Variable region gene expression and structural motifs of human polyreactive immunoglobulins

Paul Allen Ramsland

A thesis submitted for the degree of Doctor of Philosophy (PhD) at the University of Technology, Sydney

1997

CERTIFICATE

I certify that this thesis has not already been submitted for any degree and is not being submitted as part of candidature for any other degree.

I also certify that the thesis has been written by me and that any help that I have received in preparing this thesis, and all sources used, have been acknowledged in this thesis.

Signature of Candidate

Production Note:
Signature removed prior to publication.

Acknowledgments

Of all the people whom contributed to this work, my special thanks goes to my supervisor Professor Robert L Raison for his excellent support and guidance over the years. Joshua Moses facilitated greatly the sometimes difficult molecular biology involved in this project. Josh's particular attention to detail in every manipulation is a great scientific attribute. Professor Allen B Edmundson provided many stimulating discussions regarding the nature of immunoglobulins and Dr Luke W Guddat helped greatly in establishing my knowledge of protein modelling techniques. Andrew Watts and Kim Andersen were responsible for maintaining the computer workstations that I sat in front of for many hours in Sydney, Amarillo and Oklahoma. Peter Hains for his intuitive knowledge of protein separation methods. Dr Kathryn Weston helped with the use of the flow cytometer. Also, to all those members and friends within the Immunobiology Unit for their help whenever necessary and all the interesting interactions.

My personal thanks to Parisa, not only for the many intellectual forays, but, for her deep friendship, not to mention our many immunodebates. I am especially appreciative of my parents whom have supported me throughout everything. Elizabeth for simply being unique.

Publications arising from this thesis:

Ramsland, P.A., Guddat, L.W., Edmundson, A.B. and Raison, R.L. (1997). Diverse binding site structures revealed in homology models of polyreactive immunoglobulins. *Journal of Computer-Aided Molecular Design*, **11**: 453-461.

* A copy of this manuscript is bound in the back of this thesis

Table of contents

List of figures

List of tables

Abbreviations

Ab	antibody	**IgM**	immunoglobulin with μ isotype heavy chains
Ag	antigen	**IPTG**	isopropylthiogalactoside
Az	sodium azide	**J**	joining gene
B CLL	chronic B lymphocytic leukaemia	**L**	light chain
BSA	bovine serum albumin	**MFI**	mean fluorescence intensity
C	constant gene/domain	**MHC**	major histocompatibility complex
Cα	*alpha* carbon atom	**PAGE**	polyacrylamide gel electrophoresis
CDR	complementarity determining region	**PBL**	peripheral blood lymphocytes
CIAA	24:1 Choroform:isoamyl alcohol	**PBS**	phosphate buffered saline
CRI	cross reactive idiotype	**PCR**	polymerase chain reaction
D	diversity gene	**RBS**	ribosome binding site (Shine/Dalgarno sequence)
DNA	deoxyribonucleic acid	**RF**	rheumatoid factor
Fab	fragment antigen binding	**rms**	root mean square
Fc	fragment crystalline	**rmsd**	root mean square deviation
Fd	heavy chain fragment of Fab (VH–CH1)	**RNA**	ribonucleic acid
Fr	framework region	**scFv**	single chain variable region fragment (Fv)
Fv	fragment variable	**SDS**	sodium dodecyl sulfate
H	heavy chain	**TBS**	tris-HCl buffered saline
H-bond	hydrogen bond	**TdT**	terminal deoxy-transferase
HEL	hen egg-white lysozyme	**V**	variable gene/domain
Ig	immunoglobulin	**vdW**	van der Waal's
IgG	immunoglobulin with γ isotype heavy chains	**3D**	three-dimensional

* Infrequently used and common abbreviations are not included in this list

Abstract

Polyreactive immunoglobulins (Ig) exhibit a capacity to recognise multiple, structurally dissimilar antigens through a single combining site. This characteristic differentiates these Igs from monoreactive Igs which bind to a single antigen, usually with high specificity and affinity. Chronic B lymphocytic leukaemia (B CLL) is a malignancy identified by the incessant accumulation, in the peripheral circulation, of B lymphocytes of a mature and resting morphology. B CLL malignant cells generally express both surface IgM and the pan T cell antigen CD5. Moreover, the IgM on the surface of these CD5 positive B CLL cells is frequently polyreactive. This thesis examines the structural diversity found in the combining sites of B CLL derived Igs in an attempt to elucidate the structural basis of polyreactive antigen binding displayed by a significant proportion of human Igs. The genes encoding the variable (V) domains of five B CLL derived IgM antibodies (Bel, Tre, Yar, Hod and Jak) were cloned and sequenced (Chapter Two). When the light chain V domain genes were aligned with the closest germline VL and JL coding DNA sequences it was determined that there was either a complete absence of somatic mutation (Tre, Yar and Jak) or a minimal number of mutations (Bel and Hod) present in the rearranged VL domain genes. A remarkable fidelity in the splicing of VL to JL genes was noted suggesting that the diversity, normally introduced through variability of splicing VL to JL, is reduced in Igs expressed by B CLL cells. Furthermore, the markedly reduced primary structural diversity was highlighted when two of the VL domain genes (Yar and Hod) were found to be different in sequence by only four nucleotides and two amino acids. The heavy chain V domain genes of the same five Igs were sequenced in another study (Brock, 1995), however, it was interesting to analyse the sequences of the VH domain genes and compare them with the VL domain genes. The naive or germline nature of the B CLL antibodies was reflected in the VH genes by either an absence or a low frequency of mutations within these sequences compared with germline immunoglobulin gene sequences. No obvious conserved motif, which could be related to polyreactivity, was observed when the primary protein sequence was analysed for distribution of identical or similar amino acids. Thus, homology modelling was used to construct three-dimensional models of the Fv (VL–VH) portions of the five B CLL IgM molecules to examine the structures of the combining sites of these Igs (Chapter Three). Framework regions were constructed using X-ray coordinates taken from highly homologous human variable domain structures. Complementarity determining regions (CDR) were predicted by grafting loops, taken from known Ig structures, onto the Fv framework models. The CDR templates were selected, where possible, to be of the same length and of high residue identity or

similarity. If a single template CDR was not appropriate to model a particular CDR the loop was built from loop stems of known conformation, followed by chain closure with a β-turn. Template models were refined using standard molecular mechanics simulations. The binding sites were either relatively flat or contained a deep cavity at the VL–VH domain interface. Further differences in topology were the result of some CDR loops protruding into the solvent. Examination of the electrostatic molecular surface did not reveal a common structural feature within the binding sites of the five polyreactive Fv. While two of the binding cavities were positively charged the other three structures displayed either negatively charged or predominantly hydrophobic combining sites. These findings suggested that a diversity of structural mechanisms are involved in polyreactive antigen binding. Residues within CDRs which have aromatic side-chains and are partially exposed to solvent were distributed across large regions of the combining sites. It is possible that these aromatic residues are responsible for the conserved binding to mouse Igs observed (Chapter Two) for the B CLL derived polyreactive IgM molecules. Two Fv molecules (Bel and Tre) were cloned as dicistronic constructs, into the bacterial expression vector pFLAG. The expression of the Fvs was fully characterised and unfortunately the VL and VH of Bel and Tre Igs did not associate in an appropriate manner to yield large quantities of purified Fv (Chapter Four). Expression of correctly folded and stabilised fragments of human polyreactive immunoglobulins would enable the structural basis for the polyreactive binding phenomenon to be fully explored using protein crystallography.

CHAPTER ONE
Introduction

While the combining sites of monoreactive antibodies (Ab) recognise a single antigen, polyreactive immunoglobulins (Ig) are capable of binding to multiple and structurally diverse antigens through a single antibody combining site (reviewed in: Avrameas & Ternynck, 1993). The structural basis for immunoglobulin polyreactivity is not understood. However, many three-dimensional structures of antibody fragments derived from monoreactive or Ab of undetermined specificity have been solved using X-ray crystallography (reviewed in: Padlan *et al.*, 1994). The basic structure of immunoglobulins and the specific molecular characteristics of antibody/antigen recognition are reviewed here to provide a setting for this study. The research detailed within this thesis was aimed at the elucidation of the binding site structures of certain polyreactive immunoglobulins, expressed on the surface of malignant CD5 positive B cells in human chronic B lymphocytic leukaemia.

1.1: Basic structure of immunoglobulins

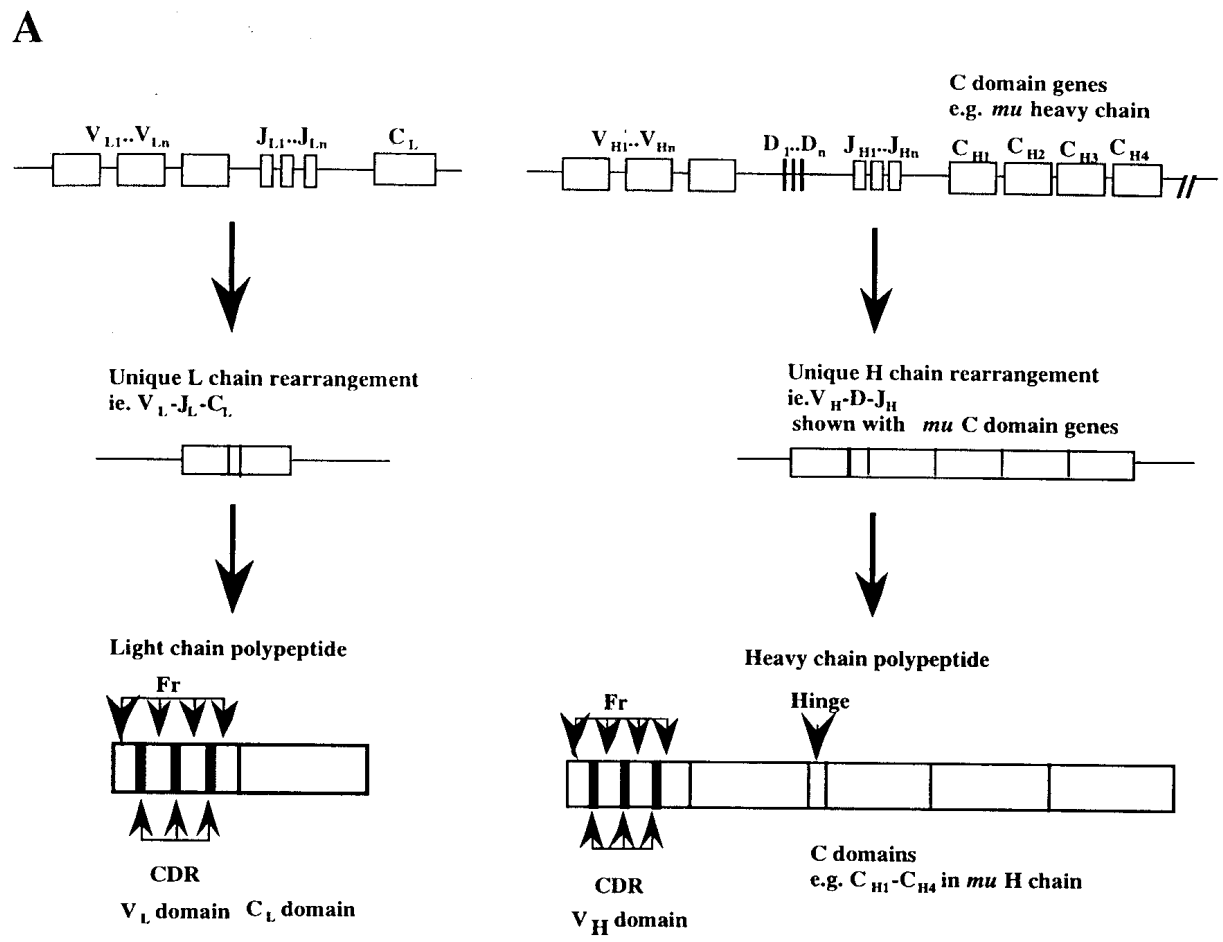
Immunoglobulins are comprised of light (L) and heavy (H) chains which are covalently linked by disulfide bridges. One light chain associates with one heavy chain to form a heterodimer, two of which form an Ab monomer. The basic structure of an antibody can be visualised as a Y-shaped molecule with two antigen binding sites located at the end of the arms and a stem which is involved in effector mechanisms (reviewed in: Edmundson *et al.*, 1993). The polypeptides (L and H chains) fold into discrete domains of approximately 110 amino acids, which all exhibit a basically similar three-dimensional structure (3D), the Ig fold (Schiffer *et al.*, 1973; Poljak *et al.*, 1973). These Ig domains are grouped into two categories; namely variable (V) and constant (C) domains. The variable domains of the light (VL) and heavy (VH) chains contain the Ab combining site. The C domains are involved in stabilising the Ab monomer and, as mentioned, the effector mechanisms. Since this thesis focuses on the structures of V domains of polyreactive Ig, the genes encoding the V domains and the three-dimensional structures of these domains will be discussed, in more detail than C domains, in the following sections.

1.1.1: Genes encoding variable domains of immunoglobulins

The genetic organisation of the human immunoglobulin loci is well described (Cox *et al.*, 1994; Cook & Tomlinson, 1995; Williams *et al.*, 1996). Light chain genes are located on chromosomes 2p11-12 and 22q11.2 for *kappa* (κ) and *lambda* (λ) isotypes respectively while the heavy chain locus, encoding all isotypes of this chain; *mu* (μ), *delta* (δ), *gamma* (γ), *alpha* (α) and *epsilon* (ε), occur on chromosome 14q32.3 (Cox *et al.*, 1982; Malcolm *et al.*, 1982; de la Chapelle *et al.*, 1983). During the development of an individual B cell the genes encoding the light and heavy chain V domains are generated from the random recombination of a large number of smaller gene segments (Figure 1.1A). These smaller gene segments are named V genes (VL and VH), diversity (D) genes (only for H chains) and joining (JL and JH) genes. Large numbers of V, D and J genes are located within the Ig gene loci and are arranged, before recombination, as discrete clusters of the respective gene segments. After recombination, the VL region is encoded by a single VL and JL gene and the VH region by a single VH, D and JH gene combination. The L and H chain genes of an antibody are then formed by the splicing of the recombined VL and VH region genes onto one or more constant domain genes (reviewed in: Schatz *et al.*, 1992).

1.1.2: Characteristics of an intact immunoglobulin molecule

In an intact antibody molecule the light chains associate with the heavy chains to form two heterodimers which are joined through single or multiple disulfide bonds within a flexible region termed the hinge. The N-terminal domains of light and heavy chains associate non-covalently to form an Fv (VL–VH) which contains the antibody combining site. The light chain C domain interacts with the CH1 domain both non-covalently and covalently (disulfide bond). Together, the V domain module (VL–VH) and the C domain module (CL and CH1) is termed the antigen binding fragment (Fab) (reviewed in: Edmundson *et al.*, 1993). The remaining portion of an antibody molecule (hinge and CH domains except for CH1) is termed the fragment crystalline (Fc) for historical reasons (reviewed in: Davies & Chacko, 1993; Edmundson *et al.*, 1993). These features of the Ig molecule are shown schematically in Figure 1.1B. The structural features of Ig molecules pertinent to this thesis will be discussed in detail in the following sections.



