UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

School of Medicine

Immunogenetic analysis of aggressive B-cell malignancies

by

Katy McCann, MBiochem (Hons)

Thesis for the degree of Doctor of Philosophy

March 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

SCHOOL OF MEDICINE

Doctor of Philosophy

IMMUNOGENETIC ANALYSIS OF AGGRESSIVE B-CELL MALIGNANCIES

by Katy McCann

Normal B cells undergo a series of recombinatorial and mutational changes during differentiation that lead to unique sequences in the immunoglobulin (Ig) heavy and light chain variable (V) region genes. The defining event is the rearrangement of the heavy and light chain genes in the bone marrow (BM), with further diversification following interaction with antigen in the secondary lymphoid organs by means of somatic mutation and class switching. In the event of neoplastic transformation, these sequence modifications are preserved and feature in every cell of the tumour clone. Since B-cell malignancies typically have functional Ig gene rearrangements and express surface Ig, analysis of the Ig V genes can provide an independent marker for a clonal B-cell population and may help to establish the developmental stage at which neoplastic transformation occurred. Therefore, Ig V gene analysis can contribute to our understanding of the differentiation and development of malignant B cells. In this study, we focus on the Ig V genes of two Non-Hodgkin's lymphomas that together encompass ~80% of all B-cell malignancies: follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL).

An unusually high incidence of *N*-glycosylation motifs, introduced by somatic mutation, has been reported to feature in the V_H genes of FL and other germinal centre (GC)-associated malignancies, and contrasts that found for normal B cells. Our primary aim was to investigate the introduction of *N*-glycosylation sites in both the V_H and V_L genes of FL, in order to gain a complete view of the B-cell receptor. Here, we confirm that the most dramatic distinction between the Ig V genes of normal B cells and those of FL lies in the accumulation of motifs available for addition of oligosaccharides in the latter, where 94% of tumours acquired at least one novel *N*-glycosylation site. Furthermore, there appears to be a drive towards the acquisition of such sites, with 21% tumours continuing to accumulate *N*-glycosylation sites as a result of ongoing mutation. Using recombinant scFv proteins derived from FL, we next confirm that these sites are functional and reveal that the added glycans are commonly oligomannose. We further show that the oligomannose glycans confer the ability to bind specifically to recombinant mannose receptor. Our data support the concept that an interaction with a mannose-binding lectin expressed by stromal cells in the GC may play a role in lymphoma pathogenesis, possibly by providing specific signals for the growth and/or survival of these tumours within the GC microenvironment.

de novo DLBCLs represent a highly heterogeneous group of lymphoid neoplasms, and for which *N*-gylcosylation sites in the Ig V genes have been reported in ~40% of cases. Our second aim was to determine whether the introduction of *N*-glycosylation sites is restricted to the Ig V genes of tumours displaying a GCB-like phenotype. Here, we report that DLBCLs classified by cDNA/tissue microarray and iFISH analysis into $t(14;18)^+$-GCB-like, $t(14;18)^-$-GCB-like and ABC-like subgroups show an incidence of *N*-glycosylation sites of 50%, 56% and 50%, respectively. Therefore, our findings do not support a link between $t(14;18)^+$-GCB-like DLBCL and FL and argue against a common pathway for pathogenesis.

Primary CNS lymphomas (PCNSL) represent highly malignant DLBCLs, which arise within and are confined to the CNS in the absence of systemic disease. Little is known about the site of origin of the malignant lymphocyte that gives rise to these tumours. It has been suggested that clonal proliferation and transformation may take place in the periphery, followed by the specific homing of tumour cells to the CNS. The third aim of our study was to investigate the existence of tumour cells at peripheral sites in PCNSL patients at the time of diagnosis. Using the unique tumour-derived Ig V gene sequences as a marker we could track tumour cells in the BM and/or blood of 3 of 3 patients and further show the existence of tumour-related subclones unique to extracerebral sites. We hypothesise that, at some stage, a proportion of tumour cells in the periphery may have migrated to the brain, with the transformation event occurring either in the periphery or after entry into the CNS.

List of Contents

List of Figures

List of Tables

Publications and presentations

Publications (see Appendix III)

Parts of this work have been published in:

- **McCann, K.J.**, Sahota, S., Stevenson, F.K. and Ottensmeier, C.H. Idiotype gene rescue in follicular lymphoma. In: Illidge T, Johnson PW, eds. Lymphoma - Methods and Protocols. Totowa: Humana Press; 2005:145-172.

- **McCann, K.J.**, Johnson, P.W.M., Stevenson, F.K. and Ottensmeier, C.H. Universal *N*-glycosylation sites introduced into the B-cell receptor of follicular lymphoma by somatic mutation: A second tumourigenic event? Leukemia. 2006 Mar;20(3):530-534.

- **McCann, K.J.**, Ottensmeier, C.H., Callard, A., Radcliffe, C.M., Harvey, D.J., Dwek, R.A., Rudd, P.M., Sutton, B.J., Hobby, P. and Stevenson, F.K. Remarkable selective glycosylation of the variable region in follicular lymphoma. Molecular Immunology. 2008 Mar; 45(6):1567-1572

Papers currently in preparation:

- **McCann, K.J.**, Stevenson, F.K. and Ottensmeier, C.H. Primary CNS lymphoma cells can be detected in the blood and bone marrow of patients and show evidence for separate clonal development.

- **McCann K.J.**, Mander, A., Low, L., Stevenson, F.K. and Ottensmeier, C.H. A novel method for the expression of recombinant idiotypic protein for immune monitoring following DNA vaccination.

Presentations

Parts of this work have been presented at:

Oral presentations

- **McCann, K.J.**, Johnson, P.W.M., Stevenson, F.K. and Ottensmeier, C.H. Immunogenetic profiling of follicular lymphoma reveals universal *N*-glycosylation sites introduced into the B-cell receptor by somatic mutation and suggests relevance to lymphoma pathogenesis. Abstract 606. The American Society of Hematology 48th Annual Meeting and Exposition. Atlanta, Dec 2005.

Poster presentations

- **McCann, K.J.**, Johnson, P.W.M., Stevenson, F.K. and Ottensmeier, C.H. Universal *N*-glycosylation sites introduced into the B-cell receptor of follicular lymphoma by somatic mutation: A second tumourigenic event? Abstract P458. NCRI Cancer Conference. Birmingham, Oct 2005

Acknowledgements

I wish to extend my sincere thanks to my supervisors, Prof. Christian Ottensmeier and Prof. Freda Stevenson, for their continued guidance, support and encouragement throughout this Ph.D. I would also like to thank our collaborators: Mrs Catherine Radcliffe, who performed the *N*-glycan profile analysis; Dr. Luisa Martinez-Pomares for allowing us to use her recombinant mannose receptor and for providing us with an assay with which to study mannose receptor binding; Prof. Izidore Lossos and his team for performing the gene expression profiling; Dr. Margaret Ashton-Key, Miss Kellyann Smith, Mrs Jean Buontempo, Dr. Matthew Hickling and Dr. Ilske Oschlies, who together performed all the immunohistochemistry; and Mr Adam Stewart for helping with the interphase FISH analysis- thanks must also go to Prof. Christine Harrison and the LFR Cytogenetics Department for allowing me to use their equipment. Thanks are also extended to Prof. Brian Sutton and Dr. Paul Hobby for their expert knowledge and contribution with regards to the functional glycosylation of motifs. I also remain indebted to the Wessex Cancer Trust and CRUK for funding this work.

A special thank you must go to all the members of Christian and Freda's teams, of which there are too many to mention individually, who have had to listen to a lot of complaining from me during the course of this Ph.D., particularly in the last few months of writing up. However, Dr. Sarah Buchan and Miss Gisella Vittes deserve a special mention as close friends and fellow Ph.D. sufferers; thanks for all your support and encouragement.

I would like to express my gratitude to my mum and dad, for ensuring that I had this opportunity in the first place, and thank them for their unyielding support. Thank you also to the rest of my family for providing plenty of welcomed distractions (when desperately needed and usually involving cake!). Finally, I would like to sincerely thank Dr. Philip Hargreaves for his love and support over the years and for his constant (!) advice, which he will be only to quick to mention was not always heeded!

Abbreviations

κ	Kappa light chain
λ	Lambda light chain
aa	Amino acid
Ab	Antibody
ACF	APOBEC-1 complementation factor
ADCC	Antibody-mediated cellular cytotoxicity
Ag	Antigen
AID	Activation-induced cytidine deaminase
AIDS	Acquired immunodeficiency virus syndrome
ALL	Acute lymphoblastic leukaemia
APC	Antigen presenting cell
APE	Apurinic-apyrimidic endonuclease
apoB	Apolipoprotein B
APOBEC-1	Apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1
BAFF	B-cell activation factor of the TNF family
BAFF-R	BAFF receptor
BBB	Blood brain barrier
BCA-1	B cell attracting chemokine 1 (also known as CXCL13)
Bcl	B-cell CLL/lymphoma
BCMA	B-cell maturation antigen
BCR	B-cell receptor
BER	Base excision repair
BL	Burkitt's lymphoma; endemic (eBL) or sporadic (sBL)
Blimp-1	B-lymphocyte-induced maturation protein 1
BM	Bone marrow
BSA	Bovine serum albumin
BSAP	B-cell specific activator protein (also known as Pax 5)
CD	Cluster designation
CDR	Complementarity-determining region
cFLIP	Cellular FLICE-like inhibitory protein
C_H	Heavy chain constant region
cIg	Cytoplasmic immunoglobulin
C_L	Light chain constant region
CLL	Chronic lymphocytic leukaemia
CLP	Common lymphoid progenitor
CMV	Cytomegalovirus
CSR	Class switch recombination
C-terminal	Carboxy terminal
Ct	Threshold cycle
CXCL	CXC-chemokine ligand
cys	Cysteine
D gene	Diversity gene
DC	Dendritic cell
DC-SIGN	DC-specific ICAM3 grabbing non-integrin
dC	Deoxycytidine
Dectin-2	DC-associated lectin 2
DIR	D genes with irregular spacers
DLBCL	Diffuse large B-cell lymphoma
DMSO	Dimethylsulphoxide
DNA pol	DNA polymerase

DNA-PK	DNA-dependent protein kinase
DNA-PK_{CS}	DNA-dependent protein kinase catalytic subunit
DSB	DNA double-stranded break
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
dU	Deoxyuracil
Eκ	Immunoglobulin kappa enhancer
Eμ	Immunoglobulin mu enhancer
E. coli	Escherichia coli
EBF	Early B-cell factor
EBV	Epstein-barr virus
ELISA	Enzyme-linked immunosorbent assay
Endo H	Endoglycosidase H
Fab	Fragment, antibody binding
FACS	Fluorescence-activated cell sorting
FAM	6-carboxy fluorescein
Fc	Fragment, crystallisable
FDC	Follicular dendritic cell
FL	Follicular lymphoma
FR	Framework region
GC	Germinal centre
H chain	Heavy chain
HCL	Hairy cell leukaemia
HD	Hodgkin's disease; nodular lymphocyte predominant (NLPHD) or classical (CHD)
HHV	Human herpes virus
HIGM	Hyper-IgM syndrome
HIV	Human immunodeficiency virus
HMG	High-mobility-group protein
HR	Homologous recombination
HRS cells	Hodgkin's Reed-Sternberg cells
HSC	Haematopoietic stem cell
iFISH	Interphase fluorescence *in situ* hybridisation
I promoter	Intronic promoter
ICAM 1	Intercellular adhesion molecule 1
IDC	Interdigitating dendritic cell
IFN-γ	Interferon gamma
Ig C_H	Immunoglobulin heavy chain constant region locus
Ig	Immunoglobulin
Igκ	Immunoglobulin kappa locus
Igλ	Immunoglobulin lambda locus
IgH	Immunoglobulin heavy chain locus
IL	Interleukin
IPI	International Prognostic Index
IRF-4	Interferon regulatory factor 4 (also known as MUM-1)
ITAM	Immunoreceptor tyrosine based activation motif
J chain	Joining chain
J gene	Joining gene
Kb	Kilobases
Ku	Heterodimer composed of Ku70 and Ku80
L chain	Light chain
LB	Luria broth
LFA-1	Lymphocyte function associated antigen 1

LPL	Lymphoplasmacytoid lymphoma
mAb	Monoclonal antibody
MALT	Mucosa-associated lymphoid tissue
MBL	Mannose-binding lectin
MCL	Mantle cell lymphoma
MGB	Minor groove binder
MGUS	Monoclonal gammopathy of undetermined significance
MHC-I/II	Major histocompatibility complex class I/II
MM	Multiple myeloma
MMR	Mismatch repair
MR	Mannose receptor
MS	Multiple sclerosis
MSH	MutS homologue
MUM-1	Multiple myeloma oncogene 1 (also known as IRF-4)
MZL	Marginal zone lymphoma
N additions	Non-templated nucleotide additions
NF-κB	Nuclear factor kappa B
NHEJ	Non-homologous end-joining
NHL	Non-Hodgkin's lymphoma
NK	Natural killer cell
NP-HPLC	Normal phase high performance liquid chromatography
N-terminal	Amino terminal
ORF	Open reading frame
P additions	Palindromic nucleotide additions
PALS	Periarteriolar lymphatic sheath
Pax 5	Paired box gene 5 (also known as BSAP)
PBS	Phoshphate buffered saline
PCR	Polymerase chain reaction
PCNSL	Primary central nervous system lymphoma
PKA	Protein kinase A
PNGase F	Peptide:N-glycosidase F
Pre-BCR	Pre B-cell receptor
Pu	Purine
Py	Pyrimidine
qPCR	Quantitative real-time PCR
RAG 1/2	Recombination-activating gene 1/2
RE	Restriction enzyme
REAL	Revised European-American Classification of Lymphoid Neoplasms
RF	Reading frame
RNA pol	RNA polymerase
RPA	Replication protein A
RSS	Recombination signal sequence
RT	Room temperature
S region	Switch region
SAg	Superantigen
SCF	Stem cell factor
scFv	Single chain variable fragment
SCID	Complete severe combined immunodeficiency
SHM	Somatic hypermutation
sIg	Surface immunoglobulin
SMZL	Splenic marginal zone lymphoma
SOE PCR	Splicing by overlapping extension PCR
SSC	Saline sodium citrate
ssDNA	Single-stranded DNA

SV40	Simian virus 40
T1/2 B cell	Transitional B cell, type 1 and 2
TBS	Tris buffered saline
T_C cell	T cytotoxic cell
TdT	Terminal deoxynucleotidyl transferase
TGF-β	Transforming growth factor beta
$T_H1/2$ cell	T helper cell, type 1 and 2
UNG	Uracil N glycosylase
V gene	Variable gene
V region	Variable region
VCAM-1	Vascular cell adhesion molecule 1
V_H	Heavy chain variable region
V_L	Light chain variable region
VLA-4	Very late antigen 4
WBRT	Whole brain radiotherapy
WHO	World Health Organisation
WM	Waldenstrom macroglobulinemia
XBP-1	X-box-binding protein 1

1. Introduction

1.1. The immune system

The role of the immune system is to recognise and dispose of foreign antigens (Ags) and invading pathogens, such as viruses, bacteria, fungi, protozoa etc. Moreover, an increasing body of evidence shows that the immune system can also target tumour cells. There are two different mechanisms by which this is accomplished, namely innate and adaptive immunity, which are distinguished by the speed and specificity of their responses.[1-4]

Innate immunity is sometimes used to describe physical, chemical and microbiological barriers, but more usually encompasses the elements of the immune system that provide immediate host defence. Innate responses are broad-spectrum (non-specific) and use phagocytic cells (e.g. neutrophils, monocytes, and macrophages), cells that release inflammatory mediators (e.g. basophils, mast cells and eosinophils) and natural killer (NK) cells. The molecular components of these responses include the complement system, acute-phase proteins and cytokines, such as the interferons and interleukins. The interaction of innate responses, including phagocytosis, opsonisation and complement-mediated lysis, allows the eradication of extracellular microorganisms, mostly bacteria. The highly conserved nature of the response confirms its important role in survival.

Whereas innate immunity may damage normal tissues through lack of specificity, the adaptive response is precise, consisting of Ag-specific reactions involving B and T lymphocytes. B cells encompass the humoral immune response.[3,4] They secrete antibody (Ab) that serves to neutralise toxins, prevent organisms adhering to mucosal surfaces, active complement, opsonise bacteria for phagocytosis and sensitise tumour and infected cells for antibody-dependent cell-mediated cytotoxic (ADCC) attack by NK cells. Thus, Ab enhances elements of the innate immune system. T cells are involved in cell-mediated immune responses.[3,4] Two major types of effector T cells have been identified, T cytotoxic (T_C) and T helper (T_H), bearing either CD8 or CD4 molecules on their surface, respectively. The binding of T_C and T_H cells to target cells is via major histocompatibiliy complex class I (MHC I) in the cases of T_C cells and class II (MHC II) in the cases of T_H cells, which is complexed with antigenic peptide. T_C cells recognise and directly kill cells infected with viruses and intracellular organisms, while T_H cells are the orchestrating cells of the immune response, recognising foreign Ag and activating other parts of the cell-mediated immune response to eradicate the pathogen. T_H cells are subdivided functionally by the pattern of cytokines they produce; upon stimulation, precursor T_{HO} cells become

either type 1 (T_{H1}) or type 2 (T_{H2}) T_H cells. T_{H1} cells secrete interleukin-2 (IL-2) and interferon-γ (IFN-γ) to induce a mainly cell-mediated inflammatory response; T_{H1} cells are potent activators of macrophages and NK cells, as well as assisting in the recruitment of lymphocytes and neutrophils to sites of inflammation. T_{H2} cells secrete IL-4, IL-5, IL-6 and IL-10 to both activate and drive the differentiation of B cells. Adaptive immunity is the hallmark of higher order animals. The precise nature of the response means that it takes several days or weeks to develop. Unlike innate immunity, the adaptive response has memory so that subsequent exposure leads to a more vigorous and rapid response.

1.2. The B lymphocyte

B lymphocytes constitute approximately 15% of the circulating lymphoid population and represent the effector cells of the humoral adaptive immune response.[1] The characteristic of a functional B cell is the ability to produce immunoglobulin (Ig) molecules. Ig molecules are glycoproteins that can either exist as membrane-bound molecules on the cell surface or as secreted Ab molecules in the serum or secretions. A mature B cell expresses approximately 10^5 to 10^6 identical Igs on its surface.

1.2.1. The immunoglobulin molecule

Ig molecules have a common structure composed of four polypeptide chains: two identical heavy (H) chains (50-70kD) and two identical light (L) chains (25kD).[5] The four polypeptide chains are assembled into a 'Y' shaped conformation (Figure 1.1.). Each arm of the 'Y' conformation is made up of a complete light chain paired with the amino (N)-terminal portion of the heavy chain, whilst the stem consists of the paired carboxy (C)-terminal portions of the two heavy chains. Digestion of an Ig molecule with the protease papain produces two identical fragments each of 45kD termed Fab fragments (fragment, Ag binding), which are responsible for Ag binding and correspond to the arms of the Y-shaped Ig. A third fragment of 50kD, termed Fc fragment (fragment, crystallisable), so named because it was found to crystalise during cold storage, corresponds to the stem of the Y-shaped Ig. Digestion with another protease, pepsin, produces the $F(ab)^2$ fragment, in which the two Ag-binding arms of the Ig molecule remain linked. Therefore, both heavy chains and light chains can be subdivided into two distinct regions. The N-terminal and 5' variable (V) region of each chain mediates the binding of Ag. As the name suggests, variable regions have a unique amino acid (aa) sequence that is specific for each Ig molecule. The C-terminal region is referred to as the constant region and is responsible for effector function.

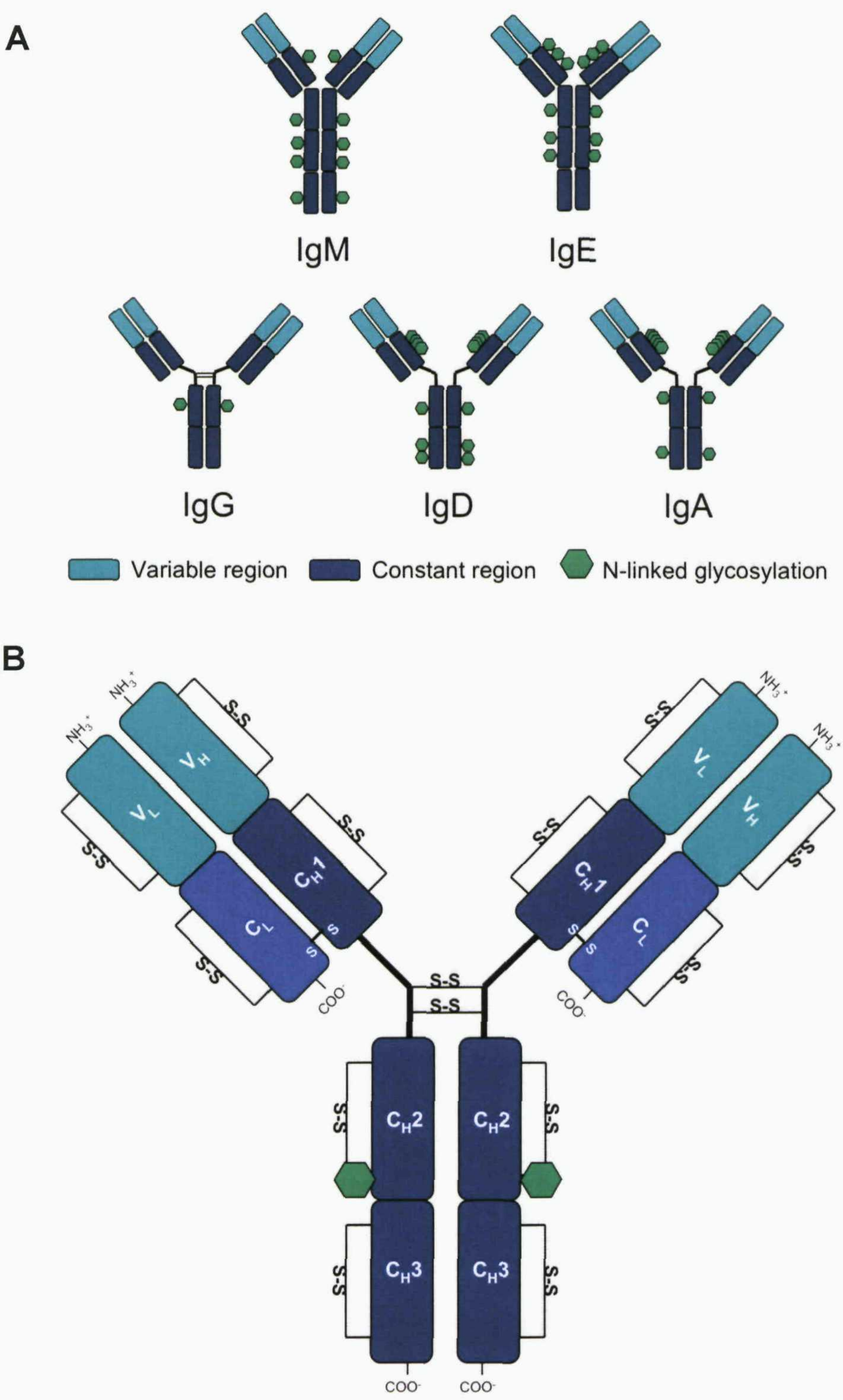


Figure 1.1. The immunoglobulin molecule. A) Schematic diagram of the structure of the five immunoglobulin (Ig) isotypes: IgM, IgD, IgG, IgA and IgE. The isotypes differ in the number of Ig domains, the presence of a hinge region and the distribution of N-linked carbohydrates in the constant regions. B) The Ig molecule (IgG shown) consists of two identical heavy chains and two identical light chains. Each heavy and light chain contains one variable (V) region. Light chains contain one constant region (C_L) while heavy chains contain three or four constant regions (C_H1-3, or C_H4) depending on the isotype. Some heavy chains contain a hinge region (IgG, IgD and IgA). The number and location of inter- and intrachain disulphide bonds differs among the isotypes. Adapted from Goldsby *et al.*, 2000.[1]

1.2.1.1. Immunoglobulin isotypes

Structural differences in the constant region of the heavy chain define five Ig classes or isotypes: IgM, IgD, IgG, IgA and IgE (Figure 1.1a.; see description below).[reviewed in 1,6] The heavy chain constant regions are denoted by μ, δ, γ, α and ε, respectively. In humans, IgG has four subclasses (IgG1, IgG2, IgG3 and IgG4) and IgA has two (IgA1 and IgA2). There are two light chain constant regions, kappa (κ) and lambda (λ), of which the latter has 4 subclasses. The light chain class does not appear to have any effector function and either light chain may be associated with any of the five heavy chain isotypes.

1.2.1.1.1. IgM

IgM can exist as a pentamer, and rarely a hexamer, in the serum or as a monomer on the cell surface. Cell surface IgM functions as a receptor for Ag and participates in the activation of B cells when cross-linked. IgM is the first Ig to be secreted by naïve B cells when stimulated by Ag during a primary response and is the third most common serum Ig; IgM tends to be of low affinity, although its ability to form pentamers greatly increases its effective affinity for Ags with multiple identical epitopes. It is highly effective in agglutination of particulate Ag, such as bacteria, and can destroy targets due to its efficient fixation of complement.

1.2.1.1.2. IgD

IgD exists only as a monomer. IgD is primarily found on the surface of B cells where it functions as a receptor for Ag and participates in the activation of B cells when cross-linked. However, it is also found in the serum at very low levels where its role is uncertain. IgD does not fix complement.

1.2.1.1.3. IgG

All four IgG subtypes exist as monomers. IgG is the most abundant Ig found in the serum where it constitutes 75% of total serum Ig and is also found in lower concentrations in tissue fluids and lymph. IgG is the only class of Ig that can cross the placenta, although not all subclasses cross equally well and IgG2 is unable to cross. IgG is the most versatile Ig as it is capable of performing all known functions of Ig molecules. It is efficient at opsonising pathogens for phagocyte engulfment, particularly IgGl and IgG3, which are primarily responsible for labelling targets for removal by the phagocytic system; both macrophages and neutrophils express receptors for the Fc portion of both subclasses so they are able to bind to and engulf organisms coated with these molecules. All subclasses except IgG4 can activate complement via the classical pathway and thereby cause the lysis

of gram negative bacteria and animal cells. Different subclasses of IgG tend to predominate during responses to different types of pathogen.

1.2.1.1.4. IgA

IgA can exist as a monomer in the serum or as a dimer in secretions. IgA is the major class of Ig found in respiratory, digestive and urogenital secretions, where it represents a key first line of defence against invasion by inhaled and ingested pathogens at the vulnerable mucosal surfaces. Monocytes, eosinophils, neutrophils and macrophages express receptors specific for the IgA Fc portion where binding mediates immune effector responses such as phagocytosis, ADCC and respiratory burst, as well as release of cytokines and inflammatory mediators. Aggregated IgA can also fix complement. IgA is the second most common serum Ig where it functions as a second line of defence mediating elimination of pathogens that have breached the mucosal surface.

1.2.1.1.5. IgE

IgE exists as a monomer. It is the least common serum Ig and is involved in parasite defence. IgE also mediates immediate hypersensitivity reactions through its binding to specific Fc receptors on basophil and mast cells; cross-linking of IgE on these cells by Ag induces degranulation and results in symptoms of allergy. IgE does not fix complement.

1.2.1.2. Immunoglobulin structure

Ig heavy and light chains contain several homologous units of 110 aa known as Ig domains (Figure 1.1b.).[1,5] Light chains contain one variable domain (V_L) and one constant domain (C_L), while heavy chains contain one variable domain (V_H) and either three or four constant domains (C_H1, C_H2, C_H3 and C_H4); IgD, IgG and IgA have 3 C_H domains, while IgM and IgE have four. The IgD, IgG and IgA chains all contain an extended peptide sequence between the C_H1 and C_H2 domains that has no homology with other domains; this region is called the hinge region and allows a certain degree of flexion. The V_H and V_L domains can be further subdivided into discrete structural sections (Table 1.1.). Each V domain contains three polypeptide loops, known as the complementarity-determining region one, two and three (CDR1, CDR2 and CDR3), which sit at the bottom of the binding groove and are crucial for Ag recognition and binding. As expected, aa sequence variability is at its highest within these regions, so named "hypervariable". Alternating with the CDRs are four framework regions (FR1, FR2, FR3 and FR4). The more conserved aa sequence of the FRs forms the basic β pleated sheet structure of the V domains and acts as a scaffold that supports the hypervariable loops.

Table 1.1. The amino acid length of complementarity determining and framework regions in the variable region of human immunoglobulin molecules

Region	Chain	Amino acid position	Length variations
FR1	Heavy	1-30	
	Light	1-23	
CDR1	Heavy	31-35	31 a-b
	Light	24-34	27 a-f
FR2	Heavy	36-49	
	Light	35-49	
CDR2	Heavy	50-65	52 a-c
	Light	50-56	
FR3	Heavy	66-94	82 a-c
	Light	57-88	
CDR3	Heavy	95-102	100 a-k
	Light	89-97	95 a-f
FR4	Heavy	103-113	
	Light	98-107	106 a

Positions and numbers relate to amino acids in the immunoglobulin protein. Length variations indicate differences found between the germline V_H and V_L gene segments; lettering serves for optimal alignment of conserved residues. Adapted from Kabat *et al.*, 1979.[496]

Within each chain, non-covalent hydrogen bonds and Van der Waals forces stabilise the spatial conformation of the polypeptide and a covalent intrachain disulphide bond within each constant domain creates peptide loops of 60 aa. In addition, a structurally important intrachain disulphide bond exists within each V_H and V_L region between aa 22 in FR1 and aa 98 in FR3. Further disulphide bonds serve to link together the heavy chains with the light chains, which occurs between the C_L and C_H1 domains. Finally, multiple cysteine (cys) residues present in the hinge region form interchain disulphide bonds that join together the two heavy chains. Although IgM and IgE chains lack a hinge region, they have an additional domain of 110 aa (C_H2/C_H2) that has hinge-like features. The exact number and position of the interchain disulphide bonds differs between the Ig classes and subclasses. For example, for IgG1, two disulfide bonds in the hinge region between cys^{235} and cys^{238} pairs unite the two heavy chains, while the light chains are coupled to the heavy chains by two additional disulfide bonds between cys^{229} in the C_H1 domains and cys^{214} in the C_L domains (Figure 1.1b.). A notable feature of IgG3 is that it contains 11 interchain disulphide bonds within the hinge region.

An Ig domain-like protein known as the joining (J) chain is associated with the secreted forms of IgM and IgA and enables the formation of pentameric and dimeric molecules, respectively. Hexameric forms of IgM do not associate with J chain. Disulphide bonds link the J chain, which possesses eight cys residues, to the penultimate cys residue of the μ or the α heavy chain. As well as the J chain, the IgA of external secretions, also called secretory IgA, contains a secretory component. This secretory component is a 70K MW consisting of 5 Ig-like domains that bind to the Fc region of the IgA dimer and is produced by the epithelial cells of mucous membranes.

Ig molecules are glycoproteins and contain 3-12% carbohydrates at conserved *N*-glycosylation sites located in the C_H domains. The precise number and location of these sites differs depending on the Ig isotype (Figure 1.1a.). The structure of the attached *N*-linked carbohydrate varies considerably depending on the degree of processing and can include high mannose, complex or hybrid oligosaccharides. Glycosylation at these sites has little or no effect on Ag binding, but variably affects protein assembly, secretion, functional activity and half-life.[7,8] Neither the κ or λ C_L domains contain a potential *N*-linked carbohydrate consensus sequence. Some antibodies have been reported to have *N*-glycosylation sites in the variable regions, which appears to influence Ag binding.[9-11] More recently, the presence of *N*-glycosylation sites in Ig variable regions has been

associated with certain types of lymphoma; this observation forms the basis of some of the data presented herein and will be discussed further in Chapter 3.[12,13]

1.2.2. Immunoglobulin variable gene organisation

B cells are capable of producing Abs that can collectively bind over 10^{15} different Ags. This is governed by the fact that the Ig variable domain can acquire an almost limitless variety of aa sequences. The vast diversity of the Ig repertoire originates from fewer than 400 genes and is partly achieved by a unique DNA recombination process involving three distinct gene segments: variable (V), diversity (D) and joining (J). Each of the three types of gene segment is a member of a multigene family (Table 1.2.). The heavy chain V domain (V_H) is formed by the rearrangement of V_H, D and J_H gene segments on chromosome 14, while the light chain V domain (V_L) is formed either by the rearrangement of Vκ and Jκ gene segments on chromosome 2 or Vλ and Jλ gene segments on chromosome 22. D gene segments do not contribute to the light chains. The exact dimensions and numbers of functional gene segments at the heavy, κ and λ chain gene loci depends upon an individuals haplotype.

1.2.2.1. V_H gene locus

In humans, the functional heavy chain locus (IgH) is located close to the telomere of the long arm of chromosome 14 (14q32.33) and spans 1250 kilobases (kb) (Figure 1.2.).[14,15] There are 51 potentially functional V_H gene segments, which are defined as those genes possessing an open reading frame (ORF) without a stop codon and having been observed rearranged *in vivo*.[16,17] Based on a nucleotide sequence homology of greater than 80%, the 51 functional gene segments can be divided into 7 V_H families: V_H1 to V_H7. A relatively high degree of sequence homology is found in the CDR1 (aa 31-35) of V_H gene segments of a particular V_H family. In contrast, the most extensive diversity lies in CDR2 (aa 50-65). The V_H3 family has the greatest number of functional members (n=22), followed by V_H1 and V_H4 (each n=11), V_H2 (n=3), V_H5 (n=2) and V_H6 and V_H7 (each n=1). Gene segments from each of the 7 V_H families are found interspersed throughout the locus (Figure 1.2.).[18,19] An equivalent number of V_H genes exist as pseudogenes or are non-functional. Pseudogenes are defined by the fact that they do not possess ORFs due to the presence of frame-shifts, stop codons or incomplete V_H exons. Non-functional V_H genes possess an ORF but have minor-defects in splice sites, recombination signal sequences (RSSs) or regulatory elements that impede their rearrangement.

Table 1.2. The number of potentially functional gene segments (V, D, J and C) for the variable regions of human heavy and light chain genes

Gene segment	V_H	Vκ	Vλ
Variable (V)	51	40	31
Diversity (D)	25	-	-
Joining (J)	6	5	4
Constant (C)	9	1	4

The numbers listed here should be regarded as the known maximum number of functional segments; some individuals exist with polymorphic deletions of functional segments or allelic differences, which can render certain segments non-functional and therefore, in these cases, certain numbers will be reduced. Adapted from Tomlinson *et al.*, 1992, Cook *et al.*, 1994, Corbett *et al.*, 1997, Ravetch *et al.*, 1981, Schable *et al.*, 1993, Williams *et al.*, 1996, and Vasicek *et al.*, 1990.[16,17,20,22,27,36,37]

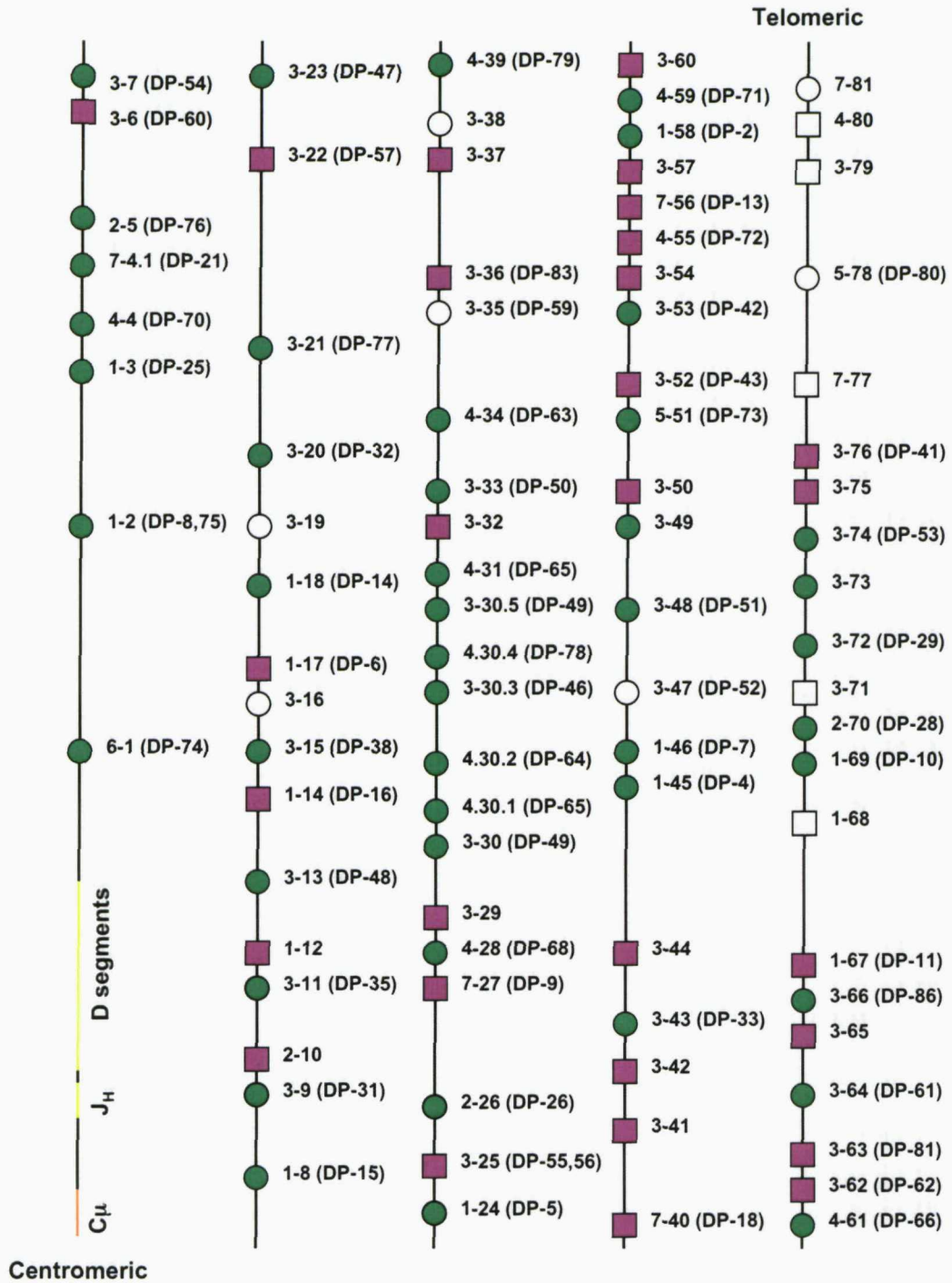


Figure 1.2. The human immunoglobulin V_H gene locus. Map of the human V_H gene locus on chromosome 14q32.33. V_H gene segments are numbered according to their family and position in the locus. Functional V_H gene segments, which have open reading frames (ORFs) and have been observed rearranged *in vivo*, are represented by green circles; non-functional V_H gene segments, which have ORFs, but have not been observed rearranged *in vivo*, are represented by open circles; Non-functional pseudogenes, which have either frame shifts, stop codons or incomplete exons, are represented by purple squares; V_H gene segments that have not yet been sequenced, but are probably pseudogenes are represented by open squares. The total distance between gene segment 7-81 and the J_H genes is approximately 1000kb. Adapted from Cook *et al.*, 1994.[17]

There are 25 functional D gene segments, which can be divided into 7 families: D1 to D7.[20] D3 has the greatest number of functional members (n=5), followed by D1, D2, D4 and D5 (each n=4), D6 (n=3) and D7 (n=1). They are arranged in 5 clusters, which are located between the single V_H6 gene segment and the J_H1 gene segment (Figure 1.3.).[20,21] The D7-27 gene segment is located immediately upstream of the J_H segments. Extra long D gene segments that are flanked by multiple RSSs, termed D segments with irregular spacers (DIR), are found interspersed among the functional D genes, specifically located immediately upstream of each D1 family gene.[21] There is no evidence that DIR genes contribute to the functional repertoire.[20] There are 6 functional J_H segments: J_H1 to J_H6.[22]

In addition to the functional locus on chromosome 14, at least one cluster of D gene segments, called minor D segments, and 25 V_H gene segments are located on chromosomes 15q11.2 and 16p11.2.[23] These genes are designated as orphons and although some V_H segments have ORFs they do not contribute to the normal B-cell repertoire.

The constant domain genes are also located on chromosome 14q32.33. The locus starts 8kb 3' of the J_H genes and extends a further 260kb incorporating the C_H genes, switch regions and control elements (Figure 1.4.). There are nine functional C_H genes (Cμ, Cδ, Cγ1, Cγ2, Cγ3, Cγ4, Cα1, Cα2 and Cε) and a further two C_H pseudogenes (ψCγ and ψCε). A third pseudogene has been mapped to chromosome 9p24.2 (ψCε2).[24,25] Structural similarities between isotypes suggests that the multiple C_H genes derive from a primordial C_H gene with four exons.

1.2.2.2. Vκ gene locus

The human kappa locus (Igκ) is located on the short arm of chromosome 2 (2p11.2) and spans approximately 1800kb (Figure 1.5.).[26] It contains 76 Vκ genes, 5 functional Jκ genes and 1 Cκ gene.[27] The 76 Vκ genes are organised into two clusters, which are separated by a ~800kb stretch of DNA that does not contain any Vκ genes. The Vκ distal cluster, which lies 5' of the locus and in the most centromeric position, spans 400kb and comprises 36 genes. The Vκ proximal cluster, which is 3' of the locus, adjacent to the Jκ and Cκ genes and in the most telomeric position, spans 600kb and comprises 40 genes. The distal region of the locus has a 5'-3' polarity opposite to that of the proximal region and Jκ and Cκ genes.[28] A comparison of the structure and Vκ gene content of the proximal and distal regions indicates that they are the result of a recent duplication and inversion event.[29] Nucleotide sequence homology between the duplicated Vκ gene locus is extremely high at 98.9% (on average).[30-32] A rare haplotype has been described that

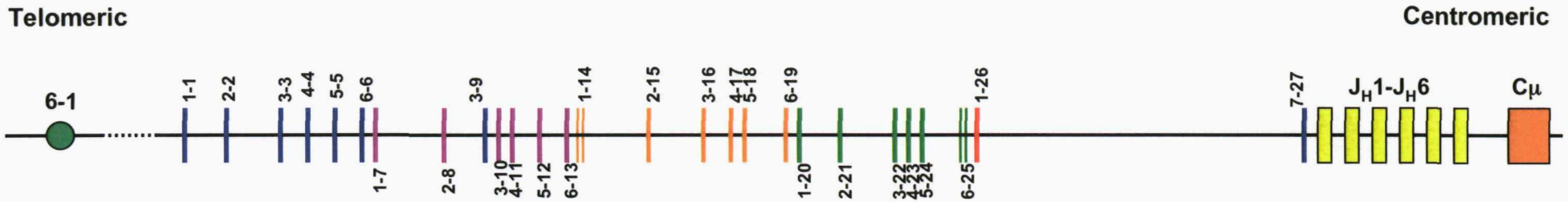


Figure 1.3. The human immunoglobulin D gene locus. Map of the human D gene locus on chromosome 14q32.33. D gene segments are numbered according to their family and relative position in the locus.[17] Functional D gene segments, which have open reading frames (ORFs) and have been observed rearranged *in vivo*, are represented by thin vertical lines; non-functional D gene segments, which have ORFs, but have not been observed rearranged *in vivo*, are represented by double vertical circles. The 5 separate D gene clusters are depicted by the colours blue, purple, orange, green and red. One DIR gene segment is located immediately upstream of each D1 family gene (not shown). The total distance between the V_H6 gene segment, 6-1, and the J_H1 gene is approximately 74kb. Adapted from Corbett *et al.*, 1997.[20]

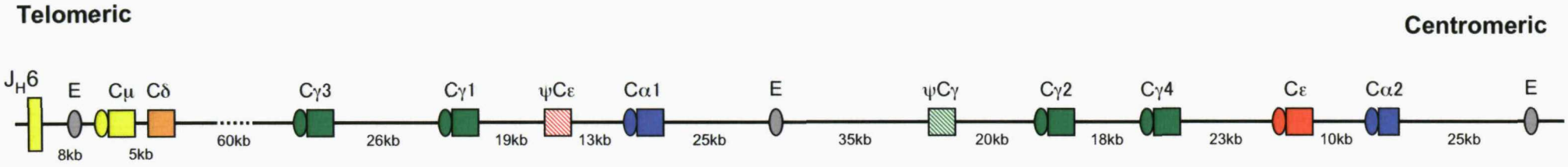


Figure 1.4. The human immunoglobulin heavy chain constant region locus. Map of the human heavy chain constant region (C_H) locus on chromosome 14q32.33. The C_H gene locus start 8kb 3' of the J_H gene segments. Nine functional C_H genes are represented by coloured squares (yellow, orange, green, blue and red); two pseudogenes are represented by striped squares (red and green). A switch (S) region lies upstream of each C_H gene and is represented by a colour matched circle. The I promoters, which lie upstream of each switch region are not shown. The total distance between the J_H6 gene segment and the Cα2 gene is approximately 260kb. Adapted from Goldsby *et al.*, 2000.[1]

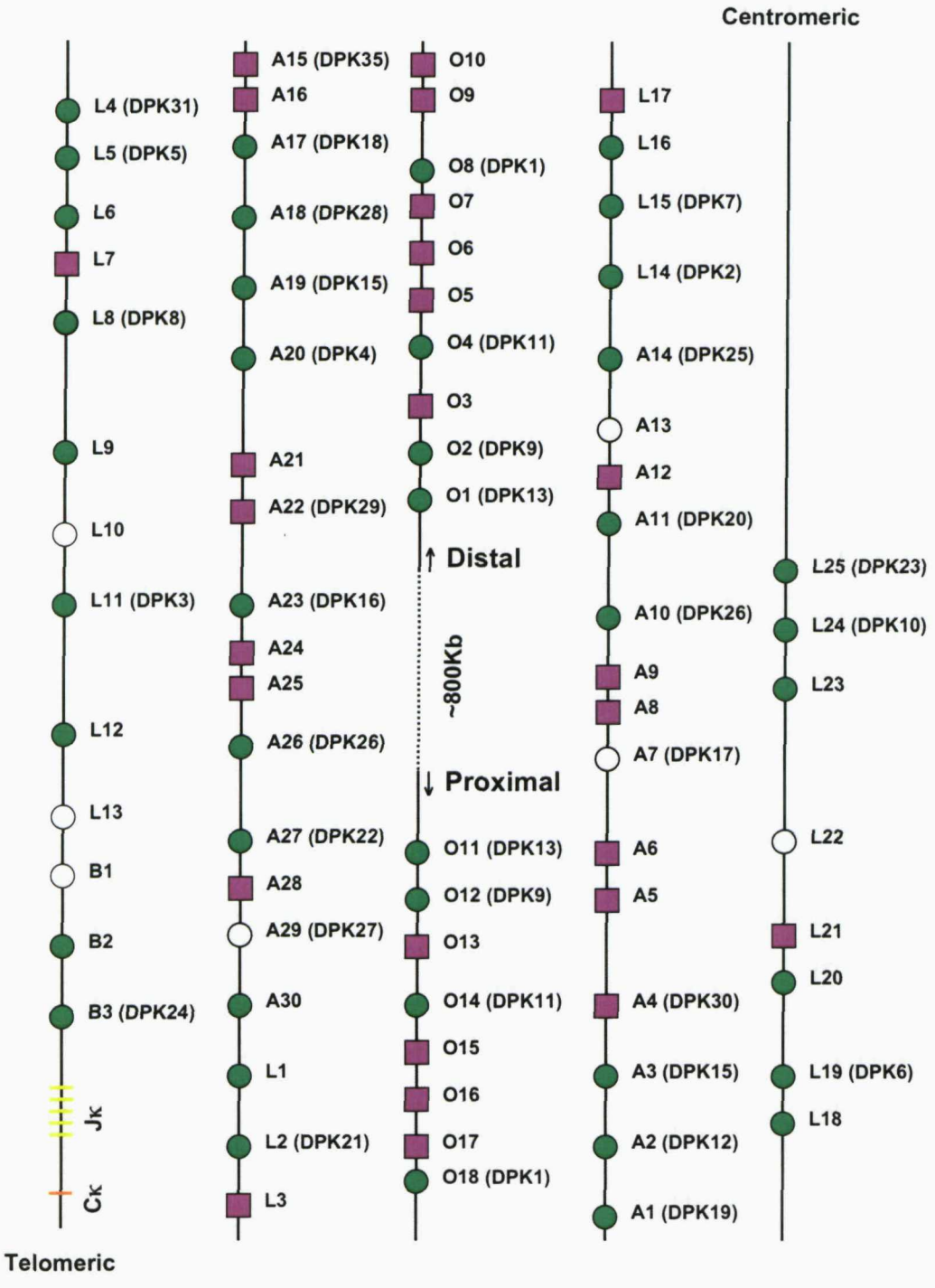


Figure 1.5. The human immunoglobulin Vκ gene locus. Map of the human Vκ gene locus on chromosome 2p11.2. Vκ gene segments are numbered according to the nomenclature of Schable and Zachau, 1993.[27] Functional Vκ gene segments, which have open reading frames (ORFs) and have been observed rearranged *in vivo*, are represented by green circles; non-functional Vκ gene segments, which have ORFs, but have not been observed rearranged *in vivo*, are represented by open circles; Non-functional pseudogenes, which have either frame shifts, stop codons or imcomplete exons, are represented by purple squares. The total distance between gene segment L25 (DPK23) and the Jκ genes is approximately 1800kb. Adapted from Lefranc, 2001.[33]

contains only the proximal cluster.[33] Among the 76 Vκ genes, 47 possess ORFs and 40 have been observed rearranged *in vivo*.[27,28] Based on the nucleotide sequence homology, the 40 functional Vκ genes can be classified into 6 Vκ families: Vκ1 to Vκ6. The Vκ1 family has the greatest number of gene segments (n=19), followed by Vκ2 (n=9), Vκ3 (n=7), Vκ6 (n=3) and Vκ4 and Vκ5 (each n=1). Almost all the rearranged Vκ genes are derived from the proximal region, which may be related to differences in recombination required to produce the rearranged gene; proximal Vκ genes must rearrange by deletion whilst distal Vκ genes must rearrange by inversion. A further 27 Vκ genes exist as pseudogenes and 7 have minor-defects that render them non-functional. There are 5 functional Jκ segments: Jκ1 to Jκ5.[27]

In addition to the Vκ genes in the functional locus on chromosome 2, 28 Vκ orphons have been mapped outside of the functional locus. One orphon has been identified on chromosome 1, 3 on the short arm of chromosome 2 and downstream of the Cκ gene, 13 on the long arm of chromosome 2, 1 on chromosome 15, 6 on chromosome 22 and a further 4 outside of chromosome 2, but which remain unlocalised.[33,34]

1.2.2.3. Vλ gene locus

The human lambda locus (Igλ) is located on the long arm of chromosome 22 (22q11.2) and spans 1050kb (Figure 1.6.).[35] It contains approximately 70 Vλ genes, 4 functional Jλ genes and 4 functional Cλ genes. Among the 70 Vλ genes, 39 possess ORFs and 31 have been observed rearranged *in vivo*.[36] The 31 functional Vλ gene segments can be divided into 10 Vλ families based on the nucleotide sequence homology: Vλ1 to Vλ10. The Vλ3 family has the greatest number of functional members (n=9), followed by Vλ1 and Vλ2 (each n=5), Vλ4 and Vλ5 (each n=3), Vλ7 (n=2) and Vλ6, Vλ8, Vλ9 and Vλ10 (each n=1). In turn, the 10 Vλ families can be organised into 3 clusters: A, B and C. Cluster A, which is proximal to Jλ genes, contains the Vλ2 and Vλ3 families and the 4b gene from the Vλ4 family; cluster B contains the Vλ1, Vλ5, Vλ7 and Vλ9 families; and cluster C, which is distal to Jλ genes, contains the Vλ6, Vλ8 and Vλ10 families and the 4a and 4c genes from the Vλ4 family. The VpreB (CD179a) gene lies between clusters B and C. A further 30 Vλ genes exist as pseudogenes and 8 are non-functional.

In contrast to the V_H and Vκ gene loci, the Jλ gene segments are arranged in four functional Jλ/Cλ pairs: Jλ1/Cλ1, Jλ2/Cλ2, Jλ3/Cλ3 and Jλ7/Cλ7.[37] A further 3 Jλ/Cλ pairs are non-functional due to minor-defects in the 5' splice sites of the Jλ gene segment.

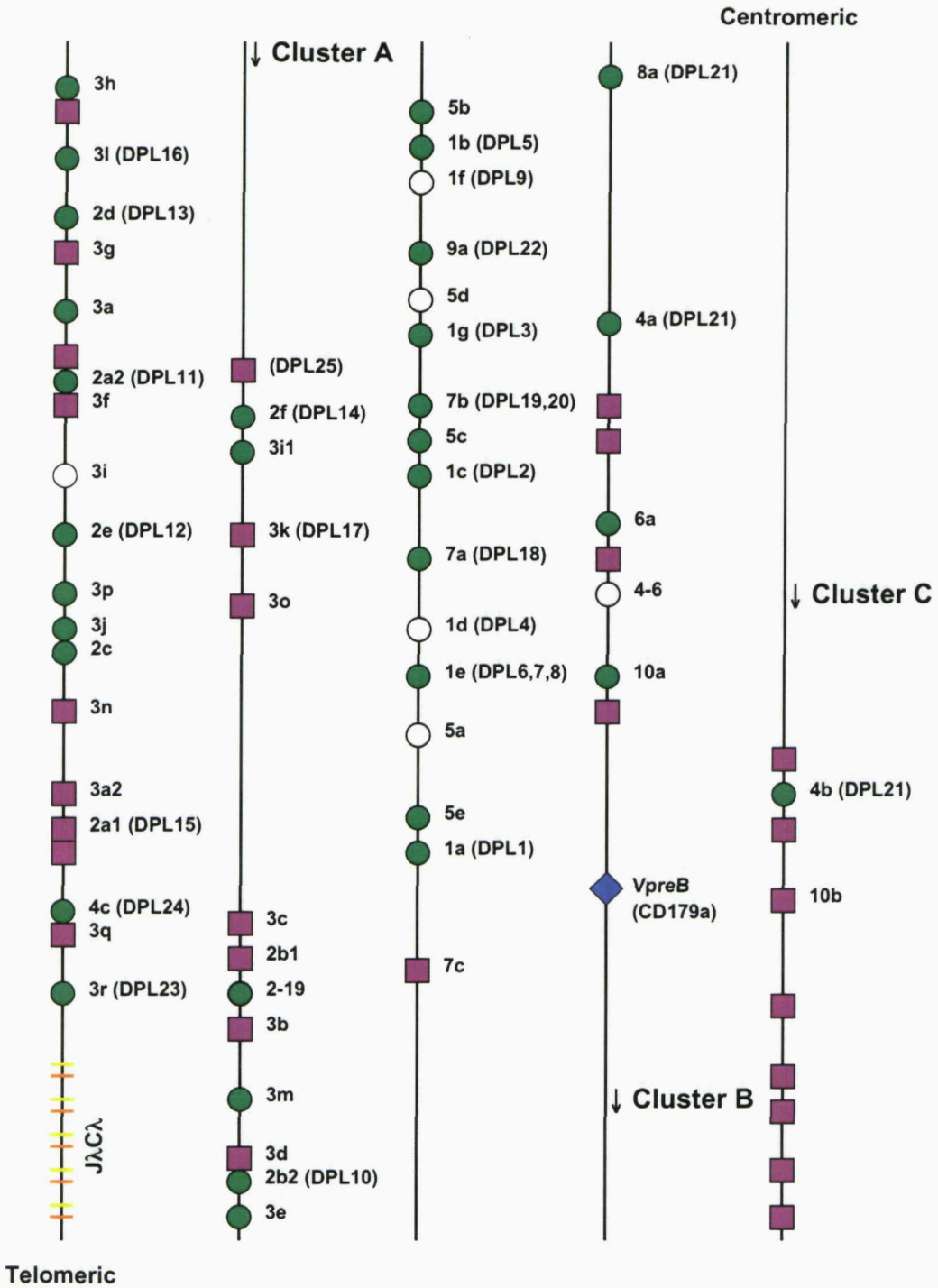


Figure 1.6. The human immunoglobulin Vλ gene locus. Map of the human Vλ gene locus on chromosome 22q11.2. Vλ gene segments are numbered according to the nomenclature of Williams *et al.*, 1996.[36] Functional Vλ gene segments, which have open reading frames (ORFs) and have been observed rearranged *in vivo*, are represented by green circles; non-functional Vλ gene segments, which have ORFs, but have not been observed rearranged *in vivo*, are represented by open circles; Non-functional pseudogenes, which have either frame shifts, stop codons or imcomplete exons, are represented by purple squares. The VpreB gene, which lies between cluster B and C, is represented by a blue diamond. The total distance between gene segment 4a and the Jλ/Cλ genes is approximately 850kb. Adapted from Williams *et al.*, 1996.[36]

1.2.3. V(D)J recombination

Ig genes are rearranged from pre-existing gene segments early in B-cell development using a specialised process of somatic DNA rearrangement, termed V(D)J recombination.[reviewed in 38] Although the targeting of genes for rearrangement is in essence random, the recombination itself is a highly ordered process whereby genes are rearranged in sequential steps, with the successful completion of each step setting the stage for the next genetic event. The multiple recombinations involved are lineage- and cell cycle-specific events. Although they occur in a rapidly cycling cell population, individual rearrangements are targeted to the G_o and G_1 phase of the cell cycle.[39,40]

1.2.3.1. An ordered process

V(D)J recombination begins at the IgH locus with the joining of a D gene segment to a J_H gene segment, which takes place in early pro-B cells (Figure 1.7.). Human pro-B cells usually rearrange D and J_H gene segments on both alleles simultaneously.[41] D segments are unusual in that they can potentially be used in all three reading frames (RFs), although one RF has a propensity to encode a stop codon, which results in the majority of D-J_H rearrangements being productive.[42] The subsequent joining of a V_H gene segment to the D-J_H on one allele generates a V_H-D-J_H unit that encodes the entire V_H region and completes the late pro-B cell stage. Only if the first allele of the IgH locus yields a non-productive or abortive rearrangement does V_H to D-J_H rearrangement proceed on the second allele. Typically, B cells with two unsuccessful V_H rearrangements die by apoptosis, although they may be rescued by secondary D-J_H rearrangements to an upstream V_H. V(D)J recombination at the IgH locus brings the Ig μ enhancer (Eμ), which is located between the J_H6 gene and the Cμ gene, to within 2kb of the Ig promoter. As a result, the rate of transcription of the rearranged V_H-D-J_H unit increases to 10^4 times that of the unrearranged germline DNA. The productive V_H-D-J_H rearrangement is transcribed as a single long mRNA that incorporates both the Cμ and Cδ genes. The primary transcript is subsequently spliced to remove intervening DNA sequences, which includes non-coding introns and J_H gene segments not lost during rearrangement, to yield a Cμ transcript. The final mRNA transcript is translated and the μ heavy chain is assembled with a surrogate light chain and the signal transduction molecules Igα (CD79a) and Igβ (CD79b) to form the pre-B-cell receptor (pre-BCR) (Figure 1.8.).[43] The surrogate light chain is identical for each pre-BCR and is composed of two polypeptides, VpreB and λ5/14.1, which resemble light chain variable and constant regions, respectively. Expression of the pre-BCR on the cell surface signals the down-regulation of the recombinase-activating genes 1 and 2 (*RAG1* and *RAG2*) and leads to the suppression of further V_H rearrangement (allelic

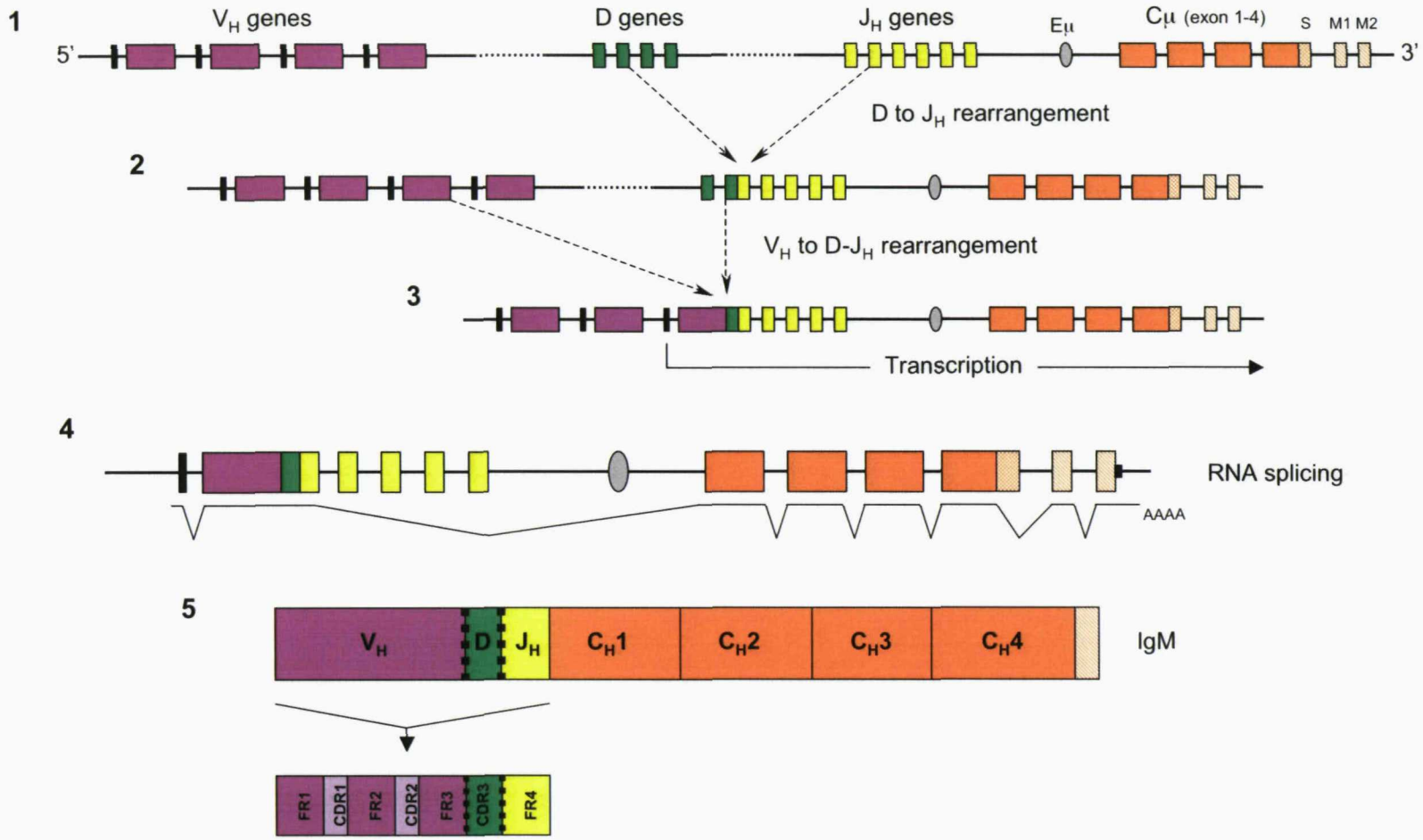


Figure 1.7. VDJ recombination at the heavy chain locus. Schematic representation of VDJ rearrangement at one allele of the Ig heavy chain locus. The top line shows the germline configuration on chromosome 14 (1). The first rearrangement joins one of 25 D gene segments to one of 6 J_H gene segments and leads to the excision of all intervening D and J_H genes (2). The second rearrangement joins the D-J_H unit to one of 51 upstream V_H genes and leads to the excision of all intervening D and V_H genes (3). Under the influenced of the μ enhancer, transcription of the functional V_H-D-J_H unit is initiated from the upstream V_H promoter and incorporates the V_H leader. The primary mRNA transcript is alternately spliced to form either membrane bound or secreted IgM (4-5). The bottom line shows the correspondence of gene segments to framework (FR) and complementarity determining regions (CDR). Adapted from Goldsby *et al.*, 2000.[1]

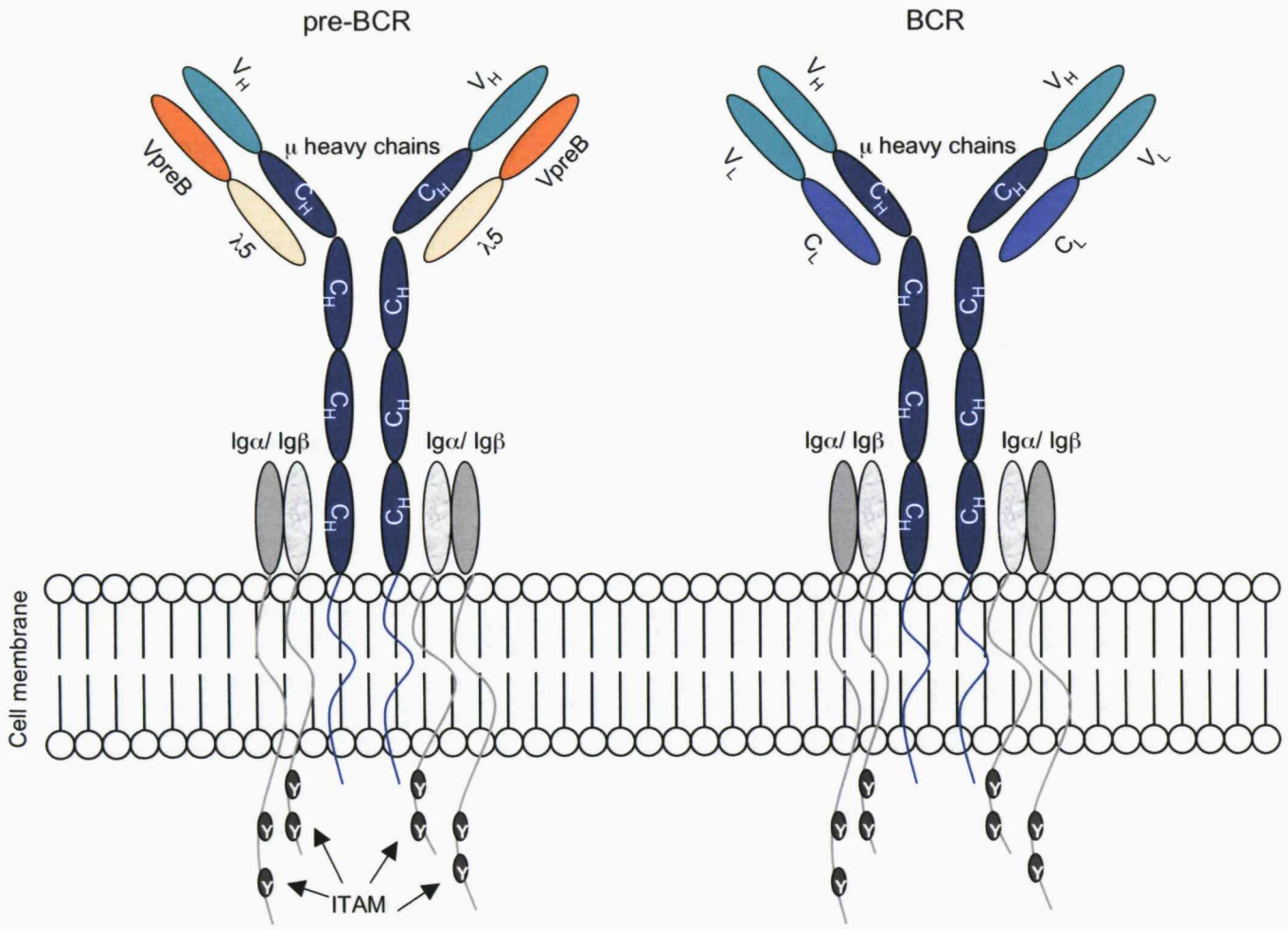
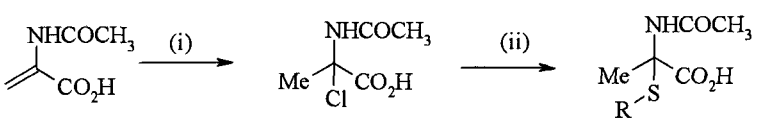


Figure 1.8. The structure of the pre B-cell receptor and the B-cell receptor. Schematic diagram to illustrate the structure of the pre-BCR (left) and the BCR (right). The pre-BCR is expressed on pre-B cells and consists of the membrane bound μ heavy chains paired with the surrogate light chains, which in turn are formed by the VpreB and λ5 chains. The BCR is first expressed on immature B cells and consists of the membrane bound μ heavy chains paired with the light chains, which can either be κ or λ chains. Both the pre-BCR and the BCR associate with two molecules of the membrane bound Igα/Igβ (CD79a/CD79b) heterodimer. The Igα and Igβ chains have long cytoplasmic tails containing 61 and 48 amino acids, respectively. The cytoplasmic tails contain immunoreceptor tyrosine based activation motifs (ITAM), which transduce the stimulus provided by the crosslinking of the membrane Ig with antigen into an effective intracellular signal. Adapted from Goldsby *et al.*, 2000.[1]

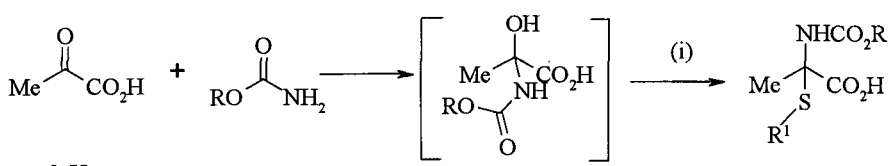
In the retrosynthesis of dithiosilvatin, fragment (**2.47**) could be disconnected in two ways. Directly to the protected α-mercapto tyrosine derived amino acid ester (**2.49**) and subsequently back to the α-mercapto α-amino acid ester (**2.45**) (represented by the double line, figure 2.5) or stepwise through the *bis* protected dithiodiketopiperazine (**2.48**) (represented by the single line, figure 2.5) and subsequently back to the α-mercapto α-amino acid ester (**2.45**). Olsen *et al.* reported the synthesis of similar α-mercapto α-amino acid esters in 1973.[67] He described the synthesis derived from acrylic acids (scheme 2.8). In this route the α-mercaptoalanine derivative (**2.52**) was prepared by treatment of the α-chloro compound (**2.51**) with hydrogen sulfide to give the α-mercaptoalanine (R = H) or by treatment with thioacetic acid to give the α-acetylthioalanine (R = CH_3CO). The α-chloro compound (**2.51**) was formed *in situ* by addition of hydrogen chloride to 2-acetylaminoacrylic acid (**2.50**).



2.50 **2.51** **2.52**

Scheme 2.8 Reagents and conditions: (i) 4 N HCl, TFA, AcOH; (ii) H_2S or AcSH.

These methods were limited to amide protecting groups, as they had to withstand harsh acid reaction conditions in the first step of the synthesis. Further advancements reported by Ottenheijm *et al.* (scheme 2.9) addressed this issue *via* the addition of pyruvic acid (**2.53**) to a carbamate (**2.54**) to form the intermediate 2-hydroxyalanine derivatives (**2.55**) in benzene.[68] The desired amino acid derivative (**2.56**) could then be prepared by the addition of an alkyl thiol, but in low yield.



2.53 **2.54** **2.55** **2.56**

Scheme 2.9 Reagents and conditions: (i) benzene, reflux, R^1-SH, 10-15%.

Although this route could be adapted to more desirable protecting groups, it was extremely low yielding. More recently α-mercapto-amino acids have been prepared by the route previously reported by Motherwell's group (scheme 2.6).[64, 65]

II.4 Synthetic Approach to Dithiosilvatin

II.4.1 Amino Acid Approach

Our initial approach was to adapt the recently published achiral route to similar α-mercapto-amino acid esters by the Motherwell group, so that we could synthesise the desired unnatural amino acid fragments (**2.45**) and (**2.49**) (figure 2.5). Initially like the Motherwell group we chose *p*-methoxybenzyl (Pmb) as our sulfur protecting group. In our desired fragment we required methylamine instead of benzylamine as our amine source in the three component reaction. This analogue was reported by Motherwell, but in very low yields.[64] In light of this publication a disconnection of fragment (**2.57**) can be carried out to give two feasible building blocks (**2.58**) and (**2.59**) (figure 2.6). When combined the target intermediate, Pmb protected dithiodiketopiperazines (**2.57**) could be made. The building block (**2.58**) had been reported by Motherwell so therefore, the route would only require optimisation. It was envisaged that building block (**2.59**) could be synthesised in a similar way, by means of the three component method starting from ketone analogue (**2.62**). This was substituted by ethyl pyruvate (**2.63**) and benzylamine (**2.32**) to investigate the feasibility of this approach. Benzylamine (**2.32**) was used as this gave consistently good yields in the Motherwell publications.[64] If this was successful the conditions would then be repeated with the ketone analogue (**2.62**) and methylamine (**2.61**) to yield the target compound (**2.59**).

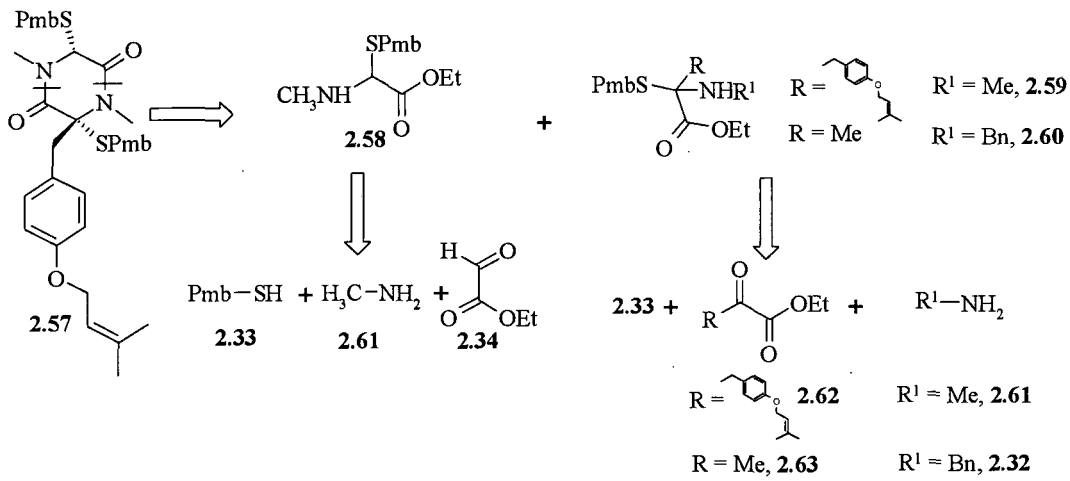
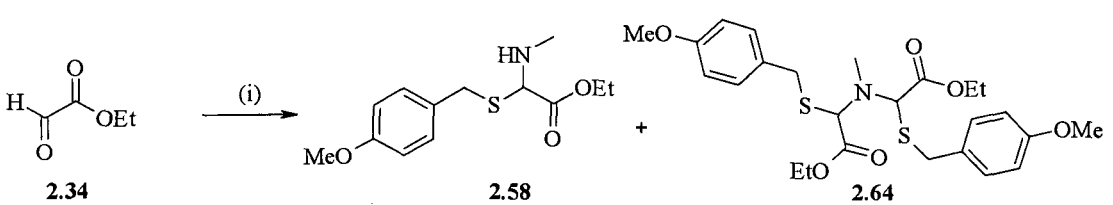


Figure 2.6 Retrosynthesis of dithiosilvatin based on the Motherwell group's findings.

We quickly established that ethyl pyruvate did not react in the desired manner. Benzylamine was treated with ethyl pyruvate and *p*-methoxybenzyl mercaptan in toluene at room temperature and at reflux but neither yielded the desired mercapto amino acid ester (**2.60**). Instead it resulted in a complex reaction mixture. This was also observed in the absence of the mercaptan. This can be explained by the fact that unlike ethyl glyoxalate, the ethyl pyruvate can exists in an enol or keto form. Both of these forms can react with benzylamine leading to a number of reactive species. Due to these findings we focused on optimising the conditions to synthesise the previously reported (4-methoxybenzylsulfanyl)methylaminoacetic acid ethyl ester (**2.58**).

Motherwell *et al.* reported that the desired (4-methoxybenzylsulfanyl)methylamino acetic acid ethyl ester (**2.58**) could be prepared via the three component reaction in 37% yield. The preferred route involved heating the reaction mixture at 110 °C for 5 h. In other examples the elevated reaction temperatures reduced the yields of the desired amino acid esters. Therefore the low yield was rationalised by the possible instability of the compound at higher temperatures.[64]

We therefore thought that microwave heating for shorter times could solve this issue. As the experimental procedure for this particular amino acid ester was not reported, methylamine was liberated *in situ via* the hydrochloride salt using triethylamine as a base. Unfortunately the amino acid ester (**2.58**) was only isolated in 15% yield when heated at 100 °C for 20 min in a CEM microwave system (scheme 2.10). The only other major product isolated was the dimer (**2.64**). There are a number of plausible reasons why this occurred. It could be that the reaction temperature was too high or the concentration of methylamine was too low.



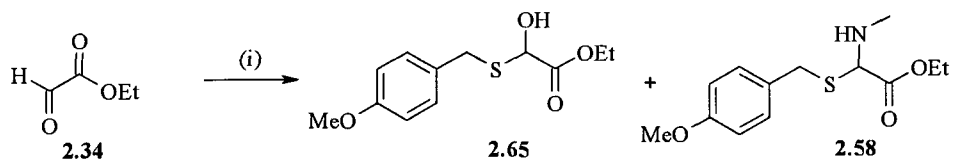
2.34 **2.58** **2.64**

Scheme 2.10 Reagents and conditions: (i) $MeNH_2.HCl$ (1 equiv), Et_3N (1 equiv), *p*-methoxybenzyl mercaptan (1 equiv), toluene, microwave heating to 110 ° C at 100 watts for 20 min, 15% (**2.58**).

In order to address these issues the reaction was repeated at room temperature (scheme

2.11). This fortunately answered both these questions, as the major product isolated was the α-hydroxy derivative (**2.65**). This was clear evidence that the concentration of methylamine was low in the reaction mixture. Therefore the reaction was essentially a two component reaction due to the insolubility of the methylamine hydrochloride in toluene.

It is feasible that the α-hydroxy derivative (**2.65**) could then react with any amino acid ester (**2.58**) formed when the reaction mixture was heated in the microwave. The heating would slightly increase the solubility of the methylamine hydrochloride, increasing the concentration of methylamine in solution but upon heating the amino acid ester (**2.58**) formed could react further with the α-hydroxy derivative (**2.65**), which is formed in excess in the reaction mixture to yield the dimer (**2.64**) as the major product.



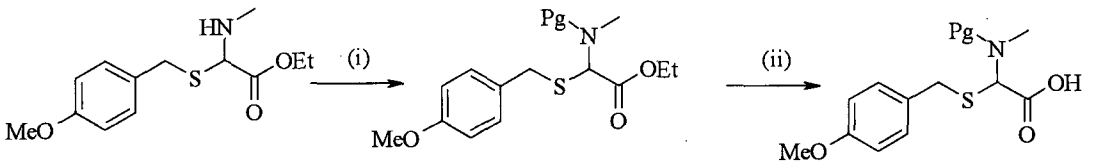
2.34 **2.65** **2.58**

Scheme 2.11 Reagents and conditions: (i) $MeNH_2.HCl$ (1 equiv), Et_3N (1 equiv), *p*-methoxybenzyl mercaptan (1 equiv), toluene, rt, 4 h. 70% (**2.65**)

The good news was the reaction did work at room temperature though, as a small trace of (**2.58**) was isolated. The reaction was therefore repeated, but the methylamine hydrochloride and triethylamine in toluene and ethyl glyoxalate were stirred at rt for 10 minutes before the addition of the *p*-methoxybenzyl mercaptan. This suspension was stirred for 2 h to give predominantly α-hydroxy derivative (**2.65**) as monitored by TLC. Excess methylamine hydrochloride and triethylamine in toluene was then added and the reaction mixture stirred at rt for 18 h to give (**2.58**) in a respectable 68% yield. The method was workable, but not ideal as the yields varied. We simply used a slight excess of 2.0 M methylamine solution in THF, instead of excess methylamine hydrochloride and triethylamine in toluene and found this consistently gave the desired compound (**2.58**) in 66% yield.

We decided that in order to maximise the potential uses of the amine (**2.58**) we would investigate a protection strategy. It was also observed that the amine (**2.58**) was not

stable following prolonged storage even at low temperature. It was postulated that protection of the amine with an amide or preferably a carbamate would solve this problem. Both the Fmoc and Boc protected ethyl esters (**2.66**, **2.67**) were synthesised in good yield (scheme 2.12) under standard conditions. The Boc protected analogue (**2.67**) was readily saponified to the acid (**2.69**) with lithium hydroxide in aqueous THF, at room temperature in almost quantitative yield. This was not the case with the Fmoc derivative (**2.66**). Even at low temperatures the Fmoc group was cleaved in preference to the saponification of the ethyl ester under basic conditions.



2.58

Pg = Fmoc, **2.66**
Pg = Boc, **2.67**

Pg = Fmoc **2.68** (failed)
Pg = Boc = **2.69**

Scheme 2.12 Reagents and conditions: (i) Fmoc-Cl (1 equiv), Na_2CO_3 (1.1 equiv), aq 1,4-dioxan, rt. 4 h, 54% (**2.66**); or (i) Boc_2O (1 equiv), NEt_3 (1.1 equiv), CH_2Cl_2, rt, 18 h, 80% (**2.67**); (ii) $LiOH.H_2O$ (1 equiv), THF, H_2O, rt, (failed) (**2.68**), 0.2M $LiOH_{(aq)}$ (1 equiv), THF,0 °C, 2 h, (failed) (**2.68**); or (ii) $LiOH.H_2O$ (1 equiv), THF, H_2O, rt. 4 h, 98% (**2.69**).

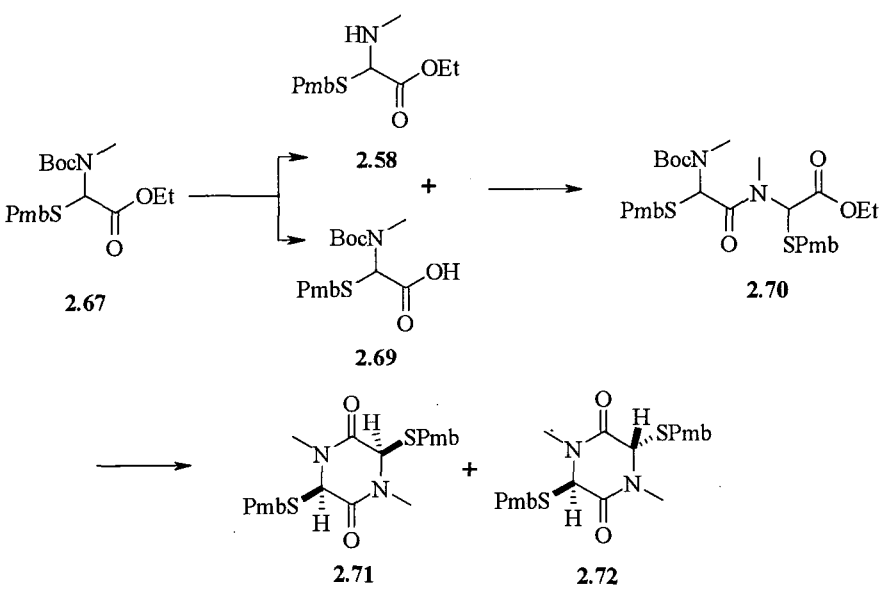
The Fmoc protecting group was the preferred protecting strategy as it was removed rapidly nonhydrolytically, by simple secondary amines, which would not affect the other protecting groups, but it is labile under basic conditions.[69]

The alternative Boc protecting group is potentially more problematic as it is removed under acid conditions. The *para*-methoxybenzyl group on the sulfur is also acid sensitive. With this in mind a number of standard Boc cleavage conditions were attempted. It was found that the Boc group could be cleanly cleaved upon the addition of 4M HCl in dioxane solution. The free amine was cleanly isolated after 30-40 minutes at room temperature, without loss of the *para*-methoxybenzyl protecting group.

The Boc amino acid ester analogue (**2.67**) was found to be stable even when stored at room temperature for prolonged periods of time. It can be selectively saponified to the acid (**2.69**) or converted back to the amine (**2.58**), making it a versatile and durable building block.

II.4.2 Synthesis of Protected Dithiodiketopiperazine

From the retrosynthesis of dithiosilvatin and the retrospective *bis* protected dithiodiketopiperazine (**2.47, 2.57**) (figure 2.5 and figure 2.6) there are two plausible disconnections. Directly to the protected α-mercapto tyrosine derived amino acid ester (**2.49**) and subsequently back to the α-mercapto α-amino acid ester (**2.45**) (represented by the double line, figure 2.5), which relies on the problematic synthesis of the substituted amino acid ester (**2.59**). Or stepwise through the *bis* protected dithiodiketopiperazine (**2.48**) (represented by the single line, figure 2.5) and subsequently back to the α-mercapto α-amino acid ester (**2.45**), which requires a good protecting group strategy. We thought the Boc protected building block (**2.67**) fulfilled these requirements and was the most advantageous approach towards the continued synthesis of the epidithiodiketopiperazine core. In the purposed scheme, (scheme 2.13) both dithio DKPs (**2.71** & **2.72**) can be prepared from the one building block (**2.67**)



Scheme 2.13 Proposed route to *bis* protected dithiodiketopiperazine.

The dipeptide (**2.70**) could be prepared *via* a number of standard peptide coupling methods, (table 2.1) from the acid (**2.69**), and the hydrochloride salt of compound (**2.58**) which in turn was prepared by acid cleavage of the Boc protected analogue (**2.68**). The best yields were achieved by first forming a mixed anhydride with *iso*butyl chloroformate in THF at -10 °C in the presence of NMM as base. After 40 minutes the reaction mixture was filtered and the filtrate added to a solution of the hydrochloride

salt (**2.58**) and NMM in dichloromethane at 0 °C. This was then stirred at room temperature for 4 hours giving the dipeptide (**2.70**) as a mixture of diasteroisomers in 58% yield. We changed from THF to dichloromethane as the solvent of choice due to the insolublity of the hydrochloride salt of (**2.58**) in THF. This resulted in an improved yield from 38% to 58%.

Reaction conditions and reagents	**Yield**
Pybrop ® (1.1 equiv), DIEA, CH_2Cl_2, rt, 18h	10%
HBTU (1.0 equiv), Et_3N, CH_2Cl_2, rt, 18 h,	12%
*iso*BuOCOCl (1.1 equiv), NMM, THF -10 °C – rt 4 h,	38%
*iso*BuOCOCl (1.1 equiv), NMM, THF & CH_2Cl_2 -10 °C – rt 4 h	58%

Table 2.1 Amino acid coupling conditions attempted.

Small scale trial reactions showed the amine of the dipeptide (**2.70**) could be unmasked in a similar manner to the Boc amino acid ester (**2.67**) previously reported. We opted not to isolate the hydrochloride salt of the amine dipeptide to limit the risk of decomposition. Compound (**2.70**) was treated with 4.0 M HCl in dioxane for 50 minutes. The solvent was removed and the residue azeotroped with toluene to remove any residual hydrochloric acid. The resultant hydrochloride salt was then heated at 50 °C in a 5% triethylamine, isopropanol solution for 18 hours, to exclusively give the *cis* dithiodiketopiperazine (**2.71**) in 45% yield as a crystalline solid. The *cis* configuration was identified by X-ray crystallography of the resultant crystals. Unfortunately the yield could not be improved, staying consistently at 45%. This made us believe that the yield (**2.71**) was determined by the subsequent stability of the deprotected dipeptide free base. This was confirmed by the presence of *p*-methoxybenzyl mercaptan by TLC of the crude reaction mixture. The *cis* isomer was exclusively isolated during the reaction. This can be explained by the *cis* isomer (**2.71**) being thermodynamically more stable than the *trans* isomer (**2.72**), consistent with Motherwell's findings. The thermodynamic preference for formation of the *cis* isomer was supported by gas phase DFT calculations using Gaussian 03 software, for a single molecule using the B3LYP/6-31G(d) level of theory which revealed that the *cis* isomer configuration is approximately 9 kJ mol^{-1} more stable than the alternative *trans* arrangement.[65]

II.4.3 Alkylation Strategy

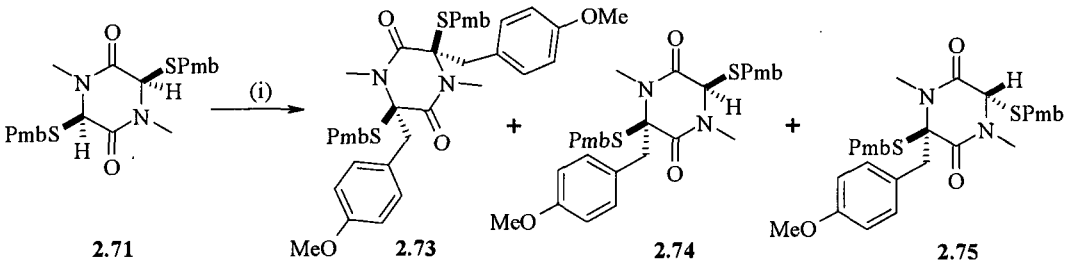
Kishi *et al.* reported the synthesis of a number of symmetrical and unsymmetrical disubstituted and mono substituted ETPs *via* clean *cis* alkylation of 3,6-*bis* methylthio-1,4-dimethylpiperazine-2,5-dione.[70] We repeated these methodologies on the 3,6-*bis*-(4-methoxybenzylsulfanyl)-1,4-dimethylpiperazine-2,5-dione (**2.71**) but our observation differed slightly from that reported by Kishi. We completed the alkylation experiments with *p*-methoxybenzyl bromide which, with further elaboration could be converted to the desired dithiosilvatin.

The diketopiperazine (**2.71**) was treated with 2.2 equiv of LDA at -78 °C in THF and stirred for one hour. One equivalent of *p*-methoxybenzyl bromide was added in THF. Kishi aways added five or more equivalents of alkylating agent in his analoguous experiments. After one hour one equivalent of acetic acid in THF was added, upon workup three new products, one major and two minor were observed. Unfortunately a lot of starting DKP was also recovered (**2.71**). Kishi only isolated the *cis* mono alkylated product under similar reaction conditions.[70]

The major product isolated was a crystalline solid. X-ray crystallography, ^{1}H and ^{13}C NMR analyses confirmed the structure to be the *cis* dialkylated analogue (**2.73**). ^{1}H, ^{13}C NMR and MS analysis identified the two remaining products to be the desired mono-alkylated analogues. The more polar of the two remaining products was again crystalline. X-ray crystallography determined it to be the *trans* isomer (**2.75**). Therefore the remaining less polar product was assigned to be the desired *cis* isomer (**2.74**).

By gradually increasing the amounts of LDA and adding the *p*-methoxybenzyl bromide slowly as a dilute solution in THF under similar reaction conditions, the amount of the two mono alkylated DKPs could be enhanced. Using LDA (3 equiv), *p*-methoxybenzyl bromide (1.5 equiv) and acetic acid (3 equiv) the two mono alkylated DKPs were isolated in 64% yield in a ratio of 3:1 (**2.74**: **2.75**). Only a trace amount of compound (**2.73**) was isolated under these conditions (scheme 2.14). The reaction was also attempted with LDA (1.2 equiv) and an excess of the *p*-methoxybenzyl bromide, but only recovered starting material was isolated. This was consistent with Kishi's

observations on the differences in the reactivity of the mono and dicarbanion, reported in the total synthesis of hyalodendrin.[58]



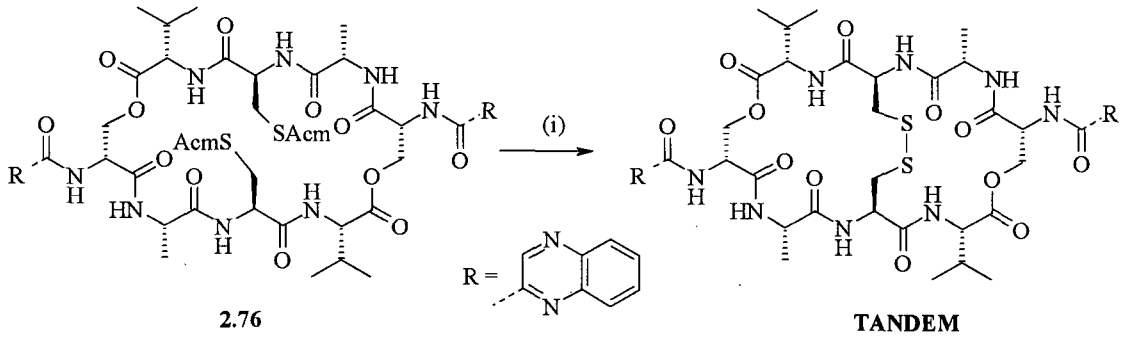
2.71 2.73 2.74 2.75

Scheme 2.14 Reagents and conditions: (i) LDA (3 equiv), THF, -78 ° C, *p*-methoxybenzyl bromide (1.5 equiv) in THF (2 ml), AcOH (3 equiv) 64 % (**2.74** & **2.75**).

II.4.4 Epidithiodiketopiperazine (ETP) Synthesis

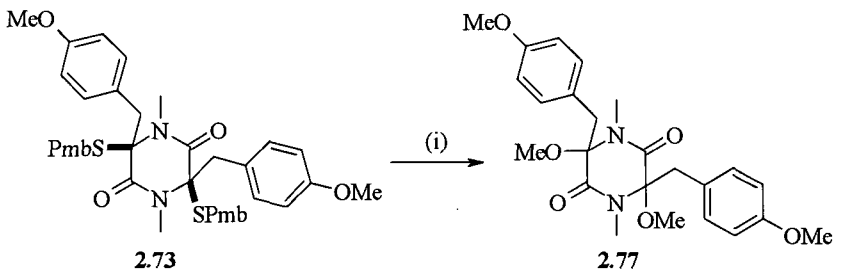
The disulfide bridge can be found in a large number of natural products. Many of these natural products and their analogues have been produced synthetically. One such example is the macrocycle TANDEM, a synthetic analogue of triostin A (Scheme 2.15).[71, 72] In TANDEM like many macrocyclic disulfide analogues, the disulfide bridge was formed by addition of iodine to the *bis*-Acm protected compound (**2.76**) in DMF. Iodine oxidation conditions of this kind have also been reported for trityl (Tr) protected analogues. These reaction conditions are a convenient method to form the disulfide without having to first cleave the protecting group. Traditionally this one step procedure, simultaneously removing the sulfur protecting groups and forming the disulfide bridge *via* treatment with iodine in DMF or MeOH, is carried out at high dilution.

No such procedures had been reported for this kind of disulfide formation *via* the Pmb protecting group, but we saw this as an easy and convenient way to form the ETP in a one step procedure.



Scheme 2.15 Reagents and conditions: (i) I_2, DMF, rt, 1 h.

This approach was therefore initially attempted on the *cis* dialkylated DKP (**2.73**) in order not to waste the *cis*-3,6-*bis*-(4-methoxybenzylsulfanyl)-3-(4-methoxybenzyl-1,4-dimethylpiperazine-2,5-dione analogue (**2.74**) (scheme 2.16). The DKP (**2.73**) was treated with iodine in a methanol, DMF solution for one hour. A rapid colour change was observed and TLC analysis showed a spot to spot conversion. Upon workup with aqueous sodium thiosuphate solution to remove any excess iodine, we isolated a single compound. Spectral analysis identified the product was not the desired ETP, but 3,6-dimethoxy-3,6-*bis*-(4-methoxybenzyl)-1,4-dimethylpiperazine-2,5-dione (**2.77**), which was isolated in near quantitative yield.



Scheme 2.16 Reagents and conditions: (i) I_2 (1.5 equiv), MeOH (10 ml), DMF (1 ml) rt 1 hr, 80%.

The mechanism outlined (figure 2.7) demonstrates the importance of the protecting group to aid the one step deprotection and oxidation to the disulfide. The Tr group stabilises the carbocation, which leads to the activated thiol and subsequent disulfide. This is a plausible reason why the *bis* Pmb compound (**2.73**) did not proceed in the desired manner. The activation with iodine resulted in the SPmb acting as a leaving group for nucleophilic attack by methanol yielding compound (**2.86**).

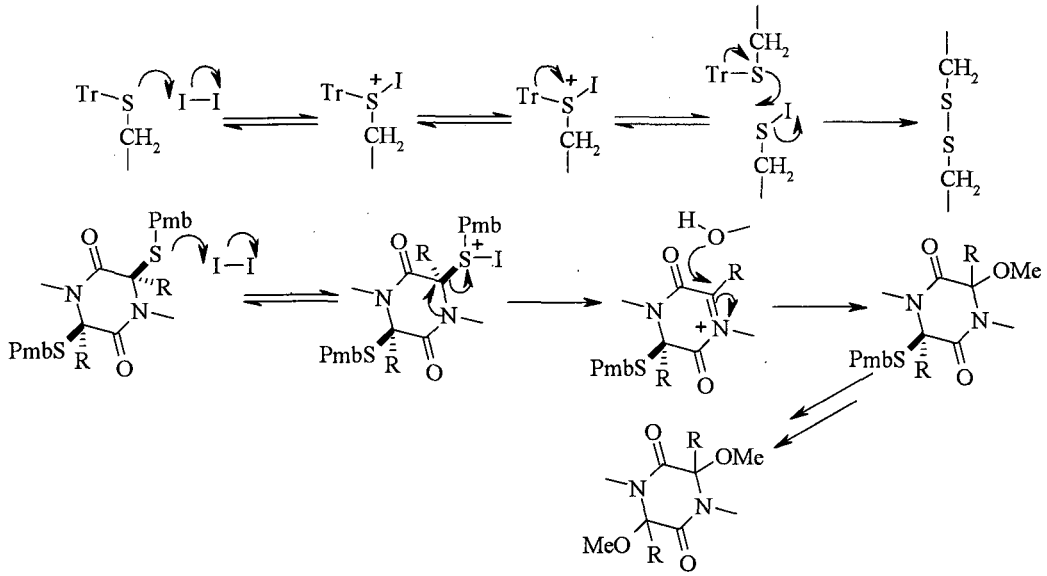
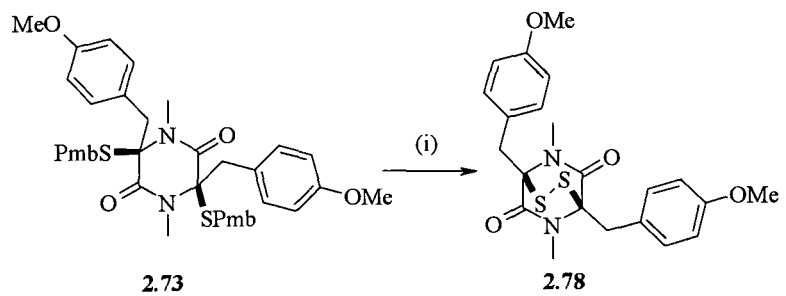


Figure 2.7 Iodine oxidation method to form disulfide bridge and plausible nucleophilic attack of activated SPmb group.

Taking this observation into consideration, an alternative approach was attempted which was to deprotect the sulfides then form the disulfide bridge *via* oxidation. As before the dialkylated DKP (**2.73**) was used to investigate the feasibility of the new reaction conditions

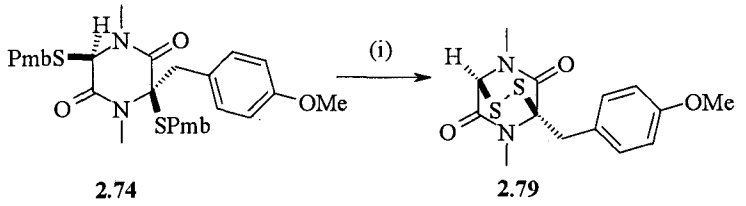
The ETP (**2.78**) was isolated in 38% yield (scheme 2.17) using the method first developed by Kishi in his synthesis of hyalodendrin.[58] The DKP (**2.73**) was first treated with 1 M boron trichloride solution in dichloromethane at 0 °C for 10 minutes to remove the sulfur protecting groups. Oxidation to the disulfide was accomplished *via* the addition of iodine in a 10% methanol dichloromethane solution, after the excess boron trichloride, dichloromethane solution had been removed by evaporation *in vacuo.*



2.73 **2.78**

Scheme 2.17 Reagents and conditions: (i) (a) 1.0 M BCl_3 (2.4 equiv) 0 °C 10 min. evaporate; (b) I_2 (2.0 equiv), 10% MeOH, CH_2Cl_2, rt, 20 min, 38%.

The method was successfully applied to the desired mono alkylated analogue (**2.79**), which was prepared in 46% yield from the *cis* DKP (**2.74**) (scheme 2.18).

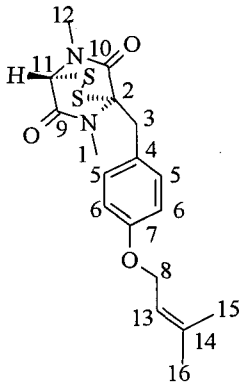


2.74 **2.79**

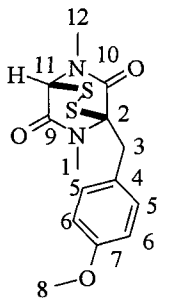
Scheme 2.18 Reagents and conditions: (i) (a) 1.0 M BCl_3 (2.4 equiv) 0 °C 10 min, evaporate; (b) I_2 (2.0 equiv), 10 % MeOH, CH_2Cl_2, rt, 20 min, 38%.

The ^1H and ^{13}C NMR spectroscopic data were comparable to that published for the isolated dithiosilvatin.[66] The ^1H signal for H11 (figure 2.8) at δ_H 5.33 ppm was characteristic for the disulfide bridge structure (table 2.3). The appropriate ^{13}C signals for the carbons found in both dithiosilvatin and compound (**2.79**) were identical (table 2.2).

The MS data were also consistent to that reported for dithiosilvatin, in that the ion minus the two sulfurs was observed in both examples: *m/z* $[M\text{-}S_2]^+$ 314 (dithiosilvatin), *m/z* $[M+H\text{-}2S]^+$ 261.1 (**2.79**).



dithiosilvatin
M.W. 378.52
$C_{18}H_{22}N_2O_3S_2$



2.79
M.W. 324.42
$C_{14}H_{16}N_2O_3S_2$

Figure 2.8 Dithiosilvatin and compound (**2.79**).

Carbon	Dithiosilvatin (δ_C)	(2.79) (δ_C)
1	28.13	28.5
2	77.14	76.9
3	36.97	36.3
4	125.83	126.1
5	130.53	130.8
6	114.78	114.4
7	158.24	159.1
8	64.80	55.4 (methyl)
9	164.89	165.1
10	165.51	165.7
11	67.14	67.3
12	32.26	32.7
13	119.76	n/a
14	137.98	n/a
15	18.17	n/a
16	25.76	n/a

Table 2.2 ^{13}C NMR chemical shifts for dithiosilvatin and compound (**2.79**).

The final steps towards dithiosilvatin require the demethylation of compound (**2.79**) to give the phenolic derivative (**2.80**). It was foreseen that this could then be further elaborated by simple alkylation of the phenol (figure 2.9).

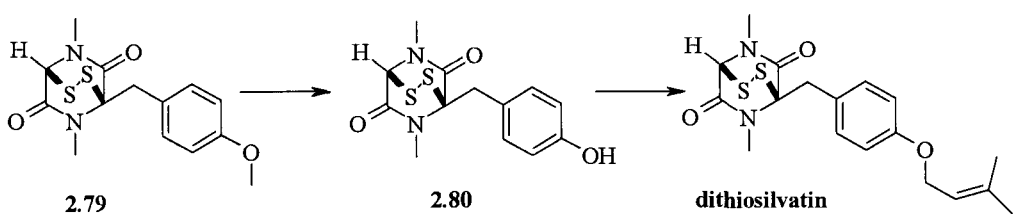


Figure 2.9 Synthetic steps required towards dithiosilvatin.

Proton type	Dithiosilvatin (δ_H)	(2.79) (δ_H)
Olefinic Me	3H, s,1.74	n/a
Olefinic Me	3H, s, 1.79	n/a
NMe	3H, s, 2.97	3H, s, 2.97
NMe	3H, s, 3.17	3H, s, 3.16
CH2	1H, d, *J* 15Hz, 3.45	1H, d, *J* 15Hz, 3.58
CH2	1H, d, *J* 15Hz, 3.94	1H, d, *J* 15Hz, 3.92
OC*H2*CH=	2H, d, *J* 7Hz, 4.48	n/a
COCHN	1H, s, 5.33	1H, s, 5.33
OCH2C*H*=	1H, br t, *J* 7Hz, 5.45	n/a
ArH	2H, d, *J* 9Hz, 6.84	2H, d, *J* 9Hz, 6.83
ArH	2H, d, *J* 9Hz, 7.24	2H, d, *J* 9Hz, 7.25

Table 2.3 ^{1}H NMR chemical shifts for dithiosilvatin and compound (**2.79**).

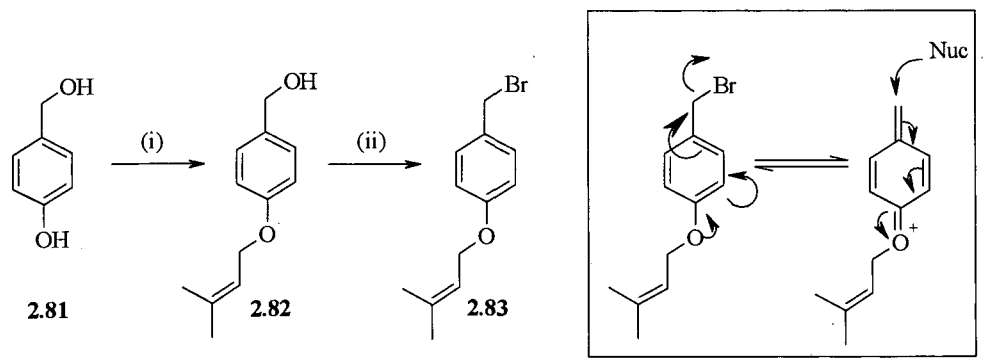
A number of demethylation conditions were attempted (table 2.4) without success. This was possibly due to the relative instability of compound (**2.79**) under standard demethylation conditions.[73]

Reagents and conditions	**Result**
1.0 M BCl_3 in CH_2Cl_2 at 0 °C, 1 h	60% recovered S.M (**2.79**)
1.0 M BCl_3 in CH_2Cl_2 at 0 °C 1 h, rt, 2 h	30% recovered S.M (**2.79**)
1.0 M BBr_3 in CH_2Cl_2 at 0 °C 1 h, rt, 2 h	No recovered S.M or products isolated

Table 2.4 Demethylation reaction conditions attempted.

II.4.5 Isoprenyl Approach

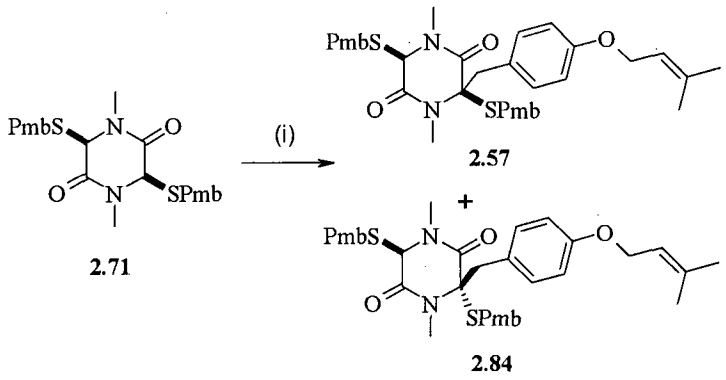
From the synthetic route developed thus far a workable route to the disulfide bridge had been successfully achieved. Unfortunately the proposed demethylation method was not viable as decomposition of the ETP (**2.79**) was observed. We thought the best approach was to have the isoprenyl group in place and form the disulfide at the final step. We first required synthesis of the relevant bromide. This proved to be difficult. The bromide (**2.83**) was extremely unstable although it could be prepared from 4-hydroxybenzoyl alcohol (**2.81**) in two steps (scheme 2.19). The more acidic phenolic group was selectively deprotonated by the addition of one equivalent of sodium hydride in DMF at 0 °C. This was then quenched with 3,3-dimethylallyl bromide to give the alcohol (**2.82**) in 61% yield. Conversion to the bromide was achieved under standard Appel reaction conditions.[74]



Scheme 2.19 Reagents and conditions: (i) NaH (1.0 equiv), DMF, 0 °C 30 min, 61% (ii) 3,3-dimethyl allyl bromide (1.0 equiv), rt, 2 h; (ii) $P(Ph)_3$ (1.27 equiv), CBr_4 (1.2 equiv), CH_2Cl_2, 0 °C – rt 6 h, crude 60%; Box = plausible decomposition of bromide.

Attempted purification of the crude product was by trituration with cold petroleum ether to remove any triphenylphosphine oxide followed by flash column chromatography through a short silica plug. However, we observed the material was still always impure, approximately 90% bromide, but quickly degraded further on standing. This could be explained by the inherent instability of the bromide. If the bromine is eliminated the resultant stabilised charged species formed can be attacked by any nucleophiles. The bromide had to be used the day it was prepared.

Alkylation of the DKP (**2.71**) with crude bromide (**2.83**) was completed as before, but not as cleanly. The *trans* isomer (**2.84**) was isolated pure, but unfortunately the desired *cis* isomer (**2.57**) could not be isolated in greater than 60% purity after several columns (scheme 2.20).



Scheme 2.20 Reagents and conditions: (i) (a) LDA (3.0 equiv), THF, -78 °C, 30 min; (i) (b) **2.83** (1.5-2.0 equiv), THF, -78 °C, 2 h; (i) (c) AcOH (2.0 equiv), THF, -78 °C rt, 1 min, 21%, (**2.57**:**2.84**, 2:1).

This was due to a co-eluting impurity from the bromide, so further purification by column chromatography was not possible. The intramolecular disulfide formation was attempted on the impure *cis* analogue without success. Once again this resulted in multiple products.

Due to the instability of the bromide, other leaving groups were investigated such as the mesylate. Unfortunately any attempts to form the mesylate were also unsuccessful. This can be explained by the mesylate being a better leaving group than the bromide thereby increasing the formation of the charged reactive species which can react with any nucleophile present such as the alcohol or water (scheme 2.19).

II.4.6 Phenol Protection Strategy

Taking into consideration the stability issues with the isoprenyl bromide (**2.83**), it was thought the best approach would be to proceed with a more stable bromide. This could be achieved by protecting the oxygen in such a manor that it could no longer stabilise a positive charge. This could be achieved *via* the carbonate, acetate or sulfonate functionality. The strategy was to revert back to the original approach of cleaving the

protecting group on the phenol after the disulfide bridge has been formed. This would leave the free phenol to be alkylated in the final step.

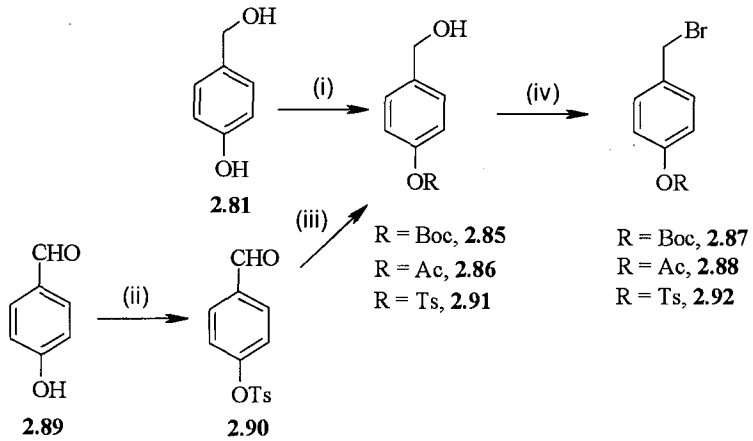
Three phenolic protecting groups were investigated; the *tert*-butyl carbonate (Boc), acetyl (Ac) and the tosyl (Ts). The desired chlorides or bromides of all three protecting groups had been reported *via* various methods.[75, 76, 77] We chose to use a standardized procedure to form the bromides.

The mono Boc protected alcohol (**2.85**) and mono Ac protected alcohol (**2.86**) were prepared *via* 4-hydroxybenzyl alcohol (**2.81**). The Ts analogue (**2.91**) was prepared from 4–hydroxybenzaldehyde (**2.89**). The alcohols were then converted to the relevant bromide in a two step procedure *via* the mesylate, which upon treatment with lithium bromide in THF yielded the desired bromide (scheme 2.21).

The Boc alcohol (**2.85**) was synthesised by the addition of one equivalent of di *tert*-butyl dicarbonate in the presence of DMAP as base in acetonitrile in 81% yield.[78] The alcohol was readily converted to the bromide (**2.87**) in 57% yield after chromatography. As predicted the bromide (**2.87**) was extremely stable and was purified by standard chromatography conditions. A stored sample remained unchanged after 1 month storage in a freezer under nitrogen.

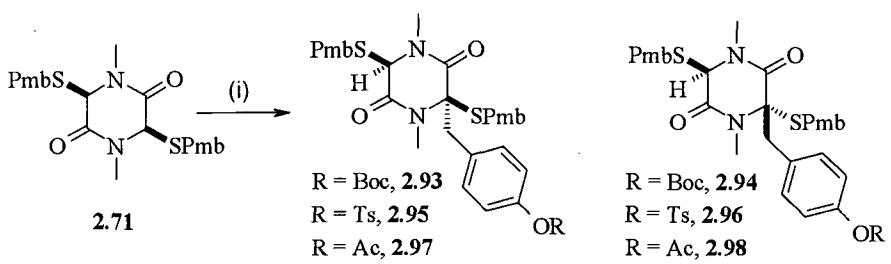
The acetyl analogue (**2.86**) was prepared by the careful addition of one equivalent of acetyl chloride to a solution of 4-hydroxybenzyl alcohol (**2.81**) and triethylamine in ethyl acetate at 0 °C. As before the alcohol was converted to the bromide (**2.88**) *via* a mesylate in 65% yield.[79] The Ac protected bromide (**2.88**) was also stable to chromatography, but does decompose after approximately two weeks storage under nitrogen in a freezer.

4-Hydroxybenzaldehyde (**2.89**) in acetonitrile and potassium carbonate was treated with tosyl chloride for 18 h at room temperature to give the aldehyde (**2.90**). Subsequent reduction with sodium borohydride in methanol yielded the alcohol (**2.91**) in 96% yield.[77] The stable bromide (**2.92**) was once again formed under the reported conditions in 62% yield.



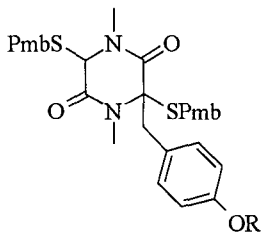
Scheme 2.21 Reagents and conditions: (i) Boc_2O (1.05 equiv), DMAP (0.05 equiv), CH_3CN, rt 1 h, 81%, (**2.85**);or (i) AcCl (1.05 equiv), TEA (1.05 equiv), EtOAc, 0 °C – rt, 5 h, 54%, (**2.86**); (ii) K_2CO_3 (1.4 equiv), tosyl chloride (1.2 equiv), CH_3CN, reflux 18 h, 96% (**2.90**); (iii) $NaBH_4$ (1.0 equiv), MeOH, 0 °C, 3 h, 96% (**2.91**); (iv) MsCl (1.1 equiv), Et_3N (2.0 equiv), THF, -40 °C, 45 min; (iv) (b) LiBr (4.0 equiv), THF, 0 °C, 1.5 h, 54 % (**2.87**); 65% (**2.88**); 62% (**2.92**).

As speculated, both the Boc and Ts protected bromide (**2.87** & **2.92**) cleanly reacted with the double deprotonated DKP (**2.71**) to give the mono alkylated products (**2.93**, **2.94** & **2.95**, **2.96**) with fewer by-products than had been previously observed (scheme 2.22). As the stable bromides could be purified it was also observed that the yields became more consistent under these reaction conditions, averaging around 50%. Unfortunately, due to the basic nature of the reaction conditions the acetyl protected bromide (**2.88**) gave a messier reaction mixture. The less polar desired *cis* DKP (**2.97**) could be isolated pure, but the more polar *trans* (**2.98**) co-eluted with the starting DKP (**2.71**) and the deprotected unassigned free phenol derivative. If the reaction time was increased the yield of the Ac analogue (**2.97**) decreased, indicating hydrolysis of the acetyl to the phenol.



Scheme 2.22 Reagents and conditions: (i) 2.0 M LDA (3.0 equiv), THF, -78 °C 30 min. (i) (a) bromide (1.2 equiv), THF, -78 °C, 2 h; (b) AcOH (1.5 equiv), THF, -78 °C rt, 44%, (**2.93**: **2.94**, 3:1); 63%, (**2.95**: **2.96**, 2:1), 46%, (**2.97** ; impure **2.98**, 2:1).

Unfortunately, no X-ray data were collected on the new analogues and therefore the configurations of these analogues, *cis* or *trans* was assigned by a combination of factors. Simply by comparing the polarity by TLC of the new analogues with that previously assigned by X-ray, the *cis* analogue is much less polar (more non polar) than the *trans*. Similar observations have been reported by Kawai *et al.*[66] This polarity difference can also be observed by LCMS, by comparison of the LC spectra (reverse phase column) of the *cis* and *trans* analogues in all the examples synthesised. The longer the retention time the more nonpolar (lipophilic) the product. (table 2.5).

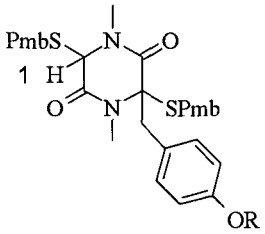


R group	*Cis* retention time	*Trans* retention time
Me	8.02 min(**2.74**)	7.58 min (**2.75**)
Isoprenyl	9.13 min (**2.57**)	8.78 min(**2.84**)
Boc	8.82 min (**2.93**)	8.25 min (**2.94**)
Ac	8.97 min (**2.97**)	(impure) (**2.98**)
Ts	8.57 min (**2.95**)	8.03 min (**2.96**)
H	7.14 min (**2.99**)	6.49 min (**2.100**)

Table 2.5 LCMS retention times recorded of cis and trans analogues synthesized.

Secondly the ^{1}H NMR chemical shift δ_H of the DKP proton (H1) is a characteristic singlet. In all examples the observed chemical shift δ_H for H1 in the *cis* isomer is higher than 4.00 ppm and the *trans* isomer is less than 4.00 ppm (table 2.6).

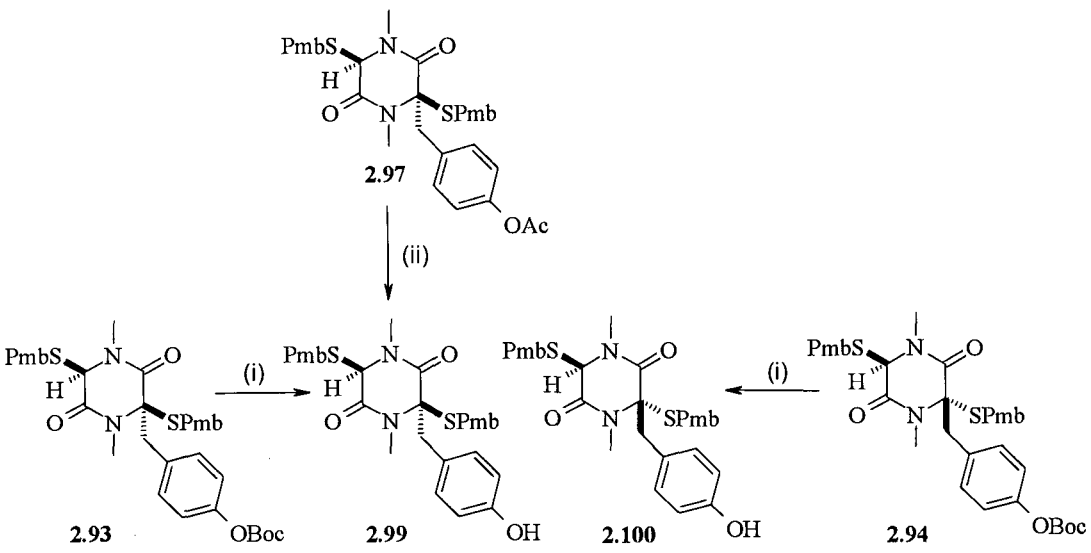
In addition they could be assigned by chemical means; the *cis* isomers, acetyl (**2.97**) and Boc (**2.93**) were deprotected. Compound (**2.93**) *via* acidic conditions gave the phenol (**2.99**) and compound (**2.97**) *via* basic conditions yielding the same product (**2.99**) as the major product plus a minor amount of a less polar product, later identified as the *trans* isomer (**2.100**). The *trans* Boc analogue (**2.94**) could also be deprotected under acid conditions and gave the less polar product (**2.100**) consistent for the expected *trans* assignment (scheme 2.23).



R group	δ_H (H1) *Cis* (ppm)	δ_H (H1) *Trans* (ppm)
Me	4.11 (**2.74**)	3.99 (**2.75**)
Isoprenyl	4.05 (**2.57**)	3.99 (**2.84**)
Boc	4.17 (**2.93**)	3.88 (**2.94**)
Ac	4.15 (**2.97**)	3.91 (impure) (**2.98**)
Ts	4.10 (**2.95**)	3.87 (**2.96**)
H	4.12 (**2.99**)	3.94 (**2.100**)

Table 2.6 Chemical shift δ_H for H1, of analogues synthesized.

The newly formed *cis* and *trans* phenols were individually treated with sodium hydride in DMF at 0 °C. Addition of the 3,3-dimethylallyl bromide yielded the previously reported isoprenyl compounds (**2.57** & **2.84**).

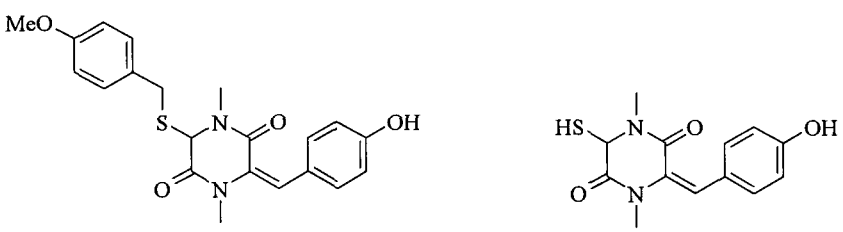


Scheme 2.23 Reagents and conditions: (i) 4 M HCl/ 1,4-dioxan, 0 °C, 18 h, 89%; (ii) 1.0 M KOH (2 equiv), MeOH, 0 °C – rt, 65% (**2.99**:**2.100**, 5:1).

We attempted to cleave the Pmb groups with a Lewis acid as before, and oxidise the resulting dithiols to the desired ETPs. We found that the *cis* acetyl (**2.97**) and Boc

(**2.93**) compounds did not give the desired disulfide products, but a complex mixture as before. Both the Boc and acetyl groups appeared to be cleaved with boron trichloride or tribromide. In fact the Boc group was cleaved at -78 °C within minutes. This discovery provided a new rapid, clean method of preparing the phenol (**2.99**).

In addition we tried the deprotection on the free phenol (**2.99**), but this was also unsuccessful. To further understand why the reaction mixtures were so complex, we first monitored the deprotection step with the Lewis acid. This was carried out on both the Ac compound (**2.97**) and the phenol (**2.99**). Both examples were treated with boron tribromide at -78 °C in dichloromethane for 10 minutes then allowed to warm to 0 °C for 20 minutes and room temperature for 10 minutes. The reactions were monitored by LCMS at each time point. For the phenol analogue (**2.99**) after 20 minutes at 0 °C only one peak was observed at 3.5 min. *m/z* $[M+H]^+$ = 279.1 and $[M+H+CH_3CN]^+$ = 320.0. This mass ion identified could be the thiol (**2.102**) Mwt = 278.33 (figure 2.10). This mass ion was also observed for the compound (**2.97**) but in addition after 10 minutes we also observed a earlier peak at 5.42 min. *m/z* $[M+H]^+$ = 399.1 & $[M+H+CH_3CN]^+$ = 440.1. This peak had completely gone after the 20 minute time point and only the one peak at 3.5 minutes was recorded. This earlier product at 5.42 minutes had been previously observed and isolated as a minor by-product during the acid deprotection of the *tert*-butyl carbonate (**2.93**) to the phenol (**2.99**). ^1H NMR and MS data of this material identified the structure to be eliminated SPmb phenolic product (**2.101**). These findings explain why the reaction is not giving the desired ETP, first the protecting group is cleaved from the phenol (in the acetyl analogue), subsquently the *p*-methoxybenzyl mercaptan group is eliminated to give the mono SPmb protected thiol (**2.101**). Finally this is deprotected to yield the mono thiol (**2.102**).



