the blood and salivary gland infiltrate of patients with primary Sjogren's syndrome. Journal of Rheumatology 1994;21(10):1832-7.

220. Pivetta B, De Vita S, Ferraccioli G, De Re V, Gloghini A, Marzotto A, et al. T cell receptor repertoire in B cell lymphoproliferative lesions in primary Sjogren's syndrome. Journal of Rheumatology 1999;26(5):1101-9.

221. Murata H, Kita Y, Sakamoto A, Matsumoto I, Matsumura R, Sugiyama T, et al. Limited TCR repertoire of infiltrating T cells in the kidneys of Sjogren's syndrome patients with interstitial nephritis. Journal of Immunology 1995;155(8):4084-9.

222. Matsumoto I, Tsubota K, Satake Y, Kita Y, Matsumura R, Murata H, et al. Common T cell receptor clonotype in lacrimal glands and labial salivary glands from patients with Sjogren's syndrome. Journal of Clinical Investigation 1996;97(8):1969-77.

223. Li Y, Szabo P, Robinson MA, Dong B, Posnett DN. Allelic variations in the human T cell receptor V beta 6.7 gene products. Journal of Experimental Medicine 1990;171(1):221-30.

224. Robinson MA. Allelic sequence variations in the hypervariable region of a T-cell receptor beta chain: correlation with restriction fragment length polymorphism in human families and populations. Proceedings of the National Academy of Sciences of the United States of America 1989;86(23):9422-6.

225. Baccala R, Kono DH, Walker S, Balderas RS, Theofilopoulos AN. Genomically imposed and somatically modified human thymocyte V beta gene repertoires. Proceedings of the National Academy of Sciences of the United States of America 1991;88(7):2908-12.

226. Kay RA, Snowden N, Hajeer AH, Boylston AW, Ollier WE. Genetic control of the human V beta 13.2 T cell repertoire: importance of allelic variation outside the coding regions of the TCRBV13S2 gene. European Journal of Immunology 1994;24(11):2863-7.

227. Kay RA, Hutchings CJ, Ollier WE. A subset of Sjogren's syndrome associates with the TCRBV13S2 locus but not the TCRBV2S1 locus. Human Immunology 1995;42(4):328-30.

228. Clarke GR, Reyburn H, Lancaster FC, Boylston AW. Bimodal distribution of V beta 2+CD4+ T cells in human peripheral blood. European Journal of Immunology 1994;24(4):837-42.

229. Boylston AW, Clarke GR, Lancaster FC, Reyburn H. Evidence for an endogenous superantigen deleting human V beta 2 positive T-lymphocytes. Annals of the New York Academy of Sciences 1995;756:113-6.

230. Kennedy N, Knedlitschek G, Groner B, Hynes NE, Herrlich P, Michalides R, et al. Long terminal repeats of endogenous mouse mammary tumour virus contain a long open reading frame which extends into adjacent sequences. Nature 1982;295(5850):622-4.

231. Queen C, Korn LJ. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Research 1984;12(1 Pt 2):581-99.

232. Karlin S, Altschul SF. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proceedings of the National Academy of Sciences of the United States of America 1990;87(6):2264-8.

233. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America 1992;89(22):10915-9.

234. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 1997;25(17):3389-402.

235. Fox RI, Saito I, Chan EK, Josephs S, Salahuddin SZ, Ahlashi DV, et al. Viral genomes in lymphomas of patients with Sjogren's syndrome. Journal of Autoimmunity 1989;2(4):449-55.

236. Flescher E, Talal N. Do viruses contribute to the development of Sjogren's syndrome? American Journal of Medicine 1991;90(3):283-5.

237. Yoshida M, Miyoshi I, Hinuma Y. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proceedings of the National Academy of Sciences of the United States of America 1982;79(6):2031-5.

238. Robert-Guroff M, Nakao Y, Notake K, Ito Y, Sliski A, Gallo RC. Natural antibodies to human retrovirus HTLV in a cluster of Japanese patients with adult T cell leukemia. Science 1982;215(4535):975-8.

239. Poiesz BJ, Ruscetti FW, Mier JW, Woods AM, Gallo RC. T-cell lines established from human T-lymphocytic neoplasias by direct response to T-cell growth factor. Proceedings of the National Academy of Sciences of the United States of America 1980;77(11):6815-9.

240. Jacobson S, Raine CS, Mingioli ES, McFarlin DE. Isolation of an HTLV-1-like retrovirus from patients with tropical spastic paraparesis. Nature 1988;331:540-3.

241. Rodgers-Johnson P, Gajdusek DC, Morgan OS, Zaninovic V, Sarin PS, Graham DS. HTLV-I and HTLV-III antibodies and tropical spastic paraparesis. Lancet 1985;2(8466):1247-8.

242. Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, et al. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet 1985;2(8452):407-10.

243. Reddy EP, Sandberg-Wollheim M, Mettus RV, Ray PE, DeFreitas E, Koprowski H. Amplification and molecular cloning of HTLV-I sequences from DNA of multiple sclerosis patients. Science 1989;243(4890):529-33.

244. Nelson PN, Lever AM, Bruckner FE, Isenberg DA, Kessaris N, Hay FC. Polymerase chain reaction fails to incriminate exogenous retroviruses HTLV-I and HIV-1 in rheumatological diseases although a minority of sera cross react with retroviral antigens. Annals of the Rheumatic Diseases 1994;53(11):749-54.

245. Hinuma Y, Komoda H, Chosa T, Kondo T, Kohakura M, Takenaka T, et al. Antibodies to adult T-cell leukemia-virus-associated antigen (ATLA) in sera from patients with ATL and controls in Japan: a nation-wide sero-epidemiologic study. International Journal of Cancer 1982;29(6):631-5.

246. Kwok S, Ehrlich G, Poiesz B, Kalish R, Sninsky JJ. Enzymatic amplification of HTLV-I viral sequences from peripheral blood mononuclear cells and infected tissues. Blood 1988;72(4):1117-23.

247. Eguchi K, Matsuoka N, Ida H, Nakashima M, Sakai M, Sakito S, et al. Primary Sjogren's syndrome with antibodies to HTLV-I: clinical and laboratory features. Annals of the Rheumatic Diseases 1992;51(6):769-76.

248. Cohen JC, Majors JE, Varmus HE. Organization of mouse mammary tumor virus-specific DNA endogenous to BALB/c mice. Journal of Virology 1979;32:483-96.

249. Temin HM. Origin of retroviruses from cellular moveable genetic elements. Cell 1980;21(3):599-600.

250. Janeway CA, Jr. Are there cellular superantigens? Immunological Reviews 1993;131:189-200.

251. Venter JC, Adams MD, Eugene WM. The sequence of the human genome. Science 2001;291(5507):1304-1351.

# Appendix

## *Abbreviations*

| | |
|---|---|
| ATP | adenosine-5'-triphosphate |
| BSA | bovine serum albumin |
| bp | base pair |
| DNA | deoxyribonucleic acid |
| DTT | dithioerythritol |
| EDTA | ethylenediaminotetraacetic acid |
| kb | kilobase |
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| RNA | ribonucleic acid |
| RT-PCR | reverse transcriptase polymerase chain reaction |
| SDS | sodium dodecyl sulphate |
| SDW | sterile distilled water |
| TAE | Tris-HCl acetate EDTA buffer |
| TBE | Tris-HCl borate EDTA buffer |
| TBS | Tris buffered saline |
| Tm | melting temperature |