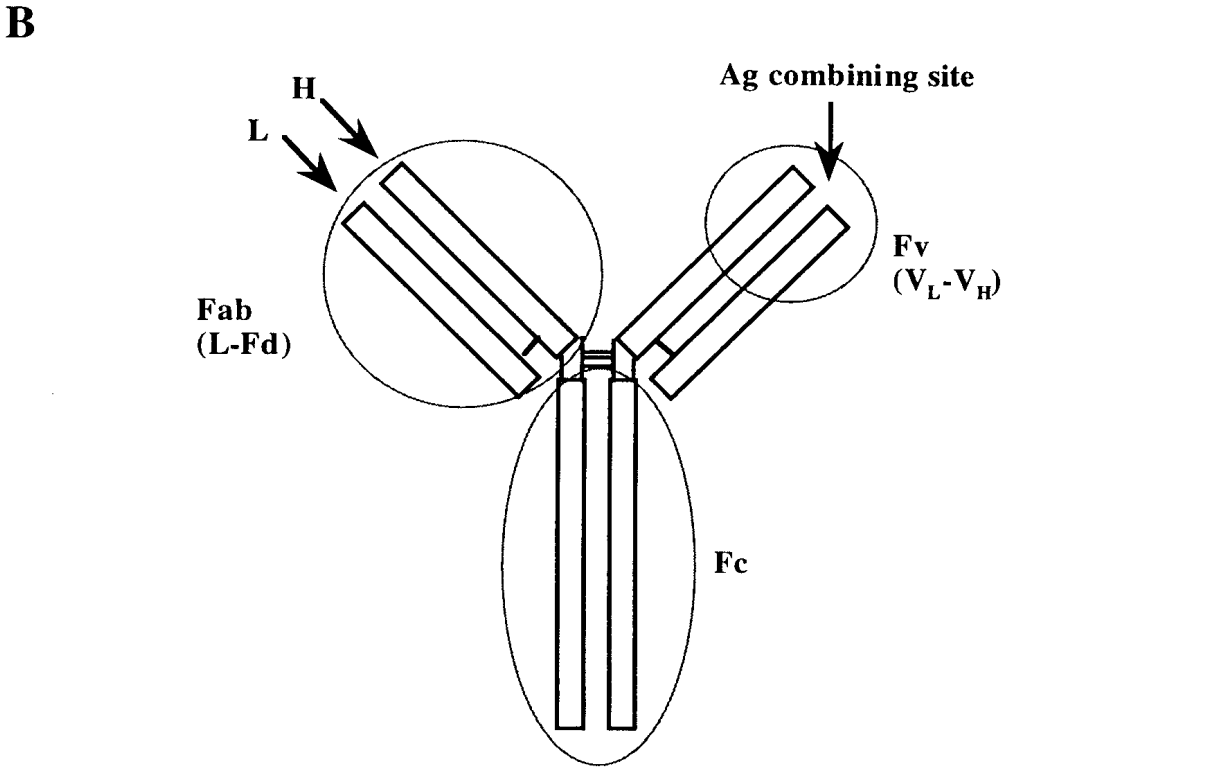


Figure 1.1. Schematic representation of stages in the production of intact antibodies by B cells. A: the gene recombinations giving rise to a unique polypeptide for light chain (L), left, and heavy chain (H), right. B: the intact Ig monomer, comprised of two light chains disulfide bonded (solid lines between chains) and two heavy chains which are also covalently bound through disulfide bonds within the hinge region.

1.1.3: The immunoglobulin fold and interactions between the domains

The conserved three-dimensional (3D) structure of Ig domains was first noted in two high resolution crystal structures of antibody fragments (Schiffer *et al.*, 1973; Poljak *et al.*, 1973) and is illustrated in Figure 1.2. Subsequent crystallographic studies describe the same basic Ig fold in numerous proteins which are evolutionally divergent and distinct from antibody molecules (reviewed in: Bork *et al.*, 1994). The Ig fold consists of two layers of antiparallel β-pleated sheets, closely associated, which are stabilised by an internal disulfide bridge. Thus, the polypeptide is folded into a compact globular domain often referred to as the β-barrel or β-sandwich. In antibodies, these Ig domains fold as independent units along the polypeptide chains and associate through one of the β-sheet layers *via* local pseudo-twofold symmetries with homologous or identical domains on their respective light or heavy chain partners (Schiffer *et al.*, 1973; Poljak *et al.*, 1973).

Although antibody domains are topologically similar there are certain differences in the structures of the variable (V) and constant (C) domains (reviewed in: Davies & Chacko, 1993). These differences are mainly related to the number of β-strands in each layer of β-sheet, although hypervariable sequences in V domains present a special case of diversity in Ab structure (discussed in detail in Section 1.2.2). As is seen with the λ light chain dimer Mcg, the two layers of the CL domain consist of four and three antiparallel β-strands respectively. The two CL domains interact through the twisted and slightly concave four strand layer (Schiffer *et al.*, 1973). This same packing of the four stranded layers of C domains is present at the interface between CL and CH1 in the crystallographic structure of the human Fab New (Poljak *et al.*, 1973). The other heavy chain C domains have this basic four/three (d-e-b-a/g-f-c) β-sheet layered structure (illustrated in Figure 1.2), although the interactions between CH2 domains are mainly through carbohydrate moieties attached to these domains. The CH3 domain interface is similar to that of CL and CH1 (reviewed in: Edmundson *et al.*, 1993; Davies & Chacko, 1993).

A

B

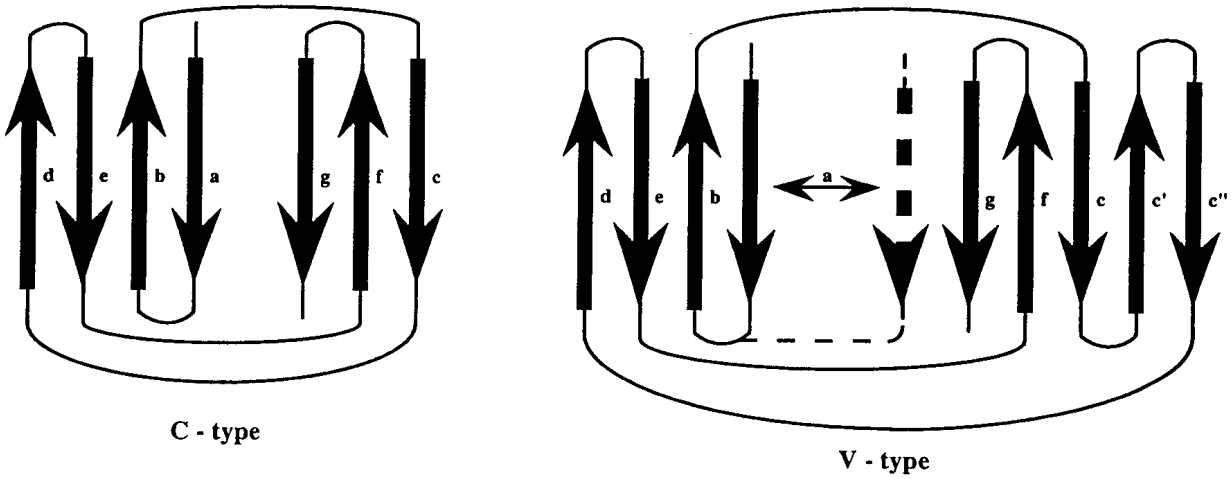


Figure 1.2. The immunoglobulin fold. A: shows a representation of a typical antibody V domain (1VFA). The two-layered anti-parallel β-sheets depicted as flat arrows. B: shows two-dimensional topology diagrams of the Ig type fold seen in antibody C domains (left) and V domains (right) (after Bork *et al.*, 1994).

Antibody variable domains have been classified into a separate topological subclass to the constant domains (Figure 1.2; Bork *et al.*, 1994). While in the majority of V domain structures the four β-strand layer (d-e-b-a) is maintained, the three strand layer (g-f-c) frequently contains an additional two strands (c'-c") (Herron *et al.*, 1989; Fan *et al.*, 1992). However, the c' and c" strands vary considerably in length and in some cases these strands cannot be assigned to the β-type secondary structure and are classified as loops in the structure (Schiffer *et al.*, 1973; Bork *et al.*, 1994). In contrast to the interdomain associations seen in C domains, the V domains of antibodies associate through a local twofold axis by the opposite β-sheet layer (g-f-c-c'-c") which was originally thought to consist of three β-strands (Schiffer *et al.*, 1973). This apparent flipping of domains is facilitated by a short region of the polypeptide termed the 'switch region'. Thus, the pseudo-twofold axes of the VL-VH and CL-CH1 domains converge at an angle in the Fab, called the elbow bend, which is related to the flexibility introduced into the structure by the switch region (Figure 1.3; reviewed in: Padlan, 1994). The elbow bend angle varies considerably between different Fab structures. In contrast to the CL-CH1 domain interface, the VL-VH interface varies considerably between different structures due to the involvement of complementarity determining regions (CDR) in the interactions between the two V domains. The contribution of CDRs to the structure of the Fv (VL-VH) and the antibody combining site will be discussed in detail in the following sections.

1.2: Structure of the immunoglobulin combining site

Paramount to the binding of antigens by antibodies is the structure of the immunoglobulin combining site. The Ig combining site is a surface of complementarity and chemical compatibility with an epitope or structure present on the antigen. Since antibody repertoires are directed against a diverse array of antigens it is clear that the Ig combining site must have the capacity to exhibit an equally diverse pattern of binding site structure (ie. diverse topology and complementary chemical environments). Factors influencing the basic structure and chemical environment of the combining site include: the VL-VH domain interface, the length and structure of CDR loops, the influence of certain framework residues on CDR conformation, the specific residue side-chains available for Ag contact and, although not as well understood, the organisation of solvent molecules within the Ab binding site (reviewed in: Davies & Chacko, 1993; Padlan 1994).

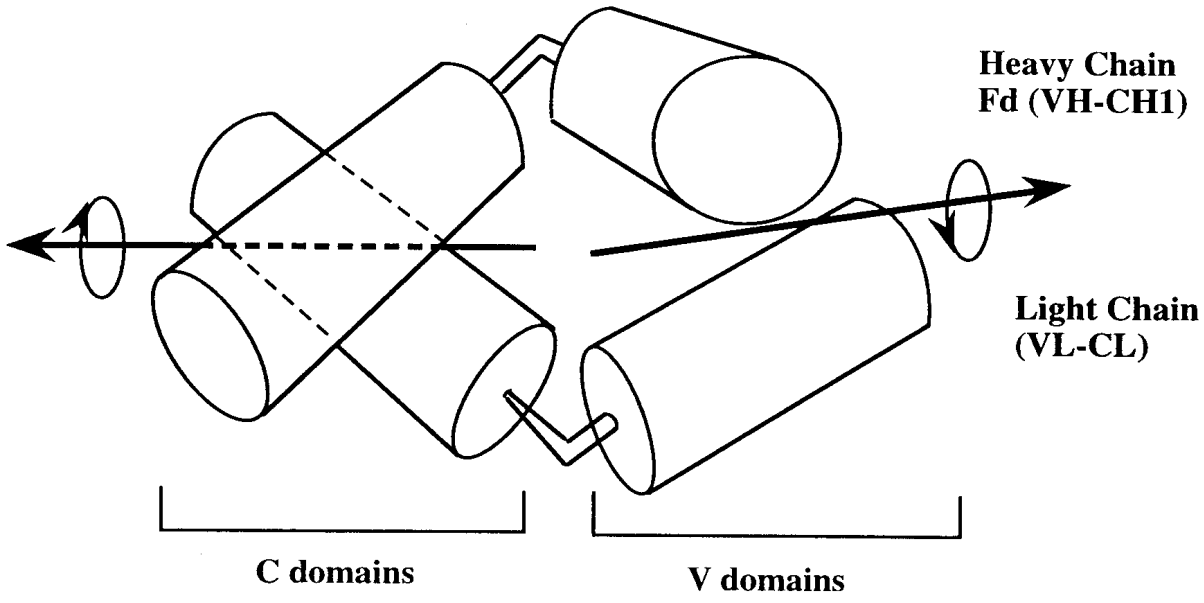


Figure 1.3. Representation of an Fab molecule. The Ig domains are represented as cylinders (after: Edmundson *et al.*, 1975) and the switch regions as bent tubes. The pseudo two-fold axes, forming the elbow bend, relating VH to VL and CL to CH1 are shown as arrows.

1.2.1: Association of VL and VH domains

The basic topology of variable domains and the general structure of the VL-VH dimer have been introduced (Section 1.1.2). The influence that specific aspects of this interaction have on the structure of the Ig combining site will be discussed. A side view of an Fv (Figure 1.4) shows the binding site located towards the top of the heterodimer where the CDR loops are clustered. The pseudo two-fold axis relating VL to VH (Figure 1.3) in a few different antibody structures has been measured to vary between 168 and 180 degrees (Fan *et al.*, 1992). This variation is greater than is the case with CL-CH1 domain pseudo-dyads (167 to 173 degrees; Davies & Chacko, 1993) and concomitantly the shape of the combining site will vary. When the V domains are simplified to cylinders (Edmundson *et al.*, 1975; Herron *et al.*, 1991) further subtle variations in the VL-VH domain interface can be quantified: a swivel of the cylinder in a horizontal plane, a 'barrel' roll of the cylinder around the long axis, a cant or tilt in the vertical direction and a simple translation of the cylinder within the plane (Figure 1.3). It is easy to envisage that even small rigid-body movements of this sort will greatly influence the geometry of the combining site.

The interface of VL and VH has a high level of surface complementarity, similar to the interior of a protein, as indicated by the presence of only a few or in most cases no crystallographic water molecules (Walls & Sternberg, 1992). The extent of this buried surface ranges from 1400-1900 $Å^2$ (Chothia *et al.*, 1985) with the majority of residue side-chains in close van der Waal's (vdW) contact with side-chains from residues of the partner V domain. Not all regions of the V domains contribute to the interface and some regions are clearly more dominant in determining quaternary structure of the Fv. The framework regions or the non-hypervariable regions are relatively conserved in three-dimensional structure and the residues involved in VL-VH contacts are generally conserved within the different crystallographic structures (reviewed in: Davies & Chacko, 1993). Residues of the second framework region in both V domains provide significantly more contacts compared with the other three framework regions. Framework region one generally provides no contacts to the interface. A conserved tyrosine or phenylalanine residue in VL and VH framework three (87L and 91H) and a similarly conserved phenylalanine or tryptophan residue in framework four (98L and 103H) stand out as generic residues of the VL-VH domain interface (reviewed in: Padlan, 1994).

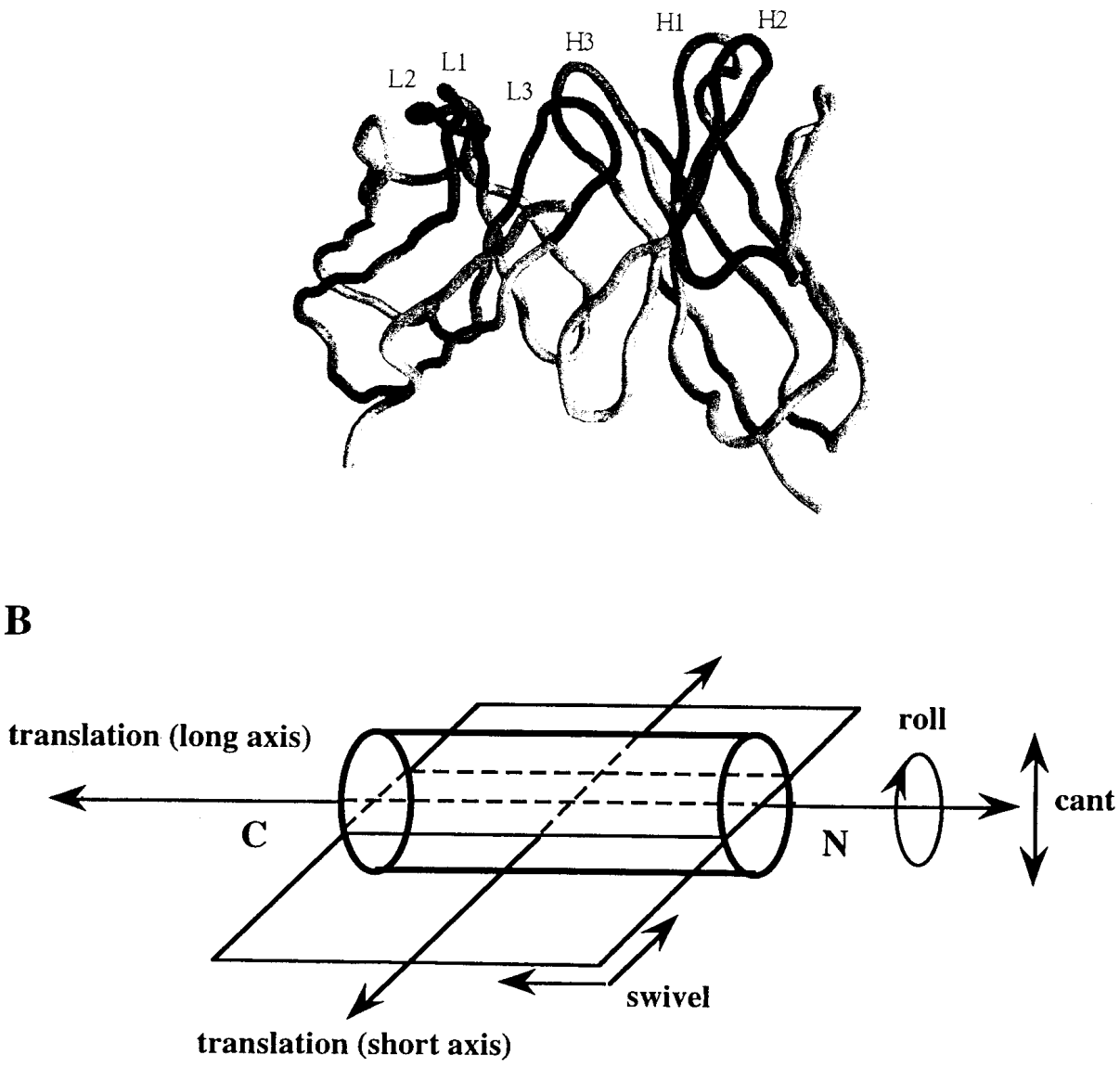


Figure 1.4. Nature of the VL-VH (Fv) association. A: shows a side-view (Cα trace) of an Fv (1VFA), VL is on left and VH on right, CDR are red and are labelled (L for light chain and H for heavy chain). Framework regions are shown in light blue. B: shows the quantitative measurements which relate VH to VL. One V domain has been depicted as a cylinder (terminology of: Edmundson *et al.*, 1975; Herron *et al.*, 1991).