2.101
M.W. 398.48
$C_{21}H_{22}N_2O_4S$

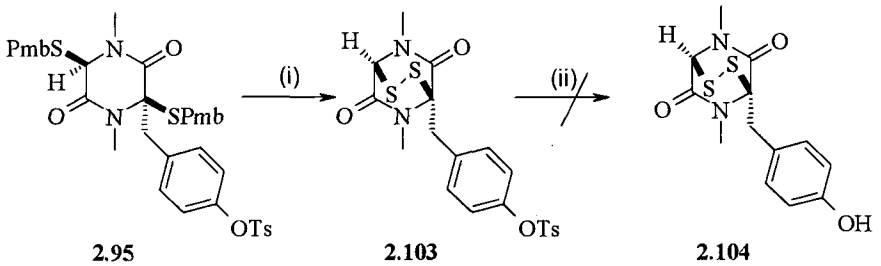
2.102
M.W. 278.33
$C_{13}H_{14}N_2O_3S$

Figure 2.10 Plausible structures for the mass ions observed during SPmb deprotection reaction.

The acid stable Ts compound (**2.95**) could be readily converted to the ETP (**2.103**) in 71% yield by this two step procedure, indicating the *cis* and *trans* isomers had been correctly assigned (scheme 2.24).

The evidence indicates thus far that in order to effect the desired deprotection of the Pmb group with boron trichloride the phenol must be protected by an acid stable group. The problem with this is that the acid stable protecting groups generally require harsh conditions to remove them as demonstrated by the tosyl ETP (**2.103**).

The standard methods for cleaving the tosyl are *via* strong base hydrolysis, refluxing the substrate in potassium hydroxide for several hours.[80] An alternative method was by reductive conditions such as magnesium in methanol.[81] Both these conditions were attempted without success. This was not unforeseen as Kishi had reported that ETPs were not stable to base or strong acid conditions. The base instability of gliotoxin was reported by Woodward *et al.*[82]



2.95 **2.103** **2.104**

Scheme 2.24 Reagents and conditions: (i) (a) 1.0 M BCl_3 (2.4 equiv) 0 °C 10 min, evaporate; (i) (b) I_2 (2.0 equiv), 10% MeOH/ CH_2Cl_2, rt, 20 min, 71%; (ii) 2.0M KOH (aq), reflux, failed; or Mg, MeOH, failed.

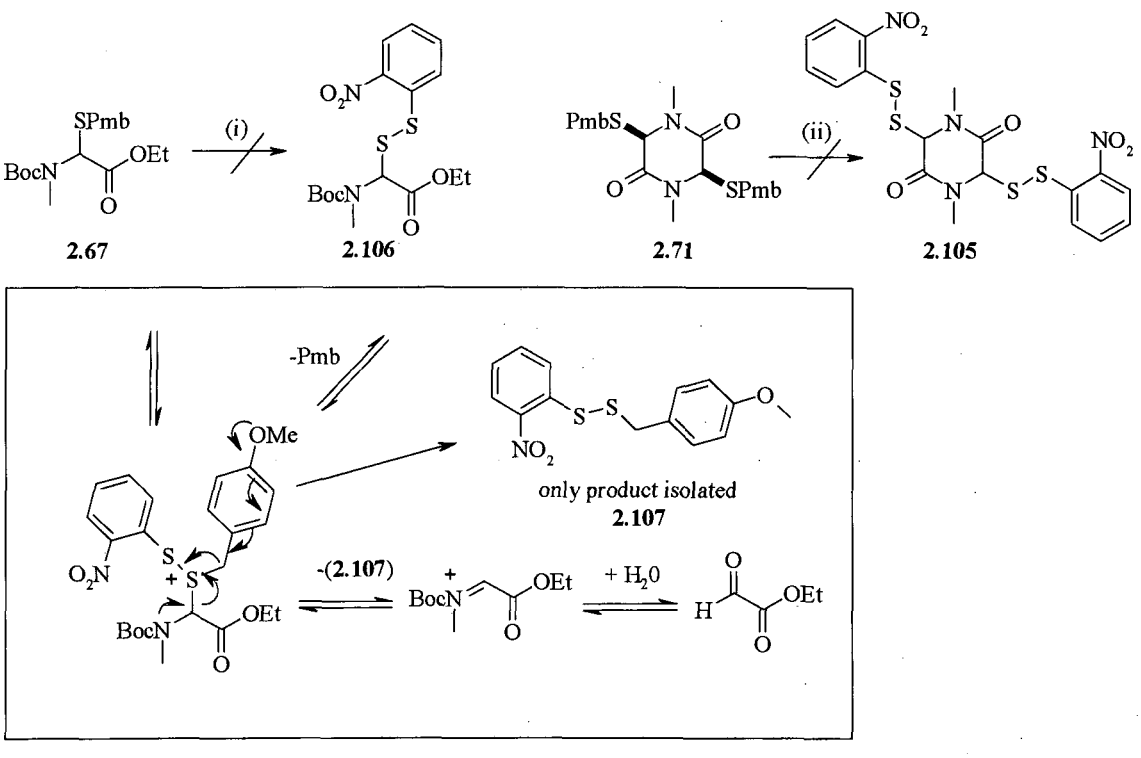
II.4.7 Synthesis of Alternative *Bis* Protected Dithiodiketopiperazines

To avoid the instability issues of the epidithiodiketopiperazine under both base and acid conditions, we concentrated on investigating new strategies in which the disulfide bridge is once again formed in the last step of the synthesis.

We investigated approaches that avoided the problematic Pmb group, which requires harsh acidic conditions to remove it. One possible route was *via* 2-nitrobenzenesulfenyl chloride as outlined by Reese *et al.*[83] They reported similar

problems with the Pmb group and simply replaced it with a disulfide analogue in a one step procedure under mild conditions.

Two trial reactions on the protected amino acid ester (**2.67**) and the DKP (**2.71**) were attempted without success. In both cases it was found that the undesired disulfide of SPmb (**2.107**) was formed when the Pmb protected substrate was reacted with 2-nitrobenezenesulfenyl chloride. No protecting group transfer had occurred to give the desired compounds (**2.105**) and (**2.106**). Only cleavage of the sulfur unit was observed. This can be explained by the Pmb protecting group on the sulfur not being eliminated as expected, as it was not sufficiently electron rich. The desired electron movement as indicated by the black arrows does not occur. Instead electron movement indicated by the blue arrows leads to the elimination of the disulfide (**2.107**) (scheme 2.25).

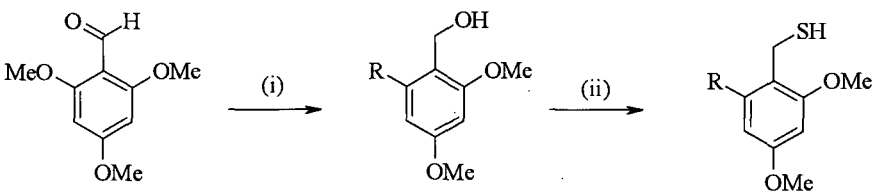


Scheme 2.25 Reagents and conditions: (i) 2-nitrobenzenesulfenyl chloride (1.1 equiv), 10% $AcOH/CH_2Cl_2$, 0 °C, 1 h; (ii) 2-nitrobenzenesulfenyl chloride (2.1 equiv), CH_2Cl_2, 0 °C, 1 h. Box: plausible reaction mechanisms

4.8 Synthesis of Electron Rich Protecting Groups

In light of these findings we chose to investigate more electron rich protecting groups. If the protecting group was more electron rich it would be possible to use milder deprotection conditions than currently used to remove the Pmb. This could be achieved by adding further methoxy groups to the phenyl ring. Therefore synthetic routes towards the 2,4-dimethoxybenzyl (Dmb) and 2,4,6-trimethoxybenzyl (Tmob) DKP's were undertaken.

Both 2,4-dimethoxybenzyl thiol (**2.111**) and 2,4,6-trimethoxybenzyl thiol (**2.112**) are not commercially available, but were readily synthesised. The 2,4-dimethoxybenzyl thiol (**2.111**) was synthesised in a one step procedure *via* commercial alcohol in 84% yield. Likewise, the 2,4,6-trimethoxybenzyl thiol (**2.112**) was synthesised in two steps from the aldehyde (**2.108**) also in good yield (scheme 2.26).[84]



2.108

R = H, **2.109**
R = OMe, **2.110**

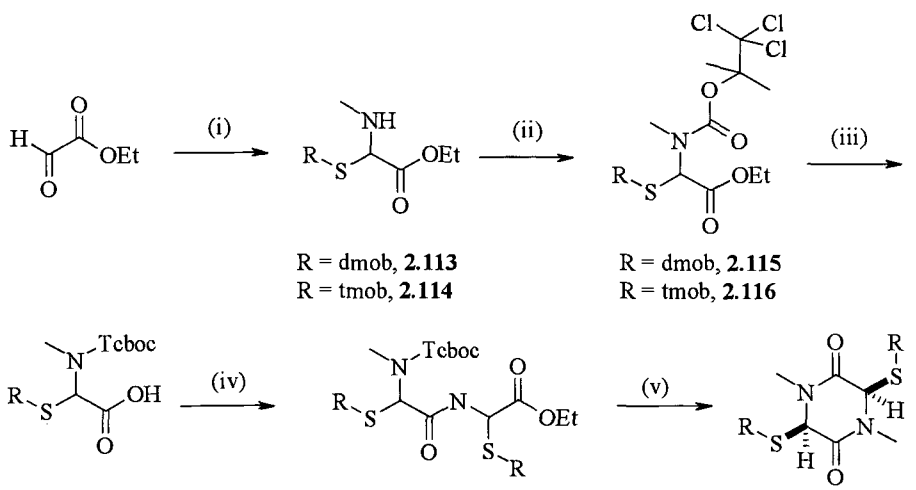
R = H, **2.111**
R = OMe, **2.112**

Scheme 2.26 Reagents and conditions: (i) $NaBH_4$ (1 equiv), MeOH, 0 °C, 3 h, 94% (**2.110**); (ii) (a) thiourea (2 equiv), 5N HCl (15 ml), acetone / water 1:1 (100 ml); (b) NaOH (3 equiv), 84% (**2.111**), 77% (**2.112**).

Both the thiols progressed smoothly to the corresponding amino acid esters (**2.113** & **2.114**) *via* the three component procedure previously reported in 62% and 38% respectively.

Both the dmob and tmob protecting groups are increasingly more acid sensitive than the Pmb. This meant the acid cleaved carbonate, Boc protection strategy could lead to undesired cleavage of the thiol protection. Therefore an alternative approach was developed (scheme 2.27). We achieved this *via* 2,2,2-trichloro-1,1-dimethylethyl carbamate (Tcboc) protection of the amine, which is readily removed under mild conditions by the addition of activated zinc powder.[85] The amino acid esters (**2.113** &

2.114) were smoothly converted to the carbonates (**2.115** & **2.116**) by the addition of 2,2,2-trichloro-1,1-dimethylethyl chloroformate in the presence of triethylamine in dichloromethane in almost quantitative yield. Further elaboration to the dipeptides (**2.119** & **2.120**) was completed under similar reaction conditions as previously reported. The acids (**1.117** & **1.118**) were synthesised by saponification of the esters (**2.113** & **2.114**) in 73% and 83% yield, then converted to the dipeptide (**2.119** & **2.120**) *via* a mixed anhydride coupling. We observed that the diketopiperazines (**2.121** & **2.122**) could be synthesized in a one step process from the starting Tcboc analogues (**2.119** & **2.120**). In both examples the DKP (**2.121** & **2.122**) were isolated when the dipeptides (**2.119** & **2.120**) were treated with activated zinc powder in 10% acetic acid, IPA solution at room temperature for 18 h. This resulted in yields of 50% and 34% respectively. Once again only one isomeric form was isolated, it was arbitrarily assigned *cis*. This could not be confirmed as crystals of good enough quality for X-ray experiments could not be obtained.

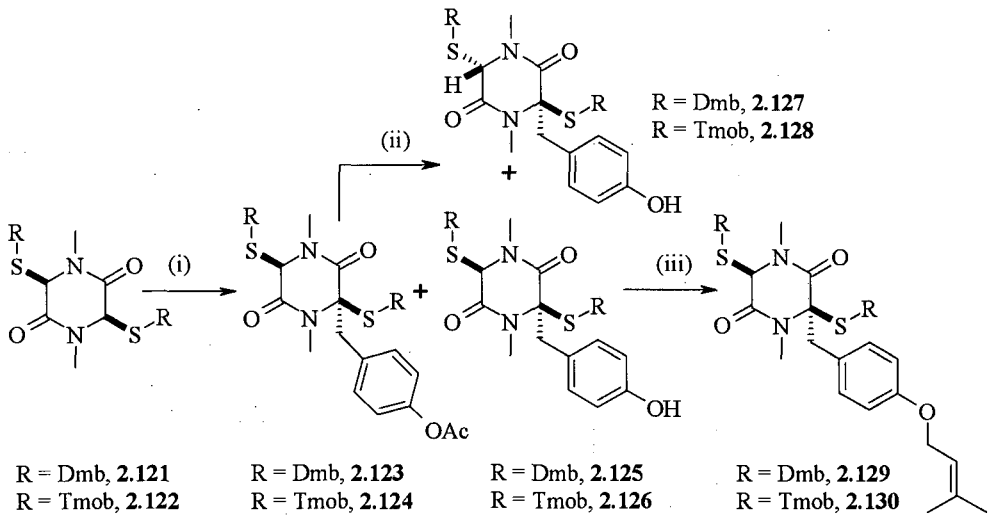


Scheme 2.27 Reagents and conditions: (i) 2.0 M methylamine in THF (1 equiv), (**2.111**) or (**2.112**) (1 equiv), Na_2SO_4 anhydrous (excess), rt, 18 h, 62% (**2.113**), 38% (**2.114**); (ii) 2,2,2 trichloro-1,1-dimethylethyl chloroformate (1.1 equiv), triethylamine (1.2 equiv), CH_2Cl_2, rt, 18 h, 94%, (**2.115**), 97% (**2.116**); (iii) $LiOH.H_2O$ (1.5 equiv), aq 1,4-dioxane, rt, 18 h, 89%, (**2.117**), 83% (**2.118**); (iv) (a) *iso*BuOCOCl (1.1 equiv), NMM (1.2 equiv), THF, -10 °C, 1 h; (iv) (b) (**2.113**) or (**2.114**) (1.1 equiv), NMM (1.2 equiv), THF, rt, 18 h, 41% (**2.119**), 45% (**1.120**); (v) zinc powder (20 equiv), 10% AcOH/IPA, rt, 18 h, 50% (**2.121**), 34% (**2.122**).

Both the dmob and tmob protected, isoprenyl analogues (**2.129** & **2.130**) were synthesised. The DKP's were treated with 3 equivalents of LDA as before and

alkylated with acetic acid 4-bromomethylphenyl ester (**2.88**), upon quenching the reaction mixture with acetic acid, to give the *cis* acetyl compounds (**2.123** & **2.124**) and the *cis* phenol analogues (**2.125** & **2.126**). In addition the *cis* acetyl analogues (**2.123** & **2.124**) were deprotected under basic conditions to give a 9:1 mixture of the *cis*:*trans* phenols (**2.125**: **2.127**) & (**2.126**:**2.128**).

As previously reported (table 2.5 & 2.6) the configuration of the isomers could be arbitrarily assigned by comparing the relative polarities of the two isomers by LCMS and TLC and the proton shifts δ_H by ^{1}H NMR. The *cis*-phenolic compounds (**2.125**) and (**2.126**) were treated with sodium hydride at -10 °C in DMF and alkylated with 3,3-dimethylallyl bromide to give isoprenyl compounds (**2.129** & **2.130**) in 41% and 45% yield (scheme 2.28).



Scheme 2.28 Reagents and conditions: (i) (a) 2.0 M LDA (3.0 equiv), THF, -78 °C, 30 min; (i) (b) Acetic acid 4-bromomethylphenyl ester (1.5 equiv), THF, -78 °C, 2 h; (i) (c) AcOH (1.5 equiv), THF, -78 °C, rt, 30% (**2.123**) & 31% (**2.125**): 21% (**2.124**) & 31% (**2.126**); (ii) 1.0 M KOH (2 equiv), MeOH, 0 °C – rt, 71% (**2.125**: **2.127**, 9:1), 96% (**2.126**: **2.128**, 9:1); (iii) (a) NaH (1.05 equiv), DMF, -10 °C, (iii) (b) 3,3-dimethylallyl bromide (1.2 equiv), rt, 4 h, 41% (**2.129**), 45% (**2.130**).

In both cases a small trace of the *trans* isomer was observed by TLC and LCMS, but material of sufficient purity for full characterisation was not isolated. A number of conditions were attempted for the final disulfide step to form dithiosilvatin. In the literature both the Tmob and Dmb can be cleaved under acid conditions.[86, 87] Therefore the acid stability of the isoprenyl group was investigated (table 2.7).

Conditions	Products observed by LCMS
50% formic acid / DCM	Loss of isoprenyl (10 min). Multiple products after 1 h
50% TFA / DCM	Loss of isoprenyl (10 min). Multiple products after 1 h
2% TFA / DCM	No change (10 min). Loss of isoprenyl after 1 h
1% TFA / DCM	No change (1 h). Loss of isoprenyl after 24 h

Table 2.7 Acid stability of Tmob protected analogues (**2.129 & 2.130**).

Unfortunately it was observed that the isoprenyl group was extremely acid labile. During all the test experiments no fragments relating to the desired cleavage of the thiol protecting group was seen, in either the Dmb (**2.129**) or Tmob (**2.130**) compounds.

Similar findings were observed when the Tmob protected analogue (**2.130**) was treated with boron tribromide (2 equiv) at -78 °C in dichloromethane. The only product observed by LCMS was the *m/z* $[M+H]^+$ 673.1, $[M+Na]^+$ 695.1 peak at 8.56 min, which is identical to that observed for the hydroxyl analogue (**2.126**), indicating the undesired loss of the isoprenyl group, even at low temperatures.

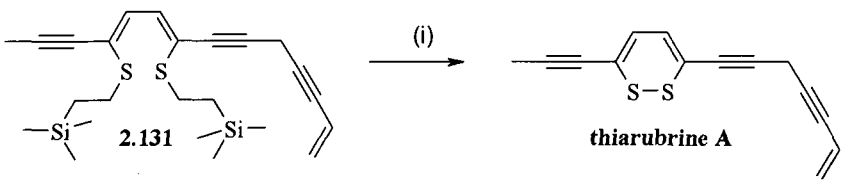
In parallel a number of oxidation methods were attempted to form the disulfide (table 2.8).[88-91] Once again at no stage was any dithiosilvatin ever observed by LCMS. Under all the conditions attempted many products were formed and no starting material remained.

Conditions	Products observed
I_2 (4 eq) $EtOH/CH_2Cl_2$, rt[88]	Multiple products after 1 h, no desired mass ion
I_2 (2-4 eq) DMF, rt	Products after 1 h loss of thiol, no desired mass ion
PhSOPh, CH_3SiCl_3, CH_2Cl_2[89]	Loss of thiol observed major product
DMSO/10% TFA[90]	Loss of isoprenyl after 10 min, multiple products
$Tl(CF_3COO)_3$, DMF, anisole[91]	Multiple products after 1 h, no desired mass ion

Table 2.8 Investigation of disulfide bond forming reactions by oxidation (reagent **2.130**).

II.4.9 Synthesis of *Bis* Silyl Analogue

It was becoming clear that no acidic reaction conditions were tolerated for the final steps of the synthesis of dithiosilvatin, as the isoprenyl group was extremely acid sensitive. This vastly limited our choice of thiol protecting groups. We chose to investigate possible silyl protection strategies. As the silicon-sulfur bond is weaker than a silicon-oxygen bond it is more susceptible to hydrolysis, therefore direct silyl protection of the sulfur was not a viable option, but indirect β-silyl protection of sulfur has been reported.[92] The novel use of the 2-(trimethylsilyl)ethyl as the thiol protecting group was reported in the construction of thiarubrine A. The *bis* silyl protected compound (**2.131**) was successfully cleaved with TBAF as a source of F^- and oxidation with iodine yielded the desired natural product (scheme 2.29).[92]



Scheme 2.29 Reagents and conditions: (i) (a) TBAF (8.0 equiv), 3 Å mol. sieves/ THF, rt, 1h; (b) I_2 (10 equiv), rt, 30 min, 53%.

Initially we attempted to adopt the same methodology we had developed in the synthesis of a number of *bis* protected dithiodiketopiperazines. Once again we utilised the three component reaction to form the key building blocks. The three component reaction progressed to the desired amino acid analogue (**2.132**) albeit in a lower yield than previous derivatives. We opted for the Tcboc route, as we presumed the silyl protecting group may be acid sensitive. The Tcboc protected ester (**2.133**) was synthesised in quantitative yield and was readily saponified to the acid (**2.134**) in 70% yield. The coupling reaction to the dipeptide (**2.135**) proved to be problematic and poor yielding. A similar result was observed for the DKP (**2.136**). The best yield obtained was an extremely poor 14% (scheme 2.30).

Due to the two successive poor yielding steps the route was not feasible to synthesise sufficient material to investigate the new protection strategy further. We therefore attempted alternative routes. Two routes were investigated, one using the Trown

exclusion), which is imposed at the V_H to D-J_H step in recombination.[44,45] Surface expression and signalling through the pre-BCR is also necessary to draw the B cell, now called a pre-B cell, from G_o phase into cell cycle. A period of cell proliferation, usually involving two to five rounds of cell division, leads to a clone of pre-B cells all producing the same μ heavy chain; during this stage the cells are called large pre-B cells.[46]

Expression of the pre-BCR initiates light chain rearrangement, which takes place first at the Igκ locus. Following proliferation, the RAG genes are re-expressed and the small pre-B cells, which are now no longer dividing, undergo Vκ to Jκ rearrangement. If the first attempts are non-productive, pre-B cells are able to repeat Vκ-Jκ joining several times until all the Jκ genes are exhausted; this process is known as light chain rescue.[28] If rearrangement is productive, the Vκ-Jκ unit is transcribed under the influence of the Igκ enhancers (intronic Eκ and 3' Eκ) as a single mRNA incorporating the Cκ gene. The primary transcript is spliced and translated and the κ chain assembled at the cell surface; the cell becomes an immature B cell expressing membrane IgM(κ) BCR (Figure 1.8.). Further rearrangement at the Ig loci is suppressed, although, allelic exclusion at the light chain loci is less stringent than for the heavy chain. Should Vκ rearrangement fail on both alleles, the kappa enhancer is deleted and rearrangement proceeds at the Igλ locus.[47,48] As with the κ locus, multiple Vλ to Jλ rearrangements are possible in order to generate a productive Vλ-Jλ transcriptional unit. Successful rearrangement leads to the production and expression of IgM(λ) BCR. In humans, approximately 60% of B cells express κ light chains and 40% express λ light chains.[49] Occasionally, rearrangement at the λ locus may occur without prior rearrangement at the κ locus.[50] Rarely, B cells may express both κ and λ protein simultaneously.[51]

The recombination process is typically inactivated from the immature B-cell stage onwards. The expression of surface IgD from the rearranged V_H gene is achieved by alternative splicing of the primary RNA transcript and marks the transition of immature B cells to naïve or virgin mature B cells, which then migrate to the secondary lymphoid organs and encounter Ag. It is estimated that the bone marrow (BM) produces about 5 x 10^7 B cells per day, but that only 5 x 10^6 are actually recruited to the circulating B-cell pool.

Failure to generate productive rearrangements and express Ig at the appropriate times during development results in the death of the developing B cell by apoptosis. The

majority of developing B cells (>75%) become apoptotic during the pre-B-cell stages. Similarly, the binding of self-Ag in the BM leads to apoptosis of the auto-reactive B cell (clonal deletion). However, self-reactive B cells may be rescued from apoptosis by further rounds of rearrangement at the light chain loci (receptor editing).[52-54] Following receptor editing, a new light chain is generated and expressed and conveys an altered BCR specificity. Salvage of auto-reactive B cells by receptor editing is also known to occur in the secondary lymphoid, where it is designated receptor revision.

1.2.3.2. Mechanism of V(D)J recombination

V(D)J recombination involves the recognition of recombination signal sequences (RSSs), which are adjacent to each germline V, D and J gene segment. RSSs are highly conserved among vertebrates. Each RSS contains a conserved palindromic heptamer sequence and a conserved AT-rich nonamer sequence separated by a spacer of non-conserved DNA, which can be either 12 or 23 nucleotides in length (Figure 1.9a.). The 12- and 23bp spacers roughly correspond to one and two turns of the DNA double helix, respectively, and for this reason, sequences are also referred to as one-turn and two-turn RSSs. The length of the spacer is important in determining the functionality of the RSS since efficient recombination can only occur between a one-turn RSS and a two-turn RSS (one-turn/two-turn joining rule).[55] The RSS spacer lengths at each gene segment are positioned so that recombination is directed towards functional products (Figure 1.9b.). At the heavy chain locus, a two-turn RSS is located 3' to each V_H gene and 5' to each J_H gene while a one-turn RSS is located both 5' and 3' of each D gene. At the kappa locus, a one-turn RSS is located 3' of each Vκ gene and a two-turn RSS 5' of each Jκ gene. At the lambda locus, this order is reversed; a two-turn RSS is located 3' of each Vλ gene and a one-turn RSS 5' of each Jλ gene.

The majority of V(D)J recombination is achieved by an intrachromosomal, deletional recombination event (Figure 1.10a.). The common arrangement of RSSs at the Ig loci is such that, following recombination, the joined coding segments remain in the chromosome and the junction of the RSSs, termed a signal joint, is excised as circular DNA, which is later lost during cell division.[56] However, some loci contain segments in an inverted orientation so that both the coding joint and signal joint are retained in the chromosome (inversion recombination; Figure 1.10b.). A notable example is the human κ locus where the distal Vκ gene segments lie in an inverted orientation relative to the Jκ and Cκ genes.[57] Inversional D to J_H rearrangements, which are technically permitted by the one-turn/two-turn rule, have been described, but the vast majority of D to J_H joints occur by deletion.[58]

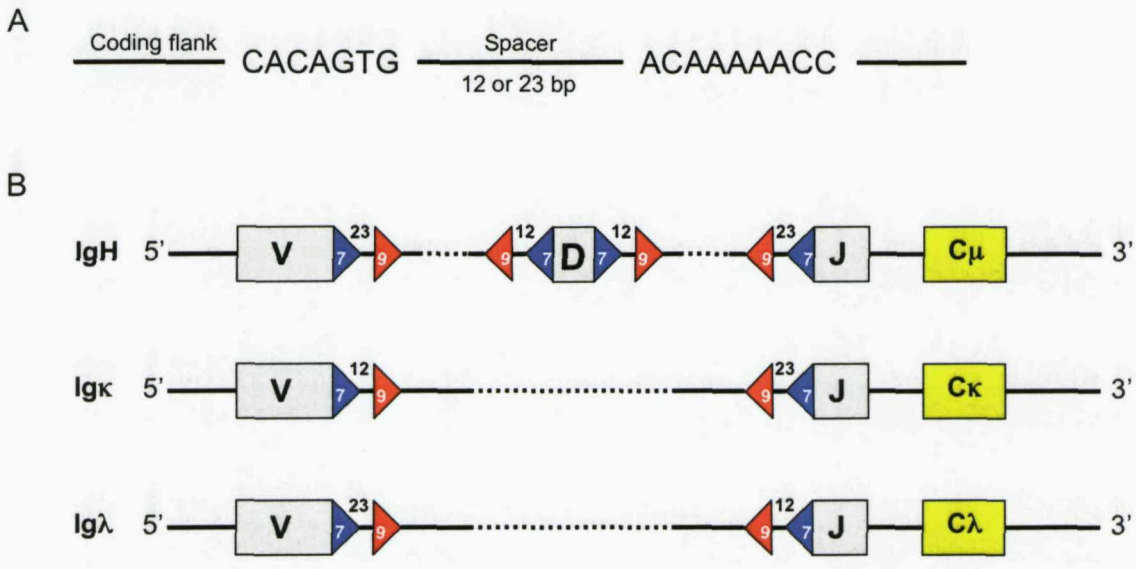


Figure 1.9. Recombination signal sequences (RSS) and their arrangement at the immunoglobulin loci. A) The consensus heptamer and nonamer sequences of a RSS with the alternative spacer lengths of 12- and 23bp. B) Schematic representation of the one-turn/two-turn joining rule governing DNA rearrangement at the immunoglobulin heavy and light chain loci. V, D and J gene segments are represented by grey rectangles; constant regions are represented by yellow rectangles; heptamers and nonamers of the RSSs are represented by red and blue triangles, respectively. The spacer lengths are indicated above the appropriate lines: 12bp (one-turn RSS) and 23bp (two-turn RSS). Dashed lines indicate intervening DNA, which is deleted during recombination. At each locus, all gene segments of one type have the same RSS arrangement. Adapted from Gellert, 2002.[38]

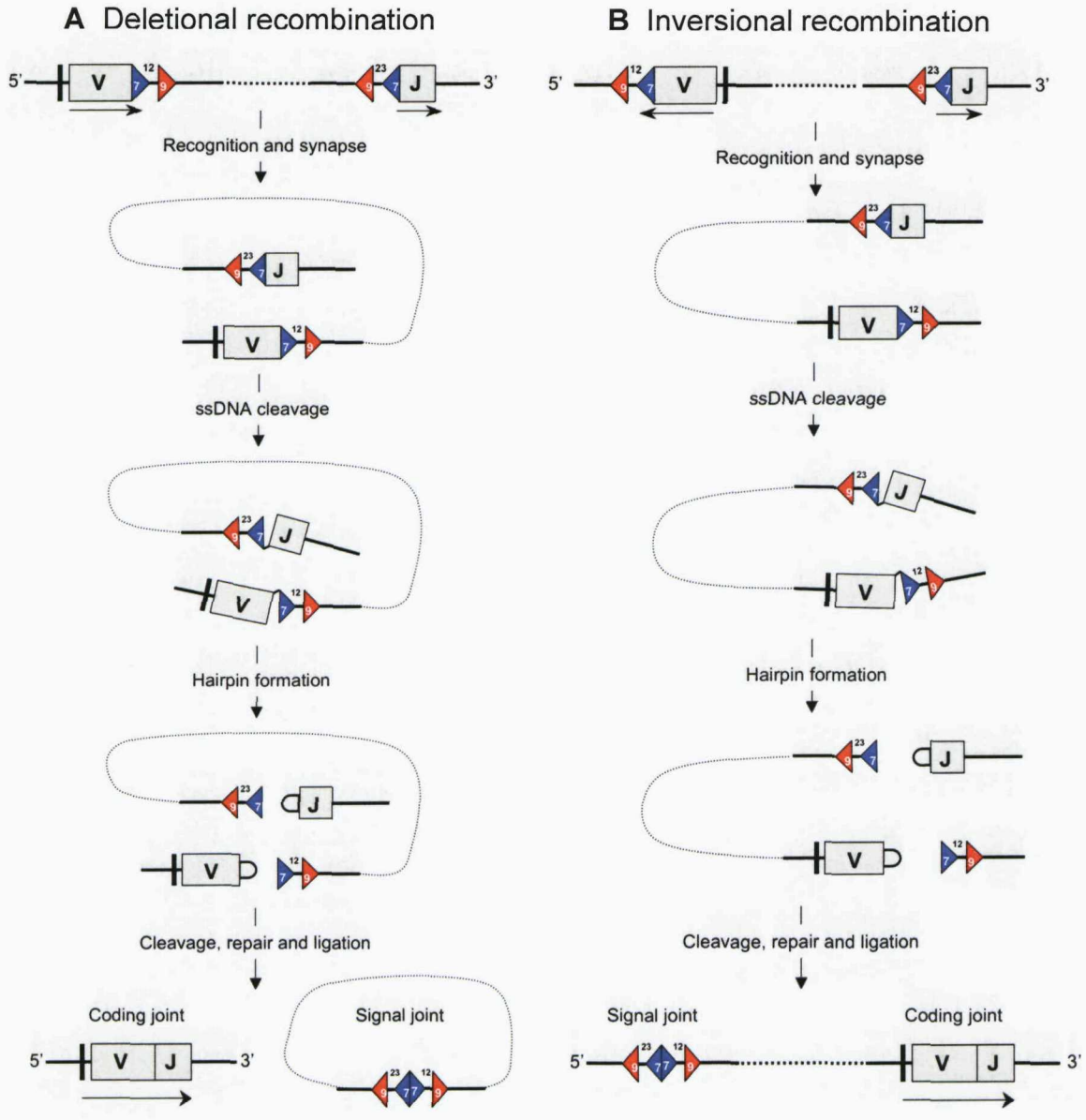


Figure 1.10. Mechanisms of V(D)J recombination. A) Deletional recombination occurs when the gene segments to be joined have the same transcriptional orientation, indicated by *black* arrows. This process yields two products: a rearranged VJ unit that includes a coding joint and a circular DNA excision product consisting of the recombination signal sequences (RRSs), signal joint and intervening DNA. B) Inversional joining occurs when the gene segments have opposite transcriptional orientations. In this case, the RRSs, signal joint and intervening DNA are retained, and the orientation of one of the joined segments is inverted. Adapted from Goldsby *et al.*, 2000.[1]

The process of V(D)J recombination can be divided into a number of stages:

1. Cleavage of RSSs and the formation of DNA hairpins

V(D)J recombination is mediated by two genes; *RAG1* and *RAG2*. The *RAG* genes are the only lymphoid specific factors required for recombination as evidenced by their ability to induce the recombination of artificial substrates in non-lymphoid cells.[59] Disruption of either the *RAG1* or *RAG2* gene renders mice completely defective in V(D)J recombination and, therefore, these mice contain no mature B or T cells in either primary or peripheral lymphoid tissue leading to a complete severe combined immunodeficiency (SCID).[60] The enzymatic function of the RAG proteins is to initiate V(D)J recombination by cleaving DNA between the RSS heptamer and the flanking coding sequence. Cleavage at the heptamer by the RAG proteins may involve both specific sequence recognition and DNA unpairing at the RSS. RAG1 is thought to be responsible for the cleaving of double-stranded (ds) DNA at the RSS. However, RAG1 is only enzymatically active when complexed with RAG2. Therefore, successful cleavage and subsequent recombination requires the cooperation of both RAG1 and RAG2.[61,62] Studies have shown that RAG1 alone binds to DNA with a moderate preference for the RSSs, specifically for the nonamer sequence. RAG2 can also bind DNA, but it has no specificity for RSSs. The complex of RAG1 and RAG2 proteins (RAG1/2) binds RSSs with a much greater affinity and specificity than either protein alone and displays contacts at both the heptamer and nonamer sequences.[63]

The RAG1/2 complex efficiently cuts an RSS in a DNA fragment to yield blunt signal ends and hairpin coding ends that retain the full coding sequence (Figure 1.11.).[64] Signal ends are cut exactly at the border between the RSS heptamer and coding sequence and are blunt ended with a 5'-phosphoryl and a 3'-hydroxyl group. Coding ends are exclusively DNA hairpins, with the 5' and 3' termini of the coding flank covalently joined. Cleavage occurs in two steps; first a single-stranded (ss) nick is made at the 5' end of the signal heptamer, leaving a 5'-phosphoryl group on the RSS and a 3'-hydroxyl group on the coding end. The second step joins this 3'-hydroxyl to the phosphoryl group at the same nucleotide position on the opposite strand, resulting in a hairpin coding end and blunt signal ends. *In vitro*, cleavage by RAG1/2 is more efficient at a one-turn RSS than a two-turn RSS. However, cleavage is increased, particularly at a two-turn RSS, by the addition of the chromosomal high-mobility-group proteins, HMG1 and HMG2.[65] HMG1 and HMG2 are non-specific DNA-binding and -bending proteins that may act to alter DNA

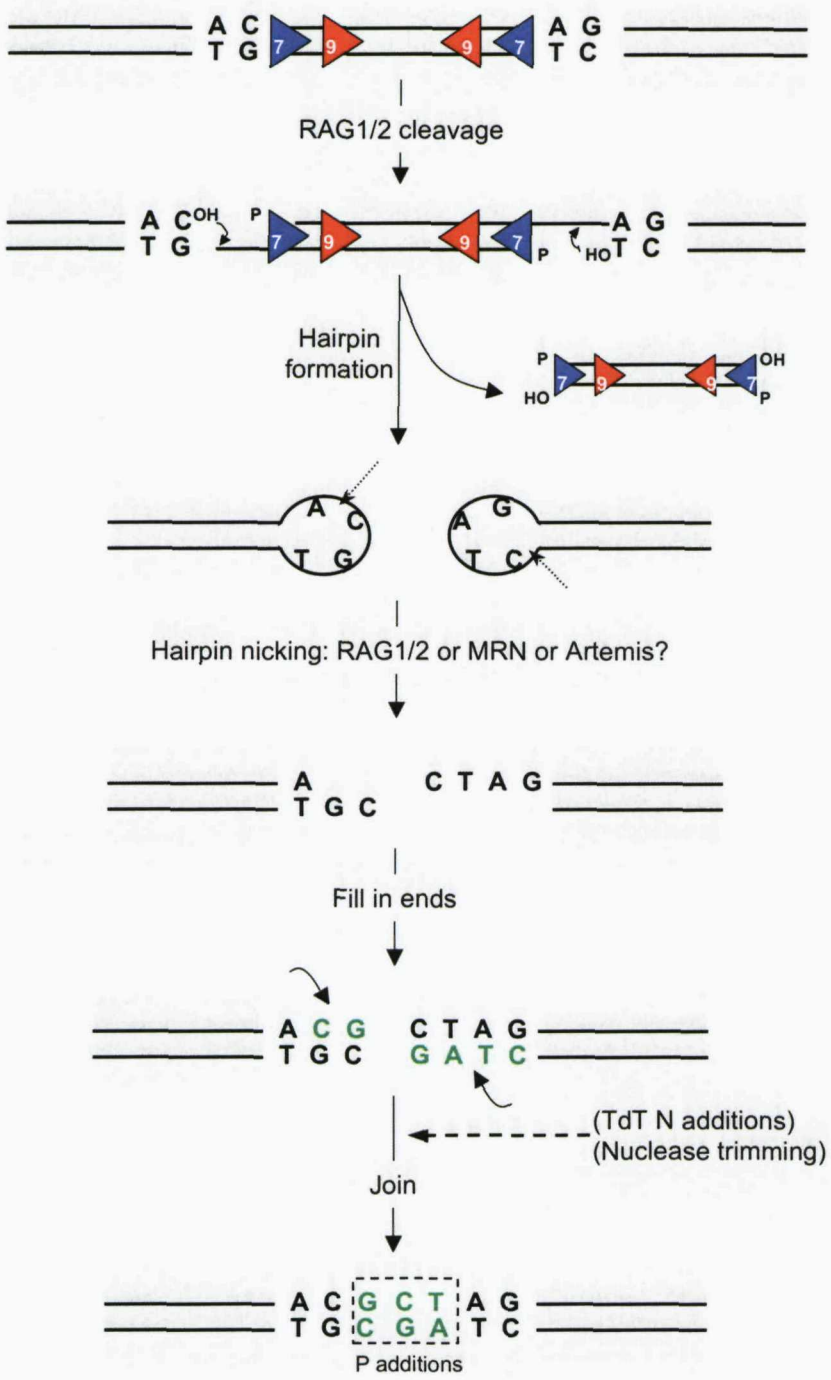


Figure 1.11. How self-complementary P nucleotide additions arise in coding joints. During DNA cleavage by the RAG1/2 complex, a single-stranded nick is made between the RSS heptamer and the coding sequence. This results in a 3' hydroxyl group on the coding flank that attacks the opposite strand to produce a hairpin coding end. These hairpins can be nicked a few bases off-centre (shown here as one base off-centre on the left and two bases off-centre on the right). This nicking leaves self-complementary ssDNA extensions, which after fill in and joining become incorporated into the junction. Trimming of ssDNA overhangs and/or N additions can occur prior to joining and further influence the junction. Adapted from Gellert, 2002.[38]

conformation and allow better RAG1/2 binding. Indeed, HMG1 and HMG2 proteins are significant V(D)J recombination cofactors *in vivo*.

Most RAG1/2-induced DNA breaks occur as coupled cuts occurring at two RSS sites simultaneously.[66] A one-turn RSS and a two-turn RSS pairing is most effective. The complex accounts for most of the one-turn/two-turn specificity of RAG cleavage since the normal 12- and 23bp spacer lengths are such that any proteins bound to the heptamer and nonamer DNA would be in the same helical phase. The preference of RAG1/2 for a one-turn/two-turn pair over a one-turn/one-turn or a two-turn/two-turn is raised by the presence of HMG1 or HMG2.[65] Following RSS cleavage, RAG1/2 and HMG proteins remain bound in a complex with the resulting 5' phosphorylated blunt signal end and the hairpin coding end in what is known as the postcleavage complex.[67]

2. DNA hairpin opening and processing

The hairpin coding ends produced by RAG1/2 cleavage must be reopened before the non-homologous end-joining (NHEJ) pathway can process and join them. Several factors have been proposed to perform this function. It has been reported that the RAG1/2 complex can produce such breaks, either on a hairpin substrate or on hairpins made in the same reaction by RAG1/2 cleavage.[68] Similarly, evidence for the participation of the MRN complex, which consists of the proteins Mre11, Rad50 and Nijmegen breakage syndrome protein 1, in hairpin opening has been published.[69] However, a gene product called Artemis, known to have nuclease activity, has become the most popular candidate for the opening of hairpin coding ends.[70,71] Defects in the *Artemis* gene lead to an accumulation of DNA hairpin intermediates and a low efficiency of coding joint formation.[70,71] *In vivo*, hairpin opening also depends upon the presence of active DNA-dependent protein kinase (DNA-PK), for which Artemis is known to associate via the catalytic subunit (DNA-PK_{CS}).[72]

The detailed structure of the coding joints provide clues as to the processing of coding sequences before joining. It appears that nicking does not target a specific site at the centre of the hairpin and that the position of the hairpin nick can vary. Off-centre nicking of the DNA hairpin intermediates results in a templated type of nucleotide addition, termed palindromic or P addition (Figure 1.11.).[73] These P nucleotide insertions add a few nucleotides complementary to the last bases of the coding end next to the RSS. The random deletion of terminal nucleotides by nucleases at the coding ends also occurs. Not all coding junctions have P nucleotide additions; if the hairpin is nicked exactly in the centre there is no self-complementary overhang, alternatively, overhangs may be removed

before the ends are joined. A second type of nucleotide addition, termed non-templated or N addition, is also found in coding joints. Non-templated tracts of up to 15 nucleotides in length are added by the enzyme terminal deoxynucleotidyl transferase (TdT), as evidenced by their absence in mice with a disruption in the *TdT* gene.[74] TdT expression adds nucleotides without a template to the ends of DNA chains, but with a preference for G residues that results in N regions being GC-rich. It was initially claimed that TdT expression is restricted to the pro-B-cell stage in normal B cells and therefore, that N additions occur only in heavy chain DNA. However, this has been revised following evidence of TdT activity in both Vκ and Vλ rearrangements.[48,75,76]

The processes discussed above leave the sequence at coding joints highly variable. This junctional diversification is significant for Ag binding (see section 1.2.3.4.). Conversely, signal joints are usually precise end-to-end fusions of two heptamer sequences; only a small fraction of signal joints have nucleotides inserted between the heptamers and loss of nucleotides is rare.[77] Junctional diversification mechanisms dictate that many junctions are non-productive since the number of nucleotides added or lost is essentially random and may lead to a change in the reading frame causing the premature termination of the protein chain. It is estimated that only one third of V(D)J joints align in the correct translational reading frame.

3. Repair and ligation

The later steps of V(D)J recombination have many aspects in common with general DNA double-stranded break (DSB) repair. They involve the processing and joining of both the two coding ends and the two signal ends by NHEJ, possibly within a complex that holds all four cleaved ends together. However, blunt signal ends are thought to persist much longer than the coding ends before they are joined.[78]

Many factors involved in DNA repair by NHEJ have been implicated in V(D)J recombination. Knockout studies have identified DNA-PK as a critical factor.[79] DNA-PK consists of at least three components, DNA-PK_{CS}, Ku70 and Ku80.[80] The Ku70 and Ku80 proteins form a heterodimer (Ku) that functions as the DNA-binding component, binding to broken DNA ends, ssDNA gaps or DNA hairpins. Ku may serve as a sensor of DSBs in the context of DSB repair mechanisms that employ DNA-PK. However, the precise role of Ku and DNA-PK in the V(D)J-joining reaction remains to be elucidated. DNA ligase IV catalyses the final step in NHEJ repair, serving to link together broken DNA strands. The DNA repair protein Xrcc4 complexes with DNA ligase IV to stabilise it *in vivo* and

increase its activity *in vitro* and may function to recruit DNA ligase IV to complete the end-joining reaction.[81] Cell lines defective in DNA ligase IV or Xrcc4 do not make either coding or signal joints in V(D)J recombination.[82] Biochemical experiments have suggested a second function of Ku in the ligation of blunt or nearly blunt DNA ends, involving the DNA ligase IV and Xrcc4 complex.[83] The MRN complex may also function in the coding end and/or signal end ligation.[84]

1.2.3.3. Regulation of V(D)J recombination.

V(D)J recombination is regulated in at least two ways; first by the expression pattern of RAG1 and RAG2 and second by the limited access of the recombination machinery to particular DNA sites. The *RAG* genes are normally only expressed in early lymphoid cells. The amount of RAG2 protein has been found to vary greatly through the cell cycle, being high in G_1 and then decreasing in the S, G_2 and M phases.[40] Since the level of RAG2 mRNA remains largely unchanged, RAG2 protein may be controlled by phosphorylation followed by degradation.[85] Since both RAG1 and RAG2 proteins are required for initiation of DNA cleavage, variations in RAG2 protein levels could explain why V(D)J recombination occurs chiefly in the G_o and G_1 phases of the cell cycle and is silenced in proliferating cells.[40] Limiting the V(D)J recombination process to non-dividing cells minimises the risk of inappropriate DNA breaks occurring during DNA replication in mitosis, which, if were to occur, may lead to chromosomal translocations (see section 1.4.1.1.).

Even in lymphoid cells expressing RAG1 and RAG2, only a small fraction of RSS sites are available for recombination at any one time due to constraints on both cell lineage and developmental stage. In normal circumstances, RSS sites are presumably inaccessible to the recombination machinery and need to be actively exposed before recombination can proceed. Recent evidence suggests that the RAG proteins themselves may play a role in controlling access to certain chromatin structures via interactions with certain remodelled forms of chromatin or with remodelling factors.[86] Notably, RAG2 has been shown to specifically interact via discrete domains outside the catalytic region with core histones.[87] Some studies report that the susceptibility of Ig gene segments to efficient rearrangement is correlated with their transcriptional activity, which in turn has been linked to the degree of cytosine methylation at CpG residues within a given locus.[88-90] Non-coding RNA transcripts generated from individual Ig gene segments have been detected prior to recombination.[90] Germline anti-sense transcription, both genic and intergenic, has been reported to occur at the V_H region prior to V_H to D-J_H recombination.[91] It has been

proposed that transcription through a Ig gene locus opens up the chromatin into a poised state to facilitate further chromatin remodelling and promote recombination. However, it remains unclear if these changes are the cause or the effect of increased locus accessibility.

A pivotal role for transcriptional cis-elements and cognate transcription factors in V(D)J recombination has been established. Enhancers are thought to control chromatin accessibility at proximal (D)J_H regions by activating associated germline promoters.[92] V(D)J rearrangement at Ig loci requires the combined activity of various transcription factors, including Pax5, E2A (E47 and E12 products) and early B-cell factor (EBF). Several transcription factors are sufficient in the presence of RAG1/2 to activate recombination in non-lymphoid cells. For example, the expression of either E12 or E47 leads to recombination at the κ loci.[93]

1.2.3.4. Immunoglobulin diversity

The combinatorial Ig diversity generated by the different permutations of germline V, (D) and J gene segments at the individual Ig loci (a) and the associations of heavy and light chains (b) can be calculated:

a) IgH: 51 V_H genes x 27 D genes x 6 J_H genes = 8262
   Igκ: 40 Vκ genes x 5 Jκ genes = 200
   Igλ: 30 Vλ genes x 4 Jλ genes= 120

b) 8262 x (200 + 120) = 2.64 x 10^6 potential combinations*

*This value is most likely higher than the actual amount of combinatorial diversity generated in an individual since it is likely that not all heavy and light chains will associate equally well due to structural constraints. Furthermore, not all gene segments are utilised at the same frequency during V(D)J recombination (see section 1.2.3.5.).

The D gene segments contribute significantly to Ig diversity. Mechanisms inherent to the rearrangement reaction ensure that the D genes can potentially be used in all three RFs depending on the V_H-D junctional sequence; one RF encodes a hydrophilic sequence, the second carries a hydrophobic aa sequence and the third has a propensity to encode a Stop codon, although this may be removed as a result of trimming and/or N addition.[42] The hydrophilic sequence promotes the formation of a loop structure and is used preferentially by the D2 and D3 families. Conversely, the hydrophobic sequence promotes a conformation that would interfere with the CDR3 structure and its ability to bind Ag. Therefore, Ig diversity may be influenced by the combined positive and negative effects of

the structural characteristics of the D gene RF-dependent aa sequences. Moreover, D to D fusions, which contradict the one-turn/two-turn rule of V(D)J joining but have been observed in ~5% of human Igs, and inverted D genes add to the somatic diversity at the IgH locus.[58] Diversity is further augmented at the D to J_H and V_H to D junctions by nuclease trimming of gene segments and the addition of P and N nucleotides; this also affects, but to a lesser extent, the diversity at V_L to J_L junctions. The junctional diversification mechanisms described above are thought to make an enormous contribution to the overall Ig diversity such that it exceeds 10^{12}.

The V_H-D-J_H and V_L-J_L joints, which correspond to the CDR3 of the V_H and V_L regions, respectively, are the most diverse regions of the Ig molecule and lie at the centre of the Ag binding site. It is estimated that less than 5 in 10^5 circulating B cells share the same CDR3.[94] As such, they represent the genetic fingerprint of an Ig molecule and therefore can identify an individual B cell or a clonal B-cell expansion. The CDR1 and CDR2, which also form the Ag binding site, have less sequence variability and are encoded in the V gene segments.

1.2.3.5. Selection of the pre-immune Ig repertoire

The pre-immune B-cell repertoire consists of naïve or virginal B cells that have not been exposed to external Ag. If gene segment selection during V(D)J recombination is an entirely random event, we would expect the pre-immune repertoire of naïve B cells to contain all germline V, (D) and J gene segments in equal frequency. However, the actual finding is a pronounced over-representation of certain gene segments and the concomitant under-representation of others.[48,76,95-98] Studies on BM-derived pre-B cells and immature B cells of healthy individuals have shown a biased usage of the V4-34, V4-59 and V3-23 gene segments and an under-representation of the V3-20 and V3-11 gene segments.[96] J_H gene segment usage is also biased; the J_H4 gene is vastly over-represented, which is likely due to the fact that J_H4 is the only J_H gene segment with a typical two-turn RSS.[94,97] The naïve repertoire has also been studied in B cells from fetal lymphoid tissue, cord blood and in pre-B cells of acute lymphoblastic leukaemia (ALL) and a consistent pattern of dominance of certain gene segments has emerged.[99-102]

Several levels of control exist which govern the usage of gene segments. In principal, repertoire bias that appears before the expression of surface Ig receptor must be due in part to intrinsic genetic mechanisms that lead to the preferential rearrangement and/or expression of certain gene segments. Preferential rearrangement as driven by (i) the

number of copies of individual V gene segments in the germline, (ii) the proximity of the two gene segments to be rearranged on the chromosome, (iii) accessibility of the DNA to the RAG1/2 complex and (iv) the presence of more efficient RSSs have all been described.[88,103-106] Similarly, preferential mRNA expression due to more efficient promoters or to the presence of gene-specific enhancer sequences has been reported.[107,108] At the protein level, sequence conservation within human V_H families also suggest a selection pressure to retain structural features of the heavy chain V region. The advantageous features of Ag binding sites encoded by the V_H genes (CDR1 and CDR2) may contribute to their over-representation in the pre-immune repertoire.[109] Additionally, the ability to bind to the surrogate light chains VpreB and λ5/14.1 at the immature B-cell stage may be a discriminating factor. Clonal selection of early B cells within the BM may also influence gene segment usage; the affinity of the surface heavy chain in the pre-BCR for self-Ags will determine if a pre-B cell is deleted by apoptosis or subject to receptor editing (high affinity) or allowed to further differentiate (low affinity).

1.2.4. Genetic maturation of the immune response

The initial generation of diversity is achieved by the successful rearrangement of the V, (D) and J gene segments to produce B cells that possess a unique sIg, as detailed above (section 1.2.3.).[55] Following exposure to Ag, further diversification is accomplished by means of somatic hypermutation (SHM) of the Ig V genes and class switch recombination (CSR) between Cμ and any of the downstream C regions (γ, ε and α). Both of these processes take place in the GC and are intricately linked to BCR cross-linking, T cell help and the cytokine milieu.[110-112] Although both SHM and CSR occur in B cells at the same stage of differentiation, they are distinct and apparently independent processes, since IgM antibodies possessing somatic mutations and IgG and IgA antibodies with unmutated V regions have been described.[111,113-115] A third B-cell maturation pathway termed Ig gene conversion, which is known to play a major role in species such as chickens, is not extensively used in humans and will not be discussed further.

1.2.4.1. Somatic hypermutation

SHM is a process whereby point mutations are introduced into the V_H and V_L region genes to increase the affinity of Ab molecules and to further diversify the B-cell repertoire. In mice and humans, SHM occurs at a rate of 10^{-5} to 10^{-3} mutations per base pair per generation, which is 10^6-fold higher than the spontaneous rate of mutation in housekeeping genes.[116] Mutations are predominately single nucleotide substitutions, although they do show a slightly increased tendency to occur in clusters. However, deletions, insertions and

duplications have been reported to occur at a very low frequency in both murine memory B cells and human tonsillar GC B cells.[117] The mutations introduced by SHM arise about 200bp downstream from the V region transcriptional start site and extend approximately 1.5-2kb further downstream of the promoter, although sequences upstream of the transcriptional start site do mutate at a lower frequency. This region spans the leader intron, the rearranged V, (D) and J exons and the intron upstream of the constant region. Mutations rarely extend into the C_H and C_L domains, although some very low-grade mutation can be detected in the first domain of the mouse lambda chain (Cλ1).[118,119] However, this is thought to be related to the short length of the JC intron.

Transition mutations (purine (Pu) to Pu and pyrimidine (Py) to Py) arise more frequently than transversion mutations (Pu to Py and Py to Pu) even though twice as many transversions than transitions are possible, suggesting that mutations are not random. Indeed, although mutations are observed along the entire length of the V genes and their immediate flanking sequences, there are a number of sequence motifs that are preferentially targeted ('hotspots') or rarely targeted ('coldspots').[120,121] Examples of recognised hotspots include mutations of G and C base pairs that are in RGYW (R=A or G; Y=C or T; W=A or T) and the complementary WRCY motifs.[122] It appears that the codon usage of Ig V genes has been selected during the course of evolution so that the most mutable residues are in positions where alterations will be most likely to lead to improved affinity, i.e. in the Ag-binding site (CDRs). Conversely, coldspots tend to be located in the FRs and serve to maintain the structural integrity of the antibody. For example, serine is unusual in that it is encoded by two types of triplet; AGY (Y=C or T) and TCN (N=A or C or G or T). The sequence AGY conforms to the hotspot consensus, whilst the TCN does not. Serine codons in the CDRs most commonly use the AGY triplet and therefore are more frequently targeted for mutation than those in the FRs, which commonly use the TCN triplet.[123] However, it appears that in any particular V region, some potential hotspots are targeted whereas others are not, even in the absence of selection for higher affinity Abs. This suggests that the neighbouring sequences or higher order DNA or chromatin structures might also influence the targeting of mutations.[124,125]

The initiation of SHM just downstream of the Ig promoter led to the suggestion that transcription has a central role in the mutational process. This has been supported by the finding that transcription is an absolute requirement for SHM and the frequency of SHM of Ig V genes directly correlates with the rate of transcription.[126-128] Furthermore, SHM of the targeted V region appears to be under the direction of transcription-related elements; in

support of this, the Ig promoter and enhancer have been identified as critical components in promoting SHM.[127,129,130] If the Ig V region promoter is removed by gene targeting, SHM of the locus is vastly diminished.[127] However, the promoter can be replaced by other RNA pol II-dependent promoters and SHM is not unduly compromised as long as the heterologous promoter is transcriptionally active.[130] However, the strength of the promoter is critical.[127,129] The Eκ enhancers are individually necessary, but together not sufficient to target SHM.[130,131] The duplication of the Ig V gene promoter upstream and the movement of the intronic Eκ enhancer downstream of the C region in a Igκ transgene targets SHM to this region.[126] Similarly, replacement of the Ig V gene with artificial substrates, such as bacterial sequences or the *human β-globulin* gene, leads to SHM of those substrates, suggesting that the sequence of the Ig V gene itself does not play a role in targeting the mutation process.[132,133] Activation-induced cytidine deaminase (AID) has been shown to have an essential role in the initiation of SHM and is discussed in section 1.2.4.3..

1.2.4.2. Class switch recombination

Post V(D)J recombination, the Ig C_H locus is organised for the surface expression of IgM and/or IgD, in which the rearranged V_H region lies upstream of the Cμ and Cδ regions. CSR results in the rearrangement of the V_H region to the downstream Cγ, Cε and Cα regions and the expression of that V_H region as an IgG, IgE or IgA isotype.[134] Each C region class mediates different combinations of effector functions, tissue localisation and persistence via the interactions of the Ig with specific cellular or soluble factors (see section 1.2.1.1.). CSR therefore permits a change of effector function of an Ig whilst maintaining Ag specificity.

The Ig C_H gene locus, positioned 3' of the J_H gene segments, consists of an ordered array of C_H genes spanning 250kb (Figure 1.4.; see section 1.2.2.1.). Each C_H gene is divided into several exons corresponding to individual Ig domains in the folded C region. CSR involves DNA regions called switch (S) regions that are located in the introns 2-3kb upstream of each C_H region, with the exception of the Cδ gene (see below). S regions are 1-12kb in length and are composed of highly repetitive palindromic sequences, which are G rich on the non-template strand. The Cμ S region (Sμ) consists of ~150 repeats of the sequence [$(GATGCT)_n(GGGGGT)$], where n is frequently 3, but can be as many as 7. Also interspersed are $(T;CAGGTTG)_n$ motifs. Although the sequences of the other S regions (Sγ, Sε and Sα) are also G rich and contain repeats of the GAGCT and GGGGGT sequences, they are not homologous to each other. Deletion of most of the Sμ region has been shown to severely impair CSR, whereas deletion of the entire Sγ1 region inhibits

switching to IgG1, indicating that the S regions are specialised targets for the CSR process.[135,136] Moreover, junctional heterogeneity in sequence and site at which both donor and recipient recombination endpoints are located demonstrate that CSR, although region-specific, is not sequence-specific.[137] However, it has been demonstrated that recombinations may occur preferentially near certain sequence motifs within the tandem repeats.[138]

The vast majority of class switching is achieved by an intrachromosomal recombination event between the donor Sμ region and one of the recipient downstream S regions (Sγ, Sε or Sα) and includes looping out and deletion of all C_H genes 5' to the C_H gene that is to be expressed (Figure 1.12.). As well as direct switching from IgM to any of the downstream C_H genes, sequential switching between downstream C_H genes can occur.[139] An exception is IgD, which is co-expressed with IgM in mature B cells by alternative splicing of primary RNA transcripts. Similarly, a deletional recombination event cannot account for B cells that express both IgM and a downstream isotypes. The likely mechanism for multiple isotype expression involves RNA splicing, which may originate either from alternatively spliced long RNA transcripts or by trans-splicing of two RNA transcripts.[140-142]

The process of CSR can be divided into a number of stages:

1. Transcription of S regions

The simultaneous transcription of both donor and recipient S regions is an absolute requirement for CSR.[143,144] An increase in transcriptional activity through the S regions is thought to confer a state of enhanced accessibility of the C_H locus for the binding of additional factors important for mediating the switch event (the 'Accessibility Model' of Ig CSR).[145] Transcription may alter the S region chromatin structure to produce a more open conformation, which would allow access to putative CSR recombinases.[145,146] Alternatively, it may generate DNA structures that are able to act as substrates for CSR.[147,148] For example, S regions are rich in palindromic repeats, which, if rendered ss, could form transient stem loop structures.[149]

A specific intronic promoter (I) located immediately upstream of each S region is involved in initiating transcription, which begins 5' to the S region, proceeds through the C_H gene and terminates at the normal poly(A) sites for secreted or membrane bound Ig heavy chain mRNAs. In mature B cells, Sμ is constitutively transcribed from its specific intronic (I) promoter. The resulting transcripts (Iμ-Cμ, Iγ-Cγ, Iε-Cε and Iα-Cα) are known as sterile

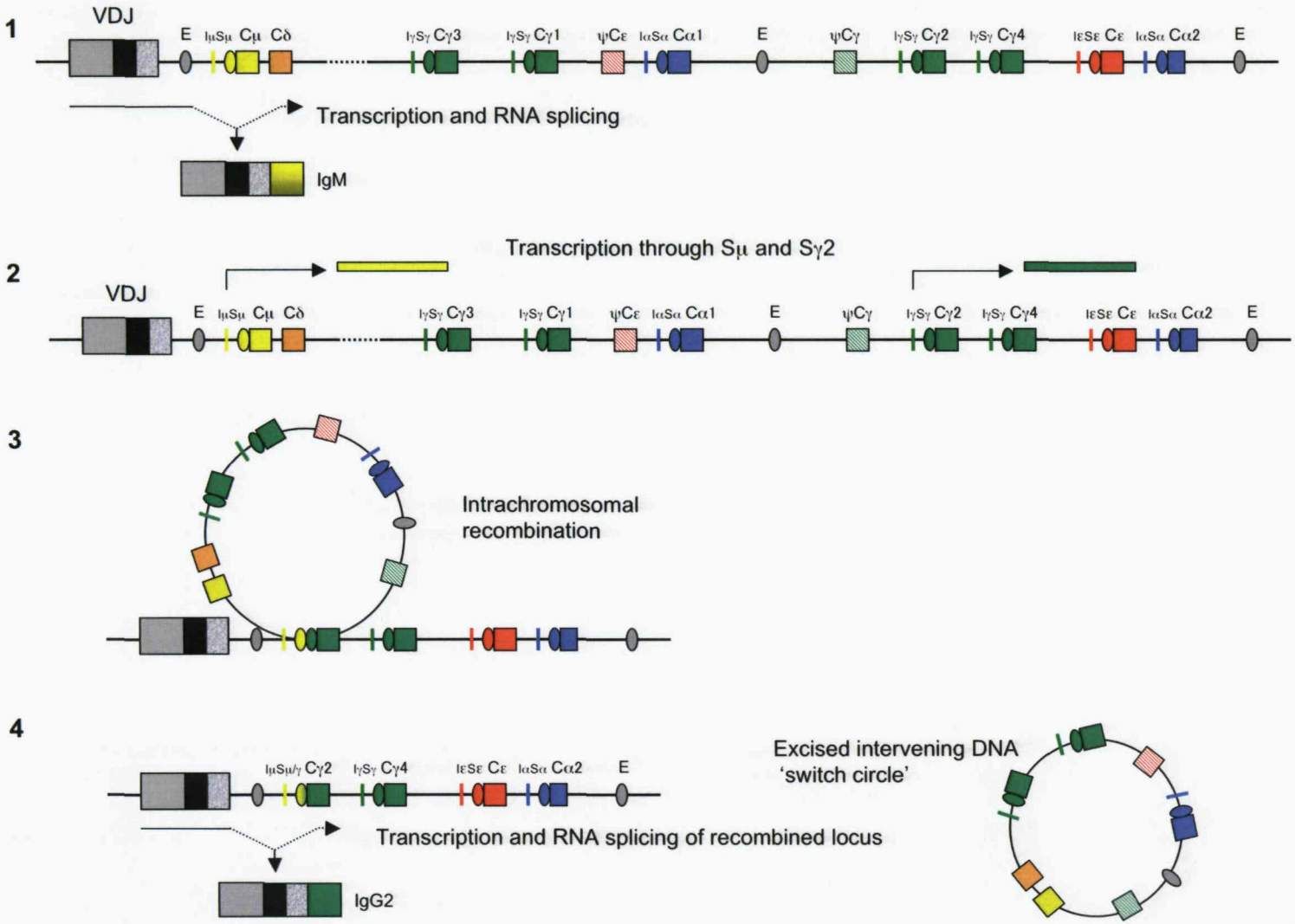


Figure 1.12. Mechanisms of class switch recombination. Schematic diagram to illustrate class switch recombination (CSR), a process whereby the constant region of the heavy chain (C_H) is replaced with any of the downstream constant regions (CSR to IgG2 is shown). Post V(D)J recombination the C_H locus is organised for the production of IgM antibodies (1). Cytokines stimulate transcription from the intronic promoter through the switch region of the appropriate downstream C_H (Iγ2-Iγ2) (2). Sμ is constitutively transcribed from its I promoter (Iμ). Intrachromosomal recombination between donor Sμ and recipient S (Sγ2) regions proceeds through DNA double-stranded break (DSB) intermediates followed by DNA repair and results in the permanent deletion of the intervening DNA sequence, termed a switch circle (3). The newly recombined C_H locus is transcribed (Iμ-Cγ2) and processed to yield IgG2 antibodies (4). The DNA switch circle is lost during subsequent cell division. Adapted from Goldsby *et al.*, 2000.[1]

germline transcripts due to the presence of multiple stop codons within the mRNA that prevent their translation.[150] Replacement of the promoter Iε or Iγ1 by heterologous strong promoters revealed that transcriptional activity through a specific locus was insufficient for CSR.[151,152] Therefore, transcription *per se* is not sufficient to direct switching to the S region. A requirement for processed (spliced) germline transcripts in CSR, where the intronic S region mRNA is removed, was demonstrated by targeted deletion of the splicing donor site in the Iγ1 exon.[152] The processed sterile germline transcript is thought to form a stable RNA-DNA hybrid with the template strand, known as an R-loop, and results in the exposure of a ssDNA on the nontemplate strand that may be as long as 1kb.[147,153]

During an immune response, different Ig isotypes are produced at different levels depending on the type of Ag and the site of infection. Cytokines secreted by T_H and other cells play a key role in the regulation of isotype specificity of switching. For example, IL-4 induces switching to IgG_4 and IgE, transforming growth factor-β (TGF-β) to IgA and IgG_{2b} and IFN-γ to IgG_{2a} and IgG_3.[154-157] The ability of any given cytokine to determine the specificity of CSR is correlated with its ability to regulate the transcription of unrearranged or germline Ig C_H genes, which is achieved via a signalling cascade that culminates in the binding of transcription factors to elements in S region promoters.[158] Several cytokine-responsive proteins have been shown to bind within or near S regions, all of which are known to serve as transcriptional activators or repressors at other binding sites in promoters or enhancers. The IgH locus contains multiple binding sites for nuclear factor κB (NF-κB) in the Sγ1, Sγ3, Sε and Sα regions, as well as in the promoters Iγ1, Iγ3 and Iε and is thought to be involved in the regulation of Cγ, Cε and Cα transcripts.[159] Lipopolysaccharide-responsive factor 1 binds to sequences within the Sγ1, Sγ3 and Sα regions where it is important for the transcription of Cγ1, Cγ3 and Cα.[160] Pax-5 has binding sites within Sμ, Sγ1, Sγ2a, Sε and Sα regions and is essential for the transcription of Cε.[161] All three proteins have also been shown to bind the IgH 3' enhancer, which is located 16kb 3' of the Cα gene, and is important in the regulation CSR.[160,162,163] It is important to note that cytokines are unable to induce CSR alone and require appropriate co-stimulation by way of CD40 signalling or sIg cross-linking.[164,165]

2. Cleavage of S regions

The generation of switch circles in CSR strongly suggests that CSR proceeds through dsDNA break intermediates and, therefore, the initiating event in switch recombination must result in the introduction of two DSBs occurring in a concerted fashion. Indeed,

DSBs have been detected in CSR.[166] In V(D)J recombination, the cleavage reaction is carried out by the RAG proteins (see section 1.2.3.2.). However, neither RAG1 nor RAG2 expression are required for efficient CSR.[167] CSR requires cell proliferation; inhibitors of DNA synthesis block CSR and suggest a role for DNA replication, which might function to convert S region ssDNA nicks into DSBs.[168,169] It has been shown that during CSR, the Sμ donor and recipient S region undergo point mutation.[170-173] These S region mutations have the same characteristics as mutations that occur in the Ig V genes during SHM in that they arise at a high rate, target hotspots and extend 1.5-2kb further downstream of the I promoter, but spare the C region. As with SHM, AID has been implicated in the initiation of CSR DSBs and is discussed in section 1.2.4.3..

The DSBs found in S regions during CSR are staggered.[174] Such lesions are thought to be either filled through short-patch DNA synthesis using error-prone DNA polymerases, such as DNA pol η, or resected by nuclease-mediated recession to generate blunt DSBs. The finding of deletions and mutations at the S region junctions would support this view.[137]

3. Repair and ligation

Completion of CSR involves the joining of the two cleaved S regions by NHEJ.[175-177] DNA-PK is a critical factor for DSB repair in V(D)J recombination and consists of at least three components, DNA-PK_{CS}, Ku70 and Ku80.[80] Several studies have demonstrated using SCID mice and mice deficient in Ku70 or Ku80 that all components of DNA-PK are required for successful CSR.[175-177] However, DNA-PK_{CS}-independent CSR activity of the Sγ1 region has been observed.[178]

Efficient recombination likely requires close juxtaposition, or synapse, between the two recombining S regions. Several DNA repair proteins have been suggested to participate in the S region synapse. The phosphorylated histone HA2X (γ-H2AX) has been shown to localise at S region DSBs during CSR.[179] It has been proposed that γ-H2AX functions as an anchor to recruit and assemble lattices of DNA repair factors at sites of DSBs and thereby hold broken chromosomal ends in close proximity to allow proper NHEJ.[180] Other DNA repair proteins include an H2AX associated DSB response factor, p53 binding protein 1, Nijmegen breakage syndrome protein 1 and ataxia telangiectasia mutated protein.[181,182]

During NHEJ, the two S regions combine and the intervening DNA, called a 'switch circle', is permanently deleted from the chromosome. The specific I promoter is still

active in DNA switch circles and directs the production of I-Cμ transcripts termed 'circle transcripts'. In the absence of switch signals, circle transcripts are rapidly degraded. Thus, circle transcripts, detectable by specific RT-PCR, may serve as a hallmark for active CSR *in vitro* and *in vivo*. [183] DNA switch circles are typically lost during cell division.

1.2.4.3. Activation-induced cytidine deaminase

In the course of seeking new factors that might be involved in CSR, Muramatsu *et al.* discovered *AID* as a differentially expressed gene in a murine B lymphoma cell line induced to undergo CSR.[184] They demonstrated that AID was selectively expressed in GC B cells and when the *AID* gene was inactivated in mice, these mice had very large GCs in their secondary lymphoid organs and were unable to perform SHM or CSR, as evidenced by the exclusive production of low-affinity unmutated IgM antibodies. Revy *et al.* later showed that a subset of patients with hyper-IgM syndrome 2 (HIGM2), which displayed an identical phenotype to that of the $AID^{-/-}$ mice, had biallelic autosomal recessive mutations in the *AID* gene.[185] AID deficiency does not appear to affect the maturation of B cells since B cells display a mature phenotype, are activated through the BCR and form GCs. Instead, AID appears to be an absolute requirement for a B-cell terminal differentiation mechanism shared by SHM and CSR. The third B-cell maturation pathway, Ig gene conversion, has also been shown to be dependent on AID activity.[186] Therefore, the *AID* gene appears to control all B-cell-specific modifications of vertebrate Ig genes.

The *AID* gene maps to the human chromosome 12p13 and is composed of 5 exons that encode a 198 aa protein.[187] Based on the sequence homology of the active site, AID has been classified as a member of the cytidine deaminase family. Like other members of the this family, AID can catalyse the deamination of cytosine to uracil *in vitro*.[184] However, its substrate *in vivo* and exact mechanisms of action and regulation remain controversial. Apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1 (APOBEC-1), a known RNA editing enzyme, is the closest homologue to AID in the cytidine deaminase family.[188] It also maps to chromosome 12p13, suggesting a common AID/APOBEC-1 ancestor and a conserved function of these proteins. An APOBEC-1 homodimer is the catalytic subunit of a multi-protein complex that deaminates cytidine 6666 in the apolipoprotein B (apoB) mRNA to a uracil, resulting in a stop codon so that the edited mRNA now encodes a truncated apoB protein with a new function.[189] The full length apoB protein (apoB 100) is involved in transporting cholesterol in the blood, whereas the truncated apoB protein (apoB 48) is a triglyceride carrier. APOBEC-1 is unable to perform a site-specific modification *in vitro*, requiring instead the APOBEC-1 complementation factor (ACF),

which forms part of the editing complex by binding to the target RNA, thus docking APOBEC-1 within close proximity to its substrate.[190] It is probable that AID also requires ACF-like factors for target specificity.

SHM and CSR are independent processes and neither is a prerequisite for the other.[125] It has been shown that separate domains of the AID protein are responsible for SHM and CSR.[191] The N-terminal end contains a bipartite nuclear localization signal and a region with a zinc coordination motif and contains the active catalytic site of the enzyme.[125,191] The C-terminal end of the protein contains a nuclear export signal that has been shown to associate with a nuclear export protein that is involved in shuttling AID from the nucleus to the cytoplasm.[192] The role of each domain has been determined by expressing mutant *AID* genes containing representative mutations in reporter assays for SHM and CSR. Sequences in the C-terminal region of the *AID* gene are essential for CSR, whilst sequences in the N-terminal are critical to SHM function.[192,193] Consistent with *in vitro* studies, a new form of hyper-IgM syndrome, HIGM4, was reported in which patient B cells undergo normal SHM but cannot perform CSR.[194] The different roles of these regions support the hypothesis that there must be factors that associate with AID and separate its function in SHM and CSR. The finding that SHM and CSR can be activated following ectopic expression of AID in non-centroblastic B cells, cells of other lineages and bacteria suggests that other factors involved in SHM and CSR are ubiquitous and highly conserved.[195-198] Therefore, AID is the only B-cell specific factor required for SHM and CSR.

1.2.4.3.1. The role of AID in SHM and CSR

Similarities between AID, APOBEC-1 and the family of RNA editing enzymes led to the suggestion that AID is acting indirectly, editing the mRNA for an unknown endonuclease and activating the protein so that it could act directly on the V and S region DNA to initiate SHM and CSR.[125] Indeed, some of the mutations present in patients with HIGM2 are in the deaminase-active site, suggesting AID is functioning as a cytidine deaminase. However, increasing evidence supports the view that AID acts directly on DNA. The over-expression of AID and APOBEC-1 in *E. coli* has shown that both enzymes can directly deaminate deoxycytidine (dC) in DNA.[199,200] Biochemical studies have shown that AID can catalyse the deamination of dC on ssDNA, but not dsDNA, RNA or DNA-RNA hybrids[201,202] However, dsDNA, such as is targeted *in vivo*, can be deaminated during transcription due to the generation of a ssDNA transcription bubble created by the interaction of DNA with RNA pol II.[203,204] The finding that ssDNA is the substrate for AID

provides reasonable explanation for the requirement of high rates of transcription for SHM and CSR.[205,206] S regions may represent special targets owing to their propensity to form stable R-loops that result from the sterile RNA transcript binding to its DNA template strand, which could collapse to provide AID targets on both DNA strands.[147,207] However, transcribed V region genes do not substantially form R-loops, which suggests AID must access them in some other way. It has been demonstrated that AID from activated B cells interacts with Replication Protein A (RPA), a ssDNA-binding protein involved in replication and repair. RPA has been hypothesised to function to expand and stabilise ssDNA loops in the context of transcription bubbles to facilitate AID access during SHM and CSR.[208] In cell-free assays, AID preferentially deaminates dC in ds-transcribed DNA on the non-transcribed strand.[203,208] However, the comparison of mutation patterns in mutated Ig genes demonstrates that the mutation bias of triplet hotspots was the same as that of their complement suggesting that both strands are equally subject to SHM.[209] The reasons for the discrepancy in the magnitude of the strand bias in AID action *in vitro* and *in vivo* are unclear. It has been suggested that *in vivo*, AID might deaminate cytosines in the negatively supercoiled DNA as it arises in the wake of transcribing RNA pol, which would expose both DNA strands to AID and hence circumvent strand bias.[210,211]

The mechanisms that target AID to the V regions and S regions remain largely unknown, but its restricted targeting results at least in part from the availability of large amounts of ssDNA in the form of transcription bubbles in the V regions and S regions. High levels of transcription alone cannot explain why AID is targeted to the V and S region of the Ig genes since some other highly transcribed genes in B cells are not targeted for SHM.[212] AID might be recruited to Ig genes in the context of transcription by way of transacting proteins that bind specific cis-elements, including Ig enhancers.[213] The E2A protein binding sequence CAGGTG, which is present in all Ig enhancers, is potentially one such element. It has also been suggested that selective increases in chromatin accessibility could be responsibly for the restriction of AID-induced mutation. Recent studies show that increases in the acetylation of histone H4 are associated with the targeting of SHM to the V region, whilst increases in the acetylation of histone H3 are associated with targeted downstream S regions.[206,214] Although there is selective targeting of AID to the Ig V region and S regions, some proto-oncogenes, such as *Bcl-6*, *c-Myc*, *Pax5* and *Fas*, also undergo SHM-like mutations in normal B cells, but at lower rates than in the V regions.[215] As with the V regions, mutations arise 200bp downstream of the promoter and extend ~2Kb further downstream.

All known factors implicated in SHM and CSR are part of the general DNA metabolism machinery.[216] Spontaneous U:G mismatches are normally repaired error free by uracil glycosylase (UNG)-mediated base excision repair (BER) and mismatch repair (MMR) involving the MutS homologue 2-MutS homologue 6 (MSH2-MSH6) heterodimer. Both UNG-mediated BER and MMR are also implicated in SHM and CSR. However, the deoxyuracil (dU) that results from AID activity during SHM or CSR does not appear to be efficiently converted back to cytosine, unlike during normal repair. The dU:dG lesion generated by AID can be repaired by a number of pathways (Figure 1.13.).[217] Firstly, it can be replicated directly to produce transition mutations (dC > dT and dG > dA). Second, it can be repaired by means of (UNG)-mediated BER. UNG is proposed to facilitate the mutation from dC and dG by removal of the dU to create an abasic site. The abasic site can subsequently be converted to a single stranded nick by apurinic-apyrimidic endonuclease (APE), which can either be repaired with error-free replication without generating mutations, or it can be bypassed by error-prone DNA polymerases, such as pol μ, pol η, pol ζ and pol ι, to generate all possible mutations. Alternatively, replication over the abasic site could yield both transition and transversion mutations from C/G. Third, it can be recognized by the MSH2-MSH6 heterodimer and the complex of other MMR proteins. The sequence surrounding the mismatch is excised and replaced by means of resynthesis with error-prone DNA polymerases, which creates additional mutations of the sequence. Finally, it can be repaired by means of homologous recombination (HR) if the mutations are present in late S and G2 when a sister chromatid is available as a template.[198]

In SHM, AID deamination activity results in the introduction of non-templated point mutations in hotspots, specifically at the RGYW and WRCY DNA motifs located within the hypervariable regions of the Ig V gene. In CSR, the role of AID is to introduce dC to dU mutations in the S regions, much as it does in the V regions during the generation of point mutations. It has been suggested that staggered DSBs in the S regions required for CSR are the result of closely spaced single-stranded nicks occurring on opposite DNA strands introduced by UNG and APE, MMR or both.[174,218] The DSBs in the donor Sμ and recipient γ ε or α S regions can be resolved by NHEJ, involving DNA-PK, or other DNA repair mechanisms.[216] The inactivation of UNG, by expression of a specific inhibitor or by the disruption of the gene, does not affect the frequency of SHM, but leads to skewing towards transition mutations at dC (dC > dT) and dG (dG > dA) residues.[217,219] In the absence of UNG, mismatched dU residues may be resolved by the replication pathway, leading to the generation of transition mutations at dC/dG residues. UNG-deficiency also

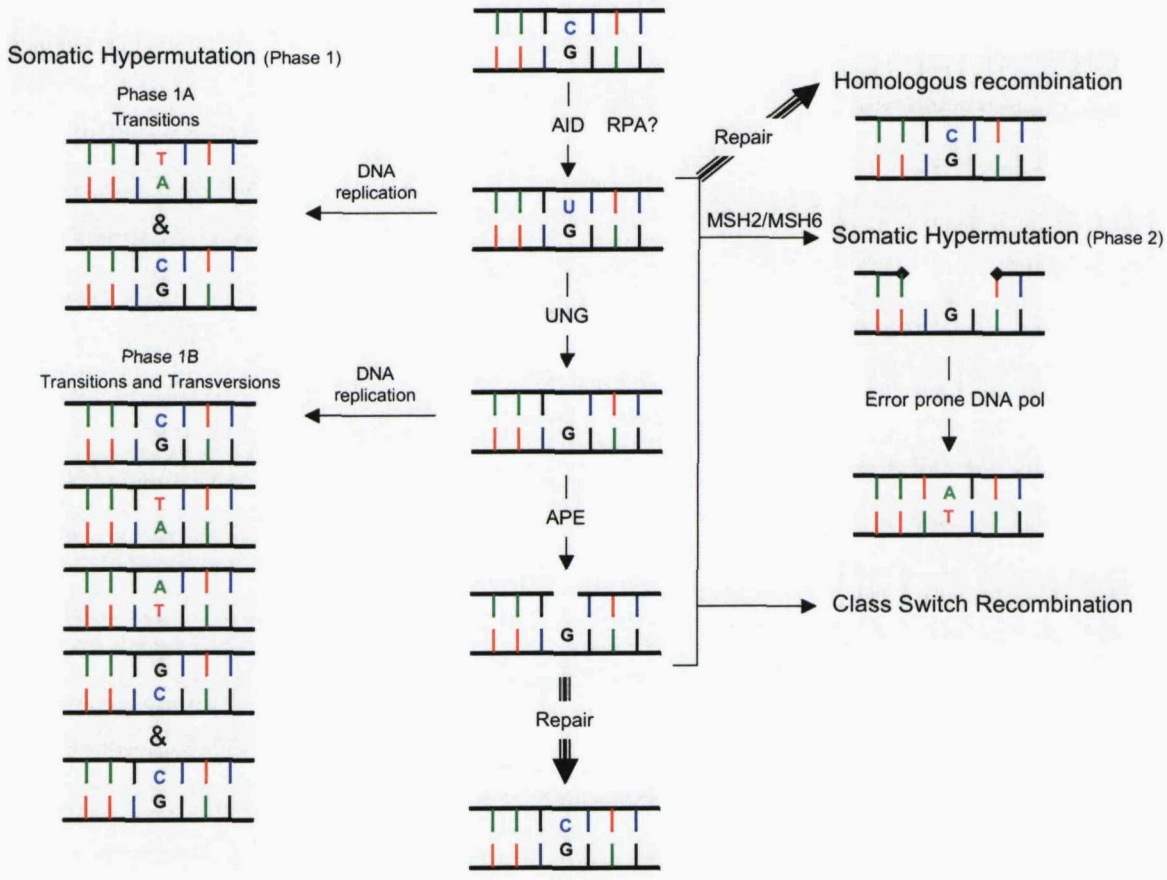


Figure 1.13. Resolution of the U:G mismatch created by AID on DNA. AID deaminates deoxycytidine (dC) to deoxyuracil (dU) to create U:G mismatches. The mismatch can be repaired by one or more of the pathways illustrated above. First, it can be replicated to produce a C to T mutation. Second, it can be repaired by means of base excision repair; the deoxyuracil can be removed by means of uracil N-glycosylase (UNG) to create an abasic site. The abasic site can be converted to a single-stranded nick by apurinic-apyrimidinic endonuclease (APE), which can either be repaired with error-free replication or bypassed by error-prone polymerases to generate all possible mutations. Third, it can be recognized by MMR proteins; the sequence surrounding the mismatch is excised and replaced with error-prone polymerases, which creates additional mutations of the sequence. Finally, it can be repaired by means of homologous recombination if the mutations are present in late S and G2 when a sister chromatid is available as a template. Adapted from Rada *et al.*, 2004.[217]

leads to the impairment of CSR, presumably due to a lack of dU residue deglycosylation and removal, which results in a lack of APE and DNA breaks. The normal frequency and pattern of SHM at dA and dT residues and residual CSR activity have been related to the MMR enzyme system, especially MSH2-MSH6.

1.2.4.3.2. Regulation of AID activity

The expression of AID has been shown to be restricted to centroblast B cells in the GC of secondary lymphoid organs, where SHM and CSR are active.[220] However, the regulation of AID expression and activity *in vivo* is not completely understood. When AID was tagged at its N-terminus with the green fluorescent protein it was demonstrated that a high proportion of AID protein was localized in the cytoplasm of B cells.[221] However, AID can accumulate in the nucleus if nuclear export is inhibited, demonstrating that AID shuttles between the cytoplasm and the nucleus.[191] It is possible that AID is sequestered in the cytoplasm until the stimulation of B cells for maturation actively translocates it to the nucleus and that this may be facilitated by the association of AID with other molecules, such as specific chaperones, or by posttranslational modifications, such as phosphorylation.[222] Indeed, it has since been demonstrated that the activity of AID is regulated by phosphorylation by protein kinase A (PKA).[223,224] Basu *et al.* reported that AID purified from activated mouse B cells is phosphorylated at serine 38 (S38) and tyrosine 184 (Y184).[223] Although little is known about Y184 phosphorylation, the S38 residue lies within a consensus motif for the cAMP-dependent PKA. Pasqualucci *et al.* reported that AID is associated with PKA *in vivo* and that the activation of PKA results in the increased phosphorylation of AID.[224] Further studies showed that while both unphosphorylated and phosphorylated AID are capable of deaminating ssDNA, only phosphorylated AID can deaminate dsDNA in a transcription-dependent fashion and that this relies on interaction with RPA.[208] Overall, these studies suggest that PKA is an AID kinase and that phosphorylation might be important in the regulation of AID activity.

1.3. B-cell development

1.3.1. Primary lymphatic organs

B-cell development begins in the placenta and embryonic blood, continues in the fetal liver and, after birth, in the fetal BM, where it continues throughout adult life.[225] In the BM, mature B cells are generated in an Ag-independent manner from haematopoietic stem cells (HSC), a process which is estimated to take 2-3 days in humans (Figure 1.14.). HSCs do not express cell surface lineage markers (Lin^-) and, therefore, they can give rise to all blood cell lineages. Several precursor cell subsets feature between HSCs and early B-cell

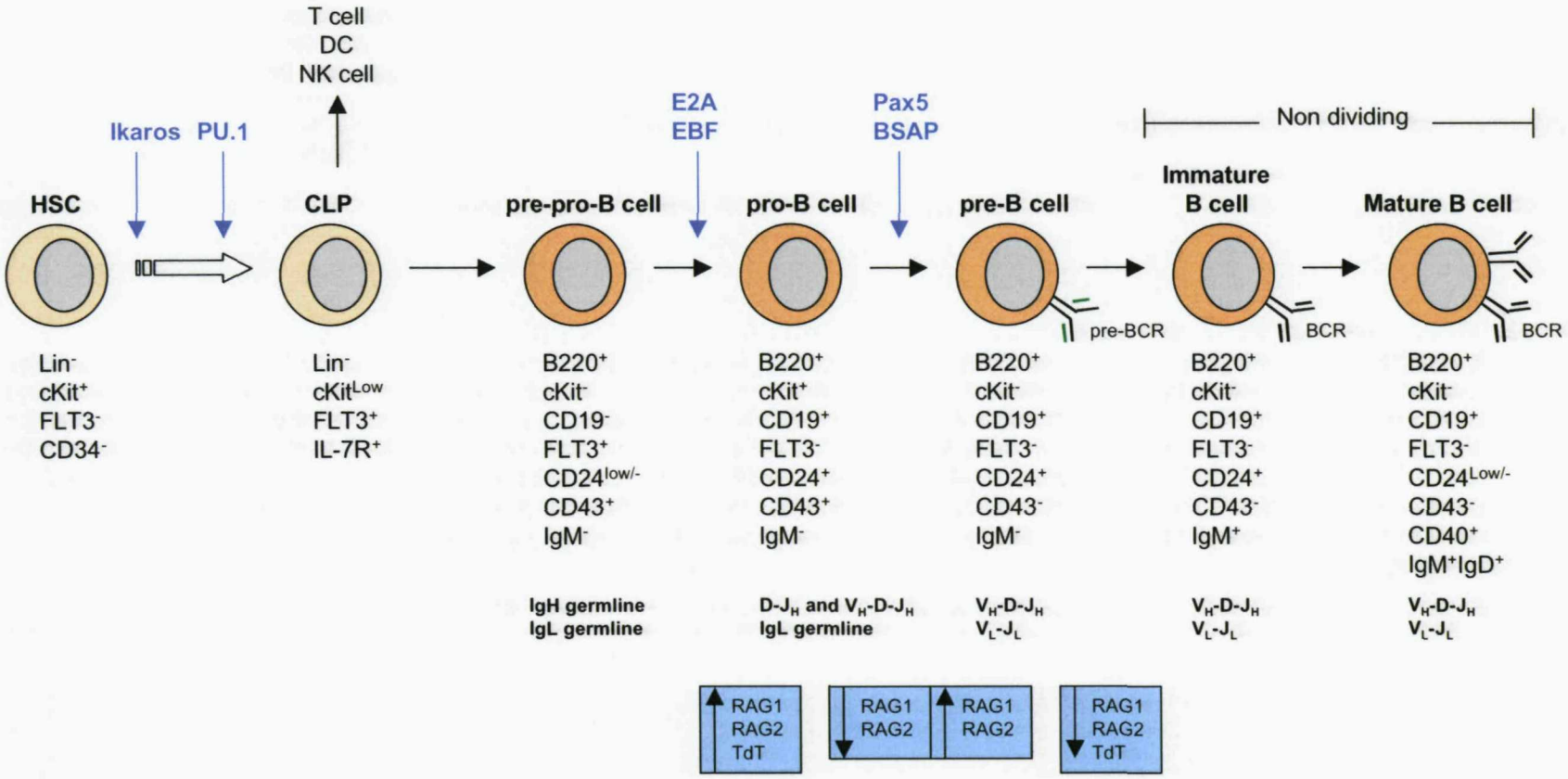


Figure 1.14. A pathway for early B-cell differentiation in the bone marrow. The progressive stages of B cell lymphopoiesis from haematopoietic stem cell (HSC), through common lymphoid progenitor (CLP) B cell precursor to mature B cell are shown. The critical points for the expression of particular transcription factors are shown in blue. The differential expression of a range of cell surface markers (listed below) determined the B cell precursor subset: pro-pre-B cell, pro-B cell, pre-B cell, immature B cell and mature B cell. The stage of V(D)J recombination that correspond to each B cell precursor subset is indicated. Changes in RAG1, RAG2 and TdT protein expression are indicated in blue boxes. Adapted from Nagasawa, 2006.[226]

precursors: these include haematopoietic multipotential progenitors, lymphoid-primed multipotential progenitors and common lymphoid progenitor (CLP) B-cell precursors.[226] CLPs lack erythroid, myeloid and megakaryocytic potential, but can give rise to B cells, T cells, NK cells and dendritic cells (DCs). B-cell precursor cells are negative for surface Ig (sIg^-) and positive for the B-cell lineage marker B220 (CD45). According to their differential expression of a range of cell surface markers they can be divided into four subsets: (1) pre-pro-B cells, (2/3) pro-B cells and (4) pre-B cells. The earliest identifiable B-cell precursors are the pre-pro-B cells, which derive from CLPs. However, recent studies have shown that this cell fraction is heterogeneous and can give rise to T cells, NK cells and DCs, as well as B cells.[227,228]

The definitive marker for commitment to the B-cell lineage comes with the initiation of Ig gene rearrangement, which is activated in early pro-B cells by the expression of RAG1/2 and TdT (see section 1.2.3.). This is followed by the expression of CD19, an early B-cell surface marker, which is maintained during subsequent B-cell differentiation until terminal differentiation results in plasma cells. Pro-B cells require direct stromal cell contact within the BM, which is mediated by the cell adhesion molecule very late Ag 4 (VLA-4, also known as integrin $\alpha4\beta1$) on the pro-B cell and its ligand vascular cell adhesion molecule 1 (VCAM-1).[1,226] Following initial contact, the pro-B-cell receptor c-Kit (CD117), a tyrosine kinase, interacts with stem-cell factor (SCF) on stromal cells leading to differentiation into pre-B cells. Pre-B cells then express the IL-7 receptor (IL-7R) and the distinctive pan B cell markers CD20, CD21, CD22 and CD72; CD10 and CD38 are transiently expressed. The binding of IL-7 to its receptor drives the proliferation of pre-B cells and, in turn, the down-regulation of c-Kit, which leads to stromal cell detachment.[226] The RAG proteins and TdT are down-regulated during this period of proliferation and Ig rearrangement is temporarily arrested. Proliferation is halted after 2-5 divisions and the RAG proteins are re-expressed, which results in the reactivation of Ig gene rearrangement. Completion of Ig rearrangement ends the pre-B cell stage. B cells are now called immature B cells and express surface IgM as part of the B-cell receptor (BCR). Following the expression of membrane-bound IgM and IgD, immature B cells develop into mature B cells, which migrate to the secondary lymphoid organs where they encounter Ag. Recently, it has been demonstrated that immature B cells can leave the BM and enter the circulation as transitional B cells, which represent an intermediate stage of development (see section 1.3.2.).

The development of B-cell precursors through these various stages requires the coordinated expression of specific transcription factors, which include Ikaros, PU.1, E2A

(E47 and E12 products), EBF and Pax 5, one isoform of which is the B-cell specific activator protein (BSAP) (Figure 1.14.).[reviewed in 39,229] For example, E2A and EBF are involved in the regulation of RAG1 and RAG2 expression and gene segment accessibility, while EBF leads to the expression of many components of the pre-BCR and BCR. In addition, several microenvironmental components that act on B-cell precursors have been identified.[reviewed in 226] IL-7 and SCF, mentioned above, are essential for the generation of pro-B and pre-B cells. Other factors include CXC-chemokine ligand 12 (CXCL12), required at the pro-pre-B cell and pro-B cell stage, FMS-like tyrosine kinase 3 ligand, required at the CPL and pro-pre-B cell stage and receptor activator of nuclear factor-κB ligand, involved in the generation of pre-B cells and immature B cells. The interaction of extracellular matrix proteins (fibronectin and laminin) with integrin receptors on the B-cell precursor is also likely to be important for supporting the growth of early B-cell progenitors.

1.3.2. Transitional B cells

Immature B cells exported from the BM enter a transitional phase during which further maturation events occur to produce mature B cells. Designated 'transitional B cells', they represent an intermediate stage of development and mark the crucial link between BM immature B cells and peripheral mature B cells. Transitional B cells were first described in mice and have now been extensively studied.[reviewed in 230] More recently, several studies have attempted to characterise transitional B cells in humans and have found that they can be readily identified as $CD19^{+}CD24^{HIGH}CD38^{HIGH}CD27^{-}CD10^{+}$ lymphocytes.[231,232] Abundant in cord blood, their percentage progressively decreases during infancy until they represent just ~4% of the total B cell pool in the adult peripheral blood. Like those of mice, human transitional B cells can be divided into two main populations, termed transitional 1 (T1) and transitional 2 (T2) B cells.[231] T1 B cells are the earliest transitional B cells and are characterised by an $IgM^{HIGH}IgD^{-}CD24^{HIGH}CD38^{HIGH}CD21^{LOW}CD23^{-}$ phenotype; in contrast to murine T1 B cells, they also express CD5. T2 B cells retain high levels of surface IgM, CD24 and CD38 but are also IgD^{+}, $CD21^{+}$ and $CD23^{+}$. It has been demonstrated that all mature B cells in the peripheral blood are derived from transitional B cells; T1 B cells give rise to T2 and fully mature B cells.[232] Furthermore, T1 and T2 B cells have been shown to display functional differences in Ag receptor-mediated signalling. In response to BCR engagement, T2 B cells up-regulate survival signals, proliferate and differentiate into follicular mature B cells, whereas T1 B cells undergo apoptosis; consistently, T1 B cells are a target for BCR-induced negative selection in the peripheral compartment.[231]

An increase in transitional B cells has been described for a number of immunodeficiency states characterised by impaired humoral immunity, such as X-linked lymphoproliferative disease, and in patients recovering from haematopoietic stem cell transplantation, as well as in patients with autoimmune diseases, such as systemic lupus erythematosus.[231-233] The precise role of transitional B cells in human B-cell differentiation has not been fully established. However, it is likely that these cells represent a critical checkpoint in establishing B-cell tolerance and that the breakdown of this checkpoint may contribute to the pathogenesis of autoimmune diseases including systemic lupus erythematosus.

1.3.3. Secondary lymphatic organs

Mature B cells express surface IgM and IgD in the context of the BCR, as well as the BCR co-receptors CD19 and CD21, the adhesion molecules CD20 and CD22, and CD40. Mature, but Ag naïve, B cells enter the secondary lymphatic organs of the spleen, tonsils, lymph nodes and mucosa-associated lymphatic tissues (MALT) where they encounter Ag. Here, interaction with Ag leads to their activation, proliferation and ultimately their terminal differentiation into either plasma B cells, which secrete Ab, or memory B cells. During this process, selection occurs, which is dependent on the affinity of a B cells sIg for Ag. Subsequently, affinity maturation is induced, whereby the affinity of the Ig for Ag is increased by successive modifications to the heavy and light chain V regions (see section 1.2.4.1.). In addition, switching of Ig class from IgM to IgG, IgA or IgE can occur, which alters Ab effector function (see section 1.2.4.2.). These processes rely on the interaction of B cells with T cells and specialised Ag presenting cells (APCs) within a defined microenvironment called the germinal centre (GC).

1.3.3.1. The primary and secondary immune response

Two major types of response to Ag can be induced in B cells; T cell-independent (TI) and T cell-dependent (TD).[1,3,4] T cell-independent activation of B cells takes place in the marginal zone of the spleen and the paracortices of the lymph node. A particular subset of B cells expressing the T cell pan marker CD5 (B1 cells) are activated in response to TI Ags. Typically, TI Ags are large, highly repetitive carbohydrate Ags, such as bacterial cell wall polysaccharides, or polymeric proteins, such as bacterial flagellin. Two types of TI Ag (TI-1 and TI-2) can elicit a response, which occur by different mechanisms. Most TI-1 Ags are polyclonal B-cell activators that activate B1 cells regardless of their specificity (polyspecific), such as lipopolysaccharide, a major component of the cell walls of gram-negative bacteria. TI-2 Ags activate B1 cells by extensively cross-linking the BCR. Although the response to TI-2 Ags does not require direct T cell involvement, cytokines

derived from T_H cells are required for class switching to isotypes other than IgM. Exposure to TI Ags generally leads to the production of low affinity Ab, of which IgM tends to predominate. No memory cells are formed.

Much more frequent is a T-cell-dependent B-cell response. This response requires T cell help and is provided by T_H cells via direct cell-cell interactions and cytokine-mediated signals. Primary exposure to a TD Ag leads to the production of short-lived, IgM-producing plasma cells, which can be detected after 7 days. The response peaks at 2 weeks and then decays. After repeated exposure, IgM Ab is generated more rapidly and followed by a sharp rise in serum IgG after 4 days. The IgG response declines slowly and leaves behind efficient memory that persists for years and even decades.

1.3.3.2. The primary response in the germinal centre

1.3.3.2.1. Primary B cell activation

When Ag is introduced into the body it becomes concentrated in various peripheral organs. For example, blood-borne Ag is filtered by the spleen, whereas tissue Ag is filtered by the regional lymph nodes. Upon leaving the BM, mature B cells are carried through the circulation to the sites of first Ag contact in the marginal zones of the secondary lymphoid tissues: paracortex of the lymph nodes or the periarteriolar lymphatic sheaths (PALS) in the spleen.[1,3,4] Both of these sites are rich in T cells and interdigitating dendritic cells (IDCs), a type of APC. Ag binds to the BCR on the mature B cell and cross-linking and activation leads to an increased expression of MHC-II and co-stimulatory B7.1 (CD80) and B7.2 (CD86) molecules. This provides a relatively weak first signal. Subsequently, B cells internalise the BCR-Ag complex and degrade Ag into peptides that are displayed at the cell surface complexed to MHC-II. In parallel, IDCs present Ag complexed to MHC-II to T_H cells, resulting in their activation and proliferation. Recognition of peptide-loaded MHC-II on B cells by activated T_H cells leads to the interaction of the two cells, forming a T-B conjugate. This interaction is strengthened by the adhesion molecules lymphocyte function-associated Ag 1 (LFA-1, also known as integrin $\alpha L\beta 2$) on the T_H cell, and intercellular adhesion molecule-1 (ICAM-1) on the B cell. The T_H cell begins to express CD40 ligand (CD40L), which can interact with CD40 on the B cell and provides a second signal for B-cell activation. The CD40-CD40L interaction simultaneously induces the polarised release of cytokines by the T_H cell and the cell surface expression of various cytokine receptors by the B cell. Following B-cell activation, small foci of proliferating B cells form at the edges of the T cell-rich zone, reaching a maximum size at day 3-4. The B cells within these foci can take one of three paths (Figure 1.15.):

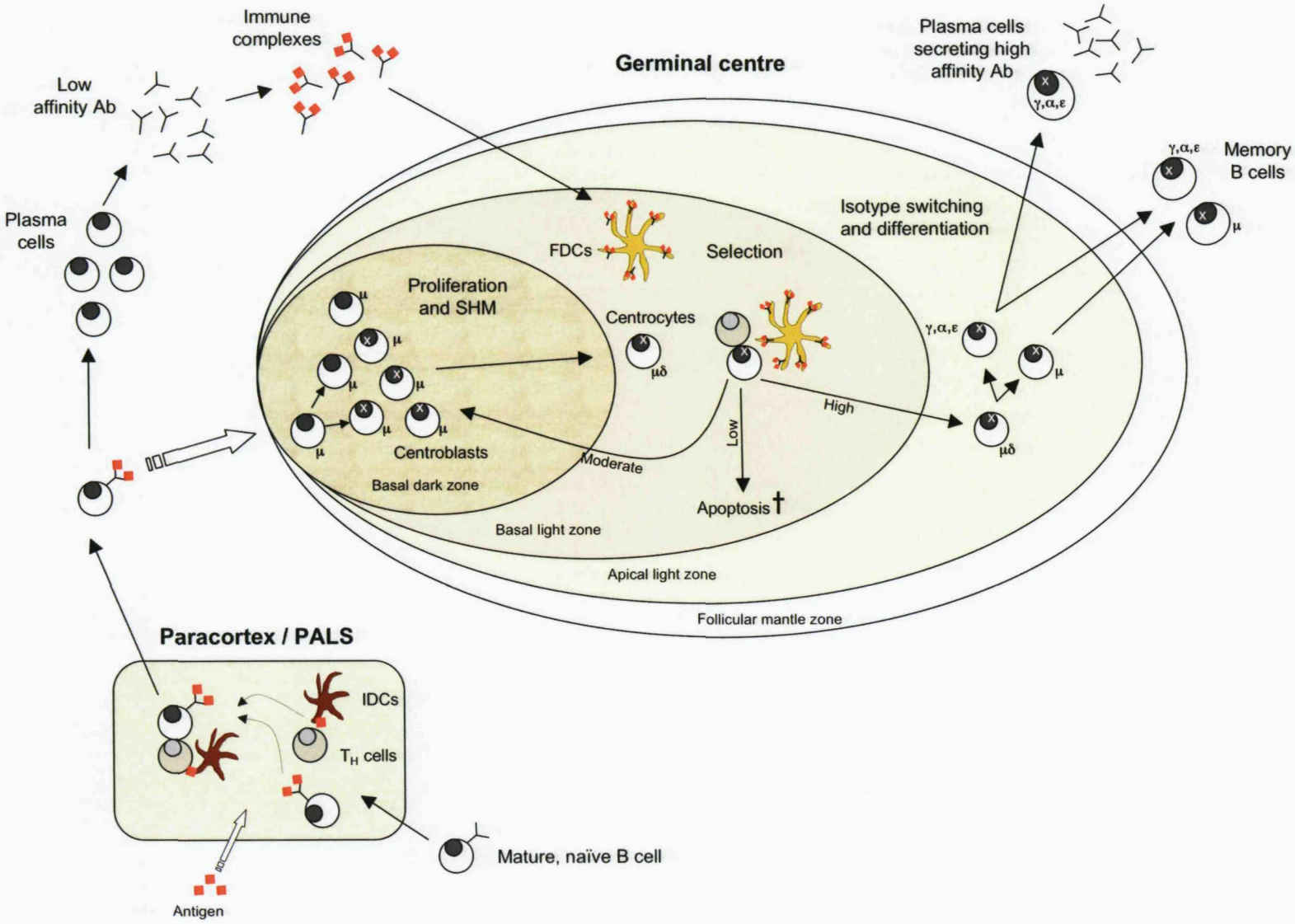


Figure 1.15. A schematic representation of the germinal centre reaction. Antigen-specific B cell activation starts outside the GC in an encounter with Ag presenting interdigitating cells (IDCs) and activated T_H cells. The majority of the B cells proliferate and differentiate into plasma cells that produce low affinity antibodies (Ab). Some activated B cells enter the primary follicles, which then become secondary follicles, and become centroblasts, that undergo proliferation and somatic hypermutation (SHM) in the basal dark zone. Non-dividing centrocytes move into the basal light zone where they interact with Ag-Ab complexes on follicular dendritic cells (FDCs). High affinity B cells can go on to isotype switch and differentiate into plasma cells or memory B cells in the apical light zone. Centrocytes can re-enter basal dark zone for further rounds of division and SHM. B cells with low affinity receptors under go apoptosis. Adapted from Han *et al.*, 1997.[238]

1. Under the influence of cytokines such as IL-6, the majority of B cells differentiate into plasmablasts and rapidly exit the marginal zone. They undergo a burst of rapid proliferation and form plasma cell foci in the extrafollicular regions. These early plasma cells usually secrete low affinity IgM Ab, although isotype switching can occur, and have a life span of around 3 days, after which they die by apoptosis.[234] Most of the Ab produced during a primary response comes from plasma cells in these foci.

2. A minority of B cells, known as GC founder cells ($CD38^{++}IgD^{+}CD27^{-}CD77^{-}$), migrate, along with T_H cells, to the primary follicles.[235] These follicles develop into the specialised GC microenvironment for the process of affinity maturation and class switching.

3. Some B cells can leave the lymphatic organs and re-circulate as resting B cells.[236]

1.3.3.2.2. The germinal centre reaction

The GC is a complex cellular microenvironment that directs post-V(D)J diversification and selection.[237] Important B-cell differentiation events take place here: somatic hypermutation (SHM), class switch recombination (CSR) and the formation of plasma cells and memory B cells (see sections 1.2.4. and 1.3.3.2.3.).

GCs form in the secondary lymphoid tissues 4-5 days after initial Ag exposure and possess a distinct morphology that includes a follicular mantle zone, a basal dark zone, a basal light zone and an apical light zone (Figure 1.15.). The process begins after a few Ag-specific B cells, together with Ag-specific T_H cells, migrate from the paracortex/PALS to the primary B-cell follicles, which then develop into secondary follicles. Initial proliferation marks the loss of surface IgD and the down-regulation of the anti-apoptotic protein B-cell CLL/lymphoma 2 (Bcl-2) in these B cells, which are now $IgM^{+}IgD^{-}CD77^{+}CD10^{+}Bcl\text{-}2^{-}$.[236,238] It has been estimated that 3-10 B cells are required to initiate the GC reaction, which leads to small foci containing ~50 rapidly proliferating B cell blasts, called centroblasts.[237,239,240] The cell cycling time of the centroblasts is around 6-7hrs.[241] Further rapid proliferation leads to the generation of the dark zone, which becomes densely packed. During proliferation, centroblasts undergo extensive SHM of their V_H and V_L genes, which occurs in a stepwise fashion between several rounds of cell division (see section 1.2.4.1.). Mutation is not entirely random, but subsequently affects the binding affinity of the Ig receptor for Ag; some cells will be generated with a higher affinity for Ag, while others will have little or no affinity.[242] SHM does not appear to commence until well into the primary response on day 7, where it continues at a steady rate until day 21. In

addition to SHM, isotype switching is likely to initiate in the dark zone (see section 1.2.4.2.).

Upon exit from the dark zone, cells move into the basal light zone and become non-dividing centrocytes, which now express sIgD. At the molecular level, the Ig V genes are somatically mutated. Selection of high affinity Ig requires follicular dendritic cells (FDCs), which represent a specialised type of APC.[reviewed in 243] FDCs differ from conventional APCs by the fact that they do not internalise, process and present antigenic peptides on MHC-II; in fact, they do not express MHC-II. Instead, FDCs capture and present Ag in the context of immune complexes, which have been formed by the binding of low affinity IgM Abs to Ag, via the Fc receptors CD21 (CR2) and CD35 (CR1).[244] These immune complexes are retained on the cell surface as iccosomes for months to years.

Centrocytes express high levels of Fas (CD95), which is capable of auto-activation.[245] The combined effects of high Fas and low Bcl-2 mean that apoptosis is inevitable unless survival signals are received. Centrocytes with a high affinity Ig receptor bind to Ag as it is displayed on FDCs. The combination of BCR cross-linking leading to Bcl-2 up-regulation and the survival signals provided by FDCs ensure that these high affinity B cells survive (positive selection); FDCs can block apoptosis by supporting high levels of cellular FLICE-like inhibitory protein (cFLIP) expression in the attached B cell, thereby preventing the expression of caspase-8.[246] Engagement of adhesion molecules plays a crucial role in maintaining cell-cell contact; ICAM-1 and VCAM-1 on FDCs and LFA-1 and VLA-4 on B cells are key elements.[242,247,248] Interaction with FDCs leads to the up-regulation of MHC-II and the co-stimulatory molecules B7.1 and B7.2 on the B cell. Internalisation of the BCR-Ag complex leads to the presentation of Ag in the context of MHC-II on the cell surface. Ag-specific T_H cells recognise Ag and stimulate centrocytes via CD40-CD40L interactions and cytokines. Moreover, interactions with Ag-specific T_H cells via CD40-CD40L can mimic cFLIP up-regulation and enhance cell survival.[246] Centrocytes selected in the basal light zone can re-enter the dark zone and undergo further rounds of proliferation and SHM; the resulting progeny again move into the basal light zone for selection.

Centrocytes with low affinity Ig receptors that do not bind Ag do not up-regulate Bcl-2, nor do they receive survival signals from FDCs or Ag-specific T_H cells. These centrocytes succumb to apoptosis and are phagocytosed by tingible body macrophages, a characteristic

feature of GCs. Approximately 90% of B cells undergo apoptosis in the basal light zone. The elimination of auto-reactive B cells by Fas-mediated apoptosis (negative selection) also occurs in the basal light zone; auto-reactive B cells cannot bind to Ag displayed on FDCs, nor do they receive help from Ag-specific T_H cells. However, re-expression of the RAG proteins has been observed in centrocytes, raising the possibility that secondary V(D)J recombination, specifically at the κ or λ light chain loci, may salvage auto-reactive B cells from apoptosis (receptor revision), so long as the new light chain restores specificity for Ag present on the FDCs.[249,250]

Once positively selected, centrocytes move into the apical light zone, where upon they complete isotype switching, as demonstrated by the expression of sterile germline Ig transcripts other than IgM, and initiate differentiation into plasma cells or memory B cells (Figure 1.15.; see section 1.3.3.2.3.).

Affinity maturation is a cyclical process of rapid cell proliferation and SHM followed by a period of cellular rest during which selection occurs. Several rounds are required to generate the 10-20 beneficial mutations needed to form high affinity Ab. The outcome of the GC reaction is the clonal expansion and differentiation of B cells with a high affinity for a specific Ag. There is no trafficking of B cells between individual GCs of a secondary lymphoid organ.[239]

1.3.3.2.2.1. Ectopic germinal centre formation

It has become clear that the formation of GCs is not restricted to the connective tissue of *bona fide* secondary lymphoid organs alone. In a number of chronic inflammatory situations the infiltration of B and T cells can result in the formation of follicular structures resembling those of the secondary lymphoid tissues, termed "ectopic" GCs. FDCs can be observed at a number of unexpected sites, including the synovial tissue of patients with rheumatoid arthritis, the salivary glands of patients with sjogren's syndrome and the thymus of patients with myasthenia gravis, as well as the vessel walls under atherosclerotic plaques and aortic aneurysms.[reviewed in 243] FDCs in the inflamed synovium of rheumatoid arthritis patients express many features characteristic of FDCs, including VCAM-1, B-cell attracting chemokine 1 (BCA-1) and CD21, and can exert typical FDC-functions *in vitro*.[251]

1.3.3.2.3. Commitment to plasma and memory B cells

In the GC, centrocytes have three developmental options: to continue further rounds of mutation and selection as GC B cells, to become memory B cells or to become plasma cells. Differentiation into plasma or memory B cells is initiated in the apical light zone and depends upon signals provided by FDCs and/or Ag-specific T_H cells. In addition, BCR affinity for Ag has also been proposed to influence cell fate, with the suggestion that cells bearing high affinity BCR are selected to differentiate into plasma cells, while cells with moderate affinity BCR differentiate into memory B cells.[252]

Plasma cells are terminally differentiated final effectors of the humoral immune response. The sole function of plasma cells is to secrete high affinity soluble Ab; in fact, membrane Ig and the BCR are low or absent. Commitment to a plasma cell fate follows the differentiation of centrocytes into plasmablasts, which occurs within the GC. Plasmablasts then migrate to the BM, which is mediated by changes in the expression of specific chemokine receptors; CXCR5 and CCR7 are decreased, while CXCR4 remains high.[253] Terminal differentiation into plasma cells occurs independently of Ag in the BM and is usually preceded by the robust proliferation of plasmablasts. However, cessation of cell division is a prerequisite for plasma cell differentiation.[254] BM stromal cells provide signals that promote plasma cell survival and possibly final differentiation of plasmablasts; one critical differentiation and survival factor provided by stromal cells is IL-6.[255] More recently, B-cell activation factor of the TNF family (BAFF), produced by cells of the myeloid cell lineage has been identified as an important factor in B-cell survival during differentiation.[reviewed in 256] BAFF has been shown to be important for the survival of long-lived plasma cells in the BM, as well as the survival of plasmablasts, through its interaction with BAFF receptor (BAFF-R) and B-cell maturation antigen (BCMA) expressed on B cells. Furthermore, a proliferation-inducing ligand (/APRIL), also a member of the TNF family, can similarly bind to BCMA and lead to plasmablasts survival. The life span of non-proliferating plasma cells varies from a few days to many months. Long-lived plasma cells predominately populate the BM and organs that are directly exposed to foreign Ag, such as the gastrointestinal tract and the lung. They continue to secrete Ab for many months in the absence of Ag or cell proliferation.[257]

Unlike plasma cells, memory B cells do not secrete Ab, but continue to express surface Ig, which may be IgM or isotype switched. Upon exiting the GC, memory B cells recirculate and are thought to preferentially home to the draining areas of the lymph node and spleen, including the marginal zones, where they reside for long periods; memory B cells are long-

lived, even in the absence of Ag.[258] Upon secondary exposure to Ag, memory B cells are biased to become plasma cells.[259] A rapid and considerable clonal expansion occurs, which generates 8- to 10-fold more plasma cells than in a primary response.[260] The heavy chain isotype of the BCR can influence the fate of the B cell, with IgG triggering a significantly larger proliferative burst than IgM, consistent with the dominance of IgG isotypes in the memory response.[261] In addition to the formation of plasma cells, some memory cells persist to replenish the memory compartment.

A number of transcription factors are thought to play a role in the differentiation of GC B cells into plasma cells or memory B cells. For example, multiple myeloma oncogene 1 (MUM-1/IRF-4), a member of the interferon regulatory factor family, appears to play a role in plasma cell commitment and is highly expressed on plasmablasts.[262] However, the transcriptional repressor B-lymphocyte-induced maturation protein 1 (Blimp-1) is known to be critical for plasma cell differentiation. Blimp-1 is expressed in all plasma cells, including those formed in a primary response to either TI- or TD-Ag, those formed from memory B cells in a secondary response and in long-lived plasma cells in the BM.[262] Blimp-1 initiates a cascade of gene expression changes that promote plasma cell differentiation, including the direct repression of other transcription factors that regulate cell proliferation, BCR signalling and GC function, and the activation of transcription factors that regulate plasma cell functions.[263] For example, Blimp-1 represses *AID*, *DNA-PK$_{CS}$*, *Ku70* and *Ku80*, and *c-Myc* gene expression leading to the cessation of SHM and CSR and cell cycle arrest, respectively.[264] Furthermore, Blimp-1 represses the transcription factor Paired box gene 5 (Pax5), an inhibitor of plasma cell differentiation and Ab secretion; Blimp-1–dependent repression of Pax5 is required for the induction of X-box-binding protein 1 (XBP-1), which is also important for plasma cell formation and Ab secretion.[265] Certain cytokines secreted by FDCs and/or T_H cells may induce Blimp-1 expression *in vivo*, these include IL-2, IL-5, IL-6 and IL-10, which have been shown to induce Blimp-1 expression and lead to plasma cell differentiation *in vitro*.[reviewed in 264] Moreover, Blimp-1 expression occurs only in the absence of CD40 engagement.

Conversely, memory B cells are generated following a T_H-derived signal triggered by the interaction of CD40 on centrocytes and CD40L on activated T_H cells, as well as signalling through the BCR. Signalling through CD40, which represses Blimp-1 expression and therefore, plasma cell differentiation, is irreversible and sufficient to induce B-cell memory.[266,267] In addition, cytokines can influence the development of memory B cells; IL-4 has been shown to direct a memory B cell fate *in vitro* via the induction of Bcl-6

expression, which negatively regulates Blimp-1 expression, thus blocking plasma cell differentiation.[268] The relative expression of Bcl-6 and Pax5, which inhibit plasma cell differentiation, and Blimp-1 and XBP-1, which promote the differentiation and function of plasma cells, is crucial in determining the GC B cell, memory B cell and plasma cell fate decisions.

1.3.3.2.4. Termination of the germinal centre response

The life span of a GC is about 3 weeks; the reaction is terminated by the withdrawal of Ag and stimuli provided by the FDCs, which is instigated after 2 weeks.[242] At one month post Ag exposure, the GC has contracted and appears as a sparse accumulation of blasts and associated FDCs.[269] Lacking any further Ag stimulation, FDCs lose the ability to produce iccosomes and instead form membrane pockets containing Ag.[244,270] In this way, Ag is removed as a stimulus, but stored long term.

1.3.3.2.5. Selection of the peripheral Ig repertoire

The immune repertoire of B cells found in the peripheral blood and lymphatic tissues is greatly influenced by the process of antigenic selection. Although, it is important to note that naïve, Ag-inexperienced B cells also contribute to the peripheral B-cell pool and, in this instance, biased usage of particular V genes is related to molecular mechanisms and not antigenic selection (see section 1.2.3.5.). Studies using human adult peripheral B cells reveal that the V_H3 family is over-represented and that this is largely a result of the preferential usage of a small number of specific individual V_H genes, including V3-23 and V3-30.3.[97,271,272] Similarly, the V_H4 family members, V4-34, V4-59 and V4-39 are also found significantly more frequently than expected. Studies investigating Vκ and Vλ gene usage find there is also a strong bias; the single Vκ4 family gene segment, B3/DPK24, is significantly over-represented, as are the L2/DPK21, A27/DPK22 and L6/VgK38 genes from the Vκ3 family and the 02-012/DPK9 and 08-018/DPK1 genes from the Vκ1 family.[48,76,105,273] The over-representation of B cells expressing specific V genes likely results from a preferential expansion of these B cells in response to positive selection by an Ag or superantigen (SAg) (see below). Indeed, during immune responses to exogenous pathogens, B cells using certain V gene segments can transiently increase. This has been demonstrated for the V4-34 gene, which is found in response to Epstein-barr virus (EBV) infection.[274] Similarly, a V4-34 gene bias has been observed in autoimmune disorders, such as cold agglutinin disease (I/i Ags) and systemic lupus erythematosus.[275-277]

Unlike classical Ags, SAgs are characterised by their ability to interact with BCR independently of CDR3, outside of the conventional Ag-binding groove. Instead, SAgs bind to BCR via the conserved FR regions of the V regions, more commonly the Ig heavy chains. In this way, SAgs are able to target and activate large numbers of B cells. SAgs are produced by many pathogens, including bacteria, mycoplasma and viruses and some examples include protein A of *Staphylococcus aureus* (SpA), gp120 of HIV-1 and protein L of *Peptostreptococcus magnus*. In addition, the red blood cell Ags, I/i, have some features of a B-cell SAg-I binding is via the FR1 region of V4-34-encoded Ig, but can be modulated by sequences in the CDR3.[278] One of the first B cell SAgs to be characterized was SpA, which was shown to bind to the conserved FR1 and FR3 regions and CDR2 of most V_H3 family genes, but not to Igs from other V_H families.[279-281] Different V_H3 Igs vary in their SpA-binding activities and analysis has revealed a hierarchy that best correlates with the usage of specific V_H3 gene segments; V3-23 has the highest SpA-binding activity and, importantly, is closest to the overall consensus sequence for all V_H3 family genes. Since the V_H3 family is the largest human V_H gene family, it is no surprise that, as a result, 30-50% of circulating human peripheral B cells have been shown to display Fab-mediated SpA binding activity.[279] Therefore, SAgs can have a profound effect on the peripheral B-cell pool.

1.4. B cell malignancy

Cancer encompasses a class of over 100 separate diseases that are characterised by cells that cannot respond to normal growth control mechanisms, including regulation of cell growth and proliferation and cell death.[1] This leads to the accumulation of clonal malignant cells that can further expand to produce a tumour or neoplasm. For malignancy, tumour cells must have the ability to invade other tissues, either by direct growth into adjacent tissues, through invasion, or by implantation into distant sites by metastasis, whereby cells are transported through the bloodstream or lymphatic system. Malignancy often leads to the death of the host. Death can occur due to the physical impairment of vital organs, such as the brain, BM or lungs, or by the production of specific cytokines or metabolic products, which in turn can upset the fine balance of salts and chemicals present in the body and lead to the fatal impairment of other crucial organs, such as the liver.

Tumours of the humoral immune system are classified as lymphomas or leukaemias (and myeloma); lymphomas proliferate as solid tumours within a lymphoid tissue such as the BM, lymph nodes or thymus, while leukaemias proliferate as single cells, circulating in the blood or lymph. However, there is substantial overlap between lymphomas and

leukaemias, as the former can develop leukaemic phases, while the latter may also involve nodal or other organ structures as solid tumours.