The involvement of the hypervariable residues or complementarity determining region (CDR) residues in the V domain interface is well recognised (reviewed in: Davies & Chacko, 1993; Padlan, 1994). However, the extent that CDR residues contribute to the interdomain contacts varies considerably, from 26 to 57 percent of all the interatomic contacts, between the different antibodies which have been crystallised. If the contribution to the interface by individual CDR loops is considered, the CDR3 regions of both heavy and light chains are the most dominant. The light chain CDR3 interacts with all the heavy chain CDR loops with an average of the contacts being 5% with HCDR1, 28% with HCDR2 and 67% with HCDR3. The heavy chain CDR3 interacts with all the light chain CDR loops with an average of the contacts being 22% with LCDR1, 18% with LCDR2 and 60% with LCDR3. Interdomain contacts between framework regions and CDR are minimal and restricted to framework one and four on the light chain and framework two on the heavy chain (Padlan, 1994). The reduced number of contacts between framework regions on one V domain and CDR residues on the other V domain suggests that the conformations of CDR loops are more influenced by CDRs on the partner domain or by intradomain interactions with framework residues. Intradomain interactions between a few particular framework residues and CDR loops is a major determinant of the conformation of these loops (Chothia *et al.*, 1989; Foote & Winter, 1992) and is the subject of the following section, 1.2.2.

1.2.2: Framework regions and CDR conformation

The concept of the division of variable domains into framework regions and hypervariable or complementarity determining regions has already been introduced. The original finding of regions of conserved primary sequence assigned as frameworks and regions of hypervariable sequence later called CDR was based solely on the comparison of amino acid sequence (Wu & Kabat, 1970; Kabat & Wu, 1971). This early analysis was supported by the subsequent detailed crystallographic description of antibody V domains (Schiffer *et al.*, 1973; Poljak *et al.*, 1973). The framework regions are, as with their primary sequence, highly conserved at the three-dimensional level yet there are certain differences emerging which significantly reflect on both the affinity for antigen and the basic shape of the antibody combining site (Foote & Winter, 1992; Chothia & Lesk, 1987; Chothia *et al.*, 1989). On the other hand, hypervariable regions, when containing the same or similar numbers of residues, were recognised to habitually adopt similar backbone conformations, when compared to the diversity

inherent at the primary sequence level within these regions (Padlan & Davies, 1975; Colman *et al.*, 1977; Fan *et al.*, 1992; Lascombe *et al.*, 1992).

In addition to CDR loops of the same length having a propensity for adopting similar conformations, it is clear that particular conserved residues or residues with similar properties drive the folding of CDR into certain conformations (Chothia & Lesk, 1987). In fact a series of 'canonical' structures has been determined for CDR which were identified by analysing the residues within these loops and neighbouring framework residues using spacial, geometric and chemical criteria (Chothia & Lesk, 1987; Chothia *et al.*, 1989; Tramontano *et al.*, 1990; Chothia *et al.*, 1992; Wu & Cygler, 1993). The canonical theory of CDR conformation evolved out of the observations of Kabat and colleagues (1977; 1978) whom suggested that certain residues, which were shown to be conserved within hypervariable regions (13 positions on light chain and 7 positions on heavy chain), are determinants of the antibody combining site structure rather than relating directly to specificity for antigen. Furthermore, the authors suggested that these residues have fixed three-dimensional positions and so could be used in antibody combining site predictions. Chothia and Lesk (1987) extended this argument by identifying a few 'key' residues which determine the three-dimensional structure of 'canonical' CDR loops through the adoption of certain crystal packing strategies, the capacity for hydrogen bonding or the adoption of unusual *phi*, *psi* or *omega* torsion angles.

Of all the CDR loops the third heavy chain CDR displays the most structural diversity. When compared to the remaining five CDR the human HCDR3 displays the greatest variability in length and primary amino acid sequence. The distribution of normal human immunoglobulin HCDR3s have a range of 2 to 26 residues with a mean length of 12 residues (Wu *et al.*, 1993). The primary amino acid sequence of HCDR3 is determined by the recombination of three gene segments VH-D-JH. Furthermore, diversity is introduced into this region by the addition of non-template (N) nucleotides and junctional variation (discussed in Section 2.1). These processes result in the HCDR3 having the most diversity in primary structure compared to the five other CDRs. It is easy to visualise how this diversity in the primary structure of HCDR3 loops is mirrored at the three-dimensional level. Not surprisingly HCDR3 has not been reliably classified into a conformational class nor predicted accurately using protein modelling techniques (de la Paz *et al.*, 1986; Bajorath & Sheriff, 1996). However, there have been some attempts to group HCDR3 loop conformations (Chothia & Lesk, 1987; Shirai *et al.*, 1996). Presence of an arginine 94H residue at the base of the HCDR3 loop has been shown to pack across the base of the loop to form a salt bridge with aspartic acid 101H, an interaction which in effect restricts the stems of the loop

(Chothia & Lesk, 1987). Furthermore, some HCDR3 will display similar take-off angles due to similar interactions occurring at the base of HCDR3 with residues both within the loop and residues of neighbouring framework regions and CDR loops (Shirai *et al.*, 1996). It remains unclear whether the conformational restrictions present at the base of HCDR3 have a dramatic effect on the overall shape of this loop. Perhaps the majority of HCDR3 conformations, longer than a particular limiting length, is driven more by residues within the loop rather than those present in the rest of the combining site. This argument does not exclude the influence of non HCDR3 residues on the conformation but implies that it is the HCDR3 which is the dominant perpetrator of the seemingly confused folding of HCDR3 loops. Diverse conformations of HCDR3s occur in a binding site environment which at the same time clearly limits the conformations of the other five CDRs (some structures showing this diversity of HCDR3 conformations: Poljak *et al.*, 1973; Lascombe *et al.*, 1992; Fan *et al.*, 1992; He *et al.*, 1992).

1.2.3: Shape (topology) of the immunoglobulin combining site

The conformation adopted by the six CDR loops within the Fv molecule determines the basic topology of the immunoglobulin combining site. In the structure of the Bence-Jones light chain dimer Mcg a large solvent-accessible cavity is present at the VL-VL domain interface (Schiffer *et al.*, 1973). Obviously this structure is not a 'classical' antibody yet this cavity-style binding site has been subsequently observed in VL-VH domain pairings (He *et al.*, 1992; Arevalo *et al.*, 1993; Bossart-Whitaker *et al.*, 1995). Poljak and coworkers (1973) describe a binding surface which is relatively flat covering an area of 25 x 20 Å with a shallow groove (15 x 6 x 6 Å) rather than a cavity at the VL-VH domain interface of the human Fab New. It is possible to group Ig combining site topology into (1) cavities or 'pockets' (Arevalo *et al.*, 1993; Bossart-Whitaker *et al.*, 1995), (2) shallow grooves (Poljak *et al.*, 1973; Sheriff *et al.*, 1987) to deeper grooves (Herron *et al.*, 1991) (3) relatively flat surfaces (Braden *et al.*, 1994), to even a slightly convex combining site surface (Fan *et al.*, 1992). These topological variations in Ig combining sites perhaps reflect the powerful discrimination and selectivity required by antibodies for different antigens.

Geometric restriction of combining site topology has been predicted to exist within the antibody specificities so far investigated at a primary sequence level (Vargas-Madrazo *et al.*, 1995). Since five out of the six CDR frequently are of a 'canonical' class,

Vargas-Madrazo *et al.* (1995) were able to approximate the combining site geometry using the 'canonical' conformation theory (Chothia & Lesk, 1987). Of the three hundred possible combinations of canonical structures the majority of the 381 sequences of VL-VH domain pairings analysed were clustered within ten distinct combining site geometries. Although the effect of HCDR3 conformation on the structure of the binding site was not accounted for in this study, it would appear that the combinatorial diversity of the other five CDR loops is restricted to ten or slightly more combinations in the effective immune repertoire. The presence of strong geometric restrictions within antibody combining sites in general requires further corroboration through crystallographic or exhaustive modelling investigations which account for the influence of HCDR3 upon the 3D structure of the Ab combining site.

1.2.4: Common residues located within immunoglobulin combining sites

Certain residues appear to have a propensity to localise to the six CDR of the Ig combining site in comparison to the framework regions of the V domains. Residues displaying increased tendency to be found in CDR loops do not necessarily participate in the antigen specificity but can sometimes play a structural role in providing combining site rigidity. Kabat and colleagues (1977) noted that there is an increased frequency of histidines and asparagines within CDR loops. This observation has been supported and extended to include tyrosine residues by Eduardo Padlan (1990) in a more thorough analysis of murine V domain sequences. This study showed that asparagines and histidines are eight, and tyrosines three, times more likely to occur within CDR loops in comparison to framework regions. The increased presence of aromatic residues within the Ab binding site (tyrosines and tryptophans in particular) was supported in two crystal structures of anti-lysozyme Fab molecules complexed with hen egg lysozyme (HEL), D1.3 (Amit *et al.*, 1986) and HyHEL-10 (Padlan *et al.*, 1989). Many of the aromatic side-chains in these combining sites were observed to interact with the HEL antigen through van der Waal's contacts and hydrogen bonding. Furthermore, the occurrence of apolar aliphatic amino acids (alanine, valine, isoleucine and leucine) is markedly decreased in CDR sequences compared with framework regions and proteins in general. Acidic and basic residues are found in similar levels within antibody combining sites and V domains when compared to other proteins (Padlan, 1990).

The side-chains of residues within Ig combining sites also differ from other proteins in the degree to which they are exposed to solvent. Comparison of the solvent accessibility of side-chains in seven antibody structures showed that the asparagines are frequently buried with their side-chains in potential hydrogen bonding distance to backbone atoms while the aromatic residues are more likely to be exposed to solvent (Padlan, 1990). Because of the unusual tendency to be buried within the combining site, asparagines are likely to provide rigidity to the combining site structure, although not to the extent provided by framework regions (Section 1.2.2). The unexpected solvent exposure of aromatic side-chains in combining sites has significant implications for the role of these residues in the process of antigen complexation which will be discussed later (Section 1.3). The unusual abundance and level of solvent exposure of certain residues in CDR sequences may provide the essential structural features which enhance the interaction of antibody combining sites with antigens.

1.3: Antibody/antigen complexation

The interaction of antibody with antigen involves thermodynamic parameters which are derived from experimental measurements or approximated from molecular structures provided by X-ray crystallography. Direct correlations between the thermodynamic measurements, gained through techniques such as microcalorimetry, and the 3D structure of the protein are difficult to obtain because of a masking effect of the reorganisation of water molecules during the process of complex formation, which is not accounted for in crystal structures (Singha *et al.*, 1996). However, interesting observations regarding antibody/antigen complexation have been gained from both thermodynamic and structural experiments.

1.3.1: Relationship of thermodynamic parameters and structure of antibody/antigen complexation

Elucidation of relationships, between the thermodynamics of an antibody binding to antigen and the X-ray structures of antibody/antigen complexes, is important in understanding the general principles of antigen recognition. The wealth of Ab/Ag complexes solved at medium to high resolution (reviewed in: Padlan, 1994) have revealed the importance of hydrogen bonding, van der Waal's contacts, salt-bridges (the neutralisation of charges) and solvent molecules (exclusion of which leads to the

hydrophobic effect and a recently discovered role of water as bridging or 'adaptor' molecules, discussed in Section 1.3.5). The most studied protein antigen in complex with a variety of Ab fragments is hen egg-white lysozyme (HEL). This antigen has now been crystallised with several antibody fragments including those which bind to different structural epitopes (e.g. HyHEL-5 and D44.1 bind to an overlapping epitope while D44.1 binds to the distal region of HEL, Braden *et al.*, 1994; Sheriff *et al.*, 1987) and combining site mutants of the wild type antibodies (Ysern *et al.*, 1994; Fields *et al.*, 1996) have been determined crystallographically in complex with HEL. Furthermore, for many of the Ab/HEL complexes there is at least some thermodynamic data available (Ysern *et al.*, 1994; Fields *et al.*, 1996) which warrants some elaboration.

Thermodynamic parameters of protein-protein interactions are described by Gibb's free energy ($\Delta G°$) which is related to the equilibrium of binding ($\Delta G° = RT - \ln K_{eq}$) and is comprised of enthalpic and entropic terms through the general equation: $\Delta G° = \Delta H - T\Delta S$, where ΔH is the change in enthalpy and $T\Delta S$ is the entropic contribution to binding. Techniques such as titration microcalorimetry enable the change in heat capacity during protein-protein interactions to be measured (Singha *et al.*, 1996). Thus changes both to free energy and enthalpy of the system are derived using titration calorimetry and the entropic term can be calculated from its relationship to these values. The formation of an antibody/antigen complex results in a decrease in the Gibb's free energy with larger decreases in free energy being correlated with increased binding constants. Since complexation involves reorganisation of both solvent and protein molecules at the antibody/antigen interface the entropy of the system must also decrease. To compensate for the energy expended in reducing the entropy a larger decrease in enthalpy is necessary. The general characteristics of antibody combining sites discussed (Sections 1.2.3-4) tend to minimise the loss of entropy and favour a sizeable reduction in the enthalpy during complexation with antigen.

1.3.2: The role of hydrogen bonds, van der Waal's interactions and neutralisation of charged side-chains in antibody/antigen complexation

The formation of hydrogen bonds between antibody and antigen polar side-chains is one of the major contributing factors to the large decrease in enthalpy. In practical terms the contribution of one weak hydrogen bond (~5.73 kJ/mol) relates to a ten fold increase in affinity, while a strong hydrogen bond (~17.24 kJ/mol, involving a charged

group) corresponds to a thousand fold increase in affinity (Fersht, 1985; Padlan, 1994). In the complex of D44.1 and HEL there are a total of five hydrogen bonds between antigen and antibody. Two water molecules in this D44.1/HEL complex are involved in a hydrogen bonding network connecting antigen and antibody residues (the role of water molecules at the antibody/antigen interface is discussed in Section 1.3.5). As mentioned the antibody HyHEL-5 binds to a very similar location on HEL and indeed a comparison of the epitope residues and the VH sequences of the two anti-HEL antibodies shows that these sequences are very similar (Braden *et al.*, 1994). However, the affinity of HyHEL-5 is substantially higher for HEL (Kd = 4 x 10^{-10} M; Benjamin *et al.*, 1992) than that of the D44.1 antibody (Kd = 1.4 x 10^{-7} M; Braden *et al.*, 1994). It remains unclear at the molecular level why this disparity in affinities exists. Only two of the five hydrogen bonds seen in the D1.3 complex are common to the HyHEL-5 complex with HEL while the three salt bridges are maintained (Braden *et al.*, 1994; Chacko *et al.*, 1995). While not discounting the possible differences in van der Waal's interactions, it is possible that the exquisite affinity of HyHEL-5 for HEL is solely due to the presence of 13 potential hydrogen bonding contacts between antigen and antibody residues (Sheriff *et al.*, 1987; Chacko *et al.*, 1995). The substantial enthalpic contribution made by hydrogen bonding as seen in the Fab/HEL complexes is supported in other crystallographic structures of Fab/Ag complexes (some examples: 12 potential H-bonds in BV04-01 Fab/dT_3 complex, Herron *et al.*, 1991; 9 potential H-bonds in N10 Fab/SNase complex, Bossart-Whitaker *et al.*, 1995; 13 potential H-bonds in NC41 Fab/neuraminidase complex, Colman *et al.*, 1987).

The complementarity of the contacting surfaces of antibody and antigen is such that many attractive van der Waal's forces exist between the residues of the combining site and the epitope. A predominance of vdW interactions is particularly obvious in combining sites which have a high affinity for haptenic antigens such as the anti-fluorescein Fab 4-4-20 (Herron *et al.*, 1989) where the xanthonyl ring of the fluorescein is closely flanked by a tyrosine (37L) and two tryptophan residues (33H and 101L) with the enolic oxygens pointing towards a histidine (31L) and an arginine (39L) residue. Also the *o*-carboxyphenyl ring of fuorescein is stacked below two tyrosine rings from HCDR3 (103H and 102H). One gap exists in an otherwise continuous vdW contact surface between fluorescein and the combining site of the 4-4-20 Fab. This gap is most likely due to the nature of the immunogen used to immunise the mouse that produced the 4-4-20 IgG. The carrier protein had a lysine residue linked to the isothiocyanate in the *para* position of the phenolic group of fluorescein. The authors point out that this unoccupied space in the crystal structure of the 4-4-20 Fab/fluorescein complex would presumably be filled by the lysine side-chain of the carrier protein (Herron *et al.*, 1989). Only two residues (histidine 31L and serine 96L)

are in positions to make hydrogen bonds and an arginine residue (39L) is positioned close enough to make an electrostatic contact with the enolic oxygens of fluorescein (2.8 Å). The dominance of vdW interactions in these hapten specific antibodies is clearly the important factor in their high binding affinities for antigen (4-4-20 has Ka = 3.4×10^{-10} M in aqueous solution, Herron *et al.*, 1989). An obviously favourable enthalpic contribution of van der Waal's contacts in the interaction of antibody combining sites with antigen is evident in other systems where the antigens are larger than the haptenic complex of 4-4-20 described above (reviewed in: Padlan, 1994; Davies & Chacko, 1993).

The contribution of electrostatic interactions to antibody/antigen complex formation is less dominant in the affinity of binding when compared to hydrogen bonding and vdW interactions (reviewed in: Padlan, 1994). However, the process which results in the neutralisation of oppositely charged side-chains may be important in determining the initial orientation and association of the Ag epitope within the Ab combining site (Stanfield *et al.*, 1990). While crystallographic structures do not reveal the nature of solution electrostatics (reviewed in: Gilson, 1995) the structures of complexes of Ab/Ag do allow the identification of specific residues involved in the neutralisation of charges and the formation of salt-bridges (Singha *et al.*, 1996; Padlan, 1994). In the structure of the BV04-01 Fab/d(pT)$_3$ complex there is only one potential electrostatic interaction between the central phosphate oxygen of the first and second thymidine nucleotides and an arginine (52H) from HCDR2 (Herron *et al.*, 1991). This ion pair interaction is towards the middle of a long binding groove in the combining site of BV04-01 Fab and so could serve initially to localise and orientate the trinucleotide within this auto-antibody's combining site. The surface of interaction between HEL and the Fab HyHEL-5 is relatively flat with only two salt-bridges formed between Ab and Ag (Sheriff *et al.*, 1987). Again the electrostatic interactions are localised to, in this case, a shallow groove between VL and VH and these charged residues may be involved in the initial association of the Ag epitope within the Ab combining site. Clearly, in comparison to other interactions such as hydrogen bonding and vdW's contacts, electrostatic interactions are relatively few and their contribution is suggested to be related to the initial dynamics of complex formation and not in the stability of the association between antibody and antigen (Padlan, 1994; Stanfield *et al.*, 1990).

1.3.3: Contribution of aromatic residues to antibody/antigen complexes

The increased likelihood of aromatic residues to be localised within antibody combining sites (Padlan, 1990; Section 1.2.4) has several implications on the energetics of antigen binding. The planar ring systems of tyrosine and tryptophan display an inclination to exclude water from their vicinity through their ability to form close van der Waal's interactions, mainly through π electron stacking, with similar ring systems. Since these residues are unusually exposed to solvent in the antibody combining site it is expected that when these aromatic side-chains are buried by the interaction with antigen they contribute substantially to the binding energy through van der Waal's and hydrophobic effects (Padlan, 1990). Hydrogen bonding can occur through the aromatic rings themselves or through polar atoms in the side-chains of these aromatic residues (Levitt *et al.*, 1988). Moreover, when compared to aliphatic side-chains with similar numbers of atoms, aromatic ring systems have a substantially lower conformational flexibility which results in an enhanced contribution of aromatic residues to binding energy through the lower entropy of heterocyclic compared to branched chain systems (Novotony *et al.*, 1989; Padlan, 1990). The substantial contribution of aromatic residues to antibody/antigen interactions has been reported in many X-ray crystallographic structures of complexes of Fab molecules with antigens ranging from proteins to oligonucleotides and to haptens (selected examples: Colman *et al.*, 1987; Herron *et al.*, 1989; Herron *et al.*, 1991; Fields *et al.*, 1996).

1.3.4: Buried surface area at the interface of antibody and antigen

In order to form a complex of antibody and antigen extensive reordering of the solvent is required. In the majority of cases most or all the solvent molecules are excluded from the interface which results in a decrease in the effective dielectric constant of the contact surfaces of antibody and antigen (reviewed in: Davies & Chacko, 1993). The packing of side-chains in this interface is not significantly different from that seen in the interior of an antibody domain (Padlan, 1994) supporting the argument that the dielectric environment at the antibody/antigen interface is similar to the interior of a protein. In comparison to a solvated protein surface, this decreased dielectric environment favours the tight association of nonpolar groups, a phenomenon often referred to as the hydrophobic effect. The strength of hydrogen bonds and vdW interactions will also be increased compared with the same interactions occurring in a

solvated environment. However, the extent of the polar and apolar surface areas buried during antibody binding is poorly correlated with thermodynamic parameters (Singha *et al.*, 1996).

The surface area of the interaction between antibody and antigen (often referred to as the buried surface area) differs significantly between the different types of antigen that a particular antibody recognises. By far the largest buried surface area occurs when an antibody complexes with a large protein antigen. The buried surface area of the Ab combining site is 537 $Å^2$ for D1.3/HEL and 882 $Å^2$ for NC41/Neuraminidase (Padlan, 1994; Amit *et al.*, 1986; Colman *et al.*, 1987). This contact surface area is reduced to 439 $Å^2$ for the myohaemerythrin peptide complexed with B13I2 Fab (Padlan, 1994; Stanfield *et al.*, 1990) and the BV04-01 Fab complex with its trinucleotide antigen is similar at 443 $Å^2$ (Padlan, 1994; Herron *et al.*, 1991). The combining site surface area involved in antigen binding is further reduced with haptenic antigens yielding values as low as 138 $Å^2$ to 247 $Å^2$ for McPC603 Fab/phosphocholine and the 4-4-20 Fab/fluorescein complexes respectively (Padlan, 1994; Herron *et al.*, 1989). These variations in the extent of buried surface areas do not necessarily correlate with binding affinity and more specifically the expected changes in enthalpy and entropy upon binding antigen (Singha *et al.*, 1996). However, if the extent to which solvent is removed from the interface (leading to an increased hydrophobic effect) and the complementarity of the surfaces are taken into account it may be possible to determine accurate binding constants and indeed thermodynamic parameters from the computational analysis of X-ray structures. It is important to note that different structural types of antigen (proteins, peptides, nucleotides, haptens) will contact different amounts of the Ab combining site surface and thus recruit all or sometimes only a few CDR loops into the interaction.

1.3.5: Role of water at the antibody/antigen interface

The interface between antibody and antigen is a highly complementary surface where indentations in one surface are generally occupied by protrusions on the other surface (reviewed in: Padlan, 1994). If there is a gap or hole in this complementary surface it is generally filled by one or two well ordered water molecules. Water molecules can provide a bridging hydrogen bonding network between antibody and antigen residues which in effect stabilises the interaction between surfaces that on first appraisal appear to lack absolute complementarity (reviewed in: Davies & Chacko, 1993). The role of water molecules as 'adaptors' of imperfections in the antibody/antigen interface was

demonstrated unequivocally in a thermodynamic and X-ray crystallographic study of three Fv mutants complexed with their lysozyme antigens (Fields *et al.*, 1996). The parent antibody, D1.3, was previously shown to have a substantial number of water molecules at the interface between an Fv of D1.3 and HEL (Bhat *et al.*, 1994). These water molecules were clearly aligned to form bridging hydrogen bonds between antibody and antigen residues in places where there would otherwise be a destabilising gap in the interface complementarity. When the combining site was selectively mutated in D1.3 there was a reorganisation of solvent molecules at the antibody/antigen interface. In a mutant Fv where tyrosine 50 of the light chain was mutated to a serine residue a direct hydrogen bond to HEL is lost and is replaced by interactions mediated by water molecules. The expected increase in enthalpy of the mutant resulting from the removal of the strong hydrogen bond between tyrosine 50L and HEL did not occur. Instead only a small increase in enthalpy of 9 kJ/mol (estimated increase of ~17.24 kJ/mol for one strong H-bond) between native D1.3 and the serine mutant Fv molecules interaction with HEL was measured (Fields *et al.*, 1996). Evidence of a compensatory hydrogen bond network provided by water molecules explains the relatively small change in enthalpy in the D1.3/HEL system (Fields *et al.*, 1996) and suggests a more general role of water molecules in providing stabilising interactions at the antibody/antigen interface.

1.3.6: Conformational changes in antibody and antigen due to complex formation

The binding of antigen by antibody can result in changes in the conformation of either molecule. It is not clear, at this stage, if these conformational changes migrate along the Ab molecule to enable effector functions or the transmission of a signal to the B cell, resulting in activation or a deleterious effect on the cell. Conformational change upon binding of haptens into the combining site of the Mcg light chain dimer was the first crystallographic evidence for an induced-fit upon antigen binding (Edmundson *et al.*, 1974). The binding of antigen can result in small adjustments in the positions of side-chains, major shifts in the position of CDR loops or rotations and translations of entire V domains (reviewed in: Davies & Chacko, 1993; Padlan, 1994). Comparison of the unliganded and liganded structures of the D1.3 anti-HEL Fab showed only minor conformational changes to the combining site structure, which were restricted to rotations of side-chains (Amit *et al.*, 1986). In contrast, the liganded form of BV0-401 shows major adjustments of the VL-VH domain interface which effectively widen the binding site to accommodate the trinucleotide ligand. The conformational change was

quantified (see Figure 1.4) as a displacement between the unliganded and liganded BV0-401 Fab structures: for VH a roll of -3.6°, a cant of 1.3°, a swivel of -1.2° and a translation of 0.1Å; for VL a roll of -0.9°, a cant of -0.2°, a swivel of -2.8° and a translation of 0.1Å (Herron *et al.*, 1991). Although these large adjustments of variable domains upon antigen binding are not always evident, CDR loops can shift dramatically as was seen with the HCDR3 of an anti-arsonate Fab R19.9 which moved away from the other CDRs by as much as 6Å to expose a binding cavity for the ligand (Lascombe *et al.*, 1992). Obviously these major adjustments of the Ab combining site require a significant enthalpic contribution. However, the increased complementarity for antigen introduced through conformational changes may more than compensate, in terms of increased affinity, for such a large enthalpic contribution to binding Ag.

While conformational changes occur within the local region of the interaction between antibody and antigen, there is less evidence for the transmission of these changes to distant regions of the antibody molecule. The only X-ray crystallographic evidence for the propagation of such changes between antibody domains is the changes to the elbow bend of an Fab (NC6.8) caused by complexation with its artificial sweetener ligand, NC174 (Guddat *et al.*, 1994). The combining site of NC6.8 showed conformational changes in line with an induced fit mechanism of binding. Furthermore, the C domain dimer of the Fab molecule shifted dramatically upon antigen binding. This was quantified as a roll of -6.0°, a cant of 35.3°, a swivel of -5.0° and a translation of 10.9 Å from its original position in the unliganded Fab and resulted in a 31° change in the elbow bend relating the V and C domain pseudodyads of which the major component was a tilting motion of the constant domains. These dramatic changes in the C domain module of the Fab upon antigen binding were not attributable to any peculiarities in the crystal packing or the protein itself (Guddat *et al.*, 1994). Clearly, if the observed changes do occur in Fab molecules upon antigen binding it is possible that conformational changes can be propagated along the antibody molecule to enable effector functions and drive intracellular signalling. Since such dramatic conformational changes have not been seen in other complexes of Fab with antigen, it remains a point of contention how signals are propagated along the antibody molecule as a result of complexation with antigen.

1.4: Polyreactive immunoglobulins

Polyreactive immunoglobulins are capable of binding multiple structurally unrelated antigens of both self and non-self origin. This property distinguishes them from monospecific Ig which bind to a single Ag or a group of structurally similar antigens. The presence of polyreactive Ig within the sera of normal and diseased humans and mice is now well established (reviewed in Avrameas & Ternynck, 1993; Coutinho *et al.*, 1995). However, there remains a lack of conclusive evidence relating the polyspecific binding phenomenon to a biological role for these antibodies. Although a clear function for polyreactive Ab has not been established, there is evidence to support several proposed functions for polyreactive Ig: (1) a role in the innate immune defence system (Hartman *et al.*, 1989; Kahn *et al.*, 1992); (2) a homeostatic role in controlling the immune system through immune networks (Adib *et al.*, 1990; Berneman *et al.*, 1992; Barbouche *et al.*, 1993), and (3) the concept of polyreactive Ig being the precursors of monospecific Ig (Ikematsu *et al.*, 1993; Houdayer *et al.*, 1993). Moreover, the structural basis of immunoglobulin polyreactivity is not understood, even though there is extensive 3D structural knowledge available on monospecific antibodies. The functional roles of polyreactive Ab and the hypotheses of the structural mechanisms of polyreacitive Ag binding will be discussed (Sections 1.4.4-5).

1.4.1: Basic immunology of polyreactive antibodies

While IgM represents the predominant isotype of polyreactive antibodies other isotypes including IgG polyreactive antibodies have been observed in both healthy and diseased humans and mice (Berneman *et al.*, 1992; Ikematsu *et al.*, 1993; Ditzel *et al.*, 1996). Human B cells which express the pan T cell antigen CD5/Leu-1 (Ly-1 in mice) have been established as the major population of B cells synthesising polyreactive antibodies (Burastero *et al.*, 1988). It has been suggested that CD5 positive B cells represent a separate population (B1a cells) from those B cells negative for the CD5 antigen (B2 cells) (Herzenberg & Kantor, 1993). Such a separation into B1 cells which predominantly express polyreactive Ig and B2 cells which predominantly express monoreactive Ig may be artificial. In fact, polyreactive Ig are also produced by B cells which do not express the CD5 antigen (B1b cells). These B1b cells often express similar levels of surface Ig and display other phenotypic characteristics which are more similar to CD5 positive than to conventional CD5 negative mature B cells (Mackenzie *et al.*, 1991). Although not exclusive in the production of polyreactive Ig, the CD5

positive B cell produces the majority of polyreactive Ig found in both healthy and diseased humans and mice.