1.4.1. The pathogenesis of B cell malignancies

The mechanism of lymphomagenesis is not completely understood. However, it is accepted that its origins lie in genetic events that are associated with DNA damage and evasion of DNA repair mechanisms. Lymphomagenesis is generally initiated by a genetic lesion that results from an error during normal cell development or from unrepaired physical or chemical damage to the genome.[282] Rarely, the abnormal gene is inherited, resulting in an increased susceptibility to cancer for those descendents who have inherited the gene.[283] This initial event provides an increased chance for additional genetic lesions to develop, which usually occur over a number of years or even decades. Epigenetic changes, such as DNA methylation and acetylation, may also present heritable, but potentially reversible, changes. When a cell acquires the proper combination of genetic/epigenetic lesions, it will have the full potential to generate a malignant tumour. As the neoplastic cells continue to divide and expand, additional genetic/epigenetic alterations may be acquired and some of these may contribute to characteristics that make the tumour more clinically aggressive, such as transformation from low- to high- grade disease, and/or resistance to treatment. Indeed, the majority of B-cell malignancies have recurrent and disease-specific clonal chromosomal abnormalities that play a pivotal role in tumour development. These comprise deletions or duplications of either entire chromosomes or subchromosomal regions, and chromosomal translocations. Malignancies of immature B cells are often associated with a single chromosomal translocation, while those of mature B cells exhibit enormous cytogenetic complexity, with multiple translocations and deletions, consistent with their slower natural development and the hypothesis that several genetic lesions are required for the full neoplastic phenotype to emerge.[reviewed in 284]

1.4.1.1. Chromosomal translocations

Chromosomal translocations are found in some, but not all, forms of B-cell malignancy and often involve the Ig loci, occurring more frequently at the IgH locus.[reviewed in 284] Multiple Ig translocations can occur, which may involve the IgH locus exclusively, or combination of the three Ig loci: IgH and/or Igκ and/or Igλ.[285] Translocations that are observed in all or the vast majority of tumours of a given subgroup of disease can be useful as diagnostic markers; examples include t(11;14), t(8;14)/t(8;22) and t(14;18), which are characteristic of mantle cell lymphoma (MCL), Burkitt's lymphoma (BL) and follicular lymphoma (FL), respectively.[286]

The principal molecular consequence of Ig translocations is the deregulated expression and, in certain instances, the mutation of the translocated gene. Deregulated expression of the incoming gene arises due to the close proximity of potent B-cell transcriptional enhancers within the Ig loci, which may result in abnormally high protein expression. In GC B cells where the SHM machinery is active, mutation of the incoming gene may be observed and may contribute to the neoplastic phenotype. For example, mutations in the promoter region of the *Bcl-6* oncogene, which may feature in the absence of translocation, can prevent the normal down-regulation of the Bcl-6 protein and result in its over-expression; *Bcl-6* is highly expressed and mutated in normal GC B cells and is over-expressed in many diffuse large B-cell lymphomas (DLBCL).[215,287-289] Similarly, mutations in the *c-Myc* oncogene that can lead to c-Myc over-expression have also been described.[290] However, most genes expressed in this context do not alone result in B-cell neoplasm. For example, the t(14;18) chromosomal translocation, which is commonly found in FL (see below), has been observed in normal individuals in the absence of malignancy.[291,292] Therefore, several genetic abnormalities, including the concurrent activation of dominant oncogenes and inactivation of tumour suppressor genes, are necessary for the full neoplastic phenotype to emerge.

The Ig loci show comparable promiscuity in their translocation partners. Many of the genes involved in Ig translocations play important roles in the development and activation of normal B cells, particularly within the GC, as well as contributing to malignancy. Most appear the have functions in at least one of four mechanisms:

1. Cellular regulation, including apoptosis (*Bcl-2*, *p53*):
For example: The t(14;18)(q32;q21) represents the cytogenetic hallmark of FL and is observed in 80% of tumours.[293,294] This translocation juxtaposes the *Bcl-2* gene to the Ig heavy chain locus and leads to its deregulated expression, which in turn results in the over-expression of the anti-apoptotic Bcl-2 protein. Therefore, t(14;18) may confer cell survival via suppression of apoptosis. The t(14;18) is also observed in a proportion of DLBCL (20%) and fraction of chronic lymphocytic leukaemia (CLL) (1-2%).[284]

2. Cell cycle progression (*c-Myc*, *Bcl-1/cyclin* D_1, *cyclin* D_2, *CDK6*)
For example: MCL is consistently associated with the t(11;14)(q13;q32) translocation.[295,296] This translocation juxtaposes the *Bcl-1* gene on chromosome 11q13 to the IgH locus and leads to its deregulation. The consequence of this translocation is the over-expression of cyclin D_1, a member of the D-type G_1 cyclins involved in cell cycle control, which leads to

the acceleration of the cell cycle through G_1 phase.[297,298] Therefore, t(11;14) may lead to uncontrolled cell proliferation. The t(11;14) is also observed in a proportion of multiple myeloma (MM) (20%).[284]

3. NF-κB activation (*NF-κB2*, *Bcl-10*, *Bcl-3*)

For example: Five percent of extranodal MALT lymphomas have translocations that disrupt the *Bcl-10* gene on chromosome 1p22, which leads to the over-expression of Bcl-10 protein.[299] These translocations typically juxtapose the *Bcl-10* to the IgH locus, t(1;14)(p22;q32), but may involve the Igκ locus t(1;2)(p22;p12). Wild-type Bcl-10 protein functions to activate NF-κB and promote apoptosis. However, studies suggest that SHM of *Bcl-10* occurs in MALT lymphomas harbouring the t(1;14), which leads to the over-expression of truncated Bcl-10 mutants.[300] It is suggested that these mutants lack the pro-apoptotic function seen in wild-type Bcl-10. Therefore, t(1;14) may confer a survival advantage, as well as leading to the constitutive NF-κB activation.

4. Signal transduction pathways (*FcγIIB*, *Pax5*)

For example: Up to 50% of lymphoplasmacytoid lymphoma (LPL) carry the t(9;14)(p13;q32).[301] This translocation juxtaposes the *Pax5* gene to the IgH locus and leads to the deregulated expression of *Pax5*.[302] The *Pax5* encodes a B-cell-specific transcription factor involved in the regulation of activation and proliferation of mature B cells. The over-expression of Pax5 protein is thought to interfere with the normal inactivation of *Pax5* transcription during plasma cell differentiation and results in enhanced proliferative signaling via the BCR.[284] Therefore, t(9;14) may lead to uncontrolled cell proliferation.

1.4.1.2. A role for RAG1/2 and AID

Many B-cell malignancies are associated with cell 'arrest' at the stages of differentiation where rearrangements and mutations in the Ig gene occur. This association has been attributed to errors in resolution of the DSBs that arise during the course of normal V(D)J recombination, CSR and the ssDNA nicks that occur during SHM.[215] Indeed, most Ig translocations are targeted to sites commonly involved in normal recombination events, such as the J segments and, in the case of IgH translocations, the S regions. For example, most *Bcl-2* translocations in FL involve the J_H gene segments, while *c-Myc* translocations in BL often involve the J_H genes or S region.[303,304] Translocations at the IgH locus may also involve the D segments and the J_H-Cμ intron.

Targeting of oncogenes to the J regions would suggest that such translocations might arise due to errors during V(D)J recombination within the BM (see section 1.2.3.).[305] Indeed, the t(14;18) translocation in FL is thought to occur early in B-cell development at the stage of D-J_H joining.[306] The RAG proteins share many similarities with DNA transposases and it has been demonstrated that RAG1/2 can perform transpositional attack *in vitro*, inserting RSS ends into a second DNA and forming hybrid joints.[307-309] This activity of the RAG proteins suggests that they might also be able to use RSSs to attack an exogenous DNA, whereby an RSS-ended chromosome fragment could be linked to an non-RSS site on the partner chromosome, thus leading to chromosomal translocations. This has been demonstrated for t(14;18) where a pseudo RSS on chromosome 18 is used by the recombinatorial machinery to join the *Bcl-2* gene to the IgH locus on chromosome 14 at a regular RSS, leading to the over-expression of Bcl-2.[294,305,306,310]

Analysis of both productive V(D)J rearrangements and both derivative chromosomes demonstrates that many translocations arise following Ag contact during a GC reaction. The observation that AID can cause mutations in bacteria and many different cell types make it clear that AID is highly mutagenic and potentially dangerous, requiring tight regulation and efficient targeting (see section 1.2.4.3.).[195-198] The deregulation and aberrant targeting of AID within the GC represents a powerful mechanism for malignant transformation, contributing to aberrant chromosomal translocations and mutation of non-Ig genes. Indeed, the constitutive and ubiquitous expression of AID in transgenic mice was shown to cause T-cell lymphomas.[311] In humans, AID-induced mutations appear to contribute to the high incidence of B-cell malignancies.[215,304] The reactivation of the *RAG* genes in GC B cells may provide a second mechanism and may in part explain the targeting of oncogenes to the J regions.[249]

1.4.2. Classification of B cell malignancies

Lymphomas and leukaemias are a complex and diverse group of cancers comprising around 30 related, but distinct diseases. Historically, the first entity to be recognised was Hodgkin's disease (HD), which was discovered by Thomas Hodgkin in 1832.[312] Other malignancies of the lymphatic system were subsequently called Non-Hodgkin's lymphomas (NHL) a distinction that remains valid today. The purpose of further classifying NHL into specific categories is to describe the individual behaviour and to develop appropriate treatment strategies for each type of lymphoma. Classification of NHL has used several different systems over the last 30 years and is accompanied by long-

term controversy, primarily owing to an evolving understanding of tumour cell physiology, immunophenotyping and genetic and molecular analyses.[reviewed in 313]

Until the 1970's, pathologists depended heavily on the morphologic appearance of the lymphomas to categorise them; the oldest classification is the Rappaport system, which was developed before lymphoid cells were divided into B cells and T cells.[314] In 1974, the Kiel classification and Lukes and Collins classification were introduced, both of which incorporated immunophenotypic, as well as morphologic, information.[315,316] The latter system provided the first distinction between B-cell and T-cell lymphomas. In 1982, the International Working Formulation (/IWF) was proposed by the National Cancer Institute to unify descriptive terminology and facilitate comparisons among 6 different classification systems for NHL in existence.[317] This lead to the designation of three major NHL categories- low, intermediate and high grade-, which were based principally on tumour cell morphology and prognosis.

In 1994, reflecting increasingly more sophisticated diagnostic tests and the growing understanding of tumour biology, the Revised European-American Classification of Lymphoid Neoplasms (REAL) was constructed.[318] This system defined a particular NHL by incorporating all available tumour information, including morphologic, immunophenotypic, genetic and clinical features. The proposed major groups included the B-cell lymphomas, the T-cell lymphomas and Hodgkin's lymphoma. A project to update and revise the REAL system was initiated by the World Health Organisation (WHO) in 1995; its consensus published in 1999.[286,319] The WHO classification, like REAL, incorporates a number of tumour characteristics and is designed to enable disease identification by pathologic examination while maintaining clinical relevance. With the use of the WHO classification, treatment is determined by identifying the specific lymphoma type and, if relevant, by considering tumour grade and other prognostic factors.

1.4.3. Immunoglobulin genes in B-cell malignancies

Beyond the morphological, immunophenotypic and genetic information used for diagnosis, molecular analyses have broadened our knowledge of the development of B-cell lymphomas.[320,321] Normal B cells undergo a series of recombinatorial and mutational changes during differentiation that lead to unique sequences in the heavy and light chain V genes, which are permanently imprinted in the genome (see sections 1.2.3. and 1.2.4.). In the event of neoplastic transformation, these sequence modifications are preserved and feature in every cell of the tumour clone. With the exception of HD, B-cell malignancies

typically have functional Ig gene rearrangements and express surface Ig. Consequently, analysis of the Ig V gene rearrangements provides an independent marker for a clonal B-cell population. Moreover, it may help to establish the developmental stage at which neoplastic transformation occurred and assign these cells to their normal counterpart. For example, the mutational status of V genes can provide knowledge about the precursor B cell; unmutated V genes are consistent with a pre-GC origin, while mutated V genes suggest a GC or post-GC origin.[321] In addition, the presence of intraclonal heterogeneity, suggestive of ongoing SHM, is considered to reflect malignant transformation within a GC. Furthermore, V gene analysis may also reveal information related to lymphoma pathogenesis; several B-cell derived leukaemias and lymphomas have been reported to display non-random V gene usage, which may suggest a role for classical Ag or SAg in the initial proliferation of B cells leading to overt neoplasm.

The molecular analysis of Ig V genes contributes to the understanding of the differentiation and development of malignant B cells and to the classification of B-cell malignancies relative to normal B-cell development (Figure 1.16.). Furthermore, molecular data have significant clinical relevance in the prognosis of B-cell tumours. In the following sections, B-cell malignancies will be reviewed with respect to Ig V gene usage, somatic mutation and class switching.

1.4.3.1. Acute lymphoblastic leukaemia

Acute lymphoblastic leukaemia (ALL) is characterised by a subset of $CD5^-$ $CD19^+CD20^+CD10^{+/-}CD23^-$ cells that do not express surface Ig, but express nuclear TdT, and are thought to derive from progenitor or precursor B cells in the BM.[286] Immature ALL can be divided into two subtypes, pre-B ALL and early pro-B ALL, which can be distinguished by the expression of CD10 and cytoplasmic Ig in the former. Studies investigating the Ig V genes of these tumours demonstrated that in the majority of cases Ig V_H gene rearrangements can be detected, which may be V_H-D-J_H or incomplete D-J_H rearrangements.[322-328] Rearrangement at the κ locus may also be evident, occurring at diagnosis or relapse.[323-325,329] Some studies report a low frequency of productive rearrangements (~20%).[324,327,328,330] Moreover, approximately 30% of ALL demonstrate multiple heavy chain recombination events, for which the majority result from unique rearrangements of specific V_H, D and J_H gene segments.[322-328,331] This suggests that the transforming event takes place before the completion of V_H-D-J_H rearrangement and that the recombinatorial machinery is still active allowing *de novo* V_H-D-J_H rearrangement, as well as V_H to D-J_H joining and V_H replacement. A bias in V_H gene usage is evident and

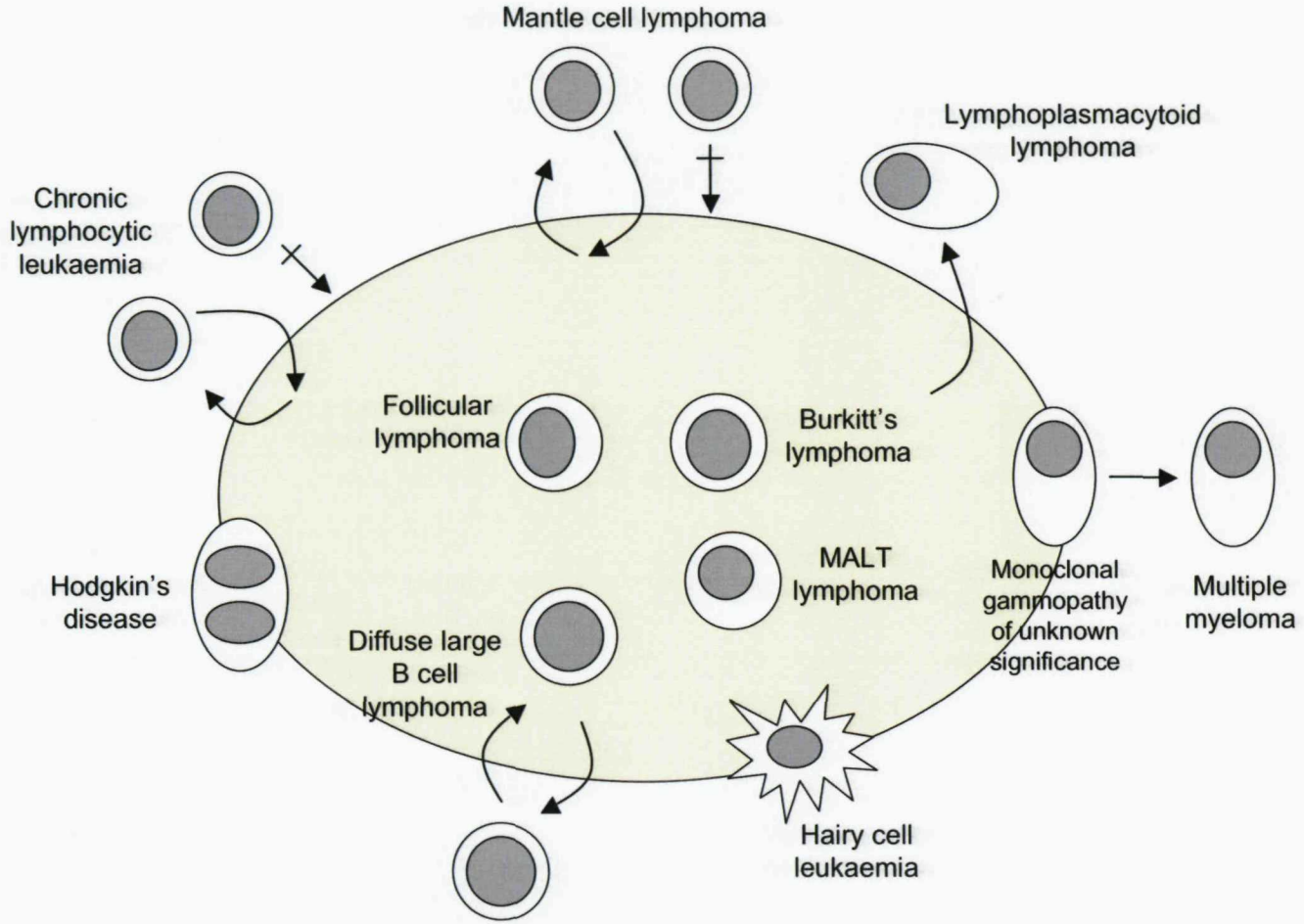


Figure 1.16. The relationship of B-cell tumours to the germinal centre. Mutational patterns in V genes allows the designation of differentiation status and subdivision of known categories. Crossed through arrows indicate that the cell of origin has not accumulated mutations, whereas arrows entering and leaving indicate a stable mutational pattern with no intraclonal heterogeneity. Tumours within the GC continue to accumulate mutations. Adapted from Stevenson *et al*, 2001.[321]

sees the preferential rearrangement of V_H genes that lie proximal to the D and J_H gene segments, such as V6-01, V3-11 and V3-13 (see Figure 1.2.).[324,328,332,333] The over-representation of these V_H genes has also been described for normal immature B cells and those of fetal development.[100,334,335] Furthermore, a restricted usage of D7-27 and J_H4 and J_H6 is described for ALL, similar to normal immature B cells.[94,100,324,327,328] No somatic mutation of the V_H genes has been described.[322,336] These findings are consistent with the presumed normal counterpart of BM precursor B cells.

1.4.3.2. Chronic lymphocytic leukaemia

Chronic lymphocytic leukaemia (CLL) represents a B-cell tumour for which knowledge of Ig V genes has had a significant impact both at the clinical level and on our understanding of disease pathogenesis. CLL is characterised by the progressive accumulation of long-lived, mature B cells, which are typically $CD5^+CD19^+CD20^+CD23^+CD10^-$ and co-express very low levels of surface IgM and IgD. Initial studies found that the Ig V_H genes of CLL were essentially unmutated (<2% somatic mutation); a 2% cut off was used to allow for disparities arising through Ig V_H gene polymorphisms.[337,338] This, together with immunophenotypic and morphological information led to the suggestion that CLL derives from mature naïve B cells, possibly arising in the primary follicle or in the mantle zone of the secondary follicle.[286] Subsequently, reports began to accumulate describing a subset of CLL with mutated V_H genes.[339-341] It is now apparent that CLL represents a heterogeneous disease with mutations in the Ig V genes defining two subsets: unmutated and mutated. Moreover, the two subsets differ in regard to V_H gene usage, which is found to be biased in both cases; the V1-69 gene segment is preferentially used by the unmutated subset, while the V4-34, V3-07 and V3-23 gene segments are preferentially used by the mutated subset.[338,341-343] An increased expression of the V3-21 gene has been described as a feature of mutated CLL in Scandinavia, although this has not been confirmed in other studies and may reflect a geographical bias.[344,345] A preferential usage of specific D and J_H gene segments has also been reported.[346,347] A λ light chain restriction, including the preferential usage of the Vλ3-14 gene segment has been associated with V3-48, V3-53 and V3-21 rearrangements in CLL.[344,348] Furthermore, isotype switching has been observed, although infrequently, featuring mainly in the unmutated subset.[349-352] Intraclonal heterogeneity, which is an indicator of ongoing SHM, is generally considered not to be a feature of IgM or isotype switched CLL of either the mutated or unmutated subsets. However, more recently, intraclonal heterogeneity has been shown in two reports, but this remains controversial.[353,354]

The different patterns of mutation and V gene usage likely reflect distinct histogenetic pathways from B cells at different stages of differentiation. It is probable that mutated CLL arises from a cell that has encountered Ag and passed through the GC (post-GC). Conversely, unmutated CLL may derive from a mature B cell that has not yet encountered Ag, although the biased V_H gene usage does not support this. However, the expression patterns of a number of genes have suggested that both unmutated and mutated CLL derive from Ag-experienced B cells.[355,356] Therefore, it has been proposed that unmutated CLL may arise from $CD5^+$ B1 cells that have been stimulated by T cell-independent Ags, which may also explain the ectopic isotype switching events that have been described in this subset.

An important consequence of Ig V gene studies in CLL is the demonstration that the two subsets behave differently clinically; unmutated CLL is associated with aggressive disease, while mutated CLL is associated with indolent disease.[343] Recent studies demonstrate that tumours displaying mutated V3-21 rearrangements are associated with an unexpectedly aggressive phenotype and may represent a third subset of CLL.[344] Clearly, these findings have significant implications for the diagnosis and treatment of CLL.

1.4.3.3. Mantle cell lymphoma

Mantle cell lymphoma (MCL) is characterised by a subset of $CD5^+CD19^+CD20^+CD10^-CD23^-$ cells that express moderate to high levels of surface IgM and IgD and are thought to derive from the mantle zone of a lymphoid follicle.[286] Initial studies investigating the Ig V_H genes of MCL demonstrated that they are unmutated, consistent with the phenotype of the presumed normal counterpart.[357,358] However, subsequent studies identified a minor (~20%) subset that displayed mutated V genes, but lacked ongoing SHM.[359,360] These findings suggest that MCL is a genetically heterogeneous disease, similar to CLL, with the majority of cases originating from unmutated pre-GC B cells residing in the follicular mantle and a subset deriving from post-GC B cells. Heterogeneity is also apparent at the level of V gene usage. The biased usage of individual V_H genes has been described by several groups and consistently involves the V3-21 and V4-34 gene segments, as well as V3-23, V4-59, V5-51 and V1-08.[358,359,361,362] V4-34 and V3-21 rearrangements were almost exclusively unmutated, while V4-59 and V3-23 were typically mutated.[361] Rearrangements involving V3-21 were almost always expressed with λ light chains and displayed a highly restricted use of the Vλ3-19 gene segment.[359] Bias in D and J_H gene usage has also been described.[361] Furthermore, isotype switching has been documented in

both mutated and unmutated MCL.[363] It is possible that in the unmutated subset switch events are occurring at ectopic sites.

1.4.3.4. Follicular lymphoma

Follicular lymphoma (FL) is characterised by a subset of $CD5^-CD19^+CD20^+CD10^+CD23^{+/-}CD38^+CD75^+$ cells that express surface IgM±IgD (or IgG or IgA) and are thought to derive from GC B cells.[286] Analysis of the Ig V_H and V_L genes demonstrate that they are extensively mutated, with evidence for ongoing SHM, consistent with a GC B cell origin.[364-370] Isotype switching is also described, with some cases expressing multiple clonally related isotypes.[371] Several studies addressing V gene usage demonstrate an unbiased V_H repertoire, with a frequency of expression of particular V_H gene family members closely similar to that of the normal B-cell repertoire.[364-370] FL is discussed further in Chapter 3.

1.4.3.5. Burkitt's lymphoma

Burkitt's lymphoma (BL) is characterised by a subset of $CD5^-CD19^+CD20^+CD10^+CD23^-$ cells that typically express monotypic surface IgM (or IgG or IgA) and are thought to derive from GC B cells.[286] There are two major subgroups: endemic BL (eBL), which mainly affects African children, and sporadic BL (sBL), which affects young adults worldwide. BL is a common neoplasm of HIV-infected patients and as such these lymphomas are currently classified as AIDS-related BL.[286] The two main forms of BL differ in EBV-infection status; eBL is almost always associated with EBV infection, whereas only ~30% of sBL are positive for EBV. EBV involvement in AIDS-related BL is variable. Studies of Ig V genes in BL have demonstrated the presence of somatic mutations in the V genes of sBL, eBL and AIDS-related BL.[372-378] Moreover, the majority report that eBL and AIDS-related BL display a significantly higher mutational frequency compared to sBL.[372-375,378] The process of ongoing SHM in BL remains controversial with some studies demonstrating intraclonal heterogeneity.[373,375,376,379] The preferential usage of V_H4 family genes has been suggested in BL, although this has not been confirmed in other studies where a V_H and Vκ usage similar to the normal B-cell repertoire is described.[373,375]

According to the mutational patterns observed it has been suggested that there are two differentiation stages to which BL may correspond: sBL likely represent early centroblasts of the GC, while eBL and AIDS-related BL may represent late GC B cells or post-GC memory B cells. It is known that memory B cells provide a reservoir for latent EBV

infection and therefore, a memory B-cell origin for eBL seems conceivable.[380] The normal counterpart of the neoplastic cells occurring in BL remains controversial.

1.4.3.6. Hodgkin's disease

The WHO classification subdivides Hodgkin's disease (HD) into two groups: nodular lymphocyte predominant HD (NLPHD) and classical HD (CHD), of which the latter consists of the three subgroups nodular sclerosing, mixed cellularity and lymphocyte depleted. NLPHD is characterised by a subset of $CD19^+CD20^+CD10^-CD30^-CD15^-CD45^+$ lymphocytic and histiocytic (L-and-H) cells that express surface Ig.[286] CHD is characterised by a subset of $CD20^{+/-}CD10^-CD30^+CD15^+CD45^-$ Hodgkin's Reed-Sternberg (HRS) cells that are unusual in that they do not express surface Ig.[286] Since the neoplastic cells of HD represent a small minority (~1%) amidst a reactive, inflammatory cell population, studies investigating the Ig V genes of HD have used single cell PCR. Sequence analysis of the clonally rearranged V_H and V_L genes in HD L-and-H and HRS cells reveals that they are both mutated, consistent with transition through the GC.[381-387] However, while the Ig V_H gene rearrangements of L-and-H cells are productive, up to 25% of V_H gene rearrangements of HRS cells are non-productive due to the presence of mutations that lead to the introduction of stop codons.[381,382,386,387] Furthermore, it has been reported that defects in transcriptional machinery render HRS cells unable to transcribe rearranged Ig genes, which likely accounts for the lack of Ig expression at both the protein and mRNA level.[388,389] Significant intraclonal heterogeneity, attributed to ongoing SHM, is observed in the V_H genes of L-and-H cells, which has lead to the suggestion that the L-and-H cells of NLPHD are derived from GC B cells at the centroblastic stage.[383-385] In contrast, there is no evidence of continuing SHM in the V_H genes of HRS cells.[381] Nonetheless, it is proposed that the HRS cells of CHD also derive from GC B cells, but in this case the GC B cells are 'crippled' by mutations and have lost the capacity to express surface Ig.[382,386,387] Since a lack of surface Ig expression on GC B cells would normally cause the death of the B cell by apoptosis, it was surmised that the HRS cells must reside within the GC and have been rescued from apoptosis by a transforming event.

1.4.3.7. Diffuse large B cell lymphoma

Diffuse large B cell lymphoma (DLBCL) represents a heterogeneous group of tumours based on significant variations in natural history, morphology, clinical presentation and response to treatment and, for this reason, is likely to constitute more than one disease entity.[286] DLBCL can develop at secondary lymphoid sites (nodal) or at extranodal sites, such as the liver, thyroid, bone, gastrointestinal tract and central nervous system (CNS); in

fact, several variants are further defined based on their site of clinical presentation and include primary mediastinal, intravascular and primary effusion lymphomas.[286] The marked heterogeneity of DLBCL is reflected in the variable expression of numerous B-cell markers; cells typically express CD19 and CD20, but the expression of CD5 (<10% of cells), CD10 (30-40%) and CD23 (~20%) are variable.[390] DLBCL typically express monotypic surface IgM (or IgG or IgA), although a lack of sIg expression is frequent in mediastinal DLBCL, the significance of which is unknown. The study of Ig V genes in DLBCL has revealed some recurring themes: the clonal V_H genes are commonly mutated, suggesting an origin from GC or post-GC B cells.[367,371,391-394] However, the issue of ongoing SHM is controversial, with some studies demonstrating no intraclonal heterogeneity and others showing ongoing mutation, although the latter seems to predominate.[367,371,391-394] Similar inconsistencies are apparent in V_H gene usage; some studies report a V4-34 gene bias, whereas others report a V_H usage similar to normal B cells.[367,394-398] Alternative isotype expression can be detected at both the mRNA and protein level, for which the pattern of mutation suggests switching is occurring in subpopulations of the tumour clone after transformation.[371,399]

Heterogeneity within DLBCL has led several groups to analyse the V genes of specific subsets; of special interest are three distinct entities, primary CNS lymphoma (PCNSL), *de novo* $CD5^+$ DLBCL and primary mediastinal B-cell lymphoma. For example, the V_H genes of PCNSL show extensive somatic mutation with intraclonal heterogeneity and a biased usage of the V4-34 gene segment in up to 50% of cases.[400-402] PCNSL is discussed further in Chapter 5. However, more recently, the study of DLBCL has been revolutionised by the introduction of gene expression profiling using DNA microarrays. Initial studies suggested that DLBCL encompasses at least two molecularly separate entities, each with distinct derivation: the first subtype expressed genes characteristic of GC B cells, termed GC B cell-like (GCB) DLBCL, while the second expressed genes normally induced during *in vitro* activation of peripheral blood B cells, termed activated B cell-like (ABC) DLBCL.[403] Importantly, patients with a GCB phenotype had a significantly better overall survival than those with ABC DLBCL. In a parallel study, it was demonstrated that ongoing SHM is a feature of GCB DLBCL, but not ABC DLBCL.[404] This supports the hypothesis that the initial transforming event leading to GCB DLBCL occurs within a GC microenvironment, whereas ABC DLBCL must derive from a post-GC B cell. This classification was conclusively demonstrated in a second study, which extended the original observations to include a third group, termed Type 3 DLBCL, for which neither a GCB or ABC signature could be applied.[405] In a concurrent study, gene expression

profiling was used to classify two DLBCL subgroups based on overall survival rates.[406] Latterly, it has been demonstrated that immunohistochemical analysis of the expression of CD10, Bcl-6 and MUM-1 using a tissue microarray can be used to determine GC and non-GC subtypes in DLBCL.[407] DLBCL is discussed further in Chapter 4.

1.4.3.8. Marginal zone lymphoma (MALT lymphoma and SMZL)

Marginal zone lymphoma (MZL) is characterised by a subset of $CD5^-CD19^+CD20^+CD10^-CD23^-$ cells that typically express monotypic surface IgM (or IgG or IgA) and are thought to derive from marginal zone B cells.[286] MZLs are subclassified into three distinct groups reflecting the site of primary tumour origin: extranodal MALT lymphomas, nodal MZL, which are very rare, and splenic marginal zone lymphomas (SMZL).

Extranodal MALT lymphomas can arise in a variety of organs, such as the gastrointestinal tract, salivary glands, lungs, thyroid gland, breast, eye and skin, and commonly arise following organ-specific infection or autoimmunity.[408-410] For example, gastric MALT lymphomas typically develop from chronic gastritis, the vast majority (>90%) of which are associated with *Helicobacter pylori* infection. Salivary and thymic MALT lymphomas are associated with Sjogren's syndrome, while thyroid MALT lymphomas are associated with Hashimotos thyroiditis. Studies investigating the Ig V_H genes of MALT lymphomas demonstrate that the majority are mutated and have frequently undergone isotype switching, consistent with a derivation from post-GC marginal zone memory B cells.[411-413] However, evidence of ongoing SHM in both gastric and salivary MALT lymphomas contradicts this view and suggests that the SHM mechanism is still active, indicative of a GC B cell origin. This finding may reflect re-entry into the permissive GC (follicular colonisation) or low level SHM occurring outside the GC, *in situ*, both of which have been reported for normal marginal zone B cells.[414-417] The biased usage of individual V_H genes, including V1-69, V1-18, V3-07 and V3-30.3 and V3-23, has been described for MALT lymphomas arising in the stomach, lungs, salivary glands and thymus, consistent with the concept that an Ag or auto-Ag may drive the lymphoproliferative process.[415,418-420] For example, the V1-69 gene segment is used in up to 50% of salivary MALT lymphomas and is also frequently observed in gastric MALT lymphomas, while thymic MALT lymphomas express highly restricted V3-23 and V3-30.3 gene segments. Furthermore, a biased usage of D and J_H gene segments has been reported for salivary MALT lymphoma and CDR3 analysis has revealed high sequence homology among V1-69 cases, further suggesting a shared Ag specificity.[418]

It has been suggested that thymic MALT lymphomas represent a distinct, homogeneous subgroup of MALT lymphomas. These tumours are unusual in that they display both homogeneously mutated and unmutated Ig V_H genes at an approximately equal frequency.[420] The unmutated subset may derive from naïve B cells, which are also found to populate the marginal zone, or may represent B cells that have traversed the GC and escaped the SHM mechanism. Interestingly, thymic MALT lymphomas almost always express an IgA phenotype irrespective of mutational status suggestive of passage through the GC.[420,421] However, ectopic CSR has been demonstrated in the marginal zone.[422]

SMZL differs phenotypically from other types of MZL by frequently co-expressing IgM and IgD.[423] Initial Ig V_H gene studies on limited numbers of SMZL reported the expression of mutated V_H genes consistent with a post-GC memory B-cell origin.[424-427] However, subsequent studies have found both unmutated and mutated V genes at an approximately equal frequency; the unmutated subset typically express IgD, while the mutated subset do not.[428,429] These findings suggest that SMZL arises from both post-GC memory B cells that have mutated V genes and lack readily detectable IgD as well as naïve B cells that have unmutated V genes and express IgD. The normal splenic marginal zone compartment appears to consist of both naïve and memory B cells, consistent with SMZL originating from different types of splenic marginal zone B cells.[430] Demonstration of ongoing SHM in the mutated subset suggests that the GC may still influence SMZL, as is described above for MALT lymphomas. The biased usage of V1-69, V4-34 and V1-02 gene segments has been described for both unmutated and mutated SMZL.[429]

1.4.3.9. Hairy cell leukaemia

Hairy cell leukaemia (HCL) is characterised by a subset of $CD5^-CD19^+CD20^+CD10^-CD23^-$ cells that typically express surface IgM±IgD (or IgG or IgA), with a high specificity for CD103 and CD11c.[431] According to the WHO classification, the postulated cell of origin is a post-GC B cell, although the exact normal counterpart is unknown.[286] Ig V gene studies reveal that the majority of V_H genes are mutated, consistent with derivation from a post-GC B cell.[432-436] This, together with evidence from expression profiling, led to the suggestion that HCL originates from a Ag-experienced memory B cell, possibly outside of the GC in the marginal zone.[434,437] Although HCL cells do not express CD27, a marker for memory B cells, recent observations suggest that a population of $CD27^-$ memory B cells exist in healthy volunteers (S. Sahota, unpublished observation 2007) and, therefore, that this cell may represent the normal counterpart.[433] Similar to MALT lymphomas, which also reside in the marginal zone compartment, low level ongoing SHM has been reported

in HCL, indicative of a GC origin.[432,433,435] However, as described above (section 1.4.3.8.), this may be explained by ectopic SHM, which is known to take place in normal marginal zone B cells.[417] A biased usage of individual V_H3 family gene segments has been described in HCL and includes V3-23, V3-30, V3-33 and V3-07.[433-436] An unusual feature of HCL is the expression of multiple clonally related isotypes by individual cells of the tumour population, which are presumably generated via the processing of long primary RNA transcripts rather than by deletional recombination, as is the usual mechanism of CSR.[433]

1.4.3.10. Lymphoplasmacytoid lymphoma/ Waldenstrom's macroglobulinaemia

Lymphoplasmacytoid lymphoma (LPL) is characterised by a subset of $CD5^-$ $CD19^+CD20^+CD10^-CD23^-$ cells that express high levels of surface IgM±IgD and are thought to derive from mature, post-GC B cells.[286] LPL is frequently associated with a monoclonal serum paraprotein, typically of IgM type. Waldenstrom's macroglobulinaemia (MW) is a specific clinical syndrome commonly associated with LPL displaying IgM monoclonal gammopathy. Ig V gene analysis has demonstrated, in the main, homogeneously mutated V_H and Vκ genes in WM, consistent with a derivation from post-GC B cells; a small number of unmutated WM have also been described.[438-443] Several studies have suggested a V_H bias towards the use of the V_H3 family gene segments, although this remains ambiguous and greater number of cases required to confidently assess V_H gene bias.[439,440,442,443] Early studies revealed a lack of isotype switch transcripts, suggesting that the transforming event takes place before the onset of CSR, but after SHM has been silenced.[440,441,443] Therefore, it had been proposed that WM is derived from a cell arrested at the stage of GC B cell to plasma cell differentiation. However, more recently, isotype switched transcripts have been demonstrated, together with AID transcripts, indicating that CSR can occur in WM at low levels and likely occurs as a post-transformation event.[444]

1.4.3.11. Multiple myeloma and MGUS

Multiple myeloma (MM) is characterised by a subset of $CD19^-CD20^-CD10^-CD23^-$ $CD45^{\pm}CD38^+CD138^+$ cells that infiltrate the BM and are thought to derive from mature, post-GC B cells.[286] MM do not express surface Ig, but secrete high levels of monoclonal Ig paraprotein of one particular isotype (clinical isotype), usually IgG or IgA, into the serum of patients. Studies investigating the Ig V genes of MM have revealed that they have undergone isotype switch and are homogeneously mutated, consistent with a post-GC B-cell origin.[445-450] With the exception of an under-representation of the V4-34 gene, there

appears to be no bias in V_H or V_L gene segment usage, which parallels that of normal B cells.[447,450,451] These findings have led to the suggestion that MM originates from a memory B cell, which, following neoplastic transformation, homes to the BM and proliferates and differentiates into plasma cells secreting high levels of Ig. CSR is thought to be an important stage in the development of MM. For rare IgM-expressing MM, it is probable that transformation occurs prior to isotype switch. Interestingly, alternative, but clonally related, Cγ transcripts have been detected in a rare IgM-expressing MM, suggesting that these cells were in the process of CSR when transformation occurred.[452] Therefore, transformation in MM may occur during isotype switch events in a IgM^+ memory B cell, with arrest possible on either side of the switch point. This is supported by reports which describe MM cells expressing other isotypes in addition to the clinical one, either pre-switched IgM or IgD or post-switched IgA or IgG; of particular interest are clonal VDJ-Cμ transcripts in isotype switched IgG^+ and IgA^+ MM.[452-456] Of note, no further SHM accumulated from pre- to post-switch, with strict intraclonal homogeneity observed. The role of the clonotypic cells that express nonclinical isotypes in malignant disease remains uncharacterised.

Monoclonal gammopathy of undetermined significance (MGUS) is an asymptomatic, pre-malignant disorder characterised by plasma cell infiltration of less than 10% of the BM mononuclear cells and the production of low, stable levels of paraprotein in the serum, usually IgG or IgA.[457] Around 1% of MGUS cases a year will become overt MM, although the prediction of which patients will undergo transformation to MM is not currently feasible. Studies investigating the Ig V genes in MGUS have shown that, like MM, the V_H genes are mutated and have undergone isotype switching events.[458,459] Similarly, pre-switched IgM-expressing precursor cells have been identified that share the clonal VDJ of the isotype-switched population.[458] However, in contrast, ongoing SHM is frequently observed in the V_H genes of both the IgM precursor and isotype switched cells, suggesting that MGUS derives from a cell at an earlier stage of differentiation than that of MM, such as GC B cells; this is consistent with their expression of sIg. Studies investigating the pathway of transformation from MGUS to MM have revealed that the same tumour clone can evolve to the malignant status.[458,459] It was also demonstrated that upon transformation, the intraclonal heterogeneity evident at the MGUS stage is lost, with apparent homogeneity in the emergent clone.

2. Materials and Methods

2.1. Patient selection

Unless otherwise stated, all tumour specimens were collected from patients attending Southampton General Hospital; informed consent was provided by each participant on study following ethical approval from the IRB.

2.1.1. Follicular lymphoma

Tumour infiltrated lymph nodes were obtained by excision biopsy from 44 patients with untreated FL during standard diagnostic procedures and/or for eligibility into a clinical trial. In all cases, the diagnosis of FL was made using histological, immunophenotypic and clinical data according to the WHO classification.[286]

2.1.2. Diffuse large B-cell lymphoma

Tumour specimens were obtained by excision biopsy from 33 patients with untreated *de novo* DLBCL during standard diagnostic procedures; 14 of these were kindly provided by Prof. S.A. Pileri, Institute of Haematology and Medical Oncology, University of Bologna, Italy. RNA was kindly provided for a further 3 *de novo* DLBCL by Prof. I. Lossos, Molecular and Cellular Pharmacology, University of Miami, USA. In all cases, the diagnosis of *de novo* DLBCL was made using histological, immunophenotypic and clinical data according to the WHO classification.[286]

2.1.3. Primary CNS lymphoma

Tumour specimens were obtained by stereotactic biopsy from 12 HIV-negative patients with untreated PCNSL. In all patients, there was no obvious immunodeficiency at the time of diagnosis. All PCNSL cases were classified as DLBCL using histological, immunophenotypic and clinical data according to the WHO classification.[286]

2.1.4. V gene sequence cohorts

Published V_H gene sequences or sequence summary data thereof were extracted from the cited references: FL- Bahler *et al.*, 1991[364], Hsu and Levy, 1995[367], Stamatopoulos *et al.*, 1997[368], Noppe *et al.*, 1999 (EMBL, GenBank DDBJ nucleotide sequence databases under accession number AJ234156-AJ234298)[369], Aarts *et al.*, 2000[370], Zabalegui *et al.*, 2004[460]; PCNSL- Thompsett *et al.*, 1999[400], Montestinos-Rongan *et al.*, 1999[401], Sekita *et al.*,

2001[402], Montestinos-Rongan *et al.*, 2005 (accession numbers AF168818 and AY598953-AY598962)[461].

Where deposited, sequence data was downloaded from the National Centre for Biotechnology Information at:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=PubMed

2.2. Patient Material

2.2.1. Processing of fresh biopsies

Processing of fresh biopsy material (44 FL and 19 DLBCL) was performed either by myself or by tumour bank staff at the Cancer Sciences Division, University of Southampton.

2.2.1.1. Preparing cell suspensions

Diagnostic biopsies were received fresh on the day of surgery and processed on the same day under sterile conditions. Biopsy tissue was cut into small pieces using a sterile blade and the cells dispersed through a fine sieve into sterile RPMI 1640 medium (Gibco, Paisley, UK). The cells were collected, centrifuged at 1200rpm, washed once and resuspended in an appropriate volume of RPMI.

2.2.1.2. Cell count

Cell viability was assessed with 0.4% trypan blue stain (Sigma-Aldrich Company Ltd, Gillingham, UK) using a haemocytometer.

2.2.1.3. Freezing and storage

Aliquots of $1x10^7$ cells/mL were resuspended in freezing medium consisting of 50% batch tested decomplemented human AB serum (Sigma-Aldrich Company Ltd, Gillingham, UK), 40% complete RPMI and 10% dimethylsulphoxide (DMSO). Freezing followed a step-wise drop in temperature to –80°C using standard cryopreservation protocols. Cells were transferred to Liquid N_2 after 24 hours.

2.2.1.4. Immunophenotyping

Where available, an aliquot of cells was used to determine the surface immunoglobulin isotype (IgM, IgD, IgG, IgA, Igκ and Igλ) in conjunction with other B-cell markers by fluorescence-activated cell sorting (FACS). All primary and secondary B-cell screens

were performed by staff at the Immunology Department, Southampton University Hospitals NHS trust.

2.2.2. Cutting of frozen biopsies

Where only archived frozen tissue blocks were available (12 PCNSL and 14 DLBCL), 10μm sections were cut on a cryostat using a disposable sterile blade and transferred directly into 1mL of TRI reagent (Sigma-Aldrich Company Ltd, Gillingham, UK) for subsequent RNA extraction (see section 2.3.1.). The number of sections varied according to the size of the tissue block, but was in the range of 5 to 20. All PCNSL biopsies were kindly sectioned by Mrs Jean Buontempo, at the Cellular Pathology Department, Southampton University Hospitals NHS trust.

2.2.3. Processing of blood

An anti-coagulated blood sample was received fresh and processed on the same day under sterile conditions. Mononuclear cells were separated by centrifugation on lymphoprep™ (Axis-Shield PoC AS, Oslo, Norway). Briefly, 10mL of blood was layered onto 10mL of lymphoprep™ and centrifuged at 2100rpm for 20 min at room temperature (RT). Mononuclear cells were collected from the interphase, washed once and resuspended in an appropriate volume of RPMI. Viable cells were assessed, counted and frozen down in aliquots of $1x10^7$ cells/mL as described in sections 2.2.1.2. and 2.2.1.3.

2.2.4. Processing of bone marrow

An anti-coagulated BM aspirate was received fresh on the day of harvest and processed on the same day under sterile conditions. Mononuclear cells were separated by centrifugation on lymphoprep™. Briefly, the sample was diluted to 15mL in sterile RPMI, layered onto 20mL of lymphoprep™ and centrifuged at 2100rpm for 20 min at RT. Mononuclear cells were collected from the interphase, washed once and resuspended in an appropriate volume of RPMI. Viable cells were assessed, counted and frozen down in aliquots of $1x10^7$ cells/mL as described in sections 2.2.1.2. and 2.2.1.3.

2.3. Preparation of cDNA

2.3.1. RNA extraction

Total cellular RNA was extracted from either $5x10^6$ cells or 5-20 10μm cut sections of frozen tissue using 1mL of TRI reagent (Sigma-Aldrich Company Ltd, Gillingham, UK) and following the manufacturers instructions in a triphasic separation. Briefly, 0.2mL of chloroform was added per mL of TRI reagent and mixed vigorously. Incubation at RT for

10 min was followed by centrifugation at 10,000rpm for 15 min at 4°C. The upper aqueous phase was removed and added to 0.5mL isopropanol. Incubation at RT for 10 min was followed by centrifugation at 10,000rpm for 15 min at 4°C. The RNA pellet was washed in 75% ethanol, briefly allowed to air dry and resuspended in an appropriate volume of RNase-free sterile water (approx. 40-60μL).

2.3.1.1. Quantification and storage of RNA

A 1 in 200 dilution of RNA in distilled water was measured for absorbance at 260nm using an Eppendorf BioPhotometer (Eppendorf AG, Hamburg, Germany). To ensure that the final RNA preparation was free of protein a 260/280nm absorbance ratio of greater than 1.7 was desired. RNAs were stored at –80°C.

2.3.2. cDNA synthesis

Two different cDNA synthesis protocols were used during this study depending on the availability of RNA. Negative controls, in which RNA was absent from the reaction mix, were run in parallel. Subsequent polymerase chain reaction (PCR) of these "blanks" using primers designed to amplify the housekeeping gene β actin (see section 2.4.1.) ensured that any contamination could be identified.

2.3.2.1. Standard cDNA synthesis

An aliquot of 5μg of RNA in a final volume of 20μL was reverse transcribed using the Superscript™ First-Strand cDNA Synthesis System II (Invitrogen Ltd, Paisley, UK) with an oligo-dT primer according to the manufacturers instructions. Briefly, 5μg of RNA was incubated with 0.5μg oligo-$dT_{12\text{-}18}$ and 500μM dNTP for 5 min at 65°C. After cooling on wet ice, the reaction mixture was added to a final volume of 20μL containing 50 Units Superscript™ II reverse transcriptase, 10mM dithiothreitol (DTT), 1X first strand buffer (20mM Tris-HCl, pH 8.4; 50mM KCl), 5mM $MgCl_2$ and 40 Units of RNaseOUT™ recombinant RNase inhibitor. The reaction was performed at 42°C for 50 min and followed by heat inactivation of enzyme at 70°C for 15 min. Finally, the reaction was incubated with 2 units of *Escherichia coli* (*E.coli*) RNase H at 37°C for 20 min to remove RNA. cDNA was stored at –80°C.

2.3.2.2. PolyA cDNA synthesis

An aliquot of 0.5-2μg of RNA in a total volume of 10μL was reverse transcribed using Superscript™ II reverse transcriptase. Briefly, 0.5-2μg of RNA was incubated with 0.5μg

oligo-dT_{12-18} at 70°C for 1 min. After cooling on ice, the reaction mixture was added to a final volume of 10μL containing 50 Units Superscript™ II reverse transcriptase, 500μM dNTP, 10mM DTT and 1X first strand buffer (20mM Tris-HCl, pH 8.4; 50mM KCl). The reaction was incubated at 37°C for 30 min followed by heat inactivation of enzyme at 65°C for 10 min.

For the addition of the 5' priming sites, a terminal transferase reaction was performed by adding 15 Units terminal deoxynucleotide transferase (Invitrogen Ltd, Paisley, UK), 1X tailing buffer (100nM potassium cacodylate, pH 7.2; 2mM $CoCl_2$; 200μM DTT) and 1mM dATP to a final volume of 20μL. The reaction was incubated at 37°C for 15 min followed by heat inactivation of enzyme at 65°C for 10 min.

Following return of the reactions to wet ice, a non-specific PCR was performed in a total volume of 100μL. It contained 5 Units HotStarTaq™ DNA polymerase (QIAgen, Crawley, UK), 1.5μg oligo-dT_{12-18}, 1X buffer (contains Tris-Cl; KCl; $(NH_4)_2SO_4$; 1.5mM $MgCl_2$, pH8.7) and 250μM dNTP. Cycling was performed on a GeneAmp® PCR System 2700 (Applied Biosystems, Warrington, UK). After an initial denaturing step at 94°C for 15 min, the first round of 25 cycles consisted of denaturation at 94°C for 1 min, primer annealing at 42°C for 2 min and extension at 72°C for 4 min. Following the addition of 12.5U HotStarTaq™ DNA polymerase, the second round of 25 cycles consisted of denaturation at 94°C for 1 min, primer annealing at 42°C for 2 min and extension at 72°C for 2 min. The non-specific PCR was repeated as often as necessary using >10μL of the tailed cDNA product to yield a further 100μL of cDNA for specific PCR. cDNA was stored at –80°C.

2.4. Polymerase chain reaction

Two μL of cDNA was used to amplify the V genes with HotStarTaq® (QIAgen, Crawley, UK) and 5' and 3' oligonucleotide primers designed specifically (Table 2.1.). PCR reactions were in a total volume of 50μL and contained 2.5 Units of HotStarTaq™ DNA polymerase, 250μM dNTP's, 20pM of each primer and 1X buffer (containing Tris-Cl; KCl; $(NH_4)_2SO_4$; 1.5mM $MgCl_2$, pH8.7). All PCR reactions were set up in a laminar flow cabinet to minimise the risk of contamination. Negative controls, in which the cDNA was absent from the reaction mix, were run in parallel for every primer pair in order that contamination of any component of the reaction mix with cDNA or DNA could be identified.

Table 2.1. Primers for V_H and V_L PCR amplification

Primer Name	Orientation	Sequence
V_H primers		
V_H Leader 1/7	Sense	CTC ACC ATG GAC TGG ACC TGG AG
V_H Leader 2	Sense	ATG GAC ATA CTT TGT TCC AGG CTC
V_H Leader 3	Sense	CCA TGG AGT TTG GGC TGA GCT GG
V_H Leader 4	Sense	ACA TGA AAC AYC TGT GGT TCT TCC
V_H Leader 5	Sense	AGT GGG TCA ACC GCC ATC CTC G
V_H Leader 6	Sense	ATG TCT GTC TCC TTC CTC ATC TTC
V_H 1/7 FR1 cons	Sense	CAG GTG CAG CTG GTG CAR YCT G
V_H 2 FR1 cons	Sense	CAG RTC ACC TTG AAG GAG TCT G
V_H 3 FR1 cons	Sense	GAG GTG CAG CTG GTG SAG TCY G
V_H 4a FR1 cons	Sense	CAG STG CAG CTG CAG GAG TCS G
V_H 4b FR1 cons	Sense	CAG GTG CAG CTA CAR CAG TGG G
V_H 5 FR1 cons	Sense	GAG GTG CAG CTG KTG CAG TCT G
V_H 6 FR1 cons	Sense	CAG GTA CAG CTG CAG CAG TCA G
FR2 b cons	Sense	TGG RTC CGM CAG SCY YCN GG
FR2 a cons	Sense	GTC CTG CAG GCY YCC GGR AAR RGT CTG GAG TGG
FR3 cons	Sense	ACA CGG CYG TRT ATT ACT GT
Constant μ	Anti-sense	GGA ATT CTC ACA GGA GAC GAG G
Constant δ	Anti-sense	CAC AGG GCT GTT ATC CTT TGG
Constant γ	Anti-sense	CTG AGT TCC ACG ACA CCG TCA
Constant α	Anti-sense	ATC TGG CTG GGT GCT GCA GAG GCT
J_Ha cons	Anti-sense	ACC TGA GGA GAC GGT GAC C
J_Hb cons	Anti-sense	GTG ACC AGG GTN CCT TGG CCC CAG
J_Hc cons	Anti-sense	TGA GGA GAC GGT GAC CAG GAT CCC TTG GCC CCA G
J_H1	Anti-sense	CAG GGC ACC CTG GTC ACC GTC TCC TCA
J_H2	Anti-sense	CGT GGC ACC CTG GTC ACT GTC TCC TCA
J_H3	Anti-sense	CAA GGG ACA ATG GTC ACC GTC TCT TCA
J_H4/5	Anti-sense	CAG GGA ACC CTG GTC ACC GTC TCC TCA
J_H6	Anti-sense	CAA GGG ACC ACG GTC ACC GTC TCC TCA
Vκ primers		
Vκ Leader 1 cons	Sense	ATR GAC ATG AGR GTS CYY GCT CAG CKC
Vκ Leader 2 cons	Sense	ATG AGG CTC CYT GCT CAG CTY CTG GGG
Vκ Leader 3 cons	Sense	ATG GAA ACC CCA GCG CAG CTT CTC TTC
Vκ Leader 4	Sense	ATG GTG TTG CAG ACC CAG GTC TTC ATT
Vκ Leader 5	Sense	ATG GGG TCC CAG GTT CAC CTC CTC AGC
Vκ Leader 6a	Sense	ATG TTG CCA TCA CAA CTC ATT GGG TTT
Vκ Leader 6b	Sense	ATG GTG TCC CCG TTG CAA TTC CTG CGG
Constant κ 27	Anti- sense	CAA CTG CTC ATC AGA TGG CGG GAA
Constant κ 69	Anti-sense	AGT TAT TCA GCA GGC ACA CAA C
Vλ primers		
Vλ Leader 1 cons	Sense	ATG RCC DGS TYY CCT CTC YTC CTC
Vλ Leader 2 cons	Sense	ATG GCC TGG GCT CTG CTS CTC CTC
Vλ Leader 3 cons	Sense	ATG GCM TGG RYC VYW CTM YKB CTS
Vλ Leader 4a	Sense	ATG GCC TGG ACC CAA CTC CTC CTC
Vλ Leader 4b	Sense	ATG GCT TGG ACC CCA CTC CTC CTC
Vλ Leader 4c	Sense	ATG GCC TGG GTC TCC TTC TAC

(Continued)

(continued)

Primer Name	Orientation	Sequence
Vλ Leader 5 cons	Sense	AGT GCC TGG ACT CYT CTY CTY CTC
Vλ Leader 6	Sense	ATG GCC TGG GCT CCA CTA CTT CTC
Vλ Leader 7	Sense	ATG GCC TGG ACT CCT CTC TTT CTG
Vλ Leader 8	Sense	ATG GCC TGG ATG ATG CTT CTC CTC
Vλ Leader 9	Sense	ATG GCC TGG GCT CCT CTG CTC CTG
Vλ Leader 10	Sense	ATG CCC TGG GCT CTG CTC CTC CTG
Constant λ 33	Anti-sense	GTT GGC TTG AAG CTC CTC AGA GGA
Constant λ 85	Anti-sense	CAC RGC TCC CGG GTA GAA GTC ACT

R= A+G; Y= C +T; K= G+T; S= C+G; N=A+C+G+T ,M=A+C, W=A+T
Where primers have greater than one consensus sequence, for example, J_H a, b and c, an equal mix is prepared.

Table 2.2. Primers for β actin PCR amplification

Primer Name	Orientation	Sequence
β actin Forward	Sense	TCA TGT TTG AGA CCT TCA A
β actin Reverse	Anti-sense	GTC TTT GCG GAT GTC CAC G

2.4.1. Assessment of cDNA quality

The integrity of the cDNA was confirmed prior to V gene rescue by PCR using primers designed to amplify the standard housekeeping gene β actin. The PCR reaction mix was as above (see section 2.4.) and used the β actin forward and reverse primers (Table 2.2.). Cycling conditions were identical to those described for the heavy chain PCR (see section 2.4.2.2.).

2.4.2. Identification of tumour-derived V_H by PCR

2.4.2.1. Primer selection

Preliminary primer selection was influenced by the availability of immunophenotypic information. For a known heavy chain isotype, e.g. IgA, multiple PCRs were performed using a specific 3' primer that hybridised the constant region (Cα) in conjunction with a variety of 5' primers; the initial round of PCRs used individual V_H leader primers in 6 separate reactions. If PCR using V_H leader primers produced only polyclonal sequences, primers designed to anneal to the FR regions, where sequence is inherently more conserved, were used; a second round of PCRs used individual FR1 primers in 6 separate reactions. For cases where the isotype was not known, a 5' FR2 region consensus primer was used in combination with all available constant region primers (Cμ, Cδ, Cγ and Cα) and a J_H consensus primer mix in 5 separate PCR reactions. This approach often led to the identification of the tumour-derived V_H (partial sequence only: FR2-FR4). To determine the full sequence, subsequent PCR used the appropriate constant region primer and V_H leader primer, as identified from sequence information. Multiple PCR reactions using all possible primer combinations were performed until the tumour-derived V_H was identified or primer combinations were exhausted. For a complete list of the primer sequences refer to Table 2.1.[462]

2.4.2.2. Cycling conditions

Cycling for V_H PCR, with the exception of the amplification of the FR3 to constant region, followed a "touchdown" protocol and was performed on a GeneAmp® PCR System 2700 (Applied Biosystems, Warrington, UK). After an initial denaturing step at 94°C for 15 min, there followed a single cycle of denaturation at 94°C for 30 sec, primer annealing at 60°C for 45 sec and extension at 72°C for 45 sec. With the denaturing and extension conditions the same, the primer annealing temperature was then reduced from 60°C by 1°C per cycle to 55°C in a further 5 cycles. Thirty cycles of denaturation at 94°C for 30 sec, primer annealing at 54°C for 45 sec and extension at 72°C for 45 sec followed with a final extension step at 72°C for 10 min and hold at 4°C.

FR3 to constant region amplifications used standard cycling conditions; after an initial denaturing step at 94°C for 15 min, 55 cycles of denaturation at 94°C for 10 sec, primer annealing at 57°C for 10 sec and extension at 72°C for 25 sec followed, with a final extension step at 72°C for 10 min and hold at 4°C.

2.4.3. Identification of tumour-derived V_L by PCR

2.4.3.1. Primer selection

Preliminary primer selection was influenced by knowledge of the light chain restriction of the tumour clone, as revealed by immunophenotyping. For a known light chain restriction, PCR was performed in duplicate using a specific 3' primer that hybridised the constant region of interest (Cκ or Cλ) in conjunction with the appropriate 5' V_L leader primer mix (Vκ1-6 or Vλ1-10). If PCR using V_L leader mix produced only polyclonal sequences, a second round of PCRs using individual V_L primers in 6(/10) separate reactions was performed. For cases where the light chain restriction was not known, both kappa and lambda PCR amplifications were performed; the initial round of PCRs used the 5' V_L leader mix, which was followed by PCR using individual V_L primers if necessary. Multiple PCR reactions using all possible primer combinations were performed until the tumour-derived V_H was identified or primer combinations were exhausted. For a complete list of the primer sequences refer to Table 2.1.[462]

2.4.3.2. Cycling conditions

Cycling for V_L PCR followed a "touchdown" protocol and was performed on a GeneAmp® PCR System 2700 (Applied Biosystems, Warrington, UK). After an initial denaturing step at 94°C for 15 min, there followed a single cycle of denaturation at 94°C for 30 sec, primer annealing at 56°C for 45 sec and extension at 72°C for 45 sec. With the denaturing and extension conditions the same, the primer annealing temperature was then reduced from 56°C by 1°C per cycle to 51°C in a further 5 cycles. Thirty cycles of denaturation at 94°C for 30 sec, primer annealing at 50°C for 45 sec and extension at 72°C for 45 sec followed with a final extension step at 72°C for 10 min and hold at 4°C.

2.4.4. Agarose gel electrophoresis

PCR amplification products were separated by 1% agarose gel electrophoresis in 1X TAE buffer against $1Kb^+$ DNA ladder (Invitrogen Ltd, Paisley, UK) and visualised with ethidium bromide under UV light. Products of the appropriate size were excised using a clean scalpel blade.

2.4.5. PCR product purification

PCR products were purified using the QIAquick® Gel Extraction Kit (QIAgen, Crawley, UK) according to the manufacturers instructions. Briefly, the gel slice was solubilised in 3 volumes of QG buffer at 50°C for 5 min and applied to a QIAquick® spin column. The DNA was incubated on the column for 1 min, washed with PE buffer and eluted into an appropriate volume (>20μL) of elution buffer (10mM Tris-Cl, pH 8.5). PCR products were stored at –20°C.

2.5. Sequencing

2.5.1. Sequencing reaction

Direct sequencing of PCR products used the ABI Prism BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied biosystems, Warrington, UK). Reactions were in a total volume of 10μL and consisted of 5μL DNA template, 2μL BigDye® Terminator Ready Reaction Mix (A-dye terminator labelled with dichloro [R6G], C-dye terminator labelled with dichloro [ROX], G-dye terminator labelled with dichloro [R110], T-dye terminator labelled with dichloro [TAMRA], dNTP [dATP, dCTP, dITP, dUTP], AmpliTaq DNA polymerase), 1.6pM appropriate oligonucleotide primer (5' primer used in the PCR amplification) and 1X sequencing buffer (80mM Tris-HCl, pH 9.0; 2mM $MgCl_2$). Cycling consisted of 25 cycles of denaturation at 94°C for 10 sec, primer annealing at 50°C for 50 sec and extension at 60°C for 4 min and was performed on a GeneAmp® PCR System 2700 version 2.04 (Applied Biosystems, Warrington, UK).

2.5.2. Precipitation of sequencing products

Reactions were precipitated on wet ice with 1μL 3M sodium acetate (pH 5.2) and 25μL 100% ethanol for 10 min, followed by centrifugation at 14,000rpm for 30 min at 4°C. DNA was washed once in 75% ethanol, air-dried and resuspended in 1μL of loading buffer (loading dye (25mM EDTA, pH 8.0, 50mg/ml Blue Dextran) and formamide at 1:5).

N.B. Post 2005 DNA was resuspended in 10μL of loading buffer (loading dye and formamide at 1:60).

2.5.3. Sequencing run

Sequencing was performed in house on a ABI 377 automatic DNA sequencer, which was replaced by a ABI Prism 3130xl Genetic analyser in 2005 (both Applied Biosystems,

Warrington, UK). DNA samples were kindly run and the files processed by Gavin Babbage, HIT Group, Cancer Sciences Division, University of Southampton.

2.5.4. Sequencing analysis

Sequence analysis used MacVector™ 7.1.1 software (Oxford Molecular, Oxford, UK) and Entrez and V-Base databases (Centre for Protein Engineering, MRC Centre, Cambridge, UK; http:// vbase.mrc-cpe.cam.ac.uk/).

2.6. Cloning of PCR products

2.6.1. Ligation

PCR products were ligated into pGEM-T® Vector System I (Promega, Southampton, UK). Ligation reactions consisted of 3μL of PCR product, 3 Weiss Units of T4 DNA ligase, 50ng pGEM-T vector and 1X ligation buffer (30mM Tris-HCl, pH 7.8; 10mM $MgCl_2$; 10mM DTT; 1mM ATP; 10% polyethylene glycol) and were incubated overnight at 4°C.

2.6.2. Transformation

Ligation products were transformed into JM109 competent *E.coli* cells (Promega, Southampton, UK). Briefly, 5μL of ligation reaction was incubated with 50μL of competent *E.coli* cells on wet ice for 20 min. Bacteria were heat shocked at 42°C for 45 sec and then immediately returned to wet ice for 2 min. One mL of luria broth (LB) (Sigma-Aldrich Company Ltd, Gillingham, UK) was added and the bacteria were incubated at 37°C for 1 hour with shaking at 200rpm. Transformed cells were then plated out onto LB-agar supplemented with 100μg/mL ampicillin (LB-agar/Amp100) and incubated at 37°C for 18 hours. The pGEM-T® Vector System I uses blue(-)/white(+) colony selection and also requires that the LB-agar be supplemented with 40μg/mL 5-bromo-3-indoyl-b-D-galactopyranoside (/X-Gal) (Promega, Southampton, UK) and 500μM isopropyl thiogalactoside (/IPTG).

2.6.3. Selection and small-scale growth of transformants

Positive white transformants (~12-25) were randomly selected and used to inoculate 2mL LB/Amp100. Bacterial cultures were incubated at 37°C for 18 hours with shaking at 200rpm.

2.6.4. Plasmid purification

Plasmid DNA was purified using the QIAprep® Spin Miniprep Kit (QIAgen, Crawley, UK) according to the manufacturers instructions. Briefly, following centrifugation at 8,000rpm for 1 min, the bacterial cell pellets were resuspended in 250μL buffer P1. An equivalent volume of lysis buffer P2 was added and mixed by inversion, followed by neutralisation with 350μL buffer N3. After centrifugation at 14,000rpm for 10 min, the supernatant was applied to a QIAprep® Spin column and incubated for 1 min. Plasmid DNA was washed once with PB buffer and twice with PE buffer before being eluted into 100μL of elution buffer (10mM Tris-Cl, pH 8.5). Plasmid DNAs were stored at –20°C.

2.6.5. Sequencing of plasmid DNA

Five μL (~0.5μg) of plasmid DNA was sequenced in both directions using the T7 and SP6 forward and reverse primers (Table 2.3.) in a total volume of 10μL, as previously described in Section 2.5.

Table 2.3. Primers for the sequencing of plasmid DNA

Primer Name	Orientation	Sequence
T7	Sense	TAA TAC GAC TCA CTA TAG GG
SP6	Anti-sense	ATT TAG GTG ACA CTA TAG AA

2.7. Analysis of sequencing data

2.7.1. Identification of tumour-derived V genes

A tumour-derived V gene was defined by the presence of repeated sequences with a clonally related CDR3, including two independent PCR products. Alignment to V-Base, which contains all known human germline Ig V region genes, identified the germline V, D and J segments with the greatest homology and allowed the construction of a presumed germline counterpart.

2.7.2. Analysis of somatic hypermutation

The level of mutation was determined by comparison of the nucleotide sequence from the beginning of FR1 to the end of FR3 to the germline V genes with the greatest homology and was expressed as a percentage of the total nucleotides. Additional mutations that featured in a proportion of V gene molecular clones, but not in the dominant consensus sequence were considered as evidence for ongoing SHM. Analysis of the intraclonal heterogeneity from FR1 to the end of CDR3 was expressed as a percentage of the total

nucleotides sequenced; ongoing SHM was confirmed if intraclonal heterogeneity exceeded that of the previously determined Taq error rate of 0.05% (data not shown).

2.7.3. Analysis of *N*-glycosylation sites

The aa sequence of the V genes were examined across their entire length for potential *N*-glycosylation sites with the motif Asn-X-Ser/Thr (N-X-S/T), where X could be any amino acid except Pro. Those *N*-glycosylation motifs using Asp (D), Glu (E), Leu (L) or Trp (W) as the intervening amino acid X, when followed by Ser (S), were included in the analysis even though they represent disfavoured aa where functional glycosylation at the protein level may be less likely.[463]

2.7.4. Statistical analyses

Data was analysed using Microsoft® Excel 2004 for Mac®, version 11.2 (Microsoft Corporation, WA, USA.). Statistical analysis was carried out with the Prism 4 software (GraphPad Software Inc., CA, USA.). An unpaired, two-tailed t-test was used to assess statistical differences between groups using a confidence interval of 95%. The Fisher exact test was used to compare the frequencies of Bcl-2 expression between groups.

2.8. Analysis of functional scFv *N*-glycosylation

2.8.1. Assembly of scFv-Cκ expression vector

2.8.1.1. scFv assembly

Tumour-derived V_H and V_L sequences were assembled as single chain variable fragment (scFv) as previously described (Figure 2.1.).[464] Briefly, in a 1st round PCR, V_H and V_L sequences were reamplified in separate PCR reactions using V gene specific primers that incorporated an SfiI restriction enzyme (RE) site 5' of V_H, a NotI RE site 3' of V_L and overlapping linker sequence [(Gly Gly Gly Gly Ser)x3] 3' of V_H and 5' of V_L. An example set of primers for scFv assembly is shown for FL14 in Table 2.4. In a 2nd round splicing by overlapping extension (SOE) PCR, the outer most primers, VH-SfiI-5' and VL-NotI-3', were used to join the two fragments together.

PCR reactions were in a total volume of 50μL and contained 5 Units of Expand High Fidelity Taq DNA polymerase (Roche Diagnostics Ltd, Lewes, UK.), 250μM dNTP's, 20pM of each primer and 1X buffer (containing Tris-Cl; KCl; $(NH_4)_2SO_4$; 1.5mM $MgCl_2$, pH8.7). Template DNA was 1μL of V_H or V_L mini prep DNA for 1st round PCRs and 5μL of both V_H and V_L 1st round PCR products for 2nd round SOE PCR. Cycling was performed on a GeneAmp® PCR System 2700 (Applied Biosystems, Warrington, UK) and consisted

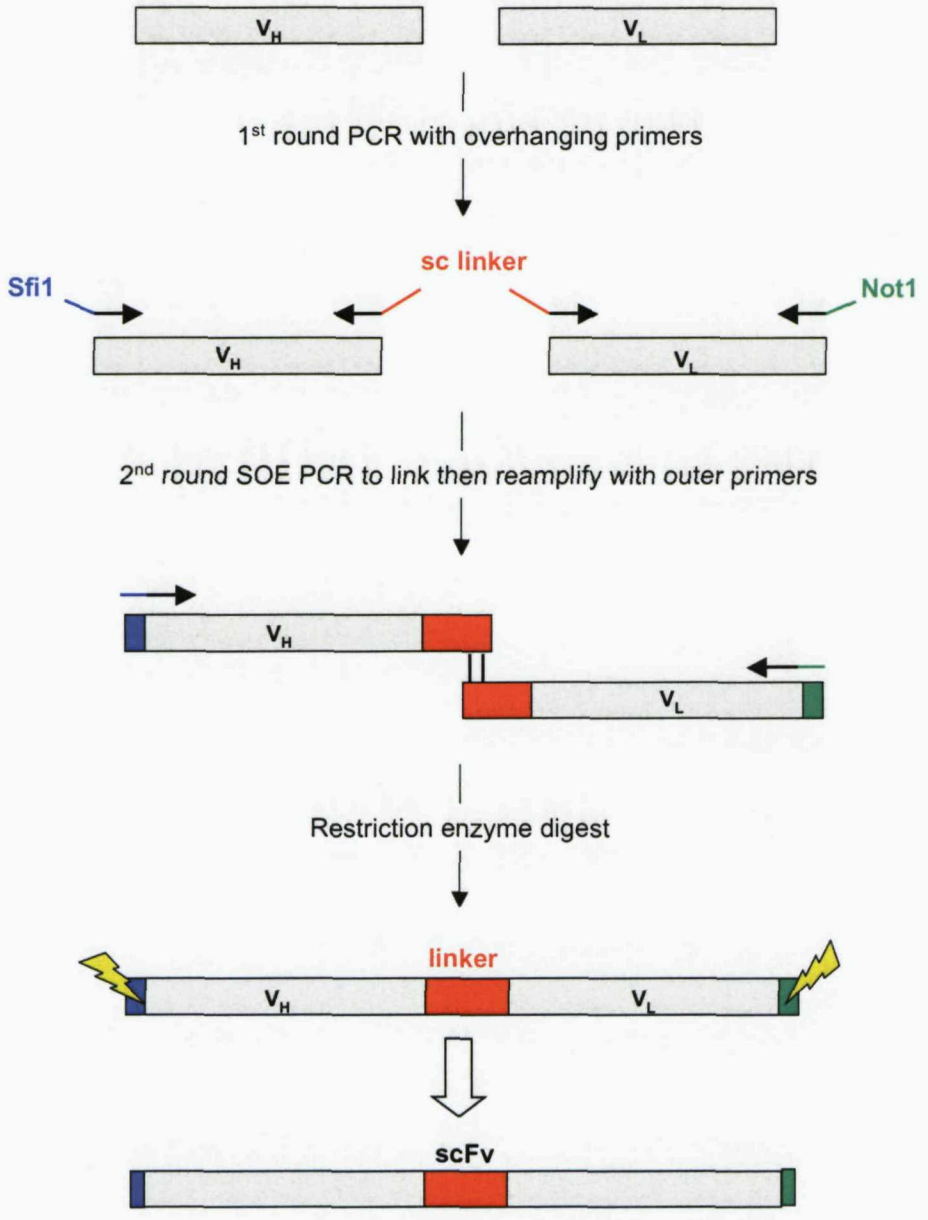


Figure 2.1. A schematic diagram to illustrate the assembly of tumour-derived immunoglobulin heavy (V_H) and light (V_L) chain variable region genes into a single chain variable fragment (scFv) by PCR. A 1st round PCR introduces an SfiI restriction enzyme site 5' and part of the linker sequence 3' of V_H and part of the linker sequence 5' and a NotI restriction enzyme site 3' of V_L. A 2nd round SOE PCR links the two fragments together to form scFv. The linker sequence is shown in *red*, the Sfi1 site in *blue* and the Not1 site in *green*.

Table 2.4. An example set of primers for scFv assembly for FL14

Primer Name	Orientation	Sequence
FL14-VH-SFiI-5'	Sense	T GTG GCC CAG CCG GCC ATG GCC CAG GTG CAG CTA CAG CAG TGG
FL14-VH3'Linker	Anti-sense	AGA GCC ACC TCC GCC TGA ACC GCC TCC ACC TGA AGA GAC GGT GGC CAG
FL14-VL5'Linker	Sense	GGC GGA GGT GGC TCT GGC GGT GGC GGA TCG GAC ATC GTG ATG ACC CAG
FL14-VL-NotI-3'	Anti-sense	TGC GGC CGC TTT GAT CTC CAG CTT GGT

Patient specific sequence is shown in *blue*. Linker sequence is shown in *red*. Restriction enzyme sites are underlined.

15aa linker sequence: GGT GGA GGC GGT TCA GGC GGA GGT GGC TCT GGC GGT GGC GGA TCG
(Gly Gly Gly Gly Ser) x3

Table 2.5. Primers for Cκ PCR amplification

Primer Name	Orientation	Sequence
Cκ-NotI-5'	Sense	TAT GCG GCC GCT ACT GTG GCT GCA CCA TCT
Cκ-XbaI-3'	Anit-sense	ATT CTA GAT CAA CAC TCT CCC CTG TTG AA

Restriction enzyme sites are underlined. A stop codon is shown in *green*.

of an initial denaturing step at 94°C for 4 min, followed by 10 cycles of denaturation at 94°C for 1 min, primer annealing at 50°C for 1 min and extension at 72°C for 1 min, and a further 20 cycles of denaturation at 94°C for 1 min, primer annealing at 60°C for 1 min and extension at 72°C for 1 min, with a final extension step at 72°C for 10 min and hold at 4°C. For the SOE PCR, extension was increased to 1 min 30 secs.

Following separation and visualisation by agarose gel electrophoresis, scFv (~750bp) was excised, purified and eluted into 30μL elution buffer (see section 2.4.4. and 2.4.5.).

2.8.1.2. Cκ amplification

Two μL of template DNA was used to amplify Cκ with HotStarTaq® (QIAgen, Crawley, UK) using Cκ specific primers that incorporated an 5' NotI RE site and a 3' XbaI RE site, which was preceded by a stop codon (Table 2.5.). PCR reactions were set up as detailed previously (see section 2.4.) and cycling consisted of an initial denaturing step at 94°C for 15 min, followed by 30 cycles of denaturation at 94°C for 1 min, primer annealing at 60°C for 1 min and extension at 72°C for 1 min with a final extension step at 72°C for 10 min and hold at 4°C.

PCR products of approximately 300bp were separated and visualised by agarose gel electrophoresis, followed by purification and elution into 30μL elution buffer (see section 2.4.4. and 2.4.5.).

2.8.1.3. DNA digest

Ten μL of scFv was digested sequentially with 10 Units of SfiI RE at 50°C for 2 hrs, followed by 20 Units of NotI RE at 37°C for 2 hrs, according to the manufacturers instructions (New England Biolabs (UK) Ltd, Hitchin, UK.). Digests were in a total volume of 20μL and contained 100μg/mL bovine serum albumin (BSA) and 1X NE buffer 2 (10mM Tris-HCl; 50mM NaCl; 10mM MgCl2; 1mM DTT, pH7.9). NotI was heat inactivated at 65°C for 20 min post-digest. Similarly, 10μL of Cκ was digested sequentially with 10 Units of NotI RE at 37°C for 2 hrs, followed by 15 Units of XbaI RE at 37°C for 2 hrs. The reaction was as detailed above except used 1X NE buffer 3 (50mM Tris-HCl; 100mM NaCl; 10mM MgCl2; 1mM DTT, pH7.9). Heat inactivation at 65°C for 20 min followed both NotI and XbaI digests.

Digested scFv and Cκ were separated and visualised by agarose gel electrophoresis, followed by purification and elution into 25μL elution buffer (see section 2.4.4. and 2.4.5.).

2.8.1.4. Cloning into pcDNA3.1

Cloning used the eukaryotic expression vector pcDNA3.1(+) (Invitrogen Ltd, Paisley, UK.). pcDNA3.1 comprises enhancer-promoter sequences from the immediate early gene of human cytomegalovirus (CMV) for high level gene transcription, a multiple-cloning site, transcription-termination and polyadenylation sequences from the bovine growth hormone gene to enhance RNA stability, a separate expression cassette conferring resistance to neomycin and pBR322 vector sequences, including sequence conferring ampicillin resistance, for replication and selection in *E. coli* (Figure 2.2.). The pcDNA3.1 vector had previously been modified within the multiple-cloning site to include a start codon followed by the V_H1 leader sequence, which incorporates an SfiI RE site at the 3' end; I will continue to refer to the modified vector as pcDNA3.1.

Digested Cκ and scFv fragments were sequentially cloned into pre-digested ('empty') pcDNA3.1 vector (Invitrogen Ltd, Paisley, UK.) using NotI/XbaI, and SfiI/NotI RE sites, respectively, to yield pcDNA3-scFv-Cκ (Figure 2.3.); scFv was linked to the Cκ at the C-terminus. Cloning was according to the methods previously described (see sections 2.6.1 to 2.6.4.). Sequencing with T7, SP6 and scFv assembly PCR primers was used to confirm the presence of the correct scFv-Cκ insert (see section 2.6.5.); an example scFv-Cκ nucleotide sequence for FL14 is shown in Figure 2.4.

2.8.1.5. Medium-scale amplification of plasmid DNA

Plasmids with the correct scFv-Cκ sequence were further amplified to yield up to 1mg plasmid DNA. A single colony, or 10μL of miniprep culture, was used to inoculate 5mL LB/Amp100 and incubated at 37°C for 8 hours with shaking at 200rpm. One mL of starter culture was then used to inoculate 200mL LB/Amp, followed by incubation at 37°C for 18 hours with shaking at 200rpm. Plasmid DNA was purified using the QIAgen™ plasmid Maxiprep Kit (QIAgen, Crawley, UK) according to the manufacturers instructions. Briefly, following centrifugation at 3,500rpm for 15 min, the bacterial cell pellet was resuspended in 10mL buffer P1. Bacterial cells were lysed with 10mL buffer P2, neutralised with 10mL buffer P3 and centrifuged at 4,000rpm for 15 min. A QIAgen-Tip500 column was equilibrated by applying 10mL buffer QBT prior to loading of the supernatant. Plasmid DNA was washed twice with buffer QC, eluted with 15mL elution

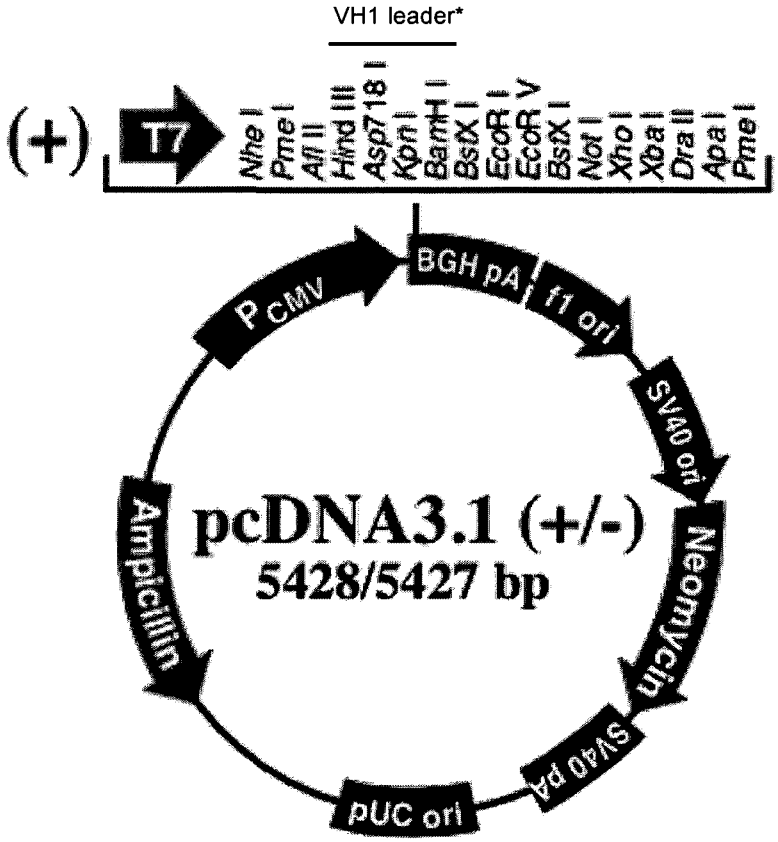


Figure 2.2. A schematic representation of the expression vector pcDNA3.1(+). pcDNA3.1(+) (Invitrogen Ltd, Paisley, UK) was previously modified within the multiple-cloning site to include a start codon followed by the V_H1 leader sequence, which also incorporated an SfiI restriction enzyme (RE) site at the 3' end (*). Patient specific scFv and Cκ DNA fragments were sequentially cloned into the modified pcDNA3.1(+) using SfiI/NotI and NotI/XbaI RE sites, respectively. The pcDNA3.1(+) vector incorporates a CMV-derived promoter sequence upstream of the polylinker site to ensure a high level of transcription *in vitro*. The ampicillin resistance gene was used for selection in bacteria, while the SV40 origin of replication ensures plasmid replication. *Image courtesy of Invitrogen Limited.*

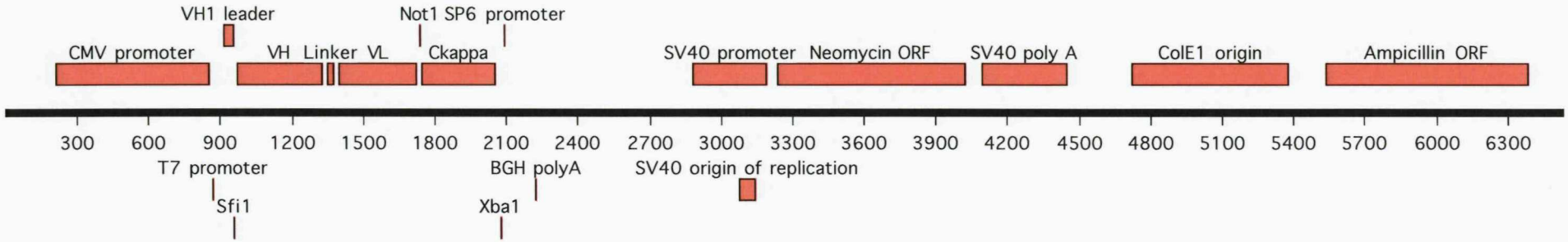


Figure 2.3. A vector map for pcDNA3.1-scFv-Cκ. A schematic diagram to show the position of scFv-Cκ in pcDNA3.1. A start codon and V_H1 leader sequence lie immediately upstream of scFv, which is fused to Cκ at the C-terminus. The presence of T7 and SP6 primer sites located 5' and 3' of the scFv-Cκ insert, respectively, enable the sequence to be verified. An upstream CMV-derived promoter sequence ensures a high level of transcription *in vitro*. The ampicillin, or neomycin, resistance genes enables selection in *E. coli*. The SV40 origin of replication ensures plasmid replication.

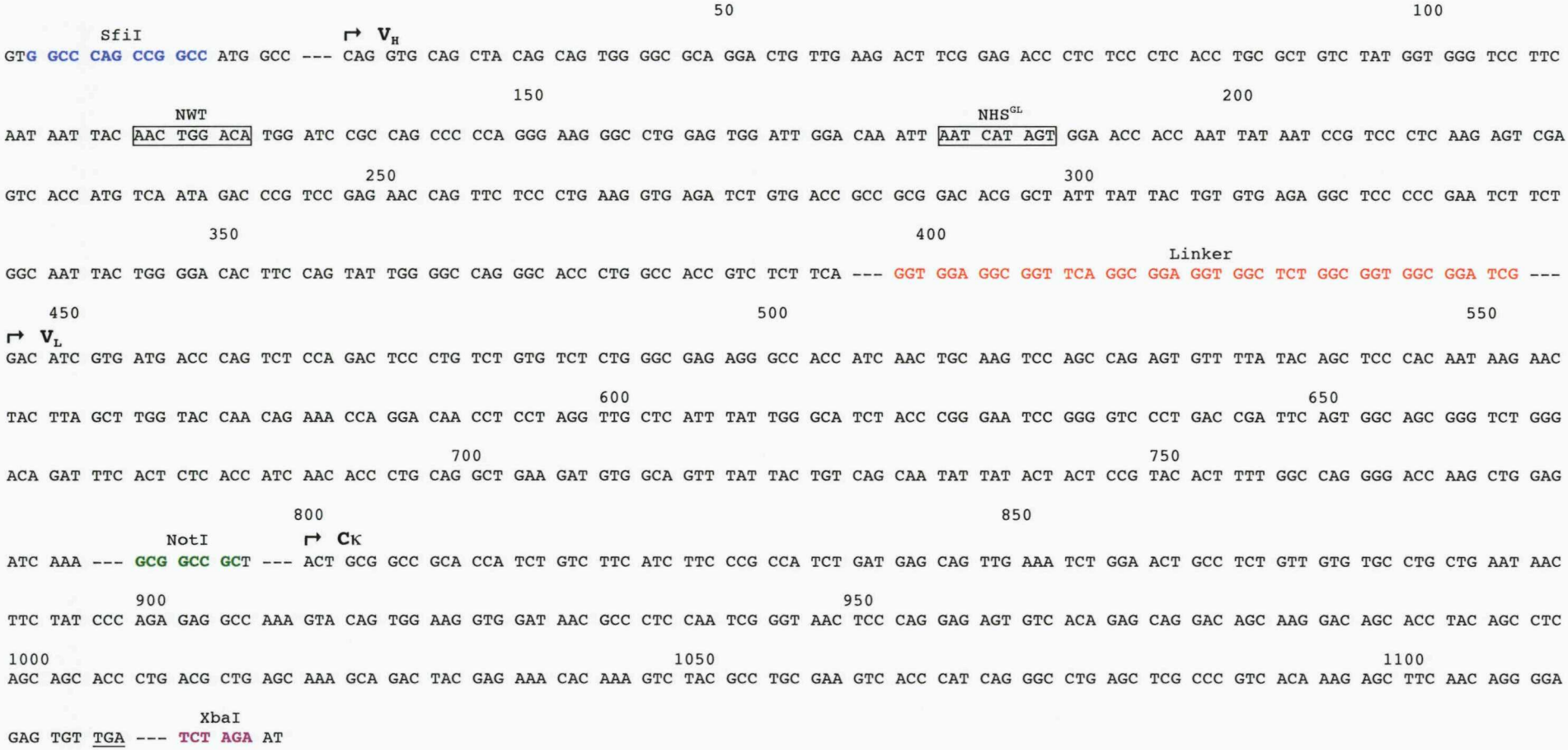


Figure 2.4. The scFv-Cκ nucleotide sequence for FL14. The scFv-Cκ nucleotide sequence for FL14. A V_H1 leader sequence (3' end shown above) is positioned upstream of V_H and incorporates the SfiI RE site. The 45bp linker sequence, which joins V_H to V_L, is shown in *red*, the SfiI RE site in *blue*, the NotI RE site in *green* and the XbaI RE site in *purple*. A stop codon (*underlined*) is shown immediately upstream of the XbaI RE site. *N*-glycosylation sites are boxed, with the aa motif shown above.

buffer QF and precipitated with 10.5mL isopropanol. Plasmid DNA was pelletted by centrifugation at 4,000rpm for 30 min at 4°C, washed once with 75% ethanol and resuspended in 1mL of elution buffer (10mM Tris-Cl, pH 8.5).

Plasmid DNA was quantified by measuring the absorbance at 260nm using an Eppendorf BioPhotometer (Eppendorf AG, Hamburg, Germany).

2.8.2. Expression of recombinant scFv-Cκ proteins

Expression used the FreeStyle™293 mammalian expression system (Invitrogen Ltd, Paisley, UK.). Plasmid DNA was used to transfect suspension 293 human embryonic kidney (293F) cells using 293Fectin™ (Invitrogen Ltd, Paisley, UK.) according to the manufacturers instructions. Briefly, 100μg of plasmid DNA and 100μL 293Fectin™ were each diluted into 3mL OPTI-MEM® I Reduced Serum Medium (Invitrogen Ltd, Paisley, UK.) and incubated for 5 min at RT. Three mL plasmid DNA was then added to 3mL 293Fectin™, mixed by inversion and incubated for a further 30 min. The DNA-293Fectin™ complex was added drop wise to $1x10^8$ 293F cells in 100mL FreeStyle™ 293 expression medium (Invitrogen Ltd, Paisley, UK.), whilst swirling. Cultures were incubated in a 37°C incubator with a humidified atmosphere of 8% CO_2 in air on an orbital shaker rotating at 125rpm. After 72 hrs, culture supernatants were harvested by centrifugation at 1700rpm for 7 min and filtered over a 0.22μm membrane under vacuum. Cell pellets were resuspended in 100mL expression medium for a second round of expression for 72 hrs. Supernatants were concentrated 10X by centrifugation at 2200rpm at 4°C using vivaspin20 10,000 MWCO concentrators (Sartorious Ltd, Epsom, UK.). Concentrated supernatants were stored at 4°C short-term.

2.8.3. Assessment of protein production by anti-Cκ ELISA

Initial assessment of Cκ-tagged recombinant protein expression was by enzyme-linked immunosorbent assay (ELISA). A 96 well Maxisorp immunoplate (Nunc, Roskilde, Denmark.) was coated with polyclonal goat anti-human κ Ab (Serotec, Oxford, UK.) at a 1:1000 dilution in coating buffer (200μL/well) and left overnight at 4°C. Plates were washed four times with phosphate buffered saline (PBS)-Tween20 (PBS-T) and blocked for 1 hr with shaking at 37°C in PBS with 2% milk. After four washes with PBS-T, cell supernatants were added to the plate at 1:50 in PBS-T, double diluted and incubated for 1 hr with shaking at 37°C. After washing, the plate was incubated with polyclonal goat anti-human κ horseradish peroxidase (HRP)-conjugated Ab (Sigma-Aldrich Company Ltd,

Gillingham, UK) at a 1:500 dilution in PBS-T for 1 hr with shaking at 37°C. After washing, the plate was incubated with 200μL/well of substrate buffer- *o*-phenylenediamine substrate (Sigma-Aldrich Company Ltd, Gillingham, UK.). The reaction was stopped by adding 80μL 2.5M H_2SO_4 to each well. The absorbance at 490nm was measured using a MRX microplate reader (Dynex Technologies (UK) Ltd, Billingshurst, UK.). Samples were quantified from a standard curve using an internal standard of Cκ-tagged protein of known concentration.

2.8.4. Purification of scFv-Cκ recombinant proteins

2.8.4.1. Anti-human κ chain column preparation

Proteins were purified using a column of polyclonal sheep anti-human κ-chain Ab linked to CNBR-activated sepharose 4B beads (Amersham Biosciences UK Ltd, Little Chalfont, UK.). Sixty mgs (5-10mg/mL) of polyclonal anti-human κ-chain Ab was coupled to CNBR-activated sepharose 4B beads according to the manufacturers instructions. Briefly, under sterile conditions, 8ml pre-prepared swollen gel (2.3g dry) was mixed with 10mL coupling buffer containing the Ab in a stoppered vessel by rotation for 4 hrs at 4°C. Following washing with coupling buffer over a sintered funnel, the medium was transferred to 10mL 1M ethanolamine, pH 8.0, and rotated for 1 hr at RT. Finally, the medium was washed with three cycles of alternating pH buffer; 1X Tris-NaCl wash buffer, pH 8.0 and ammonium thiocyanate elution buffer, pH 12.0. Four mL of medium was used to pack a sterile 25mL glass column for purification. Columns were stored at 4°C. Sheep anti-human κ-chain Ab was kindly provided by Mrs Maureen Power, Directors Group, Tenovus Laboratory, University of Southampton.

2.8.4.2. scFv-Cκ protein purification

The anti-human κ-chain column was pre washed with two column volumes (25mL) of 1X Tris-NaCl wash buffer. The concentrated supernatant was applied to the column and allowed to drip through by gravity flow, any unbound protein was collected. The column was then washed with a four volumes of wash buffer and the bound protein was eluted with two rounds of 3mL ammonium thiocyanate elution buffer, followed by two volumes wash buffer. Protein elutions were immediately dialysed into PBS and concentrated to <5mL using vivaspin20 10,000 MWCO concentrators.

2.8.5. Protein quantification and quality check

2.8.5.1. BCA assay

Protein concentration was determined by BCA™ protein assay (Perbio Science UK Ltd, Cramlington, UK.). In a microplate, 10μL of purified protein or known BSA standard (working range 2-0.025μg/mL) was added to 200μL of a 1:1 mixture of BCA™ reagent A and B. Unknown samples were double diluted. Following incubation for 45 min at 37°C the absorbance at 562nm was measured using a MRX microplate reader (Dynex Technologies (UK) Ltd, Billingshurst, UK.). Protein concentrations of unknown samples were calculated against a standard curve. Proteins were stored at -20°C in 250μL aliquots.

2.8.5.2. Western blot analysis and protein staining

Four μg of scFv-Cκ protein was denatured in X1 NuPAGE loading buffer/ 12.5% 2-mercaptoethanol at 100°C for 5 min and run through a NuPAGE Bis-Tris gradient polyacrylamide (4%-12%) gel (Invitrogen Ltd, Paisley, UK.) against full-range molecular weight rainbow marker (Amersham Biosciences UK Ltd, Little Chalfont, UK.). Separated protein bands were visualised by staining with SimplyBlue Safestain™ (Invitrogen Ltd, Paisley, UK.) for 1 hr at RT. For Western blot analysis, separated protein bands were electrotransferred for 1 hr onto Hybond-™P nitrocellulose membrane (Amersham Biosciences UK Ltd, Little Chalfont, UK.). The membrane was blocked in PBS-T with 5% milk for 1 hr with shaking at RT. The membrane was washed three times in PBS-T for 5 min then probed with polyclonal goat anti-human κ HRP-conjugated Ab (Sigma-Aldrich Company Ltd, Gillingham, UK) at a 1:1000 dilution in PBS-T for 1 hr at RT. Proteins were visualized by chemiluminescence using the enhanced chemiluminescence plus reagents (Amersham Biosciences UK Ltd, Little Chalfont, UK.).

2.8.5.3. Specificity ELISA

2.8.5.3.1. Generation of anti-Id Abs in mice

Groups of 8-10 C57/BLK6 mice, bred in house, were vaccinated at 6-10wk of age with 50μg of a DNA-fusion vaccine encoding the scFv linked at the C-terminus to the sequence of Fragment C from tetanus toxin, as previously described.[465] Injection was in saline at two sites in the quadriceps muscles, at day 0, day 21 and day 42. Mice were sacrificed between days 56-63 and the serum collected and pooled. Animal welfare and experimentation were conducted in accordance with local Ethical Committee and United Kingdom Coordinating Committee for Cancer Research guidelines, under Home Office licence.

2.8.5.3.2. Anti-Id ELISA

ELISA was used to confirm the specificity of the recombinant scFv-Cκ proteins. A 96 well Maxisorp immunoplate (Nunc, Roskilde, Denmark.) was coated with each patient-derived scFv-Cκ protein at 2ug/mL in coating buffer (200μL/well) and left overnight at 4°C. Plates were washed four times with PBS-T and blocked for 1 hr with shaking at 37°C in PBS with 1% BSA. After four washes with PBS-T, patient specific mouse anti-serum was added to the plate at a 1:10 dilution and incubated for 1 hr with shaking at 37°C; up to a further eight non-specific mouse anti-serums were included on the plate as negative controls. After washing, the plate was incubated with polyclonal sheep anti-mouse IgG HRP-conjugated Ab (The Binding Site, Birmingham, UK.) at a 1:1000 dilution in PBS-T (200μL/well) for 1 hr with shaking at 37°C. After washing, the plate was incubated with 200μL/well of substrate buffer- *o*-phenylenediamine substrate (Sigma-Aldrich Company Ltd, Gillingham, UK.). The reaction was stopped by adding 80μL 2.5M H_2SO_4 to each well. The absorbance at 490nm was measured using a MRX microplate reader (Dynex Technologies (UK) Ltd, Billingshurst, UK.).

2.8.6. Analysis of scFv *N*-linked glycosylation

To assess functional glycosylation, scFv-Cκ proteins were treated with two separate *N*-glycan removal enzymes: Peptide:*N*-glycosidase F (PNGase F) and Endoglycosidase H (Endo H). As a negative control, scFv-Cκ protein derived from a MM patient for which no *N*-glycosylation motifs were present in the nucleotide sequence was also treated.

2.8.6.1. Digestion with Peptide:*N*-glycosidase F

To remove all types of *N*-glycan chains, scFv-Cκ proteins were treated with PNGase F (New England Biolabs (UK) Ltd, Hitchin, UK.). Firstly, 20μg of protein was denatured with 1X glycoprotein denaturing buffer at 100°C for 10 min. In a final volume of 100μL, 1X G7 reaction buffer, 10% NP-40 and 2 Units PNGase F were added to the reaction mix and incubated for a further 1 hr at 37°C.

2.8.6.2. Digestion with Endoglycosidase H

To remove high mannose *N*-glycan chains, scFv-Cκ proteins were treated with Endo H (New England Biolabs (UK) Ltd, Hitchin, UK.). Firstly, 20μg of protein was denatured with 1X glycoprotein denaturing buffer at 100°C for 10 min. In a final volume of 100μL, 1X G5 reaction buffer and 2 Units Endo H were added to the reaction mix and incubated for a further 1 hr at 37°C.

2.8.6.3. Polyacrylamide gel electrophoresis

Twenty μL (4μg) of undigested, PNGase F digested and Endo H digested scFv-Cκ protein was denatured in X1 NuPAGE loading buffer/ 12.5% 2-mercaptoethanol at 100°C for 5 min and run through a NuPAGE Bis-Tris gradient polyacrylamide (4%-12%) gel (Invitrogen Ltd, Paisley, UK.) against full-range molecular weight rainbow marker (Amersham Biosciences UK Ltd, Little Chalfont, UK.). Separated protein bands were visualised by staining with SimplyBlue Safestain™ (Invitrogen Ltd, Paisley, UK.) for 1 hr at RT.

2.8.6.4. Normal phase high performance liquid chromatography

Three scFv-Cκ proteins were selected for normal phase high performance liquid chromatography (NP-HPLC) analysis. Fifteen μg of scFv-Cκ protein was separated using 10% SDS-PAGE, visualised by coomassie blue and excised. *N*-linked glycans were released *in situ* with PNGase F, fluorescently labelled with the fluorophore 2-aminobenzamide (Ludger Ltd, Oxford, UK.) and processed through NP-HPLC as previously described; exoglycosidase array digestions and Glycobase databases were complied in the Oxford Glycobiology Institute.[466] [466] Unlabelled samples were also analysed by mass spectroscopy. These analyses were kindly performed by Mrs Catherine Radcliffe, Glycobiology Institute, Department of Biochemistry, University of Oxford.

2.8.7. Analysis of binding of scFv-Cκ proteins to recombinant mannose receptor

Assessment of the binding of mannose receptor to scFv-Cκ proteins was by ELISA and used a fusion protein incorporating carbohydrate recognition domains 4-7 of the mouse mannose receptor fused to human IgG1 Fc (CRD_{4-7}-Fc); this was kindly provided by Luisa Martinez-Pomares, Institute of Infection, Immunity and Inflammation, School of Molecular Medical Sciences, University of Nottingham. Binding assays were performed as previously described.[467] [467] Briefly, a 96 well Maxisorp immunoplate (Nunc, Roskilde, Denmark.) was coated with scFv-Cκ protein at 5ug/mL in 0.154 M NaCl (50μL/well) and left overnight at 4°C. Plates were washed three times with Ca^{2+} Tris-buffered saline-Tween20 (TBS-T), followed by incubation with CRD_{4-7}-Fc at 2μg/mL in TBS-T for 2hr at RT. This was followed by further washing in TBS-T and incubation with an anti-human IgG Fc-specific alkaline phosphatase (AP) conjugate at a 1:1000 dilution in TBS-T for 1 hr at RT. After washing three times in TBS-T and twice in AP-buffer, the plate was developed with 50μL/well substrate buffer- *p*-nitrophenyl phosphate substrate (Sigma-Aldrich Company Ltd, Gillingham, UK.). The absorbance at 405nm was measured after 10 min using a MRX microplate reader (Dynex Technologies (UK) Ltd, Billingshurst,

UK.). Readings were measured against a blank of uncoated wells incubated with CRD_{4-7}-Fc protein. Yeast mannan-coated wells were used as a positive control. All assays were carried out in triplicate.

Inhibition assays were also performed using D-mannose, which included the pre-incubation of CRD_{4-7}-Fc protein with 100mM D-mannose for 30min at 4°C prior to plate application.

2.9. Assessment of AID mRNA expression

AID mRNA expression was evaluated by 2 methods, a standard PCR assay and a quantitative real-time PCR assay (qPCR). The Ramos human BL cell line was used as a positive control. Negative controls, in which the cDNA was absent from the reaction mix, were run in parallel.

2.9.1. Standard PCR

Two μL of cDNA was used to amplify the *AID* gene with HotStarTaq® (QIAgen, Crawley, UK) using 5' and 3' oligonucleotide primers designed specifically (Table 2.6.). PCR reactions were set up as detailed previously (see section 2.4.) and cycling consisted of an initial denaturing step at 94°C for 15 min, followed by 30 cycles of denaturation at 94°C for 1 min, primer annealing at 56°C for 1 min and extension at 72°C for 1 min with a final extension step at 72°C for 10 min and hold at 4°C. Amplification products of approximately 650bp were separated, purified and directly sequenced using the forward 5' PCR primer (see sections 2.4.4., 2.4.5. and 2.5.). Sequences were aligned to the AID reference sequence (EMBL, GenBank DDBJ nucleotide sequence database under accession number NM_020661) using MacVector™ 7.1.1 software (Oxford Molecular, Oxford, UK).

Table 2.6. Primers for AID PCR amplification

Primer Name	Orientation	Sequence
AID Forward	Sense	AGG CAA GAA GAC ACT CTG GGA CAC C
AID Reverse	Anti-sense	GTG ACA TTC CTG GAA GTT GC

2.9.2. Quantitative real-time PCR

The qPCR assay was performed using TaqMan technology and a 7500 real time PCR system (Applied Biosystems, Warrington, UK.). qPCR was carried out in a 20μL reaction volume that contained 2μL of cDNA, 1X TaqMan® Universal PCR Master Mix

with AmpErase® UNG, AID-specific 5' and 3' primers, and a 6-carboxy fluorescein (FAM)–labelled TaqMan minor groove binder (MGB) probe. The primers for *AID* were designed from sequences in exon 1 and exon 2 (Assays-on-Demand Gene Expression system, Applied Biosystems). Amplification of β actin was performed in all cases to normalize the *AID* values. The probe for β actin was labelled with FAM (Assays-on-Demand Gene Expression system, Applied Biosystems). After an initial incubation at 95°C for 10 minutes, the cDNA was amplified for 40 cycles of denaturation at 95°C for 15 seconds and combined annealing/extension at 60°C for 1 minute. Each sample was analysed in triplicate. Sequence Detection Software (SDS version 1.7, Applied Biosystems) was used to analyse the fluorescence emission data after PCR. The threshold cycle (Ct) values of each sample were exported to Microsoft Excel for further analysis (Microsoft, Seattle, WA). The Ct represents the cycle number at which fluorescence passes a fixed threshold. *AID* expression levels in test samples were expressed as the ratio of *AID* to β actin expression. Analyses were performed according to the $2^{-\Delta\Delta Ct}$ method.

2.10. Immunohistochemical analysis

Immunohistochemical analysis was performed using a standard streptavidin–biotin method with appropriate Ag retrieval.[468] Cases were considered to show expression of an Ag if >30% of tumour cells were immunoreactive. For Ki67, a percentage proliferative fraction was estimated. Tumours were classified as GCB-like or non-GCB-like DLBCL based on the decision tree devised by Hans *et al.*; tumours were assigned a GCB-like phenotype if they were $CD10^{+/-}$, Bcl-6^{+} and $MUM1^{-}$ or a non-GCB-like phenotype if they were $CD10^{-}$, Bcl-$6^{+/-}$ and $MUM1^{+}$ (see Chapter 4, Figure 4.1.).[407] A review of diagnostic haematoxylin and eosin (H&E) stained paraffin-embedded tissue sections prior to staining confirmed tumour origin with very little uninvolved tissue in each case.

2.10.1. Primary CNS lymphoma

A panel of monoclonal Abs (mAbs) to the cell surface immunoglobulins IgM, IgD, IgG, IgA, Igκ and Igλ were used to identify tumour immunophenotype on frozen sections. The mAb 9G4 (production in house), which recognises a W/AVY aa sequence within FR1 of the V4-34 V_H region, was used to identify tumours with V4-34 rearrangements. These analyses were kindly performed by Mrs Jean Buontempo and Miss Kellyann Smith, Cellular Pathology Department, Southampton University Hospitals NHS trust.

A panel of mAbs to the cellular markers CD5, CD20, CD79a, CD138, CD30, CD10, Bcl-2, Bcl-6, MUM-1, p53 and Ki-67 (MIB-1) were used on paraffin-embedded tissue sections.

An anti-LMP-1 mAb specific for EBV latent membrane protein-1 was used to confirm EBV negativity. These analyses were kindly performed by Dr. Matthew Hickling, Cellular Pathology Department, Southampton University Hospitals NHS trust.

2.10.2. Diffuse large B-cell lymphoma

A panel of mAbs to the cellular markers CD10, Bcl-2, Bcl-6 and MUM-1 were used on paraffin-embedded tissue sections. These analyses were kindly performed by Dr. Ilske Oschlies, Cellular Pathology Department, Southampton University Hospitals NHS trust.

2.11. Lymphochip cDNA microarray analysis

The "lymphochip" cDNA microarray, which is composed of 17,856 cDNAs whose gene products are preferentially expressed in lymphoid cells and genes thought or confirmed to play a role in cancer or immune function, was used to quantitate the expression of mRNA in 18 DLBCLs. A fluorescence ratio was quantified for each gene and reflected the relative abundance of the gene in each experimental mRNA sample compared to a reference mRNA pool. These analyses were kindly performed by Prof. Izidore Lossos, Molecular and Cellular Pharmacology, University of Miami, USA.

2.12. Fluorescence *In situ* hybridisation

Interphase fluorescence *In situ* hybridisation (iFISH) was performed on 18 DLBCLs where paraffin-embedded material was available to determine the presence of the IgH translocations, t(3;14) and t(14;18), which involve the *Bcl-6* and *Bcl-2* genes, respectively. Assistance with this analysis was kindly provided by Mr Adam Stewart, Leukaemia Research Cytogenetics Group, Cancer Sciences Division, University of Southampton. A review of diagnostic H&E stained paraffin-embedded tissue sections prior to iFISH analysis confirmed tumour origin with very little uninvolved tissue in each case.

2.12.1. Preparing the specimen slides

For each sample, four 3μm thick sections were cut from paraffin-embedded tissue blocks using a microtome. Sections were mounted on positively charged slides, air-dried and baked over night at 56°C; this was kindly performed by Miss Kellyann Smith, Cellular Pathology Department, Southampton University Hospitals NHS trust.

2.12.2. Preparing the tissue for hybridisation

Firstly, tissue sections were de-paraffinised by two times immersion in Xylene for 10 min

at RT, followed by two washes in 100% ethanol for 5 min and drying on a 50°C slide warmer for 2 min. Slides were next pre-treated by immersion in 0.2M HCl for 20 min at RT, followed by washing in dH_2O for 3 min and 2X saline sodium citrate (SSC) buffer for 3 min, and immersion in Pre-treatment solution (NaSCN) (Paraffin pre-treatment kit II; Abbott Laboratories Ltd, Dublin, Ireland.) for 30 min at 80°C, followed by one wash in dH_2O for 1 min and two washes in 2X SSC buffer for 5 min. Slides were removed from 2X SSC wash and blotted before immersion in Protease solution (0.2M HCL; 4mg/mL protease) (Paraffin pre-treatment kit II; Abbott Laboratories Ltd, Dublin, Ireland.) for 20 min at 37°C. Slides were washed twice in 2X SSC buffer for 5 min and dried on a 50°C slide warmer for 2 min. Slides were fixed by immersion in 10% buffered formalin for 10 min at RT, washed twice in 2X SSC buffer for 5 min and dehydrated with a cold ethanol series (70%, 85%, and 100%) for 2 minutes each. Finally, slides were dried on a 50°C slide warmer for 2 min.

2.12.3. Hybridisation of LSI® probes

A mixture of 0.3μL Vysis LSI® probe (Abbott Laboratories Ltd, Dublin, Ireland.) and 2.7μL hybridisation buffer (50% formamide, 2X SSC, 10% dextran sulfate) was applied to the tissue section, which was then covered with a 13mm diameter cover slip and sealed with rubber cement. Probes were hybridised for 5 min at 75°C followed by 18 hrs at 37°C in a humidified chamber (Vysis HYBrite®; Abbott Laboratories Ltd, Dublin, Ireland.).

After hybridisation, the rubber cement was peeled away and the cover slip removed by soaking in 2X SSC buffer. Slides were washed in Wash Buffer 1 (0.4X SSC/0.3%NP-40) for 2 min at 72°C, followed by Wash Buffer 2 (2X SSC/0.1% NP-40) for 2 min at RT. Chromosomes were counterstained with 10μL of a mixture of 4-6-diamidine-2-phenylindole dihydrochloride (DAPI) containing Vectashield antifade in a ratio of 1:100 (Abbott Laboratories Ltd, Dublin, Ireland.).

2.12.3.1. Dual colour, break apart rearrangement probe

A Vysis LSI® dual colour, break apart rearrangement probe is designed to detect chromosomal breakage at a given locus and consists of a mixture of two probes, a SpectrumOrange™ labelled probe and a SpectrumGreen™ labelled probe, that hybridise to either side of the gene of interest, or a known breakpoint cluster (Figure 2.5.). As a result of this probe design, any translocation with a breakpoint within the gene of interest (on one allele) should produce a separate orange signal and a separate green signal, as well as a merged orange/green (yellow) signal (representing the intact locus on the normal

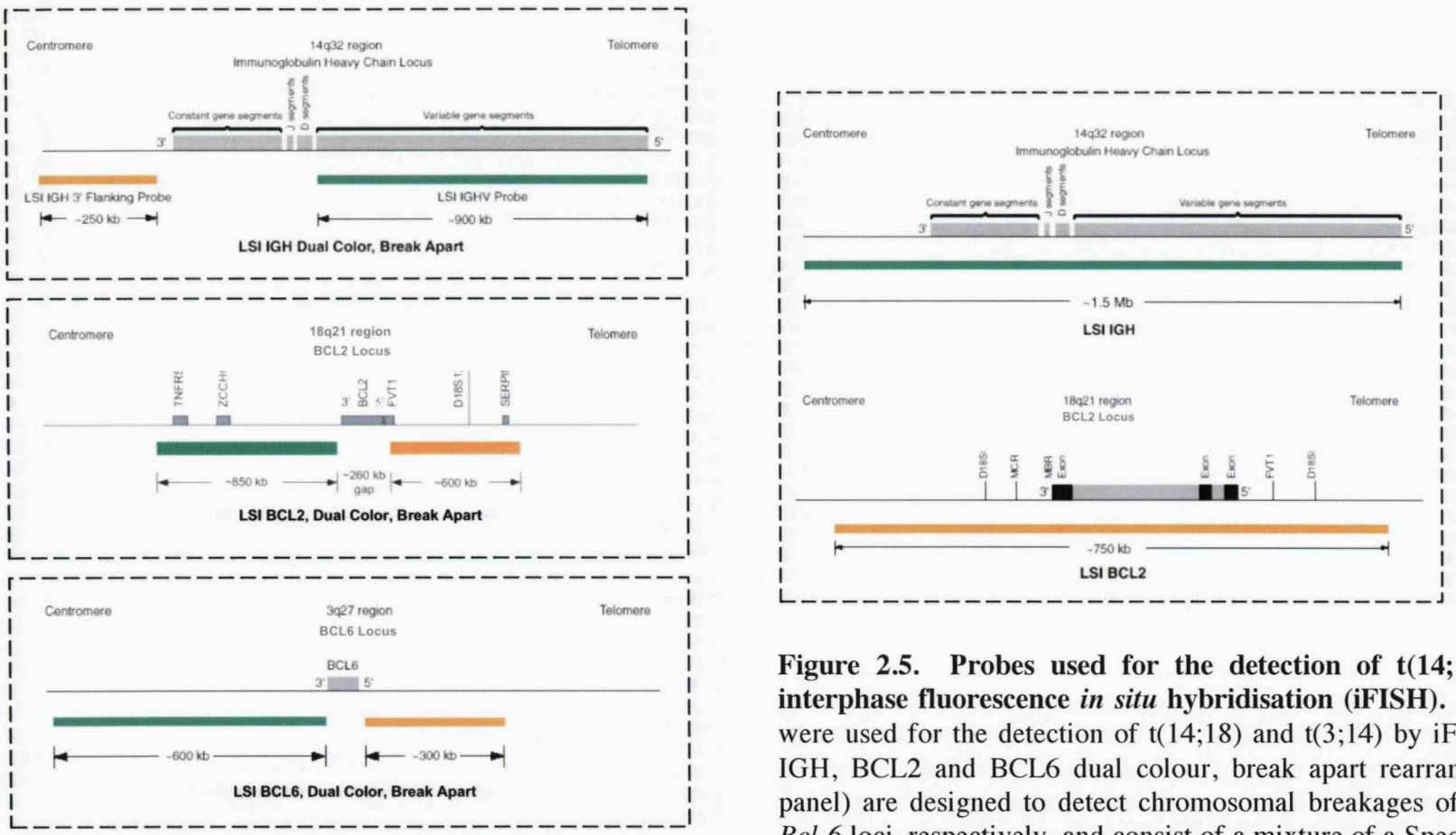


Figure 2.5. Probes used for the detection of t(14;18) and t(3;14) by interphase fluorescence *in situ* hybridisation (iFISH). Two types of probe were used for the detection of t(14;18) and t(3;14) by iFISH: (1) Vysis LSI® IGH, BCL2 and BCL6 dual colour, break apart rearrangement probes (left panel) are designed to detect chromosomal breakages of the *IgH*, *Bcl-2* and *Bcl-6* loci, respectively, and consist of a mixture of a SpectrumOrange™ probe and a SpectrumGreen™ probe, which are separated by a gap that contains the gene of interest. (2) The Vysis LSI® IGH/BCL2 dual colour, dual fusion translocation probe is designed to detect the juxtaposition of *IgH* and *Bcl2* gene (right panel). This contains a mixture of the Vysis LSI® IGH probe labeled with SpectrumGreen™, which spans the entire IgH locus, as well as sequences extending ~300kb beyond the 3' end of the IgH locus, and the Vysis LSI® BCL2 probe labeled with SpectrumOrange™, which spans a the entire *Bcl2* gene with additional sequences extending ~250kb both distal and proximal to the gene. *Image courtesy of Abbott Molecular Inc.*

homolog). This signal pattern indicates that the genomic targets for the 5' and 3' probes have been physically separated as a result of the translocation. The distance between the two probes should not span more than ~1Mb as the two probes will appear separated in normal interphase nuclei due to decondensation of the chromatin.

2.12.3.1.1. IGH dual colour, break apart rearrangement probe

The Vysis LSI® IGH dual colour, break apart rearrangement probe (Abbott Laboratories Ltd, Dublin, Ireland.) is a mixture of two probes that hybridise to either side of the J through to constant regions of the IgH locus. A ~900kb SpectrumGreen™ labelled LSI IGHV 5' probe covers the entire IgH V region, while a ~250kb SpectrumOrange™ labelled LSI IGH 3' covers the entire IgH V region, while a ~250kb SpectrumOrange™ labelled LSI IGH 3' flanking probe lies 3' to the IgH locus (Figure 2.5.).

2.12.3.1.2. BCL2 dual colour, break apart rearrangement probe

The Vysis LSI® BCL2 dual colour, break apart rearrangement probe (Abbott Laboratories Ltd, Dublin, Ireland.) is a mixture of two probes that hybridise to either side of the *Bcl-2* gene. A ~600kb SpectrumOrange™ labelled LSI BCL2 5' probe extends distally, while a ~850kb SpectrumGreen™ labelled LSI BCL2 3' probe lies proximal to the *Bcl-2* gene (Figure 2.5.).

2.12.3.1.3. BCL6 dual colour, break apart rearrangement probe

The Vysis LSI® BCL6 dual colour, break apart rearrangement probe (Abbott Laboratories Ltd, Dublin, Ireland.) is a mixture of two probes that hybridise to either side of the *Bcl-6* gene. A ~300kb SpectrumOrange™ labelled LSI BCL6 5' probe extends distally, while a ~600kb SpectrumGreen™ labelled LSI BCL6 3' probe lies proximal to the *Bcl-6* gene (Figure 2.5.).

2.12.3.2. IGH/BCL2 Dual colour, dual fusion translocation probe

A Vysis LSI® dual colour, dual fusion translocation probe is designed to detect the juxtaposition of two gene loci and consists of a mixture of two probes, a SpectrumGreen™ probe that hybridizes to one gene locus and a SpectrumOrange™ probe that hybridises to a second gene locus; both probes are extended beyond the breakpoint of each chromosome (Figure 2.5.). As a result of this probe design, a translocation resulting in a locus1/2 fusion (on one allele) should produce two yellow fusion signals- a fusion forming on both the derived chromosome and the reciprocal chromosome- and a separate orange and a separate green signal (each representing the intact locus on two normal homologs: '1' and '2').

The Vysis LSI® IGH/BCL2 dual color, dual fusion translocation probe (Abbott Laboratories Ltd, Dublin, Ireland.) consists of a mixture of a SpectrumGreen™ labelled LSI IGH probe, which spans ~1.5Mb and contains sequences homologous to the entire IgH locus as well as sequences that extend ~300kb beyond the 3' end of the IgH locus, and a SpectrumOrange™ labelled LSI BCL2 probe, which covers a ~750kb region that includes the entire *Bcl-2* gene with additional sequences extending ~250kb both distal and proximal to the gene (Figure 2.5.).

2.12.4. Interpretation of signal pattern

Slides were examined at x1,000 magnification (Zeiss Plan-Neofluar 100x, 1.30 numerical aperture, in oil) with a Zeiss Axiophot microscope equipped with single and multiple bandpass filters for DAPI-conferred fluorescence, fluoroisothiocyanate and Texas Red. A dual filter, which allowed visualization of both orange and green flourochromes at the same time, was the most frequently used. Images were captured using PowerGene MacProbe v4.3 (Perceptive Scientific International Ltd., Chester, UK).

2.12.4.1. IGH dual colour, break apart rearrangement probe

When hybridized to a normal nucleus, the LSI IGH dual colour, break apart rearrangement probe produces a two orange/green (yellow) fusion (2F) signal pattern. As there is no probe targeted to the J or constant regions, a slight gap between the two differently coloured probe signals may sometimes be observed in nuclei from normal cells. When the IGH dual colour, break apart rearrangement probe is hybridised to a nucleus containing an IgH translocation, one orange, one green, and one orange/green (yellow) fusion signal pattern is observed (1O1G1F).

2.12.4.2. BCL2 dual colour, break apart rearrangement probe

In a normal cell hybridised with the BCL2 dual colour, break apart rearrangement probe, the expected signal pattern is two orange/green (yellow) (2F) fusion signals. In a cell with a translocation breakpoint within the gap region between the LSI BCL2 5' and 3' probes, one orange, one green and one fusion pattern (1O1G1F) is observed. Some normal nuclei may appear to have a slight separation of the orange and green signals due to a gap between the hybridization targets of the two probes.

2.12.4.3. BCL6 dual colour, break apart rearrangement probe

In a normal cell hybridised with the BCL6 dual colour, break apart rearrangement probe, the expected signal pattern is two orange/green (yellow) (2F) fusion signals. In a cell with

a translocation breakpoint within the gap region between the LSI BCL6 5' and 3' probes, one orange, one green and one fusion pattern (1O1G1F) is observed. Some normal nuclei may appear to have a slight separation of the orange and green signals due to a gap between the hybridization targets of the two probes.

2.12.4.4. IGH/BCL2 Dual colour, dual fusion translocation probe

The expected pattern in a normal nucleus hybridised with the LSI IGH/BCL2 probe is a two orange, two green signal pattern (2O2G). In a nucleus harbouring a t(14;18), the most common pattern is one orange signal, one green signal (each representing the normal homolog) and two orange/green (yellow) fusion signals representing the two derivative chromosomes resulting from the reciprocal translocation (1O1G2F pattern). Patterns other than 1O1G2F may be observed in some abnormal cells including instances of nuclei containing more than two fusion signals.

2.12.5. Scoring slides

Slides were studied in random order and in a blinded fashion. The signal pattern of 100 qualifying interphase nuclei from different areas of the section was determined, avoiding damaged nuclei and clumps of cells with indistinct boundaries. Fusion signals were defined as merging (yellow) or adjacent orange and green signals. The scoring process was restricted to normal nuclei with 2F, or 2O2G, and abnormal nuclei with 1O1G1F, or 1O1G2F, for break apart rearrangement and dual fusion translocation probes, respectively. All slides were assessed independently by a second analyst (AS) and signal patterns were found to be in accordance with initial findings. Representative cells were captured using PowerGene MacProbe v4.3 (Perceptive Scientific International Ltd., Chester, UK).

Cut-off levels for false positive results arising from the moderate "accidental" juxtaposition of SpectrumOrange™ and SpectrumGreen™ probes in normal tissues were determined for each probe. Using tonsil tissue from two normal individuals, a total of 400 cells (100 cells per sample, two independent analysts) were scored as above. The cut-off level was calculated as the mean false positive rate plus 3 times standard deviations (mean ± 3SD): IGH break apart, 6.1%; BCL2 break apart, 2.2%; BCL6 break apart, 2.2%; IGH/BCL2 fusion, 4.1% (see Appendix 5.2. for calculation). Any population in the patient sample higher than the cut-off value was considered as a positive.