The incidence of polyreactive Ig secreting B cells (as characterised by $CD5^+$/sIgM B cells) is between 5-30% in the normal human peripheral blood B cell population and somewhat higher in cord blood lymphocytes (Logtenberg *et al.*, 1987; Lydyard *et al.*, 1990; Barbouche *et al.*, 1992). This high incidence of polyreactive Ig producing B cells strongly suggests that there is a function for these B cells in the healthy developing and mature immune systems. Polyreactive Ig secreting B cells are increased in certain autoimmune diseases such as rheumatoid arthritis, Sjögren's syndrome and occasionally with systemic lupus erythematosus (Burastero *et al.*, 1990; Okawa-Takatsuji *et al.*, 1992). In these auto-immune conditions the polyreactive antibodies are of the IgM isotype and frequently recognise the target of the pathogenic monospecific autoantibody, although with a lower affinity (e.g. Fc of IgG or rheumatoid factor activity) as well as binding to other auto (self) and xeno (foreign) antigens. While the significance of polyreactive Ig in normal immunity is still a mystery, an increased recruitment of B cells producing these antibodies occurs with certain chronic infections such as *Plasmodium chaubadi* in mice and leprosy in humans (Locniskar *et al.*, 1988), implying an involvement of polyreactive Ig in the skirmish against persistent infections. Interestingly, a monoclonal polyreactive Ig component is present in certain human malignancies of B cell origin such as B cell myeloma, Waldenström's macroglobulinaemia and most notably chronic B lymphocytic leukaemia (Axelrod *et al.*, 1991; reviewed in: Caligaris-Cappio, 1996). However, the significance of the monoclonal Ig displaying polyreactive antigen recognition and its potential role in the pathogenesis or evolution of the malignancy itself is basically undetermined.

1.4.2: Antigen arrays bound by polyreactive antibodies

The polyreactive antibody repertoire is skewed towards the recognition of auto-antigens (reviewed in: Avrameas & Ternynck, 1993), which differs from the situation with monoreactive Ab, where autoreactivity is suppressed by the mechanisms of deletion or anergy (Goodnow *et al.*, 1988; Goodnow *et al.*, 1989). For example Schutte and colleagues (1991) examined the polyreactive Ag binding of monoclonal polyreactive Ig secreted by EBV immortalised $CD5^+$ B cells, obtained from healthy adult PBL. Of the three polyreactive Ab examined, all three were reactive with self antigens and two recognised non-self antigens including haptens and tetanus toxoid. This basic

observation of multireactivity to self and non-self antigens has been upheld by many more extensive studies of polyreactive Ig obtained from healthy and diseased humans and mice (reviewed in: Coutinho *et al.*, 1995). The antigens recognised by a monoclonal polyreactive Ig may include globular and fibrous proteins, nucleic acids, carbohydrates and haptens; an indication that the epitope recognised is most likely not conserved between these structurally dissimilar antigens. It is remarkable that a single Ab combining site, which has been classically associated with the specific molecular recognition of a single Ag, is capable of binding several structurally dissimilar Ag, presumably through different binding modes, in a polyreactive manner.

1.4.3: Gene utilisation and somatic mutation of polyreactive immunoglobulins

While the majority of V domains of IgM polyreactive Abs show little or no mutations from germline gene nucleotide sequences (ie. no somatic mutation resulting from affinity maturation), some sequences in IgG polyreactive Ab have been shown to contain many mutations, particularly within CDR sequences (Baccala *et al.*, 1989; Schutte *et al.*, 1991; Sanz *et al.*, 1989; Ikematsu *et al.*, 1993; Ditzel *et al.*, 1996). Furthermore, at least for certain polyreactive Ig, a restricted utilisation pattern of VL, VH and to a lesser extent D and J genes has been observed (Hartman *et al.*, 1988; Schutte *et al.*, 1991; Brown *et al.*, 1992; Newkirk *et al.*, 1993; Ditzel *et al.*, 1996). However, it should be noted that in some investigations an unbiased use of these genes has been observed indicating that the polyreactive gene repertoire is potentially larger than was initially expected (Hartman *et al.*, 1989; Pascual *et al.*, 1992). The gene usage patterns and somatically generated diversity of polyreactive Ig are examined in more detail later (Chapter Two).

1.4.4: Proposed roles of polyreactive antibodies in the immune system

(1) Innate immune defence. The presence of polyreactive Ig in the sera of most phylogenetically primitive vertebrates such as fish and sharks (Avrameas & Ternynck, 1993) suggested that these antibodies play a role in the initial response to pathogens. Perhaps polyreactivity reflects an innate immune role for antibodies which is more primitive than the adaptive role played by monoreactive antibodies. Indeed early in

ontogeny (within foetal cord blood) the numbers of CD5 positive B cells expressing polyreactive Ig is higher than at later stages of development (Mackenzie *et al.*, 1991). The increased amounts of polyreactive IgM in the sera of foetuses may act in an innate capacity to strengthen the maternally derived immunity while the developing immune system is unable to fulfil its protective role. Furthermore, certain infections elicit a rise in the levels of polyreactive Ig producing B cells in the peripheral circulation of humans and mice (Ueki *et al.*, 1990; Kahn *et al.*, 1992). The 'innate' recognition by polyreactive Ig of bacterial and other pathogens may allow an infection to be kept at bay until high affinity monospecific Ig is produced.

(2) Immune networks. It has been suggested that one of the functions of polyreactive Ig is in the regulation and enhancement of the immune system. It has been demonstrated that the natural IgM in normal mouse serum can inhibit the binding of polyreactive IgG antibodies to certain self Ag. This inhibitory role was not seen in the case of pathological anti-DNA IgG antibodies (Adib *et al.*, 1990). However, the suppression of the polyreactive autoreactive IgG by normal mouse IgM supports an argument for a function of certain polyreactive IgMs in an anti-idiotypic network. Furthermore, it is recognised that polyreactive Ig can frequently recognise other molecules of the immune system, including MHC class I and II molecules, T cell membrane antigens (e.g. CD4 and CD8) as well as cytokines and their receptors (Berneman *et al.*, 1992; Hansen *et al.*, 1994). These reactivity patterns may implicate a broader role of polyreactivity in controlling the immune system than simply suppressing the autoreactivity of natural IgG antibodies.

(3) Precursors of monoreactive Ig. Rather than existing as an independent population it has been suggested that polyreactive Ab producing B cells are the precursors of high-affinity monoreactive Abs. Houdayer and colleagues (1993) observed the presence of four monoclonal antibodies which they suggested were the products of different stages of clonal evolution from a polyreactive to a monoreactive Ig component. This conclusion evolved from the strong similarities between the N-terminal sequences of the different antibody H chains and from the conservation of a private epitope on the *kappa* L chain of the antibodies. Of the four antibodies, two monospecifically bound to tubulin while a third bound to tubulin and was also polyreactive in binding to other self antigens (ie. actin, myosin and ssDNA). The fourth antibody failed to bind to the four self Ag tested. While this evidence is limited by the necessary assumption that sequence similarity is equivalent to clonal relatedness, it may support the selection of a polyreactive Ig and its subsequent Ag selected mutation into a monospecific antibody.

1.4.5: Hypothesised structure of polyreactive immunoglobulin combining sites

(1) Reduced binding site diversity. The inherent ability of polyreactive Ig to bind to multiple Ag suggests that the binding sites of these Ab are relatively conserved compared to those of monospecific Ab. This is reflected in the frequent observations of a restricted pattern of V gene usage and a lack of somatic mutation occurring in Ab sequences within the polyreactive Ig repertoire. Furthermore, it has been argued that the polyreactive Ag binding site is relatively large, plastic or 'sticky' (high number of aromatic side-chains within CDRs) and so able to accommodate several dissimilar Ag, presumably one at a time, through a variety of binding modes (Padlan, 1994).

(2) HCDR3 as the central determinant of polyreactivity. It is possible that some unusual characteristic of the HCDR3 may result in polyreactive antibody binding. At least for some polyreactive Ig the binding to multiple Ag has been shown to be dominated by the HCDR3 (Ichiyoshi & Casali, 1994; Martin *et al.*, 1994; Ditzel *et al.*, 1996). However, extensive variability in both length and sequence have been noted for HCDR3 loops of polyreactive antibodies (Pascual *et al.*, 1992) which makes it difficult to identify sequence motifs, within HCDR3s, responsible for polyreactivity. It has been suggested that polyreactive binding is a result of a high level flexibility of the HCDR3 (Adib-Conquy *et al.*, 1994; Ditzel *et al.*, 1996) or a high number of aromatic residues in this loop of polyreactive Ab compared to monoreactive Ab (Padlan, 1994). Even though the HCDR3 of an antibody may determine it being either polyreactive or monoreactive, no structural correlates have been identified which distinguish a HCDR3 sequence as polyreactive or monoreactive.

1.5: The chronic B lymphocytic leukaemia model of polyreactivity

The B cell malignancy, chronic B lymphocytic leukaemia (B CLL), is characterised by the incessant accumulation of long-lived CD5 positive B cells which exhibit the morphology of mature lymphocytes (Caligaris-Cappio *et al.*, 1993). Although the origin of the B cells comprising the malignant clone in B CLL is not definite, these B cells possess many of the properties of normal $CD5^+$ B cells (reviewed in: Caligaris-Cappio, 1996). These properties include the frequent expression of low affinity polyreactive natural autoantibodies (Broker *et al.*, 1988; Weston & Raison, 1991;

Martin *et al.*, 1992), a lack of somatic mutation within V region genes, a frequent use of restricted Ig V gene repertoires (reviewed in: Kipps, 1993) and the ability to form rosettes with mouse erythrocytes (Caligaris-Cappio *et al.*, 1993). Of particular interest to this study is the production of polyreactive IgM by B CLL cells. Weston and Raison (1991) demonstrated that mouse IgG was bound by the surface IgM of B CLL malignant cells in a low affinity fashion, where the binding was abrogated after three washes in a conventional flow cytometry staining protocol. Chen *et al.* (1995) have independently shown, in normal CD5 positive B cells, that the binding of FITC labelled antigens (self and foreign Ag, e.g. insulin, IgG Fc, β-galactosidase) is indicative of the presence of membrane-bound polyreactive IgM on these B cells. Thus, the B CLL malignant cells provide a convenient source of monoclonal polyreactive IgM.

1.6: The experimental system

Using the B CLL system as a model for antibody polyreactivity this investigation addresses the structural basis of the polyreactive antigen binding phenomenon displayed by many naturally occurring IgM antibodies. The studies detailed in this thesis were targeted to three specific aims:

(1) to sequence the VL domains of five polyreactive IgM from B CLL patients for which the sequence of the VH domains was already determined and to analyse the primary structure of these V domains;

(2) to examine the three-dimensional structure of the binding sites of the polyreactive IgM molecules of B CLL origin through computer-aided (homology) modelling of the five Fv (VL-VH) molecules, and

(3) to clone and express soluble Fv in a form which would enable X-ray crystallographic studies of native and complexed polyreactive Ig from two of the B CLL derived IgM antibodies.

CHAPTER TWO
Variable regions of immunoglobulins expressed in human chronic B lymphocytic leukaemia

2.1: Introduction

Antibody variable domains display an inherently diverse amino acid sequence which is a product of diversity in the DNA sequences encoding these regions of the molecule. The diverse sequences are generated during development of B cells and are the result of three main components: (1) the recombination of a large number of V genes with smaller gene segments which in combination encode a functional V region; (2) the junctional diversity introduced by inconsistent splicing of these Ig gene segments and the addition of extra nucleotides at the recombination site, and (3) the process of somatic hypermutation following Ag stimulation. In the monoreactive Ab repertoire these factors result in a highly diverse pattern of V domain primary structure which is especially noticeable within the CDR sequences. However, the potential diversity of primary V region structures may be reduced for polyreactive Ab such as those expressed by human B CLL cells.

2.1.1: Genes encoding human variable domains

The human germline Ig gene loci contain a large number of coding sequences which are arranged in clusters of V genes as well as the smaller diversity (D) minigenes (D genes are only found in the heavy chain locus) and joining (J) genes. Constant domain genes are also found within the Ig loci. The V genes for both light and heavy chains encode framework regions 1-3 and the hypervariable loops (CDR) 1 and 2. The recombination or splicing of an individual V gene with a single J gene for a light chain and D and J genes for a heavy chain produces the coding sequence for the CDR3 and the fourth framework region (reviewed in: Tonegawa, 1983). For a single B cell clone there is a single functional rearranged light chain, of either κ or λ isotype, consisting of VL-JL-CL and a single functional heavy chain V domain consisting of VH-D-JH joined to one of several C domain genes which encode one of the heavy chain isotypes (δ, μ, γ, α or ε). Although several rearrangements may have occurred to form the

mature Ab only those DNA sequences which contain an open reading frame through VL-JL-CL and VH-D-JH and are structurally compatible (ie. associate as a stable VL-VH heterodimer) are considered functional rearrangements and so expressed by the B cell as protein (Cook & Tomlinson, 1995).

Diversity of V domain sequences is in part due to the large numbers of V genes found on the light and heavy chain loci which have now been comprehensively mapped to chromosomes and DNA sequenced for humans (Cox *et al.*, 1994; Cook & Tomlinson, 1995; Williams *et al.*, 1996). The V genes have been grouped into families which are assigned on the basis of members of a V gene family all having at least 80% nucleotide identity to other genes in that family. The human germline Vλ repertoire consists of 30 functional genes, although this number can vary slightly according to the individuals haplotype, which can be grouped into ten gene families (Vλ1-Vλ10). The major Vλ families contain five (Vλ1) eight (Vλ2) and eight (Vλ3) genes respectively with the remaining families containing between one and three genes (Williams *et al.*, 1996). Although there are approximately 76 Vκ germline genes only 32 genes are potentially functional; of the remainder, 16 genes contain minor defects and 25 genes multiple defects and so classed as pseudogenes. Around 26 genes have been found in rearranged antibodies, indicating that the functional Vκ germline locus consists of at least 26 gene segments (Zachau, 1993). It appears, from the six Vκ gene families, the Vκ1 and Vκ3 genes are the most frequently rearranged while the Vκ2 and Vκ4 are less frequently and Vκ5 and Vκ6 genes only rarely rearranged. This pattern of Vκ gene usage is a reflection of the number of genes within each Vκ family or subgrouping (Cox *et al.*, 1994). The human VH locus is the most complex of the three Ig gene loci. It contains around 95 VH gene segments with seven VH gene families of which VH3, VH1, and VH4 are the largest in decreasing order. In contrast to the VH loci of mice, where the VH families are grouped in distinct clusters (Brodeur *et al.*, 1988), the human VH genes are arranged so that members of gene families are scattered throughout the VH locus. Of the 95 VH genes only 51 have been found in functionally rearranged Ab sequences. The remaining genes are non-functional due to either the lack of the required recombination recognition sequences (ie. pseudogenes) or perhaps the sequences are not permissive to be functionally folded into a stable Ig fold (Cook & Tomlinson, 1995).

The CDR3 and fourth framework region (Fr4) of light chain V domains are produced by the recombination of a VL gene with one of 6 Jλ or 5 Jκ gene segments (Williams *et al.*, 1996; Zachau, 1993). In the case of heavy chain V domains, the same region is the product of the recombination between a VH gene with one, or occasionally

multiple, of the 25 functional diversity (D) minigenes (Corbett *et al.*, 1997) and one of 6 JH genes (Tomlinson *et al.*, 1996). The rearrangement of a large number of V genes with the D and J genes results in a germline or 'naive' antibody repertoire containing extensive primary structural diversity within the V domains of both light and heavy chains.