3. Follicular lymphoma

3.1. Introduction

FL is one of the more common B-cell NHLs in adults and accounts for ~40% of all newly diagnosed cases.[469] In the UK, the incidence of FL is approximately 4 per 100,000 persons, increasing by 5% per year. Although rare before the age of 25, frequency increases steadily with age in both men and women, with a median presentation at 50-60 years. At diagnosis, most patients present with disseminated disease (stage III/IV using the Ann Arbor classification), commonly with nodal involvement and often accompanied by splenic and BM disease.[470,471] Curative treatment is currently only possible in patients with early disease (stage I and II). For the majority of patients a cure is not possible with the current therapeutic options, such as chlorambucil or low intensity combined chemotherapy, and the rate of relapse remains unchanged over time, even in those patients who have achieved complete responses to treatment. However, more recently, the addition of the anti-CD20 mAb, Rituximab, to a number of different chemotherapy regimens has been shown to significantly increase progression-free survival.[472] Characteristically, FL follows an indolent course with the spontaneous growth and regression at disease sites over time. Despite the advanced stage of disease at diagnosis in the majority of patients, patients have a median survival of 7-8 years.[473-475]

In a significant number of FL cases, 40-70% over nine years, histological transformation to a more aggressive form, typically DLBCL, occurs.[476-478] Transformation can happen at any time irrespective of the stage at presentation or treatment regimens and is usually heralded by rapidly increasing lymph node enlargement, elevated serum lactate dehydrogenase and the development of systemic symptoms.[476,479,480] The median survival from the time of histological transformation is 18 months.[479]

3.1.1. Morphology

FL is histologically well differentiated. It is defined as a neoplasm of follicle centre B cells, where its nodular growth pattern clearly resembles the architecture of GCs in the secondary lymphoid organs. The neoplastic follicles contain both centroblastic and centrocytic tumour cells, although centrocytes predominate, as well as non-neoplastic, activated T cells and FDCs. As a tumour of GC origin, the $CD5^-CD23^-$ peripheral B cell of the inner follicle mantle is believed to be the normal counterpart.[286] The vast majority of FL express surface Ig and are typically $sIgM^+$, either with or without IgD, and less frequently $sIgG^+$ or $sIgA^+$.[481-484]

3.1.2. Molecular analysis

Analysis of the nucleotide sequences of Ig V genes can provide important information that relates to the cellular origin of B-cell malignancies and their clonal history and provide clues as to possible pathogenic mechanisms (see section 1.4.3.).[320] Analysis of the V_H genes of FL have further supported the their derivation from GC B cells. Several studies have addressed V_H gene usage, which appears similar to the normal B-cell repertoire, and have reported a high degree of SHM.[364-370] Another well documented feature is the presence of intraclonal heterogeneity, generally considered to reflect that the tumour cells remain under the influence of the somatic mutator in the GC post-transformation.[367-370] Finally, V_H isotype switching has been described, an event that is thought to take place normally in the centrocyte stage of the GC.[371] It is becoming appreciated that disease progression, including transformation to DLBCL, or treatment may have an impact on V_H gene sequences. A narrowing of intraclonal heterogeneity has been observed after treatment and suggests that a single cell has escaped from control and that further SHM does not occur at this stage.[365,366]

The environment of the GC includes FDCs and activated T cells.[481] The presence of the third key component Ag, which is required for activation of SHM through engagement of the BCR in normal B cells, has not been functionally demonstrated in FL. Evaluation of the patterns of somatic mutation in the V_H genes has shown preferential sequence maintenance in the FRs versus clustering of replacement mutations in the CDRs and this has been interpreted as evidence for antigen driven clonal selection.[365,366,368,485] However, it is now evident that there is a natural tendency for the CDRs to accumulate replacement mutations and therefore it remains unclear as to whether Ag has a role in influencing the behaviour of tumour cells in FL.[486] Moreover, model systems are being established in which SHM can occur without external Ag and it may be that for lymphomas arising in the GC, Ag is no longer an absolute requirement.[487] Nevertheless, the importance of sIg expression is highlighted by the fact that the majority of cases retain sIg expression even though one allele of chromosome 14 is disrupted by the t(14;18) translocation, a characteristic feature observed in 80-90% of FL.[293,294,488,489] This translocation is thought to occur early in B-cell development at the stage of D-J_H joining in the BM.[306] The breakpoints within IgH all fall within J_H, while the breakpoints on chromosome 18 are clustered within the 3' region of the *Bcl-2* gene; 50% of the breaks fall within a major breakpoint region, a 150bp region within the 3' untranslated region of the gene, another 25% fall in the minor cluster region and the remaining breaks generally fall within an intermediate cluster region midway between the major breakpoint region and minor cluster

region.[490] The result of this translocation is the deregulated expression of *Bcl-2* and an increase in the anti-apoptotic Bcl-2 protein. Therefore, t(14;18) may provide one mechanism by which tumour cells survive via suppression of apoptosis and appears to provide the first step towards lymphoma pathogenesis, although it is not sufficient for tumour development as it can also occur in the B cells of normal, healthy individuals.[291,491] Conservation of the sIg may suggest a selective advantage for the tumour cells in FL, but it is not clear if this in turn is dependent on stimulation by Ag.

3.1.3. Glycosylation of V_H genes in FL

Colleagues in this laboratory have previously shown at a genetic level that the V_H genes in FL are unusual in their acquisition of *N*-glycosylation sites, which are introduced through SHM.[12] Single chain variable fragments (scFv) derived from three FL patients were expressed in a yeast system and the analysis of the recombinant proteins showed that they had undergone glycosylation, which was susceptible to digestion with PNGase F. This observation prompted the systematic analysis of the V_H sequences for potential *N*-glycosylation sites using the motif Asn-X-Ser/Thr and led to the discovery that approximately 80% of FL V_H genes acquired *N*-glycosylation sites.[12] This high incidence has also been confirmed in a second study.[460] This is in contrast to V gene rearrangements of normal plasma and memory B cells, non-functional V gene sequences and sequences from MM and the mutated subset of CLL, all of which have a low incidence of glycosylation sites (~10%).[12] Novel *N*-glycosylation sites were also common in eBL (82%) and in 4/5 patients with Iranian BL.[13] Other GC lymphomas, such as DLBCL and sBL, show an intermediate frequency of ~40%.[12,13] Since *N*-glycosylation sites do not accumulate significantly in normal B cells, the positive selection of B cells in FL and other GC-associated tumours that contain *N*-glycosylation sites suggests a potential role for the added glycans in tumourigenesis.

A recent study in collaboration with colleagues from our laboratory has now characterised the glycosylation of six FL-derived Igs.[466] Analysis has revealed that they are unusual in that they contain oligomannoses in the V regions but complex glycans in the constant regions. This consistent differential glycosylation pattern indicates normal transit through the Golgi stacks but steric blockade of sugar processing in the V regions. It also suggests a tumour-related function for the expressed oligomannose, possibly via an interaction with mannose receptors expressed by stromal cells in the GC. Indeed, further investigation has demonstrated that the oligomannose glycans occupying the V regions are accessible to mannose-binding lectin (MBL).[466] Together, these findings support a potential

contribution to FL pathogenesis involving Ag-independent interaction of sIg of the BCR with MBLs present in the GC. The t(14;18) appears to be the first step towards lymphoma pathogenesis. The acquisition of *N*-glycosylation sites during SHM and the subsequent addition of oligomannoses appears to provide a second event.

3.1.4. Aims

The aim of this study was to further our understanding of the pathogenesis of follicular lymphoma. To achieve this we first sort to build a detailed and definitive picture of V gene segment usage and mutational patterns of tumour-derived V_H and V_L to gain a complete view of the BCR. Furthermore, we aimed to assess the introduction of *N*-glycosylation sites in both the heavy and light chains at the nucleotide level. Secondly, we wished to confirm these findings at the protein level and begin to investigate further the nature of the added glycans. By incorporating the results of previously published studies into our analysis we hope to achieve a better understanding of this intriguing malignancy.

3.2. Results

3.2.1. Tumour-derived V gene usage

Using cDNA isolated from biopsy material as a source, we identified the matched V_H and V_L tumour-derived gene sequences in 33/44 (75.0%) cases (Table 3.1.). In 8 further cases we were able to identify V_L, but not V_H, and in 3 cases V_H, but not V_L; for the cases where one chain could not be identified, multiple, nonclonal PCR products were detected and most probably derived from polyclonal reactive B cells infiltrating these tumours. Since we were able to identify a clonal gene sequence for the partner chain in all 11 cases failure is unlikely to be due to low numbers of tumour cells in the population. It is most probably a consequence of somatic mutations within the regions to which the PCR primers used in this study were designed to hybridise leading, therefore, to a lower amplification efficiency and false-negative results. The higher failure rate for tumour-derived V_H gene identification compared to V_L is likely due to the intrinsically lower rates of SHM of V_L compared to V_H genes that has been observed in this study and others.[368,492]

3.2.1.1. V_H gene usage

In total, 36 potentially functional, clonal V_H gene sequences were identified (Table 3.1.). Fifteen (29.4%) of the 51 known functional V_H gene segments were found rearranged in our cohort. They derived from 4 of the 7 human V_H gene families in the following distribution: V_H3, 77.8% (n=28); V_H4, 16.7% (n=6); V_H2 and V_H7, both 2.8% (n=1). V_H segments from the V_H1, V_H5 and V_H6 families were not observed. In terms of V_H family usage, the distribution of our cases largely reflected the available germline repertoire, with some exceptions. Notably, no V_H1 family gene segments were found rearranged in our cohort. In view of the germline size of the V_H1 gene family, which contributes 11 (21.6%) gene segments to the germline repertoire, V_H1 appeared vastly under represented. Conversely, the V_H3 family, which contributes 22 of 51 (43.1%) gene segments to the repertoire, was over represented. To gain a broader view, we have compared our series with 226 cases from six published studies and similar trends are evident (Figure 3.1.).[364,367-370,460]

Individual, smaller studies have suggested that the V_H family usage in FL parallels that of the normal B-cell repertoire.[364,367,369,370] Using this extended series of 262 cases, we could now undertake a meaningful comparison to the V_H family usage in normal B cells from healthy adult donors. We can confirm that with minor exceptions this observation holds true (Figure 3.2.). Similarly, at the level of individual V_H3 and V_H4 gene segment usage,

Table 3.1. V_H and V_L gene usage in FL

	V_H					V_L			
Case	**V_H Family**	**V gene***	**D gene***	**J_H gene***	**CDR3 length (bp)**	**V_L Family**	**V gene***	**J_L gene***	**CDR3 length (bp)**
FL1	V_H3	V3-23	D6-19	J_H5a	39	Vκ1	DPK4/A20	Jκ1	27
FL2	V_H3	V3-48	-	J_H6b	45	Vκ1	A30/SG3	Jκ1	27
FL3	V_H3	V3-15	-	J_H6b	60	Vκ1	A30/SG3	Jκ4	27
FL4	V_H4	V4-59	-	J_H4b	30	Vλ1	DPL8/1e	Jλ2	33
FL5	V_H3	V3-64	-	J_H4b	48	Vκ4	DPK24/B3	Jκ1	24
FL6	V_H3	V3-48	-	J_H3a	30	Vκ3	DPK22/A27	Jκ1	24
FL7**	V_H3	V3-48	-	J_H5a	27	Vκ3	Vg38K/L6	Jκ4	24
						Vκ1	HK102/L12a	Jκ1	27
FL8	V_H3	V3-23	-	J_H4b	33	Vκ3	Vg38K/L6	Jκ4	15
FL9						Vκ4	DPK24/B3	Jκ1	27
FL10	V_H3	V3-48	-	J_H3b	39	Vκ4	DPK24/B3	Jκ4	27
FL11	V_H3	V3-21	-	J_H4b	39	Vλ1	DPL2/1c	Jλ2	36
FL12**	V_H3	V3-15	-	J_H4b	15	Vκ1	A30/SG3	Jκ2	21
						Vκ4	DPK24/B3	Jκ1	27
FL13	V_H4	V4-61	-	J_H6c	24	Vκ3	Vg38K/L6	Jκ3	27
FL14	V_H4	V4-34	-	J_H1	45	Vκ4	DPK24/B3	Jκ2	27
FL15**	V_H3	V3-48	-	J_H4b	36	Vκ3	DPK22/A27	Jκ1	33
						Vκ3	DPK22/A27	Jκ1	30
FL16	V_H3	V3-49	D2-02	J_H6b	42	Vκ1	DPK9/012	Jκ1	27
FL17	V_H3	V3-23	-	J_H5a	39	Vκ4	DPK24/B3	Jκ1	27
FL18	V_H3	V3-23	-	J_H4b	33				
FL19						Vλ2	DPL10/2b2	Jλ3b	30
FL20	V_H4	V4-39	-	J_H5b	21	Vκ6	DPK26/A26	Jκ2	24
FL21	V_H3	V3-48	-	J_H3b	45	Vκ1	DPK9/012	Jκ1	33
FL22	V_H3	V3-48	-	J_H3a	18	Vκ2	DPK18/A17	Jκ3	27
FL23						Vλ1	DPL8/1e	Jλ3b	39
FL24	V_H3	V3-11	-	J_H6c	45	Vκ3	Vg38K/L6	Jκ4	15
FL25	V_H4	V4-34	-	J_H2	33	Vκ3	DPK22/A27	Jκ2	33
FL26	V_H3	V3-64	-	J_H4b	27	Vκ4	DPK24/B3	Jκ1	27
FL27	V_H3	V3-48	D2-15	J_H4b	33	Vκ1	DPK9/012	Jκ2	24
FL28	V_H3	V3-30	D3-10	J_H4b	54	Vκ1	DPK9/012	Jκ2	30
FL29	V_H3	V3-48	D5-24	J_H4a	30	Vκ1	DPK9/012	Jκ3	27
FL30						Vκ3	DPK20/A11	Jκ2	27
FL31	V_H3	V3-07	-	J_H4b	27				
FL32	V_H3	V3-48	D3-03	J_H4b	39	Vκ3	DPK22/A27	Jκ1	27
FL33	V_H4	V4-04	-	J_H6a	57	Vκ1	DPK4/A20	Jκ2	27
FL34	V_H3	V3-48	-	J_H4b	24	Vκ3	DPK21/L2	Jκ4	24
FL35	V_H2	V2-05	-	J_H5b	27	Vκ1	HK102/L12a	Jκ1	24
FL36						Vλ1	DPL5/1b	Jλ1	27
FL37						Vκ2	DPK12/A2	Jκ4	27
FL38	V_H3	V3-23	D3-22	J_H4b	27	Vλ1	DPL8/1e	Jλ3b	33
FL39	V_H3	V3-23	-	J_H3b	45	Vλ2	DPL11/2a2	Jλ2	33
FL40						Vκ1	A30/SG3	Jκ5	27
FL41	V_H3	V3-23	-	J_H6b	18	Vκ1	A30/SG3	Jκ4	27
FL42						Vκ1	LFK431/L1	Jκ4	27
FL43	V_H3	V3-23	-	J_H5b	30	Vκ4	DPK24/B3	Jκ1	27
FL44	V_H7	V7-04.1	-	J_H4b	27				

*; Germline V, D and J genes with the greatest homology.

**; Two potentially functional clonal V_L rearrangements were identified.

-; Germline D gene could not be confidently assigned.

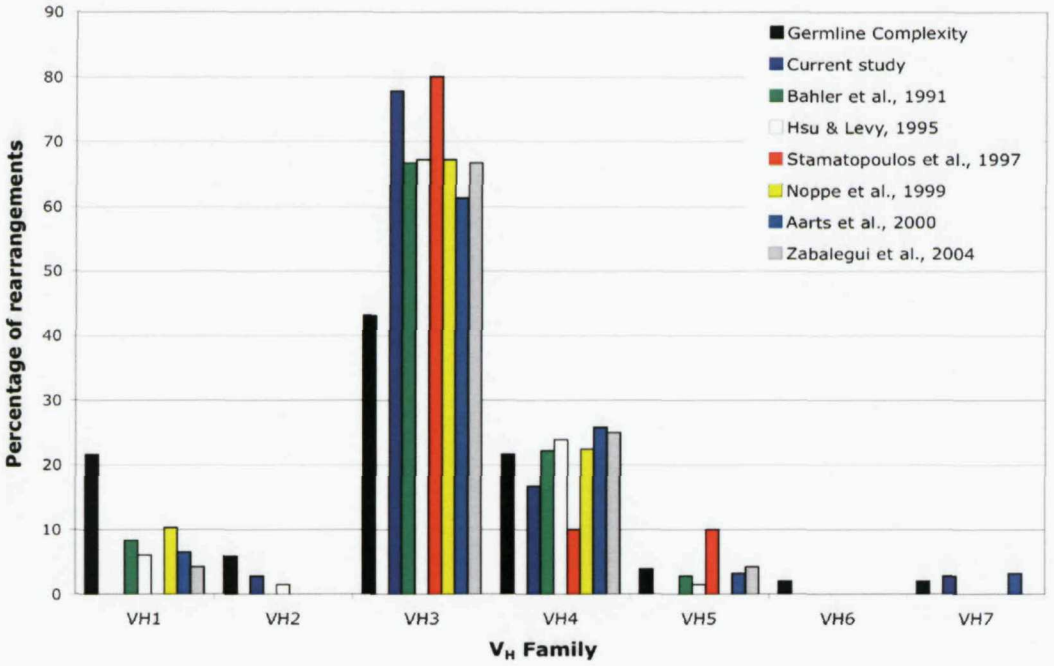


Figure 3.1. V_H gene family usage in FL. The V_H gene family usage for our 36 FL was analysed and compared to findings reported in previous studies for 226 FL: Bahler *et al.*, 1991, Hsu *et al.*, 1995, Stamatopoulos *et al.*, 1997, Noppe *et al.*, 1999, Aarts *et al.*, 2000, and Zabalegui *et al.*, 2004.[364,367-370,460] Values for each V_H family are expressed as a percentage of the total rearrangements assessed.

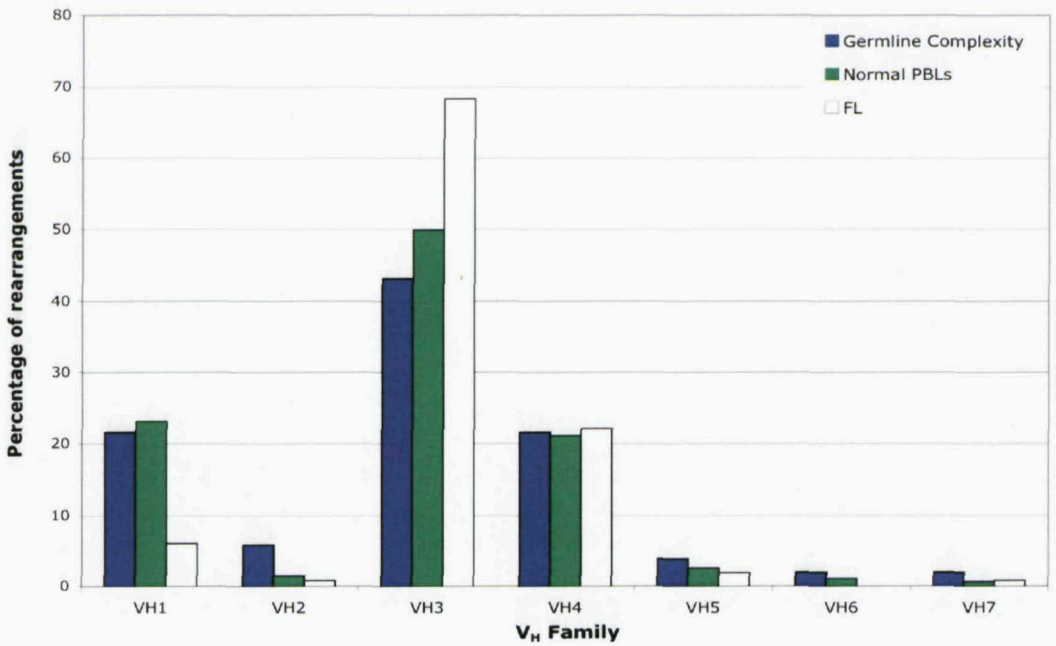


Figure 3.2. V_H gene family usage in FL compared to normal B cells. The V_H gene family usage was analysed for 262 FL and 1322 normal, healthy PBLs. Values for each V_H family are expressed as a percentage of the total rearrangements assessed. Analysis of FL included the current study and previously published data: Bahler *et al.*, 1991, Hsu *et al.*, 1995, Stamatopoulos *et al.*, 1997, Noppe *et al.*, 1999, Aarts *et al.*, 2000, and Zabalegui *et al.*, 2004.[364,367-370,460] Analysis of normal, healthy PBLs included data from Brezinschek *et al.*, 1995, Brezinschek *et al.*, 1997, Huang *et al.*, 1992, and Demaison *et al.*, 1995.[97,493-495]

a comparison has been made with the normal B-cell repertoire and with four of the published studies where individual V_H gene data for FL was available (Figure 3.3.).[97,364,368-370,460,493-495] Some asymmetries were evident. The most commonly used gene segments in our cohort were V3-48 (n=11) and V3-23 (n=8), which accounted for 52.8% of rearrangements and likely contributed to the over representation of the V_H3 family. Whereas the biased usage of the V3-23 gene segment is also found in normal B cells, the V3-48 gene segment, which was the most frequently used in our study (30.6%), is not commonly rearranged in normal B cells (<2%). A similar bias was also noted in two other FL studies, but was not observed in two further studies.[368-370,460] With this exception, the V_H gene usage in FL appears to reflect that of the normal B-cell repertoire.[97,271] J_H usage in FL is also similar to that observed in normal B cells: J_H4, 44.4% (n=16); J_H6, 19.4% (n=7); J_H5, 16.7% (n=6); J_H3, 13.8% (n=5); J_H1 and J_H2, both 2.8% (n=1).[94,97,493-495] This finding is in keeping with other studies of FL.[368-370]

The CDR3 region varied in length from 15 to 60 nucleotides (mean 34.8bp, median 33bp), according to the Kabat numbering system.[496] Assignment of the D gene segments was based on the homology between the CDR3 region and the germline D genes according to the rule proposed by Corbett *et al.*, whereby at least 10 consecutive nucleotides of identity are required for confident allocation.[20] Under these stringent conditions, D segments could only be identified in 7 (19.4%) FL. The 7 genes derived from the D segment families D3 (n=3), D2 (n=2), D5 (n=1) and D6 (n=1). A low incidence of D segment allocation has been documented previously in FL, increasing as the criteria for assigning sequences become less stringent and less reliable.[369,370] In contrast, confident assignment is possible in ~90% of normal B cell where D segments from the D3, D6 and D2 families are known to be preferentially recombined.[20,94,493,494]

3.2.1.2. V_L gene usage

Forty-four potentially functional, clonal V_L sequences were identified from 41 cases of FL (Table 3.1.); two clonal V_L sequences were found in 3 (7.3%) tumours. Of these, 37/44 (84.1%) used a κ chain and 7/44 (15.9%) used a λ chain. The majority (94.6%) of the 37 κ rearrangements were derived from the proximal region and used 5 of the 6 human Vκ gene families in the following distribution: Vκ1, 40.5% (n=15); Vκ3, 29.8% (n=11); Vκ4, 21.6% (n=8); Vκ2, 5.4% (n=2); Vκ6, 2.7% (n=1). In view of the germline size of the Vκ2 gene family, which contributes 9 (22.5%) of 40 gene segments to the germline repertoire, the Vκ2 family appears under represented in FL. Conversely, the Vκ3 and Vκ4 families, which contribute 7 (17.5%) and 1 (2.5%) gene segments to the repertoire,

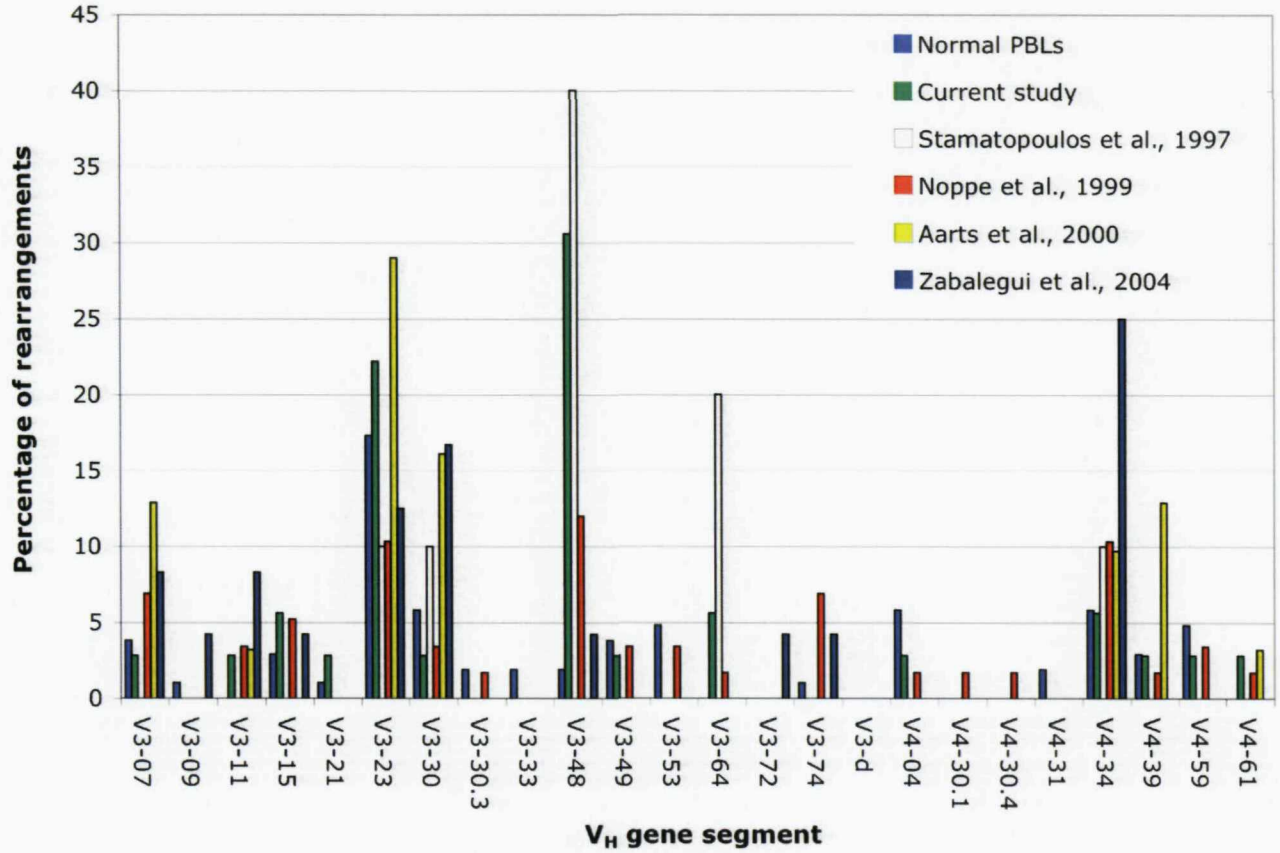


Figure 3.3. V_H3 and V_H4 gene segment usage in FL compared to normal B cells. The individual V_H gene usage for the V_H3 and V_H4 families was analysed for 159 FL and 104 normal, healthy PBLs. Values for each V_H segment are expressed as a percentage of the total rearrangements assessed. Those segments not observed in productive rearrangements in either FL or normal PBLs have not been represented (V3-13, V3-20, V3-30.5, V3-43, V3-66, V3-73, V4-28, V4-30.2 and V4-b). Analysis of FL included data from the current study and previously published data: Stamatopoulos *et al.*, 1997, Noppe *et al.*, 1999, Aarts *et al.*, 2000, and Zabalegui *et al.*, 2004.[368-370,460] Analysis of normal, healthy PBLs included data from Brezinschek *et al.*, 1995, and Demaison *et al.*, 1995.[494,495]

respectively, were over represented; an over representation of the Vκ3 family is also observed in normal B cells. The usage of Vκ1, Vκ5 and Vκ6 gene families appears to be in accordance with germline complexity (Figure 3.4.). Based on our cohort, Vκ gene family usage shows many similarities to that found for normal B cells, although the under representation of the Vκ2 family and over representation of the Vκ4 appears more pronounced in FL (Figure 3.4.).[48,75,105] A previous small study of Igκ$^+$ FL found Vκ1 family genes in 9 of 10 cases, we do not confirm this strong bias in our series.[368]

In our cohort, the Vκ3 and Vκ4 gene families were over represented largely as a result of the preferential usage of a small number of specific individual family members. Most frequently rearranged gene segments were DPK24/B3 (n=8), DPK22/A27 (n=5), A30/SG3 (n=5), DPK9/02-012 (n=5) and Vg38K/L6 (n=4), which together accounted for 72.3% of κ sequences. With the exception of A30/SG3, these genes are also found at a higher frequency in normal B cells.[48,75,105] The 37 κ rearrangements used all 5 Jκ gene segments in the following distribution: Jκ1, 43.2% (n=16); Jκ4, 24.3% (n=9); Jκ2, 21.6% (n=8); Jκ3, 8.1% (n=3) and Jκ5, 2.7% (n=1). The Jκ gene usage appears similar to that observed in normal B cells.[48,75,497]

The 7 Vλ sequences rearranged genes from the Vλ1 (n=5) and Vλ2 (n=2) families, both of which are found more commonly rearranged in the productive Vλ repertoire.[76,273] A single gene, DPL8/1e, was donor for 42.3% (n=3) of λ rearrangements. The DPL8/1e gene segment and the other 4 genes utilised here (DPL2/1c, DPL10/2b2, DPL5/1b and DPL11/2a2) have all been reported to be preferentially rearranged in normal B cells.[76,273] The 7 λ sequences used the Jλ genes Jλ2(/3a) and Jλ3(b) (both n=3) and Jλ1 (n=1). Of note, it has been reported that the Jλ2 and Jλ3 genes are used most frequently in normal B cells, followed by Jλ1.[76,273]

Analysis of all 44 functional V_L sequences revealed that the CDR3 region varied in length from 15 to 39 nucleotides (mean 27.3bp, median 27bp). Independent assessment of κ and λ rearrangements showed that the CDR3 length of the rearranged Vλ genes was 2 amino acids longer (mean 11aa) compared to Vκ rearrangements (mean 9aa), consistent with studies of κ and λ rearrangements in normal B cells other B-cell malignancies (see sections 4.2.1.2. and 5.2.1.1.).[48,75,76] This may be a result of differences in TdT activity influencing N additions at VκJκ and VλJλ joins. However, before rearrangement, the potential Vκ contribution to the CDR3 is 21 nucleotides whereas the potential Vλ

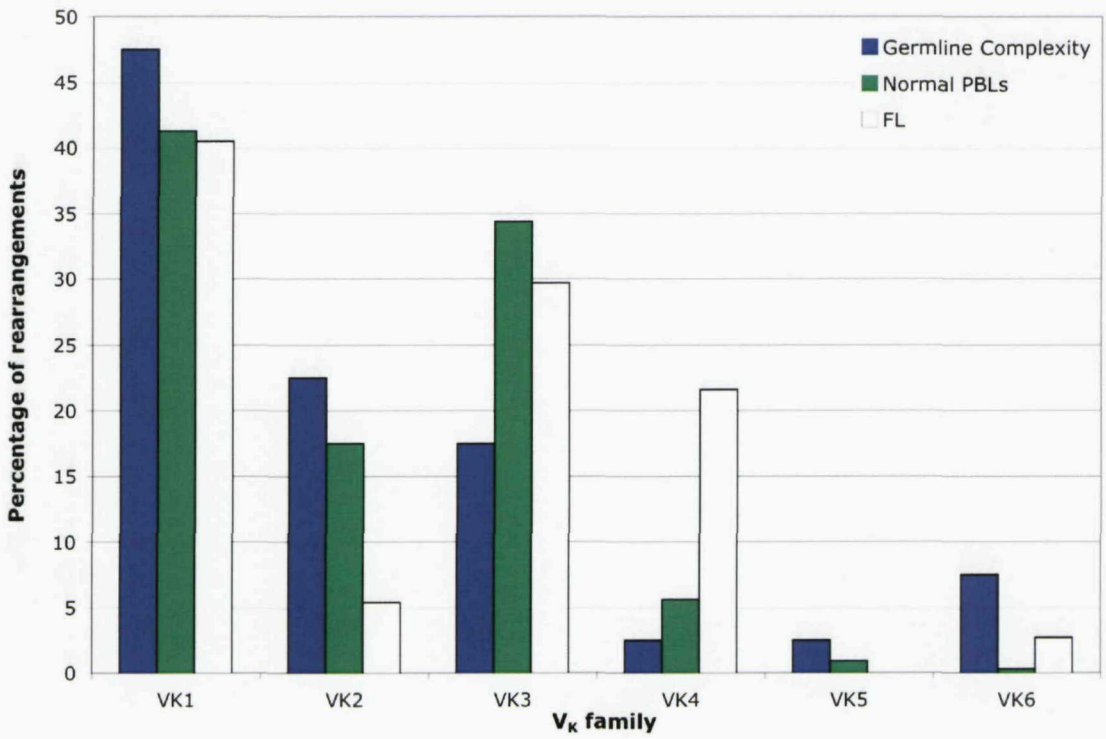


Figure 3.4. Vκ gene family usage in FL compared to normal B cells. The Vκ gene family usage was analysed for 37 FL and 950 normal, healthy PBLs. Values for each Vκ family are expressed as a percentage of the total rearrangements assessed. Analysis of normal, healthy PBLs included data from Cox *et al.*, 1994, Klein *et al.*, 1993, Foster *et al.*, 1997, and Juul *et al.*, 1997.[48,75,105,497]

contribution is 24-27 nucleotides. Thus, the increase in length of the CDR3 of VλJλ rearrangements could reflect the increased contribution of the Vλ gene segments to the CDR3 region.

Within the matched pairs of V_H and V_L there was no detectable tendency for selective combinations. For example, the V3-48 gene was found in combination with Vκ3, Vκ4 and Vκ1 family gene segments (Table 3.1.).

3.2.2. Somatic hypermutation in tumour-derived V genes

Isotype designation and mutational status are shown in Table 3.2. All V genes, with the exception of a single Vκ sequence (FL12), which could be a secondary rearrangement, showed evidence of somatic mutation (>2%). This is in contrast to that of normal PBLs where we see a lower number of B cells with mutated V genes, 40% and 23% of V_H and Vκ genes, respectively; this disparity is expected since the population of circulating PBLs in normal, healthy adult donors consists of both Ag-experienced, activated B cells and naïve B cells.[48,97] A wide range mutation levels were observed in FL: V_H, 1.0-19.4% (mean 10.6%); V_L, 0.4-16.5% (mean 6.0%). In normal PBLs, a mutation level of greater than 6% is seen in just 13.7% and 4.7% of V_H and V_L genes, respectively.[48,97] This is much higher in FL where we see ≥6% mutation in 94.4% V_H genes and 40.9% V_L genes.

No duplications or insertions were observed in our cohort of 80 V genes. However, one example of a single codon deletion was observed within the CDR1 of a Vκ gene.

In 28/36 (77.8%) V_H genes, the Ig isotype was determined by FACS and/or PCR and saw 16 IgM^+ FL, 11 IgG^+ FL and 1 IgA^+ FL. In contrast to the previous study by Aarts *et al.*, we did not observe a significant difference in the levels of somatic mutation in IgM-expressing versus isotype switched FL (P = 0.717; Figure 3.5a.).[370] Similarly, there was no significant difference between the mutation level found in κ (mean 6.0%) versus λ (mean 6.2%) rearrangements (P = 0.830; Figure 3.5b.). In the 3 cases with dual light chain rearrangements, one rearrangement showed more SHM than the other: 12.9% vs 3.0%, 6.1% vs unmutated and 8.2% vs 3.4%. In 30/33 (90.9%) cases where both tumour-derived V_H and V_L were identified, mutation rates were lower in the V_L gene compared to their partner V_H and this difference was statistically significant ($p<0.001$) in keeping with findings in FL and normal B cells (Figure 3.6.).[368,492]

Table 3.2. An analysis of somatic mutation in the V genes of FL

Case	V_H				V_L			
	V_H Isotype*	Mutation (%)≠	Intraclonal heterogeneity No. clones°	%	V_L Isotype*	Mutation (%)≠	Intraclonal heterogeneity No. clones°	%
FL1	μ	7.5	3/9	0.101	κ	4.5	2/9	0.114^R
FL2	μ	9.2	9/14	0.358	κ	3.4	12/20	0.308^R
FL3	-	11.0	3/7	0.187	κ	3.8	8/16	0.342^R
FL4	γ	13.7	na	na	λ	4.4	na	na
FL5	-	12.6	6/9	0.744^R	κ	2.8	6/9	0.390
FL6	-	11.6	na	na	κ	6.0	na	na
FL7	γ	2.0	10/13	0.362^R	κ	3.0	7/10	0.519^R
					κ	12.9	4/4	1.636^R
FL8	μ	8.2	10/10	1.682^R	κ	5.3	18/18	2.421^R
FL9					κ	6.7	12/15	1.013^R
FL10	μ	8.2	6/7	1.073^R	κ	4.6	3/6	0.095^R
FL11	μ	10.5	8/8	1.917^R	λ	5.6	6/7	1.665^R
FL12	γ	8.0	15/20	0.546^R	κ	6.1	7/10	1.076^R
					κ	0.0	0/17	0.000
FL13	μ	12.5	5/8	0.255	κ	4.5	1/7	0.046
FL14	α	8.8	12/14	1.661^R	κ	2.8	19/19	1.280^R
FL15	μ	6.5	10/11	0.670^R	κ	8.2	10/14	1.001^R
					κ	3.4#	6/6	0.577^R
FL16	γ	10.5	7/8	2.197^R	κ	13.6	10/10	2.131^R
FL17	μ	15.3	4/6	0.399^R	κ	6.7	7/8	1.612^R
FL18	-	12.9	5/12	0.153^R				
FL19					λ	5.9	17/17	4.052^R
FL20	γ	12.7	4/8	0.272^R	κ	2.7	7/14	0.297^R
FL21	μ	9.9	13/15	0.893^R	κ	9.1	24/26	1.336^R
FL22	μ	11.9	7/8	0.559^R	κ	1.4	6/6	0.505^R
FL23					λ	7.8	8/10	0.900^R
FL24	μ	9.3	4/6	0.441^R	κ	4.9	5/9	0.464^R
FL25	μ	17.2	4/5	1.438^R	κ	12.0	3/4	0.337
FL26	-	15.6	7/7	1.113^R	κ	6.7	2/4	0.647^R
FL27	γ	10.5	na	na	κ	8.3	na	na
FL28	γ	12.9	6/10	0.856^R	κ	3.8	5/8	0.171^R
FL29	μ	6.8	7/10	0.405^R	κ	15.5	1/4	0.086
FL30					κ	16.5	5/10	0.238^R
FL31	μ	8.5	4/6	1.249^R				
FL32	-	1.0	9/12	0.883^R	κ	1.1	3/5	0.742^R
FL33	μ	11.7	7/11	0.366^R	κ	0.4	7/10	0.447^R
FL34	μ	19.4	6/6	0.799^R	κ	4.5	10/11	0.535^R
FL35	γ	9.3	16/17	0.933^R	κ	8.0	10/10	1.640^R
FL36					λ	7.9	9/10	1.021^R
FL37					κ	5.0	4/7	0.213^R
FL38	γ	9.2	11/15	0.600^R	λ	3.7	9/10	0.987^R
FL39	γ	15.3	12/12	2.630^R	λ	8.1	11/11	2.069^R
FL40					κ	1.1	5/14	0.250
FL41	-	12.9	11/11	0.581^R	κ	1.9	3/12	0.256
FL42					κ	11.4	8/8	0.985^R
FL43	γ	10.5	11/11	4.020^R	κ	3.9	20/23	0.991^R
FL44	-	6.9	12/19	0.574^R				

*; Heavy and light chain isotype determined by FACS and/ or PCR.

≠; Frequency of somatic hypermutation compared to the most homologous germline V gene segment.

°; Number of clones exhibiting sequence variation from the tumour consensus sequence / total number clones sequenced.

R; Includes at least one intraclonal sequence mutation that is repeated in >1 clone.

-; Indicates isotype not determined.

na; Not assessed.

#; Rearrangement exhibits a single codon deletion at amino acid position 30.

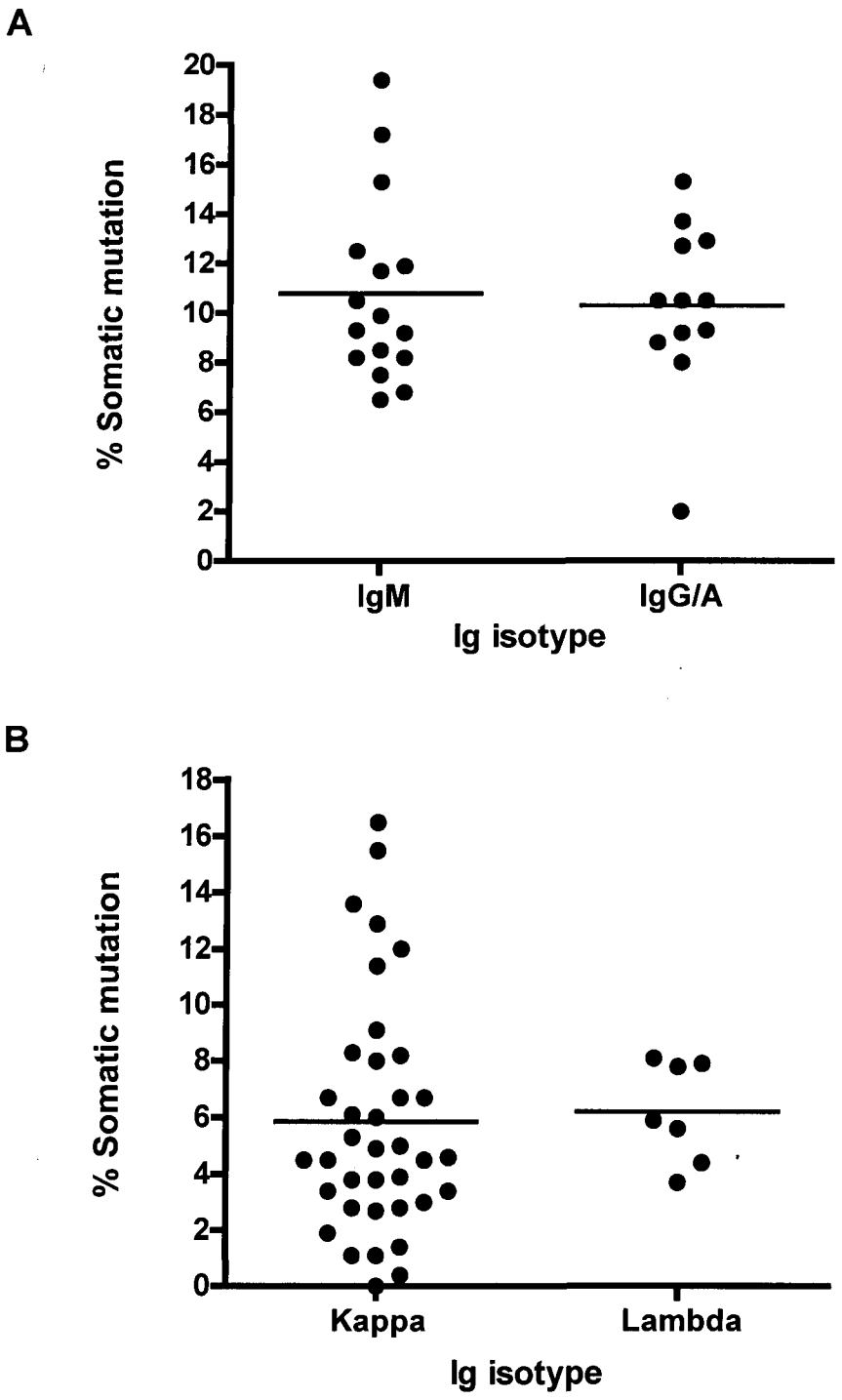


Figure 3.5. Analysis of somatic mutation load and Ig isotype in FL. A) The percentage somatic mutation in IgM versus isotype switched FL (IgG or IgA). An individual spot represents one V_H gene rearrangement. The lines represent the average percentage mutation of IgM^+ FLs (n=16) and isotype-switched FLs (n=12), respectively. P value = 0.717, by unpaired, two-tailed t-test (95% confidence interval). B) The percentage somatic mutation in Igκ versus Igλ FL. An individual spot represents one V_L gene rearrangement. The lines represent the average percentage mutation of κ^+ FL (n=37) and λ^+ FLs (n=7), respectively. P value = 0.830, by unpaired, two-tailed t-test (95% confidence interval).

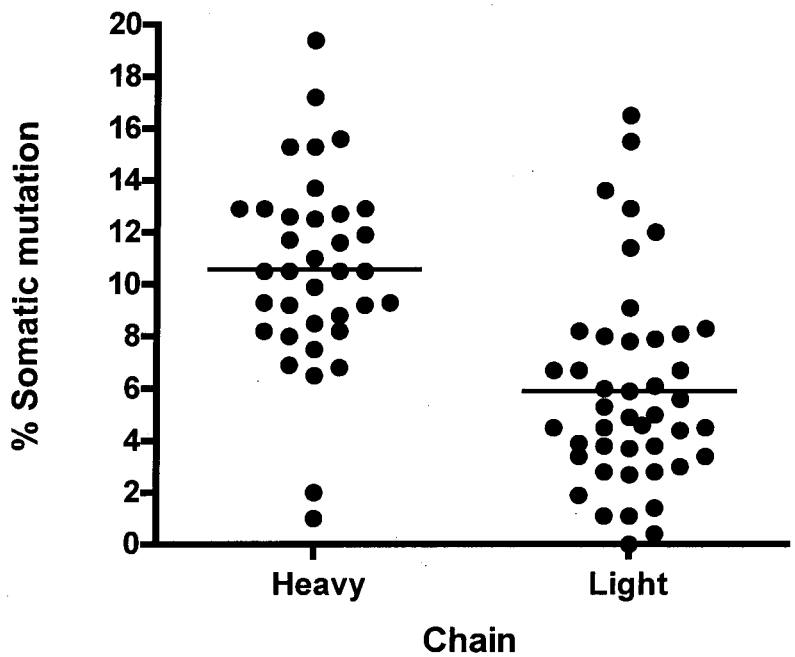


Figure 3.6. Analysis of somatic mutation load and Ig chain in FL. The percentage somatic mutation in the heavy chain versus the light chain of FL. An individual spot represents one V gene rearrangement. The lines represent the average percentage mutation of V_H (n=36) and V_L (n=44), respectively. $P < 0.0001$, by unpaired, two-tailed t-test (95% confidence interval).

Intraclonal heterogeneity was assessed in independent clones of 41 FL, including both the V_H and V_L tumour-derived genes where available (Table 3.2.). With the exception of two Vκ sequences, one of which was unmutated, all V gene sequences demonstrated levels of intraclonal heterogeneity that exceeded that of a previously determined taq error rate of 0.05%, indicating ongoing SHM. The level of intraclonal heterogeneity varied widely for V_H and V_L genes: V_H 0.101-2.630% (mean 0.845%); V_L 0.086-4.052% (0.959%). Intraclonal heterogeneity was also observed in sequences that had low overall levels of mutation, such as the V_H of FL32 (1.0%), thus verifying mutational activity.

3.2.3. Incidence of *N*-glycosylation sites in tumour-derived V genes

3.2.3.1. Incidence of novel *N*-glycosylation sites

V_H and V_L sequences were examined for potential *N*-glycosylation sites with the motif Asn-X-Ser/Thr, where X could be any aa except Pro (Table 3.3.). Disfavoured motifs using Asp (D), Glu (E), Leu (L) or Trp (W) as the intervening amino acid X, when followed by Ser (S), were considered viable in our analysis.[463] In 29/36 (80.5%) V_H sequences and 20/44 (45.5%) V_L sequences novel *N*-glycosylation sites had been acquired. Of the 33 tumours where both the V_H and V_L gene sequences were identified, 31 (93.9%) had at least one site in the tumour clone. Of these, 13 had sites in both V_H and V_L, 15 in V_H alone and 3 in V_L alone. The number of sites per V_H and V_L sequence is shown in Figure 3.7. and sees that the majority had single sites.

Two cases (FL32 and FL39) lacked a novel *N*-glycosylation site in the consensus sequences of V_H and V_L. A review of the histology and clinical data of each case confirmed the diagnosis of typical FL. The first case (FL32) exhibited an uncharacteristically low level of SHM: 1.0% (3bp) and 1.1% (3bp) in V_H and V_L, respectively. Thus, in this case, the likelihood of acquiring a site is limited. In the second case (FL39) we observed *N*-glycosylation sites in the CDR3 of 3/12 V_H molecular clones and in the CDR1 and CDR3 of 1/11 and 3/11 V_L molecular clones, respectively (Figure 3.8.). This finding could be interpreted in two ways: (i) *N*-glycosylation motifs are accumulating as a result of continued somatic mutation, or (ii) *N*-glycosylation motifs are being lost by the consensus sequence as a result of continued somatic mutation. Support for the former was provided when we assessed the conservation of *N*-glycosylation sites among molecular clones (Table 3.4.). Data from 30 FL cases with matched V_H and V_L shows that *N*-glycosylation sites were rarely lost during ongoing somatic mutation, with at least one motif conserved by all clones in 26 (86.7%) cases. In the remaining 4 cases, which each featured a single *N*-glycosylation site, motifs were conserved in 17/18, 6/7,

Table 3.3. The incidence of *N*-glycosylation sites in the V genes of FL

	V_H			V_L		
Case	**No. Sites**	**Region**	**Motif**	**No. Sites**	**Region**	**Motif**
FL1	0	-	-	1	CDR3	NTS
FL2*	2	CDR1, CDR2	NMS, NSS	0	-	-
FL3	1	CDR3	NCS	1	CDR3	NNS
FL4*	1	CDR2	NIS	1	FR4	NLT
FL5	0	-	-	1	CDR3	NYT
FL6*	1	CDR2	NIT	0	-	-
FL7*	1	CDR2	NIS	1	FR3	NDS
				1	FR1	NVT
FL8	1	CDR2	NIT	0	-	-
FL9				0	-	-
FL10	1	CDR2	NIS	1	FR3	NGS
FL11*	1	CDR3	NST	0	-	-
FL12	1	CDR2	NKS	0	-	-
				0	-	-
FL13	1	CDR2	NVS	0	-	-
FL14*	2	CDR1, CDR2^{GL}	NWT, NHS	0	-	-
FL15*	1	CDR3	NIS	0	-	-
				0	-	-
FL16	2	FR3, FR4	NDS, NIT	1	CDR3	NHT
FL17	3	CDR2x2, CDR3	NGT, NYT, NYT	1	CDR1	NLT
FL18	0	-	-			
FL19				0	-	-
FL20	1	CDR2	NVS	1	CDR3	NFT
FL21*	4	FR1, CDR1, CDR2, CDR3	NFT, NMS, NTS, NVT	1	CDR1	NIS
FL22	1	CDR2	NIS	1	FR3	NGS
FL23				1	CDR1	NSS
FL24	2	CDR1, CDR2	NMS, NIT	1	FR1	NLS
FL25*	1*	CDR2	NYS	1	FR1	NAT
FL26*	1	CDR3	NHS	0	-	-
FL27	1	CDR2	NIS	0	-	-
FL28	0	-	-	1	CDR3	NYS
FL29*	2	CDR2, FR3	NIS, NNS	1	CDR3	NYS
FL30				0	-	-
FL31	1	CDR2	NIT			
FL32	0	-	-	0	-	-
FL33	4	CDR2x2, CDR1, CDR3	NGS, NYT, NWS, NKT	0	-	-
FL34	2	CDR2, FR4	NIS, NVS	1	CDR1	NLT
FL35	2	FR1, CDR3	NFS, NSS	2	FR1, FR3	NIT, NLT
FL36				1	CDR3	NSS
FL37				0	-	-
FL38	1	CDR3	NTS	0	-	-
FL39	0	-	-	0	-	-
FL40				0	-	-
FL41*	1	CDR3	NIS	0	-	-
FL42				0	-	-
FL43	1	CDR3	NTS	0	-	-
FL44	0	-	-			

*; Functional glycosylation demonstrated by expression as scFv in 293F cells.
GL; Identifies a germline V4-34 *N*-glycosylation site.
*; Loss of germline V4-34 *N*-glycosylation site by somatic mutation.
Underlining indicates a disfavoured *N*-glycosylation motif for which functional glycosylation is questionable.

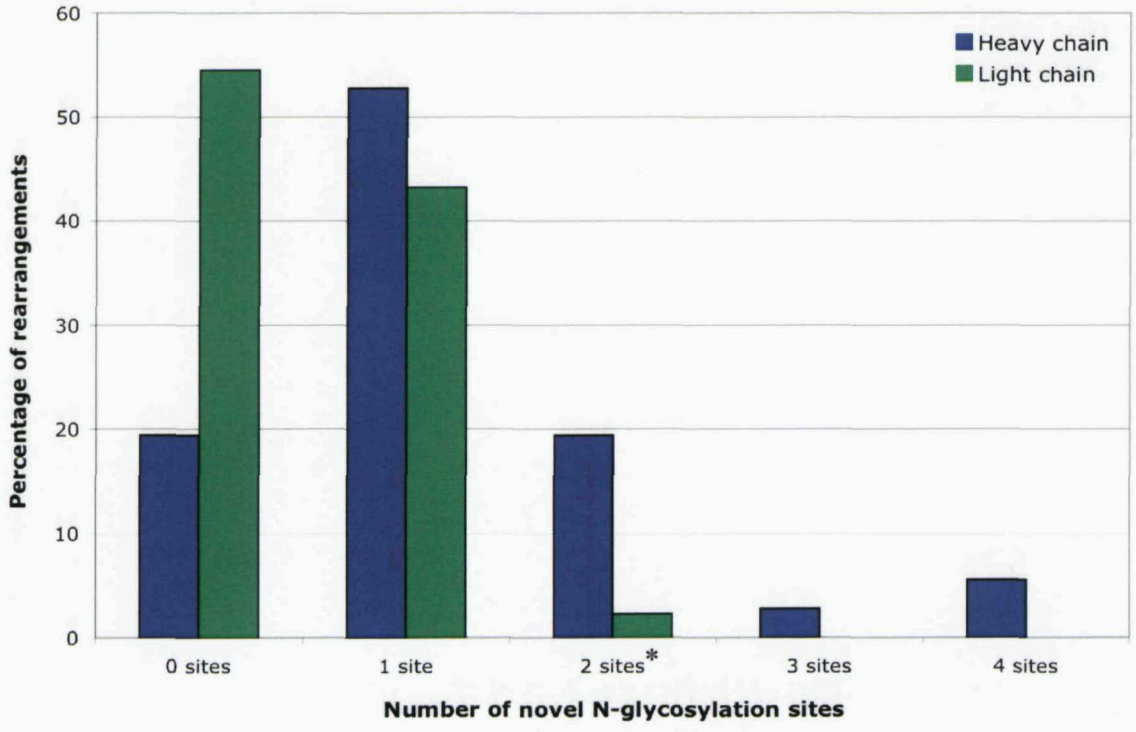


Figure 3.7. The incidence of novel *N*-glycosylation sites in the V region genes of FL. The incidence of novel *N*-glycosylation sites within the V_H and V_L gene sequences was analysed (FR1 To FR4) for 36 and 41 FL, respectively. Values are expressed as a percentage of the total number of rearrangements. *Includes a V_H gene sequence with a single novel site and the germline V4-34 site.

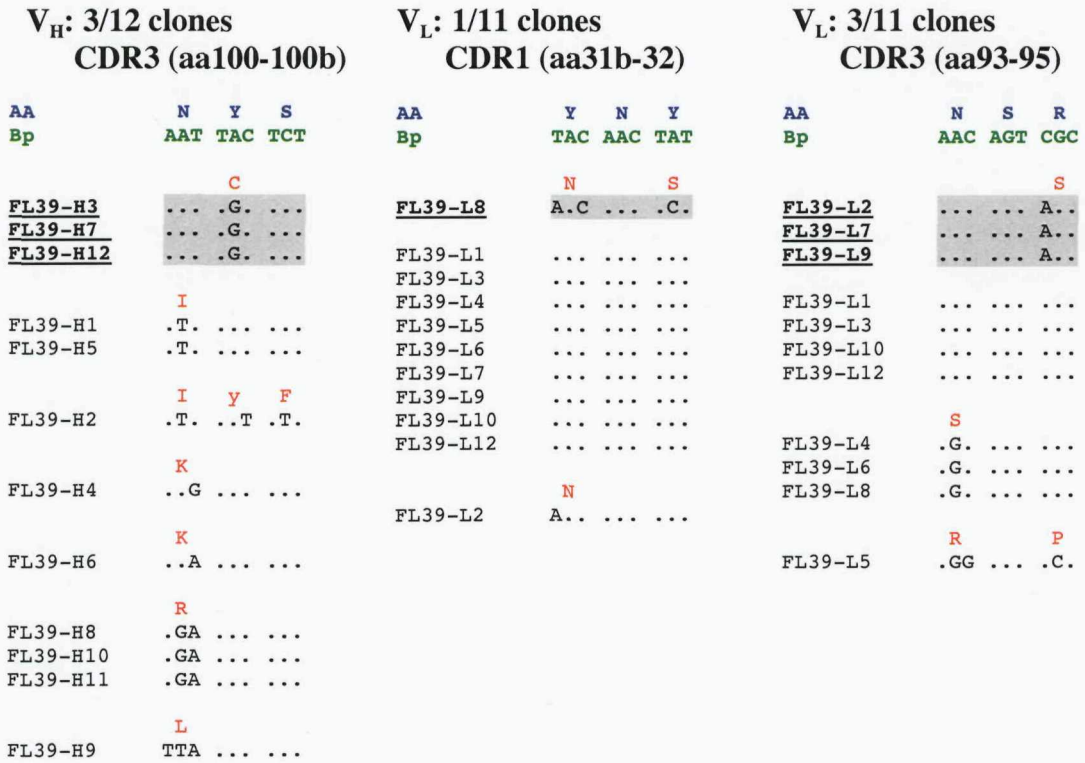


Figure 3.8. The accumulation of *N*-glycosylation sites by molecular clones during ongoing somatic hypermutation in FL39. The consensus sequences of tumour-derived V_H and V_L from FL39 do not feature an *N*-glycosylation site; amino acid sequence is shown in *blue*; nucleotide sequence is shown in *green*. The V_H and V_L molecular clones were analysed for *N*-glycosylation motifs introduced during ongoing SHM resulting in intraclonal heterogeneity. Amino acid mutations are shown in *red* (replacement, upper case; silent, lower case). Dots indicate matched nucleotides. *N*-glycosylation sites are shaded.

Table 3.4. The conservation of novel *N*-glycosylation sites among molecular clones

Case*	No. sites	Motif	No. clones with motif#
FL1	1	NTS	12/12
FL2	2	NMS, NSS	13/14, 14/14
FL3	2	NCS, NNS	7/7, 16/16
FL4	2	NIS, NLT	na
FL5	1	NYT	9/9
FL6	1	NIT	17/18
FL7°	3	NIS, NDS	13/13, 4/4
		NVT	10/10
FL8	1	NIT	10/10
FL10	2	NIS, NGS	7/7, 6/6
FL11	1	NST	8/8
FL12	1	NKS	22/22
FL13	1	NVS	8/8
FL14	2	NWT, NHS^{GL}	14/14, 10/14
FL15	1	NIS	11/11
FL16	3	NDS, NIT	6/8, 5/8
		NHT	10/10
FL17	4	NGT, NYT	6/6, 6/6
		NYT, NLT	6/6, 8/8
FL20	2	NVS, NFT	8/8, 14/14
FL21	5	NFT, NMS	15/15, 15/15
		NTS, NVT	15/15, 15/15
		NIS	20/26
FL22	2	NIS, NGS	8/8, 6/6
FL24	3	NMS, NIT	6/6, 6/6
		NLS	9/9
FL25	2	NYS, NAT	13/13, 5/5
FL26	1	NHS	6/7
FL27	1	NIS	9/11
FL28	1	NYS	8/8
FL29	3	NIS, NNS	10/10, 7/10
		NYS	4/4
FL33	4	NGS, NYT	10/11, 9/11
		NWS, NKT	11/11, 11/11
FL34	3	NIS, NVS	6/6, 6/6
		NLT	11/11
FL35	4	NFS, NSS	17/17, 16/17
		NIT, NLT	9/10, 10/10
FL38	1	NTS	14/15
FL41	1	NIS	11/11
FL43	1	NTS	11/11

*; Only cases with matched V_H and V_L gene sequences have been assessed
#; Number of clones displaying the *N*-glycosylation motif / total number clones sequenced.
°; Two potentially functional clonal V_L rearrangements were identified.
na; Not assessed.
GL; Identifies germline V4-34 *N*-glycosylation site.
Underlining indicates a disfavoured *N*-glycosylation motif for which functional glycosylation is questionable.

9/11 and 14/15 molecular clones. Furthermore, the accumulation of sites by continued SHM was also evident in cases that had sites in the consensus sequences. In these cases, motifs derived from intraclonal heterogeneity were additional. V_H genes from 4 cases (FL2, FL12, FL16, FL21) had additional motifs, which affected, on average, 13.6% (4.5-62.5% for individual V_H cases) of molecular clones. Similarly, V_L genes from a further 4 cases had additional sites arising via ongoing mutational activity, 3 in Vκ (FL8, FL17, FL42) and 1 in Vλ (FL23). The motifs were shared by 38.1% (10.0-62.5% for individual V_L genes) of molecular clones. Overall, 9/44 (20.5%) tumours continued to accumulate *N*-glycosylation sites as a result of ongoing SHM.

3.2.3.2. Incidence of germline-encoded *N*-glycosylation sites

Six V genes have a naturally occurring *N*-glycosylation site in their germline sequence. The 3 V_L gene segments, B2 (Vκ5 family), 3e (Vλ3 family) and 5e (Vλ5 family) were not detected in our cohort. Of the 3 V_H gene segments, V1-08, V4-34 and VH5a, the V4-34 gene was used in 2 of our cases. The germline NHS motif in CDR2 was retained in 1 case (FL14) and lost in the second, with both cases gaining 1 novel site. Unlike the acquired site, which was completely conserved, a loss of the germline site was observed in 4/14 FL14 V_H clones in the face of ongoing somatic mutation.

We have combined our 2 V4-34 sequences with 13 from three published studies.[368,369,460] Overall, 12/15 (80.0%) cases lost the germline site, of which 9 acquired 1 (n=7) or 2 (n=2) new sites. In 2 cases, the germline site was retained and an additional site was acquired and in 1 case the germline NHS motif was mutated to NYS and no further sites were acquired. The acquisition of novel sites was similar to that found for all V_H sequences, indicating that V4-34-encoded cases behave similarly.[12] Since only V_H sequences could be interrogated, it remains consistent with our observation that, if V_H and V_L are examined, almost all cases of FL acquire sites.[498] In summary, it appears that there is no selective pressure to maintain the germline site in V4-34, instead cases using this V_H gene acquire alternative sites.

One case in the literature used the V1-08 gene segment and in this instance the germline NTS motif in FR3 was retained and no other sites were acquired. No cases used the VH5a gene segment.

3.2.4. Location of novel *N*-glycosylation sites

The distribution of novel *N*-glycosylation sites in the V_H and V_L gene sequences is shown in Figures 3.9a. and 3.9b., and summarised in Figure 3.10. The majority (86.4%) of sites in V_H were located in the CDRs: CDR2 (n=22) > CDR3 (n=11) > CDR1 (n=5). FRs acquired fewer sites (n=6). No difference was found in the number or location of sites between tumours of different isotypes. For V_L, 12/21 (57.1%) sites were located in either CDR1 or CDR3 and 9 (42.9%) in the FR's. Strikingly, CDR2, which is a common location for motifs in V_H, shows no sites in V_L, in spite of the fact that CDR2 is not disfavoured by SHM in normal B cells.[499] The greatest number of sites was found in CDR3 (n=8). No sites were observed in FR2 in either V_H or V_L sequences, in keeping with the internal location of this region in each chain of the Ig molecule. To broaden our analysis, the location of sites in the published V_H genes of an additional 69 FL were analysed (Figure 3.11.).[368,369,460] The focusing of sites within CDR2 and CDR3 (41.1% and 29.1% of sites, respectively) was confirmed. The presence of motifs in CDR3 suggests that this region of the V gene is affected by SHM, which has previously been difficult to show directly in the absence of a germline counterpart.

3.2.5. Generation of *N*-glycosylation sites by somatic mutation

The motif Asn-X-Ser/Thr was created by a range of nucleotide changes (Table 3.5.). The codon for Asn (AAT or AAC) was generated by mutation rather than being derived from the germline sequence in 41/45 (91.1%) instances. "Starter" sequences for motif generation depended on the donor germline V_H. For example, the aa sequence YIS, present in the CDR2 (position 50-52) of V3-48 and V3-11 gene segments, was mutated to NIS in 6 cases and to NIT in 2 cases. Mutational changes of Ser to Asn were relatively common (12/45; 26.7%), likely reflecting the presence of a hot spot of SHM at codon 50 in CDR2 (see below). Other common aa replacements in codon 1 affected Tyr (15/45; 33.3%) and Thr (6/45; 13.3%), where replacement mostly required 1 nucleotide change to generate the starter Asn. Indeed, acquisition of sites was predominantly (73.3%) achieved by a single aa replacement to Asn at position 1 of the motif. At the nucleotide level, the first codon underwent mutations of 1 (68.3%) and 2 (29.3%) nucleotides, but rarely 3 (2.4%). In a proportion (24.4%) of sites, additional non-essential aa replacements in codon 2 (n=4) or codon 3 (n=7) were found. It is interesting to note that unlike the Asn in codon 1, the Ser at codon 3 of the motif was most commonly found in the germline sequence (n=20), but also found mutated from Ser to Thr (n=6); 18 motifs therefore ended in Thr, of which two thirds were present in the germline sequence. A change of Thr to Ser was only observed once. Where codon 3 was mutated (42.2%), the process had involved 1

Case	FR1	CDR1	FR2	CDR2	FR3	CDR3	FR4
FL1	EVLLLESGGGLAQPGGTLRLSCSASGFAFS	TYGMS	WVRQAPGKGLEWVS	AISGGSDGIYYADSVRG	RFIISRDNSKNTLYLQMNSLRAEDTAVYYCAK	DGDFSLGCYGGDS	WGQGILVTVSS
FL2	EVQLVESGGGLVQPGGSLRLSCTTSGFTFS	NYNMS	WVRQAPGKGLECIS	DINSSGKNIFYADSVKG	RFTISRDNVKNVVYLQMNSLRDEDTAVYYCAR	DTPEAPVSDYYDVDV	WGQGTTVTVCS
FL3	GVQLVESGGGLVKPGGSLRLSCSVSGFTFT	NAWMN	WVRQAPGKGLEWVG	RIKNRADGETTDYAAPVKG	RFSISRDESKNMLFLQMISLKPEDTAIYYCTS	NCSHSHCPKISSYNDYDMDV	WGPGILVTVSS
FL4	QVQLQESGPGLVKPSETLSLSCSVSGASVT	AYSWS	WIRQSPDKGLDWIG	NISYTGNTKYNPSVKS	RVIISLDTSKNHFSLKVTSVTVADTAIYYCAR	LRSGSSAFDY	WGQGTLVTVSS
FL5	EVQLVESGGTLVQPGGSLRLSCSASGFTFS	DYTMH	WVRQAPGKGLEYVS	SISGNGYSTYYADSVKG	RITISRDNSKKTLSLQMSTLRTEDTAVYYCVR	GSDDTGTYPQGIFFNY	WGQGTLVTVSS
FL6	EVQVVESGGRWVQPGGSLRLSCVASGFSFN	AYTMN	WVRQAPGKRLEWIS	NITSSGFTKDYADSVKG	RFTVSRDNAKNSLYLQMNGLKNEDTALYYCVR	QNFGGYVFDF	WGQGTTVTVSS
FL7	EVQLVESGGGLVQPGGSLRLSCAASGFTFS	SYSMN	WVRQAPGKGLEWIS	NISSSISNIYYADSVKG	RFTISRDNAKNSLYLQMNSLRDEDTAVYYCAR	DSLGSSLDY	WGQGTLVTVSS
FL8	DVELLESGGGLVQPGGSLRLSCAASRFTFS	TYAMT	WVRQAPGKGLEWVS	NITSSGDTTNYAESVQG	RFTISRDNSKNTVYLQMNSLRVEDTALYYCAK	EGRDSPRYFDS	FGQGTLVTVSS
FL10	EVQLVESGGGLVQPGGSLRLSCVASGFTYS	TFEMN	WVRQAPGKGLEWVS	NISSSGYSIHYIDSVKG	RFTISRDNAKNSLYLEMNNLRAEDTAIYYCAR	DLSGTSGPHVFDM	WGQGTMVTVSS
FL11	EVQLVESGGGLVKPGGSLRLFCAASGFTFR	SYTMN	WVRQAPGKGLEWVS	SISGFSSYIYYADSLRG	RFTISRDNARNSLSLQINDLRPDDTAIYYCAT	NSTSSKTSYTFDY	WGRGTLVTVSS
FL12	EVQLVESGGALVEPGESLRLSCATSGFIFP	NAWMN	WVRQAPGKGLEWLG	HIKNKSDGGTTDYAAPIKG	RFTISRDDSRNTLFLQMNSLKNEDTAMYFCTT	GDFDY	WGQGTLVTVSS
FL13	QVQLQESGPRLVKPSETLALTCTVSGDSVS	SSTSSWS	WIRQPPGKGLEWIG	NVSYTGYSNYNPSFHS	RASMSVDTSKAQLSLRLTSVTAADTAMYFCAR	DPGFRSSA	WGKGITVVVSS
FL14	QVQLQQWGAGLLKTSETLSLTCAVYGGSFN	NYNWT	WIRQPPGKGLEWIG	QINHSGTTNYNPSLKS	RVTMSIDPSENQFSLKVRSVTAADTAIYYCVR	GSPESSGNYWGHFQY	WGQGTLATVSS
FL15	EVQLVESGGGLVQSGGSLRLSCAASGFTFS	YYGMN	WVRQAPGKGLEWVS	YISGSGGAIYYADSLKG	RFIISRDNAKNSLYLQMNNLRAEDTAIYYCAS	NISTRTSYFFDS	WGQGTLVTVSS
FL16	EMILVESGGGLIQPGRSLRLSCTTSGFNFA	GYAMS	WFRQAPGKGLEWVG	FIESKAFGGTTEYAASVKG	RFTISRNDSESILYLQMTSLRIEDTAVYHCSV	YCTNITCYASGMDV	WGRGTNITVSS
FL17	EVQLKESGGGKVEPGGSLKLSCEASGFKFK	KFAMT	WVRQSPGKGLEWVS	VINGTGGNTNYTDSVKG	RFTISRDNSRNTLFLQMDSLRVEDTALYYCAK	VSNYTFGLNWFDS	WGQGTLVTVSS
FL18*	EVQVLESGGGLIQPGGSLRLSCAASGFRFS	TFGIS	WVRQAPGKGLEWVA	GIVGSGSRTYYADSVKG	RFTISRDNDQSMAFLQMSSLRHDDTAVYYCGK	ARSFGKLPLDH	WGQGTLVTVSS
FL20	HLQVQESGPGLVKPSETLSLNCKIFGDDYM	NSYNYGWD	WIRQPPGKGLEWIG	NVSYSGTTSYNPSLKS	RVTISIDMSKSQFSLNLRSVTAADTAMYYCAR	GSGWFDP	WGQGTLVTVSS
FL21	ELQLVESGGGLVQPGGSLRLSCVASGFNFT	NYNMS	WVRQAPGKGLEWIS	YISNTSTSIYYADSVKG	RFIISRDNAKESLYLQMSSLRDEDTAVYYCAR	RNVTTVNLGLDVFDI	WGQGTMVTVSS
FL22	EVQLVESGGGLVQPGGSLRLSCTASGFTFS	NYPMN	WVRQSPGKGLEWVA	NISSSGSGLHYVDCVKG	RFTISRDNAKNSLYLQMNSLRSEDTAVYYCAG	LDNLDV	WGQGTMVTVSS
FL24	LLYLEESGGGLVNPGGSLRLSCAASGLTFR	DYNMS	WIRQAPGKGLEWVS	NITTSGNTIDYADSVKG	RFTISRDNAKNSLHLQMNNLRGDDTAVYYCAT	TVPSASSPAYYYMDV	WGKGTTVTGSS
FL25	QVQLQQWGAGLLKPSETLSLICAVYGGSLN	DYYWS	WIRQPPGGGLEWLG	EVDSRGGTNYSLALTG	RVTVSLDMSKNSFSLKLTSVTVADTATYFCAR	GLGLDMDFDL	WGRGTLVTVSS
FL26	EMQLVESGGGLVQTGESLRLSCSASGFTIS	SHIMH	WVRQAPGQGLKYVS	GISAYGDKTYYADSMRD	RFAISRDNSKNTVYLQMSSLRIEDTAVYYCAT	NHSEGSFDS	WGQGTLVTVSS
FL27	EVQLVESGGGLVQPGGSLRLSCEASGFSFS	TYSMN	WVRQVSGKGLEWLS	NISSTSGSTNYADSVTA	RFTISRDNAQNTLWLQMNSLRAEDTAIYYCART	SCSGRSCHFDY	WGQGTLVTVFS
FL28	QVQLVESGGGVVQPGTSLRLSCAASGFPFI	YYGIH	WVRQAPGKGLEWVA	GLSCDGTNEYYADSLKG	RFVISRDNSKNTVSLHMNLLKTEDTAVYYCAK	DFHQHYGSGSYYQGTFDF	WGQGSLVTVSS
FL29	EVQLVESGGGLVQPGGSLRLSCAASGFTFS	SYSMT	WVRQAPGKGLEWVS	NISSSSNTAFYADSVKG	RFTISRDNVNNSLYLQMHSLRDEDTAVYFCAR	DRNGYNFFDY	WGQGTLVTVSS
FL31*	EVQLVESGGGLVQPGGSLRISCVASGFTFN	SHWMN	WVRQAPGKGLEWVA	NITQDGSDAYYVDSVKG	RFIISRDNAKNSLFLQMNSLRVEDTAVYYCVR	DDQNRIFDY	WGQGTLVTVSS
FL32	QVQLVQSGAEVKKPGSSVKVSCKASGGTFS	NYAIS	WVRQAPGQGLEWMG	GIIPIFGTANYAQKFQG	RVTITADKSTSTAYMELSSLRSEDTAVYYCAR	VFSSDFWSGYWYY	WGQGTLVTVSS
FL33	QVQLLESGPGLVKPSETLSLTCAVSGDSIR	SYNWS	WIRQPAGKGLEWIG	HIFSNGSTNYTPSLKS	RVTMSLDPSKNQLSLKLSSVTAADTAVYYCAR	DNKTYTKTWNAYDTYAMDV	WGQGTTVSVTS
FL34	EVQLVQSGGGLVQPGRSLRLSCAASGFAFS	SYTMT	WIRQTPGMGLEFIG	NISSSSTNTFYGDFVKG	RFTISRDNDKGSLFLQMNNLRDDDTALYYCAR	GGTYLDY	WGQGTRVNVSS
FL35	QITLRESGPTLVKPTQTLTLTCTFSNFSLT	TPGMGVG	WVRQPPGKALEWLA	VVYWNDYKRYSPSLKT	RLSITADISKNQVVLTMTNMDPVDTATYYCTY	NSSDRIKFDP	WGQGIRVTVSS
FL38	EVQLLESGGGLVQPGGSLRLSCAASGFPLR	SYAMT	WVRQAPGKGLEWIS	SISGGGSATYYTDSVKG	RFTISRDNSKNTLFLQMNSLRAEDTAIYYCAK	NTSDDYSAY	WGQGTLVTVSS
FL39	EKQVLESGGKLVQPGGSLRLSCVAYGFTFE	KNAIN	WVRQAPGKGLQWVS	GLSGGDSSTYYADSVKG	RFTISRDNSDNMFYLQMSSLRVEDTALYYCAT	TCPSSKYSLSYAFDI	WGRGTMVTVSS
FL41	EERLVESGGGFVQPGGSLRLSCAASGFTFN	NYDMT	WVRQAPGKGLEWVS	TISGSGVKTYYADSVKG	RFSTSRDNSKDTVNVQMNSLRVEDTAMYYCAK	NISMEI	WGQGTLVTVSS
FL43	EVQLLESGGGSVQPGGSLILSCVSSGFTFS	NYAMS	WVRQVPGKGLEWVS	TVSGSGDSTYYGDSVKG	RFTISRDNSKNTLYLQMNSVRADDTAAYYCAN	TSFSMGWFGP	WGQGTLVTVSS
FL44*	QVHLVQSGSELKKPGASVKISCKASGYTFT	RFAVN	WVRQAPGQGLEWMG	WINTNTGKPTYAQGFTG	RFVFSLDTSVTTSYLQISTLKAEDTAVYFCAR	DFFRDYFDY	WGQGTLVTVSS

Figure 3.9a. The incidence and location of novel *N*-glycosylation sites in the V_H genes of FL. The incidence and location of *N*-glycosylation sites within the deduced amino acid sequences of V_H genes for 36 FL were analysed; those cases for which the partner chain could not be identified are indicated by a *. Novel *N*-glycosylation sites are highlighted in *red*. The natural *N*-glycosylation site from the germline V4-34 sequence is highlighted in *green*. Ambiguous *N*-glycosylation motifs for which functional glycosylation at the protein level is questionable are underlined.

Case	FR1	CDR1	FR2	CDR2	FR3	CDR3	FR4
FL1	DIQMTQYPSSLSASVGDRVTIAC	RASQGISNYLA	WYQQKPGKPPKLLIY	GASTLQS	GVPSRFSGSGSGTDFTLTISCLQPEDVATYYC	QKYNTSPWT	FGQGTRVEFK
FL2	DIQMTQSPSSLSASVGDRVTITC	RASQGIRNDLV	WYQQKPGKAPKRLIY	AAYSLQA	GVPSRFSGSGSGTEFTLTVSSLQSEDFATYYC	LQHHNYPRT	FGQGTKVEIK
FL3	DVLMTQSPSSLSASVGDRVTITC	RASQGFRNDFG	WYQQIPGKAPKCLIY	AASTLQS	GVPSRFSGSGSGTEFTLTISSLQPEDFATYYC	LQRNNSALT	FGGGTKVEIK
FL4	QSVLTQPPSVSGAPGQRVTVSC	TGSSSNIGAGYDVQ	WYQQVPGTAPKLLIY	GNFNRPS	GVPDRFSGSKSATSASLAITGLQAEDEAEYYC	QSYDGSLSGVI	FGGGTNLTVL
FL5	DIVMTQSPDSLAVSLGERATINC	KSSQSLLYSINNSNYLA	WYQQKAGQPPKLLIY	WASTRAS	GVPDRFSGSGSGTDFTLTISSLQAEDVAVYYC	QQNYTTPS0	FGGGTKVEIK
FL6	EIVLTQSPGTLSLPPGERATLSC	RASQSVTNNYLA	WYQQIPGQAPRLLIS	GASSRAT	GIPDRFSGSGSGTDFTLTIARLEPEDFAVYYC	QQYGYSRT	FGQGTKVEIK
FL7-1	DIQMTQSPSTLSASIGDRVTVTV	RATPNIFSGLA	WYQQTQGRAPKVLIS	TAANLES	GVPLRFRGSGFGTRFTLTINSLQPNDSATYYC	QQYHSFPRT	FGQGTKVEIE
FL7-2	EIVLTQSPATLSLSPRENVTLSC	RASQSVTNYLA	WYQQKPGQAPRLLIY	DASNRAT	GIPARFSGSGYGTDFTLTISSLEPEDFAVYYC	QQRSNWLS	FGGGTKVEIK
FL8	HIVLTQSPATLSLSPGERATLSC	RASQSVSDFLA	WYQQKPGQAPRLLIY	DTSNRAT	GIPARFSGSGSGTNFTLTISSLEPDDFAVYYC	QQDLT	FGGGTKVEIK
FL9*	DIVMTQSPDSLAVSLGERATINC	KSSQSLLYSYNKKNYLV	WYQQKPGQPPKVLIS	WASTRES	GVPDRFSGTGSGTDFTLVISSLQAEDVAVYYC	QQYYTTPPT	FGQGTKVEVK
FL10	DIVMTQSPDSLTVSLGERATINC	KSSQSVLFSYNNKNYLG	WYQQRPGQPPKLLIY	WASTRES	GVPERFSGNGSGTDFTLIISSLQAEDVADYYC	QQYFTTPLT	FGGGTKVEIK
FL11	QSVLAQPPSASGTPGQRVIIPCS	GSSSNIGTNTVD	WYQQFPGTAPKLLIY	SNNQRPS	GVPDRFSGSNSGTSASLAISGLQYEDEADYYC	ATRDVNLDAPV	FGGGTKLTVL
FL12-1	DIQLTQSPSSLSASVGDRVTITC	RASQGIRSDLG	WFQKKPGKAPKRLIY	PASTLQS	GVPSRFSGCGSGTDFTLTINSLQPEDFATYYC	LHYKGTF	GQGTKLEIKR
FL12-2	DIVMTQSPDSLAVSLGERATINC	KSSQSVLYSSNNKNYLA	WYQQKPGQPPKLLIY	WASTRES	GVPDRFSGSGSGTDFTLTISSLQAEDVAVYYC	QQYYSTPWT	FGQGTKVEIK
FL13	EIVLTQSPATLSLSPGERVTLSC	RASQSVSSYLA	WYQQRFGQAPRLLIY	DASNRAT	GIPARFSGSGSETDFTLIISRLEPEDFAVYYC	QQYSSSFVT	FGPGTKVEIK
FL14	DIVMTQSPDSLSVSLGERATINC	KSSQSVLYSSHNKNYLA	WYQQKPGQPPRLLIY	WASTRES	GVPDRFSGSGSGTDFTLTINTLQAEDVAVYYC	QQYYTTPYT	FGQGTKLEIK
FL15-1	EIVLTQSPGTLSVSPGQRVTLSC	RASQSVISPYLA	WFQQKPGQAPSLVIY	GTSSRAT	GISDRISGSGSGTDFSLTISRLQPEDFGIYYC	QHYGSSPPMWT	FGQGTKVEIN
FL15-2	EIVLTQSPATLSLSPGERATLSC	RASQSVSSYLA	WYQQKPGQAPRLLIY	DASNRAT	GIPARFSGSGSGTDFTLTISSLEPEDFAVYYC	QQRSSWPPLT	FGGGTKVEIK
FL16	DIQMTQSPSSLSVSVGDRVTITC	RASQIIGMSLH	WYQQKIGKPPKILIY	GAVNLQR	GVPTHFSGAGSGTDFTFTISSLQPEDFAYYYC	QQNHTAPRT	FGQGTKVEMQ
FL17	DIEMTQSPDSLTVSLGERATINC	KSSQSLLHISNNKNNLT	WYQQRPGQPPKVLIY	WASTRQS	GVPDRFSGSGSGTDFTLTISSLQTEDVAVYSC	QQFHSFPWT	FGQGTKVEIK
FL19*	QSALTQPASVSGSLGQSITISC	TGSSNDIGTYKLVS	WYQKHPDKAPKLIIY	EACKRPS	GISDRFSGSKSGNTASLTISGLQAEDEADYYC	CSYGFGFTSYWV	FGEGSKLTVL
FL20	EIVLTQSPDFQSVTPKEKVTITC	RASQSIGSSLH	WYQQKPDQSPKLLIK	YASQSFS	GVPSRFSGSGSGRDFTLTINSLEAEDAATYYC	HQSSTFNT	FGQGTKVEIK
FL21	DIQMTQSPSSLSASVGDRVTVTC	RASQNISNYLN	WYQLKPGKVPKLLIF	AASSLQS	GVPSRFSGSQSGADFTLTINNLQPEDFASYYC	QQSYSSPSW	TFGQGTNVEI
FL22	DVVMTQSPLSLPVTLGQPASISC	TSSQSLVYSDGNIYLN	WFQQRPGQSPRRLIY	KVSNRDS	GVPDRFSGNGSGTDFTLKISRVEAEDVGVYYC	MQGTHWPRT	FGPGTKVDVK
FL23*	QSVLTQPPSVSGAPGLRVTISC	TGNSSNIGAGYDVH	WYRHTPGRGPELLIY	DNSNRPS	GIPDRFSGSKSGTSASLAISGLQSEDEADYYC	QSYDSSLSGSRV	FGGGTKLTVL
FL24	KIILTQSPANLSLSPGERATLSC	RASQSLSSSLA	WYQQKPGQAPRLLIH	DASNRAT	GIPARFSGSGSETDFTLTISSLEPEDIAVYYC	QQRVT	FGGGTKVEIK
FL25	KSVLTQSPDTLSLSPGENATLSC	RASQTVDGNFLA	WYQQKPGQAPRLLLY	DASTRAT	GIPDRFSGSGSGRDFTLTISSLEPEDLGIYYC	QNYDKSLMYT	FGQGTKLEIK
FL26	DIVMTQSPDSLVVSLGERATIRC	KSSHSISYYSNNKHYLG	WYQQKPGQPPKLLIN	WASTRES	GVPDRFSGSGSGTDFTLTISSLQTEDVAVYYC	QQYYFTPWT	FGQGTKVEIK
FL27	DILLTQSPSSLSASVGDRVTITC	RTSQSIYTSLN	WFQQKPGKAPKLLIY	SASTLQS	QVPSRFSGSGSGTDFTLTISGLQPEDFATYYC	QQSYSALYT	FGQGTKLEIK
FL28	DIQMTQSPSSLSASVGDRVTITC	RASQSISNNLN	WYQQKPGKPPRLLIY	VASSLQS	GVPSRFSGSGSGTDFTLTIRSLQPEDFATYYC	QQNYSLLPHT	FGQGTKLEIR
FL29	DIEMTQSPSSLSASIGDRLTITC	RASQAIRESVQ	WYQQKPGKAPKLLMF	STSILES	GVPSRFSGSGFGTHFTLTISSLQPEDFATYSC	QQNYSGPFT	FGPGTKVEMR
FL30*	EIVLTQSPESLSLSPGDRATLSC	RASQTVNDNSLA	WYRHQPGLPPRLLIY	QTSRRAT	GIADRLRGSGSGTDFTLTITGLESEDSGIYYC	HQYVIVPQT	FGRGTKLEIK
FL32	EIVLTQSPGTLSLSPGERGTLSC	RASQSVSSSYLA	WYQQKPGQAPRLLIY	GASSRAT	GIPDRFSGSGSGTDFTLTISRLEPEDFAVYYC	QQYGSSPRT	FGQGTKVEIK
FL33	DIQMTQSPSSLSASVGDRVTITC	RASQGVSNYLA	WYQQKPGKVPKLLIY	AASTLQS	GVPSRFSGSGSGTDFTLTISSLQPEDVATYYC	QKYNSAPYT	FGQGTKLEIN
FL34	EIVMTQSPATLSVSPGERATLSC	RASQSVKSNLT	WYQQKPGQAPRLLIY	AASIRVT	GIPVRFSGSGSETEFTLTISSLQSEDIAVYYC	QQYSNWLT	FGGGTKVEIK
FL35	DFVLTQSPGTLSASVGDRVNITC	RASQSVSRWLA	WYQQKPGKVPELLIY	EASNLEN	GVPSRFSGSGSGTEFNLTINSLQPDDFATYYC	QQYNSYWT	FGQGTRVEIK
FL36*	QSVLTQSPSVSAAPGQMVTISC	SGTTSNIGNHFVS	WYQQIPGTAPKLLIY	ENNKRPS	GIPDRFSGSKSGTSATLGITGLQTGDEADYYC	GTSNSSLSV	FGTGTKVTVL
FL37*	DVELTQTPLSLSVTPGQPASISC	KSSQSLLHGDGKTHLY	WYLQKPGQPPQLLIY	ETSIRFS	GVPERFSGSGSGTDFTLKISRVEAEDVGIYYC	MQSIQLPLT	FGGGTKVEIK
FL38	QSVLTQPPSVSGAPGQRVTIPC	TGSRSNIGAGYDVH	WYQHFPGTAPKVLIS	GNSNRPS	GVPDRFSGSKSGTSASLAITGLQAEDEADYYC	QSYDSSLSGWV	FGGGTKLTVL
FL39	QSALTQPASVSGSPGQSITISC	TGSRSDIGGYNYVS	WYQQHPGKVPKVIIF	EVTCRPS	GISSRFSGSKSGNTASLTISGLQADDEAHYYC	SSYSNSRTRVL	FGGGTRLTVL
FL40*	DIQMTQSPSSLSASVGDRVTITC	RASQGIRNDLG	WYQQKPGKAPKRLIY	AASSLQS	GVPSRFSGSGSGTEFTLTISSLQPEDFATYYC	LQHNTYPIT	FGQGTRLEIK
FL41	DIQMTQSPSSLSASVGDRVTITC	RAGQAIRNDLD	WYQQKPGKAPKRLIY	AASSLQS	GVPSRFSGSGSGTEFTLTISSLQPEDFATYYC	LQLNTYPLT	FGGGTNVKIN
FL42*	DIQMTQSPSSLSACVGERVTITC	RASQDITDALV	WFQQKAGQAPESLIY	AASTLQS	GVPPKFSGSGSGTDFTLTINSLQPEDSAVYFC	QQYKTFPLT	FGGGTEVELK
FL43	DIVMTQSPDSLAVSLGERATINC	KSSQSALYSSNNKKNLA	WYQLKPGQPPKLLIY	WASTRES	GVPDRISGSGSGTDFTLTISSLQAEDVAVYFC	QQYYSSPWT	FGQGTKVEIK

Figure 3.9b. The incidence and location of novel *N*-glycosylation sites in the V_L genes of FL. The incidence and location of *N*-glycosylation sites within the deduced amino acid sequences of V_L genes for 41 FL were analysed; those cases for which the partner chain could not be identified are indicated by a *. Novel *N*-glycosylation sites are highlighted in *red*. Ambiguous *N*-glycosylation motifs for which functional glycosylation at the protein level is questionable are underlined.

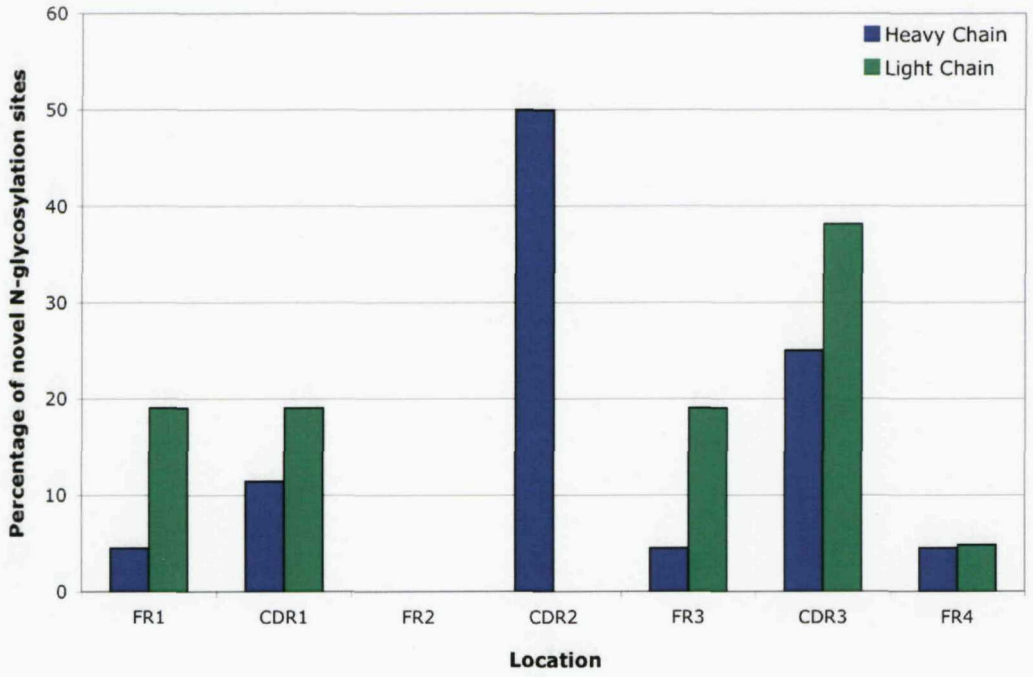


Figure 3.10. The distribution of novel *N*-glycosylation sites in the V region genes of FL. The location of novel *N*-glycosylation sites within the V_H and V_L sequence was analysed (FR1 To FR4) for 29 and 20 FL, respectively. Values are expressed for each region as a percentage of the total number of sites: V_H (n=43*); V_L (n=21). *The germline V4-34 site is not included in the analysis.

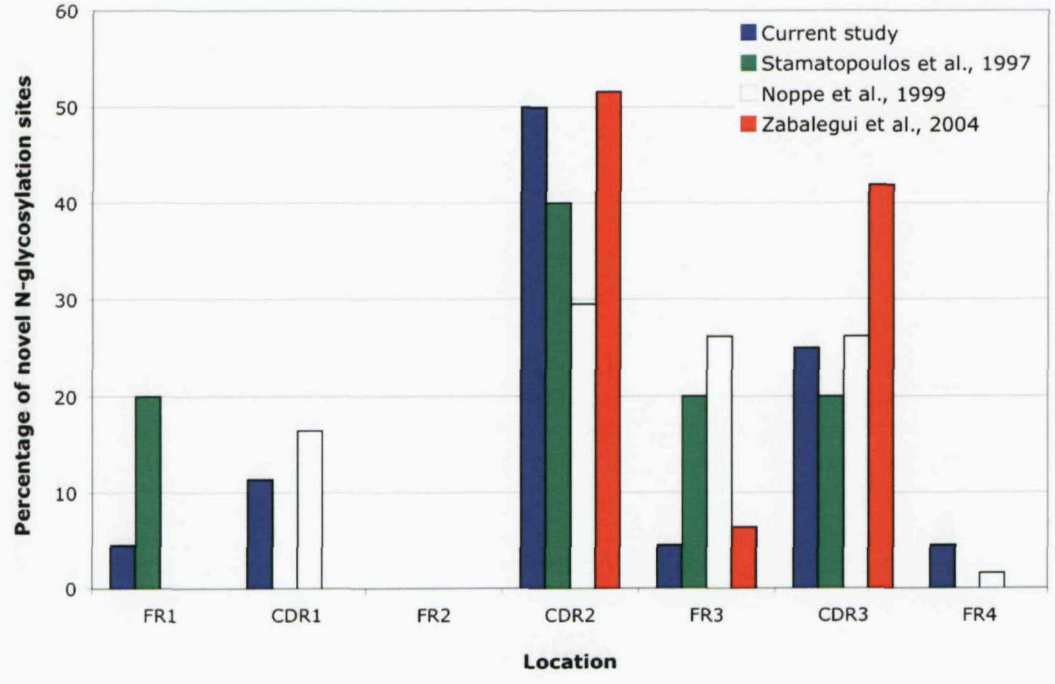


Figure 3.11. The distribution of novel *N*-glycosylation sites in the V_H genes of FL. The location of *N*-glycosylation sites observed in the V_H genes for our 29 FL was analysed and compared to other studies with published sequences for 69 FL: Stamatopoulos *et al.*, 1997, Noppe *et al.*, 1999, and Zabalegui *et al.*, 2004.[368,369,460] Values are expressed for each region as a percentage of the total number of sites (n=141).

Table 3.5. An analysis of somatic mutations giving rise to *N*-glycosylation motifs in the V genes of FL

Case	Chain	V gene*	Location	aa position	Generation of *N*-glycosylation site				
					aa change from germline			bp mutation#		
								1	2	3
FL2	V_H	V3-48	CDR1	33-35	SMN	→	NMS	AAC	ATG	AGT
			CDR2	52-53	SSS	→	NSS	AAT	AGT	AGT
FL4	V_H	V4-59	CDR2	50-52	YIY	→	NIS	AAC	ATC	TCT
	V_L	DPL8/1e	FR4	103-105	KLT	→	NLT	AAC	CTG	ACC
FL6	V_H	V3-48	CDR2	50-52	YIS	→	NIT	AAC	ATT	ACT
FL7	V_H	V3-48	CDR2	50-52	YIS	→	NIS	AAC	GAG	AGT
	V_{L1}	Vg38K/L6	FR3	81-83	DDF	→	NDS	AAT	GAC	TCT
	V_{L2}	HK102/L12a	FR1	18-20	RAT	→	NVT	AAT	GTC	ACC
FL8	V_H	V3-23	CDR2	50-52	AIS	→	NIT	AAT	ATT	ACT
FL10	V_H	V3-48	CDR2	50-52	YIS	→	NIS	AAC	ATT	AGT
	V_L	DPK24/B3	FR3	65-67	SGS	→	NGS	AAC	GGG	TCT
FL12	V_H	V3-15	CDR2	52a-52c	SKT	→	NKS	AAC	AAA	TCT
FL13	V_H	V4-61	CDR2	50-52	YIY	→	NVT	AAT	GTC	TCT
FL14	V_H	V4-34	CDR1	33-35	YWS	→	NWT	AAC	TGG	ACA
FL16	V_H	V3-49	FR3	72-74	DGS	→	NDS	AAT	GAT	TCC
			FR4	108-110	TVT	→	NIT	AAT	ATC	ACC
FL17	V_H	V3-23	CDR2	50-52	SGS	→	NGT	AAT	GGC	ACT
			CDR2	58-60	YYA	→	NYT	AAT	TAC	ACG
	V_L	DPK24/B3	CDR1	32-34	YLA	→	NLT	AAC	TTA	ACT
FL20	V_H	V4-39	CDR2	50-52	SIY	→	NVS	AAC	GTA	TCT
FL21	V_H	V3-48	FR1	28-30	TFS	→	NFT	AAT	TTC	ACG
			CDR1	33-35	SMN	→	NMS	AAC	ATG	AGT
			CDR2	52-53	SSS	→	NTS	AAC	ACT	AGT
	V_L	DPK26/A26	CDR1	28-30	SIS	→	NIS	AAT	ATA	AGT
FL22	V_H	V3-48	CDR2	50-52	YIS	→	NIS	AAT	ATT	AGT
	V_L	DPK18/A17	FR3	65-67	SGS	→	NGS	AAT	GGG	TCA
FL23	V_L	DPL8/1e	CDR1	26-28	SSS	→	NSS	AAC	AGC	TCC
FL24	V_H	V3-11	CDR1	33-35	YMS	→	NMS	AAC	ATG	AGT
			CDR2	50-52	YIS	→	NIT	AAT	ATT	ACT
	V_L	Vg38K/L6	FR1	20-22	TLS	→	NLS	AAC	CTG	TCT
FL25	V_H	V4-34	CDR2	58-60	NYN	→	NYS	AAC	TAC	AGC
	V_L	DPK22/A27	FR1	18-20	RAT	→	NAT	AAC	GCC	ACC
FL27	V_H	V3-48	CDR2	50-52	YIS	→	NIS	AAC	ATT	AGT
FL29	V_H	V3-48	CDR2	50-52	YIS	→	NIS	AAC	ATT	AGC
			FR3	75-77	KNS	→	NNS	AAC	AAT	TCA
FL31	V_H	V3-07	CDR2	50-52	NIK	→	NIT	AAC	ATA	ACC
FL33	V_H	V4-04	CDR1	33-35	YWS	→	NWS	AAT	TGG	AGT
			CDR2	54-56	SGS	→	NGS	AAT	GGG	AGC
			CDR2	58-60	NYN	→	NYT	AAC	TAC	ACG
FL34	V_H	V3-48	CDR2	50-52	YIS	→	NIS	AAC	ATC	AGT
			FR4	110-113	TVS	→	NVS	AAC	GTC	TCC
	V_L	DPK21/L2	CDR1	32-34	NLA	→	NLT	AAC	TTA	ACC
FL35	V_H	V2-05	FR1	26-28	GFS	→	NFS	AAT	TTC	TCA
	V_L	HK102/L12a	FR1	20-22	TIT	→	NIT	AAT	ATC	ACT
			FR3	72-74	TLT	→	NLT	AAT	CTC	ACC

*; Germline V gene with the greatest homology.
#; Amino acid position 1, 2 and 3 of *N*-glycosylation motif.
Underlining indicates a nucleotide replacement mutation.

(63.2%) or 2 (36.8%) nucleotide mutations. A proportion (17.8%) of sites were created through an essential aa replacement at both positions 1 and 3 of the motif and only a minority (8.9%) by an essential aa replacement at position 3 alone.

The disfavoured motifs NDS, NLS and NWS were found in 4 FL and involved 2 V_H genes and 2 V_L genes. In all cases, additional *N*-glycosylation sites were present in either the V gene featuring the disfavoured motif or its partner chain. The NES motif was not observed. In all, 26 different *N*-glycosylation motifs were observed. The most commonly used was NIS, which was observed in 9 (20.5%) V_H gene sequences, but which was only seen once in V_L. Other frequently used motifs included NIT (n=5), NSS, NTS, NYT and NLT (all n=4).

A wider analysis of the generation of *N*-glycosylation sites in the V_H CDR2 was performed and included sequences from cited references (Table 3.6.).[368,369,460] All 47 sequences with CDR2 motifs used V_H3 (n=35) and V_H4 (n=12) genes. Here, 28/47 (59.6%) had acquired a site at aa position 50-52. Codon 50 is a recognized hotspot for SHM in both productively and non-productively rearranged alleles of normal B cells and is affected in ~20% of cases.[486] However, rarely does the mutation of codon 50 lead to the introduction of an *N*-glycosylation site in normal B cells.[12] A much higher frequency (53.9%; 25/47) of FL-derived V_H sequences showed a replacement mutation at codon 50 (Table 3.6.). This may be in part be explained by the relative over representation of the V_H3 genes and in particular V3-48, for which the majority of sites were generated by aa replacement at position 50. In contrast there is no recognised hotspot for SHM at codon 52 of V_H, but we find that codon 52 is mutated in 21/34 (61.8%) cases.[486] The high incidence of codon 52 changes in CDR2 motifs further points to the selective pressures on FL to acquire *N*-glycosylation sites in their V genes. Eight further cases using gene segments V3-11, V3-23 or V3-30 and generated sites by a replacement mutation at codon 50, the majority (6/8) with an additional permissive mutation in codon 52. For segments V3-07 (n=3), all sites were generated by the replacement of Lys to either Ser or Thr at position 52. For segments V3-15 (n=3), all sites were generated by the replacement of Ser to Asn at position 52a. In the remaining gene segments, sites did not appear to focus to any one aa position. Strikingly in V_L there is a complete absence of motifs in CDR2, in spite of a known hotspot for SHM in codon 52 (Figure 3.10.). This points to a selection against further acquisition of sites here, as in normal B cells, CDR2 and CDR1 appear to be preferentially subjected to SHM.[499] Other common hotspots for SHM, such as codons 31, 55 or 82a, were not used to generate novel sites and we observed no recurring location of sites in V_L.

Table 3.6. An analysis of somatic mutations giving rise to *N*-glycosylation motifs: A focus on V_H CDR2

Case*	V gene*	aa position	Generation of *N*-glycosylation site					
			aa change from germline			bp mutation° 1	2	3
FL31	V3-07	50-52	NIK	→	NIT	AAC	ATA	ACC
FL-12[c]	V3-07	50-52	NIK	→	NIS	na	na	na
Rut[b]	V3-07	50-52	NIK	→	NIT	AAT	ATA	ACA
FL-10[c]	V3-09	50-52	GIS	→	NIS	na	na	na
FL24	V3-11	50-52	YIS	→	NIT	AAT	ATT	ACT#
FL-2[c]	V3-11	52-52b	SS-	→	NGS	na	na	-
		52c-54	-SS	→	NTS	-	na	na
FL-16[c]	V3-11	58-60	YYA	→	NYT	na	na	na
Dier[b]	V3-11	52-53	SSS	→	NGT	AAT	GGT#	ACT#
		56-58	TIY	→	NIS	AAC	ATA	TCT
Viv[b]	V3-11	50-52	YIS	→	NIT	AAC	ATT	ACT#
		56-58	TIY	→	NTS	AAC	ACA#	AGT
FL12	V3-15	52a-52c	SKT	→	NKS	AAC	AAA	TCT#
Cal[b]	V3-15	52a-52c	SKT	→	NKS	AAT	AAA	AGT#
Hou[b]	V3-15	52a-52c	SKT	→	NKT	AAT	AAA	ACT
FL8	V3-23	50-52	AIS	→	NIT	AAT	ATT	ACT#
FL17	V3-23	52-53	SGS	→	NGT	AAT	GGC	ACT#
		58-60	YYA	→	NYT	AAT	TAC	ACG
FL_5[a]	V3-23	50-52	AIS	→	NIT	AAT	ATA	ACT#
FL-17[c]	V3-23	50-52	AIS	→	NIT	na	na	na#
Timj[b]	V3-23	50-52	AIS	→	NSS	AAT	AGT#	AGT
Vit[b]	V3-23	50-52	AIS	→	NIT	AAT	ATT	ACT#
FL-5[c]	V3-30	50-52	VIS	→	NIS	na	na	na
FL-24[c]	V3-30	56-58	NKY	→	NTS	na	na#	na
Clau[b]	V3-30	58-60	YYA	→	NYS	AAT	TAT	TCA
FL2	V3-48	52-53	SSS	→	NSS	AAT	AGT	AGT
FL6	V3-48	50-52	YIS	→	NIT	AAC	ATT	ACT#
FL7	V3-48	50-52	YIS	→	NIS	AAC	GAG	AGT
FL10	V3-48	50-52	YIS	→	NIS	AAC	ATT	AGT
FL21	V3-48	52-53	SSS	→	NTS	AAC	ACT#	AGT
FL22	V3-48	50-52	YIS	→	NIS	AAT	ATT	AGT
FL27	V3-48	50-52	YIS	→	NIS	AAC	ATT	AGT
FL29	V3-48	50-52	YIS	→	NIS	AAC	ATT	AGC
FL34	V3-48	50-52	YIS	→	NIS	AAC	ATC	AGT
FL_10[a]	V3-48	50-52	YIS	→	NIS	AAT	ATT	AGT
FL-19[c]	V3-48	50-52	YIS	→	NIS	na	na	na
Tim[b]	V3-48	50-52	YIS	→	NIS	AAC	ATT	AGC
Vho[b]	V3-48	50-52	YIS	→	NIT	AAC	ATT	ACT#
		55-57	STI	→	NIT	AAT	ATA#	ACA
Lah[b]	V3-64	53-55	NGG	→	NGT	AAT	GGG	ACT
FL33	V4-04	54-56	SGS	→	NGS	AAT	GGG	AGC
		58-60	NYN	→	NYT	AAC	TAC	ACG
Reu4[b]	V4-04	60-62	NPS	→	NHS	AAT	CAC	TCC
FL25	V4-34	58-60	NYN	→	NYS	AAC	TAC	AGC
FL_3[a]	V4-34	50-52	EIN	→	NIS	na	na	na
FL-6[c]	V4-34	50-52	EIN	→	NIS	na	na	na
		60-62	NPS	→	NAS	na	na	na
FL-7[c]	V4-34	52-54	NHS	→	NYS	na	na	na
FL-22[c]	V4-34	54-56	SGS	→	NGS	na	na	na
Vwb3[b]	V4-34	58-60	NYN	→	NCS	AAC	TGC#	AGC
FL20	V4-39	50-52	SIY	→	NVS	AAC	GTA#	TCT
FL4	V4-59	50-52	YIY	→	NIS	AAC	ATC	TCT
Marn[b]	V4-59	54-56	SGS	→	NGS	AAT	GGG	AGC
FL13	V4-61	50-52	YIY	→	NVT	AAT	GTC#	TCT

*; Includes data from the current study and from published sequences: [a]Stamatopoulos *et al.*, 1997; [b]Noppe *et al.*, 1999; [c]Zabalegui *et al.*, 2004.[368,369,460]

*; Germline V gene with the greatest homology.

°; aa position 1, 2 and 3 of *N*-glycosylation motif.

#; Non-essential amino acid replacement mutation.

na; Nucleotide sequence not known.

Underlining indicates a nucleotide replacement mutation.

3.2.6. Functional glycosylation of recombinant scFv-Cκ proteins

Using a 293F mammalian expression system, recombinant scFv-Cκ proteins incorporating the paired V_H and V_L from 12 FL, 1 MM, 2 CLL and 2 PCNSL were expressed and purified. Each recombinant protein was assayed for purity and specificity by SDS-PAGE followed by protein staining and western blot analysis and specific ELISA using anti-Id Ab. All proteins were found to be both pure and specific; representative data for five FL-derived scFv-Cκ proteins is shown in Figure 3.12.

3.2.6.1. Functional glycosylation of FL-derived scFv-Cκ proteins

All 12 FL-derived scFv-Cκ proteins migrated more slowly than would expected from the molecular weight of the encoded protein, approximately ~42kD (Figure 3.13., lane A). Heterogeneity was also consistent with the variable addition of glycans. For example, FL21, which has 4 *N*-glycosylation sites in V_H and 1 in V_L migrated much more slowly than FL6, which has just 1 site in V_H. A scFv-Cκ protein derived from a patient with MM, which carried somatically mutated V region genes but no germline-encoded or introduced motifs, acted as a negative control (Figure 3.13. and Table 3.7.).

All scFv-Cκ proteins were treated with two separate deglycosylating enzymes: PNGase F and Endo H. PNGase F cleaves the bond between the innermost *N*-acetyl glucosamine (GlcNAc) and Asn residues of high mannose, hybrid and complex oligosaccharides, thus it hydrolyses nearly all types of *N*-glycan chains from glycoproteins; however, it will not cleave Asn-linked GluNAc residues that contain a α-1,3 fucosidic linkage. Endo H cleaves the bond between the two proximal GluNAc residues of high mannose oligosaccharides but not complex oligosaccharides and, therefore, cleavage indicates oligomannosylation.[500] Upon treatment with PNGase F, proteins migrated to the expected position confirming that the *N*-glycosylation sites identified in the V genes of FL can indeed be glycosylated in mammalian cells (Figure 3.13., lane B). For 8/12 (66.7%) scFv-Cκ proteins, treatment with Endo H showed an identical migration pattern to digestion with PNGase F, demonstrating that *N*-glycosylation was exclusively of the high mannose type (Figure 3.13., lane C; FL2, FL6, FL11, FL14, FL15, FL26, FL29 and FL41). For three scFv-Cκ proteins, Endo H treatment increased migration, but not to that observed following PNGase F treatment, indicating that some added glycans terminated at oligomannose while others had been processed further (Figure 3.13., lane C; FL4, FL7 and FL21). For one scFv-Cκ protein, treatment with Endo H did not affect migration and demonstrated that this protein contained only complex oligosaccharides (Figure 3.13., lane C; FL25).

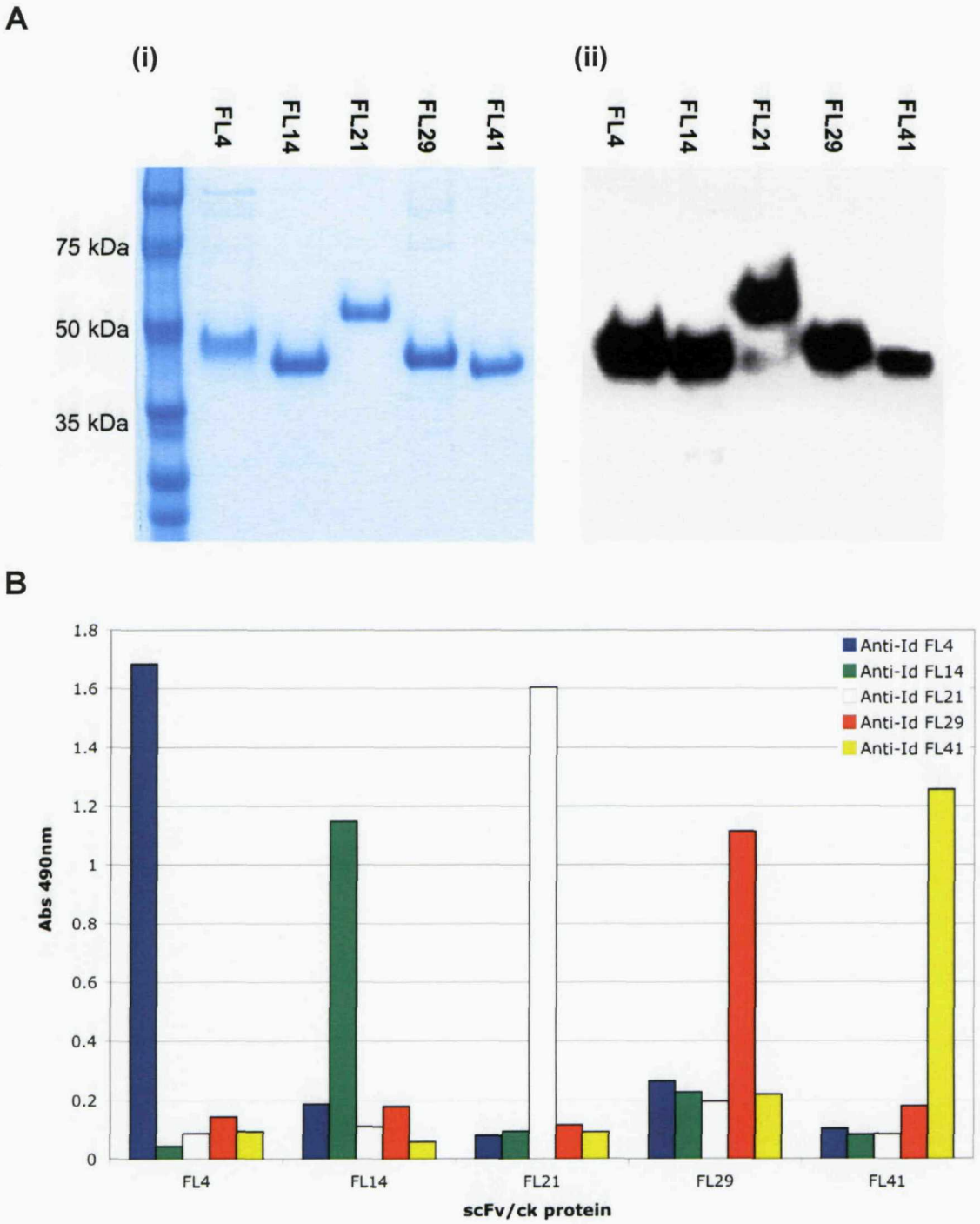


Figure 3.12. Expressed scFv-Cκ proteins were pure and specific. All recombinant proteins were assayed for their purity and specificity; representative data for five scFv-Cκ proteins is shown. A) Purity was confirmed by SDS-PAGE followed by visualisation by (i) staining with SimplyBlue Safestain™ and (ii) western blotting using polyclonal goat anti-human κ Ab. B) Specificity was confirmed by ELISA using mouse serum containing patient specific anti-idiotypic (Id) Ab: unrelated serums acted as negative controls.

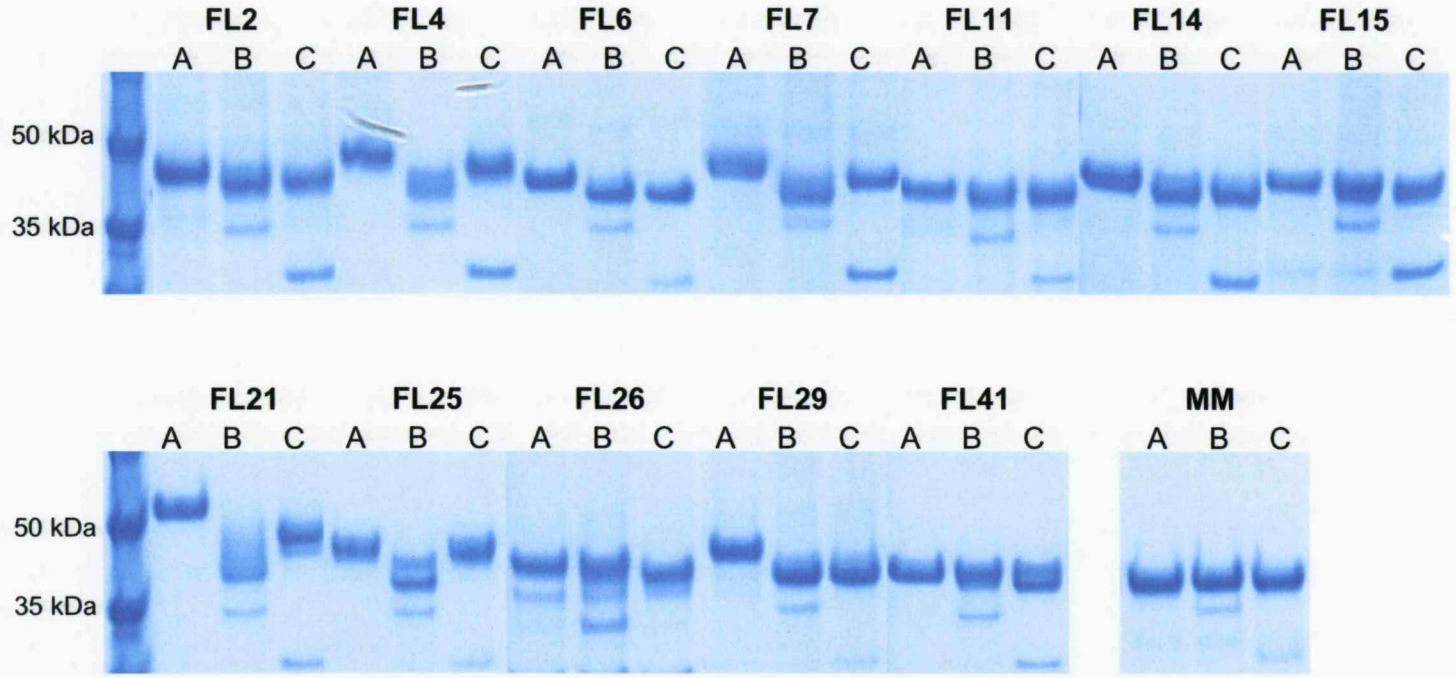


Figure 3.13. ***N*-glycosylation of scFv-Cκ proteins expressed in 293F cells.** Purified scFv-Cκ protein derived from 12 FL patients and 1 myeloma patient (MM), which served as a negative control since no *N*-glycosylation sites were present in the sequence, were separated by SDS-PAGE and visualised by staining with SimplyBlue Safestain™ (lane A). To remove *N*-linked carbohydrates, proteins were treated with PNGase F (lane B) and Endo H (lane C). Bands at ~36kDa and ~30kDa represent PNGase F and Endo H enzymes, respectively.

Table 3.7. V gene analysis, including V gene usage, somatic mutation and *N*-glycosylation, in B-cell tumours other than FL

	V_H						V_L					
Case	V_H Family	V gene*	Mutation (%)#	Glycosylation site No.	Location	Motif	V_L Family	V gene*	Mutation (%)#	Glycosylation site No.	Location	Motif
CLL1	V_H4	V4-34	9.3	1	$CDR2^{GL}$	NHS	Vκ2	DPK18/A17	2.3	0	-	-
CLL2	V_H4	V4-34	10.6	1	$CDR2^{GL}$	NHS	Vκ1	DPK5/L5	7.2	0	-	-
PCNSL1	V_H4	V4-34	1.0	1	$CDR2^{GL}$	NHS	Vλ3	DPL16/3l	1.9	0	-	-
PCNSL2	V_H3	V4-34	10.3	1	$CDR2^{GL}$	NHS	Vκ3	DPK22/A27	5.6	0	-	-
MM	V_H4	V5-51	4.8	0	-	-	Vκ2	DPK18/A17	4.5	0	-	-

*; Germline V genes with the greatest homology.
#; Frequency of somatic hypermutation compared to the most homologous germline V gene segment.
GL; Identifies a germline V4-34 *N*-glycosylation site.

Confirmation of sugar composition was obtained by analysing the complete NP-HPLC *N*-glycan profiles of FL21 and FL29 (Figure 3.14.); this was kindly performed by Mrs C. Radcliffe. The glycans isolated from FL29 confirmed an oligomannose (M) series, M5-9 (Figure 3.14a.). This oligomannose M5-9 series was also seen in glycans isolated from FL21, but this case also contained additional complex sugars, such as core fucosylated biantennary digalactosylated disialylated glycans (Figure 3.14b.). Mass spectroscopy was consistent with the HPLC results (data not shown).

3.2.6.2. Functional glycosylation of V4-34-encoded scFv-Cκ proteins

The V4-34 gene encodes a germline glycosylation site, NHS in CDR2. Four tumours expressing Ig derived from V4-34 were from CLL (CLL1 and CLL2) and PCNSL (PCNSL1 and PCNSL2). All V region genes were somatically mutated, but carried only the germline-encoded *N*-glycosylation motif in CDR2 with no introduced motifs (Table 3.7.). Expressed scFv-Cκ proteins showed single sharp bands migrating according to expected molecular mass (Figure 3.15a., lane A). Exposure to PNGase F or Endo H had no effect on migration, consistent with a lack of *N*-glycosylation of the germline-encoded motif (Figure 3.15a., lanes B and C, respectively); the pattern was similar to that of the control MM-derived scFv-Cκ (Figure 3.13.). Data obtained by NP-HPLC analysis of PCNSL1 confirmed that the germline-encoded site in V4-34 was not glycosylated (Figure 3.15b.).

The NHS motif is known to be a favoured sequence for *N*-glycosylation.[463] Although this motif was rarely observed at introduced sites, we identified it in the CDR3 of FL26 (Table 3.3.). In this case, since no other glycosylation sites existed in V_H or V_L, we were able to evaluate functional glycosylation of the motif. As described above, expressed scFv-Cκ from FL26 was susceptible to PNGase F and Endo H (Figure 3.13.), consistent with *N*-glycosylation of this single introduced motif and confirming that this sequence is a glycan acceptor. The lack of functional glycosylation of the germline NHS motif in V4-34-encoded scFv-Cκ proteins is unclear.