2.1.2: Junctional diversity within variable regions

During the recombination of V genes with D and J gene segments the diversity of the primary structure of V domains is further increased by the process of junctional diversity. Junctional diversity is achieved by the variability of the splicing of V to D and J genes which results in the deletion or occasional insertion of nucleotides or codons at these junctions (Tonegawa, 1983; Victor & Capra, 1994). Furthermore, diversity of CDR3 sequences is increased by the addition of non-template nucleotides (N-additions) at the splice sites. Heavy chain CDR3 sequences frequently contain N-additions. While most light chain VL–JL junctions display an absence of N-additions, Victor and Capra (1994) have reported that at least 20 percent of normal human foetal and adult peripheral blood B lymphocytes contain light chain rearrangements with non-germline encoded nucleotides, consistent with N-additions, at the VL–JL junction. The insertion of N-additions is catalysed by the enzyme terminal deoxy-transferase (TdT), which is expressed at low levels during light chain rearrangement and at high levels during heavy chain rearrangement (Tonegawa, 1983). As a result the CDR3s of heavy chains are more diverse at the primary sequence level when compared to light chain CDR3 sequences.

2.1.3: Somatic hypermutation

The somatically generated immune repertoire combines the diversity of the 'naive' repertoire with the process of hypermutation which occurs during affinity maturation. Somatic hypermutation results in the mutation of DNA sequences within the rearranged V regions. While mutations occur within both framework regions and CDR the changes tend to be targeted to 'hot spots' and show a propensity for the CDRs (reviewed in: Neuberger & Milstein, 1995; Betz *et al.*, 1993; Tonegawa, 1983). Mutations within V domains are selected, with a preference for affinity enhancing

mutations, in a T cell dependent manner during exposure to antigen (reviewed in: Steele *et al.*, 1997). Thus, an affinity selected antibody is 'tailored' to its Ag, by somatic hypermutation, through a series of amino acid changes from germline V sequences which have enhanced the affinity of the Ab–Ag interaction.

2.1.4: Characteristics of variable region genes utilised by polyreactive antibodies

In comparison with monoreactive Igs, the primary structural repertoire of polyreactive Igs appears to be reduced. Polyreactive Igs frequently utilise a restricted number of V and to some extent D and J genes (Sanz *et al.*, 1989; Schutte *et al.*, 1991; Brown *et al.*, 1992). Furthermore, somatic hypermutation does not occur in the majority of these antibody V region genes (Baccala *et al.*, 1989; van der Heijden *et al.*, 1991). However, there are exceptions to these dogma and so for at least some populations of polyreactive Igs an apparent unbiased usage of V genes and the presence of somatic mutation has been observed (Ikematsu *et al.*, 1993; Hartman *et al.*, 1989). This not withstanding, the increased frequency of polyreactive Igs encoded by restricted V region genes, which show little or no evidence of somatic hypermutation, suggests that the clonal diversity of B cells encoding polyreactive Ig is reduced when compared to B cells expressing monoreactive antibody.

In humans, the CD5 positive subset of B cells is responsible for producing the majority of polyreactive Igs. The malignant cells of B chronic lymphocytic leukaemia frequently express the CD5 antigen as well as membrane-bound polyreactive IgM. (Broker *et al.*, 1988). Since the malignant B cells are clonally derived, the IgM expressed by B CLL cells is monoclonal in nature, facilitating the sequencing of individual V genes of polyreactive Igs. In fact most of the V gene sequences for human polyreactive Igs were obtained using the B CLL system (Pritsch *et al.*, 1993). The restricted primary repertoire of Ig genes used by B CLL cells was first implied through the frequent presence of cross-reactive idiotypes (CRI) displayed by the surface IgM on these cells (reviewed in: Kipps & Carson, 1993). In particular the CRI 17.109 and G6 are frequently found on IgM expressed by B CLL cells. The 17.109 CRI has been revealed to be associated with the expression of a single Vκ gene (*Humkv325*) and the G6 idiotype has been linked to the expression of a heavy chain V region gene (designated 51P1) of the VH1 family (Kipps & Carson, 1993). However, the IgM expressed by B CLL cells has been shown to be of either κ or λ in isotype

with a more diverse pattern of light chain V genes being utilised than was first suggested (Pritsch *et al.*, 1993). The nonstochastic usage of *Humkv325*, 51P1 and the VH5 family gene VH251 which is also frequently rearranged in B CLL, is suggested to be linked to the frequent autoreactivity of the IgM present with this malignancy (reviewed in: Kipps & Carson, 1993). This may also be true regarding the frequent utilisation of the small VH4, 5 and 6 family genes by human polyreactive Ig expressed in both B CLL and certain autoimmune disorders such as rheumatoid arthritis (Brown *et al.*, 1992; Kipps, 1989; Kipps, 1993). While the picture of gene usage by polyreactive Ig producing B cells is not as clear as originally thought, the over representation of certain V genes may predispose this population towards the production of autoreactive/polyreactive antibodies. Although it is clear that the majority of V genes in humans may still be involved in both B CLL and normal polyreactive producing B cells.

The majority of V gene sequences obtained from B cells expressing polyreactive Ig show no evidence of somatic hypermutation suggesting that this process does not generally contribute to the polyreactive repertoire. However, in polyreactive IgG molecules low levels of somatic mutation have been observed, indicating that polyreactivity is not necessarily abrogated by somatic hypermutation (Ikematsu *et al.*, 1993). In contrast, Houdayer and colleagues (1993) suggest that polyreactive Ig undergo Ag driven somatic hypermutation at the expense of polyreactivity. This hypothesis was established from the analysis of four clonaly related serum antibodies, some of which were polyreactive and some monoreactive in specificity (Houdayer *et al.*, 1993). Although the theory of clonal evolution from polyreactivity to monoreactivity for Ab expressed by B cells has not been proven. This feature is fitting with the proposed role of polyreactive antibodies in a 'primitive' immune system, where the germline is capable of producing Ab with binding sites capable of binding to multiple Ag. Indeed the majority of gene sequences of V regions of polyreactive Ig are almost identical to the germline gene counterparts, although in some polyreactive Ab somatic hypermutation does occur (Ikematsu *et al.*, 1993; Baccala *et al.*, 1989; Sanz *et al.*, 1989). The absence of somatic hypermutation in the majority of these Ab suggests that the primary structural diversity of polyreactive Ig is reduced when compared to those of monoreactive Ig.

The light chain V region gene sequences of five IgM molecules expressed on the surface of B CLL cells are presented here and compared with the heavy chain V region gene sequences expressed by the same antibodies (Brock, 1995). Comparison of the VL and VH region gene sequences with their germline Ig gene counterparts showed either a complete absence or only a few mutations from the germline sequences indicating that somatic hypermutation was probably not a dominant process in the generation of these

antibodies. The $CD5^+$ B cells bind mouse IgG1 in a low affinity fashion, a phenomenon which has been shown within our laboratory to be due to membrane bound IgM of the B CLL cells (Weston & Raison, 1991). The sequencing of the VL and VH of the IgM is an essential step in obtaining the three-dimensional structures of the Fv molecules through homology modelling (Chapter Three) and the cloning and expression of two of the Fv for soluble protein expression (Chapter Four) which would lead into the experimental determination of their 3D structures.

2.2: Materials and methods

2.2.1: General reagents

Unless otherwise stated all reagents were reagent or molecular biology grade. General chemicals were purchased from BDH (Poole, Dorset, England) Sigma (St. Louis, MO, USA) or USB (Cleveland, USA). Restriction enzymes were purchased from New England Biolabs (Beverly, MA, USA). PolyATract™ mRNA isolation system and first-strand Reverse Transcription system were purchased from Promega (Madison, WI, USA). Sequenase™ Version 2.0 DNA sequencing kit (USB, Cleveland, USA) was used for all manual sequencing. Sequencing vectors pBluescript (SK^+) and pCR-Script™ (SK^+) were purchased from Stratagene (La Jolla, CA, USA). Stabilised α-33P Deoxyadenosine 5'-Triphosphate (33P-ATP) was purchased from (Bresatec). Saturated tris-buffered phenol (pH 8.0) and sequencing acrylamide (40%) were obtained from Amresco (Solon, OH, USA). Thermostable DNA polymerases, Taq (Perkin-Elmer, Branchburg, NJ, USA) and cloned PFU (Promega, Madison, WI, USA) were used for amplification of genes by the polymerase chain reaction. Fluorochrome conjugated antibodies were purchased from Kallestad (Austin, TX, USA) Silenus Laboratories (Hawthorn, VIC, Australia) or Becton Dickinson (San Jose, CA, USA). All water was MilliQ™ purified using a Modulab™ system and autoclaved.

2.2.2: The patients

Peripheral blood leucocytes (PBL) were collected during routine leucopheresis of five patients with chronic B lymphocytic leukaemia (B CLL). Cells were frozen and stored in liquid nitrogen at a density of $3x10^7$ cells/mL in the presence of RPMI (Trace, Sydney, NSW, Australia), 45% foetal bovine serum (FBS; CSL, Melbourne, Vic, Australia) and 10% (v/v) DMSO (BDH). Patients cells are referred to as Bel, Tre, Yar, Hod and Jak.

2.2.3: Isolation of total cellular RNA

Total cellular RNA was extracted from $1x10^7$ to $2x10^8$ PBL using a modification of the method of Chomcyznski and Sacchi (1987). Briefly, PBL were thawed from

liquid nitrogen into 3 mL of foetal bovine serum. The cells were washed twice with sterile phosphate buffered saline (PBS: 1.5 mM monobasic potassium phosphate, 8 mM dibasic sodium phosphate, 0.15 M NaCl, 2.6 mM KCl, pH 7.4) and resuspended in 2 or 3 mL of solution D (4 M guanidinium thiocyanate, 25 mM pH 7 sodium citrate, 0.5% sarcosyl, 0.1 M β-2-mercaptoethanol). The cells were then lysed by expelling the suspension repeatedly through a fine needle (18g) using a sterile 5 mL syringe. Sodium acetate (0.2 or 0.35 mL of 2 M, pH 4) water saturated phenol (2 or 3 mL) and chloroform:isoamyl alcohol (0.4 or 0.7 mL of a 24:1 mixture) was sequentially added to the lysed cells and mixed by vortexing for 10 s. After 15 minutes on ice the aqueous phase was collected by centrifugation at 11 000 g for 30 min at 4°C (RC-5 centrifuge, Sorvall). RNA was precipitated with 0.6 volumes of isopropanol at -70°C for 30 min and pelleted by centrifugation (11 000 g for 30 min, Sorvall). The RNA was dissolved in 0.5 mL of solution D and transferred to a 1.5 mL Eppendorf tube and precipitated with an equal volume of isopropanol at -70°C for 30 min. The RNA was pelleted by centrifugation at 12 000 g for 15 min at 4°C (5415C Microfuge, Eppendorf) and washed once with 0.5 mL 75% (v/v) ethanol (-20°C). The RNA was pelleted by centrifugation for 5 min, the ethanol removed and the pellet dried (GeneVac) and stored at -70°C.

2.2.4: Purification of messenger RNA and first strand cDNA synthesis

Messenger RNA was purified from 0.1 to 1 mg of total RNA by an oligo dT magnetic capture procedure using the PolyATract™ System II (Promega) according to the manufacturers instructions. The mRNA in 0.25 mL nuclease free water was lyophilised (GeneVac) and stored at -70°C. First strand cDNA was synthesised by incubating 0.1 to 1 μg mRNA (30 min at 42°C) with 0.05 to 0.5 μg oligo$(dT)_{15}$ primer and 1.5 to 15 units of AMV reverse transcriptase (USB or Promega). A typical 20 μL reaction also contained 1xRT buffer, 5 mM $MgCl_2$, 1 mM each dNTP, 20 units rRNasin (ribonuclease inhibitor). The cDNA was extracted once with equal volumes of phenol (pH 8) and CIAA (24:1) and once with CIAA and precipitated with 2.5 volumes of absolute ethanol at -20°C and washed once with 70% ethanol. The cDNA was stored as a dry pellet at -20°C and resuspended in 20 μL of sterile MilliQ™ water immediately prior to use in PCR.

2.2.5: Oligonucleotide primers

Primers for PCR of antibody light chain V domains were designed to bind to the conserved family leader sequences and to the J-proximal end of the light chain constant domain consensus sequences. *Kappa* light chain primer sequences were KF1 5'-TGC(T) TGC TCT GGC(A) TCC(T) CA(T)G GTG CC-3' and KF2 5'-CTA CTC TGG CTC CCA GAT ACC-3' leader sequence primers and KC1 5'-GAA GAC AGA TGG TGC AGC CAC-3' constant domain primer. *Lambda* light chain primers were LF1 5'-CTC(T) CTC C(A)T(C)C(T) CAC TGC ACA GG-3' leader sequence primer and LC1 5'-GG AGG GAA CAG AGT GAC CGA GG-3' constant domain primer.

2.2.6: Polymerase chain reaction (PCR)

Light chain V region genes were amplified from 1 to 10 ng of cDNA by a thermostable DNA polymerase (Taq or PFU) using a hot start PCR protocol of 1 cycle of 5 min at 90°C after which Taq or PFU was added and 1 cycle of 5 min at 60°C followed by 35 amplification cycles: 60 s at 90°C, 30 s at 65°C and 30 s at 72°C (Omingene thermocycler, Hybaid). Typical 100 µL reactions contained 1xPCR reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 10 pmol each primer and 2.5 units of Taq or PFU DNA polymerases. DMSO to a final concentration of 10% (v/v) was added for amplifications using PFU DNA polymerase.

2.2.7: Cloning of PCR products for DNA sequencing

Amplified products of 300 to 400 base pairs in size were purified using equal volumes of phenol (pH 8) and CIAA (24:1), precipitated with 2.5 volumes of absolute ethanol (-20°C) and washed once with 70% ethanol. DNA was resuspended in 10 mL of sterile MilliQ™ water and 100 to 200 ng was ligated into 100 ng of an appropriate sequencing vector (T-tailed pBluescript for Taq amplified DNA or PCRscript for PFU amplified DNA). Recombinant plasmids were selected on the basis of the disruption of the galactosidase gene (Sambrook *et al.*, 1989). Recombinant plasmids were digested with restriction enzymes on either side of the cloning site (1 unit/µg DNA for 1 h at

37°C) and electrophoresed on 10% polyacrylamide TBE gels to select clones for sequencing with appropriate size insert DNA.

2.2.8: Dideoxy DNA sequencing of cloned genes

Manual dideoxy DNA sequencing of cloned V region genes was carried out by the method of Sanger *et al.*, (1977) using the Sequenase version 2.0 kit (USB). Generally 2 µg of plasmid DNA (prepared by alkaline lysis, Sambrook *et al.*, 1989) and 4 pmol oligonucleotide primer (forward M13F, SK or reverse M13R, T7) was used for one sequencing reaction. Plasmid DNA was denatured at 85°C for 5 min in the presence of 0.2 M NaOH and 0.2 mM EDTA in a total of 20 µL of sterile MilliQ™ water. The DNA was chilled on ice and 8 µL 1 M Tris pH 4.5, 3 µL 3 M sodium acetate pH 7 and 75 µL of absolute ethanol was added sequentially. DNA was precipitated on dry ice for 20 min and pelleted by centrifugation at 12 000 g at 4°C for 15 min (Microfuge, Eppendorf). The pellet was washed once with -20°C 70% (v/v) ethanol and centrifuged at 12 000 g for 5 min after which the ethanol was decanted and the DNA pellet dried by vacuum desiccation. The DNA was resuspended in sterile MilliQ™ water containing 1x Sequenase reaction buffer and 10% (v/v) DMSO and incubated at 37°C for 15 min before the sequential addition of 1 µL 100 mM DTT, 2 µL diluted label mix (1:5 for long and 1:10 for short sequencing reactions), 0.5 to 1 µL of 33P-ATP (~10 µCi) and 3.25 units of Sequenase Version 2.0 DNA polymerase. The reaction was allowed to proceed for 5 min at 21°C before adding 3.5 µL to each termination mix (2.5 µL of a dideoxy G, A, T and C). The termination reactions were incubated at 37°C for 15 min and stopped by the addition of 4 µL of loading buffer. Immediately before electrophoresis the reactions were incubated for 3 min at 90°C and chilled on ice. Reactions were electrophoresed (2.5 µL per termination reaction) on a 6% polyacrylamide gel containing 7 M urea and 1x modified TBE (45 mM Tris-borate, 1 mM EDTA pH 8.0). Electrophoresis conditions were 1800 V, 45 mA, 60 W at a gel temperature of 45°C. Air-dried gels were exposed to X-ray film (Fuji) for 36 h before development.