3.2.7. Binding of FL-derived scFv-Cκ proteins to recombinant mannose receptor

Assessment of the binding of mannosylated scFv-Cκ to mannose receptor used a recombinant fusion protein incorporating carbohydrate recognition domain 4-7 of the mouse mannose receptor fused to human IgG1 Fc (CRD_{4-7}-Fc).[501] The direct binding of CRD_{4-7}-Fc to scFv-Cκ proteins is shown in Figure 3.16. All FL-derived scFv-Cκ proteins displaying high mannose glycans bound recombinant mannose receptor; there was no

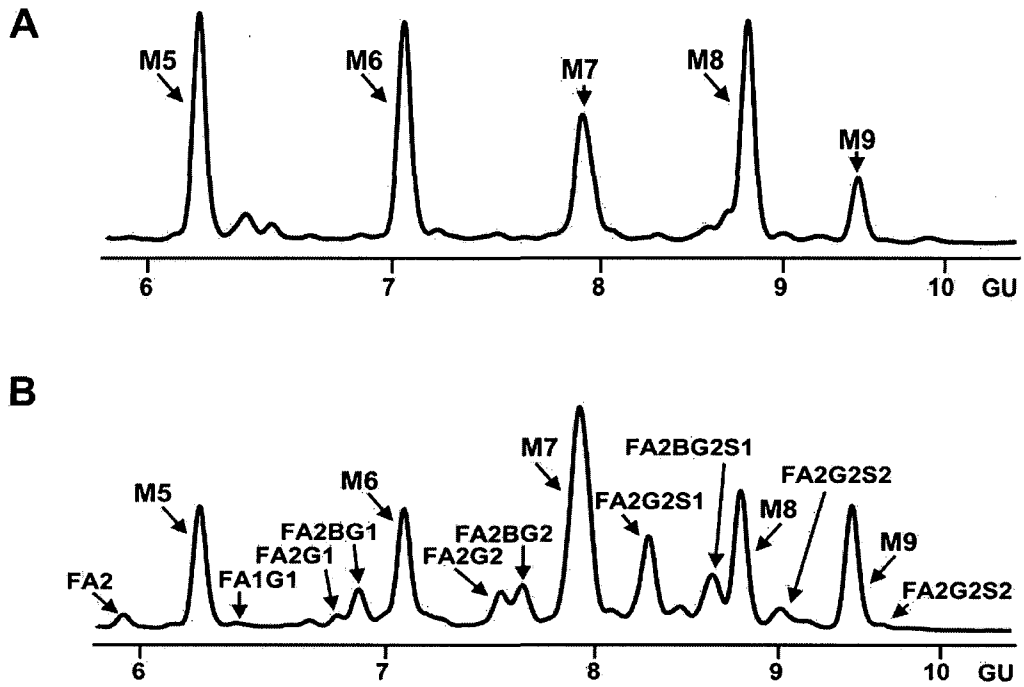


Figure 3.14. Normal phase high performance liquid chromatography (NP-HPLC) glycan profiles of scFv-Cκ from FL21 and FL29. NP-HPLC of two scFv-Cκ proteins confirms A) oligomannose glycans in FL29 and B) oligomannose and complex glycans in FL21. Glycans were assigned glucose units (GU) values, which reflect the relative size of the glycans; peak assignments are A: antennary, number of N-acetyl glucosamine residues, B: bisect, F: α1-6 linked core fucose, G: β galactose, M: mannose and S: sialic acid. *Data courtesy of Mrs C. Radcliffe.*

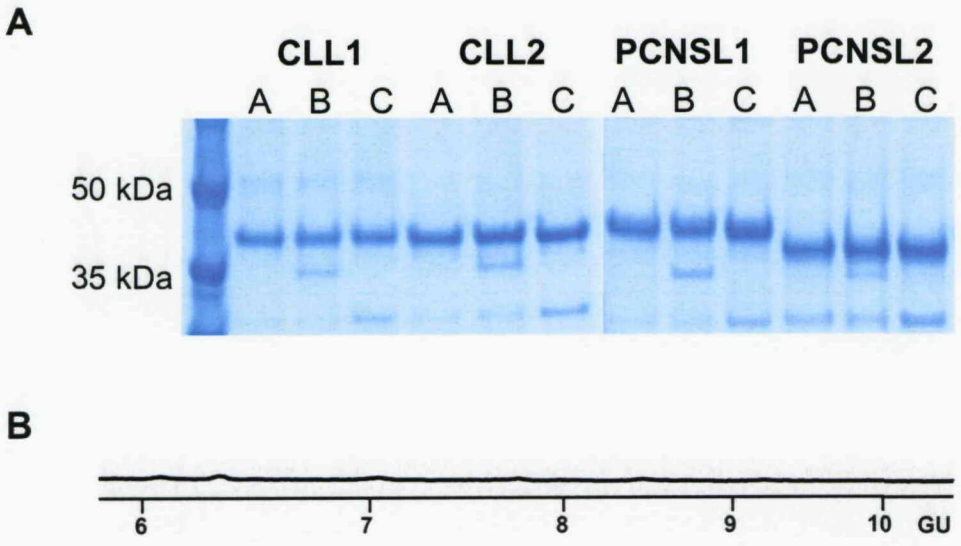


Figure 3.15. A lack of functional *N*-glycosylation of V4-34-encoded scFv-Cκ proteins. A) scFv-Cκ proteins derived from four V4-34-encoded B-cell tumours carrying only the germline NHS *N*-glycosylation site were expressed in 293F cells, purified, separated by SDS-PAGE and visualised by staining with SimplyBlue Safestain™ (lane A). Treatment with the glycan removal enzymes PNGase F (lane B) and Endo H (lane C) had no effect on protein migration revealing a lack of functional glycosylation at the germline motif. Bands at ~36kDa and ~30kDa represent PNGase F and Endo H enzymes, respectively. B) NP-HPLC of one V4-34-encoded scFv-Cκ protein (PCNSL1) confirms that no *N*-glycans are present. *Data courtesy of Mrs C. Radcliffe.*

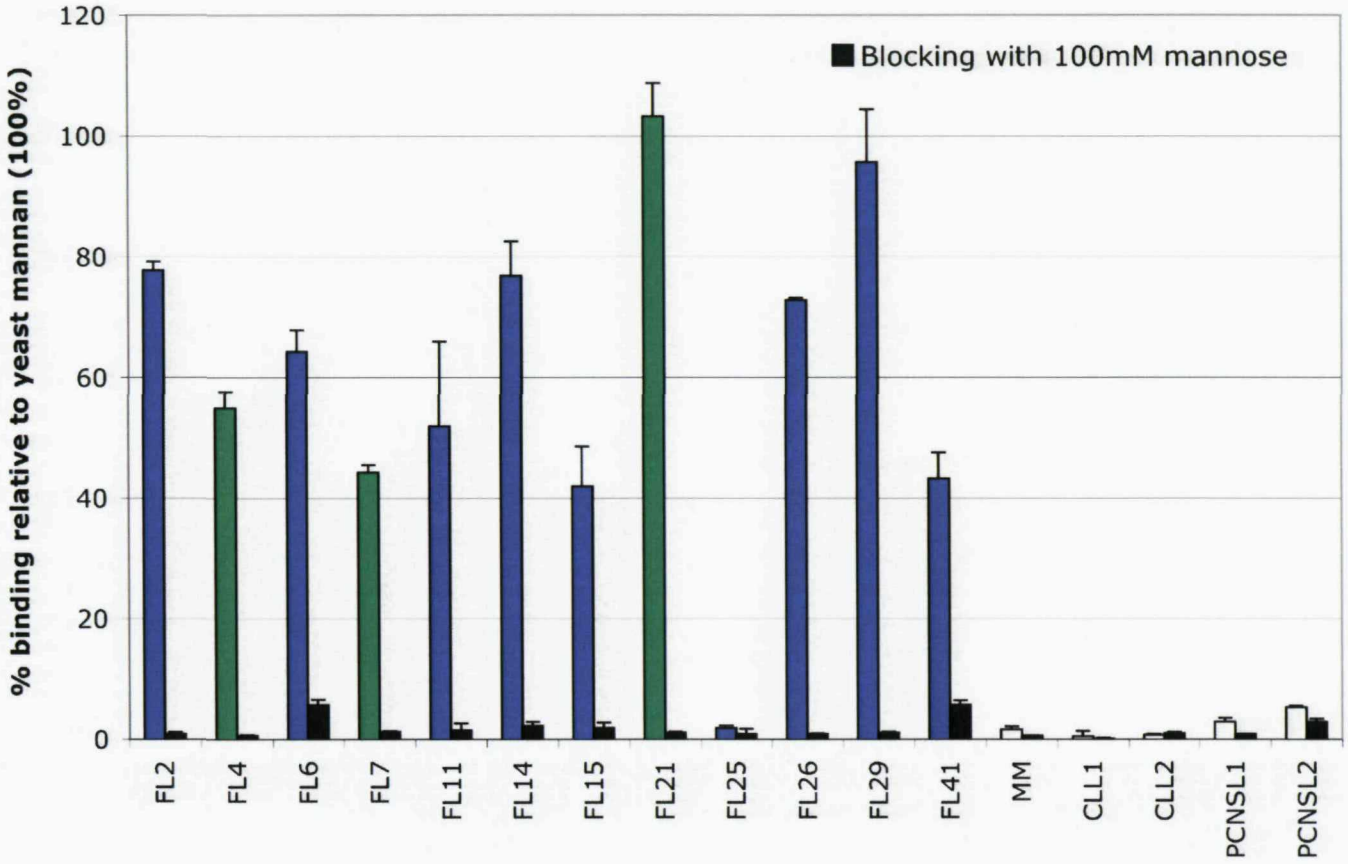


Figure 3.16. Binding of FL-derived scFv-Cκ proteins to recombinant mannose receptor. The binding of tumour-derived scFv-Cκ proteins to CRD_{4-7}-Fc was assayed by lectin ELISA. Eleven of 12 FL-derived scFv-Cκ proteins displaying high mannose glycans, with (green bars) or without (blue bars) additional complex glycans, were able to bind recombinant mannose receptor protein; the remaining case only displayed complex glycans and did not bind to to CRD_{4-7}-Fc. A scFv-Cκ protein derived from myeloma (MM), which lacked an *N*-glycosylation motif, and scFv-Cκ proteins from four V4-34-encoded tumours (CLL1&2, PCNSL1&2), which carried only the germline NHS motif, did not bind. Data are expressed as percentage binding relative to yeast mannan (100%); assays were performed in triplicate.

binding of scFv-Cκ derived from FL25, consistent with the presence of only complex glycans, as determined by digestion with PNGase F and Endo H (Figure 3.13.). The relative binding of CRD_{4-7}-Fc to the different scFv-Cκ proteins was variable (41.9-103.2%). Of the 8 scFv-Cκ proteins displaying oligomannose glycans exclusively, a higher percentage binding relative to yeast mannan was observed with those cases carrying 2 or more oligomannosylated sites (mean 86.7%; FL2, FL29) versus those with a single oligomannosylated site (mean 61.5%; FL6, FL11, FL14*, FL15, FL26, FL21); *considered as possessing a single site based on the assumption that the germline site is not functionally glycosylated (see section 3.2.6.2.). Binding was greatest for FL21-derived scFv-Cκ (103.2%), which carried both high mannose and complex glycans on 5 sites. Conversely, binding was negligible (1.6%) for the MM-derived scFv-Cκ protein, which did not carry any *N*-glycosylation motifs. Moreover, consistent with our previous analyses showing a lack of *N*-glycosylation, V4-34-encoded scFv-Cκ proteins derived from CLL and PCNSL did not bind recombinant mannose receptor (<5%).

To further investigate the specificity of the interaction of FL-derived scFv-Cκ with CRD_{4-7}-Fc, inhibition assays were carried out using D-mannose. In all cases, binding of CRD_{4-7}-Fc to scFv-Cκ could be completely inhibited by pre-incubation with100 mM D-mannose (Figure 3.16.).

3.3. Discussion

In this study we have focused on the characteristics of the V genes of FL. We firstly wanted to establish from our own and published data whether there was a bias in V_H or V_L gene usage. Such a bias is evident in other B-cell malignancies, such as CLL and MALT lymphoma, and likely reflects antigenic drive (see section 1.4.3.). It was important to assess a large cohort of cases to determine bias in gene usage and for V_H this was possible by including data from published studies to assess an extended cohort of up to 262 cases. Overall, the usage of V_H and J_H gene segments appears to reflect that observed in the normal B-cell repertoire, with only minor asymmetries. It is interesting that occasional bias in smaller series, for example to V3-48, was not confirmed in other series. While this might indicate geographical differences, it is more likely to mean that V_H gene bias is not a feature of FL.

The potential contribution of the V_L genes to antigen selection in the normal B-cell repertoire is thought to be less important than that of their partner V_H genes.[48] Unlike the V_H genes, the V_L genes of FL have not been extensively studied and our cohort contributes the largest analysis of V_L sequences to date. So far, our data suggest that the κ:λ ratio in FL is higher than expected for normal B cells, where the ratio between IgLκ and IgLλ - producing B cells is approximately 60:40.[502] Such a Vκ bias in FL is in contrast to the findings of Dogan *et al.* who have reported a κ:λ ratio similar to that of normal B cells based on the immunohistochemical analysis of 26 cases.[503] In keeping with the findings for V_H gene usage, the Vκ and Jκ gene usage appears to share many similarities with that of the normal B-cell repertoire, as does the Vλ and Jλ gene usage, although numbers are too low to be conclusive. More cases are needed to confidently assess V_L usage, however, initial indications are that, as with V_H, V_L gene bias is not a feature of FL.

In 3 (7.3%) cases, 2 potentially functional Vκ gene rearrangements were detected. Analysis on multiple RNA extracts and a common V_H sequence from the tumours on each occasion supports that in these cases the tumours have rearranged two Vκ chains. In each case, we find higher levels of SHM in one of the Vκ sequences, suggesting that a second rearrangement took place at a later time, as has been reported to be possible in normal B cells.[97] These tumours represent the first examples of V_L biclonality in FL. Biclonality for V_H has previously been reported by Noppe *et al.* who described a biclonal IgG-bearing tumour displaying clonality for both the V_H3 and V_H4 families, but without the molecular confirmation of one light chain.[369] The existence of dual receptor B cells involving the

light chains has been described in the normal human peripheral repertoire and in patients with CLL and illustrates that normal and malignant B cells can continue to rearrange κ and λ light chain gene loci even after functional rearrangement and expression on the cell surface.[51,504-507] Here we have examined V gene rearrangements from tumour cell populations. Therefore, it is impossible to know whether both Vκ genes are found rearranged in each cell of the tumour clone or if there exists different subpopulations within the tumour clone, a proportion of which express one Vκ rearrangement and the remainder which express the second. Moreover, if single cell analysis were to find the former to be true, the question would remain as to whether there is dual expression of both functional Igs on the cell surface or if phenotypic exclusion plays a role to ensure the expression of a single sIg.

Almost all V genes are somatically mutated with evidence of intraclonal heterogeneity indicating ongoing mutational activity and consistent with the GC origin of FL. As expected, more mutations are found in V_H than V_L.[368,492,499] In contrast to a previous study we find no difference in SHM levels between lymphomas of different isotypes.[370] Duplications, deletions and insertion events are thought to be linked to the hypermutation process have been observed in the V genes of normal B cells and in FL, where they are reported to occur at a higher frequency.[369,508-510] However, our findings of a single codon deletion in 80 V genes support that these are rare events in FL.

The remarkable similarity of V_H and V_L gene usage between FL cells and normal B cells suggests that the initial genetic event occurs randomly in B cells irrespective of the antigenic target the rearranged BCR recognises.[305] Indeed, the t(14;18) likely represents the first step in tumourigenesis and this occurs in the BM at the stage of Ag-independent D-J joining.[305] Since t(14;18) is not sufficient for tumour development and can also occur in the B cells of normal, healthy individuals, a second genetic event must occur.[291,491] Here we confirm that the most dramatic distinction between the V genes of normal B cells and those in FL lies in the accumulation of motifs available for addition of oligosaccharides in the latter. An unusually high incidence of novel *N*-glycosylation sites in the V_H genes of FL introduced as a result of SHM was first reported by colleagues in this laboratory.[12] Normal plasma and memory B cells have low levels of motifs, as do non-functional V_H sequences, pointing to positive selection in FL.[12] Furthermore, we have recently assessed the V_H genes of 38 polyclonal GC B cells infiltrating FL14 and found that the incidence of motifs is similarly low (~13%) (data not shown). We now reinforce our previous observation and by taking into account both the V_H and V_L sequences we have a view of

the BCR as a whole. Our data show that almost all tumours acquire a novel *N*-glycosylation site in the V genes of their BCR and that these sites cluster within the CDRs. We have found only 2 cases that lack a novel site in the V gene consensus sequences. Even here there appears to be a drive towards such mutations with the finding that in one case motifs are present in a substantial number of molecular subclones; both V_H and V_L sequences were affected such that up to 30% of tumour cells may express glycosylated sIg. Therefore, only 1 case of morphologically and immunohistochemically typical stage IVB FL completely lacked potential sites. Furthermore, we demonstrate that all 12 tumour-derived scFv-Cκ proteins that were expressed in the mammalian 293F system contained glycans, which were susceptible to enzymatic removal *in vitro*. It also appears that more than 1 site may be glycosylated in any one sequence, as is shown for FL21, which has 5 sites. This finding supports the conclusion that sites defined by molecular means are likely glycosylated in FL cells. Whether disfavoured motifs, which were observed in 4 cases, contribute to variable region glycosylation is difficult to assess at this point as in all cases additional motifs using a favoured aa sequence were detected. Importantly, we have also demonstrated that for 11/12 FL-derived scFvs the added glycans are unusual in terminating at oligomannose; either in the presence (n=3) or absence (n=8) of additional complex glycans. Our data using expressed scFv-Cκ proteins from naturally encoded human FL cells appear to reproduce the oligomannosylation pattern seen in the V regions of whole Ig rescued from FL cells of 6 previous cases.[466] These findings reveal consistent differential glycosylation of V region sequences with likely biological significance.

Functional importance for GC-associated lymphomas is questioned by the presence of natural *N*-glycosylation sites in certain germline V region genes. The main example is the V4-34 gene, used by ~4-10% of normal B cells and by a significant proportion of cases of CLL, a tumour not associated with the GC.[343] If the natural site in V4-34 were also to carry oligomannose, the importance of this feature for GC-associated lymphomas would diminish. In this study, we show that not only is there no oligomannose glycan present in the V4-34 sequence, but that no sugars are added to the nascent polypeptide. This finding is consistent with the observation that there is no selective pressure to maintain the germline site in V4-34-encoded FL, which is frequently lost by somatic mutation. Conversely, at least one acquired site is completely conserved by molecular clones in the majority of tumours. The germline V4-34-encoded motif in CDR2 is evidently unavailable for the addition of the initial dolichol phosphate-linked oligosaccharide precursor ($GlcNAc_2Man_9Glc_3$) in the ER. The NHS motif is known to be a favoured sequence for *N*-glycosylation and, as we show here, can act as an acceptor of glycans, therefore, the reason

for the lack of oligosaccharide addition is unclear.[463] A recent statistical analysis of the local sequence and structural environment of NHS motifs identified a number of factors that predispose towards or against glycosylation, including the accessibility of the Asn side-chain in the folded structure, whether Ser or Thr lies at position +3 and the presence or absence of acidic residues in the immediately preceding positions (–1 and –2).[511] Low accessibility of the site in the ER during the addition of the oligosaccharide precursor, Ser at position +3, and an acidic residue at position –1 or –2, all mitigate against glycosylation.[511] In the germline-encoded site, accessibility of the Asn residue side-chain is certainly very low; its relative accessibility (the accessible surface area in the folded structure expressed as fraction of its total surface area when fully exposed) is only 8% as calculated with a 3Å probe radius corresponding to an oligosaccharide unit and using the program NACCESS (http://wolf.bi.umist.ac.uk/unix/naccess.html) and the V4-34 crystal structure (Brian J. Sutton and Paul Hobby, Kings College London, personal communication, Aug 2006).[512,513] Furthermore, there is a glutamic acid residue in the germline sequence at position –2. Introduced sites are more variably placed and apparently do allow addition of the oligosaccharide precursor but inhibit further processing. Although the Asn residue of the somatically generated NWT motif in the immediately adjacent CDR1 loop of V4-34-encoded FL14 is similarly low, there is no preceding acidic residue and Thr is at position +3. In this study, preceding acidic residues are notably absent from 41/44 (93.2%) and 19/21 (90.5%) glycosylated sites in the V_H and V_L genes of FL cells, respectively (see Figure 3.9a. and 3.9b.); however, it is likely that a combination of factors, such as the occurrence of nearby surface-exposed hydrophobic residues, which is difficult to assess in the absence of 3D structural data, may all contribute to determining the overall propensity for glycosylation.[511]

The incidence of introduced motifs appears universal in somatically mutated FL, but the phenomenon is not confined to FL. Our colleagues have also observed sites in a proportion (~40%) of cases of DLBCL and sBL and commonly (82%) in eBL.[13] One conclusion might be that the added glycans have a role in survival or growth of B-cell malignancies in the GC (Figure 3.17.). With regard to FL, the t(14;18) likely represents the initial genetic event, but does not confer tumourigenic behaviour.[514-516] The high retention of sIg expression in FL in the face of this translocation points to a possible role for stimulation via the BCR in pathogenesis. The finding of functional *N*-glycosylation sites within the V genes of the BCR is suggestive. Since *N*-glycosylation sites were shown to cluster in the Ag-binding groove, this exposed position raises the possibility of an interaction of the added oligomannoses with some element(s) of the GC, such as the

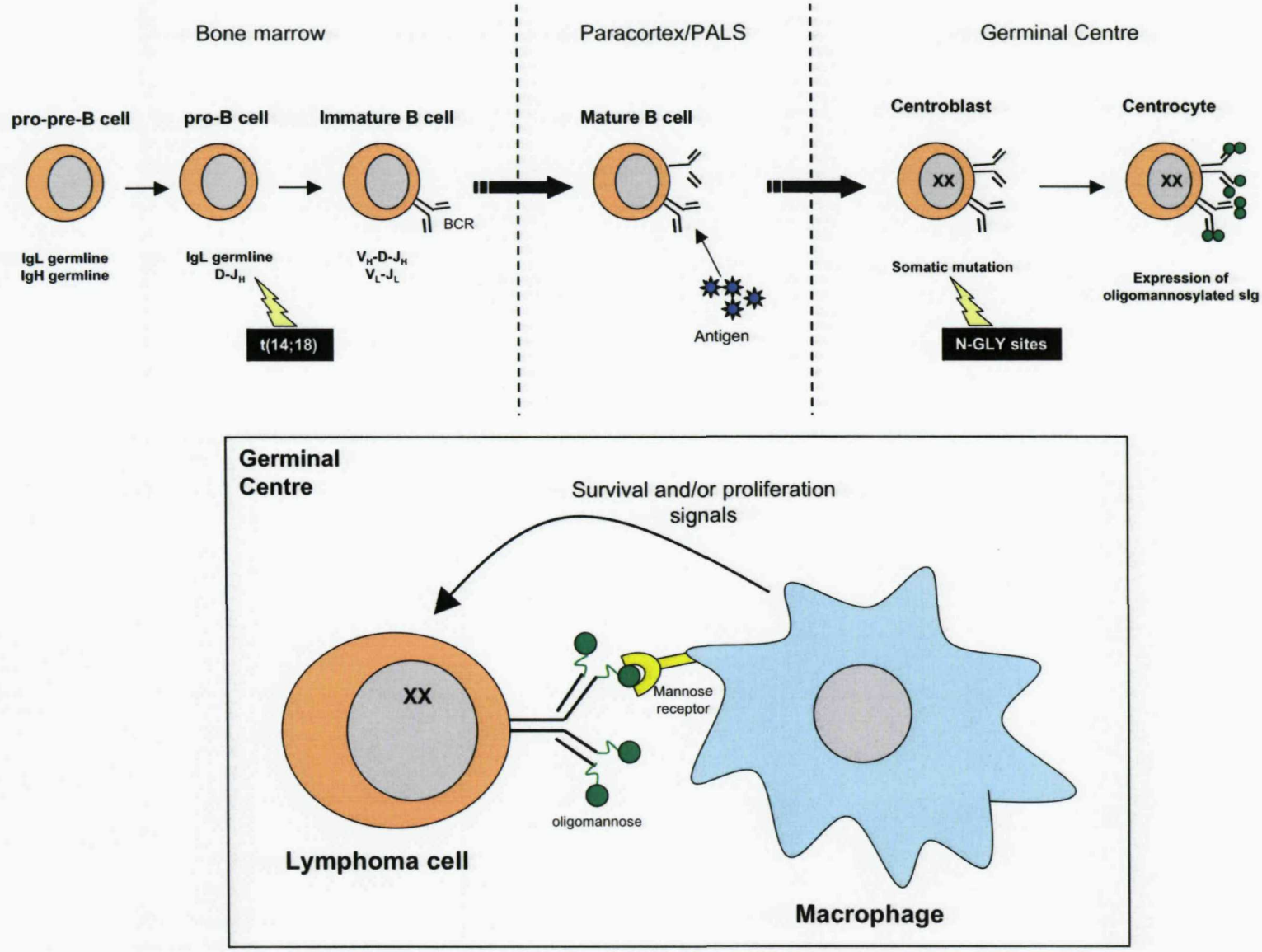


Figure 3.17. The acquisition of *N*-glycosylation sites and the subsequent expression of oligomannosylated surface Ig may provide a second step to tumourigenesis. The t(14;18) occurs during D-J_H recombination in the bone marrow and represents the initial genetic lesion in lymphoma pathogenesis. Following interaction with antigen, the B cell moves into the GC compartment where it undergoes affinity maturation. The acquisition of *N*-glycosylation sites into the V region genes by somatic mutation may represent the second genetic event. Subsequent expression of oligomannosylated sIg may permit the binding of oligomannose on the tumour cell to mannose receptor on macrophages and lead to the growth and/or survival of these tumours (inset).

mannose receptor (MR) or other MBLs. Indeed, recent data from our collaborators has shown that the oligomannose glycans occupying the V regions of FL-derived Ig are accessible and able to bind to MBL.[466] Here we have demonstrated that the oligomannose glycans displayed on FL-derived scFvs are able to bind to recombinant MR. This specific feature is associated only with motifs introduced during somatic mutation and is apparently conserved in B-cell tumours of the GC. This could replace the need for persistent Ag, which would differ in every case and is an unlikely continuing stimulant. The ongoing stimulation via a substitute MBL could account for the ongoing somatic mutation and lead to the growth and/or survival of these tumours within the GC. MR is known to be expressed on macrophages within the GC.[517,518] Consistent with our hypothesis, it has been reported that increased numbers of lymphoma-associated macrophages within the tumour denotes a worse prognosis in FL.[519] It supports further the concept that an interaction with MR expressed by stromal cells in the GC may play a role in lymphoma pathogenesis.

3.4. Conclusions

Overall, these findings reveal a remarkable similarity between V genes in FL cells and normal B cells and suggest that the genetic event leading to transformation occurs randomly in B cells. Although not exclusive to FL, it appears likely that the acquisition of *N*-glycosylation sites and the subsequent addition of oligosaccharides, more specifically oligomannoses, is important to the lifestyle of FL and may provide a critical second tumourigenic event (Figure 3.17.). The presence of oligomannose conferred the ability to bind specifically to recombinant MR protein, suggesting a possible interaction with MRs in the GC. It may be possible to exploit this seemingly essential feature to develop novel therapeutic approaches; for example, by the interruption of binding of oligomannose glycans to lectin-like receptors in the GC.

4. Diffuse large B-cell lymphoma

4.1. Introduction

DLBCL, like FL, is one of the more common B-cell NHLs in adults and accounts for 30-40% of all newly diagnosed cases.[469] The incidence is 7.1 per 100,000 persons, or 20,000 to 30,000 new cases each year, increasing by 1.4% per year in patients aged 75 years or older.[520] It can develop *de novo* or arise within the context of a pre-existing indolent lymphoma, such as FL. The median age of presentation is 65 years, with a slight male predominance. At diagnosis, one third of patients present with early-stage disease (stage I-II using the Ann Arbor classification) and may have extranodal involvement, which is observed in ~40% of cases.[318] BM involvement is seen in 15-20% of cases.[318] Approximately 40-50% of all DLBCL patients can be cured by conventional cyclophosphamide, doxorubicin, vincristine and prednisone chemotherapy.[521] The remainder have tumours that are either refractory to current available treatments or have a relapse after a period of remission and most of these patients will die of the disease. Some improvement in disease-free and overall survival may be obtained with the addition of monoclonal Ab against CD20, such as Rituximab.[reviewed in 521]

DLBCLs represent a diverse group of lymphoid neoplasms, encompassing heterogeneous clinical, histologic, immunophenotypic, cytogenetic and molecular genetic features. Equally, they exhibit widely differing responses to treatment. Previous attempts to subclassify these neoplasms have suffered from a lack of reproducibility, leading to their categorisation as a single group in the REAL classification.[318] Several clinicopathological entities have since been recognised within this group of lymphomas and, subsequently, these were listed as separate entities in the WHO classification; primary mediastinal large B-cell lymphoma, intravascular large B-cell lymphoma and primary effusion lymphoma.[286] All other large B-cell lymphomas, not fitting into one of these three entities, remain grouped under the term of DLBCL, with the understanding that several subtypes are likely to exist but clear criteria to differentiate them from each other are lacking.

It is important to identify at presentation those patients who do not benefit from current treatment regimens in order that risk-adjusted therapies can be developed. Until recently, the most effective tool for predicting outcome of patients with DLBCL was the International Prognostic Index (IPI).[522] The IPI is a clinical index based on five independent prognostic factors, including age (>60 years), Ann Arbor tumour stage (III to IV), serum lactate dehydrogenase levels (>1 times normal), performance status (≥2) and

the number of extranodal sites (>1). The subdivision of patients according to the number of prognostic factors into low risk (none or one factor), low-intermediate risk (two factors), high-intermediate risk (three factors) or high risk (four or five factors) with predicted 5 year overall survival rates of 73%, 51%, 43% and 26%, respectively, became the most widely used and accepted prognostic model for DLBCL.[522] However, the outcome of patients with identical IPI values varies considerably, likely owing to the marked genetic and molecular heterogeneity that underlies disease aggressiveness and tumour progression. Therefore, there is a need for more accurate prognostic indicators. Multiple studies have tried to associate the survival of DLBCL patients with specific molecular factors underlying the hallmark features of lymphoma cells. The clinical impact of particular chromosomal translocations, involving *Bcl-6*, *Bcl-2* and *c-Myc* proto-oncogenes, as well as the expression of several markers, including Bcl-6, Bcl-2, CD10, HGAL and p53, have been evaluated in patients with DLBCL. However, the available data are contradictory. Generally, the expression of Bcl-6 is considered to predict survival in patients, as is the presence of point mutations in the 5' non-coding region of the *Bcl-6* gene.[523,524] The prognostic significance of a rearrangement at the *Bcl-6* locus on chromosome 3q27 is less clear, with several studies suggesting it is a good prognostic factor while others claim it has a bad or indifferent effect on patient outcome.[525-527] It has since been demonstrated that *Bcl-6* translocations involving the Ig loci are associated with increased Bcl-6 protein expression and a good prognosis, whereas non-Ig/*Bcl-6* gene fusions correlate with a low Bcl-6 protein expression and a poor prognosis, which may explain the previous discrepancies.[528 529] It is uncertain how these variables interact in their effect on prognosis and there is as yet no consensus as to how this additional information can be used, in combination with the IPI, to improve the risk stratification of patients at presentation. Furthermore, it is important to note that the overall outcome of patients is determined not only by factors relating to the tumours pathogenesis, as described above, but also to the tumours response to treatment.

4.1.1. Morphology

DLBCL is a neoplasm of large (mean diameter ≥20μm), transformed B cells with a diffuse growth pattern and a high proliferative fraction. According to the WHO classification, morphological variants include centroblastic, immunoblastic, anaplastic, plasmablastic, T-cell/histiocyte rich and anaplastic lymphoma kinase (ALK)-positive DLBCL.[286] DLBCLs typically express one or more of the pan B cell markers CD19, CD20, CD22 and CD79a. The expression of CD5 (<10%), CD10 (30-40%) and CD23 (~20%) are more variable, reflecting the marked heterogeneity of this tumour; $CD5^+$ tumours are often associated

with a centroblastic morphology.[390] CD30 expression may occur, most commonly in the anaplastic variant. The majority (50-75%) of tumours express monotypic surface Ig and are typically $sIgM^+$, but may be $sIgG^+$ or $sIgA^+$; a lack of sIg expression is observed in primary mediastinal large B-cell lymphoma, the significance of which is unclear.[390] By contrast, the morphologic variant T-cell/histiocyte rich B-cell lymphoma is characterised by increased numbers of polyclonal $CD3^+$ T cells infiltrating the tumour and fewer than 10% neoplastic B cells.

4.1.2. Molecular analysis

In keeping with the marked differences in disease aggressiveness and tumour progression, DLBCLs show an underlying genetic and molecular heterogeneity. However, the study of Ig V genes in DLBCL has consistently shown that the tumour-derived V_H genes are mutated, suggestive of a GC or post-GC B cell origin.[367,371,391-394] Furthermore, alternative isotype expression has been detected at both the mRNA and protein level, for which the pattern of mutation suggests that switching is occurring in subpopulations of the tumour clone after transformation.[371,399] Conversely, the issue of ongoing SHM is controversial, with some studies demonstrating no intraclonal heterogeneity while others show ongoing mutation; although, evidence for intraclonal heterogeneity appears to predominate.[367,371,391-394] Similar inconsistencies are apparent at the level of V_H gene usage, with some studies reporting a V4-34 gene bias and others a V_H usage similar to normal B cells.[367,394-398] The vast molecular heterogeneity within DLBCL has led several groups to analyse the V genes of specific subsets, such as *de novo* $CD5^+$ DLBCL and PCNSL; PCNSL is discussed further in Chapter 5.

Approximately 50% of DLBCLs carry one of two primary molecular lesions, identified by the activation of either the *Bcl-6* or the *Bcl-2* proto-oncogene and leading to the deregulated expression of the GC transcription factor Bcl-6 and the anti-apoptotic protein Bcl-2, respectively. The t(3;14)(q27;q32) translocation is reported to be present in ~20-35% of DLBCL and is more commonly found at relapse (68%) than diagnosis (14%).[530] Similarly, the t(14;18)(q32;q21) translocation is observed in 20-30% of DLBCLs.[531] This incidence may in part be explained by the transformation of low grade FL, which commonly carries the t(14;18), to DLBCL in a significant number of cases (40-70% over nine years).[476-478] Furthermore, *Bcl-2* gene expression may be altered by other mechanisms, namely gene amplification, which is found in 20-30% of DLBCL.[532]

4.1.3. Gene expression profiling in DLBCL

DNA microarrays are a new technology used to measure the expression of tens of thousands of genes simultaneously, enabling a comprehensive evaluation of gene expression. They can be used to discriminate previously unrecognised disease subgroups with distinct biological and clinical behaviour and, as such, offer the potential to further classify DLBCL. Several microarray studies have been performed on *de novo* DLBCL and have identified two main prognostically different subgroups.[403,405,406] The founding study by Alizadeh *et al.* used a "Lymphochip" cDNA array, which incorporates 17,856 genes that are preferentially expressed by lymphoid cells and genes with known or suspected roles in cancer biology, to assess the gene expression patterns of 42 DLBCLs by an unsupervised hierarchical clustering learning method.[403] Two distinct subgroups were revealed based on the expression patterns of genes characteristic of normal GC B cells (GC signature genes) and activated peripheral blood B cells (activated B cell signature genes). Those tumours expressing GC signature genes were named GC B cell (GCB)-like DLBCL, while those expressing activated B cell signature genes were named activated B cell (ABC)-like DLBCL. These two subgroups were shown to have significantly different outcomes, with 76% of GCB-like DLBCL patients alive at 5 years compared to only 16% of ABC-like DLBCL patients. This prognostic significance was independent of IPI score, with the survival advantage for GCB-like DLBCL also observed in patients with a high-risk IPI. The demonstration that tumours with a GC gene expression profile exhibited intraclonal heterogeneity in their Ig V genes, in contrast to those of the ABC-like subgroup that do not, soon followed.[404] Thus, gene expression profiling and ongoing mutational analysis both support the idea that the initial transforming event leading to GCB-like DLBCL occurs in a B cell within the GC microenvironment and does not alter the ability of these cells to mutate their Ig V genes.

Support for the two different subgroups was provided in a second, larger study, which analysed 240 DLBCL by a similar means.[405] In addition, a third, heterogeneous group representing 22% of cases was also described and comprised tumours that did not express genes of either the GC or activated B cell signature. This subgroup, named Type 3, had a poor outcome similar to that of the ABC-like DLBCL and is to date not further studied. The ABC-like subgroup was noted for the high expression of target genes of the NK-κB transcription factors, suggesting that constitutive activity of the NF-κB pathway may contribute to the poor prognosis of these patients.[533] On the contrary, the *Bcl-2* rearrangement t(14;18)(q32;q21) was found to occur exclusively in the GCB-like subgroup; this has since been confirmed in several studies and sees ~35% t(14;18)

positivity in this subgroup.[405,534,535] Furthermore, the t(14;18) positive GCB-like tumours shared a strikingly similar gene expression profile, suggesting they represent a unique subset within the group of GCB-like DLBCL. However, the expression of Bcl-2 protein, which is seen in ~80% of DLBCL, is not significantly different between the two groups.[536] The high expression of Bcl-2 might also result from the amplification of the *Bcl-2* gene at 18q21, a phenomenon which is reported in up to 30% of DLBCL and is generally found in tumours belonging to the ABC-like subgroup.[532] Although this study failed to demonstrate any significant correlation of GCB-like and ABC-like subgroups with the morphological variants previously defined by the WHO classification, some trends were observed; GCB-like DLBCLs were commonly centroblastic and monomorphic, whereas the ABC-like subgroup comprised immunoblastic and centroblastic polymorphic tumours.[405]

A third study addressed the need to determine the relative contribution of individual genes to survival in DLBCL patients using the Affymetrix oligonucleotide array, which incorporated 6,817 human genes.[406] Two groups of patients with cured versed fatal/refractory disease were used to create a supervised learning classification approach in an effort to develop an outcome predictor model. The authors discovered a set of 13 genes that accurately predicted DLBCL outcome even within specific IPI risk categories. Of interest, there was no overlap of genes in the models derived from oligonucleotide and Lymphochip studies. Latterly, in an attempt to devise a technically simple model that would be applicable for routine clinical use, Lossos *et al.* evaluated the expression of 36 genes that had been reported to predict survival in DLBCL patients.[537] The top six genes, ranked on the basis of their predictive power, were used to construct a model based on their relative contribution: LMO2, Bcl-6 and FN1 predicted a longer survival and CCND2, SCYA3 and Bcl-2 predicted a short survival. This six gene model could identify approximately one third of all DLBCL patients with a 5 year survival of less than 27% who could benefit from intensified or investigational therapeutic modalities.

Overall, these findings support the view that that the various subgroups represent different diseases that arise as a result of distinct mechanisms of malignant transformation- GCB-like and ABC-like tumours arise from B cells at different stages of differentiation (GC versus post-GC B cells), apparently utilise different oncogenic pathways (t(14;18) versus NF-κB activation) and have distinct clinical outcomes (good versus poor).

4.1.3.1. Immunohistochemical analysis using tissue microarray

Microarray analysis is not feasible in routine diagnostic work and, as such, immunohistochemical methods to discriminate between GCB-like and ABC-like DLBCL have been investigated by several groups.[407,538] Hans *et al.* proposed a model based on the expression of three markers: CD10 and Bcl-6 (GC B-cell markers) and MUM-1 (activated B-cell marker) (Figure 4.1a.).[407] The authors demonstrated that a $CD10^{+/-}Bcl\text{-}6^{+}MUM\text{-}1^{-}$ expression pattern could identify those tumours with a GCB-like phenotype, while a $CD10^{-}$ $Bcl\text{-}6^{+/-}MUM\text{-}1^{+}$ expression pattern represented non-GCB(/ABC)-like tumours; the expression of CD10 is specific, but not sensitive, to establish GCB-like DLBCL. This model had positive predictive values of 87% and 73% for correctly identifying GCB-like and ABC-like DLBCL subgroups, respectively, and could predict patient survival. However, comparison to gene expression profiling revealed a 20% misclassification rate and suggests a need for incorporating additional markers. In a model proposed by Chang *et al.*, CD138, a marker for activated B cells, was used in combination with CD10, Bcl-6 and MUM-1 to identify a third subgroup, named 'activated' GCB-like DLBCL (Figure 4.1b.).[538] Similar to Hans *et al.*, a GCB-like phenotype was characterised by CD10 and/or Bcl-6 expression, without the expression of activation markers. In contrast, the second group was further divided into two distinct subgroups: a non-GCB(/ABC)-like subgroup, expressing activation markers MUM-1 and/or CD138 but none of the GC markers and the 'activated' GCB-like subgroup, expressing at least one GC marker and one activation marker.

4.1.4. Glycosylation of V_H genes in GC-related malignancies

First described for FL, and latterly for eBL, a high incidence of *N*-glycosylation motifs, acquired by the process of somatic mutation, has been described in the V genes of some GC-related malignancies (see Chapter 3).[12,13,460] This is in contrast to the V genes of normal plasma and memory B cells and those of MM, mutated CLL and MALT lymphomas, which have a low frequency (<10%) of sites. Further studies on FL-derived Ig have shown these sites to be functional, but unusual in their expression of oligomannose glycans.[466] The positive selection of B cells that contain *N*-glycosylation sites suggests a potential role for the added glycans in the tumourigenesis of GC-associated malignancies. We postulate that an interaction of these glycans with MBLs, such as the MR, expressed by stromal cells in the GC may lead to enhanced survival or growth of malignant B cells (see Chapter 3). It has been shown that ~40% of DLBCL cases acquire *N*-glycosylation motifs in their V_H genes, but it is not known which disease subset they may be associating with, if any, or what their role might be.[13]

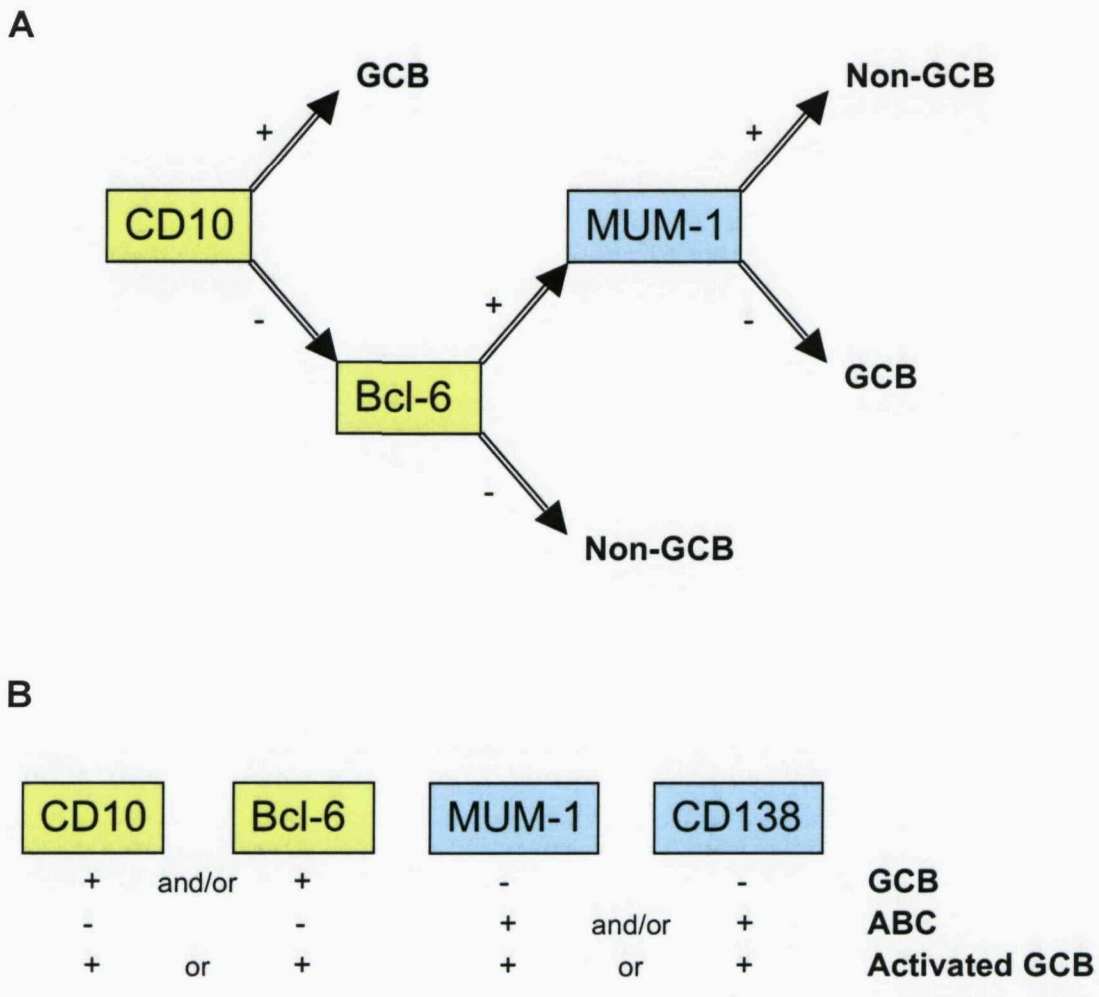


Figure 4.1. Two models for the classification of DLBCL by immunohistochemistry. A) The decision tree proposed by Hans *et al.*, 2004. for the identification of GCB-like and non-GCB-like DLBCL based on the expression of CD10 and Bcl-6 and MUM-1.[407] B) A model proposed by Chang *et al.*, 2004. for the identification of GCB-like, ABC-like and 'activated' GCB-like DLBCL based on the expression of CD10 and Bcl-6 and MUM-1 and CD138.[538] GC markers are shown in yellow boxes, while activated B cell markers are shown in blue boxes.

4.1.5. Aims

The aim of this study was to determine whether the introduction of *N*-glycosylation motifs by somatic mutation is restricted to the V genes of *de novo* DLBCLs of the GCB-like phenotype, as we might expect from previous data on FL and eBL. To achieve this, we aimed to use tissue microarray and the classification approach proposed by Hans *et al.*, as well as the lymphochip cDNA microarray, to divide our tumours into GCB-like and non-GCB-like DLBCL and, subsequently, to evaluate the presence/absence of *N*-glycosylation motifs in each subgroup. We wished to further correlate these findings with the presence of chromosomal translocations, in particular, t(14;18) and t(3;14), involving Bcl-2 and Bcl-6, respectively, which are common to DLBCL.

4.2. Results

4.2.1. Tumour-derived V gene usage

Using cDNA isolated from biopsy material as a source, we identified the matched V_H and V_L tumour-derived gene sequences in 35/36 (97.2%) cases (Table 4.1.). In the remaining case we were able to identify V_L, but not V_H. With the exception of one V gene rearrangement, all were in frame and did not harbour stop codons and were, therefore, and considered potentially functional; V_L of Ly/117/01 contained 3 stop codons.

4.2.1.1. V_H gene usage

In total, 37 potentially functional, clonal V_H gene sequences were identified from 35 cases of *de novo* DLBCL (Table 4.1.); two clonal V_H sequences were found in 2 tumours (5.5%). Biclonality for V_H, but without the molecular confirmation of one light chain, has previously been reported for 5/53 (9.4%) DLBCL.[394] Twenty-two (43.1%) of the 51 known functional V_H gene segments were found rearranged in our cohort. They derived from 6 of the 7 human V_H gene families in the following distribution: V_H3, 43.2% (n=16); V_H4, 35.1% (n=13); V_H1, 8.1% (n=3); V_H5 and V_H6, each 5.4% (n=2); V_H7, 2.7% (n=1). V_H segments from the V_H2 family were not observed. In terms of V_H family usage, the distribution of our 35 cases largely reflected the available germline repertoire and that found for normal B cells, with some exceptions (Figure 4.2.).[97,493-495] Notably, a reduction in V_H1 family genes and a corresponding increase in V_H4 family genes was found in our cohort. Our findings largely paralleled those from a key study by Lossos *et al.*, which examined V_H gene usage in 58 DLBCLs (Figure 4.2.).[394]

A biased usage of the V4-34 gene segment has been described for DLBCL in several studies.[367,395] In our cohort, the V4-34 gene segment was the most frequently used gene, observed rearranged in 5 (13.5%) tumours. Other commonly used gene segments included V3-48 (n=4) and V4-39 and V3-21 (both n=3), which, together with V4-34, accounted for 40.5% of rearrangements. Previous studies in DLBCL have also identified the frequent use of V4-34, V4-39 and V3-48, of which some of these genes, including V4-34 and V4-39, are also found at a higher than expected frequency in normal individuals.[335,394,494] In general, the V_H gene usage in DLBCL reflects that of the normal B-cell repertoire, in line with previous reports.[394,396-398]

The CDR3 region varied in length from 18 to 81 nucleotides (mean 38.4bp, median 39bp), according to the Kabat numbering system.[496] D segments could be confidently assigned in 21.6% (8/37) of tumours, according to the rule proposed by Corbett *et al.*[20] The 8 genes

Table 4.1. V_H and V_L gene usage in *de novo* DLBCL

	V_H					V_L			
Case	**V_H Family**	**V gene***	**D gene***	**J_H gene***	**CDR3 length (bp)**	**V_L Family**	**V gene***	**J_L gene***	**CDR3 length (bp)**
Ly34/01#	V_H3	V3-48	-	J_H4b	42	Vκ1	HK102/L12a	Jκ1	27
	V_H3	V3-48	-	J_H4b	30				
Ly61/01	V_H3	V3-48	-	J_H3b	45	Vκ3	DPK21/L2	Jκ4	27
Ly117/01	V_H3	V3-21	-	J_H6b	45	Vλ1	DPL5/1b°	Jλ2	33
Ly135/01	V_H3	V3-74	-	J_H3b	81	Vλ2	DPL11/2a2	Jλ2	30
Ly136/01	V_H3	V3-23	-	J_H4b	42	Vκ1	Ve+/L9	Jκ4	27
Ly157/01	V_H6	V6-01	-	J_H6b	51	Vκ6	DPK26/A26	Jκ2	27
Ly235/01	V_H4	V4-34	-	J_H4a	18	Vλ4	IGLV8A1/4b	Jλ1	27
Ly268/01	V_H4	V4-04	-	J_H6b	57	Vκ4	DPK24/B3	Jκ4	27
Ly284/02	V_H3	V3-07	-	J_H3a	36	Vκ3	DPK22/A27	Jκ1	27
L-BRA	V_H4	V4-34	-	J_H5b	57	Vκ1	Ve+/L9	Jκ2	30
L-FUL	V_H4	V4-39	-	J_H4b	27	Vκ4	DPK24/B3	Jκ4	30
L-MAY#	V_H7	V7-04.1	-	J_H4b	27	Vλ2	V1-2/2c	Jλ2	27
	V_H5	V5-51	-	J_H5a	24				
F4881	V_H3	V3-07	-	J_H6c	36	Vκ1	DPK9/012	Jκ2	27
F5049	V_H4	V4-34	-	J_H4a	27	Vκ3	DPK22/A27	Jκ1	27
F5126	V_H3	V3-11	-	J_H4a	36	Vλ1	DPL8/1e	Jλ1	33
F5184	V_H3	V3-21	D3-22	J_H4b	30	Vλ1	DPL2/1c	Jλ3b	33
F5246	V_H3	V3-48	D6-25	J_H4b	24	Vκ1	HK102/L12a	Jκ1	27
F5334	V_H1	V1-02	-	J_H5b	42	Vλ10	V1-20+/10a	Jλ3b	33
F5370	V_H3	V3-30	D3-03	J_H4a	42	Vκ4	DPK24/B3	Jκ2	27
F5434	V_H3	V3-66	D3-16	J_H3b	51	Vλ2	V1-2/2c	Jλ3b	30
F5440	V_H6	V6-01	-	J_H5b	39	Vκ4	DPK24/B3	Jκ5	27
F5481	V_H3	V3-49	-	J_H6c	21	Vλ3	DPL23/3r	Jλ2	30
E3633/93	V_H4	V4-34	-	J_H6b	45	Vλ1	DPL8/1e	Jλ2	33
E1520/94	V_H1	V1-03	-	J_H4b	21	Vκ3	Vg38K/L6	Jκ4	27
E1792/97	V_H3	V3-21	D6-19	J_H5b	48	Vλ1	DPL7/1e	Jλ2	39
E1296/98	V_H3	V3-09	-	J_H4b	45	Vλ1	DPL5/1b	Jλ2	33
E1297/98	V_H4	V4-39	-	J_H4b	45	Vλ2	DPL10/2b2	Jλ2	27
E1670/98	V_H4	V4-28	-	J_H4b	30	Vκ1	DPK8/L8	Jκ2	15
E4567/99	V_H4	V4-30.1	D3-03	J_H5b	51	Vκ6	DPK26/A26	Jκ4	15
E4708/99	V_H1	V1-03	-	J_H4b	30	Vκ6	DPK26/A26	Jκ4	27
E2009/01	V_H5	V5-51	D3-22	J_H4a	27	Vκ4	DPK24/B3	Jκ1	27
E2555/01	V_H4	V4-59	-	J_H4a	27	Vλ2	V1-2/2c	Jλ2	30
E6376/01						Vλ2	DPL11/2a2	Jλ2	30
E1437/02	V_H4	V4-61	-	J_H6b	42	Vκ4	DPK24/B3	Jκ3	27
E100/03	V_H4	V4-39	-	J_H6b	27	Vλ1	DPL2/1c	Jλ1	33
E1790/03	V_H4	V4-34	D2-02	J_H6b	54	Vκ1	DPK9/012	Jκ1	33

*; Germline V, D and J genes with the greatest homology.

-; Germline D gene could not be confidently assigned.

#; Two potentially functional clonal V_H rearrangements were identified.

°; Non functional V_L rearrangement (3 STOP codons at aa positions 2, 36 and 86).

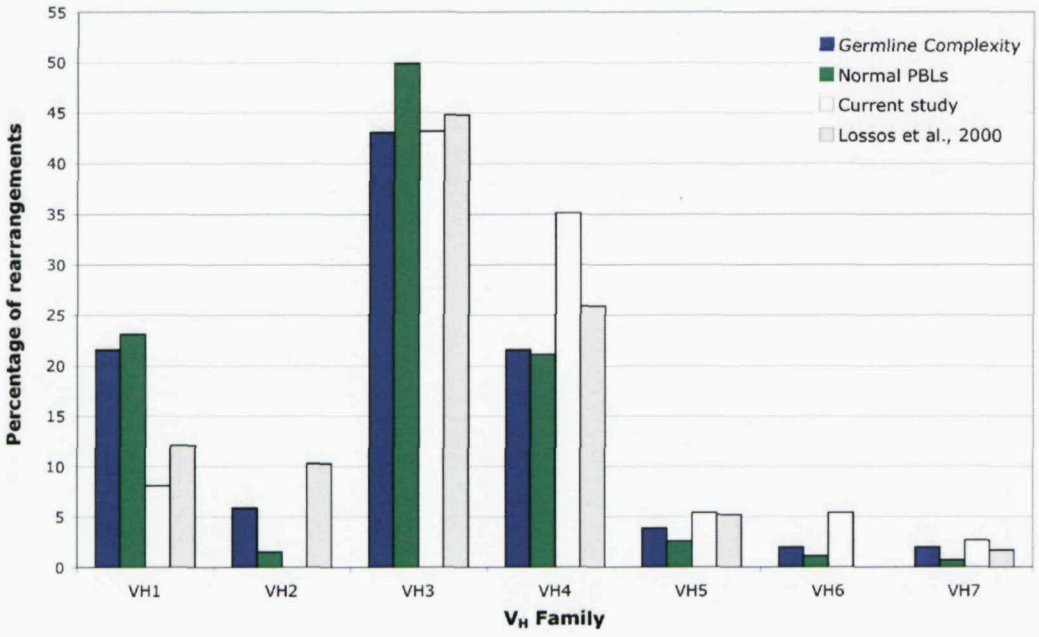


Figure 4.2. V_H gene family usage in *de novo* DLBCL compared to normal B cells. The V_H gene family usage was analysed for our 35 DLBCL and 1322 normal, healthy PBLs. Data were compared to findings previously reported in the largest study of DLBCL by Lossos *et al.*, 2000 (n=58).[394] Values for each V_H family are expressed as a percentage of the total rearrangements assessed. Analysis of normal, healthy PBLs included data from Brezinschek *et al.*, 1995, Brezinschek *et al.*, 1997, Huang *et al.*, 1992, and Demaison *et al.*, 1995.[97,493-495]

derived from the D segment families D3 (n=5), D6 (n=2) and D2 (n=1). These families are known to be preferentially recombined in normal B cells.[20,94,493,494] J_H usage in DLBCL is also similar to that observed in normal B cells: J_H4, 51.4% (n=19); J_H6, 21.6% (n=8); J_H5, 16.2% (n=6); J_H3, 10.8% (n=4).[94,97,493-495]

4.2.1.2. V_L gene usage

Thirty-five potentially functional, clonal V_L sequences were identified from 36 cases (Table 4.1.); in one tumour, the V_L sequence contained 3 stop codons rendering it non-functional. Of these, 20/36 (55.6%) used a κ chain and 16/36 (44.4%) used a λ chain. Nine (22.5%) of the 40 known functional Vκ gene segments were found rearranged in our cohort; all were derived from the proximal region and used 4 of the 6 human Vκ gene families in the following distribution: Vκ1, 35.0% (n=7); Vκ4, 30.0% (n=6); Vκ3, 20.0% (n=4); Vκ6, 15.0% (n=3). Vκ segments from the Vκ2 and Vκ5 families were not observed. In view of the germline size of the Vκ2 gene family, which contributes 9 (22.5%) of 40 gene segments to the germline repertoire, this family appears vastly under represented in DLBCL. Conversely, the Vκ4 family, which contributes only a single gene segment (2.5%) to the repertoire, was over represented. In general, the usage of Vκ1, Vκ3 and Vκ5 gene families appears to be in accordance with germline complexity. Vκ gene family usage shows some similarities to that found for normal B cells, but the over representation of Vκ4 and Vκ6 families and the under representation of the Vκ2 family is not paralleled by normal B cells (Figure 4.3.).[48,75,105] The Vκ6 gene family was over represented as a result of the preferential usage of a single gene, DPK26/A26. Most frequently rearranged gene segments were DPK24/B3 (n=6) and DPK26/A26 (n=3), which together accounted for 45.5% of κ sequences. The DPK24/B3 gene is also found at high frequency in normal B cells.[48,75,105] Vκ gene usage has not been previously studied in DLBCL. The 20 κ rearrangements used all 5 Jκ gene segments in the following distribution: Jκ4, 35.0% (n=7); Jκ1, 30.0% (n=6); Jκ2, 25.0% (n=5); Jκ3 and Jκ5, both 5.0% (n=1). The Jκ gene usage appears similar to that observed in normal B cells.[48,75,497]

Ten (32.3%) of the 31 known functional Vλ gene segments were found rearranged in our cohort; they derived from 5 of the 10 human Vλ gene families in the following distribution: Vλ1, 43.8% (n=7); Vλ2, 37.5% (n=6); Vλ3, Vλ4 and Vλ10, all 6.3% (n=1). Vλ segments from the Vλ5, Vλ6, Vλ7, Vλ8 and Vλ9 families were not observed. The Vλ3 gene family, which contributes the largest number (9; 29.0%) of genes to the germline repertoire, appears under represented in DLBCL, as do the genes of the Vλ5 family.

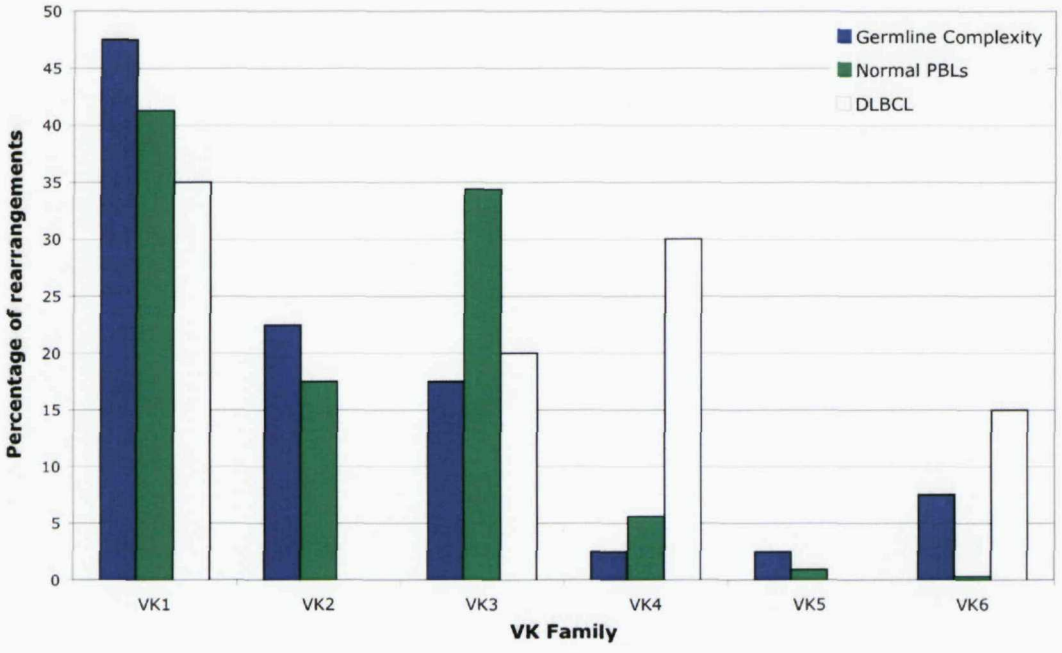


Figure 4.3. Vκ gene family usage in *de novo* DLBCL compared to normal B cells. The Vκ gene family usage was analysed for 20 DLBCL and 950 normal, healthy PBLs. Values for each Vκ family are expressed as a percentage of the total rearrangements assessed. Analysis of normal, healthy PBLs included data from Cox *et al.*, 1994, Klein *et al.*, 1993, Foster *et al.*, 1997, and Juul *et al.*, 1997.[48,75,105,497]

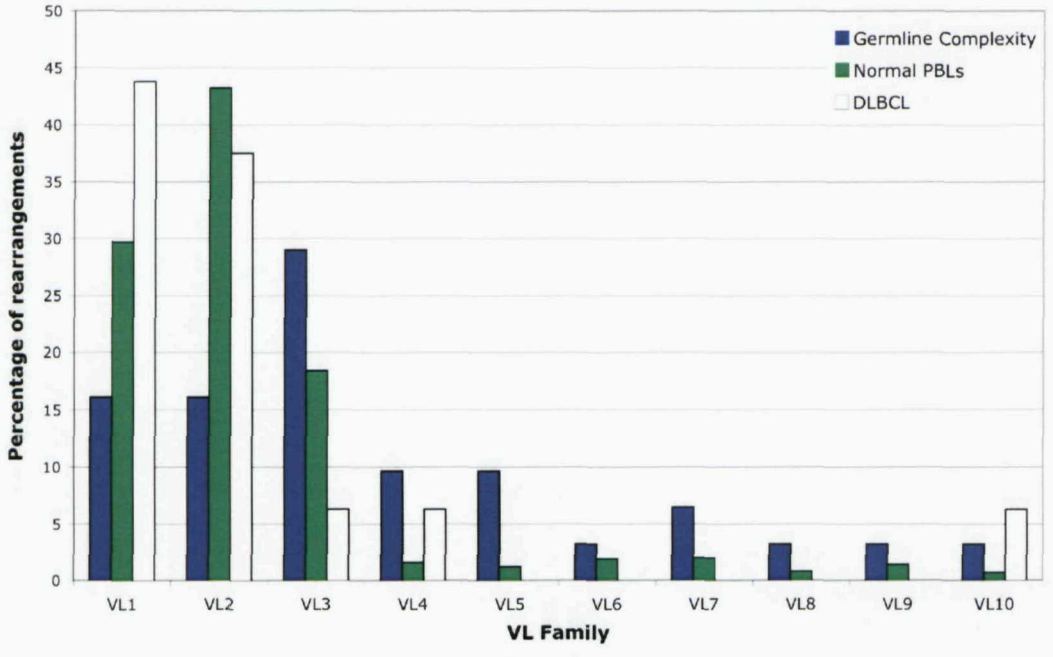


Figure 4.4. Vλ gene family usage in *de novo* DLBCL compared to normal B cells. The Vλ gene family usage was analysed for 16 DLBCL and 7771 normal, healthy PBLs. Values for each Vλ family are expressed as a percentage of the total rearrangements assessed. Analysis of normal, healthy PBLs included data from Farner *et al.*, 1999 and Ignatovich *et al.*, 1999.[76,98]

Conversely, the Vλ1 and Vλ2 family genes are over represented. In general, the usage of Vλ4, Vλ6, Vλ7, Vλ8, Vλ9 and Vλ10 gene families appear to be in accordance with germline complexity. Vλ gene family usage in DLBCL is similar to that found for normal B cells, with the exception that the extent to which Vλ1 family genes are over represented and Vλ3 family genes are under represented is more pronounced in DLBCL (Figure 4.4.).[48,75,105] The most frequently rearranged gene segment was V1-2/2c (n=3), which accounted for 18.8% of λ sequences. Eight of the ten Vλ genes found in this cohort, including V1-2/2c, have been shown to be preferentially rearranged in normal B cells.[48,75,105] Vλ gene usage has not been previously studied in DLBCL. The 16 λ sequences used the Jλ gene segments in the following distribution: Jλ2(/3a), 62.5% (n=10); Jλ3(b) and Jλ1, both 18.8% (n=3). Of note, it has been reported that the Jλ2 and Jλ3 genes are used most frequently in normal B cells, followed by Jλ1.[76,273]

Analysis of all 36 functional V_L sequences revealed that the CDR3 region varied in length from 15 to 39 nucleotides (mean 28.6 bp, median 27bp). Independent assessment of κ and λ rearrangements showed that the CDR3 length of the rearranged Vλ genes was ~2 amino acids longer (mean 10.4aa) compared to Vκ rearrangements (mean 8.8aa) consistent with studies of κ and λ rearrangements in normal B cells and other B-cell malignancies (see sections 3.2.1.2. and 5.2.1.1.).[48,75,76]

Within the matched pairs of V_H and V_L there was no detectable tendency for selective combinations. For example, the V4-34 gene was found in combination with Vκ1, Vκ3, Vλ1 and Vλ4 family gene segments (Table 4.1.).

4.2.2. Somatic hypermutation in tumour-derived V genes

Mutational status is shown in Table 4.2. Thirty-four of 37 (91.9%) V_H sequences and 33/36 (91.7%) V_L sequences showed evidence of somatic mutation (>2%). This is in contrast to that of normal PBLs where we see a lower number of B cells with mutated V genes, 40% and 23% of V_H and Vκ genes, respectively; this disparity is expected since the population of circulating PBLs in normal, healthy adult donors consists of both Ag-experienced, activated B cells and naïve B cells.[48,97] A wide range mutation levels were observed in DLBCL: V_H, 1.0-19.4% (mean 10.0%); V_L, 0.7-22.5% (mean 7.4%). We observed a mutational load of 6% or greater in 64.9% of V_H genes and 52.8% of V_L genes; this is higher than that found for normal B cells where we see 13.7% and 4.7% of V_H and V_L genes, respectively.[48,97] Although mutational load was generally higher in λ (mean

Table 4.2. An analysis of somatic mutation in the V genes of *de novo* DLBCL

Case	V_H Isotype*	V_H Mutation (%)≠	V_H Intraclonal heterogeneity No. clones°	V_H Intraclonal heterogeneity %	V_L Isotype*	V_L Mutation (%)≠	V_L Intraclonal heterogeneity No. clones°	V_L Intraclonal heterogeneity %
Ly34/01	-	16.7	14/17	0.510^R	κ	13.3	7/10	0.280^R
	γ	19.4	8/12	1.160^R				
Ly61/01	δ	11.2	1/10	0.136	κ	3.0	3/7	0.884^R
Ly117/01	α	6.8	9/11	0.684^R	λ	18.7	8/11	0.856^R
Ly135/01	-	19.4	2/12	0.083	λ	13.3	4/10	0.121
Ly136/01	γ	5.2#	1/7	0.073#	κ	9.1	7/12	0.487^R
Ly157/01	-	3.0	11/15	0.524^R	κ	8.7	4/8	0.195^R
Ly235/01	μ	16.8	5/12	0.220^R	λ	22.5	0/9	0.000
Ly268/01	μ	11.7	4/8	0.131	κ	6.0	11/11	0.197^R
Ly284/02	μ, δ	18.7	1/10	0.030	κ	10.5	7/8	0.425^R
L-BRA	δ	4.8	2/7	0.075	κ	3.0	1/12	0.026
L-FUL	δ	14.8	4/10	0.112	κ	9.9	4/12	0.146^R
L-MAY	δ	14.6	3/9	0.094	λ	14.1	1/12	0.025
	γ	4.4	4/10	0.234^R				
F4881	-	5.4	14/28	0.384^R	κ	3.4	na	na
F5049	-	10.7	0/10	0.000	κ	6.4	1/12	0.049
F5126	-	9.9	11/18	0.246^R	λ	2.6	na	na
F5184	-	14.0	10/10	2.000^R	λ	7.9	na	na
F5246	-	13.8#	6/8	1.250^R	κ	3.8	na	na
F5334	-	16.7	10/12	1.116^R	λ	3.0	na	na
F5370	-	10.5	16/24	0.719^R	κ	4.6	na	na
F5434	-	4.1	5/7	0.421^R	λ	2.2	na	na
F5440	-	6.3	10/13	0.720^R	κ	11.4	na	na
F5481	-	5.0	4/16	0.214	λ	3.8	0/7	0.000
E3633/93	γ	13.7	1/12	0.025	λ	7.0	1/10	0.066
E1520/94	γ	5.1	11/11	5.022^R	κ	8.7	3/13	1.454^R
E1792/97	μ	1.0	10/13	0.292^R	λ	5.2	2/11	0.088
E1296/98	μ, δ	6.4	3/9	0.151	λ	10.1	2/8	0.090
E1297/98	γ	16.8	3/12	0.097	λ	11.0	0/11	0.000
E1670/98	-	14.1[a]	4/8	0.191	κ	3.0	0/12	0.000
E4567/99	γ	18.9	1/7	0.042	κ	8.0	2/5	0.143^R
E4708/99	γ	12.6	0/16	0.000	κ	2.5	1/10	0.032
E2009/01	δ	1.0	0/12	0.000	κ	0.0	4/10	0.355
E2555/01	-	1.4	3/9	0.118	λ	1.9	0/10	0.000
E6376/01					λ	4.8	1/8	0.042
E1437/02	-	2.7[b]	4/11	0.134	κ	0.7	3/12	0.180
E100/03	α	8.8	1/12	0.028	λ	13.1	0/13	0.000
E1790/03	δ	3.8	7/8	0.362^R	κ	2.3	1/9	0.337

*; Heavy and light chain isotype determined by PCR.

≠; Frequency of somatic hypermutation compared to the most homologous germline V gene segment.

°; Number of clones exhibiting sequence variation from the tumour consensus sequence / total number clones sequenced.

R; Includes at least one intraclonal sequence mutation that is repeated in >1 clone.

-; Indicates isotype not determined.

na; Not assessed.

#; Sequence limited to FR2 – FR4 region.

[a]; Rearrangement exhibits a single codon insertion between amino acid 32-33.

[b]; Rearrangement exhibits 2 single codon deletions at amino acid positions 29 and 31a.

8.8%) rearrangements compared to κ (mean 6.2%) rearrangements, this did not reach statistical significance (P = 0.093; Figure 4.5.). In 25/35 (71.4%) cases where both tumour-derived V_H and V_L were identified, mutation rates were lower in the V_L gene compared to their partner V_H (Figure 4.6.) and this difference was statistically significant ($p<0.05$) in keeping with findings for normal B cells and other B-cell malignancies (see sections 3.2.2. and 5.2.2.).[368,492]

In our cohort of 73 V genes, we observed two single codon deletions and one single codon insertion, which were found in FR1 and CDR1 of the V_H gene from tumour E1437/02 and in the CDR1 of the V_H gene from tumour E1670/98, respectively.

Intraclonal heterogeneity was assessed in independent V gene clones of 36 DLBCLs, with the exception of 8 V_L gene sequences (Table 4.2.). Eighty-one % (30/37) of V_H genes and 60.7% (17/28) of V_L genes had levels of intraclonal heterogeneity that exceeded that of a previously determined taq error rate of 0.05%. Overall, 91.4% (32/35) of tumours, where both V_H and V_L genes had been identified, showed evidence for intraclonal heterogeneity indicative of ongoing SHM. The level of intraclonal heterogeneity varied widely: V_H 0.025-5.022% (mean 0.518%), V_L 0.025-1.454% (mean 0.294%). Intraclonal heterogeneity was also observed in 2/3 V_H gene sequences and 2/3 V_L gene sequences that had low overall levels of mutation (<2%); mutational activity was verified in all three tumours.

4.2.3. Assessment of *N*-glycosylation sites in tumour-derived V genes

4.2.3.1. Incidence and location of novel *N*-glycosylation sites

V_H and V_L sequences were examined for potential *N*-glycosylation sites with the motif Asn-X-Ser/Thr, where X could be any aa except Pro (Table 4.3.). In 12/36 (33.3%) V_H sequences and 6/35 (17.1%) V_L sequences novel *N*-glycosylation sites had been acquired. Twenty-one different *N*-glycosylation motifs were observed, the most common were NIT and NSS (both n=3; 2 V_H and 1 V_L). The disfavoured motifs NWS and NLS were found in 2 cases, for which the former displayed an additional 3 sites in V_H. Of the 35 tumours where both the V_H and V_L gene sequences were identified, 15 (42.9%) had at least one novel site in the tumour clone. Of these, 9 had sites in V_H alone, 3 in V_L alone and 3 in V_H and V_L. The number of sites per V_H and V_L sequence is shown in Figure 4.7. and sees that the majority had single sites.

Data shows that novel *N*-glycosylation sites were conserved among all V_H and V_L clones, with a single exception (Table 4.3.); for tumour F5246, the NSS site in CDR2 was found in

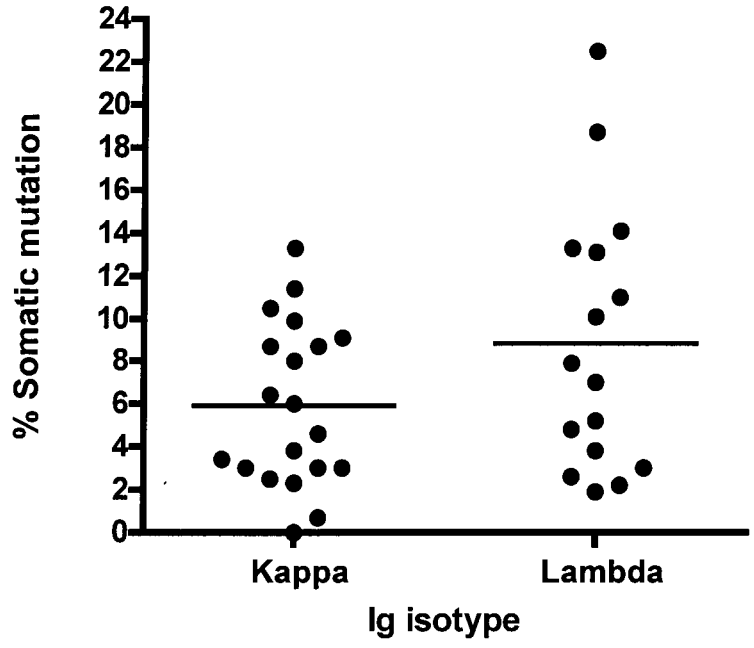


Figure 4.5. Analysis of somatic mutation load and IgL isotype in *de novo* DLBCL. The percentage somatic mutation in Igκ versus Igλ DLBCL. An individual spot represents one V_L gene rearrangement. The lines represent the average percentage mutation of κ^+ DLBCL (n=20) and λ^+ DLBCL (n=16), respectively. P value = 0.093, by unpaired, two-tailed t-test (95% confidence interval).

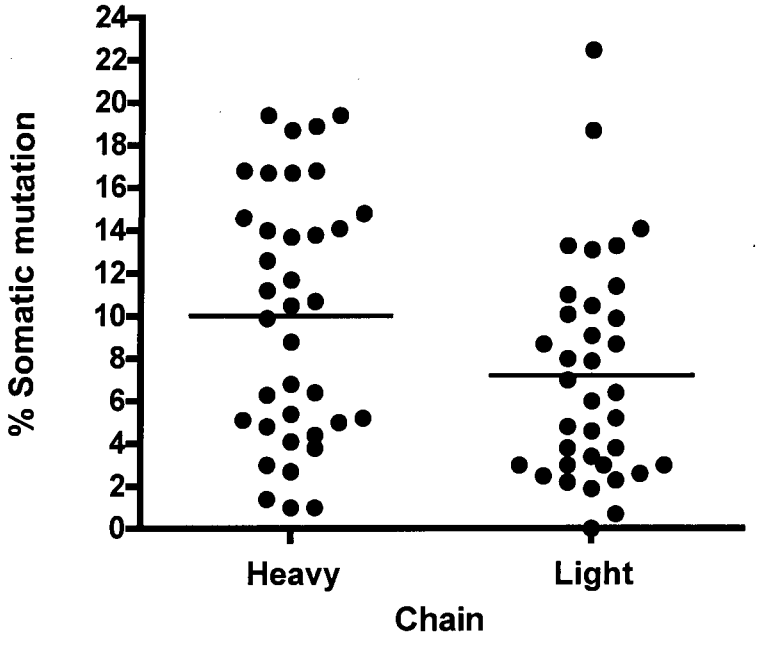


Figure 4.6. Analysis of somatic mutation load and Ig chain in *de novo* DLBCL. The percentage somatic mutation in the heavy chain versus the light chain of DLBCL. An individual spot represents one V gene rearrangement. The lines represent the average percentage mutation of V_H (n=37) and V_L (n=36), respectively. $P < 0.005$, by unpaired, two-tailed t-test (95% confidence interval).

Table 4.3. The incidence of *N*-glycosylation sites in the V genes of *de novo* DLBCL

	V_H				V_L			
Case	**No. Sites**	**Region**	**Motif**	**No. clones with motif***	**No. Sites**	**Region**	**Motif**	**No. clones with motif***
Ly34/01	0	-	-	-	1	FR1	NIT	10/10
	0	-	-	-				
Ly61/01	4	FR1, CDR1, CDR2, CDR3	NFT, NMS, NTS, NVT	10/10, 10/10, 10/10, 10/10	0	-	-	-
Ly117/01	0	-	-	-	0	-	-	-
Ly135/01	1	CDR3	NDT	12/12	0	-	-	-
Ly136/01	0"	-	-	-	0	-	-	-
Ly157/01	0	-	-	-	0	-	-	-
Ly235/01	2°	CDR1, CDR3	NWT, NIT	12/12, 12/12	0	-	-	-
Ly268/01	4	CDR1, CDR2x2, CDR3	NWS, NGS, NYT, NKT	8/8, 8/8, 8/8, 8/8	0	-	-	-
Ly284/02	1	FR1	NFS	11/11	1	CDR1	NTS	8/8
L-BRA	1	$CDR2^{GL}$	NHS	7/7	0	-	-	-
L-FUL	1	CDR3	NVS	10/10	0	-	-	-
L-MAY	0	-	-	-	0	-	-	-
	0	-	-	-				
F4881	0	-	-	-	0	-	-	-
F5049	0°	-	-	-	0	-	-	-
F5126	2	CDR1, CDR2	NMT, NIT	18/18, 18/18	1	CDR1	NSS	na
F5184	0	-	-	-	0	-	-	-
F5246	2"	CDR2, CDR3	NSS, NSS	6/8, 8/8	0	-	-	-
F5334	0	-	-	-	0	-	-	-
F5370	1	CDR1	NLS	24/24	0	-	-	-
F5434	0	-	-	-	0	-	-	-
F5440	0	-	-	-	1	CDR1	NKS	na
F5481	0	-	-	-	0	-	-	-
E3633/93	1°	CDR3	NAS	12/12	1	CDR1	NNS	10/10
E1520/94	0	-	-	-	0	-	-	-
E1792/97	0	-	-	-	0	-	-	-
E1296/98	0	-	-	-	1	FR1	NVT	6/6
E1297/98	0	-	-	-	0	-	-	-
E1670/98	1	CDR2	NIS	8/8	0	-	-	-
E4567/99	1	CDR3	NTT	5/5	0	-	-	-
E4708/99	0	-	-	-	0	-	-	-
E2009/01	0	-	-	-	0	-	-	-
E2555/01	0	-	-	-	0	-	-	-
E6376/01					0	-	-	-
E1437/02	0	-	-	-	0	-	-	-
E100/03	0	-	-	-	0	-	-	-
E1790/03	1	$CDR2^{GL}$	NHS	8/8	0	-	-	-

*; Number of clones displaying the *N*-glycosylation motif / total number clones sequenced.

"; Sequence limited to FR2 – FR4 region.

°; Loss of germline V4-34 *N*-glycosylation site by somatic mutation.

^{GL}; Identifies a germline V4-34 *N*-glycosylation site.

na; Indicates could not be assessed.

Underlining indicates a disfavoured *N*-glycosylation motif for which functional glycosylation is questionable.

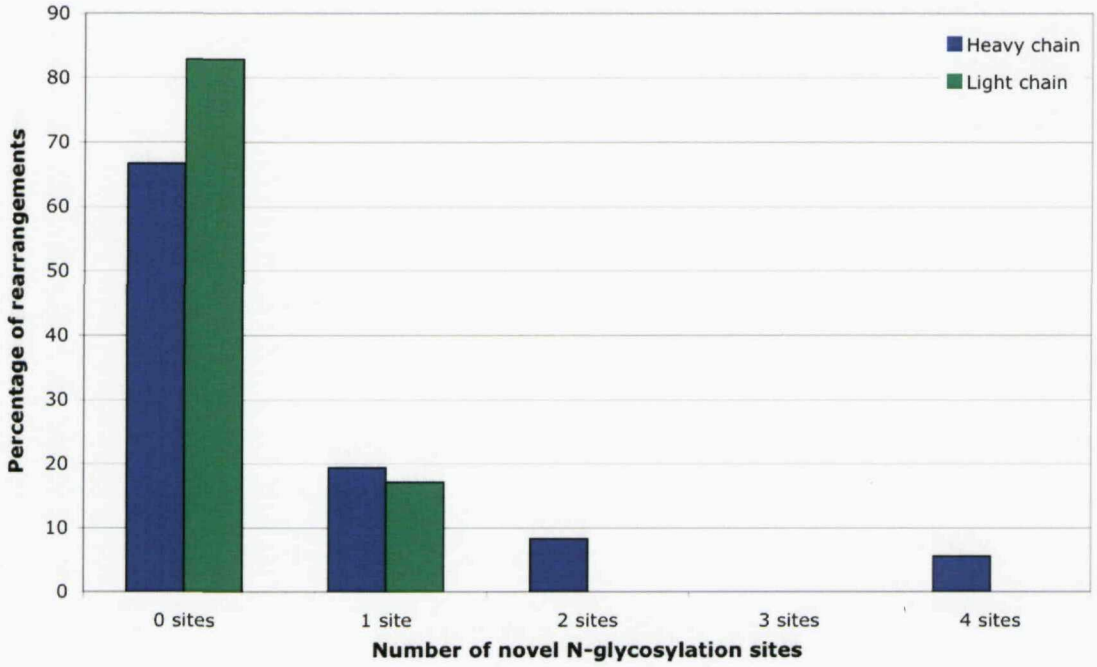


Figure 4.7. The incidence of novel *N*-glycosylation sites in the V region genes of *de novo* DLBCL. The incidence of novel *N*-glycosylation sites within the V_H and V_L gene sequences was analysed (FR1 To FR4) for 12* and 6 DLBCL, respectively. Values are expressed as a percentage of the total number of rearrangements. *Two cases carrying the germline V4-34 site alone are not included in the analysis.

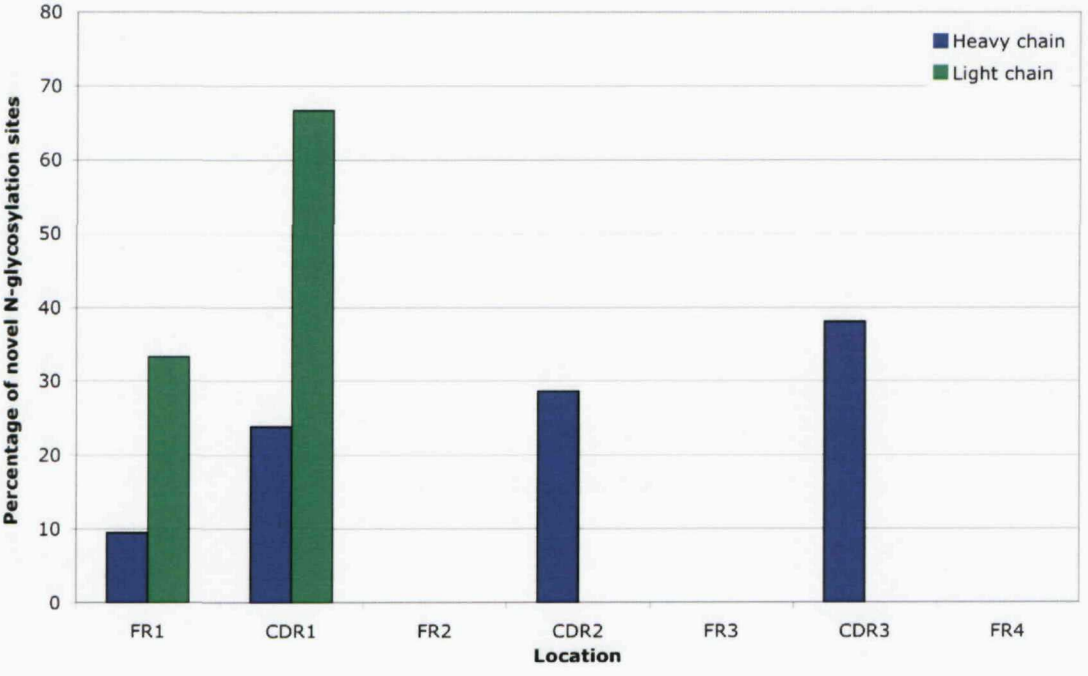


Figure 4.8. The distribution of novel *N*-glycosylation sites in the V region genes of *de novo* DLBCL. The location of novel *N*-glycosylation sites within the V_H and V_L gene sequences was analysed (FR1 To FR4) for 11* and 7 DLBCL, respectively. Values are expressed as a percentage of the total number of rearrangements. *Two cases carrying the germline V4-34 site alone are not included in the analysis.

6/8 V_H clones. However, this tumour had a second site in V_H that was completely conserved. Therefore, for all tumours displaying *N*-glycosylation site(s), at least one motif was fully conserved among molecular clones. Tumour Ly235/01, which had 2 sites in V_H, acquired a second site (NTS) in 1/12 V_H clones by continued somatic mutation. N-glycosylation sites resulting from intraclonal heterogeneity were also evident in 2 tumours that did not contain a site in their consensus V gene sequences; tumour E1520/94 acquired 2 motifs, NYS and NIS, each featuring in 1/11 V_H clones; tumour Ly117/01 acquired an NYT motif in 2/11 V_H clones.

The distribution of novel *N*-glycosylation sites in the V_H and V_L gene sequences is shown in Figure 4.8. The majority (90.0%) of sites in V_H were located in the CDRs: CDR3 (n=8) > CDR2 (n=6) > CDR1 (n=5); FR1 acquired 2 sites. For V_L, 4/6 (66.7%) sites were located in CDR1, the remainder were found in FR1. No sites were observed in FR2, FR3 or FR4 in either V_H or V_L sequences. The focusing of sites within CDR2 and CDR3, followed by CDR1, is similar to that found for FL (see section 3.2.4.).[368,369,460]

4.2.3.2. Incidence of germline-encoded *N*-glycosylation sites

Five V_H gene sequences used the V4-34 gene, which carries a natural *N*-glycosylation site in its germline sequence: NHS motif in CDR2 (Table 4.3.). The germline site was retained by 2 tumours and no other sites were acquired, either in V_H or V_L gene sequences. The remaining 3 cases lost the site by somatic mutation (12/12, 10/10 and 12/12 V_H clones, respectively), of which two gained one (n=1) or two (n=1) novel sites in V_H. The 5 remaining germline V genes that carry natural *N*-glycosylation sites were not detected in our cohort.

4.2.4. Expression of *AID* mRNA in DLBCL

AID mRNA expression was assessed in all DLBCL by standard PCR. A single amplification product of the expected size of ~646bp was detected in 41.7% (15/36) of tumours (Figure 4.9. and summarised in Table 4.4.). PCR products of the correct size were also detected in the positive control, which used cDNA from the Ramos human BL cell line. Direct sequencing of the products revealed 100% homology to the published *AID* mRNA reference sequence (GenBank accession number NM_020661) and confirmed *AID* mRNA expression in these tumours.

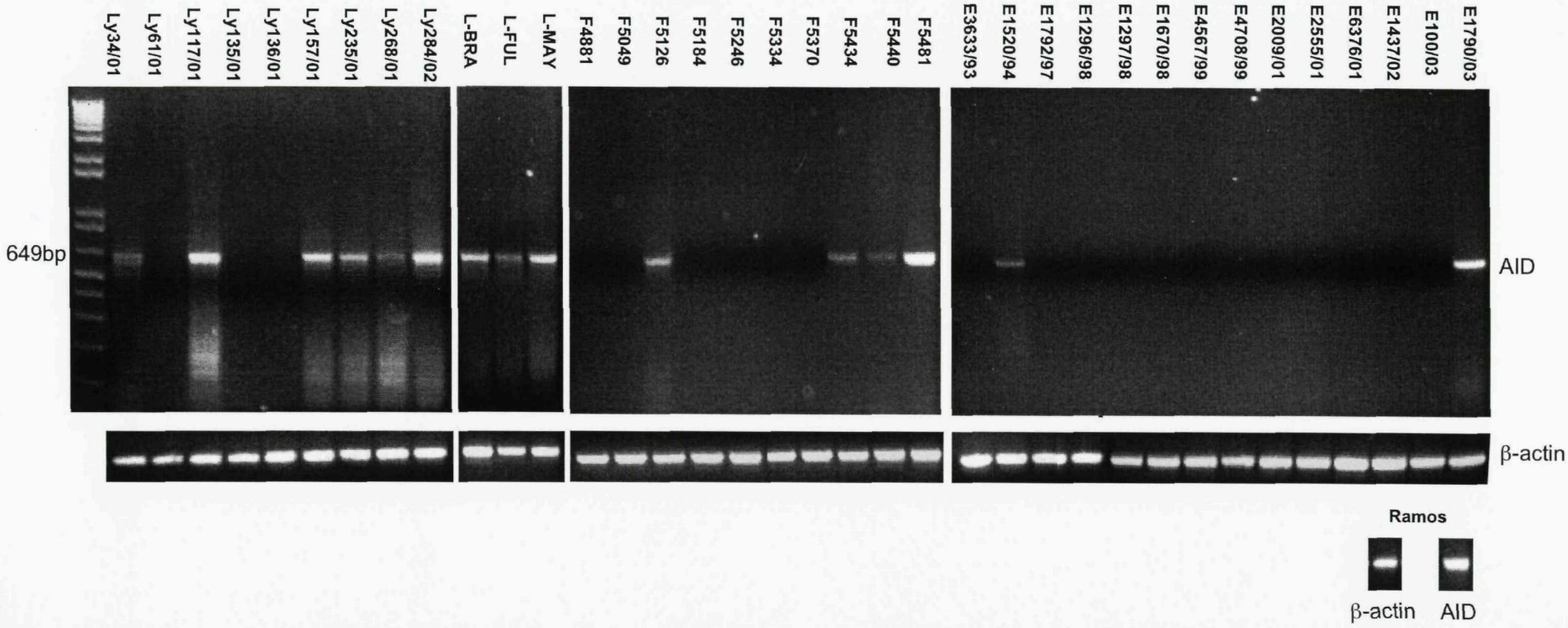


Figure 4.9. AID expression in *de novo* DLBCL assessed by RT-PCR assay. Expression of AID mRNA, corresponding to an amplification product of 649bp, was detected in 15 of 36 tumours. The amount of cDNA amplified for each sample was comparable, as shown by the β actin signal. Expression of AID mRNA in the human BL cell line Ramos was used as a positive control.

Table 4.4. A summary of AID mRNA expression and intraclonal heterogeneity in *de novo* DLBCL

Case	AID expression*	Intraclonal heterogeneity
Ly34/01	Yes	Yes
Ly61/01	No	Yes
Ly117/01	Yes	Yes
Ly135/01	No	Yes#
Ly136/01	No	Yes
Ly157/01	Yes	Yes
Ly235/01	Yes	Yes
Ly268/01	Yes	Yes
Ly284/02	Yes	Yes
L-BRA	Yes	Yes#
L-FUL	Yes	Yes
L-MAY	Yes	Yes
F4881	No	Yes
F5049	No	No
F5126	Yes	Yes
F5184	No	Yes
F5246	No	Yes
F5334	No	Yes
F5370	No	Yes
F5434	Yes	Yes
F5440	Yes	Yes
F5481	Yes	Yes#
E3633/93	No	Yes#
E1520/94	Yes	Yes
E1792/97	No	Yes
E1296/98	No	Yes#
E1297/98	No	Yes#
E1670/98	No	Yes#
E4567/99	No	Yes
E4708/99	No	No
E2009/01	No	Yes#
E2555/01	No	Yes#
E6376/01	No	No
E1437/02	No	Yes#
E100/03	No	No
E1790/03	Yes	Yes

*; *AID* mRNA expression determined by standard PCR.
#; Intraclonal heterogeneity exceeds taq error threshold (0.05%) but does not include mutations that are repeated in >1 clone.

4.2.5. Classification of GCB-like and non-GCB(/ABC)-like DLBCL

4.2.5.1. Tissue microarray

Sufficient paraffin-embedded biopsy material was available for 15/19 Southampton DLBCL cases. The expression of Bcl-2, CD10, Bcl-6 and MUM-1 was evaluated in all 15 cases by immunohistochemistry and revealed 87.5%, 53.3%, 93.3% and 73.3% positivity, respectively (Table 4.5.); this analysis was kindly performed by Dr. I. Oschlies. Tumours were classified based on CD10, Bcl-6 and MUM-1 expression into GCB-like or non-GCB-like DLBCL using the method proposed by Hans *et al.* (Figure 4.1a.).[407] Fifty-three percent (8/15) of tumours were classified as GCB-like DLBCL, while 46.7% (7/15) were classified as non-GCB-like DLBCL, which may include ABC-like or Type 3 DLBCL. The majority of GCB-like cases (7/8; 87.5%) co-expressed CD10 and Bcl-6, either with (n=3) or without (n=3) MUM-1 expression; MUM-1 expression was not evaluated in one tumour. Overall, MUM-1 expression was seen in 4/7 of the GCB-like cases, of which 2 were weakly stained ~30%. All the non-GCB-like DLBCL were characterised by the co-expression of Bcl-6 and MUM-1 in the absence of CD10. Bcl-2 expression was seen in 100% (8/8) GCB-like cases and 71.4% (5/7) of non-GCB-like cases.

4.2.5.2. cDNA microarray

Sufficient total RNA was available for 18/19 Southampton DLBCL cases. Gene expression profiling using the lymphochip cDNA microarray was performed for all 18 cases and enabled the classification of tumours into GCB-like and ABC-like DLBCL (Figure 4.10. and summarised in Table 4.6.); this analysis was kindly performed by Prof. I. Lossos. The majority (14/18; 77.8%) of tumours were classified as GCB-like DLBCL.

4.2.6. Identification of t(3;14) and t(14;18) in DLBCL

Interphase FISH analysis was performed on 18/19 Southampton DLBCL cases to determine the presence or absence of the t(3;14) and t(14;18) translocations (Figure 4.11.); the results from two independent analysts (KM and AS) are summarised in Table 4.7. The t(14;18) translocation, which involves the juxtaposition of the *Bcl-2* gene to the *IgH* gene locus, predominated; t(14;18) was found in 33.3% (6/18) of cases, in keeping with previous reports of the frequency of this translocation in *de novo* DLBCL (Figure 4.11., panel E, F and G).[531] For a single (5.6%) tumour, the t(3;14) rearrangement involving the *Bcl-6* gene was identified (Figure 4.11., panel E and H). A frequency of ~30% has been previously found for this translocation in DLBCL.[530] However, t(3;14) is known to be less common at diagnosis and the apparent under-representation here could simply reflect sample size.

Table 4.5. Classification of *de novo* DLBCL into a GCB-like and non-GCB-like phenotype by tissue microarray

Case	Bcl-2*	CD10*	Bcl-6*	MUM-1*	Phenotype#
Ly34/01	+	na	na	na	unknown
Ly61/01	+	-	+^{w}	+^{w}	non-GCB
Ly117/01	+	+	-	+	GCB
Ly135/01	+	+	+	-	GCB
Ly136/01	+	+	+	-	GCB
Ly157/01	-	-	+	+	non-GCB
Ly235/01	+	+	+	+	GCB
Ly268/01	+	-	+	+	non-GCB
Ly284/02	+	+	+	na	GCB
F4881	-	-	+	+	non-GCB
F5049	na	na	na	na	unknown
F5126	na	na	na	na	unknown
F5184	+	-	+	+	non-GCB
F5246	+	+	+	+^{w}	GCB
F5334	+	+	+	+^{w}	GCB
F5370	+	-	+	+	non-GCB
F5434	na	na	na	na	unknown
F5440	+	-	+^{w}	+	non-GCB
F5481	+	+	+	-	GCB

Data courtesy of Dr. I. Oschlies

*; Positive (+) expression: >30% immunoreactive tumour cells.
#; Phenotype determined using the decision tree devised by Hans *et al.*, 2004.[407]
na; Not assessed due to insufficient biopsy material.
^{w}; Indicates weak expression: ~30% immunoreactive tumour cells.
GCB; Germinal centre B-cell-like.
Non-GCB; Activated B-cell (ABC)-like or 'Type 3' phenotype.

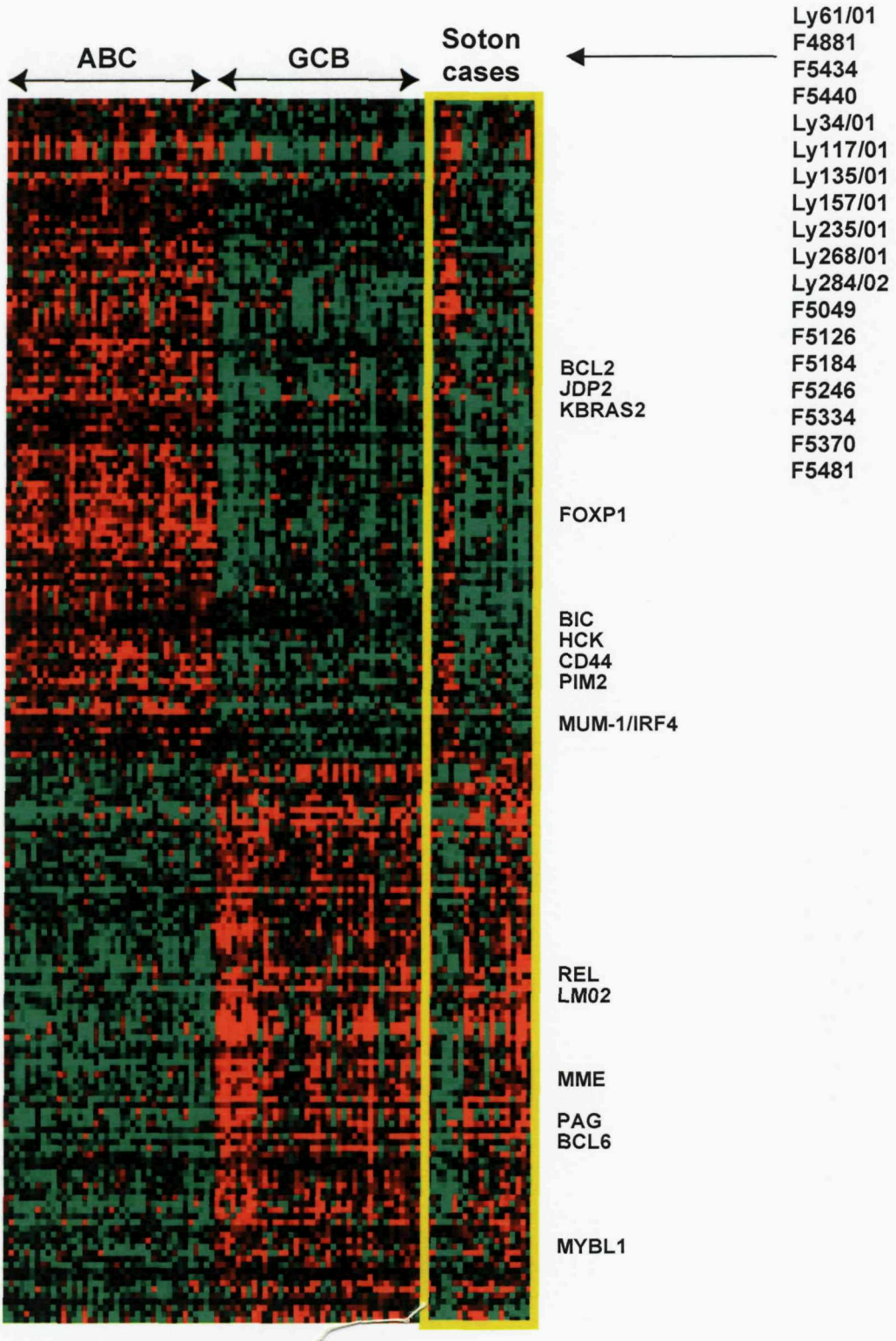


Figure 4.10. Gene expression profiling of *de novo* DLBCL. The lymphochip cDNA microarray was used to quantitate the mRNA expression of 18 DLBCL (right panel). Each row represents a separate cDNA clone on the microarray and each column a separate mRNA sample. The results presented represent the ratio of hybridisation of fluorescent cDNA probes prepared from each experimental mRNA sample to a reference mRNA sample; red indicates a high relative expression, green a low relative expression. *Data courtesy of Prof. I. Lossos.*

Table 4.6. Classification of *de novo* DLBCL into a GCB-like and ABC-like phenotype by cDNA microarray

Case	Phenotype[a]
Ly34/01	GCB
Ly61/01	ABC
Ly117/01	GCB
Ly135/01	GCB
Ly136/01	na
Ly157/01	GCB
Ly235/01	GCB
Ly268/01	GCB
Ly284/02	GCB
F4881	ABC
F5049	GCB
F5126	GCB
F5184	GCB
F5246	GCB
F5334	GCB
F5370	GCB
F5434	ABC
F5440	ABC
F5481	GCB

Data courtesy of Prof. I. Lossos

GCB; Germinal centre B-cell-like.
ABC; Activated B-cell-like.
na; Not assessed due to insufficient RNA.

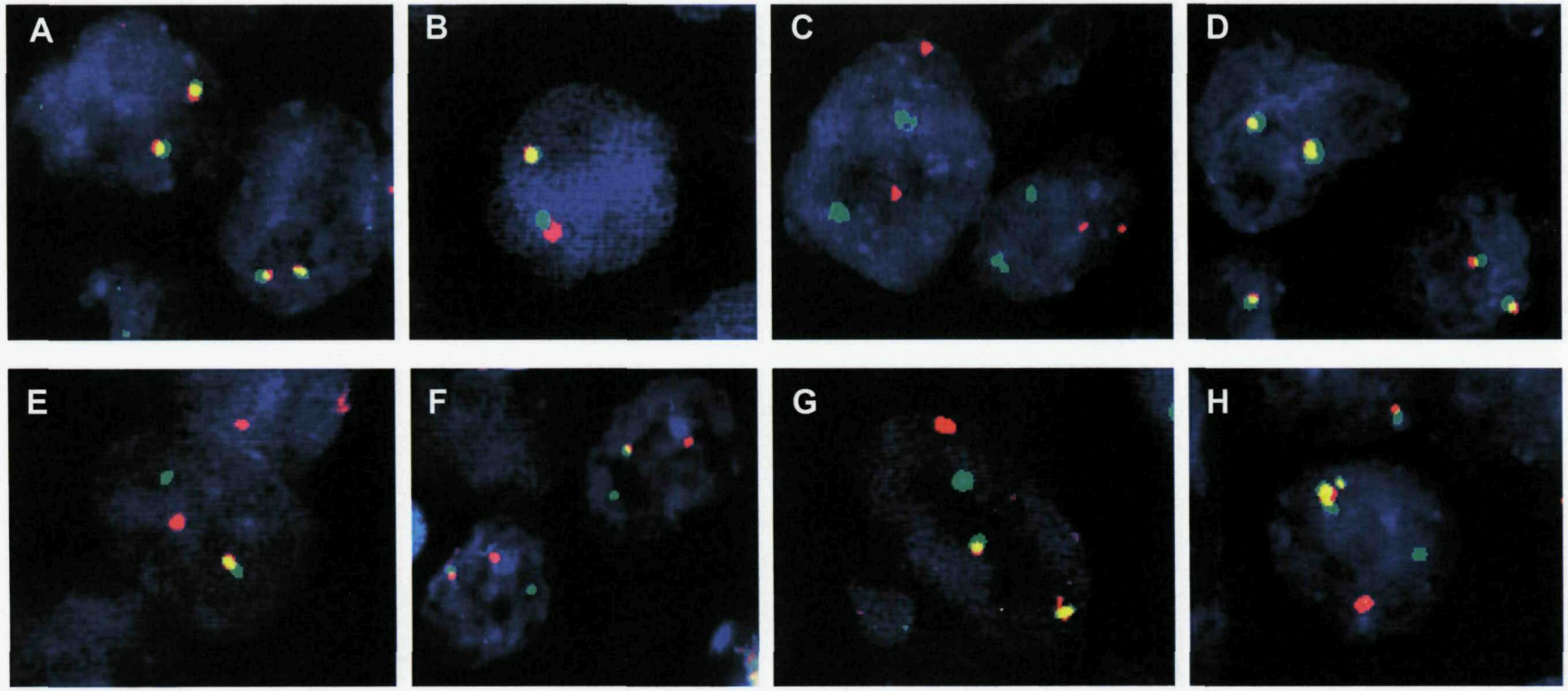


Figure 4.11. Detection of the t(14;18) and t(3;14) by interphase FISH: Representative images of the signal pattern in normal and translocated cells. A, B and D) A normal cell detected using dual colour, break apart rearrangement probes: IGH, BCL2 and BCL6. Two co-localised red and green signals (~yellow) represent the 2 alleles of the non-rearranged chromosome 14, 18 and 3, respectively. C) A normal cell detected using the IGH/BCL2 dual colour, dual fusion rearrangement probe. Two separate signals represent the non-rearranged chromosome 14 (green) and chromosome 18 (red) on 2 alleles. E, F and H) Detection of an *IgH* (F5126), *Bcl-2* (Ly135/01) and *Bcl-6* (F5184) rearrangement using the dual colour, break apart probes: IGH, BCL2 and BCL6. The green signal is separated from the red signal, localised centromeric and distal to the locus of interest. The co-localised red and green signals (~yellow) represent the non-rearranged chromosome 14, 18 and 3, respectively. G) Detection of the t(14;18) in Ly284/02 using the IGH/BCL2 dual colour, dual fusion translocation probe. The juxtaposition of the *IgH* locus on chromosome 14q32 and *Bcl-2* on chromosome 18q21 leads to two fusion signals (yellow). Two separate signals represent the non-rearranged chromosomes 14 (green) and 18 (red). The counter-stain is DAPI, which is visualised as blue fluorescence. X1000 magnification.

Table 4.7. Identification of t(14;18) and t(3;14) in *de novo* DLBCL

Case	IgH break apart*	Bcl2 break apart*	IgH/Bcl2 fusion*	Bcl6 break apart*	Translocation#	
					t(14;18)	t(3;14)
Ly34/01	na	na	na	na	unknown	
Ly61/01	1%	0%	0%	0%	No	No
Ly117/01	74%	79%	70%	0%	Yes	No
Ly135/01	90%	94%	92%	1%	Yes	No
Ly136/01	84%	80%	78%	2%	Yes	No
Ly157/01	5%	0%	1%	2%	No	No
Ly235/01	4%	2%	0%	14%	No	No
Ly268/01	12%	1%	1%	1%	No	No
Ly284/02	88%	94%	89%	2%	Yes	No
F4881	0%	0%	0%	0%	No	No
F5049	1%	0%	3%	0%	No	No
F5126	84%	80%	81%	0%	Yes	No
F5184	89%	0%	1%	88%	No	Yes
F5246	2%	0%	0%	0%	No	No
F5334	4%	2%	0%	2%	No	No
F5370	2%	0%	3%	1%	No	No
F5434	6%	2%	1%	1%	No	No
F5440	6%	0%	1%	8%	No	No
F5481	57%	52%	49%	0%	Yes	No

*; The signal pattern of 200 qualifying interphase nuclei from different areas of the section was assessed by two independent analysts; values are expressed as a % positive cells.

#; The false-positive cut-off level (mean ± 3SD): IGH break apart, 6.1%; BCL2 break apart, 2.2%; BCL6 break apart, 2.2%; IGH/BCL2 fusion, 4.1% (see Appendix 5.2. for calculation).

na; Not assessed due to insufficient biopsy material.

4.3. Discussion

In this study we have investigated the characteristics of the V genes of *de novo* DLBCL. Overall, we reveal a random usage of V_H and V_L gene segments, which appears to reflect that observed in the normal B-cell repertoire, with only minor asymmetries. Furthermore, our data support a GC-related B cell origin of DLBCL, with 94.4% of tumours displaying mutated Ig V genes and evidence for ongoing somatic mutation in 88.9% of cases. In keeping with previous reports, the present study indicates a heterogeneous pathobiology of *de novo* DLBCL.[367,371,391-394]

Consequently, where possible, we have used tissue microarray and cDNA microarray to subdivide our *de novo* DLBCL cases into those with a GCB-like phenotype and those with an ABC-like phenotype. We show that using tissue microarray and the classification model proposed by Hans *et al.* (Figure 4.1a.) we could divide our tumours into 8 GCB-like DLBCL and 6 non-GCB(/ABC)-like DLBCL; paraffin-embedded biopsy material was not available for 4 Southampton cases, which remained unclassified.[407] The majority of GCB-like cases co-expressed CD10 and Bcl-6. We next used the lymphochip cDNA microarray to classify the tumours, where we found 14 GCB-like DLBCL and 4 ABC-like DLBCL; RNA was not available for 1 Southampton case, which remained unclassified. Considering the cDNA microarray classification as the gold standard, the sensitivity of the tissue microarray was 63.6% for the GCB subgroup and 100% for the ABC subgroup; four tumours were misclassified as non-GCB-like DLBCL by tissue microarray. For the analysis that follows, the classifications obtained by cDNA microarray were used; for the single tumour (Ly136/01) where cDNA microarray data was not available we accepted the tissue microarray classification as GCB-like DLBCL. Of note, this GCB-like tumour is positive for the t(14;18) translocation and, therefore, we can be more confident of its classification since t(14;18) is universally detected in GCB-like tumours (see below).[405,534,535]

Overall, the t(14;18) translocation was detected in 31.6% (6/19) of tumours (Table 4.8.). All 6 t(14;18)-positive cases had a GCB gene expression profile, with t(14;18)-positive tumours representing 33.3% (5/15) of cases in this subgroup. This finding is in keeping with a reported association between tumours of a GCB-like phenotype and the t(14;18) translocation.[405,534,535,539] A study by Iqbal *et al.* has found that these tumours show decreased expression of proliferation signature genes when compared with their t(14;18)-negative counterparts, suggesting that t(14;18) defines a unique subset of DLBCL within the GCB-like subgroup.[535] Five of 5 t(14;18)-positive, GCB-like tumours expressed Bcl-2

Table 4.8. A summary of cytogenetic and molecular features of GCB-like and ABC-like *de novo* DLBCL

Case	Phenotype		Translocation	Glycosylation site (number)	AID expression	Intraclonal heterogeneity
	cDNA array	Tissue array*				
Ly117/01	GCB	GCB	t(14;18)	No	Yes	Yes
Ly135/01	GCB	GCB	t(14;18)	Yes (1)	No	Yes°
Ly136/01	na	GCB	t(14;18)	No	No	Yes
Ly284/02	GCB	GCB	t(14;18)	Yes (1)	Yes	Yes
F5126	GCB	na	t(14;18)	Yes (3)	Yes	Yes
F5481	GCB	GCB	t(14;18)	No	Yes	Yes°
Ly34/01	GCB	na	na	Yes (1)	Yes	Yes
Ly235/01	GCB	GCB	No	Yes (2)	Yes	Yes
F5049	GCB	na	No	No	No	No
F5246	GCB	GCB	No	Yes (2)	No	Yes
F5334	GCB	GCB	No	No	No	Yes
F5184	GCB	non-GCB	t(3;14)	No	No	Yes
Ly157/01	GCB	non-GCB	No	No	Yes	Yes
Ly268/01	GCB	non-GCB	No	Yes (4)	Yes	Yes
F5370	GCB	non-GCB	No	Yes (1ᵉ)	No	Yes
Ly61/01	ABC	non-GCB	No	Yes (4)	No	Yes
F4881	ABC	non-GCB	No	No	No	Yes
F5434	ABC	na	No	No	Yes	Yes
F5440	ABC	non-GCB	No	Yes (1)	Yes	Yes

*; Phenotype determined using the decision tree devised by Hans *et al.*, 2004.[407]
na; Not assessed due to insufficient biopsy material.
ᵉ; Single disfavoured *N*-glycosylation motif
GCB; Germinal centre B-cell-like.
ABC; Activated B-cell (ABC)-like.
Non-GCB-like; ABC-like or 'Type 3' phenotype.
°; Intraclonal heterogeneity exceeds Taq error threshold but does not include repeated mutations.

protein, consistent with the deregulated expression Bcl-2 protein in these tumours (Table 4.5.); expression could not be assessed in the remaining tumour due to a lack of biopsy material. It has been reported previously that a high percentage of t(14;18)-positive *de novo* DLBCL have a low or absent expression of Bcl-2 protein, our data do not support these findings.[526,534] The frequencies of Bcl-2 protein expression was not significantly different between t(14;18)-positive GCB-like and t(14;18)-negative GCB-like tumours, where we observed Bcl-2 expression in 100.0% (5/5) and 85.7% (6/7) of cases, respectively (P = 0.583). Overall, there was no difference between the expression of Bcl-2 in GCB-like (91.7%; 11/12) and ABC-like (66.7%; 2/3) tumours (P = 0.371). This finding indicates that Bcl-2 expression is often present in *de novo* DLBCL in the absence of t(14;18). Of note, amplification of the *Bcl-2* gene on 18q21 was not detected by iFISH in any ABC-like or GCB-like tumours in this study.[532] The t(3;14) translocation was observed in a single (5.3%) DLBCL tumour of the GCB-like subset, in keeping with the seemingly low frequency of this translocation at diagnosis.[530] Although constitutive Bcl-6 expression in disease ontogeny would suggest a GC association, the t(3;14) translocation has been found more frequently in ABC-like DLBCL.[539]

We observed *AID* mRNA expression in 41.7% of cases; this finding is similar to that previously reported for *de novo* DLBCL.[540] AID is specifically expressed in normal GC B cells and is required for SHM and class switch. Therefore, we would expect AID to be highly expressed in GCB-like but not ABC-like DLBCL and that its expression should correlate with the presence of intraclonal heterogeneity in these tumours. Our data show that 53.3% (8/15) of GCB-like DLBCL expressed *AID* and that this correlated with the presence of intraclonal heterogeneity in these tumours (Table 4.8.). Although the remaining 46.7% (7/15) of GCB-like tumours did not express *AID* mRNA, the majority (n=6) did exhibit intraclonal heterogeneity. Since our analysis was limited to mRNA, it is possible that the lack of expression seen in these tumours is related to a poor correlation between *AID* mRNA and AID protein expression. However, it has been shown previously that the detection of *AID* mRNA transcripts and AID protein is well correlated, indicating that RNA-based studies yield valuable information about AID expression.[220] It is important to distinguish between the detection of intraclonal heterogeneity and active SHM. AID expression may not remain constant over time within malignant cells and a transient increase in AID expression in response to extracellular stimuli might lead to the mutations found later at the time of analysis, while AID expression might have decreased to its base line level in the mean time. Conversely, *AID* expression was observed in 50.0% (2/4) of ABC-like DLBCL in this study, in both cases this correlated with the presence of

intraclonal heterogeneity; the remaining two tumours displayed intraclonal heterogeneity in the absence of *AID* (Table 4.8.). Expression of *AID* mRNA in both GCB-like and ABC-like DLBCL has been reported previously and was found to be significantly higher in the latter subset.[220,541] Here, the authors suggested that the high *AID* mRNA levels might be due to the activation of the NF-κB pathway, which has been shown to be associated with this subset. These observations suggest that in contrast to normal B cells in which AID is highly expressed only in GC B cells, *AID* is also expressed in ABC-like tumours in which the expression of other GC signature genes is down-regulated. Although the level of intraclonal heterogeneity in ABC-like tumours was not specified in this study by Lossos *et al.*, a previous report by the same group has shown that tumours with a GC gene expression profile exhibit intraclonal heterogeneity indicative of active somatic mutation, while the majority (5/7) of ABC-like tumours do not; in the 2 cases exhibiting ongoing mutation, intraclonal heterogeneity was low: 1bp mutation featuring in 2 clones [404] This observation does not fit with our data where we see 4/4 tumours displaying intermediate levels of intraclonal heterogeneity (mean 0.509%). Thus, our data support the concept that ongoing somatic mutation is a property of malignant cells irrespective of their developmental origin. Overall, *AID* expression and intraclonal heterogeneity correlated in 52.8% (19/36) of cases, with the remaining tumours (47.2%; 17/36) showing intraclonal heterogeneity in the absence of *AID* expression (Table 4.4.). A lack of correlation between *AID* expression and intraclonal heterogeneity has been reported previously for *de novo* DLBCL and FL; although a more recent report suggests a that there is a strong association in FL.[220,540-542]

It has been suggested that the accumulation of *N*-glycosylation sites in the Ig V genes of some tumours is an adaptive response to facilitate survival in the GC microenvironment. Novel *N*-glycosylation sites that have been introduced by somatic mutation are universally present in FL (Chapter 3) and feature in a high proportion of V_H genes of eBL.[12,13,460] In contrast, the V genes of normal plasma and memory B cells and those of MM, mutated CLL and MALT lymphomas have a low frequency (<10%) of sites. Interestingly, the V_H genes of *de novo* DLBCL have been reported to have an intermediate frequency of sites (~40%), however, it is not known which subset they may be associating with, if any.[12] We have assessed the incidence of *N*-glycosylation sites in *de novo* DLBCL and found that 42.9% of tumours had at least one novel site, in keeping with published reports.[12] The t(14;18) translocation is a characteristic feature of FL and is considered to be the initiating event in lymphomagenesis. The fact that *de novo* DLBCLs with the t(14;18) translocation have gene expression profiles similar to normal GC B cells would suggest that the t(14;18)

plays an important role in the pathogenesis of this subset of DLBCL, just as it does in FL. Moreover, we might expect these tumours to have universally acquired *N*-glycosylation sites in their V genes, as is seen in FL. However, our data show that just 50.0% (3/6) of t(14;18)-positive, GCB-like DLBCL cases acquire a site. Overall, *N*-glycosylation sites are observed in 53.3% (8/15) and 50.0% (2/4) of GCB-like and ABC-like DLBCL, respectively. Therefore, our data do not support a common cell of origin linking *de novo* t(14;18)-positive DLBCL and t(14;18)-positive FL. However, in the course of performing this study, Jardin *et al.* have also addressed the question of *N*-glycosylation in specific subgroups of *de novo* DLBCL.[539] Here, the authors report a high prevalence of *N*-glycosylation motifs in t(14;18)-positive GCB-like DLBCL (7/8), which is comparable to the acquisition of sites in FL. Furthermore, the frequency of sites in non-GCB-like DLBCL appears to be correspondingly low (3/8), although slightly higher than that seen for normal B cells. Therefore, these data reinforce a link between t(14;18)-positive GCB-like DLBCL and FL and suggest a common pathway of pathogenesis. The reason for the disparity between data sets is unclear. In the study by Jardin *et al.*, tissue microarray and the classification model proposed by Hans *et al.* was used to divided DLBCL into the two subgroups. Therefore, it may be possible that some tumours were misclassified. However, if we reassess our data using the subgroups established following tissue microarray, the discrepancy remains; 50.0% (4/8) and 66.7% (4/6) of GCB-like and non-GCB-like DLBCL, respectively, acquire *N*-glycosylation sites. It is possible that a low number of tumours in each subgroup, particularly in the case of the ABC-like DLBCLs (n=4), has led to skewing of the data. However, Jardin *at al.* assessed a similarly low number of tumours; GCB-like DLBCL (n=10), non-GCB-like DLBCL (n=11).

4.4. Conclusions

Overall, these findings do not support a link between t(14;18)-positive, GCB-like *de novo* DLBCL and t(14;18)-positive FL and argue against a common pathway for pathogenesis. Although we confirm an association of the t(14;18) translocation and DLBCL of the GCB-like subgroup, our data show that just half of these tumours acquire *N*-glycosylation sites. The frequency with which *N*-glycosylation sites are acquired in t(14;18)-positive, GCB-like DLBCL, although higher than that seen for DLBCL as a whole, is not comparable to that of FL, in which *N*-glycosylation sites are almost universally present. This suggests that for t(14;18)-positive, GCB-like *de novo* DLBCL *N*-glycosylation is not required for tumour cell survival in the GC microenvironment, in contrast to FL. Furthermore, *N*-glycosylation sites feature equally frequently in the ABC-like subgroup, although the numbers of cases in this group was very low and may have lead to skewing of the data.

We also find that the expression of *AID* mRNA and intraclonal heterogeneity feature in both the GCB-like and ABC-like subgroups, suggesting that ongoing somatic mutation is a property of malignant cells irrespective of their developmental origin.

5. Primary central nervous system lymphoma

5.1. Introduction

PCNSLs represent highly malignant NHLs, which arise within and are confined to the CNS in the absence of systemic disease. However, occult systemic disease has been reported in ~4% of patients in one study.[543] The vast majority (98%) are of the B-cell type and many (>75%) are categorised as DLBCL based on the WHO classification system.[286] However, PCNSLs differ in their biological behaviour from extracerebral DLBCLs, which has lead to the question as to whether PCNSLs comprise a distinct disease entity. PCNSL is a rare neoplasm and represents just 2-3% of NHLs and ~3% of all malignant primary brain tumours, although the incidence has been steadily increasing over the last two decades.[544] PCNSL may affect all age groups, but sees a peak incidence at 50-70 years, with a male to female ratio of 1.5.[545] While the majority of PCNSLs are sporadic, some are associated with congenital or acquired immunodeficiency. An increase in incidence may be partly attributed to the growing population of immunocompromised patients, in particular as a result of the human immunodeficiency virus (HIV) epidemic; these tumours are referred to as AIDS-related PCNSLs and are associated with EBV infection.[546] The increased incidence of sporadic, or AIDS-unrelated, PCNSL in immunocompetent individuals remains unexplained to date. While the pathogeneisis of PCNSL is still poorly defined, it has become evident that PCNSLs arising in patients with immunodeficiencies must be differentiated from those affecting immunocompetent individuals since the two forms display very different pathogenetic and clinical aspects. From herein I will focus on AIDS-unrelated PCNSL.

At diagnosis, most immunocompetent patients present with a solitary lymphomatous lesion affecting the supratentorial brain parenchyma and with a predilection for the periventricular spaces.[545] The leptomeninges are involved in up to 40% of patients and ocular involvement is found in 15-20% of cases and may consist of infiltration of the retina, the choroid or the vitreal fluid.[545] In the majority of cases, contrast-enhanced CT scan of the chest, abdomen and pelvis of patients and a biopsy of the BM do not reveal evidence of disease outside the CNS at diagnosis. In the later stages, PCNSL frequently become diffuse and multifocal; however, systemic dissemination occurs only rarely.[547] The prognosis of PCNSL is poor and the median survival is only 3 to 5 months in untreated patients.[548] Treatment with whole brain radiotherapy (WBRT) alone can prolong mean survival to 12 to 18 months. However, in patients receiving methotrexate-based

chemotherapy, mean survival has increased to up to 5 years and high-dose methotrexate is now considered an essential component of any treatment regimen for PCNSL.[545] Relapse of PCNSL almost invariably occurs within the CNS.