2.2.9: Analysis of variable region gene expression

Primary DNA sequence obtained for VL and VH (Appendix A and Brock, 1995) domains were compared with all available immunoglobulin germline gene sequences. Sequences with high nucleotide homology with the VL, VH, JL, JH and D gene

segments were extracted using the Fasta algorithm from GENBANK (release 88.0), or KABAT immunological (release 5.0) databases which were accessed through the Australian Genomic Information Service (ANGIS, Sydney Australia). Gene assignments were verified and/or updated using the V BASE PACKAGE (MRC Centre for Protein Engineering, Cambridge, UK). Translated amino acid sequences for V domains were obtained with DNA strider version 1.2 with framework and CDR residues assigned according to Kabat (1991).

2.2.10: Phenotyping of B CLL cells

The PBL obtained from B CLL patients were phenotyped for the surface expression of μ, κ or λ Ig chains. Cells were probed with a 1:100 dilution (prepared in staining wash: 0.1% BSA (w/v) in PBS containing 0.02% sodium azide) of sheep polyclonal antibodies to μ or γ (Kallestad) and κ, λ or mouse Ig (Silenus) conjugated with FITC for 20 min at 4°C. Washed three times with staining wash. Data was acquired for SSC, FSC and Fl-1 using a FACscan flow cytometer (Becton Dickinson, Sunny Vale, CA, USA). Data were analysed using the software package CellQuest™ v1.1.1 (Becton Dickinson).

2.2.11: Binding of mouse IgG by B CLL cells

The low affinity binding of murine IgG by the B CLL cells was assessed by the method of Weston and Raison (1991). Briefly, $5x10^5$ to $1x10^6$ PBL were incubated with 1, 10 or 100 μg of a murine IgG1/κ monoclonal, K.1.21 (Boux *et al.*, 1983) for 20 min at 4°C. Duplicate reactions were washed (0.5 mL of staining wash) either once or three times prior to incubation (20 min at 4°C) with a sheep α-mouse Ig (H & L) FITC conjugated polyclonal antibody. Samples were washed three times with staining wash. Background corrected ratios of Mean fluorescent intensity (MFI) were calculated (CellQuest™, Becton Dickinson) by normalising the MFI of samples containing K.1.21 to the MFI resulting from incubation of the sheep α-mouse Ig FITC antibody (Silenus) in the absence of mouse Ig.

2.3: Results

2.3.1: Amplification, cloning and sequencing of B CLL variable region genes

The VL and VH region genes have been sequenced from the monoclonal IgM expressed by the malignant cell clones of five B CLL patients: Bel, Tre, Yar, Hod and Jak. The VH region genes were previously sequenced (Brock, 1995) and the DNA sequences are reproduced in this thesis (Appendix A) because they are utilised for homology modelling and cloning for protein expression of the B CLL Fv molecules. The sequences of the VL region genes were determined for each IgM and are described here (Figures 2.1-5).

The VL region genes were PCR amplified as a single band of approximately 300 to 350 base pairs with either λ (Bel) or κ (Tre, Yar, Hod and Jak) light chain specific primers designed to bind to the leader sequence and the J-proximal region of the constant domain for forward and reverse primers respectively. The κ and λ primers were used in separate PCR amplifications on each sample and product was only obtained with one of the primer sets corresponding to the light chain phenotypes of the B CLL cells (2.3.4). This suggests that the rearranged product was not derived from residual normal B cells in the PBLs obtained from the B CLL patients. The PCR products were cloned into sequencing vectors and sequenced completely in both directions. A total of 3 independent clones were sequenced for each antibody and at least one product was amplified with the high fidelity DNA polymerase PFU. For each B CLL VL region, sequenced, all clones contained an identical nucleotide sequence with any initial ambiguous nucleotide assignments confirmed by resequencing the clone at least twice. The consensus DNA sequences for the five B CLL VL region genes are shown in comparison to the closest germline VL and JL genes (Figures 2.1-5). None of the recombinant plasmids sequenced from cloned PCR product obtained from a given patient's PBL yielded an alternative V region gene sequence, indicating that in all cases the VL regions cloned were derived from the expressed mRNA of the malignant B CLL cells and not from any residual normal B cells.

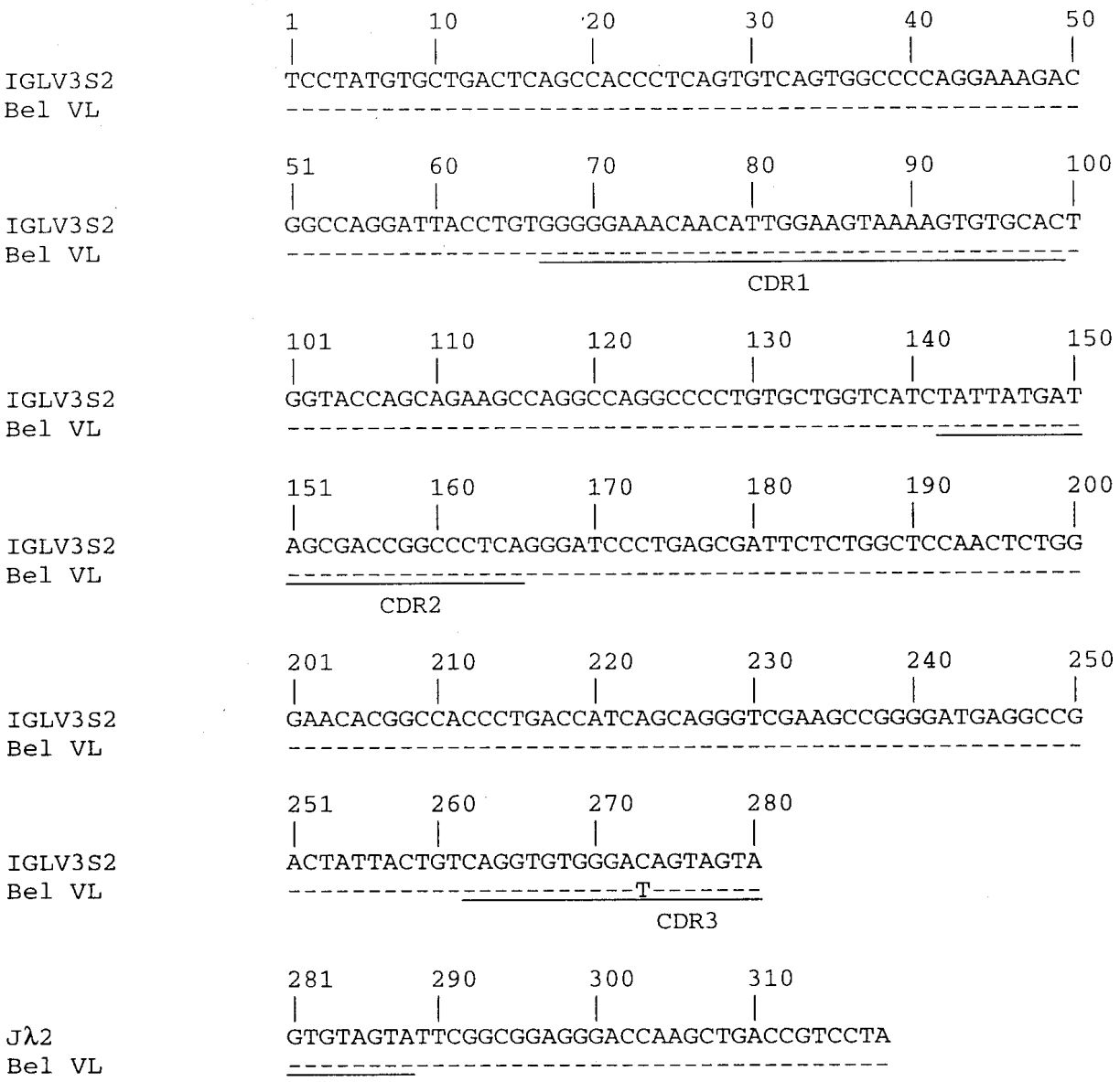


Figure 2.1: Consensus nucleotide sequence of Bel VL region. The closest germline gene elements are aligned with the Bel DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Germline gene sequences were extracted from V BASE.

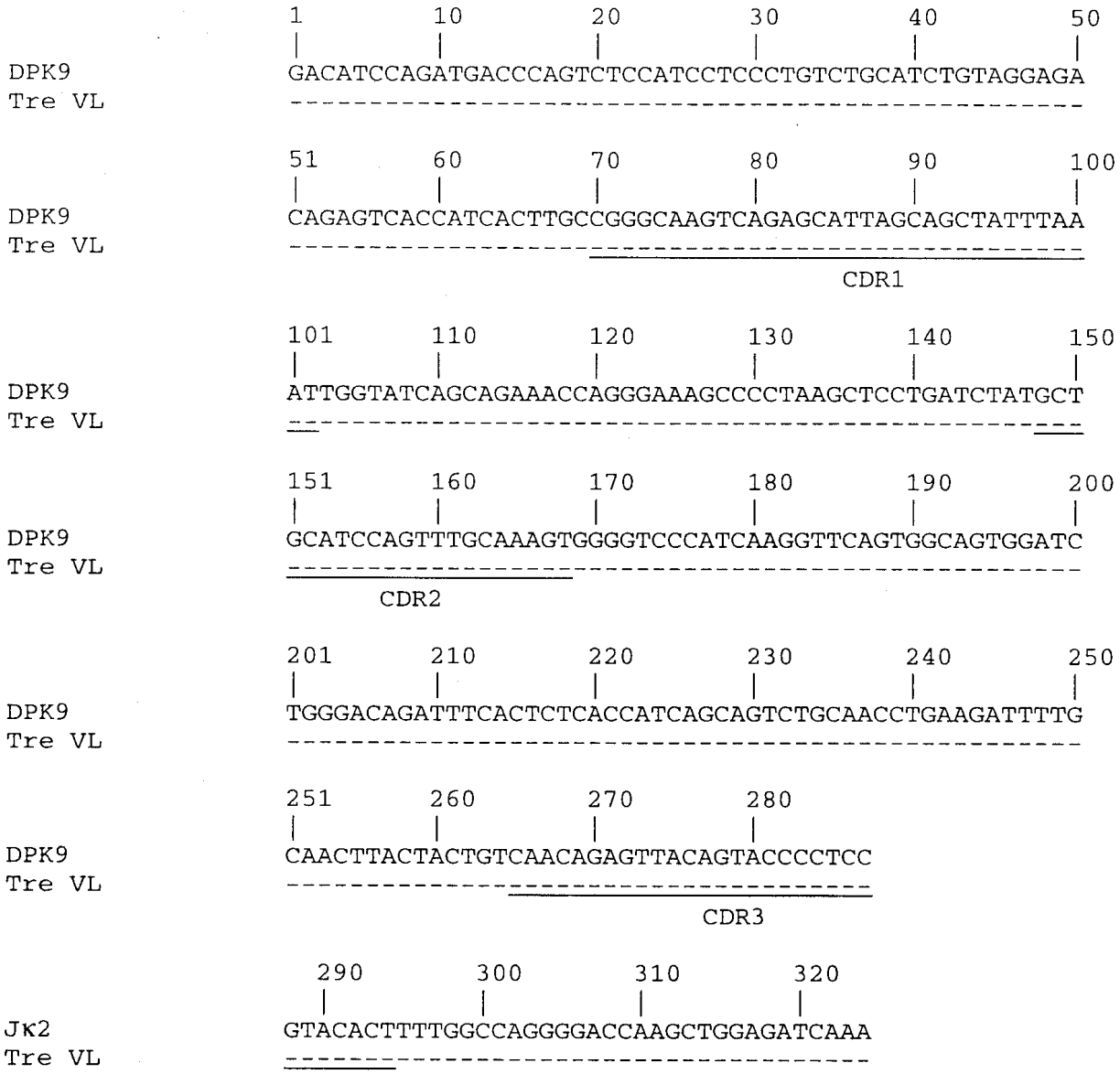


Figure 2.2: Consensus nucleotide sequence of Tre VL region. The closest germline gene elements are aligned with the Tre DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Germline gene sequences were extracted from V BASE. DPK9 is identical to genes designated O12, O2 and HSIGKLCO2.

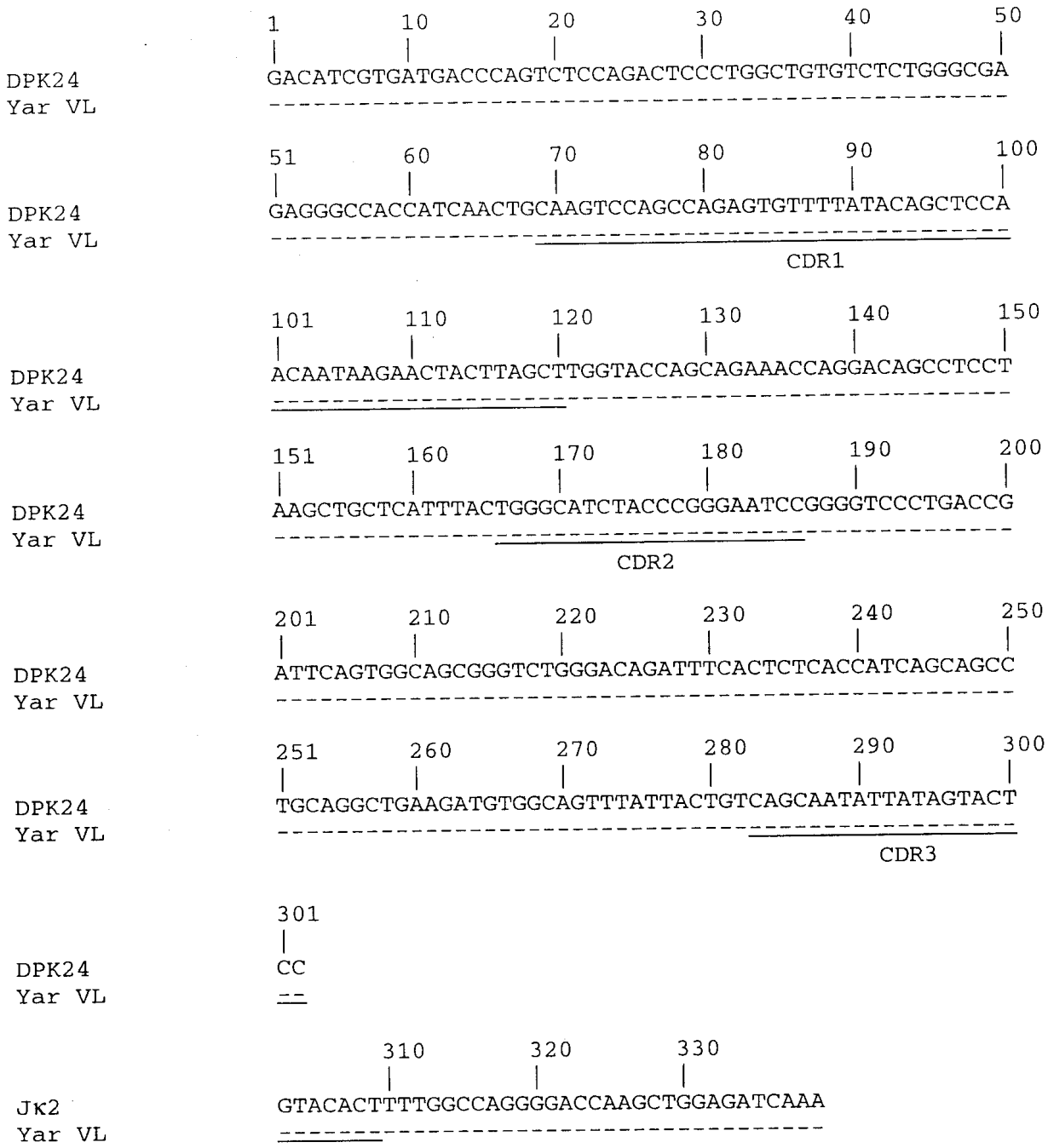


Figure 2.3: Consensus nucleotide sequence of Yar VL region. The closest germline gene elements are aligned with the Yar DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Germline gene sequences were extracted from V BASE. DPK24 is identical to genes designated VKIVKlob and HSIGK18.