5.1.1. Morphology

The majority of PCNSLs are B-cell lymphomas, defined by surface expression of the pan B-cell markers CD19, CD20 and CD79a.[545] Microscopically, tumours show a characteristic angiocentric growth pattern, typically forming cuffs of tumour cells within and around cerebral blood vessels ('perivascular cuffs'). Tumours are highly cellular and composed of lymphoma cells with a centroblastic or centrocytic morphology. Tumour cells are characteristically embedded within an infiltrate composed of reactive T cells ($CD3^+$), macrophages and small, mature, polyclonal B cells. In addition, the surrounding oedematous brain tissue contains resident cell populations like microglia and astrocytes, which are markedly activated. Tumours express monoclonal sIg or cytoplasmic Ig (cIg), most commonly $IgM^+\kappa^+$.[549]

5.1.2. Molecular analysis

Small studies on the V_H genes in PCNSLs have demonstrated that they are clonally rearranged and have acquired somatic mutations with an unusually high frequency, which exceeds other lymphoma entities, including extracerebral DLBCLs.[400-402,550] This suggests that PCNSLs are related to GC or post-GC B cells. This is further supported by the observation of somatic mutations in the 5' non-coding regions of the *Bcl-6* gene and the high expression of the Bcl-6 protein by these tumours, which is regarded as a marker for B cell transition through the GC.[551] Moreover, low level intraclonal heterogeneity has also been observed and indicates that the tumour cell of origin has been exposed to the mutational machinery both before and after neoplastic transformation, characteristic of tumours located within the GC, such as FL and a proportion of extracerebral DLBCL.[400,401] Aberrant ongoing somatic mutation can also be detected in some proto-oncogenes, including *c-Myc* and *Pax5*.[289] Indeed, a prolonged interaction of the tumour cell or its precursor in the GC microenvironment may be responsible for the high mutation frequency of the V_H genes. A recent study of 10 tumours has indicated that PCNSLs are unable to isotype switch, as evidenced by the exclusive transcription of IgM and IgD mRNA.[461] However, IgG-expressing tumours have been described, albeit at a low frequency.[549] Perhaps unexpectedly, the expression of *AID* mRNA, which is shown to be associated with GC centroblasts and lymphomas derived from GC B cells, has been reported to be uncommon in PCNSL.[461] More recently, the immunohistochemical classification of 83

PCNSLs into GCB-like and ABC-like DLBCL based on their expression of CD10, Bcl-6 and MUM-1 (see Figure 4.1a.) has revealed a uniform non-GCB-like/ABC-like phenotype.[407,552] This finding, which seems paradoxical in comparison to earlier studies of tumour-derived V_H genes, led authors to suggest that PCNSLs occupy a histogenetic time-slot overlapping late GC and early post-GC. This hypothesis is in agreement with an earlier study looking at the expression of the tyrosine phosphatase SHP-1, a post-GC marker; strong SHP-1 expression was observed in PCNSL, either with or without Bcl-6 and/or CD10 expression and in the absence of CD138 expression.[553]

Another revealing finding from V_H gene studies is that of the preferential usage of the V_H gene segment V4-34, which was observed rearranged in up to 50% of PCNSL.[400,401,461,550] The high frequency of usage of the V4-34 gene suggests a functional role of the Ig encoded by this gene in the development of PCNSL. This has lead to speculation that sustained antigenic stimulation by either an Ag or SAg in the CNS may drive the proliferation of the malignant B cells or their precursors. Indeed, infectious agents can promote lymphomagenesis either through direct transforming properties or sustained antigenic stimulation. In this regard, proteins of neurotropic viruses with the capacity to persist in the CNS, in particular viruses of the herpes group and polyomaviruses, represent attractive candidates. Infection with various viruses, including EBV and human herpes virus 8 (HHV-8), is assumed to play a role in the pathogenesis of some lymphomas. Indeed, EBV infection is commonly observed in patients with AIDS-related PCNSL, indicating a pathogenic role for the virus in these patients.[546,554] In contrast, EVB infection is rarely involved in AIDS-unrelated PCNSL and suggests a different pathogenic mechanism is operating in these cases.[555] Similarly, roles for HHV-8, HHV-6, CMV and the polyomaviruses SV40, JC and BK in PCNSL have been ruled out and all attempts to identify a causative pathogen have failed thus far.[556-562]

5.1.3. PCNSL and the CNS

The CNS has traditionally been considered an immunologically privileged site, construed as CNS isolation from the immune system by the blood brain barrier (BBB), with a lack of draining lymphatics and an apparent immunoincompetence of microglia, the resident CNS macrophage.[reviewed in 563] Modern neuroimmunology has modified this concept and it is now clear that immune privilege is limited to the CNS parenchyma proper, with the immune reactivity of the ventricles, containing cerebrospinal fluid (CSF), meninges and circumventricular organs similar to that of the periphery. Microglia, are immunocompetant but differ from other macrophages/DCs in their ability to direct

neuroprotective lymphocyte responses.[563] Moreover, the drainage of soluble Ag from the brain parenchyma to the cervical lymph nodes can occur along the perivascular spaces of capillaries and arteries. Furthermore, the 'privilege' described does not relate to the absolute absence of immunological components but rather to their elaborate regulation. It is accepted that monocytes and activated T cells (both $CD4^+$ and $CD8^+$) and B cells may cross the BBB and enter the CNS to perform routine surveillance functions and that this represents normal trafficking.[564-566] A study of normal and AIDS brains showed that low numbers of activated B cells are able to enter all parts of the brain, including the brain parenchyma, suggesting that there is no impediment to the movement of B cells from the perivascular to the parenchymal compartments.[566] However, it is not known if these B cells can then leave the CNS via the circulation or whether they are destined to die there. A more recent study has addressed the issue of immune cell migration out of the CNS and found that DCs injected within the brain parenchyma migrate little from their site of injection and do not reach cervical lymph nodes.[567] Conversely, intra-CFS-injected DCs are able to exit the CNS and reach the cervical lymph nodes, as well as being able to infiltrate the brain parenchyma.

The finding that highly malignant lymphomas can develop exclusively in the brain is conceivable. However, the site of origin of the malignant lymphocyte that results in PCNSL is unknown. One theory is that transformation and subsequent clonal proliferation of an activated B cell trafficking normally through the CNS gives rise to the malignant clone. A circulating memory B cell could be stimulated by Ag in the brain, with subsequent transformation. Molecular evidence, including a high frequency of somatic mutations and evidence of ongoing SHM, suggests that the tumour cell origin may be related to a GC B cell. One possibility is that SHM in the CNS is occurring independently of a GC structure. Support for GC-independent SHM of V genes is provided in mice deficient in lymphotoxin-α and Lyn kinase, although the mutational levels observed are low and would not be expected to reach the very high levels found in PCNSL.[568,569] Alternatively, it is possible that the tumour cells, which are typically embedded in a reactive infiltrate composed of activated T cells and resident astrocytes, microglia, endothelial cells and pericytes can 'mimic' the GC microenvironment and somatically hypermutate intracerebrally. Ectopic structures resembling classical lymphoid follicles have been described in non-lymphoid tissues during the course of both autoimmune and chronic infectious diseases (see section 1.3.3.2.2.1.). Importantly, the detection of ectopic B-cell follicles has been described in the brains of patients with secondary progressive multiple sclerosis (MS).[570] Here, it was shown that follicles containing activated B cells,

activated T cells and FDCs expressing B-cell-attracting chemokine 1 (BCA-1, also known as CXCL13) can form in close association with inflamed blood vessels in the cerebral leptomeninges. The presence of proliferating B cells in these structures, they propose, was strongly suggestive of GC formation. Furthermore, a second study has provided evidence for the recapitulation of all stages of B-cell differentiation observed in the secondary lymphoid tissues in the CNS by the identification of centroblasts and centrocytes, memory B cells and plasma cells in the CSF of MS patients.[571] Therefore, the meninges, which represent a site of intense immune traffic in normal brains, can provide a favourable microenvironment for ectopic lymphoid follicle formation. Of note, ~40% of PCNSLs have tumour infiltration of the leptomeninges.[545]

An alternative theory is that the tumour B cells of PCNSL undergo SHM and neoplastic transformation outside the CNS within the GC of the secondary lymphoid organs and then selectively 'home' to the CNS. This would suggest that tumour cells display unique surface markers to induce migration to the CNS. Cellular adhesion molecules have been implicated in this homing process by the demonstration that lymphocyte function-associated antigen (LFA) 1a, LFA1b and intercellular adhesion molecule 1 (ICAM1) mediate interaction between lymphoma cells and tumour endothelial cells.[572] The adhesion of clonal cells to the CNS has also been shown to involve the family of N-cadherins and the adhesion molecule CD44.[573] A recent study using cDNA microarray analysis has highlighted the importance of genes implicated in CNS tropism, with the finding that PCNSLs express a unique set of extracellular matix and adhesion-related genes when contrasted with extracerebral DLBCL.[574] Furthermore, chemokines play a pivotal role in the regulation of homing and spread of lymphocytes in the CNS. The tumour cells of PCNSL have been shown to consistently express CXCR4 and its ligand CXCL12, as well as CXCR5 and its ligand BCA-1/CXCL13.[575,576] Moreover, the vascular endothelial cells, reactive astrocytes, microglia and activated T cells in PCNSLs also show distinctive expression profiles of these chemokines and their receptors, suggesting chemokine-mediated regulation of B-cell homing to the CNS and their dissemination within the brain.[576] While the B cell clone may have continued to acquire somatic mutations post-transformation and before leaving the GC, the intraclonal heterogeneity observed in these tumours would suggest that several different tumour clones simultaneously made the journey to the CNS. However, since intraclonal heterogeneity is low, it is possible that GC-independent SHM within the brain plays a role or that the transforming event renders the tumour cells unable to inactivate their hypermutational machinery.

5.1.4. Aims

The aim of this study was to confirm the over-representation of the V4-34 gene segment and to further characterise mutational patterns in tumour-derived V genes, incorporating the results of previously published studies into our analysis to obtain a clearer understanding. Furthermore, we aimed to address the question of tumour cell origin, which may have important implications for pathogenesis. Firstly, we wished to correlate V gene information with other GC markers, such as Bcl-6 and CD10. Secondly, we wished to investigate the existence of tumour cells outside of the CNS in the periphery using the unique CDR3 sequence identified from biopsy material to search for the tumour cell clone in the peripheral blood and BM of patients at the time of diagnosis.

5.2. Results

5.2.1. Tumour-derived V gene usage

Using cDNA isolated from biopsy material as a source, we identified the tumour-derived V_H and V_L gene sequences in 12/12 and 11/12 cases, respectively (Table 5.1.). In one case, the V_H gene could only be identified using a FR2 to constant region PCR and, therefore, the first 44 codons from the 5' end, which incorporates FR1 and CDR1, was absent. All V gene rearrangements were in frame and did not harbour stop codons and were, therefore, considered potentially functional.

5.2.1.1. V_H gene usage

Twelve potentially functional, clonal V_H sequences were identified from 12 cases (Table 5.1.). Eight (15.7%) of the 51 known functional V_H gene segments were found rearranged in our cohort. Fifty percent (6/12) of tumours used gene segments from the V_H4 family, a further 41.7% (5/12) used genes from the V_H3 family and the remaining tumour used the V_H2 family segment V2-26. Gene segments from the V_H1, V_H5, V_H6 and V_H7 families were not observed. The V4-34 gene constituted 4 of the 6 V_H4 rearrangements and was used in 33.3% of tumours, in keeping with earlier reports that this gene segment is preferentially rearranged in PCNSL.[400-402,461] The CDR3 region varied in length from 27 to 63 nucleotides (mean 40.3bp, median 39bp), according to the Kabat numbering system.[496] D segments could be confidently assigned in 25.0% (3/12) of tumours, according to the rule proposed by Corbett *et al.*; D segments D3-22 (n=2) and D2-02 (n=1) were found rearranged.[20] All 6 J_H gene segments were used with J_H4 genes most common (n=6), followed by J_H5 (n=2) and J_H1, J_H2, J_H3 and J_H6 (all n=1).

Combining our series with the 49 cases from 5 published studies reveals that overall the V4-34 gene is used by 44.3% (27/61) of tumours.[400-402,461,550] The majority (93.4%; 57/61) of cases used genes from the V_H3 or V_H4 families, with the V_H2 and V_H5 families each represented once and the V_H1 family twice. Frequently used V_H gene segments other than V4-34 included V3-07 (n=6), which is not seen over-represented in normal B cells, and V3-23 and V4-39 (both n=4). An aa alignment of 19 of 27 V4-34-encoding tumours where the sequence was available is shown Figure 5.1. No conserved sequences were identified within the CDR3, which were found to display dissimilar aa composition- at the nucleotide level, no bias was observed in D or J_H gene usage. On average, we observed 84.3% (range 66.7-96.7%) and 77.0% (range 56.3-100.0%) conservation of the germline V4-34 sequence in FR1 and FR3, respectively; this translates to an average of 4.7 aa changes in FR1 and 7.4 aa changes in FR3. Assessment of V4-34 rearrangements isolated

Table 5.1. V_H and V_L gene usage in PCNSL

	V_H					V_L			
Case	**V_H Family**	**V gene***	**D gene***	**J_H gene***	**CDR3 length (bp)**	**V_L Family**	**V gene***	**J_L gene***	**CDR3 length (bp)**
1053-98	V_H4	V4-34#	-	J_H4b	36	Vκ3	DPK22/A27	Jκ1	27
1126-98	V_H4	V4-34	-	J_H2	48	Vκ3	DPK22/A27	Jκ2	24
1408-98	V_H4	V4-34	-	J_H5b	45				
0875-99	V_H3	V3-23	-	J_H4b	39	Vκ2	DPK18/A17	Jκ1	27
1002-99	V_H4	V4-34	D2-02	J_H4b	48	Vλ3	DPL16/3l	Jλ2/3a	39
0637-00	V_H3	V3-07	-	J_H6b	63	Vλ3	IGGLL15/3m	Jλ2/3a	33
0647-00	V_H3	V3-30.5	D3-22	J_H4b	51	Vκ3	DPK21/L2	Jκ4	33
0712-00	V_H3	V3-72	-	J_H4b	27	Vκ1	LFK431/L1	Jκ5	27
0836-05	V_H4	V4-39	-	J_H5b	30	Vκ4	DPK24/B3	Jκ1	27
AJAC	V_H3	V3-30.5	D3-22	J_H1	36	Vκ3	DPK21/L2	Jκ2	33
TREA	V_H2	V2-26	-	J_H3a	30	Vκ3	DPK22/A27	Jκ1	27
SLAW	V_H4	V4-61	-	J_H4b	30	Vκ2	DPK18/A17	Jκ2	33

*; Germline V, D and J genes with the greatest homology.
#; Only partial sequence identified (FR2-FR4).
-; Germline D gene could not be confidently assigned.

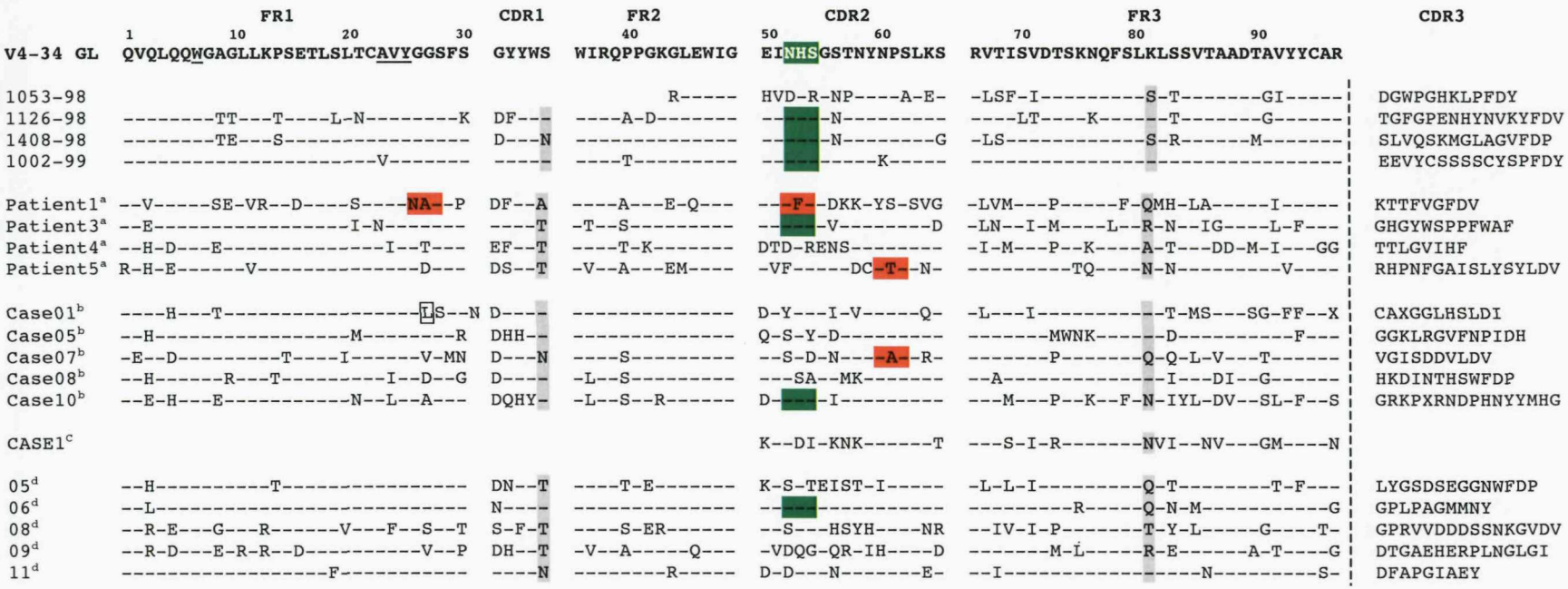


Figure 5.1. An amino acid alignment of all published V4-34 V_H gene segments in PCNSL. The aa sequences of 19 V4-34-utilising PCNSLs from the current study and 4 published studies are shown: [a]Thompsett *et al.*, 1999, [b]Montesinos-Rongen *et al.*, 1999, [c]Sekita *et al.*, 2001, [d]Montesinos-Rongen *et al.*, 2005.[400-402,461] Novel *N*-glycosylation sites are highlighted in *red*. The natural *N*-glycosylation site from the germline V4-34 sequence is highlighted in *green*. The W/AVY aa motif recognised by the monoclonal 9G4 Ab is underlined in the germline sequence. Dashed lines represent aa that are in the germline configuration. *Boxed* aa are insertions. Shading indicates a recognised mutational hotspot.

from normal B cells shows that a number of codons are particularly prone to mutational change ('hot spots') and often result in the same aa substitution.[577] For example, Ser35 and Lys81 are both commonly mutated to Thr, Asn or Arg, with the latter also mutating frequently to Gln. We observe similar aa substitutions at these positions in PCNSL-derived V4-34 sequences (Figure 5.1.). Other codons displaying frequent aa substitutions, but not highlighted in normal B cells, include Gln3, Gly31, Pro40 Ser56 and Ser84; of note, 11 of 14 substitutions at Gly31 are to Asp.

5.2.1.2. V_L gene usage

Eleven potentially functional, clonal V_L sequences were identified from 12 cases (Table 5.1.). Seventy-five percent (9/12) of PCNSL rearranged the κ chain. Gene segments from the Vκ3 family were used most frequently (56%; 5/9) and included DPK22/A27 (n=3), which constituted 33.3% of kappa rearrangements, and DPK21/L2 (n=2). The Vκ2 family member DPK18/A17 was used in 22.2% (2/9) of tumours and the remaining 2 tumours used gene segments from the Vκ1 and Vκ4 families. Gene segments from the Vκ5 and Vκ6 families were not observed. Four of the 5 Jκ genes were used in the following distribution: Jκ1, 33.3% (n=4); Jκ2, 25.0% (n=3); Jκ4 and Jκ5, both 8.3% (n=1). Of the 2 identified λ rearrangements, both used genes from the Vλ3 family (DPL16/3l and IGGLL15/3m) in combination with the Jλ2/3a gene segment.

Analysis of all 11 functional V_L sequences revealed that the CDR3 region varied in length from 24 to 39 nucleotides (mean 30bp, median 27bp). Independent assessment of κ and λ rearrangements showed that the CDR3 length of the rearranged Vλ genes was ~2 amino acids longer (mean 12aa) compared to Vκ rearrangements (mean 9.6aa) consistent with studies of κ and λ rearrangements in normal B cells and other B-cell malignancies (see section 3.2.1.2. and 4.2.1.2.).[48,75,76]

5.2.2 Somatic hypermutation in tumour-derived V genes

Isotype designation and mutational status are shown in Table 5.2. Our data reveal that although IgM (with IgD) mRNA transcripts are most common, IgG transcripts were expressed in 25.0% (3/12) of tumours, either exclusively (n=1) or in addition to IgM/IgD transcripts (n=2). All V genes showed evidence of somatic mutation. With the exception of one tumour, the V_H gene sequences exhibited intermediate to high levels of mutation (range: 1.0-23.0%, mean: 13.2%). Ninety-two percent of V_H sequences had a mutation level of greater than 5%, 75.0% of greater than 10% and 25.0% of greater than 20%. Comparatively high levels of mutation were also seen in V_L (range: 1.9-19.7%,

Table 5.2. An analysis of somatic mutations in the V genes of PCNSL

Case	V_H Isotype*	V_H Mutation (%)≠	V_H Intraclonal heterogeneity No. clones°	V_H Intraclonal heterogeneity %	V_L Isotype*	V_L Mutation (%)≠	V_L Intraclonal heterogeneity No. clones°	V_L Intraclonal heterogeneity %
1053-98	μ, δ	17.6#	3/11	0.140#	κ	9.4	4/11	0.124R
1126-98	μ, δ	10.3	3/12	0.074	κ	5.6	2/12	0.066
1408-98	μ, δ, γ	10.3	9/11	0.406R	λ§			
0875-99	μ, δ	20.1	1/12	0.025	κ	12.9	3/11	0.094
1002-99	μ, δ	1.0	6/12	0.197R	λ	1.9	2/12	0.056R
0637-00	μ, δ	20.7	10/11	0.484R	λ	13.8	9/12	0.425R
0647-00	μ, δ	12.6	3/12	0.073	κ	8.3	0/11	0.000
0712-00	μ, δ	23.0	7/12	0.229R	κ	19.7	3/10	0.103
0836-05	μ, δ, γ	16.5	1/11	0.028	κ	5.0	0/12	0.000
AJAC	γ	5.8	5/15	0.170R	κ	4.2	0/12	0.000
TREA	μ, δ	10.7	3/7	0.130R	κ	3.4	2/11	0.062
SLAW	μ, δ	9.8	1/11	0.028	κ	4.3	2/12	0.080

*; Heavy and light chain isotype determined by PCR.
≠; Frequency of somatic hypermutation compared to the most homologous germline V gene segment.
°; Number of clones exhibiting sequence variation from the tumour consensus sequence / total number clones sequenced.
#; Somatic mutation and intraclonal heterogeneity expressed for partial sequence (FR2-FR4).
R; Includes at least one intraclonal sequence mutation that is repeated in >1 clone.
§; Light chain isotype identified by immunohistochemistry.

mean: 8.0%), with 63.6% of sequences having a mutation level of greater than 5% and 27.3% of greater than 10%. Mutation rates were lower in the V_L gene compared to their partner V_H in keeping with findings in normal B cells and other B-cell malignancies (see section 3.2.2. and 4.2.2.).[368,492].

The measure of intraclonal heterogeneity for each tumour-derived V gene was assessed in, on average, 12 independent clones (range 7-15) (Table 5.2.). Seventy-five percent (9/12) of V_H genes and 72.7% (8/11) of V_L genes had levels of intraclonal heterogeneity that exceeded that of a previously determined taq error rate of 0.05%. Intraclonal heterogeneity was observed in the V genes of tumour 1002-99 that had unmutated V genes (1.0% and 1.9% in V_H and V_L, respectively), thus verifying mutational activity. Overall, using this cut off, 91.7% (11/12) of tumours showed evidence for intraclonal heterogeneity indicative of ongoing SHM. The level of intraclonal heterogeneity was generally low, mean: V_H 0.211% (range 0.073-0.484%); V_L 0.126% (range 0.056-0.425%). If a more stringent definition of intraclonal heterogeneity were applied- the presence of sequence mutation(s) in more than one molecular clone- intraclonal heterogeneity was reduced to 50.0% in the V_H genes and 25.0% in the V_L genes, with 58.3% (7/12) of tumours displaying intraclonal heterogeneity overall. Therefore, intraclonal heterogeneity may be considered as present in 7/12 cases, absent in 1/12 cases and borderline in 4/12 cases.

5.2.3. Incidence of *N*-glycosylation sites in tumour-derived V genes

V_H and V_L gene sequences were examined for potential *N*-glycosylation sites with the motif Asn-X-Ser/Thr (Table 5.3.). Four tumours used the V4-34 gene, which contains a naturally occurring *N*-glycosylation site in its germline sequence. Of these, 3 tumours retained the natural site and 1 tumour lost the site; no additional sites were acquired by SHM in any of the tumours. Of the remaining tumours, three (25.0%) gained 1 novel *N*-glycosylation site in their V_H gene (2 in CDR2 and 1 in FR1). No sites were acquired in the V_L genes. Combining our series with 30 published cases for which sequence information was available reveals that 31.6% (6/19) of V4-34 tumours retained the germline *N*-glycosylation site (Figure 5.1.) and 21.4% (9/42) of tumours acquired 1 novel site.[400-402,461]

5.2.4. Expression of *AID* mRNA in PCNSL

AID mRNA expression was assessed in 12 PCNSL by standard PCR (Figure 5.2.). A single amplification product of the expected size of ~646bp was detected in 91.6% (11/12) of tumours. PCR products of the correct size were also detected in the positive control,

Table 5.3. The incidence of *N*-glycosylation sites in the V genes of PCNSL

	V_H			V_L		
Case	**No. Sites**	**Region**	**Motif**	**No. Sites**	**Region**	**Motif**
1053-98	0[*≠]	-	-	0	-	-
1126-98	1	CDR2[GL]	NHS	0	-	-
1408-98	1	CDR2[GL]	NHS			
0875-99	0	-	-	0	-	-
1002-99	1	CDR2[GL]	NHS	0	-	-
0637-00	1	FR1	NSS	0	-	-
0647-00	0	-	-	0	-	-
0712-00	1	CDR2	NFS	0	-	-
0836-05	0	-	-	0	-	-
AJAC	0	-	-	0	-	-
TREA	1	CDR2	NTS	0	-	-
SLAW	0	-	-	0	-	-

[*]; Loss of germline V4-34 *N*-glycosylation site by somatic mutation.
[GL]; Identifies a germline V4-34 *N*-glycosylation site.
[≠]; Incidence of *N*-glycosylation sites assessed for partial sequence (FR2-FR4).

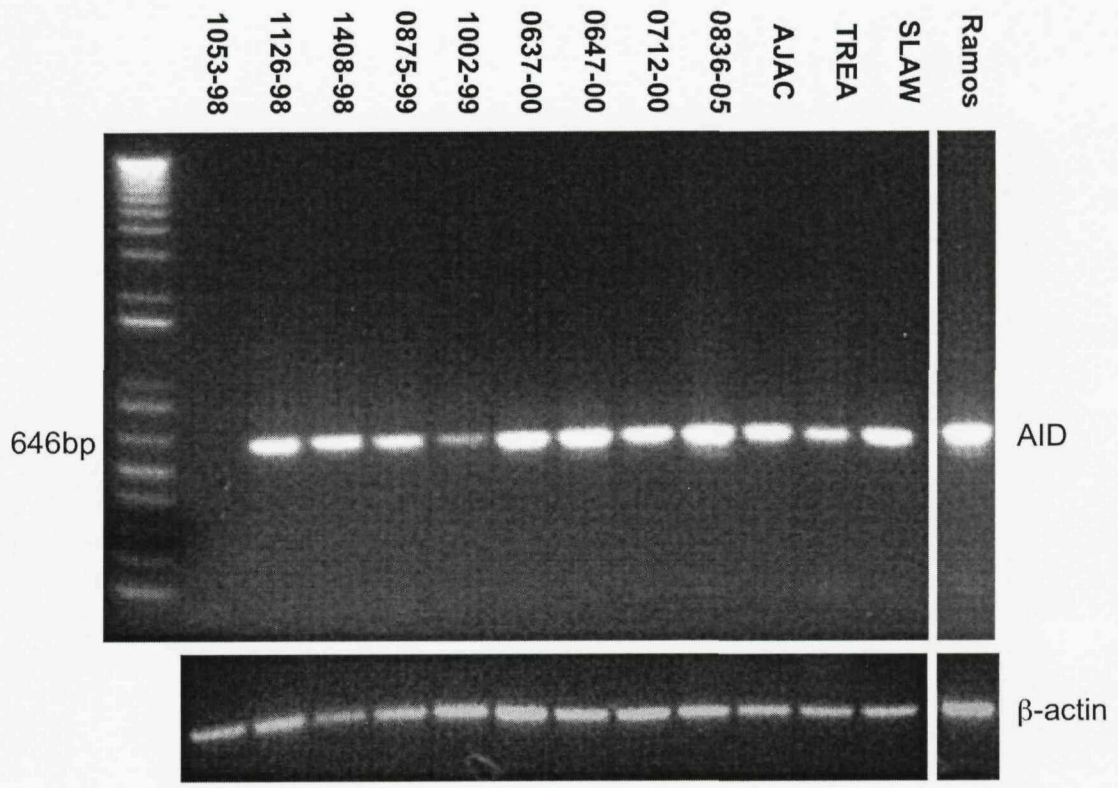


Figure 5.2. AID expression in PCNSL assessed by RT-PCR assay. Expression of AID mRNA, corresponding to an amplification product of 649bp, was detected in 11 of 12 tumours. The amount of cDNA amplified for each sample was comparable, as shown by the β actin signal. Expression of AID mRNA in the human BL cell line Ramos was used as a positive control.

which used cDNA from the Ramos human BL cell line. Direct sequencing of the products revealed 100% homology to the published *AID* mRNA reference sequence (GenBank accession number NM_020661) and confirmed *AID* mRNA expression in these tumours. In order to quantitatively assess *AID* mRNA expression in PCNSL we then performed qPCR using TaqMan technology. The relative expression of *AID* mRNA, normalised to β actin, is shown in Figure 5.3. AID mRNA transcripts were detected in 11 of 12 PCNSL in accordance with the standard PCR results.

5.2.5. Immunohistochemical analysis

Sufficient paraffin-embedded biopsy material was available for 10 of 12 PCNSLs and enabled the tumour Ig isotype at the protein level to be determined by immunohistochemisty (Table 5.4.); this analysis was kindly performed by Mrs J. Buontempo and Miss K. Smith. Seven tumours were found to express surface IgM, of which 5 were IgMκ$^+$ and 2 were IgMλ$^+$. These findings were in accordance with molecular data (Table 5.2.), with the exception of tumour 1408-98 where only surface IgM was detected in spite of the identification of both VDJ-Cμ and switched VDJ-Cγ transcripts by PCR. One tumour (AJAC) expressed surface IgM and IgG; of note only VDJ-Cγ transcripts could be detected by PCR. All 8 tumours were IgD negative. Immunohistochemistry to determine Ig isotype was inconclusive for 2 tumours. Four tumours were found to express the V4-34 gene segment, as detected my 9G4 mAb; this was in agreement with molecular data (Table 5.1.).

Immunohistochemistry confirmed a B-cell origin of the neoplastic cells in 10 cases where biopsy material was available, with tumours expressing the pan B-cell markers CD20 and CD79a, generally in the absence of CD5 (Table 5.5.); this analysis was kindly performed by Dr. M. Hickling. The mitotic activity was generally high with the majority of tumours displaying greater than 70% Ki-67 positive proliferating cells. The expression of CD10, Bcl-6 and MUM-1 enabled the classification of PCNSL into GCB-like or non-GCB-like DLBCL using the method proposed by Hans *et al.* (see Figure 4.1a.).[407] All tumours were characterised by the co-expression of Bcl-6 (GC marker) and MUM-1 (activation marker). Two tumours also expressed CD10 and, therefore, were classified as GCB-like DLBCL. The majority (8/10) of tumours displayed a non-GCB-like phenotype. The activation marker CD138 was not expressed by any of the tumours, including the 8 classified as non-GCB-like DLBCL. Finally, EBV negativity in was confirmed in each case.

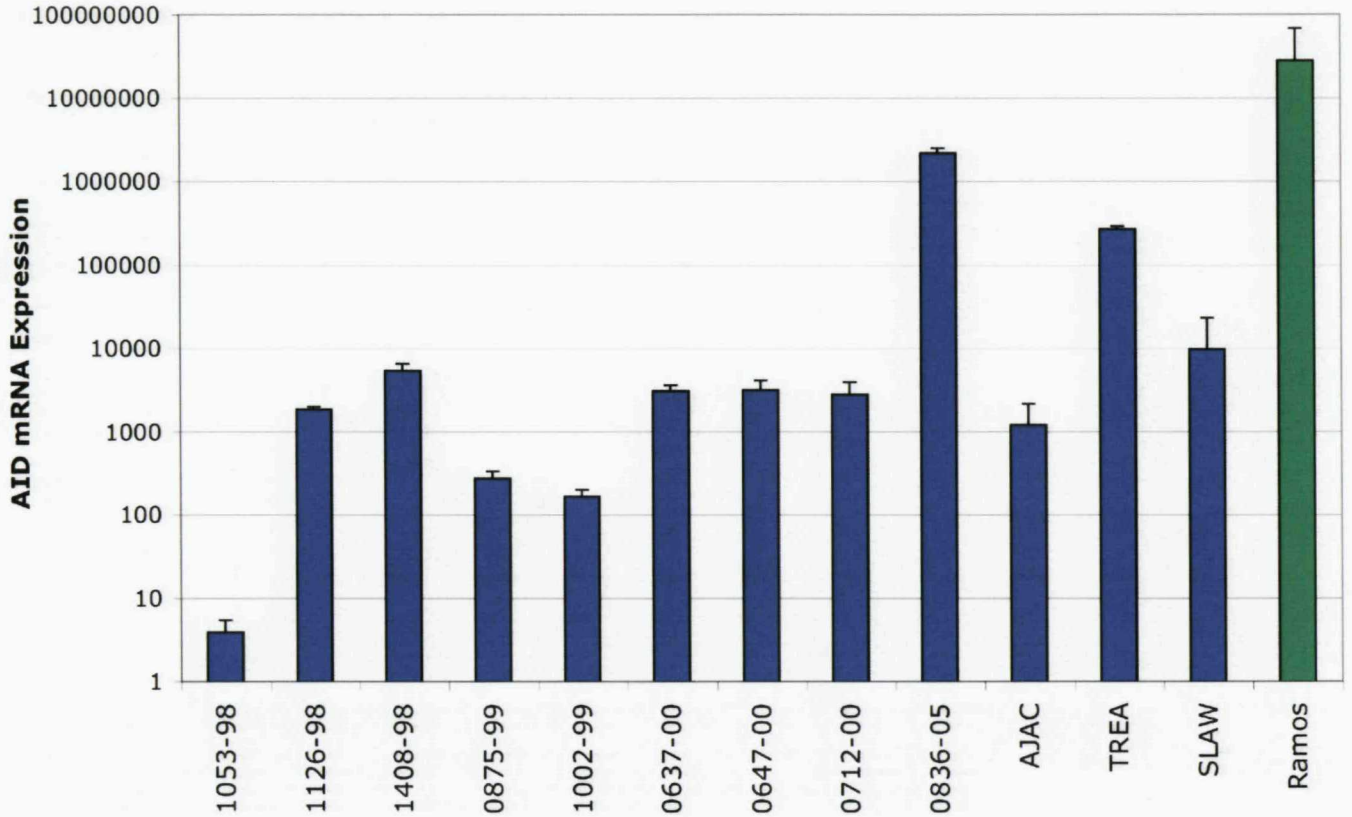


Figure 5.3. AID expression in PCNSL assessed by RT-qPCR assay. Expression of AID mRNA was normalised to β actin mRNA. AID mRNA transcripts were observed in 11 of 12 tumours and in the human BL cell line Ramos, which was used as a positive control. Analyses were performed according to the $2^{-\Delta\Delta CT}$ method. Samples were analysed in triplicate.

Table 5.4. Immunophenotyping of PCNSL

Case	Heavy chain isotype	Light chain isotype	V4-34$^+$
1053-98	IgM	Igκ	+
1126-98	IgM	Igκ	+
1408-98	IgM	Igλ	+
0875-99	ND	ND	-
1002-99	ND	ND	+
0637-00	IgM	Igλ	-
0647-00	IgM	Igκ	-
0712-00	IgM	Igκ	-
0836-05	na	na	na
AJAC	IgM, IgG	ND	na
TREA	na	na	na
SLAW	IgM	Igκ	na

Data courtesy of Mrs J. Buontempo and Miss K. Smith

ND; Not determined: immunohistochemistry was inconclusive.
na; Not assessed due to insufficient biopsy material.

Table 5.5. Classification of PCNSL into germinal centre B-cell (GCB)-like and Non-GCB-like phenotype

Case	CD5*	CD20*	CD79a*	CD138	CD30*	Bcl-2*	p53*	Ki-67≠	EBV	CD10*	Bcl-6*	MUM-1*	Phenotype°
1053-98	-	+	+	-	-	+	+	na	-	-	+	+	Non-GCB
1126-98	+F	+	+	-	-	+	+	90%	-	-	+	+	Non-GCB
1408-98	-	+	+	-	-	+	+	70%	-	-	+	+	Non-GCB
0875-99	+	+	+	-	+F	+	+	70%	-	+	+	+	GCB
1002-99	na	na	na	na	na	na	na	na	na	na	na	na	Unknown
0637-00	-	+	+	-	-	+	+	70%	-	+	+	+	GCB
0647-00	-	+	+	-	-	+	+	80%	-	-	+	+	Non-GCB
0712-00	-	+	+	-	-	+	+	80%	-	-	+	+	Non-GCB
0836-05	na	na	na	na	na	na	na	na	na	na	na	na	Unknown
AJAC	-	+	+	-	+F	+	+	80%	-	-	+	+	Non-GCB
TREA	-	+	+	-	-	+	+	90%	-	-	+	+	Non-GCB
SLAW	-	+	+	-	-	+	+	60%	-	-	+	+	Non-GCB

Data courtesy of Dr. M. Hickling

*; Positive (+) expression: >30% immunoreactive tumour cells.
≠; Percentage proliferative fraction was estimated.
°; Phenotype determined using the decision tree devised by Hans *et al.*, 2004.[407]
na; Not assessed due to insufficient biopsy material.
F; Indicates focal expression.
GCB; Germinal centre B-cell-like.
Non-GCB; Non germinal centre B-cell-like, i.e. activated B-cell (ABC)-like or 'Type 3' phenotype

5.2.6. Tracking of tumour-derived V genes in the blood and bone marrow of PCNSL patients

An anti-coagulated blood sample and BM aspirate were obtained for 3 PCNSL patients at the time of diagnosis. The V_H and V_L genes were subsequently identified from biopsy material (Table 5.1.). CDR3-specific reverse primers were designed for both the tumour-derived V_H (Figure 5.4a.) and V_L (Figures 5.4b.) gene of each patient and subsequently used in PCR with an appropriate leader primer. The results of separate heavy and light chain-specific PCRs using cDNA isolated from biopsy material, which served as a positive control, blood and BM are shown in Figure 5.5. All PCR products were directly sequenced and cloned to confirm their clonal relationship to the tumour; summarised in Table 5.6. and Table 5.7.

Tumour-specific V_H and V_L PCRs using cDNA isolated from biopsy material were successful in all 3 patients with each resulting in a single product of the expected size (~350bp). Sequence analysis of the amplification products obtained for each patient confirmed the tumour origin of these sequences, which retained 100% homology to the clonal consensus sequence, and served to validate the primers. Tumour-derived Ig V genes were detected in the blood and BM of 2 patients (TREA and AJAC) and in the blood alone for 1 patient (SLAW):

5.2.6.1. Patient AJAC

Amplification products of the expected size (~350bp) resulted from the light chain specific PCR when using cDNA isolated from the blood and the BM of the patient (Figure 5.5.). No amplification products resulted from the heavy chain specific PCR. Direct sequencing revealed that the V_L sequence identified from the blood was identical (11/11 nucleotide mutations) to that of the consensus sequence of the tumour clone identified from biopsy material (Figure 5.6.). However, direct sequencing was less conclusive for the sequence identified from the BM, where 4 mutations were shared with the tumour clone, but the remaining mutations were either absent (n=2), altered (n=1) or could not be resolved by direct sequencing (n=4); an additional mutation was also observed in FR1. PCR products were cloned and up to 32 individual sequences were assessed (Table 5.7.). Cloning served to confirm the presence of tumour-derived V_L in the BM, where a minority (5/30) of sequences had 100% (11/11 nucleotide mutations) homology to the tumour consensus sequence with a further 4 intraclonal variants, which differed from the tumour sequence by 1, 2 or 3 nucleotides (Figure 5.7a.). The dominant (21/30) subset of sequences represented a tumour-related V_L subclone, which differed from the tumour consensus sequence isolated

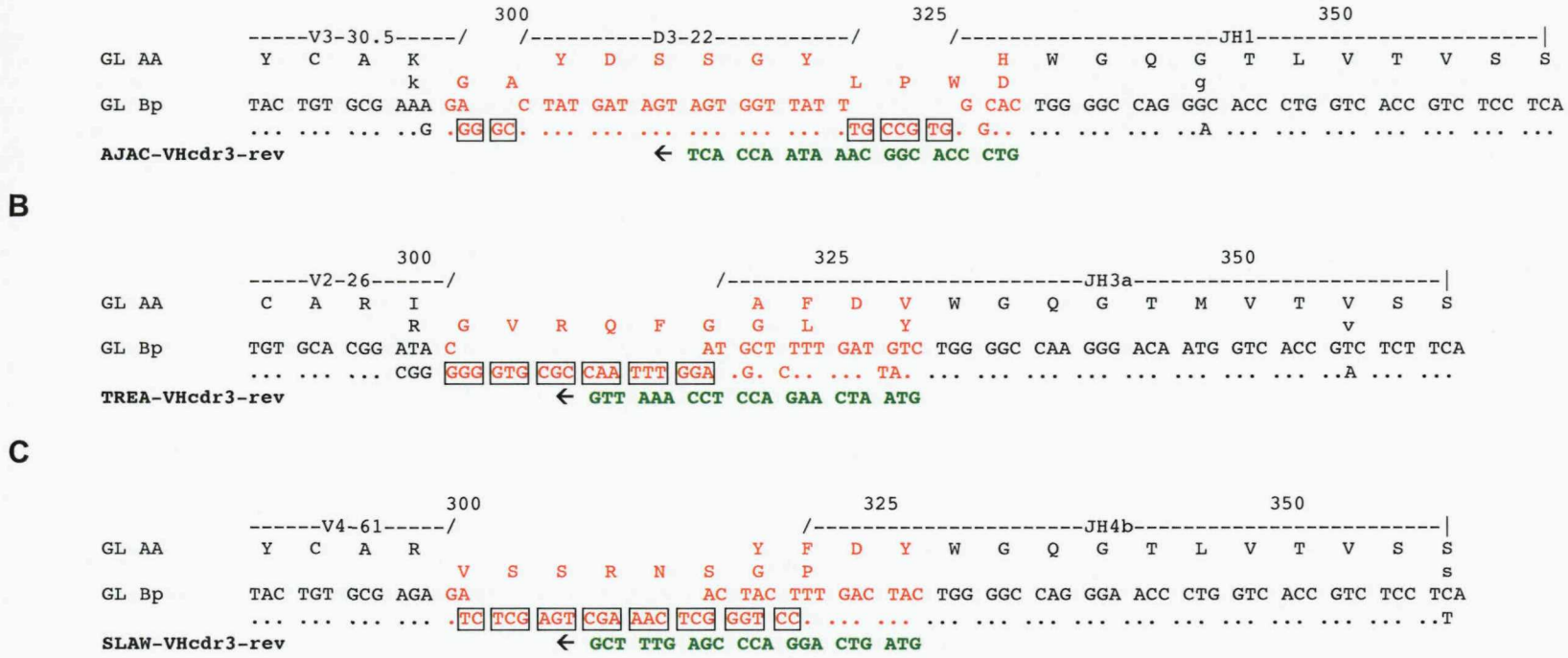


Figure 5.4a. Heavy chain CDR3 specific primer design. Individual reverse primers were designed to anneal the unique CDR3 region of the tumour-derived V_H genes in three PCNSL patients; AJAC (a), TREA (b) and SLAW (c). The CDR3 sequence to which the specific primer was designed to anneal is shown in *red*. Primer sequence is shown in *green*. Dots represent identity to germline. Mutations at the aa level are shown as upper case letters for replacement mutations or as lower case letters for silent mutations. Forward slash indicates that the germline donor gene has been trimmed. Nucleotide additions are *boxed*.

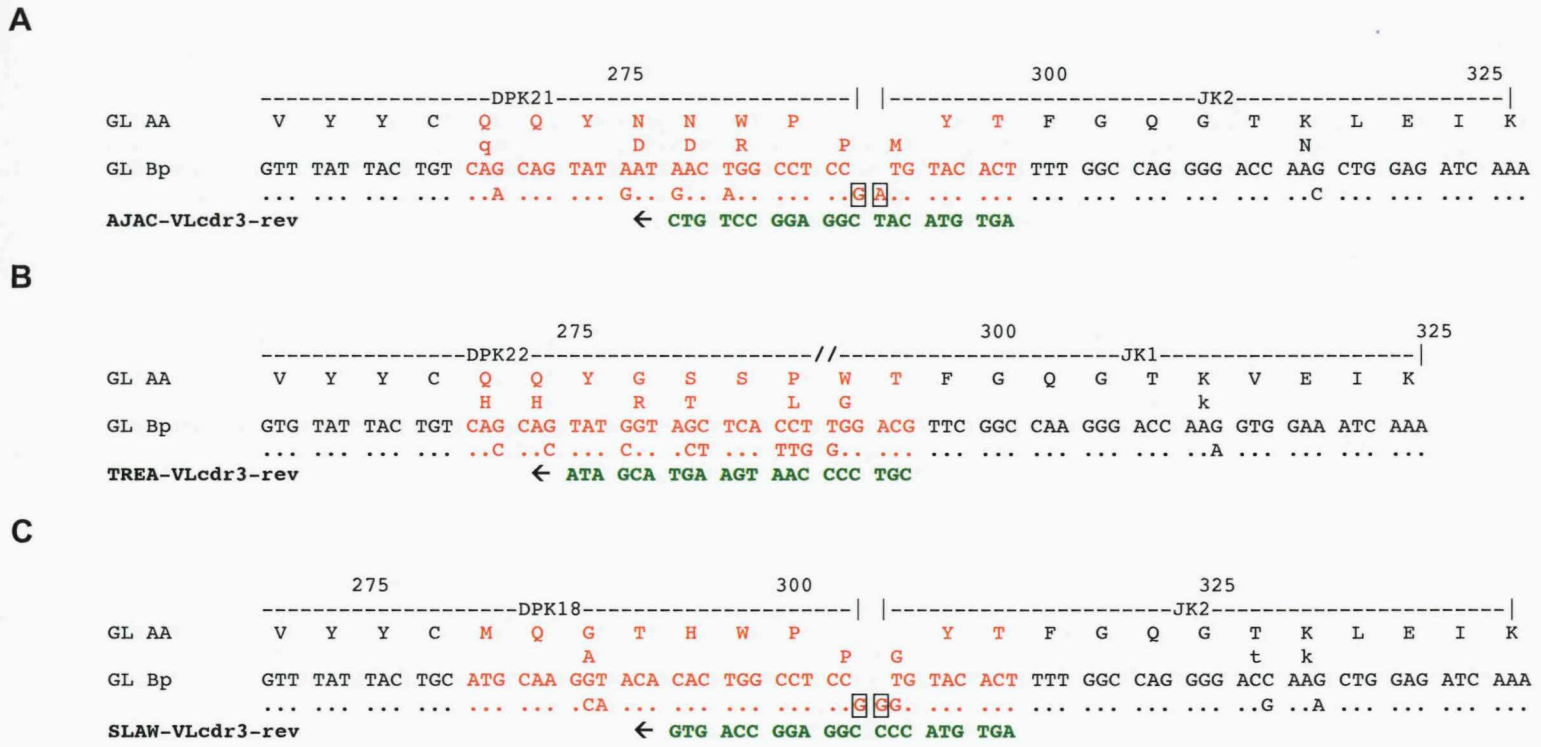


Figure 5.4b. Light chain CDR3 specific primer design. Individual reverse primers were designed to anneal the unique CDR3 region of the tumour-derived V_L genes in three PCNSL patients; AJAC (a), TREA (b) and SLAW (c). The CDR3 sequence to which the specific primer was designed to anneal is shown in *red*. Primer sequence is shown in *green*. Dots represent identity to germline. Mutations at the aa level are shown as upper case letters for replacement mutations or as lower case letters for silent mutations. Forward slash indicates that the germline donor gene has been trimmed. Nucleotide additions are *boxed*.

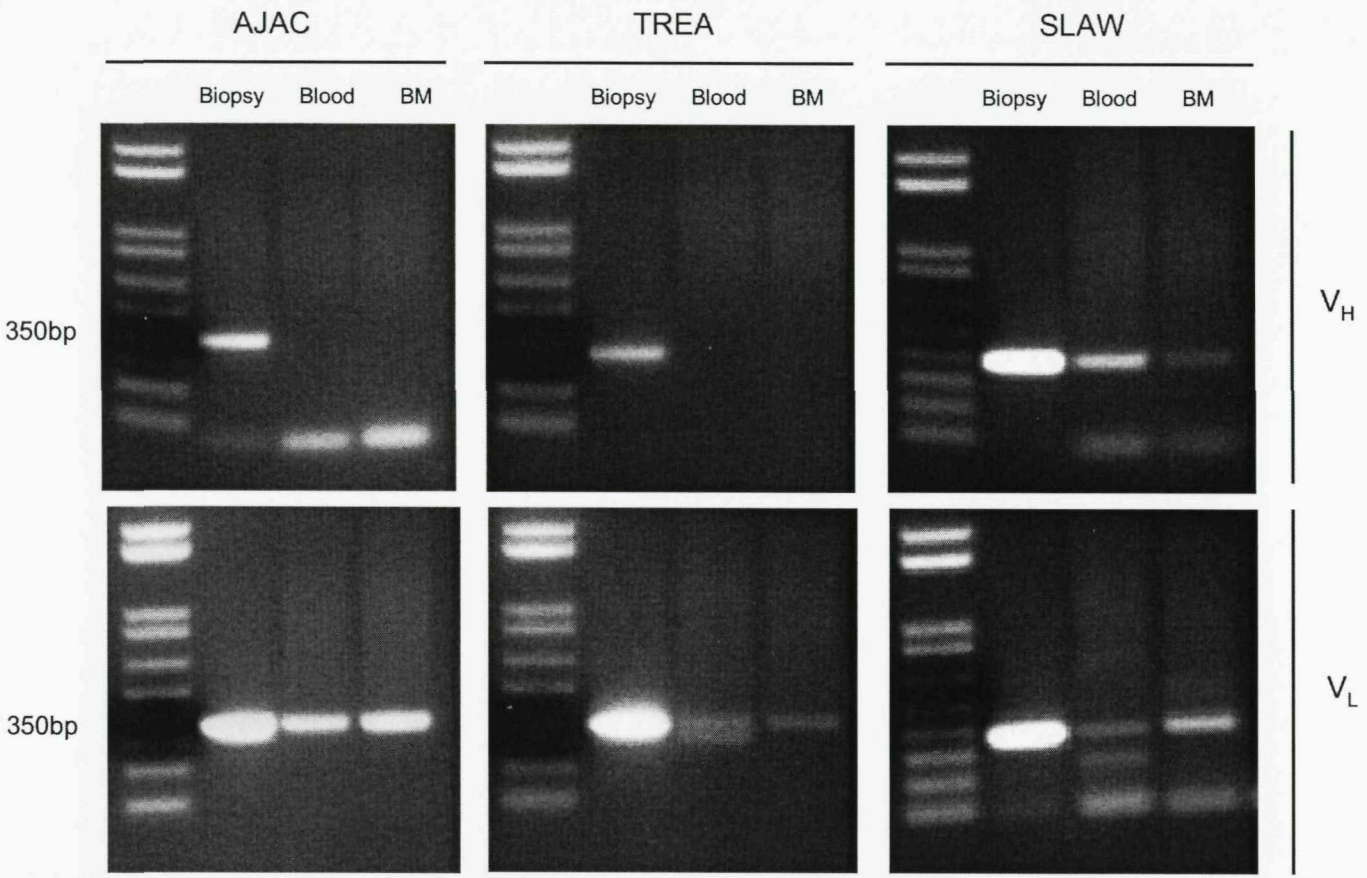


Figure 5.5. Amplification products resulting from tumour-specific PCR on cDNA isolated from the biopsy, blood and bone marrow of PCNSL patients. Individual CDR3-specific PCR for the tumour-derived V_H and V_L genes was performed on cDNA isolated from the biopsy material, blood and BM of 3 PCNSL patients (AJAC, TREA and SLAW). Amplification products of ~350bp were separated by agarose gel electrophoresis and visualised with ethidium bromide.

Table 5.6. Tracking of tumour-derived V genes by CDR3-specific PCR in the blood and bone marrow of patients with PCNSL

Case	Chain	Location	PCR amplification*
AJAC	V_H	Biopsy	Clonal
		Blood	No product
		BM	No product
	V_L	Biopsy	Clonal
		Blood	Clonal
		BM	Clonal
TREA	V_H	Biopsy	Clonal
		Blood	No product
		BM	Clonal≠
	V_L	Biopsy	Clonal
		Blood	Clonal
		BM	Clonal
SLAW	V_H	Biopsy	Clonal
		Blood	Clonal
		BM	Non-clonal
	V_L	Biopsy	Clonal
		Blood	Clonal
		BM	Non-clonal

*; Clonality assessed by direct sequencing.
≠; Tumour-derived V_H identified by CDR2-specific PCR.

Table 5.7. An analysis of somatic mutations in the tumour-derived V genes isolated from the biopsy, blood and bone marrow

Case	Chain	Location	No. tumour-derived clones/ total clones	No. bp mutations	Intraclonal heterogeneity	
					No. of clones*	%
AJAC	V_L	Biopsy	25/25	11	4/25	0.054
		Blood	25/25≠			
			↘ 20/25	11/11	7/20	0.118
			5/25	4/11[a]	5/5	0.673[R]
		BM	30/32≠			
			↘ 9/30	11/11	4/9	0.374[R]
			21/30	4/11[a]	6/21	0.144[R]
TREA	V_H	Biopsy	23/23	24°	5/23	0.134[R]
		BM	24/24	24/24°	5/24	0.129[R]
	V_L	Biopsy	25/25	9	5/25	0.095
		Blood	25/25	9/9[b]	5/25	0.245[R]
		BM	25/25	9/9	1/25	0.014
SLAW	V_H	Biopsy	25/25	29	1/25	0.012
		Blood	25/25	29/29	9/25	0.171[R]
	V_L	Biopsy	25/25	12	3/25	0.038
		Blood	25/26	12/12	3/25	0.051

*; Number of clones exhibiting sequence variation from the tumour consensus sequence / total number clones.
≠; A tumour-related V_L subclone forms a dominant (21/30) and a minor (5/25) subset of blood and BM sequences, respectively; intraclonal heterogeneity assessed separately.
[a]; Sequence includes 3 bp mutations (S) that are not present in the tumour consensus sequence and 1 bp mutation alteration (S).
°; Tumour-derived V_H identified by CDR2-specific PCR (FR2-FR4).
[b]; Sequence includes 3 bp mutations (R) that are not present in the tumour consensus sequence.
[R]; Includes intraclonal sequence mutation(s) that is repeated in >1 clone.

Figure 5.6. Tumour-derived V_L can be detected in the blood and bone marrow of PCNSL patient AJAC. CDR3-specific V_L PCR was performed on cDNA isolated from biopsy material (positive control), blood and BM of patient AJAC. Resulting amplification products were directly sequenced and aligned to the identified tumour-derived V_L gene sequence. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline. N represents unknown nucleotides, which cannot be resolved by direct sequencing.

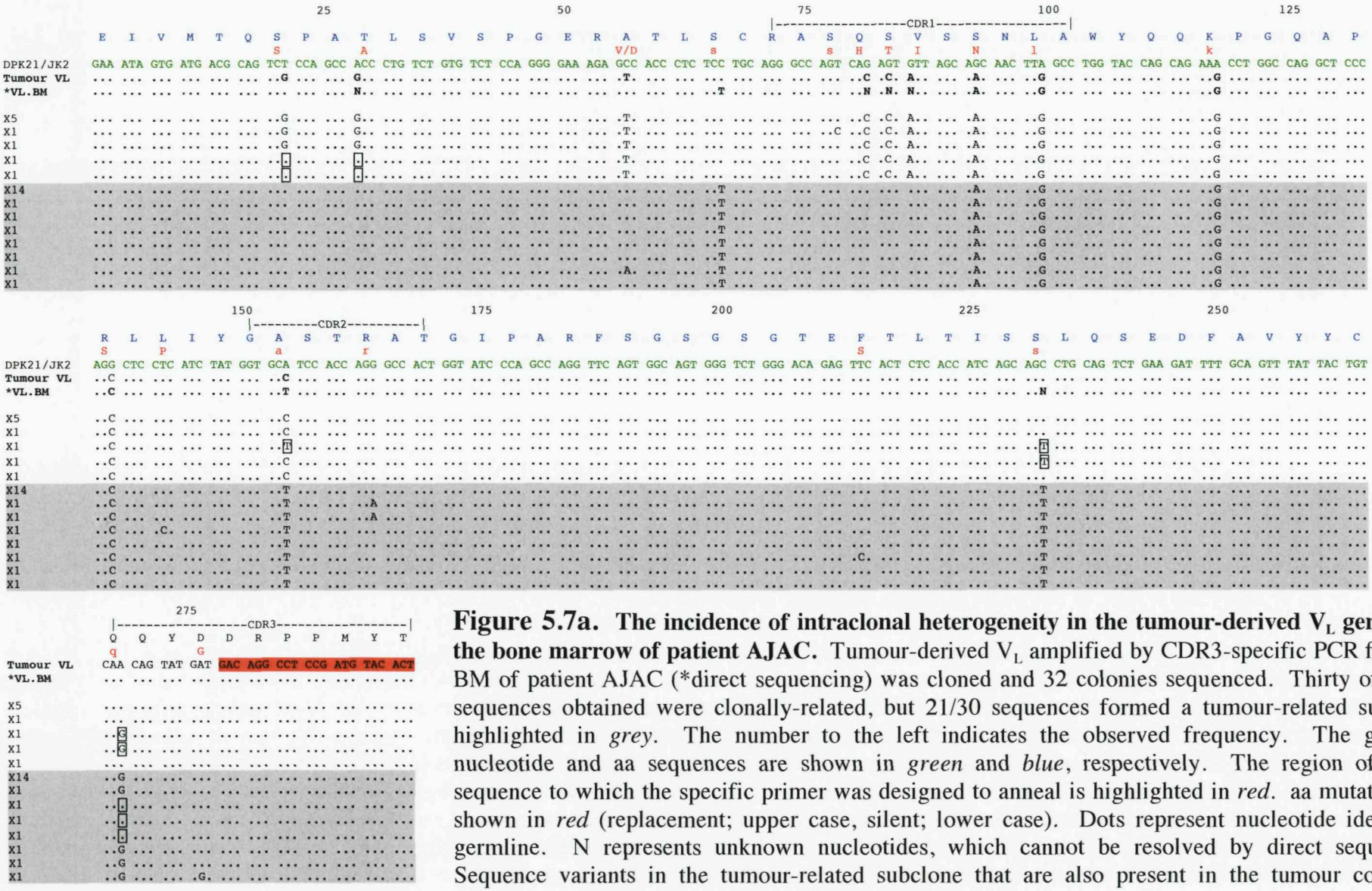


Figure 5.7a. The incidence of intraclonal heterogeneity in the tumour-derived V_L gene from the bone marrow of patient AJAC. Tumour-derived V_L amplified by CDR3-specific PCR from the BM of patient AJAC (*direct sequencing) was cloned and 32 colonies sequenced. Thirty of the 32 sequences obtained were clonally-related, but 21/30 sequences formed a tumour-related subclone, highlighted in *grey*. The number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline. N represents unknown nucleotides, which cannot be resolved by direct sequencing. Sequence variants in the tumour-related subclone that are also present in the tumour consensus sequence, or vice versa, are indicated by *boxed* nucleotides.

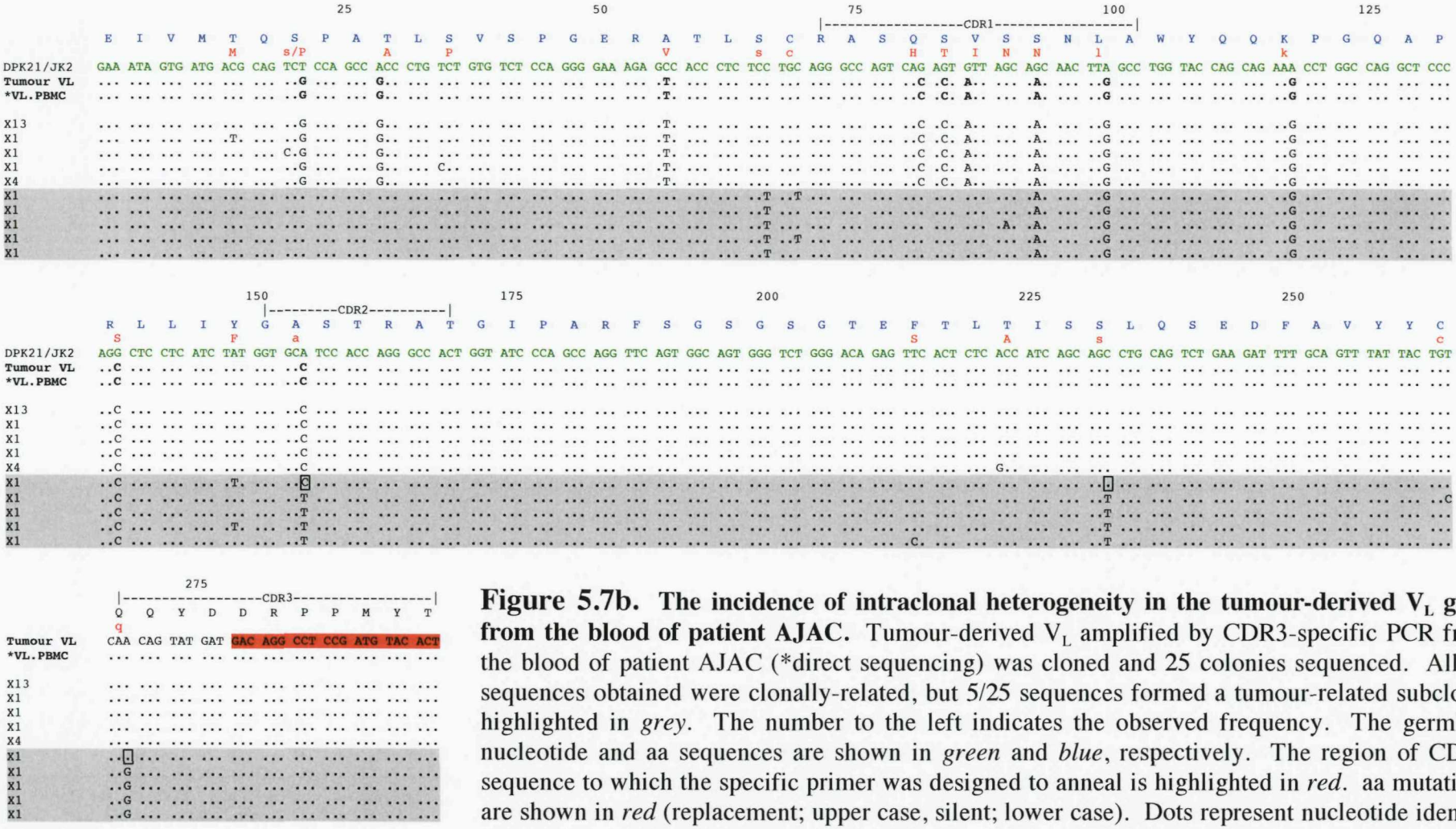


Figure 5.7b. The incidence of intraclonal heterogeneity in the tumour-derived V_L gene from the blood of patient AJAC. Tumour-derived V_L amplified by CDR3-specific PCR from the blood of patient AJAC (*direct sequencing) was cloned and 25 colonies sequenced. All 25 sequences obtained were clonally-related, but 5/25 sequences formed a tumour-related subclone, highlighted in *grey*. The number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline. Sequence variants in the tumour-related subclone that are also present in the tumour consensus sequence are indicated by *boxed* nucleotides.

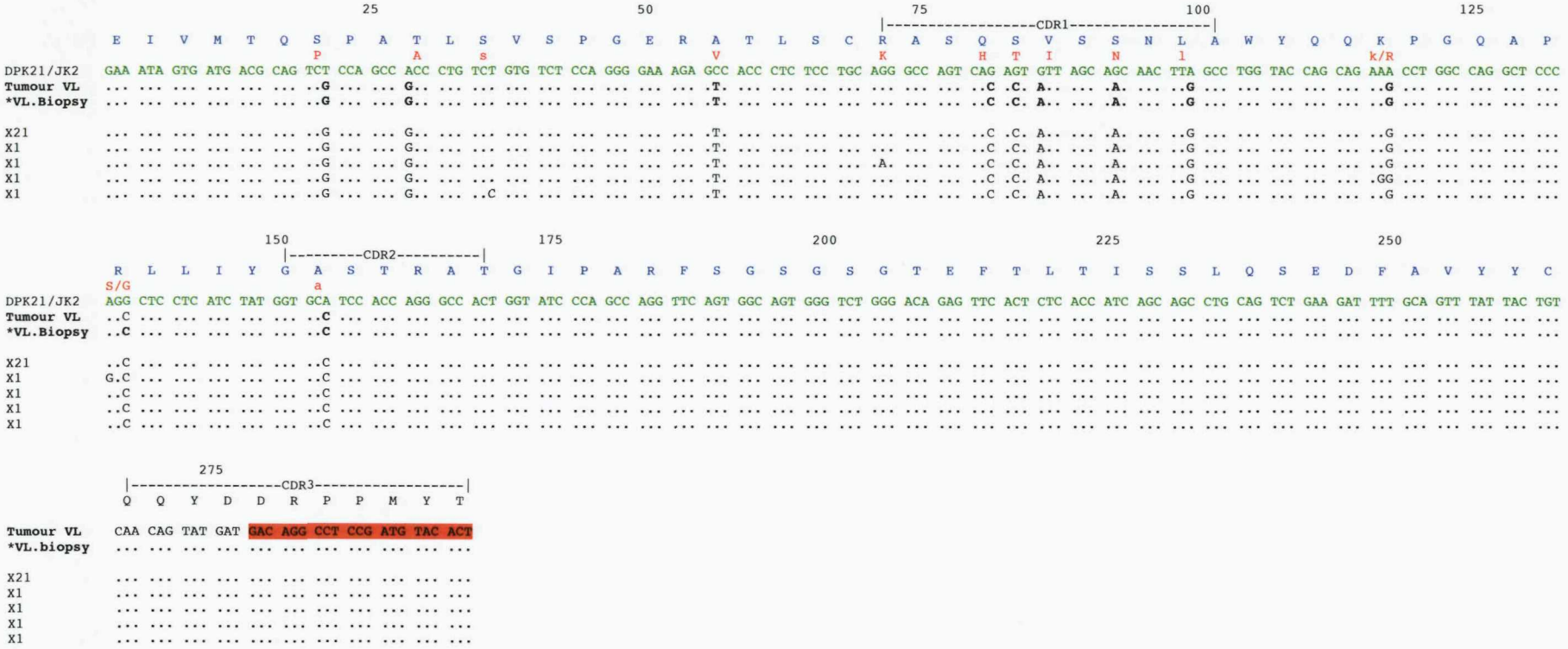


Figure 5.7c. The incidence of intraclonal heterogeneity in the tumour-derived V_L gene of patient AJAC. Tumour-derived V_L amplified by CDR3-specific PCR from biopsy material of patient AJAC (*direct sequencing) was cloned and 25 colonies sequenced. All 25 sequences obtained were clonal; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

from biopsy material by the loss of 5 replacement mutations and 1 silent mutation, the addition of 3 silent mutations and the silent change of 1 mutation from C to T. Overall, 5 nucleotide changes impacted the aa sequence. Cloning revealed that the same tumour-related V_L subclone formed a minor (5/25) subset of sequences identified from the blood (Figure 5.7b.), but was not detected in sequences from the biopsy material (Figure 5.7c.). Evidence that this subclone is truly tumour-related is strengthened by the presence of intraclonal variants in the BM and the blood, although less pronounced in the latter, that share mutations present in the tumour consensus sequence and vice versa (Figure 5.7a. and 5.7b.).

In a further attempt to identify the tumour-derived V_H gene in the blood and BM, a second PCR using a CDR2-specific forward primer [5'-TTT TAT GCA GAC TCT GTG AAG-3'] together with the Cγ primer was performed. PCR on cDNA from biopsy material was successful producing a product of the expected size (~180bp); direct sequencing revealed 100% homology to the clonal consensus sequence (data not shown). However, as with the CDR3-specific PCR, no amplification products resulted from the blood or BM.

5.2.6.2. Patient TREA

Amplification products of the expected size (~350bp) resulted from the light chain specific PCR when using cDNA isolated from the blood and the BM of the patient (Figure 5.5.). PCR using cDNA from the blood also produced a second, smaller (~250bp) product. No amplification products resulted from the heavy chain specific PCR. Direct sequencing of all 3 V_L products confirmed their tumour origin (Figure 5.8.); the ~250bp product isolated from the blood was shown to lack the first 45bp of FR1. The V_L sequence obtained from the BM showed 100% (9/9 nucleotide mutations) homology to that of the clonal consensus sequence. However, in addition to the 9 consensus mutations, 3 replacement mutations (1 in FR1 and 2 in CDR1) were observed in the sequence identified from the blood. Subsequent cloning of this product revealed that the majority of sequences featured the 3 additional mutations (n=20), with a few intraclonal variants that lacked 1 (n=1) or all 3 mutations (n=4) (Figure 5.9a. and Table 5.7.). Of note, none of these mutations were present in sequences obtained from the biopsy (n=25) (Figure 5.9b.) or BM (n=25) (Figure 5.9c.).

In a further attempt to identify the tumour-derived V_H gene in the blood and BM, a second PCR using a CDR2-specific forward primer [5'-GAG GAC ATA TCA TAC AAC ACA-3'] together with the Cμ primer was performed. PCR using cDNA from biopsy material

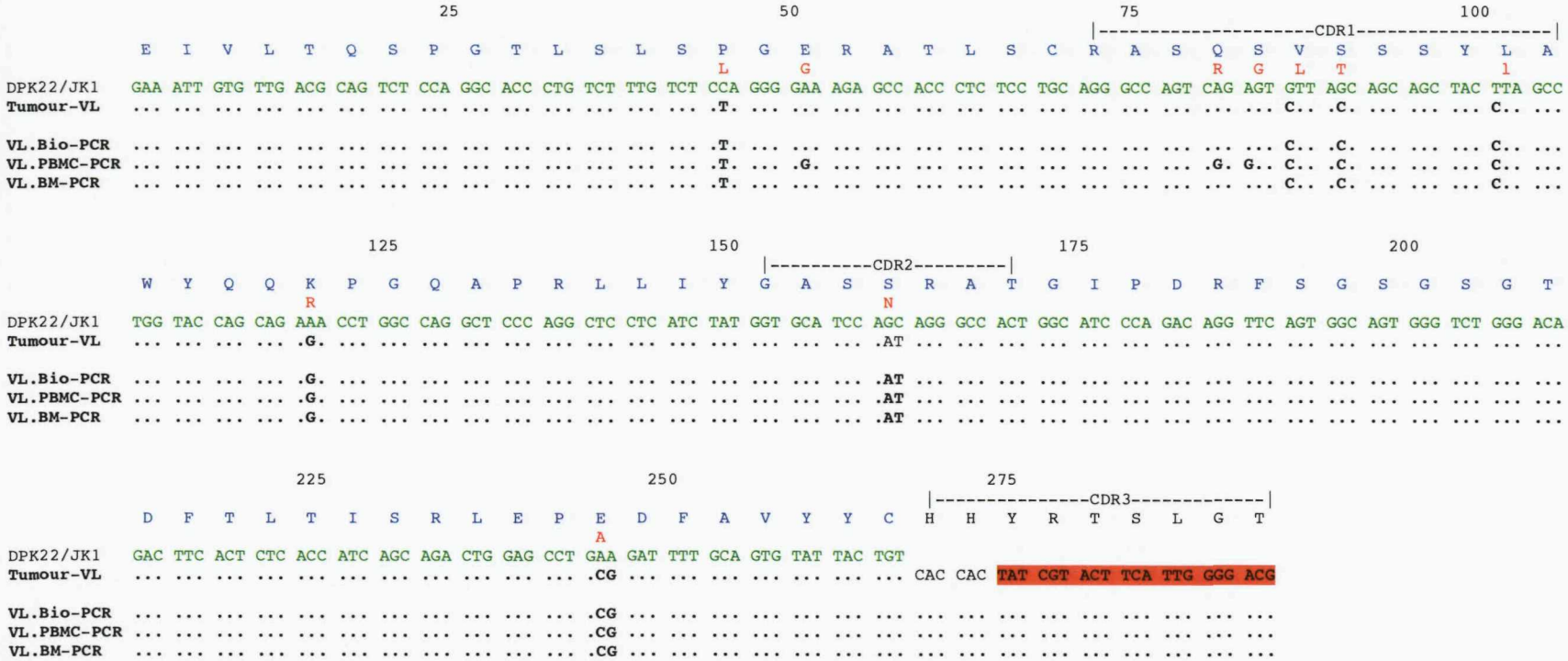


Figure 5.8. Tumour-derived V_L can be detected in the blood and bone marrow of PCNSL patient TREA. CDR3-specific V_L PCR was performed on cDNA isolated from biopsy material (positive control), blood and BM of patient TREA. Resulting amplification products were directly sequenced and aligned to the identified tumour-derived V_L gene sequence. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

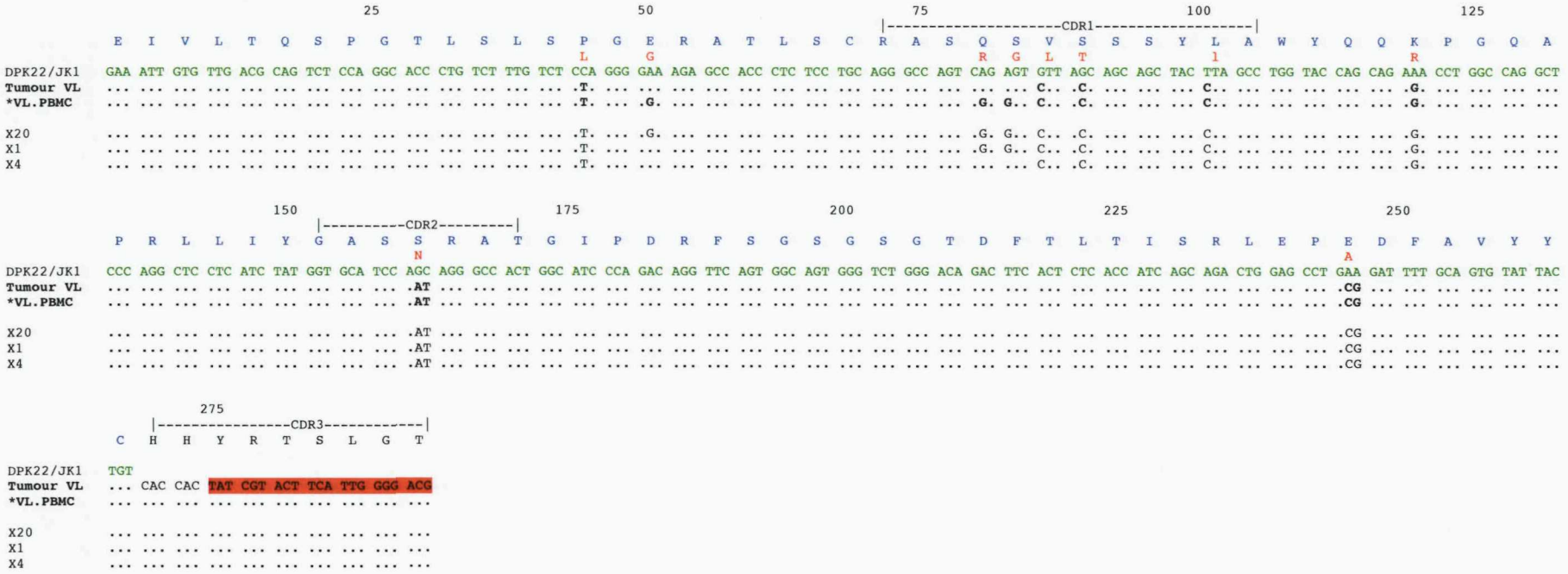


Figure 5.9a. The incidence of intraclonal heterogeneity in the tumour-derived V_L gene from the blood of patient TREA. Tumour-derived V_L amplified by CDR3-specific PCR from the blood of patient TREA (*direct sequencing) was cloned and 25 colonies sequenced. All 25 sequences obtained were clonally-related; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

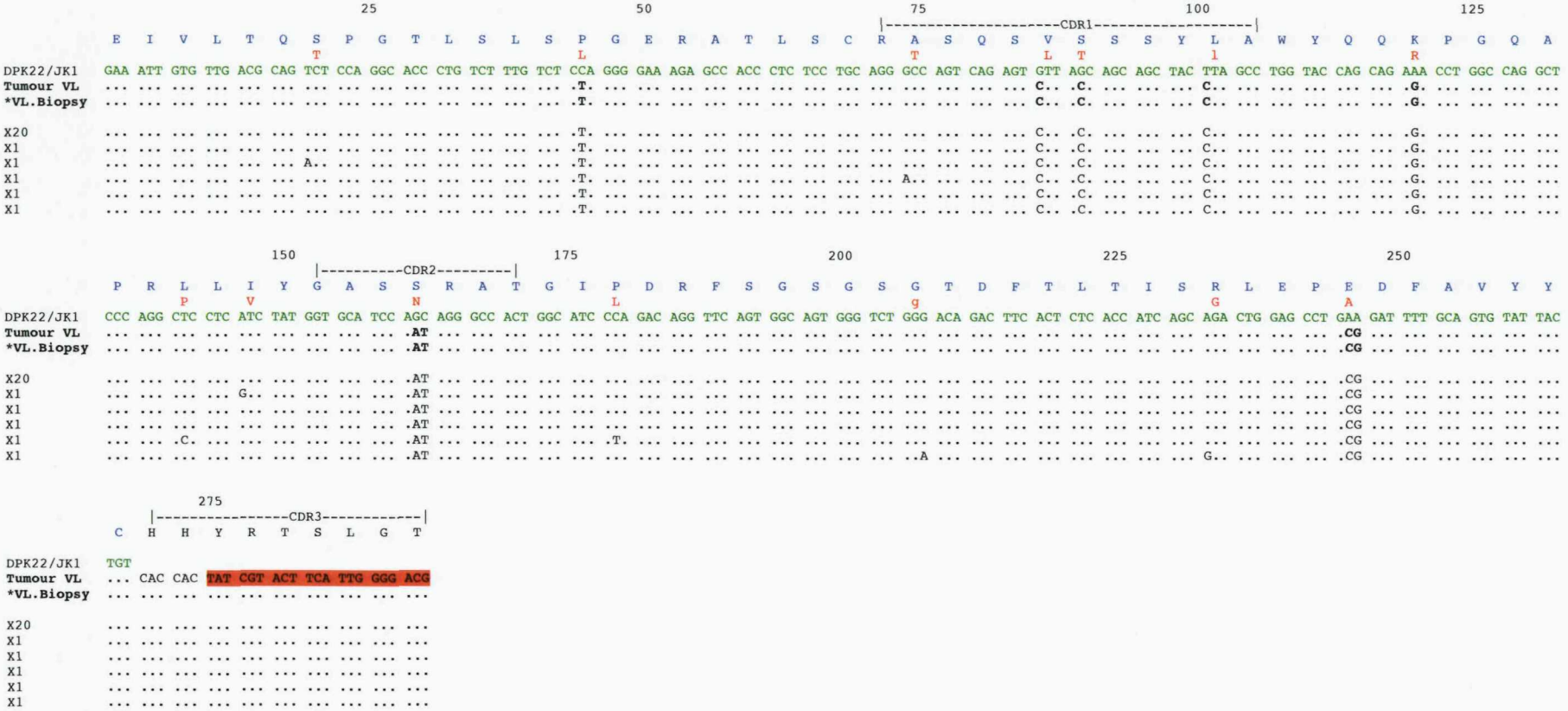


Figure 5.9b. The incidence of intraclonal heterogeneity in the tumour-derived V_L gene of patient TREA. Tumour-derived V_L amplified by CDR3-specific PCR from biopsy material of patient TREA (*direct sequencing) was cloned and 25 colonies sequenced. All 25 sequences obtained were clonal; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

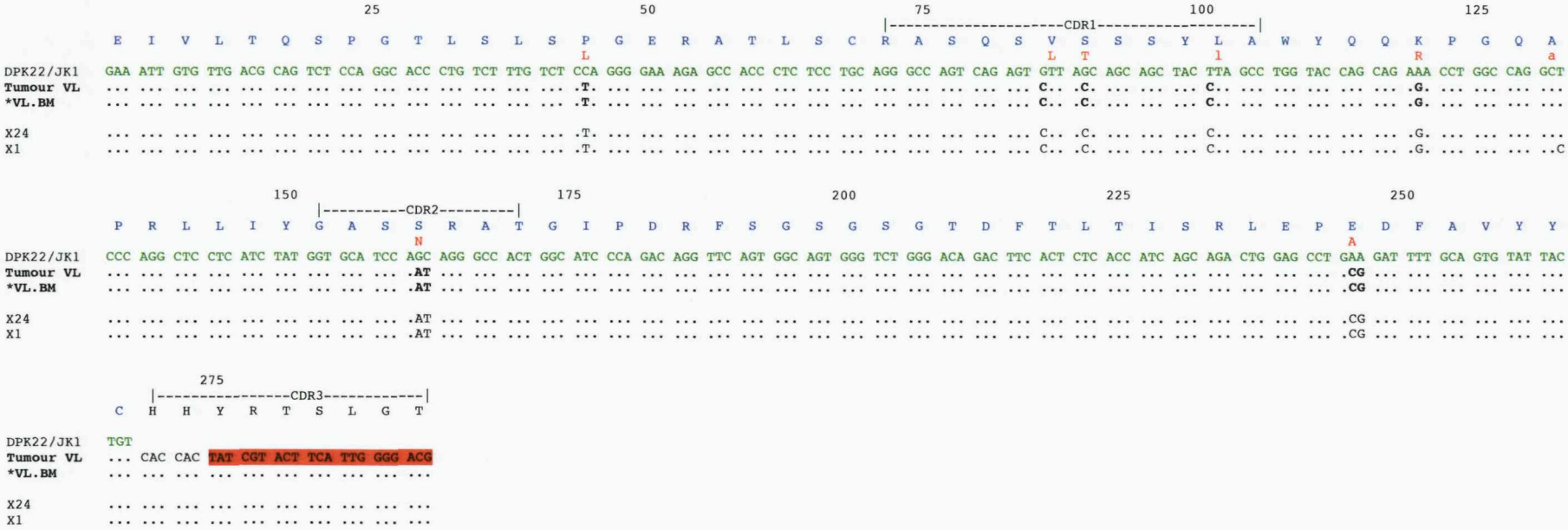


Figure 5.9c. The incidence of intraclonal heterogeneity in the tumour-derived V_L gene from the bone marrow of patient TREA. Tumour-related V_L amplified by CDR3-specific PCR from the BM of patient TREA (*direct sequencing) was cloned and 25 colonies sequenced. All 25 sequences obtained were clonal; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

was successful producing a product of the expected size (~180bp); direct sequencing and cloning revealed 100% homology to the clonal consensus sequence (Figure 5.10a.). PCR was also successful using cDNA isolated from the BM and direct sequencing and cloning confirmed 100% homology to the clonal V_H consensus sequence (Figure 5.10b.). No amplification products resulted from the blood.

5.2.6.3. Patient SLAW

Amplification products of the expected size (~350bp) resulted from both the heavy and light chain specific PCR when using cDNA isolated from the blood and the BM of the patient (Figure 5.5.). Light chain PCR on cDNA from the blood also produced a second, smaller (~200pb) product, which was later shown not to be Ig gene sequence. The reduced specificity of the PCR may be related to the fact that the CDR3 of the V_L clonal consensus sequence, for which the primer was designed to hybridise, displayed limited uniqueness, i.e. a low number of nucleotide additions and no somatic mutations at the 3' and 5' ends of DPK18 and Jκ2, respectively (see Figure 5.4b.). Direct sequencing of the V_H and V_L products amplified from the blood confirmed their tumour origin with sequences showing 100% homology to the clonal consensus sequence; 29/29 and 12/12 nucleotide mutations in V_H and V_L, respectively (Figures 5.11a. and 5.11b.). However, tumour-derived V genes were not detected in the BM of this patient and both the V_H and V_L sequences obtained were unrelated. All PCR products were cloned (Figures 5.12a., 5.12b., 5.13a. and 5.13b. and Table 5.7.). Following the assessment of a total of 50 sequences, we were able to confirm the absence of tumour-derived V_H and V_L in the BM of this patient since all V_H (n=25) and V_L (n=25) sequences obtained were unrelated (data not shown).

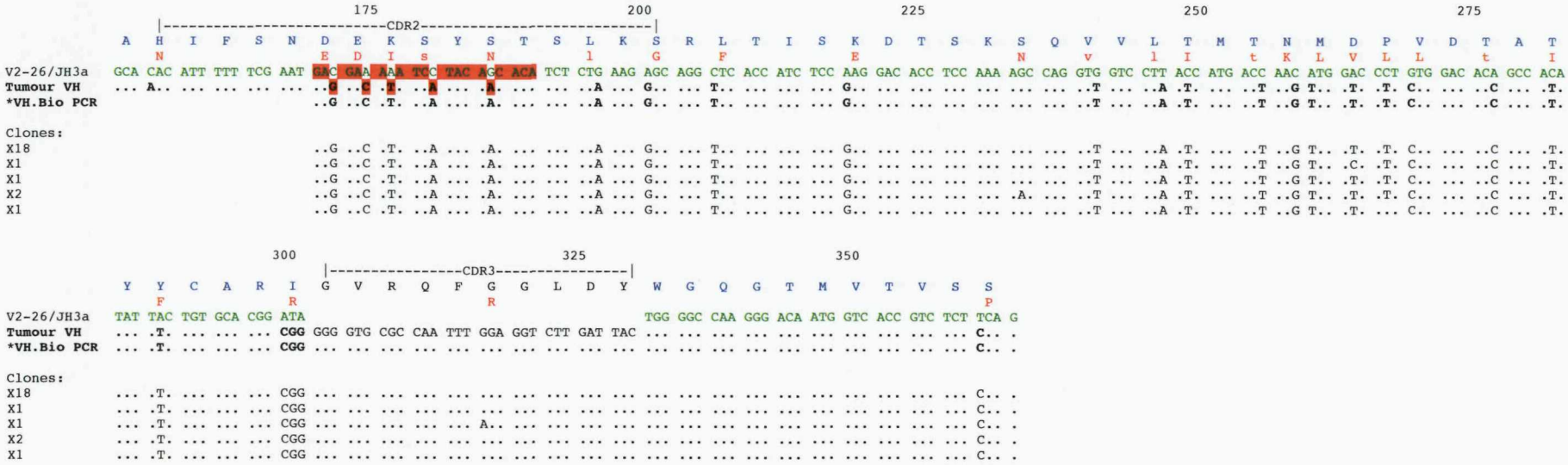


Figure 5.10a. CDR2-specific primers can be used to amplify tumour-derived V_H in patient TREA. CDR2-specific V_H PCR was successful using cDNA isolated from the biopsy material of patient TREA. The resulting amplification product was directly sequenced (*) and cloned (n=23) and the sequences aligned to the identified tumour-derived V_H gene sequence. All 23 sequences obtained were clonally related; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR2 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

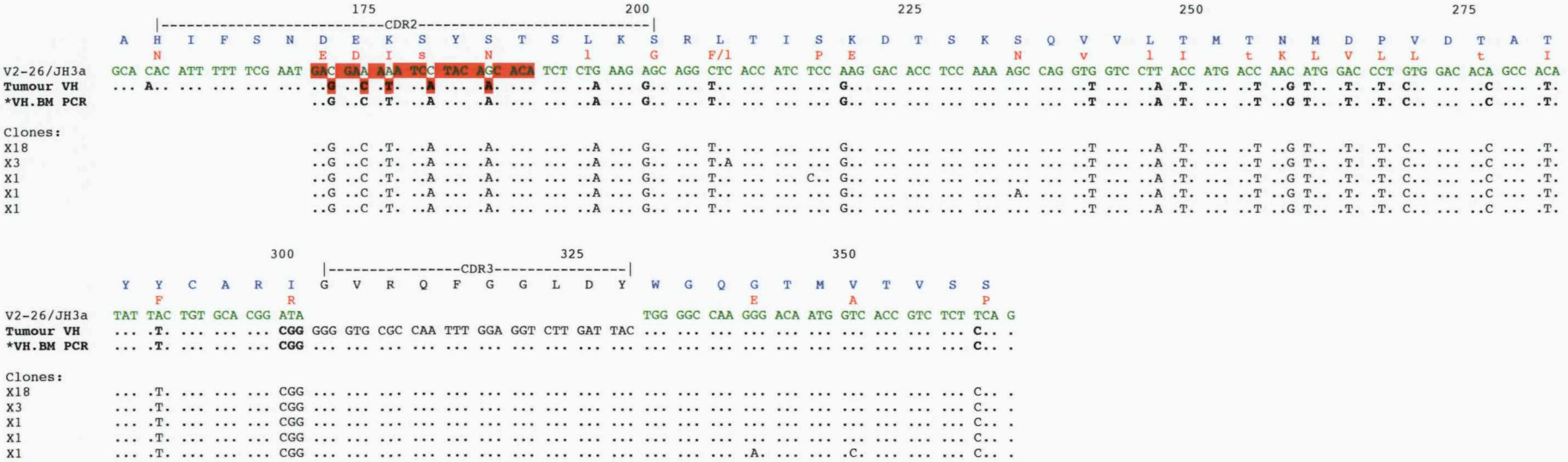


Figure 5.10b. Tumour-derived V_H can be detected in the bone marrow of PCNSL patient TREA. CDR2-specific V_H PCR was successful using cDNA isolated from the BM of patient TREA. The resulting amplification product was directly sequenced (*) and cloned (n=24) and the sequences aligned to the identified tumour-derived V_H gene sequence. All 24 sequences obtained were clonally related; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR2 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

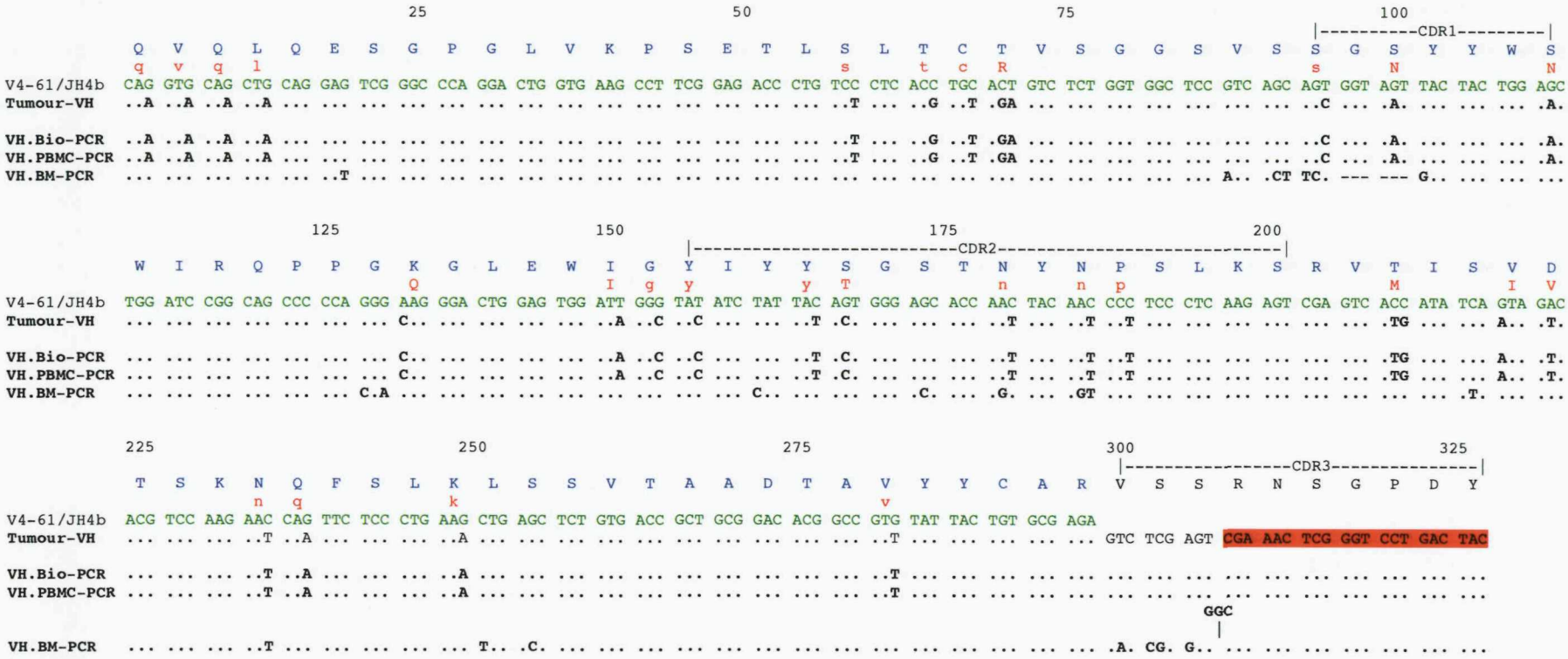


Figure 5.11a. Tumour-derived V_H can be detected in the blood but not the bone marrow of PCNSL patient SLAW. CDR3-specific V_H PCR was performed using cDNA isolated from biopsy material (positive control), blood and BM of patient SLAW. Resulting amplification products were directly sequenced and aligned to the identified tumour-derived V_H gene sequence. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline. Slashes represent nucleotide deletions.

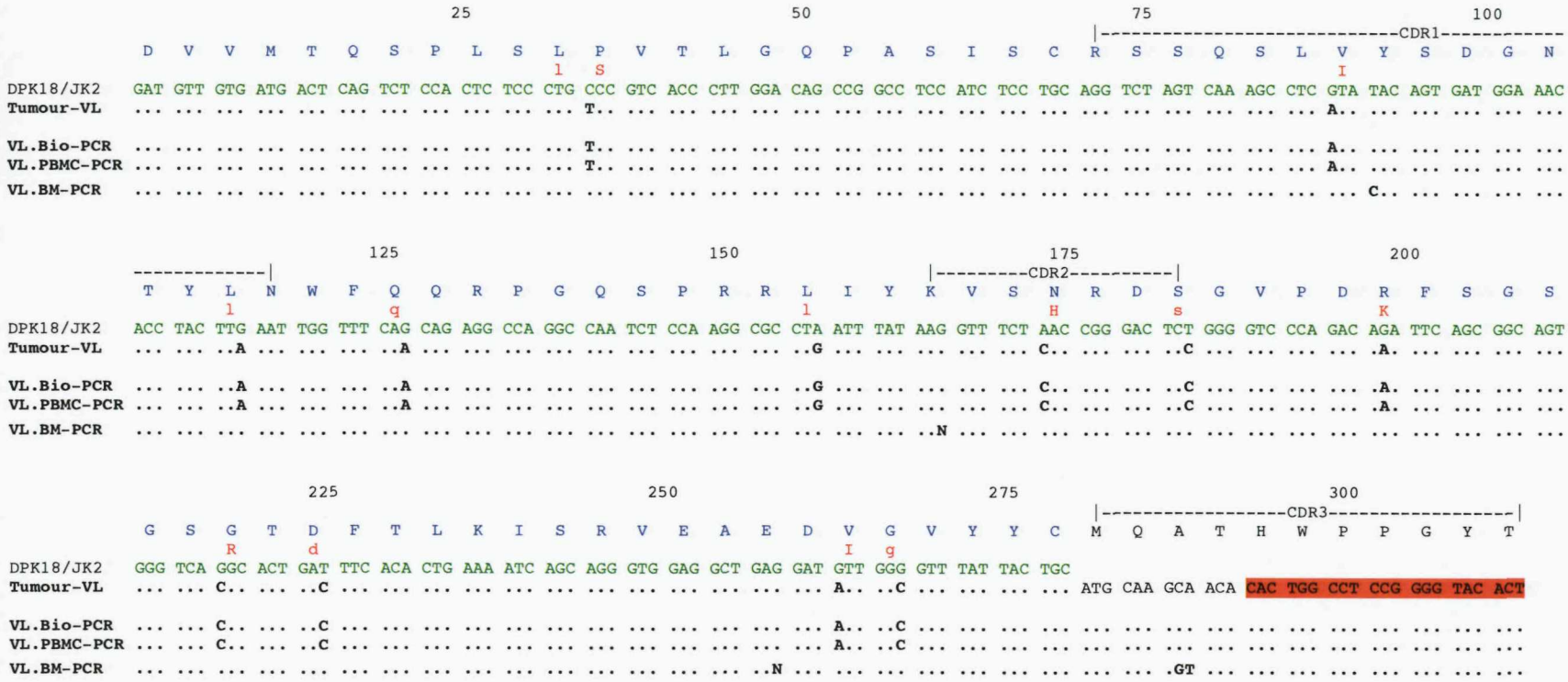


Figure 5.11b. Tumour-derived V_L can be detected in the blood but not the bone marrow of PCNSL patient SLAW. CDR3-specific V_L PCR was performed using cDNA isolated from biopsy material (positive control), blood and BM of patient SLAW. Resulting amplification products were directly sequenced and aligned to the identified tumour-derived V_L gene sequence. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline. N represents unknown nucleotides, which cannot be resolved by direct sequencing.

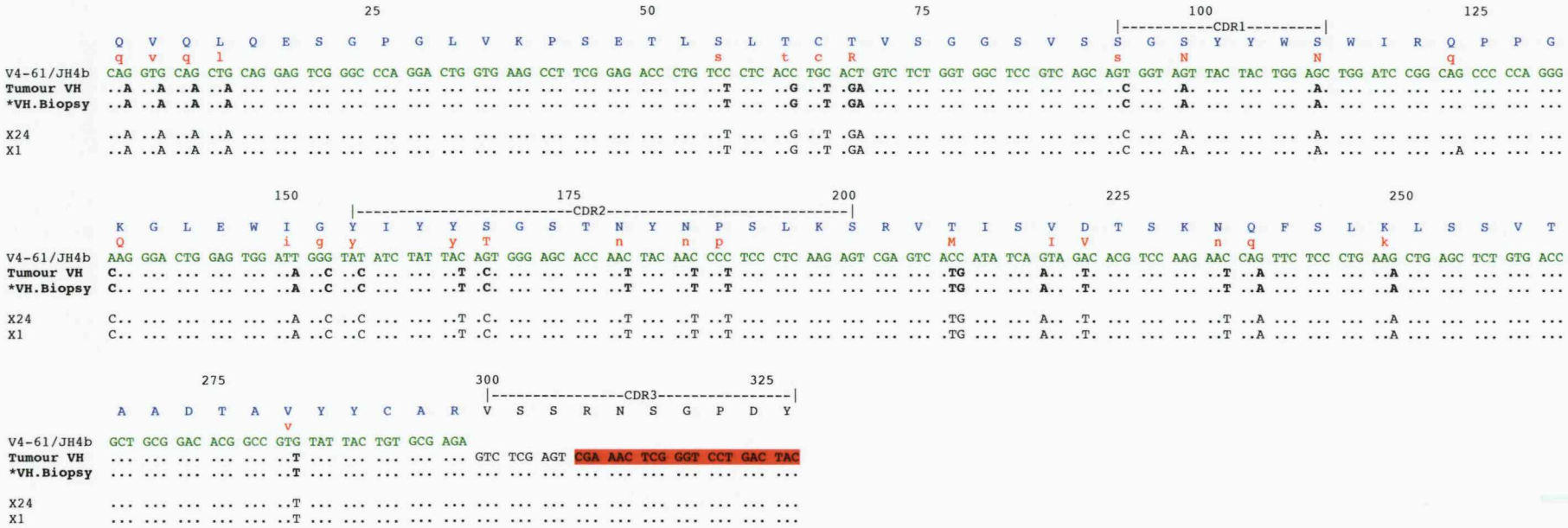


Figure 5.12a. The incidence of intraclonal heterogeneity in the tumour-derived V_H gene of patient SLAW. Tumour-derived V_H amplified by CDR3-specific PCR from biopsy material of patient SLAW (*direct sequencing) was cloned and 25 colonies sequenced. All 25 sequences obtained were clonal; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

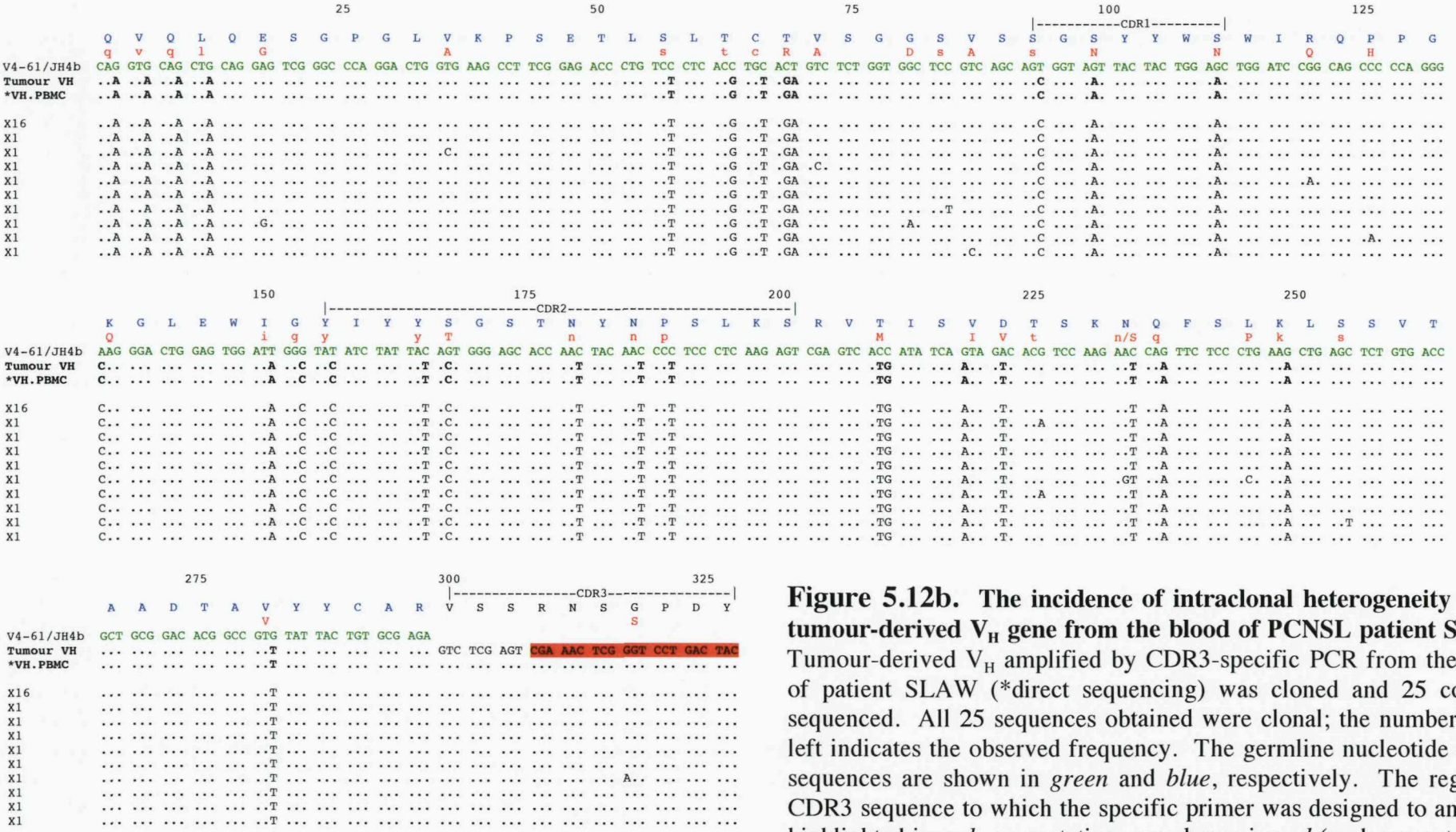


Figure 5.12b. The incidence of intraclonal heterogeneity in the tumour-derived V_H gene from the blood of PCNSL patient SLAW. Tumour-derived V_H amplified by CDR3-specific PCR from the blood of patient SLAW (*direct sequencing) was cloned and 25 colonies sequenced. All 25 sequences obtained were clonal; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

```
                                               25                              50                              75                                 100                             125
                                                                                                          |-----------------------------CDR1-----------------------------|
               D   V   V   M   T   Q   S   P   L   S   L   P   V   T   L   G   Q   P   A   S   I   S   C   R   S   S   Q   S   L   V   Y   S   D   G   N   T   Y   L   N   W   F   Q   Q   R
                                                       l   S                                                                       I                               l               q
DPK18/JK2     GAT GTT GTG ATG ACT CAG TCT CCA CTC TCC CTG CCC GTC ACC CTT GGA CAG CCG GCC TCC ATC TCC TGC AGG TCT AGT CAA AGC CTC GTA TAC AGT GAT GGA AAC ACC TAC TTG AAT TGG TTT CAG CAG AGG
Tumour VL     ... ... ... ... ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ..A ... ... ... ..A ... ...
*VL.Biopsy    ... ... ... ... ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ..A ... ... ... ..A ... ...

X22           ... ... ... ... ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ..A ... ... ... ..A ... ...
X1            ... ... ... ... ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ..A ... ... ... ..A ... ...
X1            ... ... ... ... ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ..A ... ... ... ..A ... ...
X1            ... ... ... ... ... ... ... ... ... ... ..A T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... A.. ... ... ... ... ... ... ... ..A ... ... ... ..A ... ...

                                  150                                 175                             200                             225                                 250
                                                      |-----------CDR2-----------|
               P   G   Q   S   P   R   R   L   I   Y   K   V   S   N   R   D   S   G   V   P   D   R   F   S   G   S   G   S   G   T   D   F   T   L   K   I   S   R   V   E   A   E   D   V
                                           l                       H   r       s                   K                           R       d                                                   I
DPK18/JK2     CCA GGC CAA TCT CCA AGG CGC CTA ATT TAT AAG GTT TCT AAC CGG GAC TCT GGG GTC CCA GAC AGA TTC AGC GGC AGT GGG TCA GGC ACT GAT TTC ACA CTG AAA ATC AGC AGG GTG GAG GCT GAG GAT GTT
Tumour VL     ... ... ... ... ... ... ... ..G ... ... ... ... ... C.. ... ... ..C ... ... ... ... .A. ... ... ... ... ... ... C.. ... ..C ... ... ... ... ... ... ... ... ... ... ... ... A..
*VL.Biopsy    ... ... ... ... ... ... ... ..G ... ... ... ... ... C.. ... ... ..C ... ... ... ... .A. ... ... ... ... ... ... C.. ... ..C ... ... ... ... ... ... ... ... ... ... ... ... A..

X22           ... ... ... ... ... ... ... ..G ... ... ... ... ... C.. ... ... ..C ... ... ... ... .A. ... ... ... ... ... ... C.. ... ..C ... ... ... ... ... ... ... ... ... ... ... ... A..
X1            ... ... ... ... ... ... ... ..G ... ... ... ... ... C.. ... ... ..C ... ... ... ... .A. ... ... ... ... ... ... C.. ... ..C ... ... ... ... ... ... ... ... ... ... ... ... A..
X1            ... ... ... ... ... ... ... ..G ... ... ... ... ... C.. ..A ... ..C ... ... ... ... .A. ... ... ... ... ... ... C.. ... ..C ... ... ... ... ... ... ... ... ... ... ... ... A..
X1            ... ... ... ... ... ... ... ..G ... ... ... ... ... C.. ... ... ..C ... ... ... ... .A. ... ... ... ... ... ... C.. ... ..C ... ... ... ... ... ... ... ... ... ... ... ... A..

                          275                             300
                                  |-------------------CDR3-------------------|
               G   V   Y   Y   C   M   Q   A   T   H   W   P   P   G   Y   T
              g/S
DPK18/JK2     GGG GTT TAT TAC TGC
Tumour VL     ..C ... ... ... ... ATG CAA GCA ACA CAC TGG CCT CCG GGG TAC ACT
*VL.Biopsy    ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

X22           ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
X1            A.C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
X1            ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
X1            ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
```

Figure 5.13a. The incidence of intraclonal heterogeneity in the tumour-derived V_L gene of patient SLAW. Tumour-derived V_L amplified by CDR3-specific PCR from biopsy material of patient SLAW (*direct sequencing) was cloned and 25 colonies sequenced. All 25 sequences obtained were clonal; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

Figure 5.13b. The incidence of intraclonal heterogeneity in the tumour-derived V_L gene from the blood of PCNSL patient SLAW. Tumour-derived V_L amplified by CDR3-specific PCR from the blood of patient SLAW (*direct sequencing) was cloned and 26 colonies sequenced. Twenty-five of the 26 sequences obtained were clonal; the number to the left indicates the observed frequency. The germline nucleotide and aa sequences are shown in *green* and *blue*, respectively. The region of CDR3 sequence to which the specific primer was designed to anneal is highlighted in *red*. aa mutations are shown in *red* (replacement; upper case, silent; lower case). Dots represent nucleotide identity to germline.

5.3. Discussion

In the first part of this study we have focused on the characteristics of the V genes of PCNSL. We have confirmed that there is a striking bias in V_H gene usage in these tumours towards the V4-34 gene segment and genes of the V_H3 family. Biased usage of the V4-34 gene strongly suggests a functional role of the Ig encoded by V4-34 in the development of PCNSL, such that sustained antigenic stimulation may drive the proliferation of the malignant B cells or their precursors. A biased usage of the V4-34 gene has been reported in mutated CLL and coincides with evidence for a unique conserved CDR3 sequence.[578] These findings would indicate that stimulation by a particular Ag plays an important role in the development of mutated CLL. Closer examination of the aa sequence of V4-34 genes rearranged by 19 PCNSLs reveals that there are no conserved sequences within the CDR3 and provides evidence against such a role for Ag in this case (Figure 5.1.). Alternatively, it is possible that a SAg-driven clonal expansion plays a role in PCNSL lymphomagenesis, targeting V4-34 and V_H3 gene family members. Unlike conventional Ags, SAgs interact with the more conserved sequences of the FR1 and FR3 regions, but not the FR2 region which is found buried within the Ig molecule, to lead to the activation of large numbers of B cells. We assessed the degree of germline sequence conservation found in the FR1 and FR3 regions of V4-34 genes used by PCNSL and found a high level of conservation, which was highest in FR1, in the face of a high somatic mutation load. It is important to note that although somatic mutation is known to preferentially target hotspots in the CDRs, we observe increased somatic mutation in the FRs in PCNSL where overall somatic mutation is unusually high. Furthermore, we observe that a number of codons are particularly prone to mutational change, and that this often result in the same aa substitution. Therefore, we suggest it is likely that a SAg and not an Ag plays a role in the development of PCNSL.

In keeping with previous reports, our data support a GC-related B cell origin for PCNSL; tumour-derived V genes display a high mutational load and we find evidence for ongoing somatic mutation in >58.3% of cases, indicating engagement of the BCR post transformation. Furthermore, we observed the expression of *AID* mRNA in 11 of 12 tumours. AID has been shown to be selectively expressed in centroblast B cells in the GC of secondary lymphoid organs, where CSR and SHM may be active, and in GC-derived malignancies. Although we have not yet confirmed AID expression at the protein level in our study, it has been shown previously that the detection of *AID* transcripts and AID protein is well correlated, indicating that RNA-based studies yield valuable information about AID expression.[220] Our findings are in contrast to those of Montesinos-Rongen *et al.*

who reported that *AID* mRNA expression is uncommon in PCNSL.[461] This report appears inconsistent owing to increasing evidence for ongoing SHM in both tumour-derived V genes and proto-oncogenes.[289,400,401] Tumour 1053-98 was the only tumour not to express *AID* mRNA in our study, but it did show intraclonal heterogeneity indicative of ongoing mutation. It is important to distinguish between the detection of intraclonal heterogeneity and active SHM. In this case, it is possible that SHM is still active in an early phase of lymphoma clone expansion leading to intraclonal heterogeneity, but that in later phases of disease, SHM activity and, therefore, AID expression, is arrested. Vice versa, *AID* was expressed in the absence of intraclonal heterogeneity in tumour 0836-05 (and 4 additional tumours if more stringent criteria are applied). Although *AID* expression and intraclonal heterogeneity have been shown to be tightly associated in FL, the dissociation we describe here has been reported previously in extracerebral DLBCL.[540-542] Indeed, we have also observed a dissociation in extracerebral DLBCL in this study (see Chapter 4).

Our data reveal that although VDJ-Cμ/Cδ transcripts were most common, isotype switching to VDJ-Cγ can occur, as was seen in 3 tumours. Furthermore, we were able to confirm IgG expression at the protein level in 1 of 2 tumours where biopsy material was available; for tumour 1408-98, the surface expression of IgM, but not IgG, was observed in spite of the detection of μ, δ and switched γ transcripts by PCR. This suggests that the tumour cells of PCNSL are capable of further maturation steps. Our findings are in agreement with an early immunohistochemical study that found 2/18 IgG-expressing tumours.[549] However, the data are in contrast to a more recent study that reported the exclusive transcription of IgM and IgD mRNA in 12 tumours.[461] *AID* mRNA was expressed in all 3 tumours displaying VDJ-Cγ transcripts, although, it is important to stress that as with SHM we must distinguish between the detection of switched transcripts and active CSR. We observed the expression of multiple clonally related μ, δ and γ transcripts in 2 of 3 switched cases. Such expression might suggest that these cells were in the process of CSR when transformation occurred. However, since we have identified the tumour-derived V genes from total tumour cell populations, it is impossible to know whether both transcripts are expressed in each cell of the tumour clone or if there exists distinct subpopulations of cells within the tumour clone, a proportion expressing IgM/D and the remainder IgG; this question can only be answered by single cell PCR.

In contrast to the molecular data discussed above, immunohistochemical analysis did not support a GC B cell origin for PCNSL. CD10 was rarely expressed and, thus, only two tumours were assigned to the GCB-like subgroup. Our findings suggest that PCNSL

should be classified as non-GCB-like and not GCB-like DLBCL as would be predicted from molecular data and *AID* expression. Of note, one of the GCB-like cases (0875-99) showed only dubious evidence for ongoing mutation, with low levels of intraclonal heterogeneity and an absence of repeated point mutations among molecular clones. Conversely, five of the non-GCB-like tumours showed clear evidence for ongoing mutation. It has been shown in extracerebral DLBCL that tumours with a GC gene expression profile exhibit intraclonal heterogeneity indicative of active somatic mutation, while ABC-like tumours do not.[404] Therefore, our data do not fit with previous reports. However, with regards to *AID* expression, a dissociation with other GC genes has been demonstrated for extracerebral DLBCL where *AID* expression has been found in both GCB-like and ABC-like DLBCL.[541] All 8 tumours that were classified as non-GCB-like DLBCL were characterised by the co-expression of a GC marker (Bcl-6) and an activation marker (MUM-1), a situation that does not exist in the reactive GC where Bcl-6 and MUM-1 are mutually exclusive, but has been described for extracerebral DLBCL.[536] These findings are in agreement with those of Camilleri-Broët *et al.* who found 79 of 82 PCNSLs were of the non-GCB-like DLBCL and of which ~50% co-expressed Bcl-6 and MUM-1.[552] In this report, the authors suggested that since MUM-1 may also be expressed in a subset of cells located in the GC, preferentially in the light zone (late GCB), an "activated GC" pattern would seem to reflect more accurately the histogenesis of Bcl-6$^+$MUM-1$^+$ lymphomas. The activated GC B cell pattern was first described by Chang *et al.* for the classification of *de novo* DLBCL and was characterised by the expression of at least one GC marker (CD10 or Bcl-6) and one activation marker (MUM-1 or CD138).[538] Indeed, there was no expression of CD138 by our Bcl-6$^+$MUM-1$^+$ tumours. The proposal of an "activated GCB-like" phenotype, corresponding to a late GC/early post-GC B cell, may explain previous findings, such as active hypermutation and *AID* expression. A more recent study has investigated the gene expression profile of 21 PCNSLs using the Affymetrix oligonucleotide array. Here, the authors confirm a gene expression pattern most closely related to late GC B cells and suggest these tumours are more closely related to memory B cells than to GC B cells.[579]

A high incidence of *N*-glycosylation sites in the tumour-derived V genes has been shown to be associated with GC-derived malignancies, such as FL and eBL (see Chapter 3).[12,13,460] In our study, half (50.0%) of the PCNSL cases displayed either a novel *N*-glycosylation site (n=3) or retained the germline V4-34 site (n=3). Based on our findings in Chapter 3, it is likely that the germline site is not functionally glycosylated and holds no importance for GC-related tumours. Therefore, only 3 (25.0%) tumours displayed a novel *N*-

glycosylation site in their V_H gene; of these one was classified as GCB-like DLBCL by immunohistochemistry. Overall, 1 of 2 (50.0%) GCB-like DLBCL and 2 of 8 (25.0%) of non-GCB-like DLBCL gained a N-glycosylation site. It is difficult to assess the significance of these findings since a low number if cases were examined. Furthermore, it is unclear if the lower frequency of motifs in non-GCB-like PCNSL reflects the background level (~10%) seen in normal B cells.

Typically, PCNSLs are characterised by the strict confinement of tumours to the CNS, as revealed by contrast-enhanced CT scan of patients at diagnosis. In the second part of this study we have investigated the existence of tumour cells outside of the CNS in 3 patients at the time of diagnosis; each patient was confirmed to be free of systemic disease by CT scan. Using the unique tumour-derived V gene sequences to probe the blood and BM, we have shown that tumour cells can be detected at peripheral sites in 3 of 3 patients and assume that this must be a common feature in patients with PCNSL. Indeed, whilst performing this study, Jahnke *et al.* reported the presence of clonal V gene sequences in the blood and BM of two PCNSL patients.[580] Although significant, the finding of tumour cells in the periphery does not provide important information regarding the direction of movement of the tumour cells and leaves us with two possible situations: (i) that tumour cells present in the periphery are entering the CNS, which may be via transient migration or a continuous "feeding" of the CNS-localised clone by inward migration, or (ii) that there is dissemination of tumour cells from the primary site in the CNS to the periphery. In view of the evidence that immune cells in the brain parenchyma are unable to migrate out of the brain, instead migrating only a short distance along adjacent white matter tracts, the second scenario appears less likely.[567] Conversely, the expression of a variety of adhesion molecules and chemokines/chemokine receptors has been reported for tumour cells of PCNSL, as well as for tumour infiltrating vascular endothelial cells, astrocytes, microglia and activated T cells, and could facilitate the transient or continued entry of malignant B cells into the CNS.[575,576] We have used intraclonal variant analysis to assess the behaviour of tumour cells at the different sites. In general, intraclonal heterogeneity was highest in the periphery and showed evidence for separate clonal development in the blood and BM, possibly arguing for a greater influence of Ag at systemic sites. For one patient (AJAC), a tumour-related subclone was detected in the blood and BM, but which was not observed in the primary tumour in the brain. Its presence suggests diversification of the tumour-clone in systemic sites, but not in the CNS, with free interchange between blood and BM. This was confirmed for a second patient (TREA), where a tumour-related subclone was detected in the blood but not in the primary tumour; the subclone could not be detected in the BM,

which may suggest that, in this case, there is no movement between systemic compartments. These findings support the scenario that PCNSL undergoes at least some SHM extracerebrally and that a proportion of tumour cells in the periphery may have at some stage migrated to the brain, with the transformation event occurring either in the periphery or after entry into the CNS. Furthermore, the finding of clonal variants at peripheral sites that cannot be detected in the CNS would suggest that migration is transient, with separate clonal development following at the two sites.

Therefore, PCNSL appears to have both CNS-specific and systemic components at diagnosis and our observations raise the possibility that PCNSL should be considered a systemic disease at the outset. However, significant malignant growth does not occur at peripheral sites, and the limited diversity of the CNS tumour indicates that there is little or no exchange with peripheral tumour cells. It is possible that the final transforming event has occurred in a member of the clone that has entered the CNS. Alternatively, if transformation were to have occurred prior to entry into the CNS, some Ag-independent CNS-specific factor(s) must be required for the growth and establishment of secondary lesions in the periphery. It has been reported in PCNSL that the transcript levels of the BAFF receptors BAFF-R and BCMA greatly exceeds that in normal brain, reaching the same level as in lymphatic tissues.[581] Similar findings were been observed in the lesions of MS brains. Furthermore, in the CNS, BAFF is constitutively present and is produced by astrocytes.[581] Therefore, one possible scenario is that BAFF released by resident astrocytes supports the survival of BAFF-receptor-expressing malignant B cells in PCNSL and that this local supply of BAFF together with the shielding of the brain microenvironment from the immune system might provide a favourable basis for tumour development. Whatever the pathway of events, the malignant tumour appears to be locked into the CNS, leaving the more indolent clonal members in the periphery. Treatment regimens for PCNSL traditionally consisted of WBRT alone, but now universally incorporate the systemic delivery of high-dose methotrexate. We assume that by targeting tumour cells in the periphery, migration to the CNS could be reduced, leading to an improved outcome for patients. Indeed, survival rates have increased to ~5 years in patients receiving methotrexate.[545] In our study, 2 of the 3 patients, who all received methotrexate, remain in complete remission 37 and 41 months post-diagnosis; the third died 8 months post treatment of unrelated causes. However, Jahnke *et al.* have demonstrated that tumour cells can still be detected in the peripheral blood and BM of one patient 24 months post diagnosis and following treatment with high-dose methotrexate and WBRT.[580] This finding indicates that methotrexate is insufficient to eradicate peripheral tumour cells,

which may indicate a low proliferative rate of these cells since this agent targets tumour cells in a cell cycle dependent manner. Conversely, proliferation of tumour cells in the CNS is high (Ki-67 of 60-90% in our study) and therefore, highly susceptible to chemotherapy.[401,550] For FL and extracerebral DLBCL, positive tumour-specific PCR in the peripheral blood is considered a reliable predictor for relapse.[582] However, the patient described by Jahnke *et al.* remained in complete remission, highlighting the quiescent nature of these systemic tumour cells.

5.4. Conclusions

Overall, these findings confirm a biased usage of V4-34 gene in PCNSL and give strength to the hypothesis that sustained antigenic stimulation by a SAg in the CNS may drive proliferation of the malignant B cells. Our finding of continuing somatic mutation and isotype switch variants indicates that engagement of the BCR occurs post transformation. We also demonstrate that tumour cells are present in the periphery of PCNSL patients at diagnosis and assume that this is a common phenomenon. Furthermore, we show the existence of tumour-related subclones unique to extracerebral sites and hypothesise that a proportion of tumour cells in the periphery may have migrated to the CNS, with transformation occurring either in the periphery or after entry into the CNS. The more malignant behaviour of the tumour cells in the CNS suggests either a local environmental influence or a less malignant phenotype of the peripheral clone. These observations indicate that PCNSL should be considered a systemic disease at the outset and may have important implications for the understanding of PCNSL pathogenesis and treatment.

6. Concluding remarks

In the first part of this study we have investigated the acquisition of *N*-glycosylation sites in the Ig V genes of FL. A high incidence if *N*-glycosylation sites in the V_H genes of FL compared to normal B cells was first reported by colleagues in this laboratory.[12] It followed that this high frequency of sites could also be found in other GC-associated malignancies, such as eBL.[13] In this study, we have assessed the introduction of *N*-glycosylation sites in both the V_H and V_L genes of FL. We have demonstrated that, upon examination of the BCR as a whole, *N*-glycosylation sites are almost universally present in FL, and that this positive selection suggests an important role for these sites in tumour pathogenesis. We have further explored the nature of the added glycans using FL-derived recombinant scFv proteins expressed in a mammalian 293F expression system. We find that the glycans displayed on recombinant scFv proteins are unusual in terminating at oligomannose and that these high-mannose glycans can bind specifically to recombinant MR. We conclude that the presence of oligomannosylated Ig as it is displayed on the surface of FL cells may mediate binding to specific lectins present in the GC microenvironment, and that such an interaction may provide ongoing stimulation, replacing the need for persistent Ag, and lead to the growth and/or survival of these tumours within the GC. Ultimately, are goal will be to identify the carbohydrate binding lectin responsible for this interaction and, further, to ascertain the cell type that expresses this lectin in the GC microenvironment. Given the central role of FDCs in promoting B-cell survival and growth in GCs, FDCs represent one possible candidate for the cell that promotes survival and proliferation of B cells in FL. However, lymphoma-associated macrophages may also provide these important signals and increased numbers of these cells has been shown to denote a worse prognosis in FL.[519] It may be possible to exploit this seemingly essential feature to develop novel therapeutic modalities; the disruption of this interaction would be one such approach to treating the disease.

We have already established that oligomannose glycans, as they are displayed on FL-derived recombinant scFv proteins, can bind to recombinant MR, which is known to be expressed on macrophages within the GC.[517,518] We wish to further investigate binding to other carbohydrate binding lectins, such as DC-specific ICAM3 grabbing non-integrin (DC-SIGN) and DC-associated lectin-2 (Dectin-2). Both DC-SIGN and Dectin-2 bind to *N*-linked oligomannose with specific recognition of the structure Man9GlcNAc2 and are found on myeloid cells of the macrophage and DC lineages, as such they represent ideal candidates.[583,584] Interaction of our oligomannosylated-scFv proteins with DC-SIGN and

Dectin-2 will be assessed *in vitro* by ELISA using recombinant DC-SIGN and Dectin-2 proteins provided by Dr. L. Martinez-Pomares. Furthermore, we will investigate the ability of mannan, fucose and galactose to block this interaction. Lectin binding studies can also be used to confirm the presence of oligomannosylated sIg on primary tumour cells and FL cell lines. Recombinant MR, DC-SIGN and/or Dectin-2 proteins can be used *ex vivo* to assess binding to tumour cells, using either primary tumour cell suspensions or cell lines, by FACS analysis. Alternatively, they can be used to probe frozen/paraffin-embedded FL tissue sections by immunohistochemistry. Similarly, oligomannosylated-scFv proteins, which have been biotinylated, can be used to probe frozen/paraffin-embedded tissue sections of normal reactive lymph nodes to establish the expression of lectins specific for mannose in the GC microenvironment. All the work described above will be performed by Dr. V. Coelho. With the help of the Leukaemia Research Fund Monoclonal Antibody Facility, we are currently trying to raise a monoclonal Ab against the oligomannose glycans. If successful, this will provide a tool that can be used to recognise human FL cells in tissue sections and to probe the interaction of the mannosylated Ig with receptors.

Following on from our work in FL, we next investigated the acquisition of *N*-glycosylation sites in the Ig V genes of *de novo* DLBCL. DLBCLs represent a diverse group of lymphoid neoplasms, which are highly heterogeneous both clinically and morphologically. Recently, gene expression profiling has enabled the sub-classification of DLBCLs in to two main subgroups: GCB-like DLBCL and ABC-like DLBCL.[403,405] The former subgroup can be further divided based on the presence or absence of the t(14;18) translocation.[405,534,535] Since FL is also associated with the t(14;18) translocation and a GC B cell origin, we wished to investigate if t(14;18)-positive GCB-like DLBCL and FL share a common pathway of lymphomagenesis, starting with the t(14;18) translocation and followed by the acquisition of *N*-glycosylation sites in the GC. The frequency of *N*-glycosylation sites in the V_H genes of *de novo* DLBCL has been reported to be ~40%, although it is not known which subgroup the sites may be associating with, if any.[12] We demonstrate that just 50.0% of t(14;18)-positive GCB-like DLBCLs acquire an *N*-glycosylation site. Furthermore, we also show that an equal proportion of ABC-like tumours acquire sites. Therefore, our data suggests that, unlike FL, the acquisition of *N*-glycosylation sites is not important for the growth and/or survival of t(14;18)-positive GCB-like DLBCL. These findings are in stark contrast to those of Jardin *et al.* who report strong evidence for a common cell of origin linking *de novo* t(14;18)-positive GCB-like DLBCL and FL.[539] The reason for the disparity between data sets is unclear and would

suggest that our study might benefit from the assessment of further cases of *de novo* DLBCL to reinforce our findings.

In the final part of this study, we have investigated the presence of tumour cells at peripheral sites in patients with PCNSL. PCNSLs represent highly malignant DLBCLs, which arise within and are confined to the CNS in the absence of systemic disease. Although molecular analysis of the Ig V genes has suggested that these tumours are related to GC B cells, there is yet no consensus as to the site of origin of the malignant B cell that gives rise to PCNSLs.[400-402,550] One theory is that transformation and subsequent clonal proliferation of an activated B cell trafficking normally through the CNS gives rise to the malignant clone. Alternatively, it has been suggested that the tumour B cells undergo proliferation and neoplastic transformation outside the CNS within the GC microenvironment of the secondary lymphoid organs and then selectively 'home' to the CNS. Indeed, a unique expression of cellular adhesion molecules and chemokines and their receptors has been describe for these tumours.[572-576] We have investigated the existence of tumour cells outside the CNS in PCNSL patients at the time of diagnosis. Using the unique tumour-derived Ig V gene sequences as a marker, we demonstrate the presence of tumour cells in the BM and/or blood in 3 of 3 patients. Furthermore, we show the existence of tumour-related subclones that are unique to extracerebral sites and hypothesise that PCNSL undergoes at least some SHM extracerebrally and that a proportion of tumour cells in the periphery may have at some stage migrated to the brain, with the transformation event occurring either in the periphery or after entry into the CNS. The more malignant behaviour of the tumour cells in the CNS suggests either a local environmental influence or a less malignant phenotype of the peripheral clone. This study warrants the confirmation of subclinical disease in a larger cohort of patients, although a study by Jahnke *et al.* has shown similar findings.[580] Furthermore, a longer prospective evaluation of patients is necessary to determine the clinical significance of these seemingly quiescent tumour cells. We aim to correlate our extended findings with clinical information, such as disease progression and outcome, as well as investigate the effect of treatment on longevity of tumour cells residing in the periphery. Our findings have important implications for the treatment of PCNSL.

Appendices

Appendix I Buffers and reagents

Complete RPMI

500mL RPMI 1640, 25mM HEPES, with L-glutamine supplemented with:

Sodium pyruvate	100mM
Sodium Citrate	10mM
Penicillin	10,000U/mL
Streptomycin	10,000U/mL
L-glutamine	29.2mg/mL

Store at RT.

Freezing medium

human AB serum*	5.0mL
complete RPMI	4.0mL
DMSO	1.0mL

Make up fresh; discard any unused medium.
*batch tested and decomplemented (56°C for 1 hour).

50X TAE Buffer

Tris	242.0g
Acetic acid	57.1mL
0.5M EDTA	100.0mL

Make up to a volume of 1L with dH_2O

1X TAE buffer was made by dilution of 50X TAE buffer at a ratio of 1:50 with dH_2O.
Store at RT.

5X Sequencing Buffer, pH 9.0

1M Tris HCl, pH 8.0	400μL, final concentration 400mM
1M $MgCl_2$	10μL, final concentration10mM
dH_2O	590μL

Store at -20°C.

Luria Bertani Broth (LB) (Sigma; L3152)

25g powder required to prepare 1L with dH_2O:

NaCl	5.0g/L
Bacto-tryptone	10.0g/L
Bacto-yeast extract	5.0g/L

Autoclave, cool and store at 4°C

Luria Bertani Agar (LB-Agar) (Sigma; L3027)

40g powder required to prepare 1L with dH_2O:

NaCl	5.0g/L
Bacto-tryptone	10.0g/L
Bacto-yeast extract	5.0g/L
Agar	15.0g/L

Autoclave, cool and store at 4°C

Ampicillin, 100µg/mL (Amp100)

Ampicillin	100µg
dH_2O	1.0mL

Use at 1:1000 with LB and LB-Agar. Store at -20°C.

Ab Coupling buffer, pH 8.0

$NaHCO_3$	8.4g
NaCl	29.2g

Make up to a volume of 1L with dH_2O

Store at RT.

2X Tris-NaCl Wash Buffer, pH 8.0

NaCl	58.0g
Tris	24.3g
Na_2EDTA	3.5g
5M HCl	25.0mL

Make up to a volume of 1L with endotoxin-free H_2O

1X Tris/NaCl buffer was made by dilution of 2X Tris/NaCl buffer at a ratio of 1:2 with endotoxin-free H_2O. Store at RT.

Ammonium Thiocyanate Elution Buffer, pH 12.0

KSCN	97.2g
Ammonia (880)	40.0mL

Make up to a volume of 1L with endotoxin-free H_2O

Store at RT.

Phosphate Buffered Saline (PBS), pH 7.0

NaCl	7.0g
NA_2HPO_4	3.4g
KH_2PO_4	0.8g

Make up to a volume of 1L with (endotoxin-free/) dH_2O

Store at RT.

Ca^{2+} Tris Buffered Saline (TBS), pH 7.5

NaCl	9.0g
$CaCl_2$	1.1g
Tris-HCl	1.2g

Make up to a volume of 1L with dH_2O

Store at RT.

PBS/TBS-Tween20 (0.001%)

PBS/TBS	1.0L
Tween20	1.0mL

Store at RT.

ELISA Coating Buffer, pH 9.5

Na_2CO_3 (anhydrous)	1.6g
$NaHCO_3$	2.9g

Make up to a volume of 1L with dH_2O

Store at 4°C.

ELISA Substrate Buffer, pH 5.5

Citric acid (anhydrous)	4.7g
Na_2HPO_4	7.3g
Make up to a volume of 1L with dH_2O	

Store at 4°C.

Immediately prior to use, add the following to 25mLs of substrate buffer:

H_2O_2	25µL
o-Phenylenediamine (OPD)	1 tablet

AP-buffer, pH 9.5

$MgCl_2 \cdot 6H_2O$	0.4g
NaCl	5.8g
Tris-HCl	12.1g
Make up to a volume of 1L with dH_2O	

Store at 4°C.

Lectin ELISA Substrate Buffer

Immediately prior to use, add the following to 15mLs of AP-buffer:

p-nitrophenyl phosphate (1mg/mL)	1 tablet

20X SSC buffer, pH 5.3

NaCl	175.3g
Na citrate	88.2g
Make up to a volume of 1L with dH_2O	

Store at RT.

2X SSC buffer was made by dilution of 20X SSC buffer at a ratio of 1:10 with dH_2O. Store at RT.

FISH Wash Buffer 1, pH 7.5

20X SSC	20mL
NP-40	3mL
Make up to a volume of 1L with dH_2O	

Store at RT.

FISH Wash Buffer 2, pH 7.5

20X SSC	100mL
NP-40	1mL
Make up to a volume of 1L with dH_2O	

Store at RT.

Appendix II Calculation of iFISH probe false-positive cut off values

Vysis LSI® probe	Normal tonsil 1		Normal tonsil 2		Mean	SD	3X SD	Cut-off
	KM	AS	KM	AS				
IGH	3	4	4	2	3.25	0.957	2.871	6.1
BCL2	1	0	1	0	0.5	0.577	1.731	2.2
BCL6	1	0	0	1	0.5	0.577	1.731	2.2
IGH/BCL2	2	2	1	0	1.25	0.957	2.871	4.1

The signal pattern of 100 qualifying interphase nuclei from different areas of the section was assessed by two independent analysts (KM and AS); values are expressed as a percentage positive cells.

Appendix III Publications

McCann, K.J., Sahota, S., Stevenson, F.K. and Ottensmeier, C.H. Idiotype gene rescue in follicular lymphoma. In: Illidge T, Johnson PW, eds. Lymphoma - Methods and Protocols. Totowa: Humana Press; 2005:145-172.

McCann, K.J., Johnson, P.W.M., Stevenson, F.K. and Ottensmeier, C.H. Universal *N*-glycosylation sites introduced into the B-cell receptor of follicular lymphoma by somatic mutation: A second tumourigenic event? Leukemia. 2006 Mar;20(3):530-534.

McCann, K.J., Ottensmeier, C.H., Callard, A., Radcliffe, C.M., Harvey, D.J., Dwek, R.A., Rudd, P.M., Sutton, B.J., Hobby, P. and Stevenson, F.K. Remarkable selective glycosylation of the variable region in follicular lymphoma. Molecular Immunology. 2008 Mar; 45(6):1567-1572

Jbb: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

7

Idiotype Gene Rescue in Follicular Lymphoma

Katy McCann, Surinder Sahota, Freda Stevenson, and Christian Ottensmeier

Summary

Beyond the morphological, immunophenotypic, and genetic information used for the diagnosis of lymphoid malignancies, molecular analyses have deepened our insights into the development of B cell lymphomas. We have learned that B cell tumors can be grouped according to the mutational status of their immunoglobulin variable (V_H) region genes, and this has become an important prognostic tool in chronic lymphocytic leukemia. The analysis of V genes also has allowed us to more precisely place B cell lymphomas relative to their normal B cell counterparts and to the germinal center where somatic hypermutation takes place. It has become evident that many of the common B cell tumors arise at this site and are able to respond to stimuli, which govern normal B cells. In this chapter, we focus on the analysis of V genes in follicular lymphomas based on the experience in our laboratory and provide a detailed guide for this analysis.

Key Words: Follicular lymphoma; non-Hodgkin's lymphoma; immunoglobulin; V-gene analysis; immunogenetics; PCR; sequencing; cloning.

Au: Please cite Fig. 1 in the text.

1. Introduction

Follicular lymphomas (FLs) account for approx 30–40% of all newly diagnosed cases of non-Hodgkin's lymphoma (NHL) *(1)* and display a wide spectrum of clinical presentations. Although rare in patients younger than the age of 25, the frequency of these lymphomas increases steadily with age, with a median presentation at 50–60 years. They occur as frequently in men as in women, in contrast to many other NHLs, and their incidence is increasing. FL typically follows an indolent course, with waxing and waning of the disease. In most cases, standard treatments cannot secure a prolonged remission or cure, and the median survival is 7–8 years *(2–4)*.

From: *Methods in Molecular Medicine, Vol. 115: Lymphoma: Methods and Protocols*
Edited by: T. Illidge and P. W. M. Johnson © Humana Press Inc., Totowa, NJ

Jbb: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

1.1. Morphology

FL is defined as a tumor in which the normal architecture of the lymph node has been destroyed and is replaced by malignant cells, which grow, at least in part, in patterns reminiscent of lymphoid follicles. The tumor cell population consists of variable mixtures of cells, although centrocytes predominate and, by definition, centroblasts form only a minority in the population.

FL is believed to be of a germinal center (GC) origin with the $CD5^-$ $CD23^-$ peripheral B cell of the inner follicle mantle being the presumed normal counterpart. More than 90% of FLs express surface immunoglobulins (Ig)—in the absence of cytoplasmic Ig—and typically are surface IgM^+ but also are $sIgD^+$, $sIgG^+$, or $sIgA^+$, albeit less frequently *(5–8)* Characteristically, tumor cells are CD5 and CD43 negative, and CD10 and CD23 are variably expressed *(9)*. Genetic analysis reveals a t(14;18)(q32;q21) translocation in the most cases *(10–12)*.

1.2 Immunoglobulin Gene Rearrangement

Ig gene rearrangement is the defining molecular event in the development of a B cell. In the early stages of B cell development, the Ig heavy chain locus on chromosome 14 brings together a variable (V_H), a diversity (D), and a joining segment (J_H) in a process of somatic deoxyribonucleic acid (DNA) rearrangement known as V(D)J recombination. If successful, rearrangement of the light chain at the kappa locus on chromosome 2 and the lambda locus on chromosome 22 follows, respectively, until a functional light chain terminates the recombination process. Subsequent somatic hypermutation (SHM) and class switch recombination provide mechanisms by which Ig is further diversified and constitute central events in the life of a B cell tightly associated with developmental stages. Scrutiny of the evidence these processes leave in the Ig gene sequence reveals the stage of development that the cells have reached at a particular time point compared with normal B cells.

With the exception of Hodgkin's disease, B cell malignancies typically have functional Ig genes, which may reflect the requirement for Ig expression in normal B cells *(13)*. As is the case with normal B cells, SHM and class switch recombination events will leave a lasting imprint in the genome, which will be present in each cell of a tumor clone. Therefore, V genes can serve as specific markers for B cell lymphomas. With the development of modern molecular techniques, this genetic information has become accessible and now analysis of Ig genes can provide details of the molecular events that a malignant B cell has been exposed to, revealing information about the clonal history of the neoplastic cell *(14,15)*.

V gene analysis also provides an independent marker for the presence of a clonal population. This markers permits one the ability to distinguish neo-

plasms from reactive processes and, therefore, it contributes to the interpretation of clinical and histological findings in difficult situations.

Genetic analysis also may reveal information related to pathogenesis. Preferential usage of individual V_H gene segments suggest the role of classical antigens (Ag) or superantigens (SAg) in the initial proliferation of B cells, leading to overt neoplasia, because it has been shown that SAgs can bind to the (unmutated) framework (FR) regions of an antibody molecule *(16–18)*. Because the human V gene repertoire is known, it is now possible to confidently identify somatic mutations in individual B cells, which in turn allows assessment of whether a tumor cell, or its parent, has undergone SHM and affinity maturation. In tumors thought to be of GC origin, heterogeneity of the somatic mutations in the V genes within the tumor cell population indicates that SHM is still ongoing. Interestingly, in some cases, a narrowing of this heterogeneity has been found after treatment, implying selection of treatment resistant cells *(14,19–21)*. More recently, isotype-switching events have been analyzed and reveal the presence of multiple isotypes in some tumors *(22–24)*. The presence of these isotypes provides more evidence that the tumor cells can to some degree respond to stimuli that govern the life of the normal cellular counterparts.

The findings of these studies contribute to the understanding of differentiation and development of malignant B cells and to the classification of these malignancies relative to normal B cell development *(9,14,15,25)*. Within some disease entities individual subgroups exist, which might arise at different time points in B cell development *(26,27)*.

1.3. Tumors of the Germinal Center

Most B cell lymphomas, that is, FL, diffuse large B cell lymphomas (DLBCL), MALT lymphomas, and Burkitt lymphoma, arise from cells that have undergone SHM. In combination with the morphological and immunophenotypic evidence, this progress is suggestive of a GC origin.

Normal cells undergo SHM of their V_H and V_L genes and may additionally class switch. They then leave the GC *(28)*. Some cells re-enter the GC dark zone for further rounds of mutation, or alternatively become plasma cells. Although V gene analysis shows that many B cell tumors have undergone, and in some cases can continue to undergo SHM, it is possible that prolonged cell survival and residence in the GC may blur the information that V gene analysis generates. Factors that are known to increase cell survival, such as deregulation of bcl-2, are in fact observed in the vast majority of FL with a t(14;18), and p53 mutations commonly are observed. Although the initial t(14;18) rearrangement likely occurs in the bone marrow *(29)*, it may exert its influence only after the tumor cell has arrived in the GC and has additionally undergone some yet undefined additional transforming event.

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

In mouse models, it has been shown that the RAG1 and RAG2 genes can be reactivated in the GC, opening a new window for susceptibility to (abnormal) rearrangements (*30*; reviewed in **ref. *31***). It appears likely that such mechanisms can be active in humans too, because the salvage of cells from death by receptor editing in the GC requires reactivation of the recombinase machinery *(32–34)*. An open question remains whether lymphomas, like normal B cells, need the presence of Ag or of SAgs for ongoing somatic hypermutation.

1.4. V Gene Analysis in FL

Several studies have addressed V gene usage in FL and have reported a high degree of somatic mutations in the V genes of these lymphomas *(20,21,35–40)*. V_H gene usage appears to be similar to that of the normal B cell repertoire, with the possibility of a small excess of V_H4 genes. A well-documented feature is the presence of intraclonal heterogeneity, implying that the cells of this lymphoma remain under the influence of the somatic mutator in the GC, and it often is possible to deduce genealogical trees depicting the relationship between individual cell clones *(41)*. In these lymphomas, the presence of antigen-presenting cells in the form of follicular dendritic cells is well recognized *(8)* and T cells are invariably present. The presence of the third key component antigen, which drives somatic mutation in normal B cells, has not been functionally demonstrated in FL. Nonetheless, in some cases, the evaluation of the patterns of somatic mutations has shown preferential sequence maintenance in FRs vs clustering of mutations in the complementarity-determining regions (CDRs) and this has been interpreted as evidence for antigen-driven clonal selection *(19,21,38,42,43)*. Model systems are being established in which somatic mutations occur without external antigen *(44)*, and it may be that for lymphomas arising in the GC, antigen is no longer an absolute requirement for SHM.

Most studies have assessed V_H genes only. However, light chain gene analysis has been reported in two studies *(21,38)*. The latter study investigated light chain V_κ genes and V_H genes in 10 FL cases and found that the V_κ genes fell into two groups: one that demonstrated extensively mutated V_κ genes (6/10 cases) and a second with a very low degree of V_κ mutations (4/10 cases). To date this observation has not been confirmed and its relevance remains to be assessed.

More recently, it has become appreciated that disease progression or treatment may have an impact on V_H gene sequences, and loss of intraclonal heterogeneity has been observed *(20,21)*.

As part of our clinical trial of anti-idiotypic DNA vaccination against FL, we have sequenced 36 cases, identifying both the V_H and V_L sequences from the tumor clone. All cases demonstrated mutated V genes, with higher levels of

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

SHM in V_H genes as compared with V_L. We could consistently identify J_H segments. However, if the stringent criteria of Corbett et al. ***(45)*** are applied, which require at least 10 nucleotides identity to a germline D segment,the identification of D region segments was ambiguous in most cases. ,

To measure the immune responses of the patients in our study, it was necessary to express idiotypic protein for use in immunological assays ***(40)***. Analysis of the recombinant proteins showed that they had undergone glycosylation. This observation prompted the systematic analysis of the V_H and V_L sequences for potential N-linked glycosylation sites ***(40)***. The motif for N-glycosylation is AsnXSer/Thr, for which X can be any amino acid except Pro, Glu, or Asp. Natural sites occur in three V_H germline segments (V1-08, V4-34, and VH5a) and in three V_L germline segments (VKV21(1), VL514115e, and 1173e). Sequence analysis of our own data show that 34/36 FL acquired one or more new glycosylation sites in either V_H, V_L, or both. In contrast, normal B cell V gene rearrangements, nonfunctional V gene sequences, and sequences from multiple myeloma and the mutated subset of chronic lymphocytic leukemia have a low incidence of approx 10% glycosylation sites ***(40)***. Other germinal center lymphomas, such as DLBCL or sporadic Burkitt lymphoma, show an intermediate frequency of approx 40%; 90% of sites were located in the CDRs, particularly CDR2; FR2 and FR4 had no sites. Novel sites are limited to a few positions in the CDRs and focussed on codon 33-35 in CDR1 (14/14) and codon 50 at the N-terminus of CDR2 (20/36). In addition, replacement Asn residues commonly were acquired at or near the N-terminus of the CDR3 sequence. These may be derived from N-addition or from D-segment genes. Naturally occurring sites showed no preference as to retention or loss.

Together, these data suggest that the V gene sequences in FL are unusual in the acquisition of glycosylation sites and that these sites may be functional. The high incidence in these tumors suggests a positive selection and that glycosylation of V genes in these tumors may play an important role in the pathogenesis of FL and it is possible that they functionally take on the role of the "missing antigen" in the germinal center.

1.5. Technical Aspects of V-Gene Analysis From FL Cells

Ribonucleic acid (RNA) is extracted from tumor cells and followed by complimentary DNA (cDNA) synthesis using oligo (dT) primers. For this purpose, single-cell suspensions or frozen tissue can be used. For the analysis of heavy chain V genes, initial PCR reactions use individual family specific leader primers (1-6), and a mix thereof, together with an appropriate constant region primer (*see* **Table 1**). If dealing with a tumor of unknown Ig isotype, it is judicious to begin investigation with a FR2 consensus primer to constant regions Cμ, Cγ, and Cα. The outcome of these initial PCRs will indicate which heavy chain

T1

Table 1
Primers for V_H and V_L PCR

Primer Name	Sequence
V_H primers	
V_H Leader 1/7	CTC ACC ATG GAC TGG ACC TGG AG
V_H Leader 2	ATG GAC ATA CTT TGT TCC AGG CTC
V_H Leader 3	CCA TGG AGT TTG GGC TGA GCT GG
V_H Leader 4	ACA TGA AAC AYC TGT GGT TCT TCC
V_H Leader 5	AGT GGG TCA ACC GCC ATC CTC G
V_H Leader 6	ATG TCT GTC TCC TTC CTC ATC TTC
V_H 1/7 FR1 consensus	CAG GTG CAG CTG GTG CAR YCT G
V_H 2 FR1 consensus	CAG RTC ACC TTG AAG GAG TCT G
V_H 3 FR1 consensus	GAG GTG CAG CTG GTG SAG TCY G
V_H 4a FR1 consensus	CAG STG CAG CTG CAG GAG TCS G
V_H 4b FR1 consensus	CAG GTG CAG CTA CAR CAG TGG G
V_H 5 FR1 consensus	GAG GTG CAG CTG KTG CAG TCT G
V_H 6 FR1 consensus	CAG GTA CAG CTG CAG CAG TCA G
FR2 a consensus	TGG RTC CGM CAG SCY YCN GG
FR2 b consensus	GTC CTG CAG GGY YCC GGR AAR RGT CTG GAG TGG
FR3 consensus	ACA CGG CYG TRT ATT ACT GT
Constant μ	GGA ATT CTC ACA GGA GAC GAG G
Constant γ	CTG AGT TCC ACG ACA CCG TCA
Constant α	ATC TGG CTG GGT GCT GCA GAG GCT
J_Ha consensus	ACC TGA GGA GAC GGT GAC C
J_Hb consensus	GTG ACC AGG GTN CCT TGG CCC CAG
J_Hc consensus	TGA GGA GAC GGT GAC CAG GAT CCC TTG GCC CCA G
J_H1	CAG GGC ACC CTG GTC ACC GTC TCC TCA
J_H2	CGT GGC ACC CTG GTC ACT GTC TCC TCA
J_H3	CAA GGG ACA ATG GTC ACC GTC TCT TCA
JH4/5	CAG GGA ACC CTG GTC ACC GTC TCC TCA
JH6	CAA GGG ACC ACG GTC ACC GTC TCC TCA
$V_κ$ primers	
Leader 1 consensus	ATR GAC ATG AGR GTS CYY GCT CAG CKC
Leader 2 consensus	ATG AGG CTC CYT GCT CAG CTY CTG GGG
Leader 3 consensus	ATG GAA ACC CCA GCG CAG CTT CTC TTC
Leader 4	ATG GTG TTG CAG ACC CAG GTC TTC ATT
Leader 5	ATG GGG TCC CAG GTT CAC CTC CTC AGC
Leader 6a	ATG TTG CCA TCA CAA CTC ATT GGG TTT
Leader 6b	ATG GTG TCC CCG TTG CAA TTC CTG CGG
Constant κ27	CAA CTG CTC ATC AGA TGG CGG GAA
Constant κ69	AGT TAT TCA GCA GGC ACA CAA C

(continued)

.bb: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

Table 1 (Continued)

Primer Name	Sequence
V_λ primers	
Leader 1 consensus	ATG RCC DGS TYY CCT CTC YTC CTC
Leader 2 consensus	ATG GCC TGG GCT CTG CTS CTC CTC
Leader 3 consensus	ATG GCM TGG RYC VYW CTM YKB CTS
Leader 4a	ATG GCC TGG ACC CAA CTC CTC CTC
Leader 4b	ATG GCT TGG ACC CCA CTC CTC CTC
Leader 4c	ATG GCC TGG GTC TCC TTC TAC
Leader 5 consensus	AGT GCC TGG ACT CYT CTY CTY CTC
Leader 6	ATG GCC TGG GCT CCA CTA CTT CTC
Leader 7	ATG GCC TGG ACT CCT CTC TTT CTG
Leader 8	ATG GCC TGG ATG ATG CTT CTC CTC
Leader 9	ATG GCC TGG GCT CCT CTG CTC CTG
Leader 10	ATG CCC TGG GCT CTG CTC CTC CTG
V_λ FR1-1a	CAG TGT GTG CTG ACT CAG
V_λ FR1-1b	CAG TCT GTG YTG ACG CAG
V_λ FR1-1c	CAG TCT GTC GTG ACG CAG
V_λ FR1-2	CAG TCT GCC CTG ACT CAG
V_λ FR1-3a	TCC TAT GWG CTG ACT CAG
V_λ FR1-3b	TCC TAT GAG CTG ACA CAG
V_λ FR1-3c	TCT TCT GAG CTG ACT CAG
V_λ FR1-3d	TCC TAT GAG CTG ATG CAG
V_λ FR1-4	CAG CYT GTG CTG ACT CAA
V_λ FR1-5	CAG SCT GTG CTG ACT CAG
V_λ FR1-6	AAT TTT ATG CTG ACT CAG
V_λ FR1-7	CAG TCT GTG GTG ACT CAG
V_λ FR1-8	CAG ACT GTG GTG ACC CAG
V_λ FR1-9	CWG CCT GTG CTG ACT CAG
V_λ FR1-10	CAG GCA GGG CTG ACT CAG
Constant λ33	GTT GGC TTG AAG CTC CTC AGA GGA
Constant λ85	CAC RGC TCC CGG GTA GAA GTC ACT
Cloning	
T7	TAA TAC GAC TCA CTA TAG GG
SP6	ATT TAG GTG ACA CTA TAG AA
β actin	
β actin F	TCA TGT TTG AGA CCT TCA A
β actin R	GTC TTT GCG GAT GTC CAC G

R = A+G; Y = C +T; K = G+T; S = C+G; N = A+C+G+T, M = A+C, W = A+T

Where primers have more than 1 consensus sequence, for example,. JH a, b and c, an equal mix is prepared.

constant region is used by the tumor clone and will focus subsequent efforts with individual leader primers. It also is possible to use a J_H consensus primer as the 3' primer, but this is more likely to fail due to the presence of mutations in J_H.

For the amplification of the light chain, PCRs using family specific leader primers (Vκ1-6 or Vλ1-10), either individually or as a mix, together with a constant region primer is usually sufficient to identify the tumor clone. In general, we find that the V_L gene of a tumor clone is identified more easily than the V_H gene. This is most likely caused by the intrinsically lower mutational rates found in these genes.

Ideally, we seek to obtain a single product produced from any particular primer pair, although multiple products occur. Amplified products are gel extracted, directly sequenced, and aligned to the V Base database. For those sequences that align cleanly to V and J segments, a proposed germ line can be constructed.

PCR products that are polyclonal, either across the entire sequence or over the CDR3 alone as identified through analysis of the electropherogram (EPG), are not pursued. Accordingly, and for cases in which the initial PCR fails, revised cycling conditions for the PCR with a touch-down protocol or using existing primer pairs or with different primer combinations are likely to give a PCR product. For V_H genes, "reserve primers" are designed to anneal to FR1, FR2 and FR3 regions and the individual J_H segments.

Potentially clonal sequences are always confirmed by a second independent PCR. After two PCR products with identical sequence have been identified, one or both PCR products are cloned and 12–24 colonies sequenced. 10 tumor derived sequences are sufficient to assess the intraclonal variation of the tumor clone.

Although this approach is successful in the majority of cases, there are instances when it is not possible to define the V_H and/or V_L clone. This may be either because there are not enough tumor cells in the population or perhaps because the mutation in the leader or framework regions is too extensive to permit binding of the primers. In such cases, it can be useful to employ gene scan of FR3-$J_H$consensus PCR products to confirm a clonal population is present. Polyclonality, as revealed by gene scan, makes it less likely that the clonal signature of a tumor sample will be determined and therefore gene-scanning can avoid extensive and unsuccessful PCR and cloning.

Very occasionally, cloning reveals more than one sequence, suggesting the presence of multiple clones. We have observed this phenomenon in the light chains of a small number of FL cases, which can be explained by dual light chain rearrangement or receptor editing. In these cases, we have increased our cloning to determine and select the dominant clone. However, we have not

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

experienced this when rescuing heavy chain variable regions in FL. We think this is related to the use of cDNA rather than genomic DNA, which avoids amplification of the second allelically excluded and nonfunctional rearranged V_H genes. In the single instance during which we observed two sequences in multiple clones identified from a single patient, it proved to be the result of contamination with material from a different patient and was resolved, but highlighted the potential risk of working with samples from multiple patients at one time. Theoretically, it is possible that a small minority of tumors express two V_H sequences in the same tumor cell population or that tumors may be truly biclonal. In other B cell malignancies, this has been reported, and we have observed this occasionally in DLBCL, but in germinal center type B cell lymphomas we would expect to find this in less than 5% of cases. In practical terms such a result is highly suspicious of contamination.

2. Materials

2.1. Cell Storage and Freezing

2.1.1. Preparing the Cell Suspension

All work should be conducted in a laminar flow tissue culture hood.

1. Patient biopsy.
2. Sterile tissue plates (10 cm in diameter).
3. Sterile scalpel.
4. Sterile tweezers.
5. Sterile metal tea strainer.
6. Sterile RMPI, store at 4°C.
7. Sterile 10- or 20-mL syringes.
8. 50-mL Falcon tubes.

2.1.2. Cell Counting

Additional items required are a hemocytometer, light microscope, and counter.

1. Trypan blue stain 0.4% (0.85% saline, Sigma).

2.1.3. Culture Media

RPMI 1640 Medium + 25 m*M* HEPES + L-glutamine (Gibco, Paisley, UK). Complete RPMI contents are listed below:

1. 500 mL of RPMI 1640 Medium + 25 m*M* HEPES + L-glutamine.
2. 5 mL of PSG, 100X (penicillin, 10,000 U/mL, streptomycin sulfate, 10,000U/mL; L-glutamine, 29.2mg/mL; Sodium citrate, 10 m*M*; Gibco)
3. 5 mL of sodium pyruvate, 100 m*M* (Gibco).

2.1.4. Freezing of Cells

Additional items required are 1.6-mL sterile cryovials and a cryo 1°C freezing container.

1. Freezing media: 50% human Ab serum, complement free, batch tested (Sigma); 40% complete RPMI (*see* **Subheading 2.1.3.**); and 10% dimethylsulfoxide (DMSO).
2. Isopropanol (warning: flammable).

2.2. Preparation of cDNA

2.2.1. RNA Extraction

TRI Reagent (Sigma). In addition to the items listed below, a tabletop temperature-controlled microcentrifuge, a vortex, and 1.6-mL polypropylene Eppendorf (RNase-free) tubes are required.

NOTE: Phenol and TRI reagent are toxic carcinogens and irritant. Meticulous handling and adherence to laboratory safety and good laboratory practice is essential; *see* also **Note 1.**

1. TRI Reagent™; store at 4°C.
2. Chloroform.
3. Isopropanol (warning: flammable).
4. Ethanol (warning: flammable).
5. RNase-free water (as a cheap and easy alternative we use standard water for injection, available in 5-mL or 10-mL aliquots from NHS hospital supplies).

2.2.2. cDNA Synthesis

Superscript™ First-Strand Synthesis system for RT-PCR (Invitrogen, Paisley, UK). In addition to the items listed below, a tabletop microcentrifuge, wet ice, a thermocycler and the appropriate thin-walled polypropylene tubes are also required.

1. Diethylpyrocarbonate-treated water/water for injection.
2. 10 mM dNTP mix (10 m*M* each of dATP, dCTP, dGTP, and dTTP).
3. Oligo$(dT)_{12-18}$, 0.5 µg/µL.
4. 10X RT buffer (200 m*M* Tris-HCl, pH 8.4, 500 n*M* KCl).
5. 25 m*M* $MgCl_2$.
6. 0.1 *M* dithiothreitol (DTT).
7. RNaseOUT™ Recombinant RNase Inhibitor (40 U/mL).
8. Superscript™ II Reverse Transcriptase (50 U/µL).
9. *Escherichia coli* RNase H (2 U/µL).
10. 1 µL of control GSP.

Jbb: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

11. 1–5 μg of total RNA from the sample.
12. 1 μL of control RNA.

2.3. PCR and Related Procedures

2.3.1. PCR

HotStartTaq™ (QIAGEN, Crawley, UK). The PCR is performed using a thermocycler with the corresponding thin-walled polypropylene tubes. Additional items required are wet ice, a laminar flow cabinet, and a tabletop microcentrifuge.

1. 10X PCR buffer (15 m*M* $MgCl_2$, Tris-HCl, KCl, $(NH_4)SO_4$).
2. HotStartTaq DNA Polymerase (2.5 U/μL).
3. 2.5 m*M* dNTP mix (2.5 m*M* each dATP, dCTP, dGTP, dTTP).
4. 1–5 μL template cDNA.
5. 1 μL 5' Primer(20 pmol/μL).
6. 1 μL 3' Primer(20 pmol/μL).
7. Diethylpyrocarbonate-treated water.

Primers are diluted to working concentration from stocks solutions and stored as aliquots at –20°C to minimize the possibility of contamination.

2.3.2. Agarose gel Electrophoresis

Items required are a gel tray and comb, a microwave, a gel tank, a power source and a UV light source, sterile scalpels, gloves, and face shield.

NOTE: Ethidium bromide is a carcinogen. Meticulous handling and adherence to laboratory safety and good laboratory practice is essential.

1. Agarose (Sigma).
2. Ethidium bromide, 10 mg/mL (Sigma).
3. 1X TAE buffer (40 m*M* Tris-acetate, 10 m*M* ethylenediamine tetraacetic acid).
4. 2μL of 1 kb + DNA ladder, 1 μg/μL (Invitrogen).

2.3.3. PCR Purification

Purification requires the use of the QIAquick Gel Extraction kit (QIAGEN). Other items required are 1.6-mL polypropylene Eppendorf tubes, a sterile scalpel, a precision scale, a heat block at 50°C, and a tabletop microcentrifuge.

1. Gel solubilization buffer QG.
2. Isopropanol (warning: flammable).
3. QIAquick spin column and 2-mL collection tube.
4. Ethanol (warning: flammable).
5. Wash buffer phycoerythrin (PE).
6. Elution buffer (EB).

Au: Does EB stand for elution buffer or ethidium bromide?

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

2.4. Sequencing

2.4.1. Sequencing Reaction

BigDye® Terminator v1.1 Cycle Sequencing Kit and ABI 377 automatic DNA sequencer (Applied Biosystems, Warrington, UK). The sequencing reaction is performed using a thermocycler with the corresponding thin-walled polypropylene tubes.

1. BigDye Terminator Ready Reaction Mix (A-dye terminator labeled with dichloro [R6G], C-dye terminator labeled with dichloro [R0X], G-dye terminator labeled with dichloro [R110], T-dye terminator labeled with dichloro [TAMRA], dNTP [dATP, dCTP, dITP, dUTP], and AmpliTaq DNA polymerase).
2. 5X sequencing buffer (400 m*M* Tris-HCl, 10 m*M* $MgCl_2$; pH 9.0).
3. 1 µL of 5' or 3' primer, 1.6 pmol.
4. 5 µL of template DNA (PCR product or purified plasmid).

Additional items required for precipitation of sequencing reactions after cycling include 1.6-mL polypropylene Eppendorf tubes, wet ice, and a temperature-controlled tabletop microcentrifuge.

1. 3 *M* sodium acetate (pH 4.2).
2. Ethanol, 100% and 75% (warning: flammable).
3. Loading buffer; Loading dye:formamide (1:5).

2.4.2. Sequencing Analysis

MacVector software (Oxford Molecular, Oxford, UK) and Entrez and V-Base databases (Center for Protein Engineering, MRC Center, Cambridge, UK; http://www.mrc-cpe.cam.ac.uk/vbase-ok) are required for analysis.

2.5. Cloning

2.5.1. Ligation

pGEM®-T Vector System I (Promega, Southampton, UK). Additional items required are 1.6-mL polypropylene Eppendorf tubes and a fridge at 4°C.

1. pGEM-T empty vector (50 ng/µL).
2. T4 DNA Ligase (3 Weiss U/µL).
3. 2X rapid ligation buffer (60 m*M* Tris-HCl, pH 7.8; 20 m*M* $MgCl_2$; 20 m*M* DTT; 2 m*M* ATP 10% polyethylene glycol).
4. 3 µL of purified PCR product.

Store 2X buffer in single use aliquots at –20°C. Avoid multiple freeze-thaw cycles.

2.5.2. Transformation

JM109 Competent cells (Promega). Additional items required are 20-mL universal tubes, 1.6-mL polypropylene Eppendorf tubes, wet ice, a waterbath

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

set at 42°C, an orbital shaker set at 37°C, a tabletop microcentrifuge, spreaders, and a warm room set at 37°C.

1. 50 µL of JM109 competent cells.
2. 5 µL of ligate.
3. 1 mL of LB.
4. LB/Agar Amp, 100 plates.

2.5.2.1. LB

Millers LB broth (Sigma) is used. Additional items required include glassware, weighing scales, a flea and stirring plate, and an autoclave.

1. 25 g/L LB.
2. Autoclave and cool.

LB Amp 100 contains: Penbritin® Ampicillin, 100 mg/mL; add at 1 µL/mL to cool LB.

2.5.2.2. LB/Agar

Millers LB broth (Sigma.) and Bacto®-Agar (BD Biosciences, Oxford, UK) should be used. Additional items required include glassware, weighing scales, a flea and stirring plate, and an autoclave.

1. 25g/L LB.
2. 15g/L Agar.
3. Autoclave and cool.

LB/Agar Amp 100 contains: Penbritin® ampicillin, 100 mg/mL; add at 1 µL/mL to LB/Agar cooled to less than 50°C.

2.5.2.3. LB/Agar Amp100 Plates

Additional items required include a microwave, sterile Petri dishes, and a spreader.

1. LB/Agar Amp100 plates.
2. X-Gal (5-bromo-4-chloro-3indolyl-b-D-galactoside), 50 mg/mL; use 20 µL/plate.
3. Isopropyl β-D-thiogalactopyranoside, 100 m*M*; use 100 µL/plate.

2.5.3. Plasmid Purification

QIAprep® Miniprep (QIAGEN) is needed, as well as 1.6-mL polypropylene Eppendorf tubes, a tabletop microcentrifuge, a vortex, and a QIAvac vacuum manifold (optional).

1. Resuspension buffer P1.
2. Lysis buffer P2.
3. Neutralization buffer N3.

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

4. QIAprep spin column and 2-mL collection tube.
5. Binding buffer PB.
6. Ethanol, flammable.
7. Wash buffer PE.
8. Elution buffer (10 m*M* Tris-HCl, pH 8.5).

3. Methods

3.1. Technical Comments

The immunophenotype of the tumor provides a useful starting point for analysis. Expression of IgM, IgD, IgG, and IgA is evaluated and can be used to determine which constant region primer is most likely to identify the tumor clone. Knowledge of kappa and lambda expression can identify the light chain used by the tumor clone and, thus, the appropriate primers can be selected. In conjunction with other B cell markers, including CD20 and CD19, immunophenotypic information can give an indication as to the percentage of tumor cells in the cell population. A low percentage can be a reason for failure to identify a tumor-derived clone.

Fig. 1

Au: please call out figure 1 in the text.

3.2. Cell Storage and Freezing

Diagnostic lymph node biopsies are received fresh on the day of surgery and processed on the same day. An aliquot of cells is used for immunophenotyping by fluorescence-activated cell sorting analysis for IgM, IgD, IgG, and IgA and kappa and lambda expression.

1. Using a sterile scalpel, lymph nodes are cut into small pieces and the cells dispersed by forcing through a fine sieve (e.g., a sterile metal tea strainer) into sterile RPMI medium.
2. Cells are collected, centrifuged, and washed once in RPMI.
3. After resuspension, cell viability is assessed with 0.4% trypan blue stain and aliquots of 1×10^7 cells/mL are resuspended in freezing medium.
4. Freezing follows a step-wise drop in temperature to –80°C using cryopreservation protocols.
5. Cells are transferred to Liquid N_2 after 24 h.

Alternatively, material from frozen tissue blocks can be used. 5- to 10-μm sections are cut using a cryostat and transferred directly into 1mL TRI Reagent with a sterile needle. Five to 20 sections per milliliter of TRI reagent should be sufficient to obtain enough RNA for analysis, but the number of sections needed ultimately is dependent on the size of the tissue block. Care should be taken to avoid thawing of the tissue block prior to, during and after cutting.

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

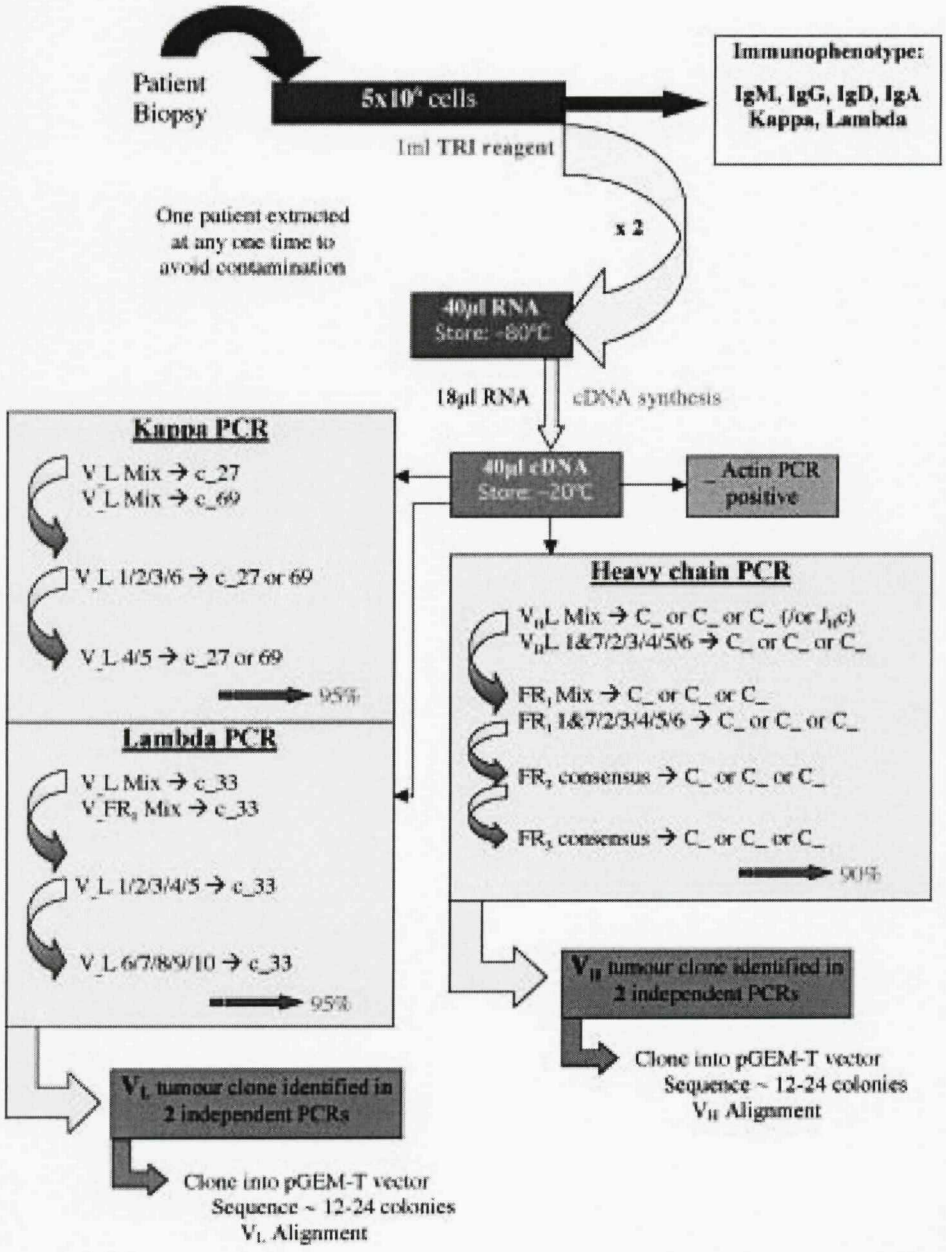


Fig. 1. Flowchart of V gene analysis.

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

3.3. Preparation of cDNA

3.3.1. RNA Extraction

RNA is extracted from 5×10^6 to 1×10^7 cells using 1 mL of TRI reagent (Sigma) and following the manufacturers instructions in a triphasic separation (*see* **Note 1** regarding RNA handling).

1. Rapidly thaw one aliquot of cells at 37°C and pellet by touch spin.
2. Resuspend cells in approx 100 µL of RPMI medium before adding 1 mL of TRI reagent; Mix thoroughly and allow the sample to stand for 5 min at room temperature.
3. Add 0.2 mL of chloroform, shake vigorously, and allow to stand for 15 min.
4. Centrifuge at 12000*g* for 15 min at 4°C.
5. Transfer the colorless upper aqueous phase containing the RNA into a fresh tube with 0.5 mL isopropanol and allow to stand for 5–10 min.
6. Centrifuge at 12000*g* for 10 min at 4°C.
7. Remove the supernatant and wash once with 75% ethanol.
8. After a final centrifugation step at 12,000*g* for 5 min, air dry the resulting RNA pellet for a few minutes and resuspend into an appropriate volume of RNase-free sterile water (approx 40–80 µL).
9. Quantitative analysis of the RNA is performed using a spectrophotometer (A_{260}) and the RNA is then stored at –80°C.

3.3.2. cDNA Synthesis

First-strand cDNA is synthesized from total RNA using the Superscript II system (Invitrogen) with an oligo-dT primer according to the manufacturer's instructions (*see* **Note 1**).

1. Up to 5 µg of total RNA is transferred to an RNase-free sterile tube, mixed with 2 µg of oligo-dT primer and incubated at 65°C for 5 min.
2. A mix of 10X RT buffer, 25 m*M* $MgCl_2$, 0.1 *M* DTT, and RNaseOUT-RNase inhibitor is added and incubated at 42°C for 2 min.
3. 50 units of Superscript II reverse transcriptase is then added and incubated for a further 50 min at 42°C.
4. The reaction is terminated by heating to 70°C for 15 min and placed on ice.
5. After a final addition of 1 µL of RNaseH and incubation at 37°C for 20 min, the synthesized cDNA can be immediately amplified or stored at –20°C.

3.4. Heavy-chain PCR

For more information *see* **Notes 2–6.**

3.4.1. V_H PCR - Set Up and Cycling Conditions

V_H genes are amplified from the cDNA by PCR with HotStart Taq (QIAGEN) using 5' and 3' oligonucleotide primers designed specifically (*see* **Subheading 3.4.2.**). It is advisable to check the quality of the cDNA before

Jbb: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

attempting V gene rescue using primers designed to amplify a standard housekeeping gene, such as β actin.

1. 1–5 μL of cDNA template.
2. 1 μL of 5' primer, 20 pmol.
3. 1μL of 3' primer, 20 pmol.
4. 10 μL of dNTPs, 2.5 m*M*.
5. 1 μL (2.5U) HotStart Taq DNA polymerase.
6. 10 μL of 10X PCR buffer.
7. Up to 50 μL of total volume with sterile water.

All PCRs are set up in a laminar flow cabinet to avoid contamination. Negative controls, in which the cDNA is absent from the reaction, are run in parallel for every primer pair. Thus contamination of any component of the PCR reaction with cDNA or DNA can be identified immediately (*see* also **Notes 2–6**).

"Touchdown" PCR cycling conditions for the amplification of V_H genes and β actin (with the exception of FRβ-constant region primer combinations; see below) consists of the following:

1. Initial denaturing step at 94°C for 15 min.
2. A single cycle of 94°C for 30 s, 60°C for 45 s, and 72°C for 45 s followed by a single cycle of 94°C for 30 s, 59.5°C for 45 s, and 72°C for 45 s followed by a single cycle of 94°C for 30 s, 59°C for 45 s, and 72°C for 45 s. In the touchdown PCR, the primer annealing temperature is reduced from 60°C by 0.5°C per cycle, to 54°C.
3. 30 cycles of 94°C for 30 s, 54°C for 45 s, and 72°C for 45 s.
4. A final extension step is performed for 10 min at 72°C.

Standard PCR cycling conditions for the amplification of FR3-constant region are as follows:

1. Initial denaturing step at 94°C for 15 min.
2. 55 cycles of 94°C for 10 s, 57°C for 10 s, and 72°C for 25 s.
3. Final extension step of 10 min at 72°C.

3.4.2. Primer Combinations

3.4.2.1. If the Constant Region Phenotype Is Known

Begin V_H gene rescue with the following:

1. V_H leader mix: constant (Cμ or Cγ or Cα , as indicated by immunophenotype) and
2. Individual V_H leader 1/7 and 2 and 3 and 4 and 5 and 6: constant.

Au: Leader and constant have been set in lowercase throughout because they do not appear to be proprietary names; please confirm.

3.4.2.2. If the Constant Region Phenotype Is NOT Known

In this situation, it is beneficial to perform a small number of preliminary PCRs to identify the constant region using a 5' primer designed to anneal to the FR2 region together with all available 3' constant region and J_H primers:

FR2 consensus - Cμ and Cγ and Cα and J_H consensus and J_H mix

FR2 is selected because it is less likely to be mutated than the leader and FR1 regions. Therefore, the chance of mismatches between the V_H gene and the primers designed to hybridize to this region is lower resulting in a greater chance of amplification.

Which subheading details gel analysis and sequencing? Please supply

3.4.3. Outcome of First PCR (see also Gel Analysis and Sequencing)

3.4.3.1. If the Constant Region Phenotype Is Known

See also **Notes 2** to **6**.

1. If two or more PCR products have similar or repeated sequences with a clonally related CDR3, then the tumor-related V_H gene has been identified and confirmed; proceed to clone one PCR product.
2. If one PCR product has a clonal sequence through V_H CDR3 and J_H that can be aligned successfully to VBase, attempt to confirm this sequence by repeating the PCR using the same primer pair and/or set up a second PCR using the family specific leader primer for the identified germ line donor together with the constant region primer. One or both of these PCR products should confirm the sequence and clonally related CDR3. Proceed to cloning.
3. If multiple PCR products have clonal sequence through V_H CDR3 to J, but the sequences are not similar and do not share a clonally related CDR3, then attempt to identify the "true" tumor-related V_H gene by repeating all the relevant PCRs. If more than one sequence is repeated, pursue both sequences via cloning. It is rare to find true V_H biclonality in FL; therefore, rigorous checks for contamination need to be performed at this point to eliminate this possibility.
4. If the V_H leader PCRs failed to recover an unambiguous sequence, then, in this case, the primer pairs used for the PCR need to be revised. It is useful to perform FR2 to all possible constant region primers and to J_H consensus and J_H mix primers as a next step, since we have found that these primer combinations have a high hit rate for identification of V_H (for outcome, *see* **Subheading 3.4.3.2.**).

3.4.3.2. If the Constant Region Phenotype Is NOT Known:

In this instance, the preliminary PCR serves to identify the heavy chain constant region used by the tumor clone and to determine which constant region primers to use in later PCRs.

1. If one FR2 PCR product has a clonal sequence through the end of V_H CDR3 and J_H that can be aligned successfully to VBase, then a partial sequence for the tumor-related V_H gene (FR2-constant) and the constant region it uses has been identified. Proceed to perform PCRs using the appropriate individual family-specific V_H leader primers to the identified constant primer as in **Subheading 3.4.2.1.** Once confirmed, the leader PCR product (full-length V_H gene) can be cloned.
2. Two or more FR2 PCR products amplified from different constant region primers have similar or repeated sequences with a clonally related CDR3: A partial

sequence for the tumor-related V_H gene (FR2 constant) is identified and is shown to express multiple isotypes. Proceed to perform PCRs using the appropriate individual family specific V_H leader primers to one or all of the identified constant primer, as in 3.4.2.1. Once confirmed a single leader PCR product (full-length V_H gene) can be cloned.

3. If multiple FR2 PCR products have clonal sequence through the end of V, CDR3 to J but the sequences are not similar and do not share a clonally related CDR3, attempt to identify the "true" tumor-related V_H gene by performing PCRs using the appropriate individual family specific V_H leader primers to the identified constant primer, as in **Subheading 3.4.2.1.** If more than one sequence is repeated, pursue both sequences via cloning if contamination has been ruled out.
4. If the FR2 PCRs failed to recover an unambiguous sequence, then experience has shown that FR2/constant primer pairs for PCR usually will lead to the identification of the tumor-related V_H if it is there to be found, but this is not always the case and, therefore, further primer combinations should be investigated (*see* **Subheading 3.4.4.**). However, if failure to detect a clonal V_H sequence from a FR2 PCR is accompanied by failure to identify the partner V_L, it is likely that further investigation will be fruitless and may indicate problems with the tumor load. Genescan of FR3-$J_H$consensus PCR products can be employed the confirm polyclonality of the sample (*see* **Subheading 3.7.3.**).

3.4.4. If V_H Leader PCRs Yield Polyclonal Sequences

It is possible that extensive mutation of the tumor-related V_H gene within the leader region causes failure of V_H leader primers to hybridize to this region and results in a lower amplification efficiency. This is masked by the amplification of normal B cell V_H rearrangements and leads to polyclonal products. The next step is to use primers designed to anneal to the framework regions, where sequence is inherently conserved. The following primer pairs are used until the clone is identified or the primer combinations are exhausted. In our laboratory the failure rate for the identification of tumor-related V_H genes using these primer sets is approx 10%.

3.4.4.1. Framework 1 Region Primers

1. V_H FR1 Mix: Cμ and Cγ and Cα and $J_H$consensus and $J_H$mix.
2. V_H FR1 1/7and 2, 3, 4, 5, and 6 - Cμ and Cγ and Cα and $J_H$consensus and $J_H$mix.

3.4.4.2. Framework 2 Region Primers

V_H FR2 consensus: Cμ and Cγ and Cα and J_H consensus and J_H mix.

3.4.4.3. Framework 3 Region Primers

V_H FR3 consensus - Cμ and Cγ and Cα and J_H consensus and J_H mix.

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

3.5. Light-Chain PCR

3.5.1. V_L PCR: Set Up and Cycling Conditions

V_L genes are amplified by PCR in the same way as for V_H genes (refer to **Subheading 3.4.1.**) with the exception of the "Touchdown" PCR cycling conditions, which are listed below.

1. Initial denaturing step at 94°C for 15 min.
2. A single cycle of 94°C for 30 s, 56°C for 45 s, and 72°C for 45 s followed by a single cycle of 94°C for 30 s, 55.5°C for 45 s, and 72°C for 45 s followed by a single cycle of 94°C for 30 s, 55°C for 45 s. and 72°C for 45 s. In the touchdown PCR, the primer annealing temperature is reduced from 56°C by 0.5°C per cycle to 50°C.
3. 30 cycles of 94°C for 30 s, 50°C for 45 s, and 72°C for 45 s.
4. Final extension step at 72°C for 10 min.

3.5.2. Primer Combinations

3.5.2.1. If The Light Chain Restriction Is Known

Begin Vκ gene rescue with: Vκ leader mix (constantκ27 and constantκ69).
Begin Vλ gene rescue with: Vλ leader mix (constantλ33 and constantλ85).
Duplicating PCRs with two different constant region primers increases the likelihood of identifying and confirming the V_L tumor clone in a single PCR.

3.5.2.2. If The Light Chain Restriction Is NOT Known

Au: Do you mean "add" leader mix to constant PCRs? Please clarify.

For an unknown light chain restriction, both Vκ and Vλ leader mix to constant PCRs (as listed above) are performed simultaneously.

3.5.3. Outcome of First PCR (see also Gel Analysis and Sequencing)

For more information, also *See* **Notes 2–6**.

1. If both light-chain PCR products have similar or repeated sequences with a clonally related CDR3, then Tumor-related V_L gene has been identified and confirmed; proceed to clone one PCR product.
2. If only one PCR product has a clonal sequence through V_L, CDR3 to J_L that can be aligned successfully to VBase, then attempt to confirm this sequence by repeating the PCR using the same primer combination and/or set up a second PCR using the family-specific leader primer for the identified germ line donor. It is likely that one or both of these PCRs will confirm the sequence and clonally related CDR3.
3. If both PCR products have clonal sequence through V_L, CDR3 to J_L but the sequences are not similar and do not share a clonally related CDR3, then attempt to confirm both the V_L gene sequences by repeating the relevant PCRs and in addition set up a second PCR using the family-specific leader primer for each of the identified

germ line donors. If both sequences are validated by a second independent PCR, pursue both via cloning. In our experience, dual light chain rearrangements occur in about 8% of FL; however, be sure to rule out possible PCR contamination.

4. If both PCR products fail to recover an unambiguous sequence then, in this case, the primer pairs used for the PCR need to be revised (*see* **Subheading 3.5.4.**)

The outcomes listed above also apply in cases where the light chain restriction is not known. Both Vκ and Vλ PCRs are performed and the products analyzed independently. In the majority of cases, either one or both of the Vκ or Vλ PCR products will have a clonal sequence that can be aligned to VBase, with polyclonal products for the second light chain. However, it is possible that both Vκ and Vλ PCR products yield polyclonal sequences and, therefore, it is necessary to use individual family-specific primers for both Vκ and Vλ, sequentially until the tumor-related V_L is identified. Because Vκ rearrangements are found more frequently than Vλ (60:40; for normal B cells), it is logical to start with the investigation of Vκ followed by Vλ. For more information, *see* **Notes 2–6.**

3.5.4. If V_L Leader Mix PCRs Yield Polyclonal Sequences

In many cases, the tumor-related V_L gene can be readily identified using a leader mix together with a constant region primer. However, if this primer pair yields only polyclonal sequences, individual family-specific leader primers can be used. Because Vκ and Vλ have numerous families, each with a disproportionate number of germ line donors, we begin investigation with the largest families first before moving on to those families with fewer germ line donors until the clone is identified or the primer combinations are exhausted. In our laboratory the failure rate for the identification of tumor-related V_L using these primer sets is less than 5%. This is lower than for the V_H genes and is likely a consequence of an intrinsically lower level of somatic mutation in V_L genes.

3.5.4.1. Individual Family-Specific V κ Leader Primers

Vκ Leader 1 and 2 and 3 (constant κ27 or constant κ69) followed by Vκ Leader 4 and 5 and 6 (constant κ27 or constant κ69).

3.5.4.2. Families With Fewer Members

Vλ Leader 1 and 2 and 3 and 4 and 5 (constant λ33 or constant λ85) followed by Vλ Leader 6 and 7 and 8 and 9 and 10 (constant λ33 or constant λ85).

3.6. Agarose Gel Electrophoresis

PCR products are separated by 1% agarose gel electrophoresis against 1 kb+ DNA ladder (Life Technologies, Paisley, UK) and visualized with ethidium

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

bromide under UV light. Bands of the appropriate size (350–400 bp) were excised and purified using the QIAquick Gel Extraction Kit (QIAGEN). Remember to take a photograph for documentation!

1. Excise band and weigh.
2. Add 3 volumes of buffer QG to 1 volume gel (100 mg~100 µL) and incubate at 50°C for 10 min.
3. Add 1 volume of isopropanol, mix, and apply to the column.
4. Centrifuge at 14,000 rpm for 1 min.
5. Wash once with buffer QG (optional) followed by one wash with buffer PE.
6. Elute DNA in an appropriate volume of buffer EB.

Au: Please set centrifugation speed in gravity (g) throughout chapter.

3.7. Sequencing

3.7.1. Sequencing Reaction

Direct DNA sequencing of PCR products is performed on an ABI 377 automatic DNA sequencer (Applied Biosystems, Warrington, UK) using ABI Prism BigDye Terminator Kit (Applied Biosystems) in a final volume of 10 µL.

1. 5 µL of PCR product.
2. 1 µL 5' primer, 1.6 pmol.
3. 2 µL Big Dye.
4. 2 µL 5X sequencing buffer.

Sequencing PCR cycling conditions consist of 25 cycles of 96°C for 10 s, 50°C for 50 s, and 60°C for 4 min. This is followed by Sodium Acetate and ethanol precipitation.

1. 10 µL of sequencing reaction is added to 1 µL of 3 *M* sodium acetate (pH 5.2) and 25 µL 100% ethanol and placed on ice for 15 min.
2. Samples are centrifuged at 14,000 rpm for 30 min at 4°C.
3. Samples are washed once with 75% ethanol.
4. After air drying, pellets are resuspended in 1 µL of loading buffer (1:5 of loading dye:formamide) for immediate analysis or storage at –20°C.

3.7.2. Sequence Analysis

For sequence alignment analysis, use MacVector software (Oxford Molecular, Oxford, UK) aligned to Entrez and V-BASE databases (Center for Protein Engineering, MRC Center, Cambridge, UK; http://www.mrc-cpe.cam.ac.uk./vbase-ok), which contain all known human germ line Ig V region genes.

Tumor-related V genes were defined by the presence of repeated sequences with a clonally related CDR3. Somatic hypermutation is determined by comparison of the sequence to the germline V genes with the greatest homology. The percentage of somatic mutation is calculated from the aligned sequences from the beginning of FR1 to the end of FR3 (*see* **Notes 7–9**).

3.7.2. Genescan Analysis

In the event that all PCR attempts using the available primer sets fail to identify the tumor-related V_H and V_L clones, GeneScan™ can be used to assess the polyclonality of a sample. This technique uses a simple V_H PCR from FR3 to a 6-FAM-labeled J_H consensus primer according to the set up and cycling conditions listed in **Subheading 3.4.1.** The resulting ~100 bp PCR product is run directly on an ABI 377 automatic DNA sequencer (Applied Biosystems) incorporating the GeneScan™ Analysis software. If the product is polyclonal, it is very unlikely that the clonal signature of the tumor will be identified with further investigation. Alternatively, GeneScan™ may reveal a clonal population. The FR3-J_H PCR can then be sequenced to identify the clonal CDR3 and subsequent PCR using a CDR3-specific reverse primer. This, together with individual leader and FR1 primers, may resolve the full-length V_H gene sequence.

3.8. Cloning

3.8.1. Ligation

Clonal PCR products are ligated into pGEM-T vector system I (Promega).

1. 3 µL of PCR product.
2. 1 µL of Empty pGEM-T vector.
3. 1 µL of T4 DNA ligase.
4. 5 µL of 2X buffer.

Ligations reactions αρε περφορμεδ ατ 4ϒC overnight.

3.8.2. Transformation

Transformation into JM109 competent *Escherichia coli* cells (Promega) follows:

1. Gently mix 5 µL of ligation product with 50 µL of JM109 cells and leave on ice for 20 min.
2. Heat shock cells at 42°C for *exactly* (use a timer!) 45 s and return to ice for 5 min.
3. Add 1 mL of LB and shake at 37°C for 1 hour.
4. Plate cells onto X-Gal/0.1 *M* Isopropyl b-D-thiogalactopyranoside, Amp100 Agar and leave overnight at 37°C.

The next day, 12 randomly picked bacterial colonies are each used to spike 2mLs Amp100 LB. Grow bacterial cultures at 37ϒC overnight.

3.8.3. Plasmid Purification

Plasmid is purified from the bacterial cultures using the QIAprep Miniprep kit (QIAGEN,).

1. Pellet cultures by centrifugation and resusupend in 250µL of buffer P1.
2. Add 250 µL of lysis buffer P2 and mix by inversion.
3. Neutralize the reaction within 5 min by the addition of 350 µL of buffer N3. Leaving the alkaline solution for longer will hydrolyze the DNA (see manufacturer's instructions).
4. After centrifugation at 14,000 rpm for 10 min, apply the supernatant to the column and leave for 1 min.
5. Centrifuge at 14,000 rpm for 1 min
6. Wash once with buffer PB, followed by two washes with buffer PE.
7. Ensure a final centrifugation at 14,000 rpm for 3 min before eluting DNA with 100 µL of buffer EB

All plasmids are sequenced in both directions using T7 and SP6 forward and reverse primers. We aim to collect approx 10 clonally related sequences to confidently assess intraclonal variation in FL.

4. Notes

1. When working with RNA, gloves should be worn at all times and touching contaminated surfaces and equipment should be avoided; RNases can be reintroduced by contact with ungloved hands and equipment or with unfiltered air. Use sterile, RNase-free tubes and pipet tips and purchase reagents that are free of RNases. Separate reagents and plasticware used for RNA work from general-use reagents in the laboratory. Place all RNA samples on ice; RNA is susceptible to degradation when left at room temperature.
2. If no amplified products result from the first round of PCR, check cDNA integrity with β-actin primers if not previously performed. If, subsequently, 1) β-actin PCR fails, repeat cDNA synthesis, ensuring that positive and negative controls are run in parallel, and repeat using new cDNA synthesis reagents if necessary. In the event that the β-actin PCR fails again and assuming that cDNA synthesis reagents are viable, proceed to check the quality of the RNA and assess sample collection and storage. 2) If β-actin PCR successful, then the cDNA is viable. Repeat PCR using more cDNA template OR increase the number of PCR cycles OR lower primer annealing temperatures OR try alternative primer combinations OR use the above in combination.
3. If amplified products are weak, repeat PCR using more cDNA template OR increase the number of PCR cycles OR use in combination.
4. If multiple products are amplified from a single primer pair,excise, purify, and sequence each individual band that falls within the expected size range. Alternatively, repeat the PCR using standard (i.e., not "touchdown") cycling conditions and an increased annealing temperature.
5. If the sequencing of PCR product fails, check that the sequencing primer is correct. If not, repeat the sequencing reaction using more template DNA. If this coincides with a low yield of the original PCR product, disregard this product and attempt to increase the yield of product in a repeat PCR (refer to **Subheading 4.2.**) before resequencing.

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

6. If the sequencing of the PCR product appears clean but does not align to the VBase database, disregard this product. It is not Ig and most probably results from contamination for example from bacterial DNA.
7. If the sequence of the PCR product will 1) align to a V segment but not a J segment or 2) not align to a V or a J segment, examine the electropherogram for the presence of "double peaks" which indicate the existence of more than one sequence. Pay particular attention to the CDR3 region, from the end of V to constant, which may be polyclonal even if the V segment appears clonal. This PCR product is polyclonal and likely results from the amplification of V genes from normal B cells, which exist within the tumor cell population.
8. If the cloning of the PCR product yields a low number of related sequences, purify more plasmids from the transformation plate for this PCR product AND/OR clone out the second confirmatory PCR.
9. If cloning of the PCR product yields clonally related sequences but also a second group of sequences (>2) that share a clonally related CDR3 different to that already identified and you are confident that this is not the result of contamination and the tumor is indeed biclonal, purify more plasmids from the transformation plate for this PCR product and determine the dominant clone. We have observed dual light chain rearrangements in ~8% of FL, whereas we have not found any tumors that proved to be biclonal for V_H.

References

1. NCI Non-Hodgkin's Classification Project Writing Committee. (1985) Classification of non-Hodgkin's lymphomas: reproducibility of major classification systems: *Cancer* **55,** 457–481.
2. Horning, S. J. (1993) Natural history of and therapy for the indolent non-Hodgkins's lymphoma. *Semin. Oncol.* **20,** 75–88.
3. Ersbøll, J., et al. (1989) Follicular low grade non-Hodgkin's lymphoma: long-term outcome with or without tumor progression: *Eur. J. Haematol.* **42,** 155–163.
4. Simon, R., et al. (1988) The Non-Hodgkin Lymphoma Pathologic Classification Project. Long-term follow-up of 1153 patients with non-Hodgkin lymphomas. *Ann. Intern. Med.* **109,** 939–945.
5. Zukerberg, L. R., et al. (1993) Diffuse low-grade B cell lymphomas. Four clinically distinct subtypes defined by a combination of morphologic and immunophenotypic features. *Am. J. Clin. Pathol.* **100,** 373–385.
6. Stein, H., et al. (1984) Immunohistological analysis of human lymphoma: correlation of histological and immunological categories. *Adv. Cancer Res.* **42,** 67–147.
7. Harris, N. L., et al. (1984) Immunohistologic characterization of two malignant lymphomas of germinal center type (centroblastic/centrocytic and centrocytic) with monoclonal antibodies. Follicular and diffuse lymphomas of small- cleaved-cell type are related but distinct entities. *Am. J. Pathol.* **117,** 262–272.
8. Stein, H., et al. (1982) The normal and malignant germinal center. *Clin. Haematol.* **11,** 531–559.

Job: Illidge (159-6)
Chapter: 07/McCann
Template: MiMM/6x9/TempW/S(Rev.06.04)

Operator: SV
Date: 12/04
Revision: 1st Pass

9. Harris, N. L., et al. (1994) A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood* **84,** 1361–1392.
10. Johnson, P. W., et al. (1994) Detection of cells bearing the t(14;18) translocation following myeloablative treatment and autologous bone marrow transplantation for follicular lymphoma. *J. Clin. Oncol.* **12,** 798–805.
11. Stamatopoulos, K., et al. (2000) Molecular insights into the immunopathogenesis of follicular lymphoma. *Immunol. Today* **21,** 298–305.
12. Willis, T. G., et al. (2000) The role of immunoglobulin translocations in the pathogenesis of B cell malignancies. *Blood* 96, 808–822.
13. Lam, K. P., et al. (1997) In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell* **90,** 1073–1083.
14. Levy, R., et al. (1987) Somatic mutation in human B cell tumors. *Immunol. Rev.* 43–58.
15. Schroeder, H. J., et al. (1994) The pathogenesis of chronic lymphocytic leukemia: Analysis of the antibody repertoire. *Immunol. Today* **15,** 288–294.
16. Li, Y., et al. (1996) The B binding specificity of human VH4–34 (VH4–21) encoded antibodies is determined by both VH framework region 1 and complementarity determining region 3. *J. Mol. Biol.* **256,** 577–589.
17. Pospsil, R., et al. (1998) CD5 and othr superantigens as 'ticklers' of the B cell receptor. *Immunol. Today* **19,** 106–108.
18. Silverman, G. J. (1992) Human antibody responses to bacterial antigens: studies of a model conventional antigen and a proposed model B cell superantigen. *Int. Rev. Immunol.* **9,** 57–78.
19. Bahler, D. W., et al. (1992) Clonal evolution of a follicular lymphoma: evidence for antigen selection. *Proc. Natl. Acad. Sci. USA.* **89,** 6770–6774.
20. Zelenetz, A. D., et al. (1991) Histologic transformation of follicular lymphoma to diffuse lymphoma represents tumor progression by a single malignant B cell. *J. Exp. Med.* **173,** 197–207.
21. Zhu, D., et al. (1994) Clonal history of a human follicular lymphoma as revealed in the immunoglobulin variable region genes. *Br. J. Haematol.* **86,** 505–512.
22. Malisan, F., et al. (1996) B-Chronic lymphocytic leukemias can undergo isotype switching in vivo and can be induced to differentiate and switch in vitro. *Blood* **87,** 717–724.
23. Efremov, D. G., et al. (1996) IgM-producing chronic lymphocytic leukemia cells undergo immunoglobulin isotype-switching without acquiring somatic mutations. *J. Clin. Invest.* **98,** 290–298.
24. Ottensmeier, C. H., et al. (1998) Analysis of VH genes in follicular and diffuse lymphoma shows ongoing somatic mutation and multiple isotype transcripts in early disease with changes during disease progression. *Blood* **91,** 4292–4299.
25. Stevenson, F., et al. (1998) Insight into the origin and clonal history of B cell tumors as revealed by analysis of immunoglobulin variable region genes. *Immunol. Rev.* **162,** 247–259.
26. Alizadeh, A. A., et al. (2000) Distinct types of diffuse large B cell lymphoma identified by gene expression profiling. *Nature* **403,** 503–511.

27. Hamblin, T. J., et al. (2000) Immunoglobulin V genes and CD38 expression in CLL. *Blood* **95,** 2455–2457.
28. Berek, C. (1992) The development of B cells and the B cell repertoire in the microenvironment of the germinal center. *Immunol. Rev.* **5,** 5–19.
29. Limpens, J., et al. (1995) Lymphoma-associated translocation t(14;18) in blood B cells of normal individuals. *Blood* **85,** 2528–2536.
30. Han, S., et al. (1996) Neoteny in lymphocytes: Rag1 and Rag2 expression in germinal center B cells. *Science* **274,** 2094–2097.
31. Henderson, A., et al. (1998) Transcriptional regulation during B cell developmen. *Annu. Rev. Immunol.* **16,** 163–200.
32. Chen, C., et al. (1997) Editing disease-associated autoantibodies. *Immunity* **6,** 97–105.
33. Pelanda, R., et al. (1997) Receptor editing in a transgenic mouse model: site, efficiency, and role in B cell tolerance and antibody diversification. *Immunity* **7,** 765–775.
34. Radic, M. Z., et al. (1996) Receptor editing, immune diversification, and self-tolerance. *Immunity* **5,** 505–511.
35. Bahler, D., et al. (1991) Ig VH gene expression among human follicular lymphomas. *Blood* **78,** 1561–1568.
36. Stevenson, F. K., et al. (1993) Differential usage of an Ig heavy chain variable region gene by human B cell tumors. *Blood* **82,** 224–230.
37. Hummel, M., et al. (1994) Mantle cell (previously centrocytic) lymphomas express V(H) genes with no or very little somatic mutations like the physiologic cells of the follicle mantle. *Blood* **84,** 403–407.
38. Stamatopoulos, K., et al. (1997) Follicular lymphoma immunoglobulin kappa light chains are affected by the antigen selection process, but to a lesser degree than their partner heavy chains. *Br. J. Haematol.* **96,** 132–146.
39. Noppe, S. M., et al. (1999) The genetic variability of the VH genes in follicular lymphoma: the impact of the hypermutation mechanism. *Br. J. Haematol.* **107,** 625–640
40. Zhu, D., et al. (2002) Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. *Blood* **99,** 2562–2568.
41. Cleary, M. L., et al. (1986) Clustering of extensive somatic mutations in the variable region of an immunoglobulin heavy chain gene from a human B cell lymphoma. *Cell* **44,** 97–106.
42. Zelenetz, A. D., et al. (1992) Clonal expansion in follicular lymphoma occurs subsequent to antigenic selection. *J. Exp. Med.* **176,** 1137–1148.
43. Zelenetz, A. D., et al. (1993) A submicroscopic interstitial deletion of chromosome 14 frequently occurs adjacent to the t(14;18) translocation breakpoint in human follicular lymphoma. *Genes Chromosom. Cancer* **6,** 140–150.
44. Wu, H., et al. (1995) A human follicular lymphoma B cell line hypermutates its functional immunoglobulin genes in vitro. *Eur. J. Immunol.* **25,** 3263–3269.
45. Corbett, S. J., et al. (1997) Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. J. *Mol. Biol.* **270,** 587–597.

Universal N-glycosylation sites introduced into the B-cell receptor of follicular lymphoma by somatic mutation: a second tumorigenic event?

Leukemia (2006) **20,** 530–534. doi:10.1038/sj.leu.2404095; published online 26 January 2006

We have previously shown that variable region heavy chain (V_H) genes in FL commonly acquire sequence motifs by somatic mutation that serve as sites for N-glycosylation.[1] We further showed that subsets of aggressive B NHL acquire sites[2] in contrast to sequences of myeloma, CLL and normal B cells where this is rare.[1] These data suggest positive selection and a critical role for the introduced glycosylation sites in the pathogenesis of FL.

To provide a context for interpreting our own data, we assessed V_H gene usage in our own and published sequences.[3–8]

Table 1 The incidence of novel N-glycosylation sites in the V genes of patients with follicular lymphoma

Case	V_H			V_L		
	No. of sites	*Region*	*Motif*	*No. of sites*	*Region*	*Motif*
P1	0	—	—	1	CDR3	NTS
P2	2	CDR1, CDR2	NMS, NSS	0	—	—
P3	1	CDR3	NCS	1	CDR3	NNS
P4[a]	1	CDR2	NIS	1	FR4	NLT
P5	0	—	—	1	CDR3	NYT
P6[a]	1	CDR2	NIT	0	—	—
P7	1	CDR2	NIS	1	FR3	NDS
				1	FR1	NVT
P8	1	CDR2	NIT	0	—	—
P9				0	—	—
P10	1	CDR2	NIS	1	FR3	NGS
P11	1	CDR3	NST	0	—	—
P12	1	CDR2	NKS	0	—	—
				0	—	—
P13	1	CDR2	NVS	0	—	—
P14[a]	2	CDR1, CDR2[b]	NWT, NHS	0	—	—
P15	1	CDR3	NIS	0	—	—
				0	—	—
P16	2	FR3, FR4	NDS, NIT	1	CDR3	NHT
P17	3	CDR2x2, CDR3	NGT, NYT, NYT	1	CDR1	NLT
P18	0	—	—			
P19				0	—	—
P20	1	CDR2	NVS	1	CDR3	NFT
P21[a]	4	FR1, CDR1, CDR2, CDR3	NFT, NMS, NTS, NVT	1	CDR1	NIS
P22	1	CDR2	NIS	1	FR3	NGS
P23				1	CDR1	NSS
P24	2	CDR1, CDR2	NMS, NIT	1	FR1	NLS
P25[a]	1	CDR2	NYS	1	FR1	NAT
P26	1	CDR3	NHS	0	—	—
P27[a]	1	CDR2	NIS	0	—	—
P28	0	—	—	1	CDR3	NYS
P29[a]	2	CDR2, FR3	NIS, NNS	1	CDR3	NYS
P30				0	—	—
P31	1	CDR2	NIT			
P32	0	—	—	0	—	—
P33	4	CDR2x2, CDR1, CDR3	NGS, NYT, NWS, NKT	0	—	—
P34	2	CDR2, FR4	NIS, NVS	1	CDR1	NLT
P35	2	FR1, CDR3	NFS, NSS	2	FR1, FR3	NIT, NLT
P36				1	CDR3	NSS
P37				0	—	—
P38	1	CDR3	NTS	0	—	—
P39	0	—	—	0	—	—
P40				0	—	—
P41[a]	1	CDR3	NIS	0	—	—
P42				0	—	—
P43	1	CDR3	NTS	0	—	—
P44	0	—	—			

[a]Expressed as recombinant scFv in 293F cells and functional glycosylation demonstrated.
[b]Naturally occurring N-glycosylation site from the germline gene V4-34.
Underlining indicates ambiguous N-glycosylation motifs for which functional glycosylation at the protein level is questionable.
Shading indicates tumors without a N-glycosylation motif in the V gene consensus sequences.

Analysis of the V_H and/or V_L genes in an extended cohort of 262 cases of FL revealed no significant bias compared to normal B cells. We next analysed matched V_H and V_L sequences to provide a complete view of the B-cell receptor obtained from biopsies after written informed consent of the patients. Sequence analysis of FL V genes shows that the most dramatic distinction to normal B cells lies in the accumulation of motifs available for addition of oligosaccharides. V_H and V_L sequences were examined for potential N-glycosylation sites with the motif Asn-X-Ser/Thr (N-X-S/T), where X could be any amino acid except Pro (Table 1). Motifs using Asp, Glu, Leu or Trp as the middle amino acid X and when followed by Ser were included in this analysis even though they represent disfavoured amino acids where functional glycosylation at the protein level may be less likely.[9] Twenty-nine of 36 V_H rearrangements and 20 of 44 V_L rearrangements acquired new sites. Of note, the naturally occurring site in V4-34 appears irrelevant and is commonly lost in FL. The paired analysis of 33 FL shows that 94% of tumours acquire at least one and up to five novel potential N-glycosylation sites and of these 13 have sites in both V_H and V_L, 15 in V_H alone and three in V_L alone. A motif was not found in two cases. In case P32, it was likely due to an unusually low level of somatic hypermutation (SHM). We observed 1.0 and 1.1% mutation of V_H and V_L, respectively, compared to other FL (V_H: mean 10.6%; V_L mean 6.0%). In P39, motifs are present in some members of the clone, but perhaps have not yet assumed dominance. Here, we observed novel sites in the CDR3 of 3/12 of the V_H clones and in the CDR1 and CDR3 of 1/11 and 3/11 of the V_L clones, respectively (Figure 1). Creation of additional motifs by SHM in subclones was also evident in cases with motifs in the consensus sequences found in all clones. V_H genes from four cases (P2, P12, P16, P21) had additional motifs arising in this way, affecting on average 14% of clones. Similarly, V_L genes from a further four cases (three in V_κ (P8, P17, P42) and one in V_λ (P23)) had additional sites which were shared by 38% of clones. Overall, nine of 44 (21%) tumours continued to accumulate N-glycosylation sites as a result of ongoing SHM.

The distribution of novel N-glycosylation sites in the V_H and V_L gene sequences is shown in Figure 2a. The majority (87%) of sites in V_H were located in the complementarity determining regions (CDRs): CDR2 ($n=22$) >CDR3 ($n=11$) >CDR1 ($n=5$). Framework regions (FRs) acquired fewer sites ($n=6$).

V_H: 3/12 clones CDR3 (aa100-100b)

AA	N	Y	S
Bp	AAT	TAC	TCT
		C	
P39-H7	...	.G.	...
P39-H12	...	.G.	...
P39-H3	...	.G.	...
	K		
P39-H4	..G	...	...
P39-H6	..A	...	...
	L		
P39-H9	TTA	...	...
	R		
P39-H8	.GA	...	...
P39-H10	.GA	...	...
P39-H11	.GA	...	...
	I	*y*	*F*
P39-H5	.T.	...	...
P39-H1	.T.	...	...
P39-H2	.T.	..T	.T.

V_L: 3/11 clones CDR3 (aa93-95)

AA	N	S	R
Bp	AAC	AGT	CGC
			S
P39-L9	...	...	A..
P39-L7	...	...	A..
P39-L2	...	...	A..
P39-L12	...	...	...
P39-L1	...	...	...
P39-L10	...	...	...
P39-L3	...	...	...
		S	
P39-L8	.G.	...	...
P39-L6	.G.	...	...
P39-L4	.G.	...	...
		R	*P*
P39-L5	.GG	...	.C.

V_L: 1/11 clones CDR1 (aa31b-32)

AA	Y	N	Y
Bp	TAT	AAC	TAT
	N		*S*
P39-L8	A.C	...	.C.
	y		
P39-L9	..C	...	...
P39-L5	..C	...	...
P39-L12	..C	...	...
P39-L1	..C	...	...
P39-L10	..C	...	...
P39-L6	..C	...	...
P39-L7	..C	...	...
P39-L4	..C	...	...
P39-L3	..C	...	...
	N		
P39-L2	A.C	...	...

Figure 1 The accumulation of N-glycosylation sites during ongoing SHM in P39. The V_H and V_L clones from P39 were analysed (FR1 to FR4) for N-glycosylation motifs introduced during ongoing SHM resulting in intraclonal heterogeneity. The consensus sequences of tumour-derived V_H and V_L from P39 do not feature a site.

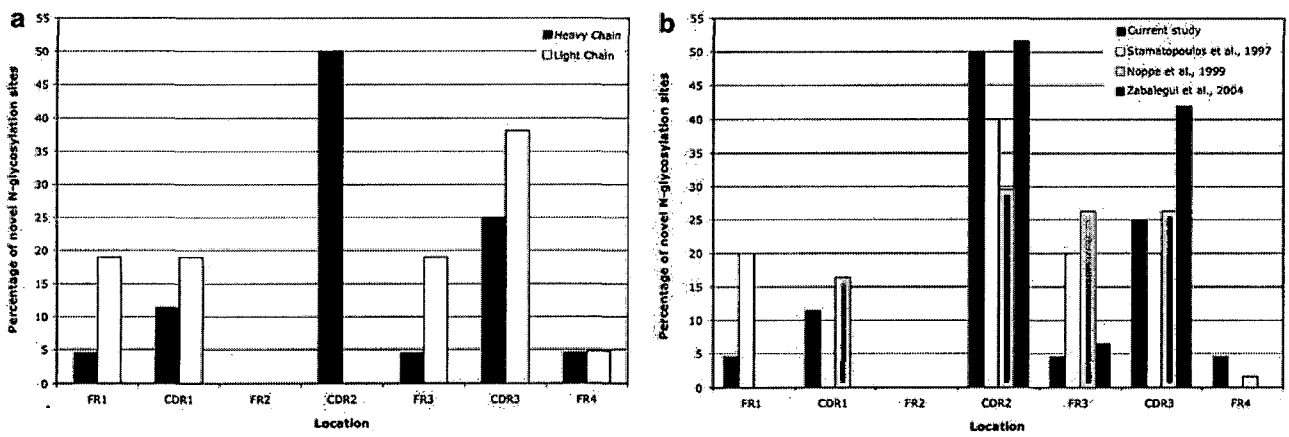


Figure 2 *Distribution of novel N-glycosylation sites in the heavy and light chain variable region genes of follicular lymphoma.* (**a**) *The location* of novel N-glycosylation sites within the V_H and V_L sequences of 29 and 20 FL, respectively, were analysed (FR1 to FR4). Values are expressed for each region as a percentage of the total number of sites: V_H ($n=44$); V_L ($n=21$) (**b**) The location of N-glycosylation sites observed in the V_H genes of our 29 FL was compared to other studies with published sequences from 69 FL.[5,6,8] Values are expressed for each region as a percentage of the total number of sites ($n=141$).

Table 2 Analysis of somatic mutations giving rise to N-glycosylation motifs: a focus on V_H CDR2

Case[a]	*GL donor*	*AA position*	*Generation of N-glycosylation site*			
			AA change	*Bp mutation*[b]		
				1	*2*	*3*
P31	V3-07	50–52	NIK→NIT	AAC	ATA	ACC
FL12	V3-07	50–52	NIK→NIS	na	na	na
Rut	V3-07	50–52	NIK→NIT	AAT	ATA	ACA
FL10	V3-09	50–52	GIS→NIS	na	na	na
P24	V3-11	50–52	YIS→NIT	AAT	ATT	ACT[c]
FL2	V3-11	52–52b	SS-→NGS	na	Na	—
		52c–54	-SS→NTS	—	na	na
FL16	V3-11	58–60	YYA→NYT	na	na	na
Dier	V3-11	52–53	SSS→NGT	AAT	GGT[c]	ACT[c]
		56–58	TIY→NIS	AAC	ATA	TCT
Viv	V3-11	50–52	YIS→NIT	AAC	ATT	ACT[c]
		56–58	TIY→NTS	AAC	ACA[c]	AGT
P12	V3-15	52a–52c	SKT→NKS	AAC	AAA	TCT[c]
Cal	V3-15	52a–52c	SKT→NKS	AAT	AAA	AGT[c]
Hou	V3-15	52a–52c	SKT→NKT	AAT	AAA	ACT
P8	V3-23	50–52	AIS→NIT	AAT	ATT	ACT[c]
P17	V3-23	52–53	SGS→NGT	AAT	GGC	ACT[c]
		58–60	YYA→NYT	AAT	TAC	ACG
FL_5	V3-23	50–52	AIS→NIT	AAT	ATA	ACT[c]
FL17	V3-23	50–52	AIS→NIT	na	na	na[c]
Timj	V3-23	50–52	AIS→NSS	AAT	AGT[c]	AGT
Vit	V3-23	50–52	AIS→NIT	AAT	ATT	ACT[c]
FL5	V3-30	50–52	VIS→NIS	na	na	na
FL24	V3-30	56–58	NKY→NTS	na	na[c]	na
Clau	V3-30	58–60	YYA→NYS	AAT	TAT	TCA
P2	V3-48	52–53	SSS→NSS	AAT	AGT	AGT
P6	V3-48	50–52	YIS→NIT	AAC	ATT	ACT[c]
P7	V3-48	50–52	YIS→NIS	AAC	GAG	AGT
P10	V3-48	50–52	YIS→NIS	AAC	ATT	AGT
P21	V3-48	52–53	SSS→NTS	AAC	ACT[c]	AGT
P22	V3-48	50–52	YIS→NIS	AAT	ATT	AGT
P27	V3-48	50–52	YIS→NIS	AAC	ATT	AGT
P29	V3-48	50–52	YIS→NIS	AAC	ATT	AGC
P34	V3-48	50–52	YIS→NIS	AAC	ATC	AGT
FL_10	V3-48	50–52	YIS→NIS	AAT	ATT	AGT
FL19	V3-48	50–52	YIS→NIS	na	na	na
Tim	V3-48	50–52	YIS→NIS	AAC	ATT	AGC
Vho	V3-48	50–52	YIS→NIT	AAC	ATT	ACT[c]
		55–57	STI→NIT	AAT	ATA[c]	ACA
Lah	V3-64	53–55	NGG→NGT	AAT	GGG	ACT
P33	V4-04	54–56	SGS→NGS	AAT	GGG	AGC
		58–60	NYN→NYT	AAC	TAC	ACG
Reu4	V4-04	60–62	NPS→NHS	AAT	CAC	TCC
P25	V4-34	58–60	NYN→NYS	AAC	TAC	AGC
FL_3	V4-34	50–52	EIN→NIS	na	na	na
FL6	V4-34	50–52	EIN→NIS	na	na	na
		60–62	NPS→NAS	na	na	na
FL7	V4-34	52–54	NHS→NYS	na	na	na
FL22	V4-34	54–56	SGS→NGS	na	na	na
VWB3	V4-34	58–60	NYN→NCS	AAC	TGC[c]	AGC
P20	V4-39	50–52	SIY→NVS	AAC	GTA[c]	TCT
P4	V4-59	50–52	YIY→NIS	AAC	ATC	TCT
Marn	V4-59	54–56	SGS→NGS	AAT	GGG	AGC
P13	V4-61	50–52	YIY→NVT	AAT	GTC[c]	TCT

[a]Included data from the current study and from published sequences.[5,6,8]
[b]Amino-acid position 1, 2 and 3 of N-glycosylation motif.
[c]Non-essential amino-acid replacement mutation.
na; nucleotide sequence not known.
Underlining indicates a nucleotide replacement mutation.
Shading indicates an amino-acid replacement.

No difference was found in the number or location of sites between tumours of different isotypes. For V_L, the motifs are focused mainly in CDR3, with a notable absence in CDR2. No sites were found in FR2 in either V_H or V_L in keeping with the internal location of this region in the antibody protein. To broaden the analysis, the location of sites in the published V_H genes of an additional 69 FL[5,6,8] were analysed (Figure 2b). The focusing of V_H sites within CDR2 and CDR3 (41 and 29% of sites, respectively) was confirmed. Considering also that in normal B cells with similar levels of SHM motifs are observed in

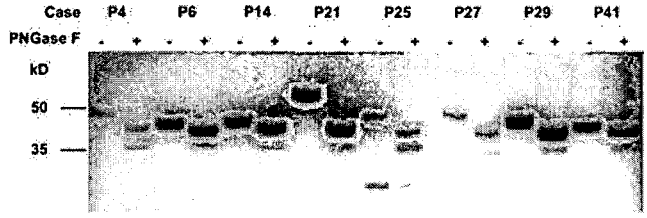


Figure 3 Glycosylation of scFv proteins expressed in 293F cells. Tumour-derived V_H and V_L sequences from eight patients (P4, P6, P14, P21, P25, P27, P29 and P41) were assembled as single chain Fv (scFv), linked to a human kappa constant (Cκ) region and cloned into the expression vector pcDNA3. Proteins were expressed using the FreeStyle™293 expression system (Invitrogen), purified on a sheep anti-human free κ chain column, separated by SDS-PAGE and visualized by staining with Coomassie blue. To remove N-linked carbohydrates, scFv-Cκ proteins were treated with Peptide:N-glycosidase F (+) (New England Biolabs).

only a small minority of sequences,[1] it appears that the motifs in CDR3 of FL are acquired by SHM rather than by the VDJ (V_H) or VJ (V_L) recombination.

In all, 26 different N-glycosylation motifs were observed. The most commonly used was NIS, which was observed in nine (21%) of V_H genes, but which was only seen once in V_L. Other repeatedly used motifs included NIT ($n=5$), NSS, NTS, NYT and NLT (all $n=4$). The disfavoured motifs NDS, NLS and NWS were found in four FL and involved two V_H genes and two V_L genes, but in all cases additional sites were present in either the V gene itself or its partner chain. Motifs were created by a range of nucleotide changes. The codon for Asn (AAU or AAC) was generated by mutation rather than being derived from germline sequence in 41 of 45 (91%) instances. 'Starter' sequences for motif generation depended on the donor germline V_H. For example, YIS present in V3-48 and V3-11 gene segments at CDR2 was mutated in six sequences to NIS and in two cases to NIT. Mutational changes of Ser to Asn were relatively common (12/45), likely reflecting the presence of a hot spot of SHM at codon 50 in CDR2 (see below). Other common amino-acid (aa) replacements in codon 1 affected Tyr (15/45) and Thr (6/45), where replacement mostly required one nucleotide change to generate the starter Asn. Acquisition of sites was indeed predominantly (73%) achieved by a single aa replacement to Asn at position 1 of the motif. At the nucleotide level, the first codon underwent mutations of one (68%), two (29%) and rarely three (2%) nucleotides. In a substantial proportion (24%), additional, non-essential aa replacements in codon 2 ($n=4$) or codon 3 ($n=7$) were found. It is interesting to note that unlike the Asn in codon 1, the C-terminal Ser of the motif was most commonly found either in the natural sequence ($n=20$) or mutated from Ser to Thr ($n=6$); 18 motifs therefore ended in Thr, of which only one-third were present in the germline sequence. A change of Thr to Ser was only observed once. Where codon 3 was mutated (42%), SHM had affected one (63%) or two (37%) nucleotide mutations. In all, 18% of sites were created through an essential aa replacement at both positions 1 and 3 of the motif, but only in a minority (9%) by an essential aa replacement at position 3 alone. A more detailed analysis of the generation of N-glycosylation sites in the V_H CDR2 was performed and included sequences from cited references[5,6,8] (Table 2). All 47 sequences with CDR2 motifs used V_H3 ($n=35$) and V_H4 ($n=12$) genes. Here, 60% had acquired a site at aa position 50–52. Codon 50 is a recognized hotspot for SHM in both productively and non-productively rearranged alleles of normal B cells and affected ~20% of cases.[10] A total of 60% of FL sequences showed a replacement mutation here (28/47). This may, in part, be explained by the relative over representation of the V_H3 genes and in particular V3-48, for which the majority of sites were generated by aa replacement at position 50. In contrast, there is no recognized hotspot for SHM at codon 52 of V_H,[10] but we find that codon 52 is mutated in 21 of 34 cases. The high incidence of codon 52 changes in CDR2 motifs further point to selective pressure on FL to acquire N-glycosylation sites in their V genes. Eight further cases using the gene segments V3-11, V3-23 or V3-30 generated sites by a replacement mutation at codon 50, the majority (6/8) with an additional permissive mutation in codon 52. For segments V3-07, all sites were generated by the replacement of Lys to either Ser or Thr at position 52. For segments V3-15, all sites were generated by the replacement of Ser to Asn at position 52a. In the remaining gene segments, sites did not appear to focus to any one aa position. Strikingly in V_L there is a complete absence of motifs in CDR2 in spite of a known hotspot for SHM in codon 52. This points to a selection against further acquisition of sites here, as in normal B cells CDR2 (and CDR1) appear to be preferentially subjected to SHM.[11] Other common hotspots for SHM such as codons 31, 55 or 82a were not used to generate novel sites and we observed no repeated location of sites in V_L. Overall analysis of the affected codons provides further evidence for positive selection. Additionally, the accumulation of motifs in CDR3 of FL compared to normal B cells provides new data that this region in the V genes is affected by SHM, which had previously been difficult to show directly in the absence of a germline counterpart.

Using a 293F mammalian expression system, recombinant scFv proteins incorporating the paired V_H and V_L from eight FL were expressed and purified. We show that in all eight cases, the expressed tumour-derived scFv contains glycans, which are susceptible to enzymatic removal *in vitro* (Figure 3). Heterogeneity was also consistent with variable addition of carbohydrate. For example, P21, which has four N-glycosylation sites in V_H and one in V_L migrated much more slowly than P6, which has just one site in V_H. This finding supports the conclusion that sites defined by molecular means are likely glycosylated in FL cells. Whether disfavoured motifs contribute to variable region glycosylation is difficult to assess at this point as in all cases additional motifs using a favoured aa sequence were detected.

The incidence of introduced motifs appears universal in somatically mutated FL, but the phenomenon is not confined to FL. We have also observed sites in a proportion (~40%) of cases of DLBCL, and commonly (82%) in endemic Burkitt's lymphoma.[2] However, sites are only occasionally seen in mutated CLL and multiple myeloma. The same is true for available sequences of normal B cells (memory B cells or plasma cells) with comparable levels of SHM to FL.[1] One conclusion might be that the added oligosaccharides have a role in survival or growth of B-cell malignancies in the germinal centre. With regard to FL, it is likely that the characteristic t(14;18) translocation occurs in the bone marrow during maturation of the B cells and clearly this alone does not confer tumorigenic behaviour since it is found in healthy individuals. However, retention of sIg expression is high in FL, pointing to a possible role for stimulation via the B-cell receptor in pathogenesis. The finding of functional N-glycosylation motifs suggests a strategy for the tumour cell to interact with a lectin in the germinal centre. This could replace the need for persistent antigen which would differ in every case and is an unlikely continuing

stimulant. The ongoing stimulation via a substitute lectin could account for the ongoing somatic mutation. Our data support the concept that glycosylation could contribute to the development of FL. The characteristic t(14;18) chromosomal translocation may be only a first step to tumorigenesis, and we propose that the universal acquisition of N-glycosylation sites in the GC contributes a second critical factor. The nature of the added oligosaccharide is at present under investigation, with preliminary experiments suggesting an unusual composition (manuscript in preparation). If this is indeed a mechanism required for maintenance of FL, interruption would provide a clear therapeutic goal.

Acknowledgements

This work was funded by Cancer Research UK and the Nessex Cancer Trust.

KJ McCann, PWM Johnson, FK Stevenson and CH Ottensmeier
Cancer Sciences Division, Somers Cancer Sciences Building, University of Southampton, Southampton, UK
E-mail: cho@soton.ac.uk

References

1 Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK. Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. *Blood* 2002; **99**: 2562–2568.
2 Zhu D, Ottensmeier CH, Du MQ, McCarthy H, Stevenson FK. Incidence of potential glycosylation sites in immunoglobulin variable regions distinguishes between subsets of Burkitt's lymphoma and mucosa-associated lymphoid tissue lymphoma. *Br J Haematol* 2003; **120**: 217–222.
3 Bahler DW, Campbell MJ, Hart S, Miller RA, Levy S, Levy R. Ig VH gene expression among human follicular lymphomas. *Blood* 1991; **78**: 1561–1568.
4 Hsu FJ, Levy R. Preferential use of the VH4 Ig gene family by diffuse large-cell lymphoma. *Blood* 1995; **86**: 3072–3082.
5 Stamatopoulos K, Kosmas C, Papadaki T, Pouliou E, Belessi C, Afendaki S *et al.* Follicular lymphoma immunoglobulin kappa light chains are affected by the antigen selection process, but to a lesser degree than their partner heavy chains. *Br J Haematol* 1997; **96**: 132–146.
6 Noppe SM, Heirman C, Bakkus MH, Brissinck J, Schots R, Thielemans K. The genetic variability of the VH genes in follicular lymphoma: the impact of the hypermutation mechanism. *Br J Haematol* 1999; **107**: 625–640.
7 Aarts WM, Bende RJ, Steenbergen EJ, Kluin PM, Ooms EC, Pals ST *et al.* Variable heavy chain gene analysis of follicular lymphomas: correlation between heavy chain isotype expression and somatic mutation load. *Blood* 2000; **95**: 2922–2929.
8 Zabalegui N, de Cerio AL, Inoges S, Rodriguez-Calvillo M, Perez-Calvo J, Hernandez M *et al.* Acquired potential N-glycosylation sites within the tumor-specific immunoglobulin heavy chains of B-cell malignancies. *Haematologica* 2004; **89**: 541–546.
9 Kasturi L, Chen H, Shakin-Eshleman SH. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem J* 1997; **323** (Part 2): 415–419.
10 Dorner T, Brezinschek HP, Brezinschek RI, Foster SJ, Domiati-Saad R, Lipsky PE. Analysis of the frequency and pattern of somatic mutations within nonproductively rearranged human variable heavy chain genes. *J Immunol* 1997; **158**: 2779–2789.
11 Foster SJ, Dorner T, Lipsky PE. Targeting and subsequent selection of somatic hypermutations in the human V kappa repertoire. *Eur J Immunol* 1999; **29**: 3122–3132.

JAK2 V617F mutation in classic chronic myeloproliferative diseases: a report on a series of 349 patients

Leukemia (2006) **20**, 534–535. doi:10.1038/sj.leu.2404086; published online 12 January 2006

Several groups have recently reported the presence of a unique gain-of-function *JAK2* point mutation (V617F) in the majority of cases of classic cMPD as well as in a minority of atypical MPDs and in occasional cases of AML. To date, this mutation has not been reported in normal controls, lymphoid disorders or in patients with secondary erythrocytosis.[1–10]

In this study, we have analyzed bone marrow or peripheral blood samples from 349 patients (PV, $n=84$; ET, $n=243$; IMF, $n=22$) referred to our diagnostic service for cytogenetic analysis and/or exclusion of the *BCR-ABL* fusion from several hospitals in the north of Spain between 1996 and 2005. Our study group thus includes the largest series of ET cases reported to date for *JAK2* analysis. DNA was extracted and tested for V617F by a sensitive amplification refractory mutation system as described previously.[6] Results were obtained from all samples and V617F was detected in 80, 62 and 68% of cases of PV, ET and IMF, respectively (Table 1). These frequencies are comparable to other studies for PV and IMF, but somewhat higher than expected for ET, for which 23–57% of cases have been reported as V617F positive. Part of the reason for this is probably attributable to technical differences in the sensitivity of the V617F detection. Most other studies used sequencing, which can underestimate the proportion of positive patients,[2,6] whereas allele-specific PCR assay is capable of detecting a mutation present in 1–2% of cells.[6]

Of the 17 V617F-negative PV cases of our series, only two were female ($P=0.044$, contingency $\chi^2=4.058$). This supports the previous observation of a significant male excess in mutation-negative PV cases.[6] In addition, a relative risk ratio estimate (RR = 1.26; confidence interval 95% = 1.04–1.52) of our data indicated that V617F is 26% more frequent in PV females compared to PV males (Table 1).

PV and IMF are both considered to be clonal diseases, but for ET only approximately half of the cases are clonal, the remainder being polyclonal and therefore presumably reactive. It has been suggested[6] that the great majority of clonal ET cases must be V617F positive, but it remains to be determined if any V617F ET cases are also clonal. To address this question, we compared karyotypes between mutation-negative and mutation-positive cases. Of the 191 ET cases for whom cytogenetic analysis had been successfully performed, clonal cytogenetic abnormalities were observed in 11 of the 118 (9%) V617F-positive and five of the 73 (7%) V617F-negative patients (Table 2). Although the frequency of chromosome abnormalities

Available online at www.sciencedirect.com
ScienceDirect
Molecular Immunology 45 (2008) 1567–1572
Molecular Immunology
www.elsevier.com/locate/molimm

Remarkable selective glycosylation of the immunoglobulin variable region in follicular lymphoma

Katy J. McCann[a,*], Christian H. Ottensmeier[a], Alice Callard[a], Catherine M. Radcliffe[b,1], David J. Harvey[b], Raymond A. Dwek[b], Pauline M. Rudd[b,2], Brian J. Sutton[c], Paul Hobby[c], Freda K. Stevenson[a]

[a] *Genetic Vaccine Group, Cancer Sciences Division, University of Southampton School of Medicine, Southampton, UK*
[b] *Glycobiology Institute, Department of Biochemistry, University of Oxford, Oxford, UK*
[c] *The Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Hospital Campus, London, UK*

Received 4 October 2007; accepted 5 October 2007
Available online 19 November 2007

Abstract

Follicular lymphoma (FL) generally expresses immunoglobulin (Ig) with somatically mutated variable (V) region genes. Surprisingly, these almost always carry introduced motifs available for *N*-glycosylation (Asn-X-Ser/Thr). Introduced motifs are uncommon on normal B cells, but are on other germinal center (GC)-associated B-cell malignancies suggesting a site-specific role. They are not evident in mutated chronic lymphocytic leukemia (CLL) or myeloma. Recently, we found that the glycosylation sites are unusual in containing oligomannose glycans, which are apparently displayed on tumor cell surface IgM. This suggests a potential interaction with a mannose receptor in the GC. However, natural *N*-glycosylation sites exist in germline (GL) V region genes, particularly the V_{4-34} gene expressed by normal B cells and by some malignancies, including CLL, potentially undermining the selective importance for FL. To compare oligosaccharide addition at the introduced and natural sites, we expressed V region genes as single chain Fv (scFv) and analyzed the added glycans. In contrast to introduced sites, which were oligomannosylated, the natural GL motif in the V_{4-34} sequence had no added sugars. The remarkable selective glycosylation within the heavy chain V region gene of FL apparently permits only limited processing to oligomannose at somatically mutated motifs, creating a feature exploitable by GC lymphomas.

Keywords: Lymphoma; Immunoglobulin variable region; V_{4-34}; Glycosylation

1. Introduction

Strikingly, follicular lymphoma (FL) cells can be distinguished from normal germinal center (GC) B cells by the acquisition of sequence motifs available for *N*-glycosylation (Asn-X-Ser/Thr) (Zhu et al., 2002). These motifs arise during somatic mutation and appear to be positively selected by FL and certain other GC-located B-cell tumors, including Burkitt's lymphoma (Zhu et al., 2003). Analysis of the added glycans of FL-derived immunoglobulin (Ig) revealed that they are unusual as they contain oligomannoses in the variable (V) regions but complex sugars in the constant regions (Radcliffe et al., 2007). This consistent differential glycosylation pattern indicates normal transit through the Golgi stacks but steric blockade of sugar processing in the V regions. It also suggests a tumor-related function for the expressed oligomannoses,

Abbreviations: CDR, complementarity-determining region; C_K, constant kappa region; CLL, chronic lymphocytic leukaemia; Endo H, endoglycosidase H; FL, follicular lymphoma; GC, germinal center; GL, germline; Ig, immunoglobulin; PCNSL, primary CNS lymphoma; PNGase F, peptide*N*-glycosidase F; scFv, single chain variable fragment; V_H, heavy chain variable gene; V_L, light chain variable gene.

* Corresponding author at: Genetic Vaccine Group, Cancer Sciences Division, Southampton University Hospitals Trust, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK. Tel.: +44 2380 794863; fax: +44 2380 795152.
E-mail address: kjm8@soton.ac.uk (K.J. McCann).
[1] Current address: Lonza Biologics Plc, 228 Bath Road, Slough, UK.
[2] Current address: Dublin-Oxford Glycobiology Laboratory, NIBRT, Conway Institute, University College Dublin, Dublin 4, Ireland.

0161-5890/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.molimm.2007.10.009

possibly via interaction with mannose receptors expressed by stromal cells in the GC. Importantly, oligomannose glycans are also presented at the cell surface of a Burkitt's lymphoma cell line carrying the motif, as revealed by specific binding to mannose-binding lectin (Radcliffe et al., 2007).

However, functional importance for GC-associated lymphomas was questioned by the presence of natural *N*-glycosylation sites in certain germline (GL) V region genes. The main example is the V_{4-34} gene, used by ~4–10% of normal B cells and by a significant proportion of cases of chronic lymphocytic leukemia (CLL), a tumor not associated with the GC (Hamblin et al., 1999). If the natural site in V_{4-34} were also to carry oligomannoses, the importance of this feature for GC-associated lymphomas would diminish. We now report that single chain Fv (scFv) molecules derived from tumor cells and expressed in mammalian cells mimic whole Ig in acquiring oligomannoses at the introduced motifs. However, the neighbouring natural motif in the GL sequence of V_{4-34}-encoded scFv not only lacks oligomannoses, but the nascent polypeptide fails to acquire any added glycan. The importance of introduced motifs in FL is supported by their conservation during ongoing somatic mutation *in vivo*. In contrast, apparent random loss of the natural site occurs. These findings reveal consistent differential glycosylation of V region sequences with likely biological significance.

2. Materials and methods

2.1. Patient material and V region gene analysis

Tumor specimens were obtained from 12 patients with either untreated FL ($n=7$), primary CNS lymphoma (PCNSL) ($n=2$), mutated CLL ($n=2$) or myeloma ($n=1$). Informed consent was provided by all participants following ethical approval from the IRB. Identification of tumor-derived V region genes and single chain variable fragment (scFv) assembly were as described (Hawkins et al., 1994; McCann et al., 2005; Ottensmeier et al., 1998).

2.2. Assembly and expression of scFv–Cκ proteins

scFvs were cloned into pcDNA3 vector (Invitrogen Limited, Paisley, UK) together with the constant kappa (Cκ) sequence, which was linked at the carboxy-terminus. Plasmid DNA (100 μg) was transfected into 1×10^8 293F cells using 293Fectin™ (Invitrogen Limited). Cultures were incubated at 37 °C in humidified 8% CO_2 on an orbital shaker at 125 rpm. After 72 h, cultures were centrifuged and supernatants filtered (0.22 μm) and concentrated 10× using vivaspin20 10,000 MWCO concentrators (Sartorious Limited, Epsom, UK). Proteins were purified by immunosorption using polyclonal sheep anti-human κ chain linked to Sepharose 4B (Amersham Biosciences UK Limited, Little Chalfont, UK). Eluted proteins were dialysed into PBS and the concentration determined by BCA™ protein assay (Perbio Science UK Limited, Cramlington, UK).

2.3. Analysis of scFv N-linked glycosylation

scFv-Cκ proteins (4 μg) were electrophoresed in a NuPAGE Bis–Tris gradient polyacrylamide (4–12%) gel (Invitrogen Limited). To remove N-linked glycans, proteins were treated separately with peptide: *N*-glycosidase F (PNGase F) and endoglycosidase H (Endo H) (both New England Biolabs (UK) limited, Hitchin, UK). Separated protein bands were visu-

Table 1
Incidence of *N*-glycosylation sites in tumor-derived Ig V region genes

Case	V_H					V_L				
	V_H	SHM[a] (%)	No.	Glycosylation site		V_L	SHM[a] (%)	No.	Glycosylation site	
				Location[b]	Motif				Location	Motif
FL2	IGHV3–48	9.2	2	CDR1	NMS	IGKV1-17	3.4	0	–	–
				CDR2	NSS					
FL6	IGHV3–48	11.6	1	CDR2	NIT	IGKV3–20	6.0	0	–	–
FL11	IGHV3–21	10.5	1	CDR3	NST	1–44	5.6	0	–	–
FL14	IGHV4–34	8.8	2	CDR1	NWT	IGKV4–1	2.8	0	–	–
				$CDR2^{GL}$	NHS					
FL21	IGHV3–48	9.9	4	FR1	NFT	IGKV1–39	9.1	1	CDR1	NIS
				CDR1	NMS					
				CDR2	NTS					
				CDR3	NVT					
FL26	IGHV3–64	15.6	1	CDR3	NHS	IGKV4–1	6.7	0	–	–
FL29	IGHV3–48	6.8	1	CDR2	NIS	IGKV1–39	15.5	1	CDR3	NYS
				FR3	NNS					
CLL1	IGHV4–34	9.3	1	$CDR2^{GL}$	NHS	IGKV2–30	2.3	0	–	–
CLL2	IGHV4–34	10.6	1	$CDR2^{GL}$	NHS	IGKV1–12	7.2	0	–	–
PCNSL1	IGHV4–34	1.0	1	$CDR2^{GL}$	NHS	3–19	1.9	0	–	–
PCNSL2	IGHV4–34	10.3	1	$CDR2^{GL}$	NHS	IGKV3–20	5.6	0	–	–
MM1	IGHV5–51	4.8	0	–	–	IGKV1–33	4.5	0	–	–

[a] Frequency of somatic hypermutation (SHM) compared to the most homologous GL V gene segment.
[b] GL: indicates the natural GL-encoded V_{4-34} *N*-glycosylation site.

alised by staining with SimplyBlue Safestain™ (Invitrogen Limited). Three scFv-Cκ proteins (15 μg) were also separated on 10% SDS–PAGE, visualised by Coomassie Blue and excised. Glycans were extracted, fluorescently labelled with 2-aminobenzamide and analyzed by normal phase HPLC; exoglycosidase array digestions and Glycobase databases were complied in the Oxford Glycobiology Institute (Radcliffe et al., 2007). Unlabelled samples were also analyzed by mass spectroscopy.

3. Results

3.1. Glycans at introduced motifs

Seven cases of FL carrying Asn-X-Ser/Thr motifs introduced by somatic mutation were analyzed; one case also contained a GL-encoded site (FL14). All expressed Ig with somatically mutated V region genes (Table 1). Each carried ≥1 motif in the heavy chain V region gene (V_H), mainly in the complementarity-determining regions (CDRs), and 2/7 had additional motifs in the light chain V region gene (V_L). FL-derived V_H and V_L genes were expressed as scFv linked to Cκ. Glycosylation was demonstrated in all expressed scFv–Cκ proteins by the slowed migration on SDS/PAGE (apparent MW > 45 kDa, calculated MW ~ 42 kDa) (Fig. 1A–D, lane 1). Following treatment with PNGase F, which removes all N-linked oligosaccharides, migration was increased to that expected for scFv–Cκ (Fig. 1A–D, lane 2). In 6/7 cases (Fig. 1A, C, D, lane 3), treatment with Endo H had the same effect. Endo H cleaves the bond between the two proximal *N*-acetyl glucosamine residues of oligomannose but not complex oligosaccharides and, therefore, cleavage indicates oligomannosylation (Tarentino et al., 1974). FL21 has five motifs, including one in V_L. In this case, Endo H treatment increased migration, but not to that observed following

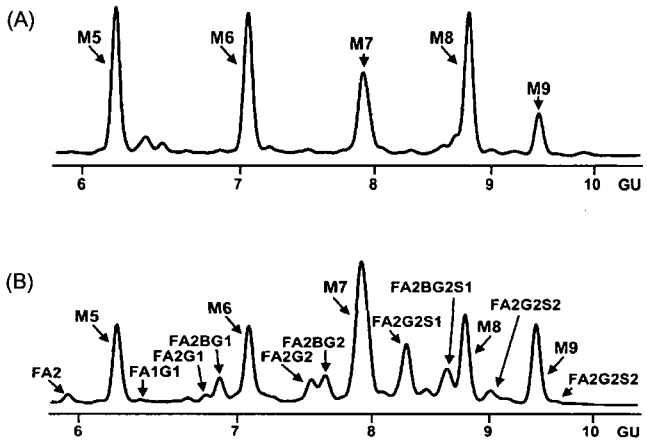


Fig. 2. HPLC glycan profiles of scFv–Cκ from FL21 and FL29. HPLC confirms (A) oligomannose glycans in FL29 and (B) oligomannose and complex glycans in FL21. Peak assignments are A: antennary, B: bisect, F: core fucose, G: galactose, M: mannose and S: sialic acid.

PNGase F treatment (Fig. 1B, lane 3), indicating that some glycans remained as oligomannose while others had been processed further. Confirmation of sugar composition was obtained by analyzing the scFv–Cκ of one case (FL29) by HPLC, which showed only oligomannose glycans, M5–M9 (Fig. 2A). In contrast, FL21 had additional complex sugars, such as core fucosylated biantennary digalactosylated disialylated glycans, FA2G2S2 (Fig. 2B). Mass spectroscopy was consistent with the HPLC results (data not shown).

3.2. Glycans at natural GL-encoded motifs

The V_{4-34} gene encodes a natural glycosylation site, Asn-His-Ser in CDR2. Four cases expressing Ig derived from V_{4-34} were from CLL (CLL 1 and 2) and PCNSL (PCNSL1 and

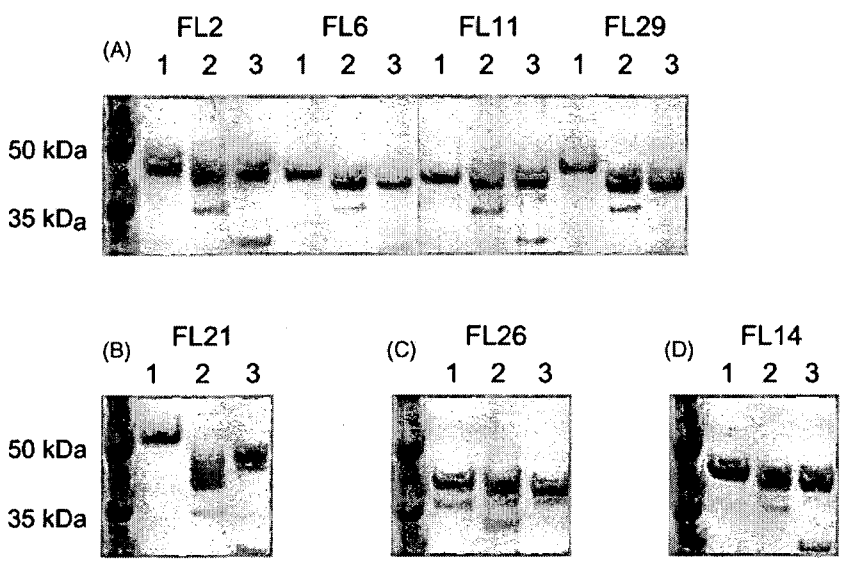


Fig. 1. Glycosylation of FL-derived scFv–Cκ proteins. Using the FreeStyle™ 293 expression system, tumor-derived scFv–Cκ proteins were expressed and purified. Proteins were separated by SDS–PAGE and visualised by staining with SimplyBlue Safestain™ (lane 1). To remove N-linked carbohydrates, proteins were treated with two separate *N*-glycan removal enzymes: PNGase F (lane 2) and Endo H (lane 3). (A, C, D) scFv–Cκ proteins derived from six FL display N-linked glycosylation of the high mannose type, which is susceptible to both enzymes. (B) scFv–Cκ protein derived from FL21 displays N-linked glycosylation of both the high mannose and complex type.

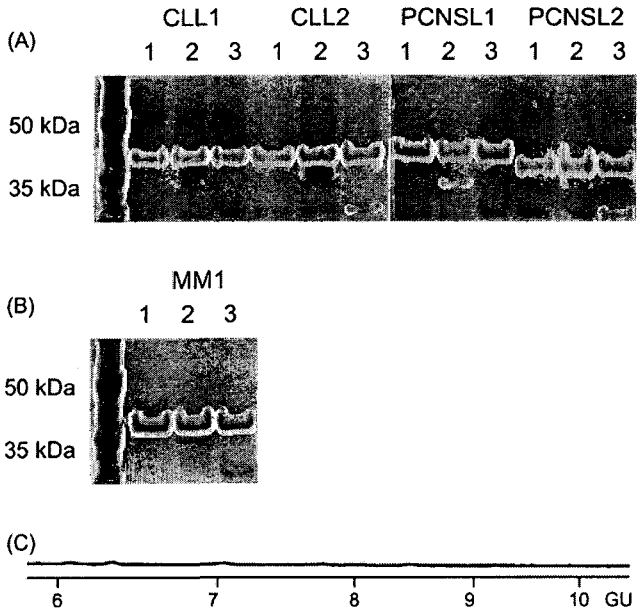


Fig. 3. Absence of glycosylation in V_{4-34}-encoded scFv–Cκ proteins. N-Linked glycosylation is absent for (A) four B-cell tumors that express the V_{4-34} gene containing the GL-encoded motif but no other motifs and (B) one control B-cell tumor that does not contain a motif. (C) HPLC confirms a lack of glycosylation in PCNSL1.

	Novel site				GL site		
AA position	33	34	35		52	53	54
AA GL	Y	W	S		N	H	S
	N		T			h	R
V4-34 GL	TAC	TGG	AGC	---	AAT	CAT	AGT
FL14-vh1	A..	...	.CA	---	...	...	...
FL14-vh2	A..	...	.CA	---	...	...	...
FL14-vh3	A..	...	.CA	---	...	...	...
FL14-vh4	A..	...	.CA	---	...	...	...
FL14-vh5	A..	...	.CA	---	...	...	...
FL14-vh6	A..	...	.CA	---	...	...	..G
FL14-vh7	A..	...	.CA	---	...	...	...
FL14-vh8	A..	...	.CA	---	...	...	..G
FL14-vh9	A..	...	.CA	---	...	..C	...
FL14-vh10	A..	...	.CA	---	...	...	..G
FL14-vh11	A..	...	.CA	---	...	...	...
FL14-vh12	A..	...	.CA	---	...	...	..G
FL14-vh13	A..	...	.CA	---	...	...	...
FL14-vh14	A..	...	.CA	---	...	...	...

Fig. 4. Conservation of novel N-glycosylation sites, but not the GL V_{4-34}-encoded site, in the face of continued somatic mutation. The effect of ongoing somatic mutation resulting in intraclonal heterogeneity on introduced motifs and the GL-encoded V_{4-34} motif was assessed in FL14. The single novel site was conserved in all clones while the GL-encoded site was lost in 4/14 clones.

2). All V region genes were somatically mutated, but carried only the GL-encoded *N*-glycosylation motif in CDR2 with no introduced motifs. Expressed scFv–Cκ molecules showed single sharp bands migrating according to expected molecular mass (Fig. 3A, lane 1). Exposure to PNGase F or Endo H had no effect on migration, consistent with a lack of *N*-glycosylation of the GL-encoded motif (Fig. 3A, lanes 2 and 3, respectively). The pattern was similar to that of a control scFv–Cκ derived from a patient with myeloma (MM1) (Fig. 3B), which carried somatically mutated V region genes but no GL-encoded or introduced motifs (Table 1). Data obtained by HPLC analysis of PCNSL1 confirmed that the GL-encoded site in V_{4-34} was not glycosylated (Fig. 3C).

3.3. Asn-His-Ser as a functional acceptor motif

The CLL and PCNSL cases reveal that the natural Asn-His-Ser motif in the GL CDR2 sequence is not glycosylated whereas the introduced sites in FL are. One question is whether Asn-His-Ser, although a favored sequence for *N*-glycosylation, can in fact be glycosylated in introduced sites (Kasturi et al., 1997). Table 1 shows that His is rarely involved in mutated sites, but we identified it in CDR3 of FL26. Expressed scFv–Cκ from this case was susceptible to Endo H (Fig. 1C), consistent with *N*-glycosylation of this single introduced motif and confirming that this sequence is an acceptor.

3.4. Differential fate of natural and introduced motifs in V_{4-34}-encoded FL

To assess the fate of the natural site in V_{4-34}-encoded FL during somatic mutation we interrogated the V_H sequence database (Noppe et al., 1999; Stamatopoulos et al., 1997; Zabalegui et al., 2004). We found that 12/15 (80%) lost the GL-encoded site, indicating no positive selection. In contrast, 80% acquired new motifs by somatic mutation. The acquisition rate is similar to that which we reported for all available V_H sequences, indicating that V_{4-34}-encoded cases behave similarly (Zhu et al., 2002). Since only V_H sequences could be interrogated, it remains consistent with our observation that, if V_H and V_L are examined, almost all cases of FL acquire sites (McCann et al., 2006).

Distinct behavior of the GL-encoded and acquired sites was also evident during ongoing somatic mutation. For example, cloning and sequencing of V_H of V_{4-34}-encoded FL14 (Table 1) revealed a loss of the natural Asn-His-Ser site in 4/14 clones, but complete conservation of the single acquired Asn-Trp-Thr site (Fig. 4). Among 30 FL cases, introduced motifs in the V region genes were rarely lost during intraclonal mutational changes and at least one motif was conserved by all clones in 27 cases (Table 2).

4. Discussion

Analysis of the added glycans in the V regions of FL reveal that they are commonly oligomannosylated, in contrast to the constant regions, which contain complex sugars (Radcliffe et al., 2007). However, our findings now reveal that this is a feature only of somatically mutated sites. Differential glycosylation of Ig V regions has been reported in a mouse anti-dextran antibody (Gala and Morrison, 2004). In that case, complex oligosaccharides were detected at Asn^{58} in CDR2, but at a second Asn residue, introduced at position 60 by site-directed mutagenesis, only oligomannose glycans were observed. Clearly differential *N*-glycosylation can occur in V regions, and in human

Table 2
The conservation of novel *N*-glycosylation sites among molecular clones

Case	No. of sites	Motif[a]	No. of clones with motif[b]
FL1	1	NTS	12/12
FL2	2	NMS, NSS	13/14, 14/14
FL3	2	NCS, NNS	7/7, 16/16
FL5	1	NYT	9/9
FL6	1	NIT	17/18
FL7	3	NIS, NDS	13/13, 4/4
		NVT	10/10
FL8	1	NIT	10/10
FL10	2	NIS, NGS	7/7, 6/6
FL11	1	NST	8/8
FL12	1	NKS	22/22
FL13	1	NVS	8/8
FL14	2	NWT, NHS^{GL}	14/14, 10/14
FL15	1	NIS	11/11
FL16	3	NDS, NIT	6/8, 5/8
		NHT	10/10
FL17	4	NGT, NYT	6/6, 6/6
		NYT, NLT	6/6, 8/8
FL20	2	NVS, NFT	8/8, 14/14
FL21	5	NFT, NMS	15/15, 15/15
		NTS, NVT	15/15, 15/15
		NIS	20/26
FL22	2	NIS, NGS	8/8, 6/6
FL24	3	NMS, NIT	6/6, 6/6
		NLS	9/9
FL25	2	NYS, NAT	13/13, 5/5
FL26	1	NHS	6/7
FL27	1	NIS	9/11
FL28	1	NYS	8/8
FL29	3	NIS, NNS	10/10, 7/10
		NYS	4/4
FL33	4	NGS, NYT	10/11, 9/11
		NWS, NKT	11/11, 11/11
FL34	3	NIS, NVS	6/6, 6/6
		NLT	11/11
FL35	4	NFS, NSS	17/17, 16/17
		NIT, NLT	9/10, 10/10
FL38	1	NTS	14/15
FL41	1	NIS	11/11
FL43	1	NTS	11/11

[a] GL: indicates the natural GL-encoded V_{4-34} *N*-glycosylation site.

[b] Number of clones displaying the *N*-glycosylation motif/total number clones sequenced.

FL occurs in sequences that have not been manipulated *in vitro*. Our data using expressed scFv–Cκ proteins from naturally encoded human FL cells reproduce the oligomannosylation pattern seen in the V regions of whole Ig rescued from FL cells of six previous cases (Radcliffe et al., 2007). Moreover, it also reveals selective *N*-glycosylation, but with a consistent pattern.

We first confirm that all the somatically mutated sites in FL are oligomannosylated. It appears that within the V region, motifs introduced by somatic mutation, although located at multiple positions, mainly within the CDRs and commonly in CDR2, are oligomannosylated. In contrast, an adjacent GL-encoded motif in CDR2 is evidently unavailable for addition of the initial dolichol phosphate-linked oligosaccharide precursor ($GlcNAc_2Man_9Glc_3$) in the ER and, therefore, no *N*-glycosylation is detectable at this position. Since the Asn-His-Ser motif is apparently an acceptor, the reason for the lack of oligosaccharide addition is unclear.

A recent statistical analysis of the local sequence and structural environment of Asn-X-Ser/Thr motifs identified a number of factors that predispose towards or against glycosylation, including the accessibility of the Asn side-chain in the folded structure, whether Ser or Thr lies at position +3 and the presence or absence of acidic residues in the immediately preceding positions (−1 and −2) (Petrescu et al., 2004). Low accessibility of the site in the ER during the addition of the oligosaccharide precursor, Ser at position +3, and an acidic residue at position −1 or −2, all mitigate against glycosylation (Petrescu et al., 2004). In the natural GL-encoded site, accessibility of the Asn residue side-chain is certainly very low; its relative accessibility (the accessible surface area in the folded structure expressed as fraction of its total surface area when fully exposed) is only 8% as calculated with a 3 Å probe radius corresponding to an oligosaccharide unit and using the program NACCESS (http://wolf.bi.umist.ac.uk/unix/naccess.html) and the V_{4-34} crystal structure (Cauerhff et al., 2000; Hubbard and Thornton, 1993). Furthermore, there is a glutamic acid residue in the GL sequence at position −2. Introduced sites are more variably placed and apparently do allow addition of the oligosaccharide precursor but inhibit further processing. Although the Asn residue of the somatically generated Asn-Trp-Thr motif in the immediately adjacent CDR1 loop of FL14 is similarly low, there is no preceding acidic residue and Thr is at position +3. Preceding acidic residues are notably absent from 11 of the 12 glycosylated sites in the V_H genes of FL cells reported in this paper; however, it is likely that a combination of factors, such as the occurrence of nearby surface-exposed hydrophobic residues, which is difficult to assess in the absence of 3D structural data, may all contribute to determining the overall propensity for glycosylation (Petrescu et al., 2004).

B-cell malignancies located in the GC, including FL, tend to undergo continuing low-level somatic mutation (Stevenson et al., 2001). The V_{4-34} gene allowed a comparison of the fate of the introduced and natural GL-encoded *N*-glycosylation sites. We found a striking difference, with the GL-encoded site frequently lost during ongoing mutation while the acquired sites were conserved. Positive selection and conservation of the acquired oligomannosylated sites during intraclonal variation underlines the importance to the FL clones. Apparent random loss of the GL-encoded but non-glycosylated site in V_{4-34} indicates redundancy and focuses attention onto somatically mutated motifs. It appears that B-cell malignancies have been opportunistic in exploiting unprocessed glycosylation. The oligomannosylated V regions may be sensing a receptor, and if so, offer the possibility of therapeutic blockade.

Acknowledgements

This study was supported by Cancer Research UK and Tenovus UK. The glycan analysis was supported by the Oxford Glycobiology Institute endowment.

References

Cauerhff, A., Braden, B.C., Carvalho, J.G., Aparicio, R., Polikarpov, I., Leoni, J., Goldbaum, F.A., 2000. Three-dimensional structure of the Fab from a human IgM cold agglutinin. J. Immunol. 165, 6422–6428.

Gala, F.A., Morrison, S.L., 2004. V region carbohydrate and antibody expression. J. Immunol. 172, 5489–5494.

Hamblin, T.J., Davis, Z., Gardiner, A., Oscier, D.G., Stevenson, F.K., 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 94, 1848–1854.

Hawkins, R.E., Zhu, D., Ovecka, M., Winter, G., Hamblin, T.J., Long, A., Stevenson, F.K., 1994. Idiotypic vaccination against human B-cell lymphoma. Rescue of variable region gene sequences from biopsy material for assembly as single-chain Fv personal vaccines. Blood 83, 3279–3288.

Hubbard, S.J., Thornton, J.M., 1993. NACCESS, Department of Biochemistry and Molecular Biology, University College London.

Kasturi, L., Chen, H., Shakin-Eshleman, S.H., 1997. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem. J. 323 (Pt 2), 415–419.

McCann, K.J., Sahota, S.S., Stevenson, F.K., Ottensmeier, C.H., 2005. Idiotype gene rescue in Follicular Lymphoma. In: Illidge, T., Johnson, M.W. (Eds.), Methods in Molecular Medicine; Lymphoma: Methods and Protocols. Humana Press, Totowa, NJ.

McCann, K.J., Johnson, P.W., Stevenson, F.K., Ottensmeier, C.H., 2006. Universal *N*-glycosylation sites introduced into the B-cell receptor of follicular lymphoma by somatic mutation: a second tumorigenic event? Leukemia 20, 530–534.

Noppe, S.M., Heirman, C., Bakkus, M.H., Brissinck, J., Schots, R., Thielemans, K., 1999. The genetic variability of the VH genes in follicular lymphoma: the impact of the hypermutation mechanism. Br. J. Haematol. 107, 625–640.

Ottensmeier, C.H., Thompsett, A.R., Zhu, D., Wilkins, B.S., Sweetenham, J.W., Stevenson, F.K., 1998. Analysis of VH genes in follicular and diffuse lymphoma shows ongoing somatic mutation and multiple isotype transcripts in early disease with changes during disease progression. Blood 91, 4292–4299.

Petrescu, A.J., Milac, A.L., Petrescu, S.M., Dwek, R.A., Wormald, M.R., 2004. Statistical analysis of the protein environment of *N*-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology 14, 103–114.

Radcliffe, C.M., Arnold, J.N., Suter, D.M., Wormald, M.R., Harvey, D.J., Royle, L., Mimura, Y., Kimura, Y., Sim, R.B., Inoges, S., Rodriguez-Calvillo, M., Zabalegui, N., de Cerio, A.L., Potter, K.N., Mockridge, C.I., Dwek, R.A., Bendandi, M., Rudd, P.M., Stevenson, F.K., 2007. Human follicular lymphoma cells contain oligomannose glycans in the antigen-binding site of the B-cell receptor. J. Biol. Chem. 282, 7405–7415.

Stamatopoulos, K., Kosmas, C., Papadaki, T., Pouliou, E., Belessi, C., Afendaki, S., Anagnostou, D., Loukopoulos, D., 1997. Follicular lymphoma immunoglobulin kappa light chains are affected by the antigen selection process, but to a lesser degree than their partner heavy chains. Br. J. Haematol. 96, 132–146.

Stevenson, F.K., Sahota, S.S., Ottensmeier, C.H., Zhu, D., Forconi, F., Hamblin, T.J., 2001. The occurrence and significance of V gene mutations in B cell-derived human malignancy. Adv. Cancer Res. 83, 81–116.

Tarentino, A.L., Plummer Jr., T.H., Maley, F., 1974. The release of intact oligosaccharides from specific glycoproteins by endo-beta-*N*-acetylglucosaminidase H. J. Biol. Chem. 249, 818–824.

Zabalegui, N., de Cerio, A.L., Inoges, S., Rodriguez-Calvillo, M., Perez-Calvo, J., Hernandez, M., Garcia-Foncillas, J., Martin-Algarra, S., Rocha, E., Bendandi, M., 2004. Acquired potential *N*-glycosylation sites within the tumor-specific immunoglobulin heavy chains of B-cell malignancies. Haematologica 89, 541–546.

Zhu, D., McCarthy, H., Ottensmeier, C.H., Johnson, P., Hamblin, T.J., Stevenson, F.K., 2002. Acquisition of potential *N*-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. Blood 99, 2562–2568.

Zhu, D., Ottensmeier, C.H., Du, M.Q., McCarthy, H., Stevenson, F.K., 2003. Incidence of potential glycosylation sites in immunoglobulin variable regions distinguishes between subsets of Burkitt's lymphoma and mucosa-associated lymphoid tissue lymphoma. Br. J. Haematol. 120, 217–222.

References

1. Goldsby RA, Kindt TJ, Osborne BA. Kuby Immunology (ed 4th). New York: W. H. Freeman and Co; 2000.
2. Delves PJ, Roitt IM. The immune system. First of two parts. N Engl J Med. 2000;343:37-49.
3. Delves PJ, Roitt IM. The immune system. Second of two parts. N Engl J Med. 2000;343:108-117.
4. Parkin J, Cohen B. An overview of the immune system. Lancet. 2001;357:1777-1789.
5. Carayannopoulos L, Capra JD. Immunoglobulins: structure and function (ed 3rd). New York: Raven; 1993.
6. Bengten E, Wilson M, Miller N, Clem LW, Pilstrom L, Warr GW. Immunoglobulin isotypes: structure, function, and genetics. Curr Top Microbiol Immunol. 2000;248:189-219.
7. Donadel G, Calabro A, Sigounas G, Hascall VC, Notkins AL, Harindranath N. Human polyreactive and monoreactive antibodies: effect of glycosylation on antigen binding. Glycobiology. 1994;4:491-496.
8. Chuang PD, Morrison SL. Elimination of N-linked glycosylation sites from the human IgA1 constant region: effects on structure and function. J Immunol. 1997;158:724-732.
9. Wright A, Tao MH, Kabat EA, Morrison SL. Antibody variable region glycosylation: position effects on antigen binding and carbohydrate structure. Embo J. 1991;10:2717-2723.
10. Leung SO, Goldenberg DM, Dion AS, et al. Construction and characterization of a humanized, internalizing, B-cell (CD22)-specific, leukemia/lymphoma antibody, LL2. Mol Immunol. 1995;32:1413-1427.
11. Khurana S, Raghunathan V, Salunke DM. The variable domain glycosylation in a monoclonal antibody specific to GnRH modulates antigen binding. Biochem Biophys Res Commun. 1997;234:465-469.
12. Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK. Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. Blood. 2002;99:2562-2568.
13. Zhu D, Ottensmeier CH, Du MQ, McCarthy H, Stevenson FK. Incidence of potential glycosylation sites in immunoglobulin variable regions distinguishes between subsets of Burkitt's lymphoma and mucosa-associated lymphoid tissue lymphoma. Br J Haematol. 2003;120:217-222.
14. McBride OW, Battey J, Hollis GF, Swan DC, Siebenlist U, Leder P. Localization of human variable and constant region immunoglobulin heavy chain genes on subtelomeric band q32 of chromosome 14. Nucleic Acids Res. 1982;10:8155-8170.
15. Walter MA, Surti U, Hofker MH, Cox DW. The physical organization of the human immunoglobulin heavy chain gene complex. Embo J. 1990;9:3303-3313.
16. Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. The repertoire of human germline V(H) sequences reveals about fifty groups of V(H) segments with different hypervariable loops. J Mol Biol. 1992;227:776-798.
17. Cook GP, Tomlinson IM, Walter G, et al. A map of the human immunoglobulin V(H) locus completed by analysis of the telomeric region of chromosome 14q. Nature Genetics. 1994;7:162-168.

18. Kodaira M, Kinashi T, Umemura I, et al. Organization and evolution of variable region genes of the human immunoglobulin heavy chain. J Mol Biol. 1986;190:529-541.
19. Berman JE, Mellis SJ, Pollock R, et al. Content and organization of the human Ig VH locus: definition of three new VH families and linkage to the Ig CH locus. Embo J. 1988;7:727-738.
20. Corbett SJ, Tomlinson IM, Sonnhammer ELL, Buck D, Winter G. Sequence of the human immunoglobulin diversity (D) segment locus: A systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. J Mol Biol. 1997;270:587-597.
21. Ichihara Y, Matsuoka H, Kurosawa Y. Organization of human immunoglobulin heavy chain diversity gene loci. Embo J. 1988;7:4141-4150.
22. Ravetch JV, Siebenlist U, Korsmeyer S, Waldmann T, Leder P. Structure of the human immunoglobulin mu locus: characterization of embryonic and rearranged J and D genes. Cell. 1981;27:583-591.
23. Tomlinson IM, Cook GP, Carter NP, et al. Human immunoglobulin VH and D segments on chromosomes 15q11.2 and 16p11.2. Hum Mol Genet. 1994;3:853-860.
24. Zhang J, Alt FW, Honjo T. Regulation of class switch recombination of the immunoglobulin heavy chain genes. In: Honjo T, Alt FW, eds. Immunoglobulin Genes (ed 2nd). London: Academic Press; 1995:235-265.
25. Rothman P, Coffman RL. Immunoglobulin heavy chain class-switching. In: Herzenberg LA, Weir DM, Herzenberg L, Blackwell C, eds. Weir's handbook of Experimental immunology. Vol. I (ed 5). Cambridge, MA: Blackwell Science; 1997:19.11-19.14.
26. Malcolm S, Barton P, Murphy C, Ferguson-Smith MA, Bentley DL, Rabbitts TH. Localization of human immunoglobulin kappa light chain variable region genes to the short arm of chromosome 2 by in situ hybridization. Proc Natl Acad Sci U S A. 1982;79:4957-4961.
27. Schable KF, Zachau HG. The variable genes of the human immunoglobulin kappa locus. Biol Chem Hoppe Seyler. 1993;374:1001-1022.
28. Zachau HG. The human immunoglobulin κ genes. In: Honjo T, Alt FW, eds. Immunoglobulin Genes (ed 2nd). London: Academic Press; 1995:173-191.
29. Bentley DL, Rabbitts TH. Evolution of immunoglobulin V genes: evidence indicating that recently duplicated human V kappa sequences have diverged by gene conversion. Cell. 1983;32:181-189.
30. Pargent W, Meindl A, Thiebe R, Mitzel S, Zachau HG. The human immunoglobulin kappa locus. Characterization of the duplicated O regions. Eur J Immunol. 1991;21:1821-1827.
31. Lautner-Rieske A, Huber C, Meindl A, et al. The human immunoglobulin kappa locus. Characterization of the duplicated A regions. Eur J Immunol. 1992;22:1023-1029.
32. Huber C, Huber E, Lautner-Rieske A, Schable KF, Zachau HG. The human immunoglobulin kappa locus. Characterization of the partially duplicated L regions. Eur J Immunol. 1993;23:2860-2867.
33. Lefranc MP. Nomenclature of the human immunoglobulin kappa (IGK) genes. Exp Clin Immunogenet. 2001;18:161-174.
34. Barbie V, Lefranc MP. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp Clin Immunogenet. 1998;15:171-183.
35. de la Chapelle A, Lenoir G, Boue J, et al. Lambda Ig constant region genes are translocated to chromosome 8 in Burkitt's lymphoma with t(8;22). Nucleic Acids Res. 1983;11:1133-1142.

36. Williams SC, Frippiat JP, Tomlinson IM, Ignatovich O, Lefranca MP, Winter G. Sequence and evolution of the human germline Vlambda repertoire. J Mol Biol. 1996;264:220-232.
37. Vasicek TJ, Leder P. Structure and expression of the human immunoglobulin lambda genes. J Exp Med. 1990;172:609-620.
38. Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. Annu Rev Biochem. 2002;71:101-132.
39. Henderson A, Calame K. Transcriptional regulation during B cell development. Annu Rev Immunol. 1998;16:163-200.
40. Lin W-C, Desiderio S. V(D)J recombination and the cell cycle. Immunol Today. 1995;16:279-289.
41. Alt FW, Yancopoulos GD, Blackwell TK, et al. Ordered rearrangement of immunoglobulin heavy chain variable region segments. Embo J. 1984;3:1209-1219.
42. Raaphorst FM, Raman CS, Nall BT, Teale JM. Molecular mechanisms governing reading frame choice of immunoglobulin diversity genes. Immunol Today. 1997;18:37-43.
43. Rolink A, Melchers F. Molecular and cellular origins of B lymphocyte diversity. Cell. 1991;66:1081-1094.
44. Loffert D, Ehlich A, Muller W, Rajewsky K. Surrogate light chain expression is required to establish immunoglobulin heavy chain allelic exclusion during early B cell development. Immunity. 1996;4:133-144.
45. Picarella D, Serunian L, Rosenberg N. Allelic exclusion of membrane but not secreted Ig in a mature B cell line. Eur J Immunol. 1991;21:55-62.
46. Geier JK, Schlissel MS. Pre-BCR signals and the control of Ig gene rearrangements. Semin Immunol. 2006;18:31-39.
47. Graninger WB, Goldman PL, Morton CC, O'Brien SJ, Korsmeyer SJ. The kappa-deleting element. Germline and rearranged, duplicated and dispersed forms. J Exp Med. 1988;167:488-501.
48. Foster SJ, Brezinschek HP, Brezinschek RI, Lipsky PE. Molecular mechanisms and selective influences that shape the kappa gene repertoire of IgM(+) B cells. J Clin Invest. 1997;99:1614-1627.
49. Hood L, Gray WR, Sanders BG, Dreyer WY. Light chain evolution: antibodies. Cold Spring Harbor Symp Quant Biol. 1967;32:133-146.
50. Zou YR, Takeda S, Rajewsky K. Gene targeting in the Ig kappa locus: efficient generation of lambda chain-expressing B cells, independent of gene rearrangements in Ig kappa. EMBO J. 1993;12:811-820.
51. Pauza ME, Rehman JA, Tucker WL. Unusual patterns of immunoglobulin gene rearrangement and expression during human B cell ontogeny: human B cells can simultaneously express cell surface kappa and lambda light chains. J Exp Med. 1993;178:139-149.
52. Tiegs SL, Russell DM, Namazee D. Receptor editing in self-reactive bone marrow B cells. J Exp Med. 1993;177:1009-1020.
53. Radic MZ, Erikson J, Litwin S, Weigert M. B lymphocytes may escape tolerance by revising their antigen receptors. J Exp Med. 1993;177:1165-1173.
54. Chen C, Radic MZ, Erikson J, et al. Deletion and editing of B cells that express antibodies to DNA. J Immunol. 1994;152:1970-1982.
55. Tonegawa S. Somatic generation of antibody diversity. Nature. 1983;302:575-581.
56. Okazaki K, Davis DD, Sakano H. T cell receptor beta gene sequences in the circular DNA of thymocyte nuclei: direct evidence for intramolecular DNA deletion in V-D-J joining. Cell. 1987;49:477-485.

57. Zachau HG. The immunoglobulin kappa locus-or-what has been learned from looking closely at one-tenth of a percent of the human genome. Gene. 1993;135:167-173.
58. Meek KD, Hasemann CA, Capra JD. Novel rearrangements at the immunoglobulin D locus. Inversions and fusions add to IgH somatic diversity. J Exp Med. 1989;170:39-57.
59. Schatz DG, Oettinger MA, Schlissel MS. V(D)J recombination: molecular biology and regulation. Annu Rev Immunol. 1992;10:359-383.
60. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, Papaioannou VE. RAG-1-deficient mice have no mature B and T lymphocytes. Cell. 1992;68:869-877.
61. Cuomo CA, Oettinger MA. Analysis of regions of RAG-2 important for V(D)J recombination. Nucleic Acids Res. 1994;22:1810-1814.
62. Kirch SA, Sudarsanam P, Oettinger MA. Regions of RAG1 protein critical for V(D)J recombination. Eur J Immunol. 1996;26:886-891.
63. Hiom K, Gellert M. A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. Cell. 1997;88:65-72.
64. McBlane JF, van Gent DC, Ramsden DA, et al. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell. 1995;83:387-395.
65. van Gent DC, Hiom K, Paull TT, Gellert M. Stimulation of V(D)J cleavage by high mobility group proteins. Embo J. 1997;16:2665-2670.
66. van Gent DC, Ramsden DA, Gellert M. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell. 1996;85:107-113.
67. Hiom K, Gellert M. Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. Mol Cell. 1998;1:1011-1019.
68. Besmer E, Mansilla-Soto J, Cassard S, et al. Hairpin coding end opening is mediated by RAG1 and RAG2 proteins. Mol Cell. 1998;2:817-828.
69. Paull TT, Gellert M. Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. Genes Dev. 1999;13:1276-1288.
70. Moshous D, Callebaut I, de Chasseval R, et al. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. Cell. 2001;105:177-186.
71. Rooney S, Alt FW, Lombard D, et al. Defective DNA repair and increased genomic instability in Artemis-deficient murine cells. J Exp Med. 2003;197:553-565.
72. Roth DB, Menetski JP, Nakajima PB, Bosma MJ, Gellert M. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. Cell. 1992;70:983-991.
73. Lafaille JJ, DeCloux A, Bonneville M, Takagaki Y, Tonegawa S. Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. Cell. 1989;59:859-870.
74. Komori T, Okada A, Stewart V, Alt FW. Lack of N regions in antigen receptor variable region genes of TdT- deficient lymphocytes [published erratum appears in Science 1993 Dec 24;262(5142):1957]. Science. 1993;261:1171-1175.
75. Juul L, Hougs L, Andersen V, Svejgaard A, Barington T. The normally expressed kappa immunoglobulin light chain gene repertoire and somatic mutations studied by single-sided specific polymerase chain reaction (PCR); frequent occurrence of features often assigned to autoimmunity. Clin Exp Immunol. 1997;109:194-203.
76. Farner NL, T Dr, Lipsky PE. Molecular Mechanisms and Selection Influence the Generation of the Human VlambdaJlambda Repertoire. J Immunol. 1999;162:2137-2145.
77. Lieber MR, Hesse JE, Mizuuchi K, Gellert M. Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. Proc Natl Acad Sci U S A. 1988;85:8588-8592.

78. Zhu C, Roth DB. Characterization of coding ends in thymocytes of scid mice: implications for the mechanism of V(D)J recombination. Immunity. 1995;2:101-112.
79. Bosma MJ, Carroll AM. The SCID mouse mutant: definition, characterization, and potential uses. Annu Rev Immunol. 1991;9:323-350.
80. Nussenzweig A, Chen C, da Costa Soares V, et al. Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. Nature. 1996;382:551-555.
81. Critchlow SE, Bowater RP, Jackson SP. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. Curr Biol. 1997;7:588-598.
82. Taccioli GE, Rathbun G, Oltz E, Stamato T, Jeggo PA, Alt FW. Impairment of V(D)J recombination in double-strand break repair mutants. Science. 1993;260:207-210.
83. Ramsden DA, Gellert M. Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. Embo J. 1998;17:609-614.
84. Chen L, Trujillo K, Ramos W, Sung P, Tomkinson AE. Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. Mol Cell. 2001;8:1105-1115.
85. Lin WC, Desiderio S. Regulation of V(D)J recombination activator protein RAG-2 by phosphorylation. Science. 1993;260:953-959.
86. Kirch SA, Rathbun GA, Oettinger MA. Dual role of RAG2 in V(D)J recombination: catalysis and regulation of ordered Ig gene assembly. Embo J. 1998;17:4881-4886.
87. West KL, Singha NC, De Ioannes P, et al. A direct interaction between the RAG2 C terminus and the core histones is required for efficient V(D)J recombination. Immunity. 2005;23:203-212.
88. Schlissel MS, Baltimore D. Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. Cell. 1989;58:1001-1007.
89. Hsieh CL, Lieber MR. CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication. Embo J. 1992;11:315-325.
90. Sleckman BP, Gorman JR, Alt FW. Accessibility control of antigen-receptor variable-region gene assembly: role of cis-acting elements. Annu Rev Immunol. 1996;14:459-481.
91. Bolland DJ, Wood AL, Johnston CM, et al. Antisense intergenic transcription in V(D)J recombination. Nat Immunol. 2004;5:630-637.
92. Bassing CH, Swat W, Alt FW. The mechanism and regulation of chromosomal V(D)J recombination. Cell. 2002;109 Suppl:S45-55.
93. Romanow WJ, Langerak AW, Goebel P, et al. E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells. Mol Cell. 2000;5:343-353.
94. Yamada M, Wasserman R, Reichard BA, Shane S, Caton AJ, Rovera G. Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. J Exp Med. 1991;173:395-407.
95. Huang SC, Jiang R, Glas AM, Milner EC. Non-stochastic utilization of Ig V region genes in unselected human peripheral B cells. Mol Immunol. 1996;33:553-560.
96. Kraj P, Rao SP, Glas AM, Hardy RR, Milner EC, Silberstein LE. The human heavy chain Ig V region gene repertoire is biased at all stages of B cell ontogeny, including early pre-B cells. J Immunol. 1997;158:5824-5832.
97. Brezinschek HP, Foster SJ, Brezinschek RI, Dorner T, Domiati-Saad R, Lipsky PE. Analysis of the human VH gene repertoire. Differential effects of selection and somatic

hypermutation on human peripheral CD5(+)/IgM+ and CD5(-)/IgM+ B cells. J Clin Invest. 1997;99:2488-2501.
98. Ignatovich O, Tomlinson IM, Popov AV, Bruggemann M, Winter G. Dominance of intrinsic genetic factors in shaping the human immunoglobulin Vlambda repertoire. J Mol Biol. 1999;294:457-465.
99. Schroeder HWJ, Hillison JL, Perlmutter RJ. Early restriction of the human antibody repertoire. Science. 1987;238:791-?
100. Schroeder HJ, Wang JY. Preferential utilization of conserved immunoglobulin heavy chain variable gene segments during human fetal life. Proc Natl Acad Sci USA. 1990;87:6146-6150.
101. Deane M, Norton JD. Preferential rearrangement of developmentally regulated immunoglobulin VH1 genes in human B-lineage leukaemias. Leukemia. 1991;5:646-650.
102. Mortari F, Newton JA, Wang JY, Schroeder Hw J. The human cord blood antibody repertoire. Frequent usage of the V(H)7 gene family. Eur J Immunol. 1992;22:241-245.
103. Yancopoulos GD, Desiderio SV, Paskind M, Kearney JF, Baltimore D, Alt FW. Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. Nature. 1984;311:727-733.
104. Hesse JE, Lieber MR, Mizuuchi K, Gellert M. V(D)J recombination: a functional definition of the joining signals. Genes Dev. 1989;3:1053-1061.
105. Cox JPL, Tomlinson IM, Winter G. A directory of human germ-line V(kappa) segments reveals a strong bias in their usage. Eur J Immunol. 1994;24:827-836.
106. Akamatsu Y, Tsurushita N, Nagawa F, et al. Essential residues in V(D)J recombination signals. J Immunol. 1994;153:4520-4529.
107. Chen PP, Soto-Gil RW, Carson DA. The early expression of some human autoantibody-associated heavy chain variable region genes is controlled by specific regulatory elements. Scand J Immunol. 1990;31:673-678.
108. Stiernholm BJN, Berinstein NL. Mutations in immunoglobulin V gene promoters may cause reduced germline transcription and diminished recombination frequencies. Ann N Y Acad Sci. 1995:116-120.
109. Schroeder HW, Jr., Hillson JL, Perlmutter RM. Structure and evolution of mammalian VH families. Int Immunol. 1990;2:41-50.
110. Pascual V, Liu Y-J, Magalski A, de Bouteiller O, Banchereau J, Capra JD. Analysis of Somatic Mutation in five B Cell Subsets of Human Tonsil. J Exp Med. 1994;180:329-339.
111. Liu YJ, Malisan F, De BO, et al. Within germinal centers, isotype switching of immunoglobulin genes occurs after the onset of somatic mutation. Immunity 4. 1996:241-250.
112. Cerutti A, Zan H, Schaffer A, et al. CD40 ligand and appropriate cytokines induce switching to IgG, IgA, and IgE and coordinated germinal center and plasmacytoid phenotypic differentiation in a human monoclonal IgM+IgD+ B cell line. J Immunol. 1998;160:2145-2157.
113. Shan H, Schlomchik M, Weigert M. Heavy-chain class switch does not terminate somatic mutation. J Exp Med. 1990;172:531-536.
114. van Es JH, Meyling FH, Logtenberg T. High frequency of somatically mutated IgM molecules in the human adult blood B cell repertoire. Eur J Immunol. 1992;22:2761-2764.
115. Varade WS, Insel RA. Isolation of germinal centerlike events from human spleen RNA. Somatic hypermutation of a clonally related VH6DJH rearrangement expressed with IgM, IgG, and IgA. J Clin Invest. 1993;91:1838-1842.
116. Rajewsky K, Forster I, Cumano A. Evolutionary and somatic selection of the antibody repertoire in the mouse. Science. 1987;238:1088-1094.

117. Neuberger MS, Ehrenstein MR, Klix N, et al. Monitoring and interpreting the intrinsic features of somatic hypermutation. Immunol Rev. 1998;162:107-116.
118. Rada C, Yelamos J, Dean W, Milstein C. The 5' hypermutation boundary of kappa chains is independent of local and neighbouring sequences and related to the distance from the initiation of transcription. Eur J Immunol. 1997;27:3115-3120.
119. Motoyama N, Okada H, Azuma T. Somatic mutation in constant regions of mouse lambda 1 light chains. Proc Natl Acad Sci USA. 1991;88:7933-7937.
120. Betz AG, Rada C, Pannell R, Milstein C, Neuberger MS. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. Proc Natl Acad Sci U S A. 1993;90:2385-2388.
121. Betz AG, Neuberger MS, Milstein C. Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes. Immunol today. 1993;14:405-411.
122. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. J Immunol. 2004;172:3382-3384.
123. Wagner SD, Milstein C, Neuberger MS. Codon bias targets mutation [letter]. Nature. 1995;376:732.
124. Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim Biophys Acta. 1992;1171:11-18.
125. Kinoshita K, Honjo T. Linking class-switch recombination with somatic hypermutation. Nat Rev Mol Cell Biol. 2001;2:493-503.
126. Peters A, Storb U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. Immunity. 1996;4:57-65.
127. Fukita Y, Jacobs H, Rajewsky K. Somatic hypermutation in the heavy chain locus correlates with transcription. Immunity. 1998;9:105-114.
128. Bachl J, Carlson C, Gray-Schopfer V, Dessing M, Olsson C. Increased transcription levels induce higher mutation rates in a hypermutating cell line. J Immunol. 2001;166:5051-5057.
129. Jacobs H, Puglisi A, Rajewsky K, Fukita Y. Tuning somatic hypermutation by transcription. Curr Top Microbiol Immunol. 1999;246:149-158; discussion 159.
130. Betz AG, Milstein C, Gonzalez Fernandez A, Pannell R, Larson T, Neuberger MS. Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. Cell. 1994;77:239-248.
131. Klix N, Jolly CJ, Davies SL, Bruggemann M, Williams GT, Neuberger MS. Multiple sequences from downstream of the J kappa cluster can combine to recruit somatic hypermutation to a heterologous, upstream mutation domain. Eur J Immunol. 1998;28:317-326.
132. Azuma T, Motoyama N, Fields LE, Loh DY. Mutations of the chloramphenicol acetyl transferase transgene driven by the immunoglobulin promoter and intron enhancer. Int Immunol. 1993;5:121-130.
133. Yelamos J, Klix N, Goyenechea B, et al. Targeting of non-Ig sequences in place of the V segment by somatic hypermutation. Nature. 1995;376:225-229.
134. Stavnezer J. Molecular processes that regulate class switching. Curr Top Microbiol Immunol. 2000;245:127-168.
135. Luby TM, Schrader CE, Stavnezer J, Selsing E. The mu switch region tandem repeats are important, but not required, for antibody class switch recombination. J Exp Med. 2001;193:159-168.

136. Shinkura R, Tian M, Smith M, Chua K, Fujiwara Y, Alt FW. The influence of transcriptional orientation on endogenous switch region function. Nat Immunol. 2003;4:435-441.
137. Dunnick W, Hertz GZ, Scappino L, Gritzmacher C. DNA sequences at immunoglobulin switch region recombination sites. Nucleic Acids Res. 1993;21:365-372.
138. Kenter AL, Wuerffel R, Sen R, Jamieson CE, Merkulov GV. Switch recombination breakpoints occur at nonrandom positions in the S gamma tandem repeat. J Immunol. 1993;151:4718-4731.
139. Zhang K, Mills FC, Saxon A. Switch circles from IL-4-directed epsilon class switching from human B lymphocytes. Evidence for direct, sequential, and multiple step sequential switch from mu to epsilon Ig heavy chain gene. J Immunol. 1994;152:3427-3435.
140. Perlmutter AP, Gilbert W. Antibodies of the secondary response can be expressed without switch recombination in normal mouse B cells. Proc Natl Acad Sci USA. 1984;81:7189-7193.
141. Shimizu A, Nussenzweig MC, Han H, Sanchez M, Honjo T. Trans-splicing as a possible molecular mechanism for the multiple isotype expression of the immunoglobulin gene. J Exp Med. 1991;173:1385-1394.
142. Nolan-Willard M, Berton MT, Tucker P. Coexpression of mu and gamma 1 heavy chains can occur by a discontinuous transcription mechanism from the same unrearranged chromosome. Proc Natl Acad Sci USA. 1992;89:1234-1238.
143. Ott DE, Alt FW, Marcu KB. Immunoglobulin heavy chain switch region recombination within a retroviral vector in murine pre-B cells. Embo J. 1987;6:577-584.
144. Ott DE, Marcu KB. Molecular requirements for immunoglobulin heavy chain constant region gene switch-recombination revealed with switch-substrate retroviruses. Int Immunol. 1989;1:582-591.
145. Yancopoulos GD, DePinho RA, Zimmerman KA, Lutzker SG, Rosenberg N, Alt FW. Secondary genomic rearrangement events in pre-B cells: VHDJH replacement by a LINE-1 sequence and directed class switching. Embo J. 1986;5:3259-3266.
146. Stavnezer-Nordgren J, Sirlin S. Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. Embo J. 1986;5:95-102.
147. Reaban M, Griffin JA. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. Nature. 1990;348:342-344.
148. Reaban ME, Lebowitz J, Griffin JA. Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. J Biol Chem. 1994;269:21850-21857.
149. Tashiro J, Kinoshita K, Honjo T. Palindromic but not G-rich sequences are targets of class switch recombination. Int Immunol. 2001;13:495-505.
150. Manis JP, Tian M, Alt FW. Mechanism and control of class-switch recombination. Trends Immunol. 2002;23:31-39.
151. Bottaro A, Lansford R, Xu L, Zhang J, Rothman P, Alt FW. S region transcription per se promotes basal IgE class switch recombination but additional factors regulate the efficiency of the process. Embo J. 1994;13:665-674.
152. Lorenz M, Jung S, Radbruch A. Switch transcripts in immunoglobulin class switching. Science. 1995;267:1825-1828.
153. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature. 1988;334:364-366.
154. Lundgren M, Persson U, Larsson P, et al. Interleukin 4 induces synthesis of IgE and IgG4 in human B cells. Eur J Immunol. 1989;19:1311-1315.

155. Islam KB, Nilsson L, Sideras P, Hammarstrom L, Smith CI. TGF-beta 1 induces germ-line transcripts of both IgA subclasses in human B lymphocytes. Int Immunol. 1991;3:1099-1106.
156. McIntyre TM, Klinman DR, Rothman P, et al. Transforming growth factor beta 1 selectivity stimulates immunoglobulin G2b secretion by lipopolysaccharide-activated murine B cells. J Exp Med. 1993;177:1031-1037.
157. Snapper CM, Paul WE. Interferon-gamma and B cell stimulatory factor-1 reciprocally regulate Ig isotype production. Science. 1987;236:944-947.
158. Stavnezer J. Immunoglobulin class switching. Curr Opin Immunol. 1996;8:199-205.
159. Wuerffel R, Jamieson CE, Morgan L, Merkulov GV, Sen R, Kenter AL. Switch recombination breakpoints are strictly correlated with DNA recognition motifs for immunoglobulin S gamma 3 DNA-binding proteins. J Exp Med. 1992;176:339-349.
160. Williams M, Maizels N. LR1, a lipopolysaccharide-responsive factor with binding sites in the immunoglobulin switch regions and heavy-chain enhancer. Genes Dev. 1991;5:2353-2361.
161. Liao F, Birshtein BK, Busslinger M, Rothman P. The transcription factor BSAP (NF-HB) is essential for immunoglobulin germ-line epsilon transcription. J Immunol. 1994;152:2904-2911.
162. Neurath MF, Max EE, Strober W. Pax5 (BSAP) regulates the murine immunoglobulin 3' alpha enhancer by suppressing binding of NF-alpha P, a protein that controls heavy chain transcription. Proc Natl Acad Sci U S A. 1995;92:5336-5340.
163. Wuerffel RA, Kenter AL. Protein recognition motifs of S gamma 3 DNA are statistically correlated with switch recombination breakpoints. Curr Top Microbiol Immunol. 1992;182:149-156.
164. Warren WD, Berton MT. Induction of germ-line gamma 1 and epsilon Ig gene expression in murine B cells. IL-4 and the CD40 ligand-CD40 interaction provide distinct but synergistic signals. J Immunol. 1995;155:5637-5646.
165. Zelazowski P, Collins JT, Dunnick W, Snapper CM. Antigen receptor cross-linking differentially regulates germ-line CH ribonucleic acid expression in murine B cells. J Immunol. 1995;154:1223-1231.
166. Catalan N, Selz F, Imai K, Revy P, Fischer A, Durandy A. The block in immunoglobulin class switch recombination caused by activation-induced cytidine deaminase deficiency occurs prior to the generation of DNA double strand breaks in switch mu region. J Immunol. 2003;171:2504-2509.
167. Lansford R, Manis JP, Sonoda E, Rajewsky K, Alt FW. Ig heavy chain class switching in Rag-deficient mice [In Process Citation]. Int Immunol. 1998;10:325-332.
168. Hodgkin PD, Lee JH, Lyons AB. B cell differentiation and isotype switching is related to division cycle number. J Exp Med. 1996;184:277-281.
169. Kuzminov A. Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proc Natl Acad Sci U S A. 2001;98:8241-8246.
170. Bemark M, Sale JE, Kim HJ, Berek C, Cosgrove RA, Neuberger MS. Somatic hypermutation in the absence of DNA-dependent protein kinase catalytic subunit (DNA-PK(cs)) or recombination-activating gene (RAG)1 activity. J Exp Med. 2000;192:1509-1514.
171. Nagaoka H, Muramatsu M, Yamamura N, Kinoshita K, Honjo T. Activation-induced deaminase (AID)-directed hypermutation in the immunoglobulin Smu region: implication of AID involvement in a common step of class switch recombination and somatic hypermutation. J Exp Med. 2002;195:529-534.

172. Reina-San-Martin B, Difilippantonio S, Hanitsch L, Masilamani RF, Nussenzweig A, Nussenzweig MC. H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. J Exp Med. 2003;197:1767-1778.
173. Schrader CE, Bradley SP, Vardo J, Mochegova SN, Flanagan E, Stavnezer J. Mutations occur in the Ig Smu region but rarely in Sgamma regions prior to class switch recombination. Embo J. 2003;22:5893-5903.
174. Rush JS, Fugmann SD, Schatz DG. Staggered AID-dependent DNA double strand breaks are the predominant DNA lesions targeted to S mu in Ig class switch recombination. Int Immunol. 2004;16:549-557.
175. Rolink A, Melchers F, Andersson J. The SCID but not the RAG-2 gene product is required for S mu-S epsilon heavy chain class switching. Immunity. 1996;5:319-330.
176. Manis JP, Gu Y, Lansford R, et al. Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. J Exp Med. 1998;187:2081-2089.
177. Casellas R, Nussenzweig A, Wuerffel R, et al. Ku80 is required for immunoglobulin isotype switching. Embo J. 1998;17:2404-2411.
178. Manis JP, Dudley D, Kaylor L, Alt FW. IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells. Immunity. 2002;16:607-617.
179. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem. 1998;273:5858-5868.
180. Bassing CH, Alt FW. H2AX may function as an anchor to hold broken chromosomal DNA ends in close proximity. Cell Cycle. 2004;3:149-153.
181. Dempsey LA, Sun H, Hanakahi LA, Maizels N. G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. J Biol Chem. 1999;274:1066-1071.
182. Manis JP, Morales JC, Xia Z, Kutok JL, Alt FW, Carpenter PB. 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. Nat Immunol. 2004;5:481-487.
183. Kinoshita K, Harigai M, Fagarasan S, Muramatsu M, Honjo T. A hallmark of active class switch recombination: transcripts directed by I promoters on looped-out circular DNAs. Proc Natl Acad Sci U S A. 2001;98:12620-12623.
184. Muramatsu M, Sankaranand VS, Anant S, et al. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem. 1999;274:18470-18476.
185. Revy P, Muto T, Levy Y, et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell. 2000;102:565-575.
186. Arakawa H, Hauschild J, Buerstedde JM. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science. 2002;295:1301-1306.
187. Muto T, Muramatsu M, Taniwaki M, Kinoshita K, Honjo T. Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. Genomics. 2000;68:85-88.
188. Papavasiliou FN, Schatz DG. The activation-induced deaminase functions in a postcleavage step of the somatic hypermutation process. J Exp Med. 2002;195:1193-1198.
189. Navaratnam N, Morrison JR, Bhattacharya S, et al. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem. 1993;268:20709-20712.

190. Blanc V, Henderson JO, Kennedy S, Davidson NO. Mutagenesis of apobec-1 complementation factor reveals distinct domains that modulate RNA binding, protein-protein interaction with apobec-1, and complementation of C to U RNA-editing activity. J Biol Chem. 2001;276:46386-46393.
191. Shinkura R, Ito S, Begum NA, et al. Separate domains of AID are required for somatic hypermutation and class-switch recombination. Nat Immunol. 2004;5:707-712.
192. Ta VT, Nagaoka H, Catalan N, et al. AID mutant analyses indicate requirement for class-switch-specific cofactors. Nat Immunol. 2003;4:843-848.
193. Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Mol Cell. 2003;12:501-508.
194. Imai K, Catalan N, Plebani A, et al. Hyper-IgM syndrome type 4 with a B lymphocyte-intrinsic selective deficiency in Ig class-switch recombination. J Clin Invest. 2003;112:136-142.
195. Okazaki IM, Kinoshita K, Muramatsu M, Yoshikawa K, Honjo T. The AID enzyme induces class switch recombination in fibroblasts. Nature. 2002;416:340-345.
196. Yoshikawa K, Okazaki IM, Eto T, et al. AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. Science. 2002;296:2033-2036.
197. Martin A, Bardwell PD, Woo CJ, Fan M, Shulman MJ, Scharff MD. Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. Nature. 2002;415:802-806.
198. Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature. 2002;418:99-103.
199. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol Cell. 2002;10:1247-1253.
200. Petersen-Mahrt SK, Neuberger MS. In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). J Biol Chem. 2003;278:19583-19586.
201. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A. 2003;100:4102-4107.
202. Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. J Exp Med. 2003;197:1291-1296.
203. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature. 2003;422:726-730.
204. Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. Nat Immunol. 2003;4:452-456.
205. Daniels GA, Lieber MR. Transcription targets recombination at immunoglobulin switch sequences in a strand-specific manner. Curr Top Microbiol Immunol. 1996;217:171-189.
206. Nambu Y, Sugai M, Gonda H, et al. Transcription-coupled events associating with immunoglobulin switch region chromatin. Science. 2003;302:2137-2140.
207. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol. 2003;4:442-451.
208. Chaudhuri J, Khuong C, Alt FW. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature. 2004;430:992-998.

209. Milstein C, Neuberger MS, Staden R. Both DNA strands of antibody genes are hypermutation targets. Proc Natl Acad Sci U S A. 1998;95:8791-8794.
210. Shen HM, Storb U. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. Proc Natl Acad Sci U S A. 2004;101:12997-13002.
211. Shen HM, Ratnam S, Storb U. Targeting of the activation-induced cytosine deaminase is strongly influenced by the sequence and structure of the targeted DNA. Mol Cell Biol. 2005;25:10815-10821.
212. Shen HM, Michael N, Kim N, Storb U. The TATA binding protein, c-Myc and survivin genes are not somatically hypermutated, while Ig and BCL6 genes are hypermutated in human memory B cells. Int Immunol. 2000;12:1085-1093.
213. Storb U, Shen HM, Michael N, Kim N. Somatic hypermutation of immunoglobulin and non-immunoglobulin genes. Philos Trans R Soc Lond B Biol Sci. 2001;356:13-19.
214. Woo CJ, Martin A, Scharff MD. Induction of somatic hypermutation is associated with modifications in immunoglobulin variable region chromatin. Immunity. 2003;19:479-489.
215. Pasqualucci L, Neumeister P, Goossens T, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001;412:341-346.
216. Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol. 2004;4:541-552.
217. Rada C, Di Noia JM, Neuberger MS. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol Cell. 2004;16:163-171.
218. Chen X, Kinoshita K, Honjo T. Variable deletion and duplication at recombination junction ends: implication for staggered double-strand cleavage in class-switch recombination. Proc Natl Acad Sci U S A. 2001;98:13860-13865.
219. Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr Biol. 2002;12:1748-1755.
220. Pasqualucci L, Guglielmino R, Houldsworth J, et al. Expression of the AID protein in normal and neoplastic B cells. Blood. 2004;104:3318-3325.
221. Ito S, Nagaoka H, Shinkura R, et al. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. Proc Natl Acad Sci U S A. 2004;101:1975-1980.
222. Reynaud CA, Aoufouchi S, Faili A, Weill JC. What role for AID: mutator, or assembler of the immunoglobulin mutasome? Nat Immunol. 2003;4:631-638.
223. Basu U, Chaudhuri J, Alpert C, et al. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature. 2005;438:508-511.
224. Pasqualucci L, Kitaura Y, Gu H, Dalla-Favera R. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. Proc Natl Acad Sci U S A. 2006;103:395-400.
225. Faust EA, Saffran DC, Witte ON. Regulation of B-cell differentiation by stromal cells. In: Honjo T, Alt FW, eds. Immunoglobulin Genes (ed 2nd). London: Academic Press; 1995:103-128.
226. Nagasawa T. Microenvironmental niches in the bone marrow required for B-cell development. Nat Rev Immunol. 2006;6:107-116.
227. Rolink A, ten Boekel E, Melchers F, Fearon DT, Krop I, Andersson J. A subpopulation of B220+ cells in murine bone marrow does not express CD19 and contains natural killer cell progenitors. J Exp Med. 1996;183:187-194.

228. Diao J, Winter E, Chen W, Cantin C, Cattral MS. Characterization of distinct conventional and plasmacytoid dendritic cell-committed precursors in murine bone marrow. J Immunol. 2004;173:1826-1833.
229. Hagman J, Lukin K. Transcription factors drive B cell development. Curr Opin Immunol. 2006;18:127-134.
230. Chung JB, Silverman M, Monroe JG. Transitional B cells: step by step towards immune competence. Trends Immunol. 2003;24:343-349.
231. Sims GP, Ettinger R, Shirota Y, Yarboro CH, Illei GG, Lipsky PE. Identification and characterization of circulating human transitional B cells. Blood. 2005;105:4390-4398.
232. Marie-Cardine A, Divay F, Dutot I, et al. Transitional B cells in humans: Characterization and insight from B lymphocyte reconstitution after hematopoietic stem cell transplantation. Clin Immunol. 2008.
233. Cuss AK, Avery DT, Cannons JL, et al. Expansion of functionally immature transitional B cells is associated with human-immunodeficient states characterized by impaired humoral immunity. J Immunol. 2006;176:1506-1516.
234. Smith KG, Hewitson TD, Nossal GJ, Tarlinton DM. The phenotype and fate of the antibody-forming cells of the splenic foci. Eur J Immunol. 1996;26:444-448.
235. Bohnhorst JO, Bjorgan MB, Thoen JE, Natvig JB, Thompson KM. Bm1-Bm5 classification of peripheral blood B cells reveals circulating germinal center founder cells in healthy individuals and disturbance in the B cell subpopulations in patients with primary Sjogren's syndrome. J Immunol. 2001;167:3610-3618.
236. Liu YJ, Banchereau J. Human peripheral B-cell subsets. In: Herzenberg LA, Weir DM, Herzenberg L, Blackwell C, eds. Weir's handbook of Experimental immunology. Vol. III (ed 5). Cambridge, MA: Blackwell Science; 1997:93.91-93.99.
237. MacLennan IC. Germinal centers. Annu Rev Immunol. 1994;12:117-139.
238. Han S, Zheng B, Takahashi Y, Kelsoe G. Distinctive characteristics of germinal center B cells. Semin Immunol. 1997;9:255-260.
239. Jacob J, Przylepa J, Miller C, Kelsoe G. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. J Exp Med. 1993;178:1293-1307.
240. Kuppers R, Zhao M, Hansmann ML, Rajewsky K. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. EMBO Journal. 1993;12:4955-4967.
241. Milstein C, Neuberger MS. Maturation of the immune response. Adv Protein Chem. 1996;49:451-485.
242. Kosco Vilbois MH, Zentgraf H, Gerdes J, Bonnefoy JY. To 'B' or not to 'B' a germinal center? Immunol today. 1997;18:225-230.
243. van Nierop K, de Groot C. Human follicular dendritic cells: function, origin and development. Semin Immunol. 2002;14:251-257.
244. Szakal AK, Kosco MH, Tew JG. A novel in vivo follicular dendritic cell-dependent iccosome-mediated mechanism for delivery of antigen to antigen-processing cells. J Immunol. 1988;140.
245. Smith KG, Nossal GJ, Tarlinton DM. FAS is highly expressed in the germinal center but is not required for regulation of the B-cell response to antigen. Proc Natl Acad Sci U S A. 1995;92:11628-11632.
246. van Eijk M, Defrance T, Hennino A, de Groot C. Death-receptor contribution to the germinal-center reaction. Trends Immunol. 2001;22:677-682.
247. Kosco Vilbois MH, Gray D, Scheidegger D, Julius M. Follicular dendritic cells help resting B cells to become effective antigen-presenting cells: Induction of B7/BB1 and

upregulation of major histocompatibility complex class II molecules. J Exp Med. 1993;178:2055-2066.
248. Maeda K, Kosco Vilbois MH, Burton GF, Szakal AK, Tew JG. Expression of the intercellular adhesion molecule-1 on high endothelial venules and on non-lymphoid antigen handling cells: Interdigitating cells, antigen transporting cells and follicular dendritic cells. Cell Tissue Res. 1995;279:47-54.
249. Han S, Dillon SR, Zheng B, Shimoda M, Schlissel MS, Kelsoe G. V(D)J recombinase activity in a subset of germinal center B lymphocytes. Science. 1997;278:301-305.
250. Giachino C, Padovan E, Lanzavecchia A. Re-expression of RAG-1 and RAG-2 genes and evidence for secondary rearrangements in human germinal center B lymphocytes. Eur J Immunol. 1998;28:3506-3513.
251. Lindhout E, van Eijk M, van Pel M, Lindeman J, Dinant HJ, de Groot C. Fibroblast-like synoviocytes from rheumatoid arthritis patients have intrinsic properties of follicular dendritic cells. J Immunol. 1999;162:5949-5956.
252. Smith KG, Weiss U, Rajewsky K, Nossal GJ, Tarlinton DM. Bcl-2 increases memory B cell recruitment but does not perturb selection in germinal centers. Immunity. 1994;1:803-813.
253. Hargreaves DC, Hyman PL, Lu TT, et al. A coordinated change in chemokine responsiveness guides plasma cell movements. J Exp Med. 2001;194:45-56.
254. Jego G, Robillard N, Puthier D, et al. Reactive plasmacytoses are expansions of plasmablasts retaining the capacity to differentiate into plasma cells. Blood. 1999;94:701-712.
255. Kawano MM, Mihara K, Huang N, Tsujimoto T, Kuramoto A. Differentiation of early plasma cells on bone marrow stromal cells requires interleukin-6 for escaping from apoptosis. Blood. 1995;85:487-494.
256. Bossen C, Schneider P. BAFF, APRIL and their receptors: structure, function and signaling. Semin Immunol. 2006;18:263-275.
257. O'Connor BP, Cascalho M, Noelle RJ. Short-lived and long-lived bone marrow plasma cells are derived from a novel precursor population. J Exp Med. 2002;195:737-745.
258. Maruyama M, Lam KP, Rajewsky K. Memory B-cell persistence is independent of persisting immunizing antigen. Nature. 2000;407:636-642.
259. Arpin C, Banchereau J, Liu YJ. Memory B cells are biased towards terminal differentiation: a strategy that may prevent repertoire freezing. J Exp Med. 1997;186:931-940.
260. McHeyzer-Williams MG, Ahmed R. B cell memory and the long-lived plasma cell. Curr Opin Immunol. 1999;11:172-179.
261. Martin SW, Goodnow CC. Burst-enhancing role of the IgG membrane tail as a molecular determinant of memory. Nat Immunol. 2002;3:182-188.
262. Angelin-Duclos C, Cattoretti G, Lin KI, Calame K. Commitment of B lymphocytes to a plasma cell fate is associated with Blimp-1 expression in vivo. J Immunol. 2000;165:5462-5471.
263. Shaffer AL, Lin KI, Kuo TC, et al. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. Immunity. 2002;17:51-62.
264. Calame KL, Lin KI, Tunyaplin C. Regulatory mechanisms that determine the development and function of plasma cells. Annu Rev Immunol. 2003;21:205-230.
265. Reimold AM, Iwakoshi NN, Manis J, et al. Plasma cell differentiation requires the transcription factor XBP-1. Nature. 2001;412:300-307.
266. Gray D, Siepmann K, Wohlleben G. CD40 ligation in B cell activation, isotype switching and memory development. Semin Immunol. 1994;6:303-310.

267. Randall TD, Heath AW, Santos-Argumedo L, Howard MC, Weissman IL, Lund FE. Arrest of B lymphocyte terminal differentiation by CD40 signaling: mechanism for lack of antibody-secreting cells in germinal centers. Immunity. 1998;8:733-742.
268. Choe J, Kim HS, Armitage RJ, Choi YS. The functional role of B cell antigen receptor stimulation and IL-4 in the generation of human memory B cells from germinal center B cells. J Immunol. 1997;159:3757-3766.
269. Kelsoe G. Life and death in germinal centers (redux). Immunity. 1996;4:107-111.
270. Burton GF, Kosco MH, Szakal AK, Tew JG. Iccosomes and the secondary antibody response. Immunology. 1991;73:271-276.
271. Suzuki I, Pfister L, Glas A, Nottenburg C, Milner EC. Representation of rearranged VH gene segments in the human adult antibody repertoire. J Immunol. 1995;154:3902-3911.
272. Davidkova G, Pettersson S, Holmberg D, Lundkvist I. Selective usage of VH genes in adult human B lymphocyte repertoires. Scand J Immunol. 1997;45:62-73.
273. Ignatovich O, Tomlinson IM, Jones PT, Winter G. The creation of diversity in the human immunoglobulin V(lambda) repertoire. J Mol Biol. 1997;268:69-77.
274. Chapman C, Spellerberg M, Hamblin T, Stevenson F. Pattern of usage of the V(H)4-21 gene by B lymphocytes in a patient with EBV infection indicates ongoing mutation and class switching. Ann N Y Acad Sci. 1995:576-579.
275. van Es JH, Gmelig Meyling FH, van de Akker WR, Aanstoot H, Derksen RH, Logtenberg T. Somatic mutations in the variable regions of a human IgG anti-double-stranded DNA autoantibody suggest a role for antigen in the induction of systemic lupus erythematosus. J Exp Med. 1991;173:461-470.
276. Dighiero G. Autoantibody activity and V gene usage by B-cell malignancies. Leuk Lymphoma. 1992;8:345-351.
277. Mockridge CI, Rahman A, Buchan S, et al. Common patterns of B cell perturbation and expanded V4-34 immunoglobulin gene usage in autoimmunity and infection. Autoimmunity. 2004;37:9-15.
278. Potter KN, Hobby P, Klijn S, Stevenson FK, Sutton BJ. Evidence for involvement of a hydrophobic patch in framework region 1 of human v4-34-encoded igs in recognition of the red blood cell I antigen. J Immunol. 2002;169:3777-3782.
279. Silverman GJ, Sasano M, Wormsley SB. Age-associated changes in binding of human B lymphocytes to a VH3-restricted unconventional bacterial antigen. J Immunol. 1993;151:5840-5855.
280. Potter KN, Li Y, Capra JD. Staphylococcal protein A simultaneously interacts with framework region 1, complementarity-determining region 2, and framework region 3 on human VH3-encoded Igs. J Immunol. 1996;157:2982-2988.
281. Graille M, Stura EA, Corper AL, et al. Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. Proc Natl Acad Sci U S A. 2000;97:5399-5404.
282. Pearson PL, Van der Luijt RB. The genetic analysis of cancer. J Intern Med. 1998;243:413-417.
283. Knudson AG. Hereditary predisposition to cancer. Ann N Y Acad Sci. 1997;833:58-67.
284. Willis TG, Dyer MJ. The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies. Blood. 2000;96:808-822.
285. Dyer MJ, Lillington DM, Bastard C, et al. Concurrent activation of MYC and BCL2 in B cell non-Hodgkin lymphoma cell lines by translocation of both oncogenes to the same immunoglobulin heavy chain locus. Leukemia. 1996;10:1198-1208.

286. Harris NL, Jaffe ES, Diebold J, et al. World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. J Clin Oncol. 1999;17:3835-3849.
287. Pasqualucci L, Migliazza A, Fracchiolla N, et al. BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. Proc Natl Acad Sci U S A. 1998;95:11816-11821.
288. Shen HM, Peters A, Baron B, Zhu X, Storb U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science. 1998;280:1750-1752.
289. Montesinos-Rongen M, Van Roost D, Schaller C, Wiestler OD, Deckert M. Primary diffuse large B-cell lymphomas of the central nervous system are targeted by aberrant somatic hypermutation. Blood. 2004;103:1869-1875.
290. Johnston JM, Carroll WL. c-myc hypermutation in Burkitt's lymphoma. Leuk Lymphoma. 1992;8:431-439.
291. Limpens J, Stad R, Vos C, et al. Lymphoma-associated translocation t(14;18) in blood B cells of normal individuals. Blood. 1995;85:2528-2536.
292. Fuscoe JC, Setzer RW, Collard DD, Moore MM. Quantification of t(14;18) in the lymphocytes of healthy adult humans as a possible biomarker for environmental exposures to carcinogens. Carcinogenesis. 1996;17:1013-1020.
293. Yunis JJ, Frizzera G, Oken MM, McKenna J, Theologides A, Arnesen M. Multiple recurrent genomic defects in follicular lymphoma. A possible model for cancer. N Engl J Med. 1987;316:79-84.
294. Tsujimoto Y, Gorham J, Cossman J, Jaffe E, Croce CM. The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. Science. 1985;229:1390-1393.
295. Raffeld M, Jaffe ES. bcl-1, t(11;14), and mantle cell-derived lymphomas. Blood. 1991;78:259-263.
296. Welzel N, Le T, Marculescu R, et al. Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. Cancer Res. 2001;61:1629-1636.
297. Jiang W, Kahn SM, Zhou P, et al. Overexpression of cyclin D1 in rat fibroblasts causes abnormalities in growth control, cell cycle progression and gene expression. Oncogene. 1993;8:3447-3457.
298. Hashimoto Y, Nakamura N, Kuze T, Abe M, Wakasa H. Intranuclear expression of cyclin D1 protein as a useful prognostic marker for mantle cell lymphoma. Fukushima J Med Sci. 1997;43:87-98.
299. Willis TG, Jadayel DM, Du MQ, et al. Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. Cell. 1999;96:35-45.
300. Zhang Q, Siebert R, Yan M, et al. Inactivating mutations and overexpression of BCL10, a caspase recruitment domain-containing gene, in MALT lymphoma with t(1;14)(p22;q32). Nat Genet. 1999;22:63-68.
301. Offit K, Parsa NZ, Filippa D, Jhanwar SC, Chaganti RS. t(9;14)(p13;q32) denotes a subset of low-grade non-Hodgkin's lymphoma with plasmacytoid differentiation. Blood. 1992;80:2594-2599.
302. Iida S, Rao PH, Nallasivam P, et al. The t(9;14)(p13;q32) chromosomal translocation associated with lymphoplasmacytoid lymphoma involves the PAX-5 gene. Blood. 1996;88:4110-4117.
303. Stamatopoulos K, Kosmas C, Belessi C, et al. t(14;18) chromosomal translocation in follicular lymphoma: an event occurring with almost equal frequency both at the D to

J(H) and at later stages in the rearrangement process of the immunoglobulin heavy chain gene locus. Br J Haematol. 1997;99:866-872.
304. Kuppers R, Goossens T, Klein U. The role of somatic hypermutation in the generation of deletions and duplications in human Ig V region genes and chromosomal translocations. Curr Top Microbiol Immunol. 1999;246:193-198.
305. Kluin PM, Van Krieken J. The molecular biology of B-cell lymphoma: Clinicopathologic implications. Ann Hematol. 1991;62:95-102.
306. Haluska FG, Finver S, Tsujimoto Y, Croce CM. The t(8; 14) chromosomal translocation occurring in B-cell malignancies results from mistakes in V-D-J joining. Nature. 1986;324:158-161.
307. Oettinger MA, Schatz DG, Gorka C, Baltimore D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science. 1990;248:1517-1523.
308. Kennedy AK, Guhathakurta A, Kleckner N, Haniford DB. Tn10 transposition via a DNA hairpin intermediate. Cell. 1998;95:125-134.
309. Hiom K, Melek M, Gellert M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell. 1998;94:463-470.
310. Vaandrager JW, Schuuring E, Philippo K, Kluin PM. V(D)J recombinase-mediated transposition of the BCL2 gene to the IGH locus in follicular lymphoma. Blood. 2000;96:1947-1952.
311. Okazaki IM, Hiai H, Kakazu N, et al. Constitutive expression of AID leads to tumorigenesis. J Exp Med. 2003;197:1173-1181.
312. Wilks S. Cases of enlargement of lymphatic glands and spleen (or Hodgkin's disease). Guy's Hosp Rep. 1865;11:56.
313. Ottensmeier C. The classification of lymphomas and leukemias. Chem Biol Interact. 2001;135-136:653-664.
314. Rappaport H. Tumors of the hematopoietic system. In: Atlas of Tumor Pathology Washinghton DC: Armed Forces Institute of Pathology. 1966;Sec. 3, Fasc.8.
315. Lukes RI, Collins FD. Immunological characterization of human malignant lymphomas. Cancer. 1974;22:1087-1088.
316. Lennert K, Mohri N. Histopathology and diagnosis of non-Hodgkin's lymphomas. In: Lennert K, ed. Malignant Lymphomas Other than Hodgkin's Disease. Berlin: Springer Verlag; 1978:284-302.
317. The Non-Hodgkin's Lymphoma Pathological Classification Project. NCI sponsored study of classifications of non Hodgkin's Lymphoma: summary and description of a working formulation for clinical usage. Cancer. 1982;49:2112-2135.
318. Harris NL, Jaffe ES, Stein H, et al. A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. Blood. 1994;84:1361-1392.
319. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J. Lymphoma classification--from controversy to consensus: the R.E.A.L. and WHO Classification of lymphoid neoplasms. Ann Oncol. 2000;11:3-10.
320. Stevenson F, Sahota S, Zhu D, et al. Insight into the origin and clonal history of B-cell tumors as revealed by analysis of immunoglobulin variable region genes. Immunol Rev. 1998;162:247-259.
321. Stevenson FK, Sahota SS, Ottensmeier CH, Zhu D, Forconi F, Hamblin TJ. The occurrence and significance of V gene mutations in B cell-derived human malignancy. Adv Cancer Res. 2001;83:81-116.
322. Bird J, Galili N, Link M, Stites D, Sklar J. Continuing rearrangement but absence of somatic hypermutation in immunoglobulin genes of human B cell precursor leukemia. J Exp Med. 1988;168:229-245.

323. Kawakami K, Kita K, Miwa H, et al. Rearrangement patterns of immunoglobulin heavy chain (IgH) and light chain genes in acute lymphoblastic leukemia and chronic myelocytic leukemia lymphoid crisis cells showing oligoclonal IgH gene rearrangements. Int J Hematol. 1992;55:61-69.
324. Davi F, Gocke C, Smith S, Sklar J. Lymphocytic progenitor cell origin and clonal evolution of human B-lineage acute lymphoblastic leukemia. Blood. 1996;88:609-621.
325. Stolz F, Panzer S, Fischer S, Panzer-Grumayer ER. Oligoclonal immunoglobulin heavy-chain and T-cell receptor delta rearrangements persist in a recurrent acute lymphoblastic leukemia with one immunoglobulin kappa rearrangement as a clonal marker. Mod Pathol. 1999;12:819-826.
326. Szczepanski T, Pongers-Willemse MJ, Langerak AW, van Dongen JJ. Unusual immunoglobulin and T-cell receptor gene rearrangement patterns in acute lymphoblastic leukemias. Curr Top Microbiol Immunol. 1999;246:205-213; discussion 214-205.
327. Liu Y, Zhu P, Hu YM. [Characteristics of immunoglobulin heavy chain variable region genes in childhood B-cell acute lymphoblastic leukemia]. Zhonghua Xue Ye Xue Za Zhi. 2004;25:8-12.
328. Li A, Rue M, Zhou J, et al. Utilization of Ig heavy chain variable, diversity, and joining gene segments in children with B-lineage acute lymphoblastic leukemia: implications for the mechanisms of VDJ recombination and for pathogenesis. Blood. 2004;103:4602-4609.
329. Beishuizen A, Verhoeven MA, Hahlen K, van Wering ER, van Dongen JJ. Differences in immunoglobulin heavy chain gene rearrangmeent patterns between bone marrow and blood samples in childhood precursor B-acute lymphoblastic leaukemia at diagnosis [corrected and republished with original paging, article originally printed in Leukemia 1992 Jan;6(1):60-3]. Leukemia. 1993;7:60-63.
330. Height SE, Swansbury GJ, Matutes E, Treleaven JG, Catovsky D, Dyer MJ. Analysis of clonal rearrangements of the Ig heavy chain locus in acute leukemia. Blood. 1996;87:5242-5250.
331. Beishuizen A, Hahlen K, Hagemeijer A, et al. Multiple rearranged immunoglobulin genes in childhood acute lymphoblastic leukemia of precursor B-cell origin. Leukemia. 1991;5:657-667.
332. Wasserman R, Galili N, Ito Y, Reichard BA, Shane S, Rovera G. Predominance of fetal type DJH joining in young children with B precursor lymphoblastic leukemia as evidence for an in utero transforming event. J Exp Med. 1992;176:1577-1581.
333. Mortuza FY, Papaioannou M, Moreira IM, et al. Minimal residual disease tests provide an independent predictor of clinical outcome in adult acute lymphoblastic leukemia. J Clin Oncol. 2002;20:1094-1104.
334. Berman JE, Nickerson KG, Pollock RR, et al. VH gene usage in humans: biased usage of the VH6 gene in immature B lymphoid cells. Eur J Immunol. 1991;21:1311-1314.
335. Kraj P, Friedman DF, Stevenson F, Silberstein LE. Evidence for the overexpression of the VH4-34 (VH4.21) Ig gene segment in the normal adult human peripheral blood B cell repertoire. J Immunol. 1995;154:6406-6420.
336. Mori H, Kiyoi H, Horibe K, Ohno R, Naoe T. Comparison of the immunoglobulin gene transcripts between immature B lineage acute lymphoblastic leukemia and the normal phenotypic counterparts in the bone marrow. Leukemia. 1997;11:1274-1280.
337. Ebeling SB, Schutte MEM, AkkermansKoolhaas KE, Bloem AC, GmeligMeyling FHJ, Logtenberg T. Expression of members of the immunoglobulin V(H)3 gene families is not restricted at the level of individual genes in human chronic lymphocytic leukemia. International Immunology. 1992;4:313-320.

338. Kipps TJ, Tomhave E, Pratt LF, Duffy S, Chen PP, Carson DA. Developmentally restricted immunoglobulin heavy chain variable region gene expressed at high frequency in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 1989;86:5913-5917.
339. Schroeder HJ, Dighiero G. The pathogenesis of chronic lymphocytic leukemia: Analysis of the antibody repertoire. Immunol Today. 1994;15:288-294.
340. Maloum K, Pritsch O, Magnac C, et al. VH gene expression in CD5 positive and CD5 negative B cell chronic lymphoid malignancies. Leuk Lymphoma. 1997;24:437-448.
341. Oscier DG, Thompsett A, Zhu D, Stevenson FK. Differential rates of somatic hypermutation in V(H) genes among subsets of chronic lymphocytic leukemia defined by chromosomal abnormalities. Blood. 1997;89:4153-4160.
342. Logtenberg T, Schutte ME, Inghirami G, et al. Immunoglobulin VH gene expression in human B cell lines and tumors: biased VH gene expression in chronic lymphocytic leukemia. Int Immunol. 1989;1:362-366.
343. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999;94:1848-1854.
344. Tobin G, Thunberg U, Johnson A, et al. Somatically mutated Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. Blood. 2002;99:2262-2264.
345. Bomben R, Dal Bo M, Capello D, et al. Comprehensive characterization of IGHV3-21-expressing B-cell chronic lymphocytic leukemia: an Italian multicenter study. Blood. 2007;109:2989-2998.
346. Johnson TA, Rassenti LZ, Kipps TJ. Ig VH1 genes expressed in B cell chronic lymphocytic leukemia exhibit distinctive molecular features. J Immunol. 1997;158:235-246.
347. Potter KN, Orchard J, Critchley E, Mockridge CI, Jose A, Stevenson FK. Features of the overexpressed V1-69 genes in the unmutated subset of chronic lymphocytic leukemia are distinct from those in the healthy elderly repertoire. Blood. 2003;101:3082-3084.
348. Matthews C, Catherwood MA, Morris TC, Alexander HD. V(H)3-48 and V(H)3-53, as well as V(H)3-21, gene rearrangements define unique subgroups in CLL and are associated with biased lambda light chain restriction, homologous LCDR3 sequences and poor prognosis. Leuk Res. 2006.
349. Ebeling SB, Schutte MEM, Logtenberg T. Molecular analysis of V(H) and V(L) regions expressed in IgG-bearing chronic lymphocytic leukemia (CLL): Further evidence that CLL is a heterogeneous group of tumors. Blood. 1993;82:1626-1631.
350. Malisan F, Fluckiger AC, Ho S, Guret C, Banchereau J, MartinezValdez H. B-Chronic lymphocytic leukemias can undergo isotype switching in vivo and can be induced to differentiate and switch in vitro. Blood. 1996;87:717-724.
351. Efremov DG, Ivanovski M, Batista FD, Pozzato G, Burrone OR. IgM-producing chronic lymphocytic leukemia cells undergo immunoglobulin isotype-switching without acquiring somatic mutations. J Clin Invest. 1996;98:290-298.
352. Cerutti A, Zan H, Kim EC, et al. Ongoing in vivo immunoglobulin class switch DNA recombination in chronic lymphocytic leukemia B cells. J Immunol. 2002;169:6594-6603.
353. Bagnara D, Callea V, Stelitano C, et al. IgV gene intraclonal diversification and clonal evolution in B-cell chronic lymphocytic leukaemia. Br J Haematol. 2006;133:50-58.
354. Volkheimer AD, Weinberg JB, Beasley BE, et al. Progressive immunoglobulin gene mutations in chronic lymphocytic leukemia: evidence for antigen-driven intraclonal diversification. Blood. 2007;109:1559-1567.

355. Konikova E, Babusikova O, Mesarosova A, Kusenda J, Glasova M. Cytoplasmic and surface membrane phenotypic markers in cells of B cell chronic lymphocytic leukemia. Neoplasma. 1994;41:69-74.
356. Klein U, Tu Y, Stolovitzky GA, et al. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. J Exp Med. 2001;194:1625-1638.
357. Hummel M, Tamaru J, Kalvelage B, Stein H. Mantle cell (previously centrocytic) lymphomas express VH genes with no or very little somatic mutations like the physiologic cells of the follicle mantle. Blood. 1994;84:403-407.
358. Thorselius M, Walsh S, Eriksson I, et al. Somatic hypermutation and VH gene usage in mantle cell lymphoma. Eur J Haematol. 2002;68:217-224.
359. Walsh SH, Thorselius M, Johnson A, et al. Mutated VH genes and preferential VH3-21 use define new subsets of mantle cell lymphoma. Blood. 2003;101:4047-4054.
360. Orchard J, Garand R, Davis Z, et al. A subset of t(11;14) lymphoma with mantle cell features displays mutated IgVH genes and includes patients with good prognosis, nonnodal disease. Blood. 2003;101:4975-4981.
361. Kienle D, Krober A, Katzenberger T, et al. VH mutation status and VDJ rearrangement structure in mantle cell lymphoma: correlation with genomic aberrations, clinical characteristics, and outcome. Blood. 2003;102:3003-3009.
362. Camacho FI, Algara P, Rodriguez A, et al. Molecular heterogeneity in MCL defined by the use of specific VH genes and the frequency of somatic mutations. Blood. 2003;101:4042-4046.
363. Babbage G, Garand R, Robillard N, Zojer N, Stevenson FK, Sahota SS. Mantle cell lymphoma with t(11;14) and unmutated or mutated VH genes expresses AID and undergoes isotype switch events. Blood. 2004;103:2795-2798.
364. Bahler DW, Campbell MJ, Hart S, Miller RA, Levy S, Levy R. Ig VH gene expression among human follicular lymphomas. Blood. 1991;78:1561-1568.
365. Zelenetz AD, Chen TT, Levy R. Clonal expansion in follicular lymphoma occurs subsequent to antigenic selection. J Exp Med. 1992;176:1137-1148.
366. Zhu D, Hawkins RE, Hamblin TJ, Stevenson FK. Clonal history of a human follicular lymphoma as revealed in the immunoglobulin variable region genes. Br J Haematol. 1994;86:505-512.
367. Hsu FJ, Levy R. Preferential use of the VH4 Ig gene family by diffuse large-cell lymphoma. Blood. 1995;86:3072-3082.
368. Stamatopoulos K, Kosmas C, Papadaki T, et al. Follicular lymphoma immunoglobulin kappa light chains are affected by the antigen selection process, but to a lesser degree than their partner heavy chains. Br J Haematol. 1997;96:132-146.
369. Noppe SM, Heirman C, Bakkus MH, Brissinck J, Schots R, Thielemans K. The genetic variability of the VH genes in follicular lymphoma: the impact of the hypermutation mechanism. Br J Haematol. 1999;107:625-640.
370. Aarts WM, Bende RJ, Steenbergen EJ, et al. Variable heavy chain gene analysis of follicular lymphomas: correlation between heavy chain isotype expression and somatic mutation load. Blood. 2000;95:2922-2929.
371. Ottensmeier CH, Thompsett AR, Zhu D, Wilkins BS, Sweetenham JW, Stevenson FK. Analysis of VH genes in follicular and diffuse lymphoma shows ongoing somatic mutation and multiple isotype transcripts in early disease with changes during disease progression. Blood. 1998;91:4292-4299.
372. Tamaru J, Hummel M, Marafioti T, et al. Burkitt's lymphomas express VH genes with a moderate number of antigen-selected somatic mutations. Am J Pathol. 1995;147:1398-1407.

373. Chapman CJ, Mockridge CI, Rowe M, Rickinson AB, Stevenson FK. Analysis of V(H) genes used by neoplastic B cells in endemic Burkitt's lymphoma shows somatic hypermutation and intraclonal heterogeneity. Blood. 1995;85:2176-2181.
374. Klein U, Klein G, Ehlin-Henriksson B, Rajewsky K, Kuppers R. Burkitt's lymphoma is a malignancy of mature B cells expressing somatically mutated V region genes. Mol Med. 1995;1:495-505.
375. Chapman CJ, Zhou JX, Gregory C, Rickinson AB, Stevenson FK. V(H) and V(L) gene analysis in sporadic Burkitt's lymphoma shows somatic hypermutation, intraclonal heterogeneity, and a role for antigen selection. Blood. 1996;88:3562-3568.
376. Isobe K, Tamaru J, Nakamura S, Harigaya K, Mikata A, Ito H. VH gene analysis in sporadic Burkitt's lymphoma: somatic mutation and intraclonal diversity with special reference to the tumor cells involving germinal center. Leuk Lymphoma. 2002;43:159-164.
377. Riboldi P, Ikematsu W, Brambilla B, Caprani C, Gerosa M, Casali P. Diversity and somatic hypermutation of the Ig VHDJH, V kappa J kappa, and V lambda J lambda gene segments in lymphoma B cells: relevance to the origin of the neoplastic B cell clone. Hum Immunol. 2003;64:69-81.
378. Bellan C, Lazzi S, Hummel M, et al. Immunoglobulin gene analysis reveals 2 distinct cells of origin for EBV-positive and EBV-negative Burkitt lymphomas. Blood. 2005;106:1031-1036.
379. Ikematsu H, Cerutti A, Zan H, Knowles DM, Ikematsu W, Casali P. Ongoing hypermutation in the Ig V(D)J gene segments and c-myc proto-oncogene of an AIDS lymphoma segregates with neoplastic B cells at different sites: implications for clonal evolution. Hum Immunol. 2000;61:1242-1253.
380. Babcock GJ, Decker LL, Volk M, Thorley-Lawson DA. EBV persistence in memory B cells in vivo. Immunity. 1998;9:395-404.
381. Kanzler H, Kuppers R, Hansmann ML, Rajewsky K. Hodgkin and Reed-Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells. J Exp Med. 1996;184:1495-1505.
382. Rajewsky K, Kanzler H, Hansmann ML, Kuppers R. Normal and malignant B-cell development with special reference to Hodgkin's disease. Ann Oncol. 1997;8:79-81.
383. Marafioti T, Hummel M, Anagnostopoulos I, et al. Origin of nodular lymphocyte-predominant Hodgkin's disease from a clonal expansion of highly mutated germinal-center B cells. N Engl J Med. 1997;337:453-458.
384. Ohno T, Stribley JA, Wu G, Hinrichs SH, Weisenburger DD, Chan WC. Clonality in nodular lymphocyte-predominant Hodgkin's disease. N Engl J Med. 1997;337:459-465.
385. Braeuninger A, Kuppers R, Strickler JG, Wacker HH, Rajewsky K, Hansmann ML. Hodgkin and Reed-Sternberg cells in lymphocyte predominant Hodgkin disease represent clonal populations of germinal center-derived tumor B cells [published erratum appears in Proc Natl Acad Sci U S A 1997 Dec 9;94(25):14211]. Proc Natl Acad Sci U S A. 1997;94:9337-9342.
386. Klein U, Goossens T, Fischer M, et al. Somatic hypermutation in normal and transformed human B cells [In Process Citation]. Immunol Rev. 1998;162:261-280.
387. Kuppers R. Identifying the precursors of Hodgkin and Reed-Sternberg cells in Hodgkin's disease: role of the germinal center in B-cell lymphomagenesis. J Acquir Immune Defic Syndr. 1999;21 Suppl 1:S74-79.
388. Marafioti T, Hummel M, Foss HD, et al. Hodgkin and reed-sternberg cells represent an expansion of a single clone originating from a germinal center B-cell with functional immunoglobulin gene rearrangements but defective immunoglobulin transcription. Blood. 2000;95:1443-1450.

389. Stein H, Marafioti T, Foss HD, et al. Down-regulation of BOB.1/OBF.1 and Oct2 in classical Hodgkin disease but not in lymphocyte predominant Hodgkin disease correlates with immunoglobulin transcription. Blood. 2001;97:496-501.
390. de Leval L, Harris NL. Variability in immunophenotype in diffuse large B-cell lymphoma and its clinical relevance. Histopathology. 2003;43:509-528.
391. Hyland J, Lasota J, Jasinski M, Petersen RO, Nordling S, Miettinen M. Molecular pathological analysis of testicular diffuse large cell lymphomas. Hum Pathol. 1998;29:1231-1239.
392. Taniguchi M, Oka K, Hiasa A, et al. De novo CD5+ diffuse large cell lymphomas express VH genes with somatic mutation. Blood. 1998;91:1145-1151.
393. Delecluse HJ, Hummel M, Marafioti T, Anagnostopoulos I, Stein H. Common and HIV-related diffuse large B-cell lymphomas differ in their immunoglobulin gene mutation pattern. J Pathol. 1999;188:133-138.
394. Lossos IS, Okada CY, Tibshirani R, et al. Molecular analysis of immunoglobulin genes in diffuse large B-cell lymphomas. Blood. 2000;95:1797-1803.
395. Funkhouser WK, Warnke RA. Preferential IgH V4-34 gene segment usage in particular subtypes of B- cell lymphoma detected by antibody 9G4. Hum Pathol. 1998;29:1317-1321.
396. Daley MD, Berinstein NL, Siminovitch KA. Immunoglobulin heavy chain variable gene utilization in human large cell and Burkitt's lymphoma cell lines. Clin Invest Med. 1994;17:522-530.
397. Kuppers R, Rajewsky K, Hansmann ML. Diffuse large cell lymphomas are derived from mature B cells carrying V region genes with a high load of somatic mutation and evidence of selection for antibody expression. Eur J Immunol. 1997;27:1398-1405.
398. Rosenquist R, Lindstrom A, Holmberg D, Lindh J, Roos G. V(H) gene family utilization in different B-cell lymphoma subgroups. Eur J Haematol. 1999;62:123-128.
399. Ottensmeier CH, Stevenson FK. Isotype switch variants reveal clonally related subpopulations in diffuse large B-cell lymphoma. Blood. 2000;96:2550-2556.
400. Thompsett AR, Ellison DW, Stevenson FK, Zhu D. V(H) gene sequences from primary central nervous system lymphomas indicate derivation from highly mutated germinal center B cells with ongoing mutational activity. Blood. 1999;94:1738-1746.
401. Montesinos-Rongen M, Kuppers R, Schluter D, et al. Primary central nervous system lymphomas are derived from germinal- center B cells and show a preferential usage of the V4-34 gene segment. Am J Pathol. 1999;155:2077-2086.
402. Sekita T, Tamaru JI, Kaito K, Katayama T, Kobayashi M, Mikata A. Primary central nervous system lymphomas express Vh genes with intermediate to high somatic mutations. Leuk Lymphoma. 2001;41:377-385.
403. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403:503-511.
404. Lossos IS, Alizadeh AA, Eisen MB, et al. Ongoing immunoglobulin somatic mutation in germinal center B cell-like but not in activated B cell-like diffuse large cell lymphomas. Proc Natl Acad Sci U S A. 2000;97:10209-10213.
405. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med. 2002;346:1937-1947.
406. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med. 2002;8:68-74.

407. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood. 2004;103:275-282.
408. Hyjek E, Isaacson PG. Primary B cell lymphoma of the thyroid and its relationship to Hashimoto's thyroiditis. Hum Pathol. 1988;19:1315-1326.
409. Hyjek E, Smith WJ, Isaacson PG. Primary B-cell lymphoma of salivary glands and its relationship to myoepithelial sialadenitis. Hum Pathol. 1988;19:766-776.
410. Wotherspoon AC, Ortiz-Hildago C, Falzon MR, Isaacson PG. Helicobacter pylori-associated gastritis and primary B-cell gastric lymphoma. Lancet. 1991;338:1175-1176.
411. Qin Y, Greiner A, Trunk MJ, Schmausser B, Ott MM, Muller-Hermelink HK. Somatic hypermutation in low-grade mucosa-associated lymphoid tissue-type B-cell lymphoma. Blood. 1995;86:3528-3534.
412. Hallas C, Greiner A, Peters K, Muller-Hermelink HK. Immunoglobulin VH genes of high-grade mucosa-associated lymphoid tissue lymphomas show a high load of somatic mutations and evidence of antigen-dependent affinity maturation. Lab Invest. 1998;78:277-287.
413. Kurihara H, Koda H, Asami S, Kiso Y, Tanaka T. Contribution of the antioxidative property of astaxanthin to its protective effect on the promotion of cancer metastasis in mice treated with restraint stress. Life Sci. 2002;70:2509-2520.
414. Kepler TB, Perelson AS. Cyclic re-entry of germinal center B cells and the efficiency of affinity maturation. Immunol Today. 1993;14:412-415.
415. Bahler DW, Miklos JA, Swerdlow SH. Ongoing Ig gene hypermutation in salivary gland mucosa-associated lymphoid tissue-type lymphomas. Blood. 1997;89:3335-3344.
416. Du M, Diss TC, Xu C, Peng H, Isaacson PG, Pan L. Ongoing mutation in MALT lymphoma immunoglobulin gene suggests that antigen stimulation plays a role in the clonal expansion. Leukemia. 1996;10:1190-1197.
417. William J, Euler C, Christensen S, Shlomchik MJ. Evolution of autoantibody responses via somatic hypermutation outside of germinal centers. Science. 2002;297:2066-2070.
418. Miklos JA, Swerdlow SH, Bahler DW. Salivary gland mucosa-associated lymphoid tissue lymphoma immunoglobulin V(H) genes show frequent use of V1-69 with distinctive CDR3 features. Blood. 2000;95:3878-3884.
419. Bende RJ, Aarts WM, Riedl RG, de Jong D, Pals ST, van Noesel CJ. Among B cell non-Hodgkin's lymphomas, MALT lymphomas express a unique antibody repertoire with frequent rheumatoid factor reactivity. J Exp Med. 2005;201:1229-1241.
420. Yoshida M, Okabe M, Eimoto T, et al. Immunoglobulin VH genes in thymic MALT lymphoma are biased toward a restricted repertoire and are frequently unmutated. J Pathol. 2006;208:415-422.
421. Inagaki H, Chan JK, Ng JW, et al. Primary thymic extranodal marginal-zone B-cell lymphoma of mucosa-associated lymphoid tissue type exhibits distinctive clinicopathological and molecular features. Am J Pathol. 2002;160:1435-1443.
422. Zandvoort A, Timens W. The dual function of the splenic marginal zone: essential for initiation of anti-TI-2 responses but also vital in the general first-line defense against blood-borne antigens. Clin Exp Immunol. 2002;130:4-11.
423. Mollejo M, Lloret E, Menarguez J, Piris MA, Isaacson PG. Lymph node involvement by splenic marginal zone lymphoma: morphological and immunohistochemical features. Am J Surg Pathol. 1997;21:772-780.
424. Zhu D, Oscier DG, Stevenson FK. Splenic lymphoma with villous lymphocytes involves B cells with extensively mutated Ig heavy chain variable region genes. Blood. 1995;85:1603-1607.

425. Dunn-Walters DK, Boursier L, Spencer J, Isaacson PG. Analysis of immunoglobulin genes in splenic marginal zone lymphoma suggests ongoing mutation. Hum Pathol. 1998;29:585-593.
426. Tierens A, Delabie J, Pittaluga S, Driessen A, DeWolf-Peeters C. Mutation analysis of the rearranged immunoglobulin heavy chain genes of marginal zone cell lymphomas indicates an origin from different marginal zone B lymphocyte subsets. Blood. 1998;91:2381-2386.
427. Miranda RN, Cousar JB, Hammer RD, Collins RD, Vnencak-Jones CL. Somatic mutation analysis of IgH variable regions reveals that tumor cells of most parafollicular (monocytoid) B-cell lymphoma, splenic marginal zone B-cell lymphoma, and some hairy cell leukemia are composed of memory B lymphocytes. Hum Pathol. 1999;30:306-312.
428. Camacho FI, Mollejo M, Mateo MS, et al. Progression to large B-cell lymphoma in splenic marginal zone lymphoma: a description of a series of 12 cases. Am J Surg Pathol. 2001;25:1268-1276.
429. Bahler DW, Pindzola JA, Swerdlow SH. Splenic marginal zone lymphomas appear to originate from different B cell types. Am J Pathol. 2002;161:81-88.
430. Spencer J, Perry ME, Dunn-Walters DK. Human marginal-zone B cells. Immunol Today. 1998;19:421-426.
431. Jansen J, Schuit HRE, Meijer CJLM, Van Nieuwkoop JA, Hijmans W. Cell markers in hairy cell leukemia studied in cells from 51 patients. Blood. 1982;59:52-60.
432. Maloum K, Magnac C, Azgui Z, et al. VH gene expression in hairy cell leukaemia. Br J Haematol. 1998;101:171-178.
433. Forconi F, Sahota SS, Raspadori D, Mockridge CI, Lauria F, Stevenson FK. Tumor cells of hairy cell leukemia express multiple clonally related immunoglobulin isotypes via RNA splicing. Blood. 2001;98:1174-1181.
434. Vanhentenrijk V, Tierens A, Wlodarska I, Verhoef G, Wolf-Peeters CD. V(H) gene analysis of hairy cell leukemia reveals a homogeneous mutation status and suggests its marginal zone B-cell origin. Leukemia. 2004;18:1729-1732.
435. Thorselius M, Walsh SH, Thunberg U, Hagberg H, Sundstrom C, Rosenquist R. Heterogeneous somatic hypermutation status confounds the cell of origin in hairy cell leukemia. Leuk Res. 2005;29:153-158.
436. Arons E, Sunshine J, Suntum T, Kreitman RJ. Somatic hypermutation and VH gene usage in hairy cell leukaemia. Br J Haematol. 2006;133:504-512.
437. Basso K, Liso A, Tiacci E, et al. Gene expression profiling of hairy cell leukemia reveals a phenotype related to memory B cells with altered expression of chemokine and adhesion receptors. J Exp Med. 2004;199:59-68.
438. Wagner SD, Martinelli V, Luzzatto L. Similar patterns of V kappa gene usage but different degrees of somatic mutation in hairy cell leukemia, prolymphocytic leukemia, Waldenstrom's macroglobulinemia, and myeloma. Blood. 1994;83:3647-3653.
439. Aoki H, Takishita M, Kosaka M, Saito S. Frequent somatic mutations in D and/or JH segments of Ig gene in Waldenstrom's macroglobulinemia and chronic lymphocytic leukemia (CLL) with Richter's syndrome but not in common CLL. Blood. 1995;85:1913-1919.
440. Sahota SS, Forconi F, Ottensmeier CH, et al. Typical Waldenstrom macroglobulinemia is derived from a B-cell arrested after cessation of somatic mutation but prior to isotype switch events. Blood. 2002;100:1505-1507.
441. Sahota SS, Forconi F, Ottensmeier CH, Stevenson FK. Origins of the malignant clone in typical Waldenstrom's macroglobulinemia. Semin Oncol. 2003;30:136-141.
442. Kriangkum J, Taylor BJ, Mant MJ, Treon SP, Belch AR, Pilarski LM. The malignant clone in Waldenstrom's macroglobulinemia. Semin Oncol. 2003;30:132-135.

443. Kriangkum J, Taylor BJ, Treon SP, Mant MJ, Belch AR, Pilarski LM. Clonotypic IgM V/D/J sequence analysis in Waldenstrom macroglobulinemia suggests an unusual B-cell origin and an expansion of polyclonal B cells in peripheral blood. Blood. 2004;104:2134-2142.
444. Sahota S, Babbage G, Townsend M, Weston-Bell N, Stevenson F. Tracking the malignant cell of origin in Waldenstrom's macroglobulinemia. Haematologica. 2007;92:78-79.
445. Bakkus MH, Heirman C, Van Riet I, Van Camp B, Thielemans K. Evidence that multiple myeloma Ig heavy chain VDJ genes contain somatic mutations but show no intraclonal variation. Blood. 1992;80:2326-2335.
446. Sahota S, Hamblin T, Oscier DG, Stevenson FK. Assessment of the role of clonogenic B lymphocytes in the pathogenesis of multiple myeloma. Leukemia. 1994;8:1285-1289.
447. Vescio RA, Cao J, Hong CH, et al. Myeloma Ig heavy chain V region sequences reveal prior antigenic selection and marked somatic mutation but no intraclonal diversity. J Immunol. 1995;155:2487-2497.
448. Kosmas C, Viniou NA, Stamatopoulos K, et al. Analysis of the kappa light chain variable region in multiple myeloma. Br J Haematol. 1996;94:306-317.
449. Sahota SS, Leo R, Hamblin TJ, Stevenson FK. Myeloma V(L) and V(H) gene sequences reveal a complementary imprint of antigen selection in tumor cells. Blood. 1997;89:219-226.
450. Hadzidimitriou A, Stamatopoulos K, Belessi C, et al. Immunoglobulin genes in multiple myeloma: expressed and non-expressed repertoires, heavy and light chain pairings and somatic mutation patterns in a series of 101 cases. Haematologica. 2006;91:781-787.
451. Gonzalez D, Gonzalez M, Balanzategui A, et al. Molecular characteristics and gene segment usage in IGH gene rearrangements in multiple myeloma. Haematologica. 2005;90:906-913.
452. Sahota SS, Garand R, Mahroof R, et al. V(H) gene analysis of IgM-secreting myeloma indicates an origin from a memory cell undergoing isotype switch events. Blood. 1999;94:1070-1076.
453. Corradini P, Boccadoro M, Voena C, Pileri A. Evidence for a bone marrow B cell transcribing malignant plasma cell VDJ joined to Cmu sequence in immunoglobulin (IgG)- and IgA-secreting multiple myelomas. J Exp Med. 1993;178:1091-1096.
454. Billadeau D, Ahmann G, Greipp P, Van Ness B. The bone marrow of multiple myeloma patients contains B cell populations at different stages of differentiation that are clonally related to the malignant plasma cell. J Exp Med. 1993;178:1023-1031.
455. Reiman T, Seeberger K, Taylor BJ, et al. Persistent preswitch clonotypic myeloma cells correlate with decreased survival: evidence for isotype switching within the myeloma clone. Blood. 2001;98:2791-2799.
456. Taylor BJ, Pittman JA, Seeberger K, et al. Intraclonal homogeneity of clonotypic immunoglobulin M and diversity of nonclinical post-switch isotypes in multiple myeloma: insights into the evolution of the myeloma clone. Clin Cancer Res. 2002;8:502-513.
457. Kyle RA, Robinson RA, Katzmann JA. The clinical aspects of biclonal gammopathies. Review of 57 cases. American Journal of Medicine. 1981;71:999-1008.
458. Sahota SS, Leo R, Hamblin TJ, Stevenson FK. Ig V(H) gene mutational patterns indicate different tumor cell status in human myeloma and monoclonal gammopathy of undetermined significance. Blood. 1996;87:746-755.
459. Zojer N, Ludwig H, Fiegl M, Stevenson FK, Sahota SS. Patterns of somatic mutations in VH genes reveal pathways of clonal transformation from MGUS to multiple myeloma. Blood. 2003.

460. Zabalegui N, de Cerio AL, Inoges S, et al. Acquired potential N-glycosylation sites within the tumor-specific immunoglobulin heavy chains of B-cell malignancies. Haematologica. 2004;89:541-546.
461. Montesinos-Rongen M, Schmitz R, Courts C, et al. Absence of immunoglobulin class switch in primary lymphomas of the central nervous system. Am J Pathol. 2005;166:1773-1779.
462. McCann KJ, Sahota SS, Stevenson FK, Ottensmeier CH. Idiotype gene rescue in Follicular Lymphoma. Totowa, NJ: Humana Press; 2005.
463. Kasturi L, Chen H, Shakin-Eshleman SH. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem J. 1997;323 (Pt 2):415-419.
464. Hawkins RE, Zhu D, Ovecka M, et al. Idiotypic vaccination against human B-cell lymphoma. Rescue of variable region gene sequences from biopsy material for assembly as single-chain Fv personal vaccines. Blood. 1994;83:3279-3288.
465. King CA, Spellerberg MB, Zhu D, et al. DNA vaccines with single-chain Fv fused to fragment C of tetanus toxin induce protective immunity against lymphoma and myeloma. Nat Med. 1998;4:1281-1286.
466. Radcliffe CM, Arnold JN, Suter DM, et al. Human follicular lymphoma cells contain oligomannose glycans in the antigen-binding site of the B-cell receptor. J Biol Chem. 2007;282:7405-7415.
467. Zamze S, Martinez-Pomares L, Jones H, et al. Recognition of bacterial capsular polysaccharides and lipopolysaccharides by the macrophage mannose receptor. J Biol Chem. 2002;277:41613-41623.
468. Guesdon JL, Ternynck T, Avrameas S. The use of avidin-biotin interaction in immunoenzymatic techniques. J Histochem Cytochem. 1979;27:1131-1139.
469. NCI Non-Hodgkin's Classification Project Writing Committee. Classification of Non-Hodgkin's lymphomas: Reproducibility of Major Classification Systems. Cancer. 1985;55:457-481.
470. Carbone PP, Kaplan HS, Musshoff K, al e. Report of the committee on Hodgkin's Disease staging. Cancer Res. 1971;31:1860-1861.
471. Portlock CS. Management of low-grade non-Hodgkin's lymphomas. Semin Oncol. 1990;17:51-59.
472. Marcus R. Use of rituximab in patients with follicular lymphoma. Clin Oncol (R Coll Radiol). 2007;19:38-49.
473. Simon R, Durrleman S, Hoppe RT, et al. The Non-Hodgkin Lymphoma Pathologic Classification Project. Long-term follow-up of 1153 patients with non-Hodgkin lymphomas. Annals of Internal Medicine. 1988;109:939-945.
474. Ersbøll J, Shultz HB, Pedersen-Bjergaard J, Nissen NI. Follicular low grade non-Hodgkin's lymphoma: Long-term outcome with or without tumor progression. Eur J Haematol. 1989;42:155-163.
475. Horning SJ. Natural history of and therapy for the indolent non-Hodgkins's lymphoma. Semin Oncol. 1993;20:75-88.
476. Gallagher CJ, Gregory WM, Jones AE, et al. Follicular lymphoma: prognostic factors for response and survival. J Clin Oncol. 1986;4:1470-1480.
477. Acker B, Hoppe RT, Colby TV, Cox RS, Kaplan HS, Rosenberg SA. Histologic conversion in the non-Hodgkin's lymphomas. J Clin Oncol. 1983;1:11-16.
478. Hubbard SM, Chabner BA, De Vita VT, et al. Histologic progression in non-Hodgkin's lymphoma. Blood. 1982;59:258-264.
479. Yuen AR, Kamel OW, Halpern J, Horning SJ. Long-term survival after histologic transformation of low-grade follicular lymphoma. J Clin Oncol. 1995;13:1726-1733.

480. Bastion Y, Sebban C, Berger F, et al. Incidence, predictive factors, and outcome of lymphoma transformation in follicular lymphoma patients. J Clin Oncol. 1997;15:1587-1594.
481. Stein H, Gerdes J, Mason DY. The normal and malignant germinal centre. Clin Haematol. 1982;11:531-559.
482. Stein H, Lennert K, Feller AC, Mason DY. Immunohistological analysis of human lymphoma: correlation of histological and immunological categories. Adv Cancer Res. 1984;42:67-147.
483. Harris NL, Nadler LM, Bhan AK. Immunohistologic characterization of two malignant lymphomas of germinal center type (centroblastic/centrocytic and centrocytic) with monoclonal antibodies. Follicular and diffuse lymphomas of small- cleaved-cell type are related but distinct entities. Am J Pathol. 1984;117:262-272.
484. Zukerberg LR, Medeiros LJ, Ferry JA, Harris NL. Diffuse low-grade B-cell lymphomas. Four clinically distinct subtypes defined by a combination of morphologic and immunophenotypic features. Am J Clin Pathol. 1993;100:373-385.
485. Bahler DW, Levy R. Clonal evolution of a follicular lymphoma: evidence for antigen selection. Proc Natl Acad Sci USA. 1992;89:6770-6774.
486. Dorner T, Brezinschek HP, Brezinschek RI, Foster SJ, Domiati-Saad R, Lipsky PE. Analysis of the frequency and pattern of somatic mutations within nonproductively rearranged human variable heavy chain genes. J Immunol. 1997;158:2779-2789.
487. Wu H, Pelkonen E, Knuutila S, Kaartinen M. A human follicular lymphoma B cell line hypermutates its functional immunoglobulin genes in vitro. Eur J Immunol. 1995;25:3263-3269.
488. Tsujimoto Y, Cossman J, Jaffe E, Croce CM. Involvement of the bcl-2 gene in human follicular lymphoma. Science. 1985;228:1440-1443.
489. Tsujimoto Y, Jaffe E, Cossman J, Gorham J, Nowell PC, Croce CM. Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the t(11;14) chromosome translocation. Nature. 1985;315:340-343.
490. Akasaka T, Akasaka H, Yonetani N, et al. Refinement of the BCL2/immunoglobulin heavy chain fusion gene in t(14;18)(q32;q21) by polymerase chain reaction amplification for long targets. Genes Chromosomes Cancer. 1998;21:17-29.
491. Roulland S, Navarro JM, Grenot P, et al. Follicular lymphoma-like B cells in healthy individuals: a novel intermediate step in early lymphomagenesis. J Exp Med. 2006;203:2425-2431.
492. Foster SJ, Dorner T, Lipsky PE. Somatic hypermutation of VkappaJkappa rearrangements: targeting of RGYW motifs on both DNA strands and preferential selection of mutated codons within RGYW motifs. Eur J Immunol. 1999;29:4011-4021.
493. Huang C, Stewart AK, Schwartz RS, Stollar BD. Immunoglobulin heavy chain gene expression in peripheral blood B lymphocytes. J Clin Invest. 1992;89:1331-1343.
494. Brezinschek HP, Brezinschek RI, Lipsky PE. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. J Immunol. 1995;155:190-202.
495. Demaison C, David D, Letourneur F, Theze J, Saragosti S, Zouali M. Analysis of human VH gene repertoire expression in peripheral CD19+ B cells. Immunogenetics. 1995;42:342-352.
496. Kabat EA, Wu TT, Bilofsky H. Sequence of Immunoglobulin Chains. Tabulation and analysis of Amino Acid Sequences of Precursors, V-regions, C-regions, J-chains and β2-microglobulins. Washington, D.C.: U.S. Department of Health, Education and Welfare; 1979.

497. Klein R, Jaenichen R, Zachau HG. Expressed human immunoglobulin kappa genes and their hypermutation. Eur J Immunol. 1993;23:3248-3262.
498. McCann KJ, Johnson PW, Stevenson FK, Ottensmeier CH. Universal N-glycosylation sites introduced into the B-cell receptor of follicular lymphoma by somatic mutation: a second tumorigenic event? Leukemia. 2006;20:530-534.
499. Foster SJ, Dorner T, Lipsky PE. Targeting and subsequent selection of somatic hypermutations in the human V kappa repertoire. Eur J Immunol. 1999;29:3122-3132.
500. Tarentino AL, Plummer TH, Jr., Maley F. The release of intact oligosaccharides from specific glycoproteins by endo-beta-N-acetylglucosaminidase H. J Biol Chem. 1974;249:818-824.
501. Linehan SA, Martinez-Pomares L, da Silva RP, Gordon S. Endogenous ligands of carbohydrate recognition domains of the mannose receptor in murine macrophages, endothelial cells and secretory cells; potential relevance to inflammation and immunity. Eur J Immunol. 2001;31:1857-1866.
502. Levy R, Warnke R, Dorfman RF, Haimovich J. The monoclonality of human B-cell lymphomas. J Exp Med. 1977;145:1014-1028.
503. Dogan A, Du MQ, Aiello A, et al. Follicular lymphomas contain a clonally linked but phenotypically distinct neoplastic B-cell population in the interfollicular zone. Blood. 1998;91:4708-4714.
504. del Senno L, Gandini D, Gambari R, Lanza F, Tomasi P, Castoldi G. Monoclonal origin of B cells producing k, lambda and k lambda immunoglobulin light chains in a patient with chronic lymphocytic leukemia. Leuk Res. 1987;11:1093-1098.
505. Peltomaki P, Bianchi NO, Knuutila S, et al. Immunoglobulin kappa and lambda light chain dual genotype rearrangement in a patient with kappa-secreting B-CLL. Eur J Cancer Clin Oncol. 1988;24:1233-1238.
506. Rolink A, Grawunder U, Haasner D, Strasser A, Melchers F. Immature surface Ig+ B cells can continue to rearrange kappa and lambda L chain gene loci. J Exp Med. 1993;178:1263-1270.
507. Giachino C, Padovan E, Lanzavecchia A. kappa+lambda+ dual receptor B cells are present in the human peripheral repertoire. J Exp Med. 1995;181:1245-1250.
508. Wu HY, Kaartinen M. The somatic hypermutation activity of a follicular lymphoma links to large insertions and deletions of immunoglobulin genes. Scand J Immunol. 1995;42:52-59.
509. Goossens T, Klein U, Kuppers R. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. Proc Natl Acad Sci U S A. 1998;95:2463-2468.
510. Wilson PC, de Bouteiller O, Liu YJ, et al. Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. J Exp Med. 1998;187:59-70.
511. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology. 2004;14:103-114.
512. Hubbard SJ, Thornton JM. NACCESS: Dept. of Biochemistry and Molecular Biology, University College London; 1993.
513. Cauerhff A, Braden BC, Carvalho JG, et al. Three-dimensional structure of the Fab from a human IgM cold agglutinin. J Immunol. 2000;165:6422-6428.
514. Liu Y, Hernandez AM, Shibata D, Cortopassi GA. BCL2 translocation frequency rises with age in humans. Proc Natl Acad Sci U S A. 1994;91:8910-8914.
515. Dolken G, Illerhaus G, Hirt C, Mertelsmann R. BCL-2/JH rearrangements in circulating B cells of healthy blood donors and patients with nonmalignant diseases. J Clin Oncol. 1996;14:1333-1344.

516. Schuler F, Hirt C, Dolken G. Chromosomal translocation t(14;18) in healthy individuals. Semin Cancer Biol. 2003;13:203-209.
517. Stein M, Keshav S, Harris N, Gordon S. Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. J Exp Med. 1992;176:287-292.
518. Gordon S. Alternative activation of macrophages. Nat Rev Immunol. 2003;3:23-25.
519. Farinha P, Masoudi H, Skinnider BF, et al. Analysis of multiple biomarkers shows that lymphoma-associated macrophage (LAM) content is an independent predictor of survival in follicular lymphoma (FL). Blood. 2005;106:2169-2174.
520. Morton LM, Wang SS, Devesa SS, Hartge P, Weisenburger DD, Linet MS. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. Blood. 2006;107:265-276.
521. Ng AK. Diffuse large B-cell lymphoma. Semin Radiat Oncol. 2007;17:169-175.
522. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. N Engl J Med. 1993;329:987-994.
523. Lossos IS, Jones CD, Warnke R, et al. Expression of a single gene, BCL-6, strongly predicts survival in patients with diffuse large B-cell lymphoma. Blood. 2001;98:945-951.
524. Vitolo U, Botto B, Capello D, et al. Point mutations of the BCL-6 gene: clinical and prognostic correlation in B-diffuse large cell lymphoma. Leukemia. 2002;16:268-275.
525. Offit K, Lo Coco F, Louie DC, et al. Rearrangement of the bcl-6 gene as a prognostic marker in diffuse large-cell lymphoma. N Engl J Med. 1994;331:74-80.
526. Kramer MH, Hermans J, Wijburg E, et al. Clinical relevance of BCL2, BCL6, and MYC rearrangements in diffuse large B-cell lymphoma. Blood. 1998;92:3152-3162.
527. Barrans SL, O'Connor SJ, Evans PA, et al. Rearrangement of the BCL6 locus at 3q27 is an independent poor prognostic factor in nodal diffuse large B-cell lymphoma. Br J Haematol. 2002;117:322-332.
528. Akasaka T, Ueda C, Kurata M, et al. Nonimmunoglobulin (non-Ig)/BCL6 gene fusion in diffuse large B-cell lymphoma results in worse prognosis than Ig/BCL6. Blood. 2000;96:2907-2909.
529. Ueda C, Akasaka T, Ohno H. Non-immunoglobulin/BCL6 gene fusion in diffuse large B-cell lymphoma: prognostic implications. Leuk Lymphoma. 2002;43:1375-1381.
530. Lo Coco F, Ye BH, Lista F, et al. Rearrangements of the BCL6 gene in diffuse large cell non-Hodgkin's lymphoma. Blood. 1994;83:1757-1759.
531. Volpe G, Vitolo U, Carbone A, et al. Molecular heterogeneity of B-lineage diffuse large cell lymphoma. Genes Chromosomes Cancer. 1996;16:21-30.
532. Monni O, Joensuu H, Franssila K, Klefstrom J, Alitalo K, Knuutila S. BCL2 overexpression associated with chromosomal amplification in diffuse large B-cell lymphoma. Blood. 1997;90:1168-1174.
533. Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. J Exp Med. 2001;194:1861-1874.
534. Huang JZ, Sanger WG, Greiner TC, et al. The t(14;18) defines a unique subset of diffuse large B-cell lymphoma with a germinal center B-cell gene expression profile. Blood. 2002;99:2285-2290.
535. Iqbal J, Sanger WG, Horsman DE, et al. BCL2 translocation defines a unique tumor subset within the germinal center B-cell-like diffuse large B-cell lymphoma. Am J Pathol. 2004;165:159-166.

536. Colomo L, Lopez-Guillermo A, Perales M, et al. Clinical impact of the differentiation profile assessed by immunophenotyping in patients with diffuse large B-cell lymphoma. Blood. 2003;101:78-84.
537. Lossos IS, Czerwinski DK, Alizadeh AA, et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med. 2004;350:1828-1837.
538. Chang CC, McClintock S, Cleveland RP, et al. Immunohistochemical expression patterns of germinal center and activation B-cell markers correlate with prognosis in diffuse large B-cell lymphoma. Am J Surg Pathol. 2004;28:464-470.
539. Jardin F, Sahota SS, Ruminy P, et al. Novel Ig V gene features of t(14;18) and t(3;14) de novo diffuse large B-cell lymphoma displaying germinal center-B cell like and non-germinal center-B cell like markers. Leukemia. 2006;20:2070-2074.
540. Smit LA, Bende RJ, Aten J, Guikema JE, Aarts WM, van Noesel CJ. Expression of activation-induced cytidine deaminase is confined to B-cell non-Hodgkin's lymphomas of germinal-center phenotype. Cancer Res. 2003;63:3894-3898.
541. Lossos IS, Levy R, Alizadeh AA. AID is expressed in germinal center B-cell-like and activated B-cell-like diffuse large-cell lymphomas and is not correlated with intraclonal heterogeneity. Leukemia. 2004;18:1775-1779.
542. Hardianti MS, Tatsumi E, Syampurnawati M, et al. Activation-induced cytidine deaminase expression in follicular lymphoma: association between AID expression and ongoing mutation in FL. Leukemia. 2004;18:826-831.
543. O'Neill BP, Dinapoli RP, Kurtin PJ, Habermann TM. Occult systemic non-Hodgkin's lymphoma (NHL) in patients initially diagnosed as primary central nervous system lymphoma (PCNSL): how much staging is enough? J Neurooncol. 1995;25:67-71.
544. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. CA Cancer J Clin. 2006;56:106-130.
545. Schlegel U, Schmidt-Wolf IG, Deckert M. Primary CNS lymphoma: clinical presentation, pathological classification, molecular pathogenesis and treatment. J Neurol Sci. 2000;181:1-12.
546. Bashir R, Luka J, Cheloha K, Chamberlain M, Hochberg F. Expression of Epstein-Barr virus proteins in primary CNS lymphoma in AIDS patients. Neurology. 1993;43:2358-2362.
547. Loeffler JS, Ervin TJ, Mauch P, et al. Primary lymphomas of the central nervous system: patterns of failure and factors that influence survival. J Clin Oncol. 1985;3:490-494.
548. Fine HA, Mayer RJ. Primary central nervous system lymphoma. Ann Intern Med. 1993;119:1093-1104.
549. Bashir R, Freedman A, Harris N, Bain K, Nadler L, Hochberg F. Immunophenotypic profile of CNS lymphoma: a review of eighteen cases. J Neurooncol. 1989;7:249-254.
550. Pels H, Montesinos-Rongen M, Schaller C, et al. VH gene analysis of primary CNS lymphomas. J Neurol Sci. 2005;228:143-147.
551. Larocca LM, Capello D, Rinelli A, et al. The molecular and phenotypic profile of primary central nervous system lymphoma identifies distinct categories of the disease and is consistent with histogenetic derivation from germinal center-related B cells. Blood. 1998;92:1011-1019.
552. Camilleri-Broet S, Criniere E, Broet P, et al. A uniform activated B-cell-like immunophenotype might explain the poor prognosis of primary central nervous system lymphomas: analysis of 83 cases. Blood. 2006;107:190-196.

553. Sugita Y, Tokunaga O, Nakashima A, Shigemori M. SHP-1 expression in primary central nervous system B-cell lymphomas in immunocompetent patients reflects maturation stage of normal B cell counterparts. Pathol Int. 2004;54:659-666.
554. Bossolasco S, Cinque P, Ponzoni M, et al. Epstein-Barr virus DNA load in cerebrospinal fluid and plasma of patients with AIDS-related lymphoma. J Neurovirol. 2002;8:432-438.
555. Rao CR, Jain K, Bhatia K, Laksmaiah KC, Shankar SK. Association of primary central nervous system lymphomas with the Epstein-Barr virus. Neurol India. 2003;51:237-240.
556. Paulus W, Jellinger K, Hallas C, Ott G, Muller-Hermelink HK. Human herpesvirus-6 and Epstein-Barr virus genome in primary cerebral lymphomas. Neurology. 1993;43:1591-1593.
557. Morgello S. Pathogenesis and classification of primary central nervous system lymphoma: an update. Brain Pathol. 1995;5:383-393.
558. Gaidano G, Capello D, Pastore C, et al. Analysis of human herpesvirus type 8 infection in AIDS-related and AIDS-unrelated primary central nervous system lymphoma. J Infect Dis. 1997;175:1193-1197.
559. Montesinos-Rongen M, Hans VH, Eis-Hubinger AM, et al. Human herpes virus-8 is not associated with primary central nervous system lymphoma in HIV-negative patients. Acta Neuropathol (Berl). 2001;102:489-495.
560. Montesinos-Rongen M, Besleaga R, Heinsohn S, et al. Absence of simian virus 40 DNA sequences in primary central nervous system lymphoma in HIV-negative patients. Virchows Arch. 2004;444:436-438.
561. Murray JM, Morgello S. Polyomaviruses and primary central nervous system lymphomas. Neurology. 2004;63:1299-1301.
562. Martini F, Dolcetti R, Ferreri AJ, et al. No association between polyomaviruses and primary central nervous system lymphomas of HIV-seronegative and HIV-positive patients. Cancer Epidemiol Biomarkers Prev. 2004;13:1819-1820.
563. Galea I, Bechmann I, Perry VH. What is immune privilege (not)? Trends Immunol. 2007;28:12-18.
564. Hickey WF, Hsu BL, Kimura H. T-lymphocyte entry into the central nervous system. J Neurosci Res. 1991;28:254-260.
565. Galea I, Bernardes-Silva M, Forse PA, van Rooijen N, Liblau RS, Perry VH. An antigen-specific pathway for CD8 T cells across the blood-brain barrier. J Exp Med. 2007;204:2023-2030.
566. Anthony IC, Crawford DH, Bell JE. B lymphocytes in the normal brain: contrasts with HIV-associated lymphoid infiltrates and lymphomas. Brain. 2003;126:1058-1067.
567. Hatterer E, Davoust N, Didier-Bazes M, et al. How to drain without lymphatics? Dendritic cells migrate from the cerebrospinal fluid to the B-cell follicles of cervical lymph nodes. Blood. 2006;107:806-812.
568. Matsumoto M, Lo SF, Carruthers CJ, et al. Affinity maturation without germinal centres in lymphotoxin-alpha- deficient mice. Nature. 1996;382:462-466.
569. Kato J, Motoyama N, Taniuchi I, et al. Affinity maturation in Lyn kinase-deficient mice with defective germinal center formation. J Immunol. 1998;160:4788-4795.
570. Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F. Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. Brain Pathol. 2004;14:164-174.
571. Corcione A, Casazza S, Ferretti E, et al. Recapitulation of B cell differentiation in the central nervous system of patients with multiple sclerosis. Proc Natl Acad Sci U S A. 2004;101:11064-11069.

572. Bashir R, Coakham H, Hochberg F. Expression of LFA-1/ICAM-1 in CNS lymphomas: possible mechanism for lymphoma homing into the brain. J Neurooncol. 1992;12:103-110.
573. Aho R, Kalimo H, Salmi M, Smith D, Jalkanen S. Binding of malignant lymphoid cells to the white matter of the human central nervous system: role of different CD44 isoforms, beta 1, beta 2 and beta 7 integrins, and L-selectin. J Neuropathol Exp Neurol. 1997;56:557-568.
574. Tun HW, Personett D, Baskerville KA, et al. Pathway analysis of primary central nervous system lymphoma. Blood. 2008.
575. Smith JR, Braziel RM, Paoletti S, Lipp M, Uguccioni M, Rosenbaum JT. Expression of B-cell-attracting chemokine 1 (CXCL13) by malignant lymphocytes and vascular endothelium in primary central nervous system lymphoma. Blood. 2003;101:815-821.
576. Brunn A, Montesinos-Rongen M, Strack A, et al. Expression pattern and cellular sources of chemokines in primary central nervous system lymphoma. Acta Neuropathol. 2007;114:271-276.
577. Chapman CJ, Mockridge CI, Hamblin TJ, Stevenson FK. Tracking of the V4-34 (V(H)4-21) gene in human tonsil reveals clonal isotype switch events and a highly variable degree of somatic hypermutation. Clinical and Experimental Immunology. 1996;105:360-368.
578. Potter KN, Mockridge CI, Neville L, et al. Structural and functional features of the B-cell receptor in IgG-positive chronic lymphocytic leukemia. Clin Cancer Res. 2006;12:1672-1679.
579. Montesinos-Rongen M, Brunn A, Bentink S, et al. Gene expression profiling suggests primary central nervous system lymphomas to be derived from a late germinal center B cell. Leukemia. 2008;22:400-405.
580. Jahnke K, Hummel M, Korfel A, et al. Detection of subclinical systemic disease in primary CNS lymphoma by polymerase chain reaction of the rearranged immunoglobulin heavy-chain genes. J Clin Oncol. 2006;24:4754-4757.
581. Krumbholz M, Theil D, Derfuss T, et al. BAFF is produced by astrocytes and up-regulated in multiple sclerosis lesions and primary central nervous system lymphoma. J Exp Med. 2005;201:195-200.
582. Tysarowski A, Fabisiewicz A, Paszkiewicz-Kozik E, Kulik J, Walewski J, Siedlecki JA. Usefulness of real-time PCR in long-term follow-up of follicular lymphoma patients. Acta Biochim Pol. 2007;54:135-142.
583. McGreal EP, Rosas M, Brown GD, et al. The carbohydrate-recognition domain of Dectin-2 is a C-type lectin with specificity for high mannose. Glycobiology. 2006;16:422-430.
584. van Liempt E, Bank CM, Mehta P, et al. Specificity of DC-SIGN for mannose- and fucose-containing glycans. FEBS Lett. 2006;580:6123-6131.