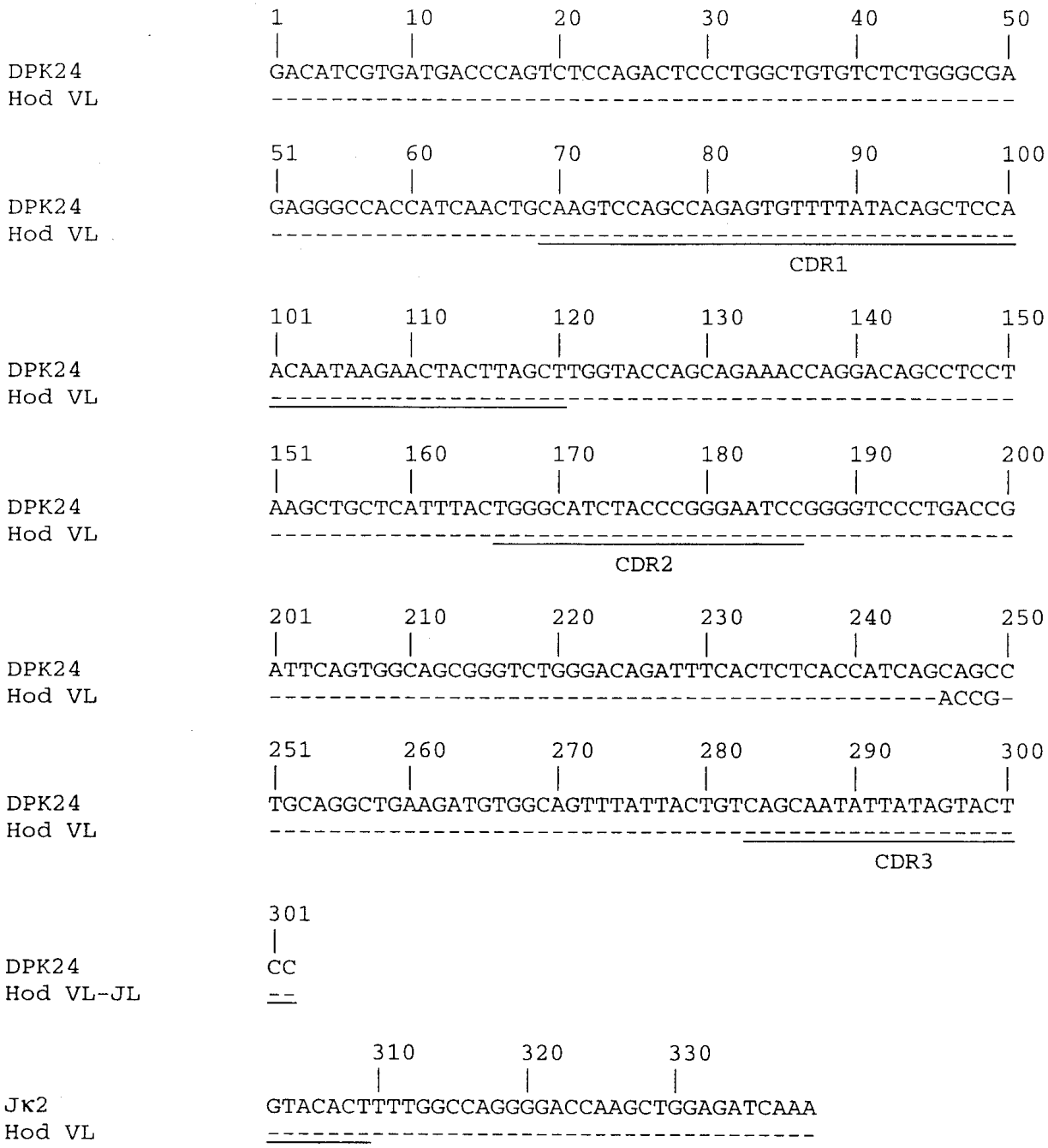


Figure 2.4: Consensus nucleotide sequence of Hod VL region. The closest germline gene elements are aligned with the Hod DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Germline gene sequences were extracted from V BASE. DPK24 is identical to genes designated VKIVKlob and HSIGK18.

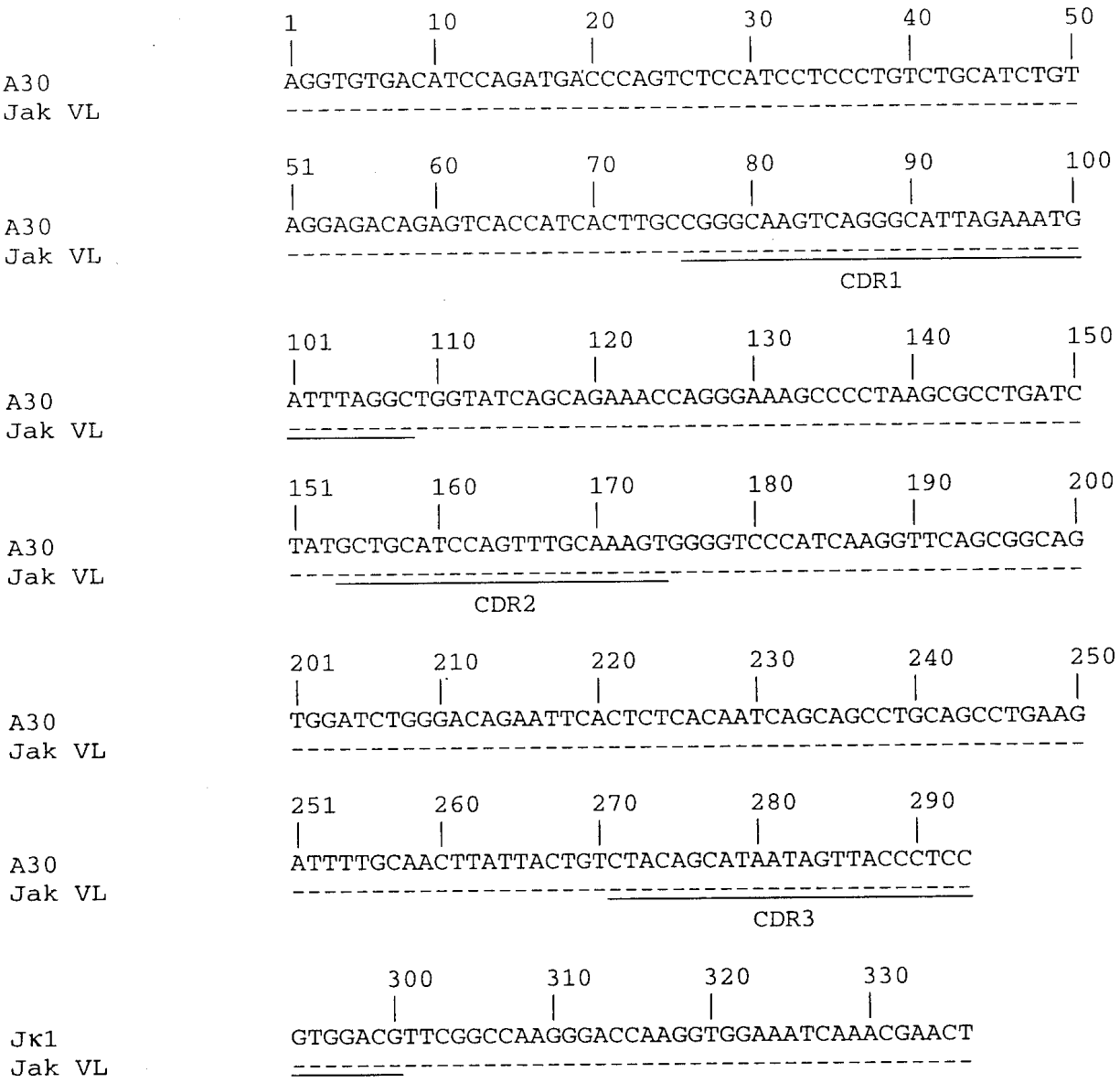


Figure 2.5: Consensus nucleotide sequence of Jak VL region. The closest germline gene elements are aligned with the Jak DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Germline gene sequences were extracted from V BASE. A30 is identical to genes designated SG3 and HSIGKLA30.

2.3.2: Analysis of variable region gene expression

All five VL regions were almost identical to the germline sequences of the gene elements comprising the recombined sequence (Figures 2.1-2.5 and Table 2.1). In fact the VL–JL sequences of Tre, Yar and Jak derived VL regions were identical to the germline sequences. The germline genes utilised in the five VL regions are summarised in Table 2.1. The mutations from germline sequences and any changes to the protein sequence are summarised in Table 2.2. Of the five antibodies two (Yar and Hod; Figures 2.3 and 2.4) utilise the same germline genes, although the two sequences contain a total of four nucleotide differences (Table 2.2) due to somatic mutations in the VL gene expressed by Hod. Table 2.3 is a summary of the VH region genes expressed by the B CLL cells (Brock, 1995) and is presented here for comparison to the VL region genes.

2.3.3: Primary structure of B CLL variable domains

The translated amino acid sequence of the Fv molecules of the B CLL antibodies is shown (Figures 2.6 and 2.7). Complementarity determining regions (CDR) and framework regions (Fr) were assigned and the residues are numbered throughout using the terminology of Kabat *et al.* (1991). Figure 2.8 is an alignment of CDR sequences showing residues shaded according to the basic chemical nature of their side-chains. The sequences contain a high number of uncharged polar side-chains which are distributed throughout the CDR sequences. The CDRs contain relatively few charged groups which are scattered across the aligned sequences with two obvious exceptions: all the κ light chains have either an arginine or lysine residue at position 24L and 101H invariably has an aspartic acid residue. Amino acids with non-polar side-chains appear to be restricted to certain positions within individual CDR sequences. Positions 33L, 34H, 51H and 63H have non-polar amino acids in all five sequences and positions 50L, 51L, 93L, 33H and 100H have non-polar amino acids in four of the aligned sequences.

Table 2.1
Analysis of light chain variable region gene expression in B CLL patients

Patient	VL family[a]	VL germline gene accession codes[b]	Identity to VL (%)	JL Gene[a]	Identity to JL (%)
Bel	Vλ3	IGLV3S2	99.6	Jλ2	100
Tre	Vκ1	DPK9	100	Jκ2	100
Yar	Vκ4	DPK24	100	Jκ2	100
Hod	Vκ4	DPK24	98.7	Jκ2	100
Jak	Vκ1	A30	100	Jκ1	100

[a] The family or subgroup of the germline gene element used in the rearranged antibody light chain V region is indicated.

[b] Alternative designations for VL germline genes are included in Figures 2.2 to 2.5.

Table 2.2
Mutations from germline Ig light chain gene sequences

Patient	Codon position mutated	Mutations DNA (amino acid)
Bel	91	GAC (D) → GAT (D)
Tre	None	None
Yar	None	None
Hod	82	AGC (S) → AGA (R)
	83	AGC (S) → CCG (P)
Jak	None	None

Table 2.3

Analysis of heavy chain variable region gene expression in B CLL patients

Patient	VH gene family	VH germline gene[a]	Identity to VH gene (%)[b]	D minigene[a]	JH gene[a]
Bel	VH5	DP-73	99.6	DXP'1	JH6c
Tre	VH5	VH32	98.5	D6-19	JH4b
Yar	VH3	p1	97.8	DXP4	JH5b
Hod	VH1	DP-10	100	DXP1	JH6a
Jak	VH3	DP-49	98.5	DXP4	JH4b

[a] The gene assignments are different from those originally assigned in the study of Brock (1995). Alternative designations for the VH genes are given in Figures A.1 to A.5.

[b] Percent identities to germline VH genes were calculated for the sequence up until the end of framework region 3 and excluding the first 23 nucleotides which were encoded by the N-terminal primers.

A

```
                1         10         20         30         40
                |         |          |          |          |
Bel VL     SYVLTQPPS VSVAPGKTARITCGGNNIGSKSVHWYQQKPGQAPVLVIY
                          Fr1                CDR1          Fr2

                50         60         70         80         90
                |          |          |          |          |
Bel VL     YDSDRPSGIPERFSGSNSGNTATLTISRVEAGDEADYYCQVWDSSSVV
            CDR2                Fr3                      CDR3

                  100
                   |
Bel VL     FGGGTKLTVL
             Fr4
```

B

```
                1         10         20       27a-f   30         40
                |         |          |          |      |          |
Tre VL     DIQMTQSPSSLSASVGDRVTITCRASQ              SISSYLNWYQQKPGK
Yar VL     --V-----D--AV-L-E-A--N-KS--SVLYSSNNKN--A-------Q
Hod VL     --V-----D--AV-L-E-A--N-KS--SVLYSSNNKN--A-------Q
Jak VL     -------------------------____       G-RND-G--------
                          Fr1                CDR1          Fr2

                       50         60         70         80         90
                       |          |          |          |          |
Tre VL     APKLLIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQ
Yar VL     P------W--TRE----D------------------A--VAV-----
Hod VL     P------W--TRE----D----------------RP--A--VAV-----
Jak VL     ---R---_________________________E-----------------L-
                    CDR2                     Fr3

                     95a 100
                      |   |
Tre VL     SYSTPPYTFGQGTKLEIK
Yar VL     Y----YT ----------
Hod VL     Y----YT ----------
Jak VL     HN-Y--W-------V---
            CDR3      Fr4
```

Figure 2.6: Translated amino acid sequences of five B CLL light chain V domains. **A:** Sequence of the λ isotype Bel VL domain. **B:** Sequences of κ isotype VL domains of Tre, Yar, Hod and Jak. Identities with the Tre sequence are indicated by dashes. Gaps in the alignments or numbering are indicated with spaces. CDRs are underlined and framework (Fr) sequences are indicated. Residues are numbered following Kabat *et al.* (1991).

```
                1         10         20         30         40
                |         |          |          |          |
Bel VH     EVQLQESGAEVKKPGESLKISCKGSGYSFTNYWIGWVRQMPGKGLEWM
Tre VH     ------------------R-----------S---S-------------
Yar VH     --------GGLVQ--G--RL--SA--FT-SS-AMH----A------YV
Hod VH     ---------------S-V-V---A--GT-SS-A-S----A--QG----
Jak VH     --------GG-VQ--R--RL--AA--FT-SS-GMH----A-------V
                           Fr1             CDR1       Fr2

                 50 52a      60         70         80 82a-c     90
                 |  |        |          |          |  |         |
Bel VH     GIIYPGDSDTRYSPSFQGQVTISADKSISTAYLQWSSLKASDTAMYYC
Tre VH     -R-D-S--Y-N-------V----S-----S------------------
Yar VH     SA-SSNGGT-Y-AD-VK-RF---R-N-KN-L---M---R-E---V---
Hod VH     -G-I-IFGTAN-AQK---R---T--E-T----MEL---RSE---V---
Jak VH     AV-WYDG-NKY-AD-VK-RF---R-N-KN-L---MN--R-D---V---
                CDR2                        Fr3

                       100,100a-k  101        110
                       |           |          |
Bel VH     ARRTGTGDPYYYYYYM    DVWGKGTTVTVSS
Tre VH     ---QWLALGHF         -Y--Q--L-----
Yar VH     VKTYYDFWSG-SPNWF    -P--Q--L-----
Hod VH     --DIYDILTG-T---YYGM----Q---------
Jak VH     --GE-GMYYDFWSGKYYF  -Y--Q--L-----
                      CDR3            Fr4
```

Figure 2.7: Translated amino acid sequences of five B CLL heavy chain V domains. DNA sequences were determined elsewhere (Brock, 1995). Identities with Bel VH are shown as dashes. Residues are numbered and CDR and Fr assigned following Kabat *et al.* (1991). The sequences have been reproduced here as they comprise a significant component of the primary sequence analysis and homology modelling of the Fv molecules

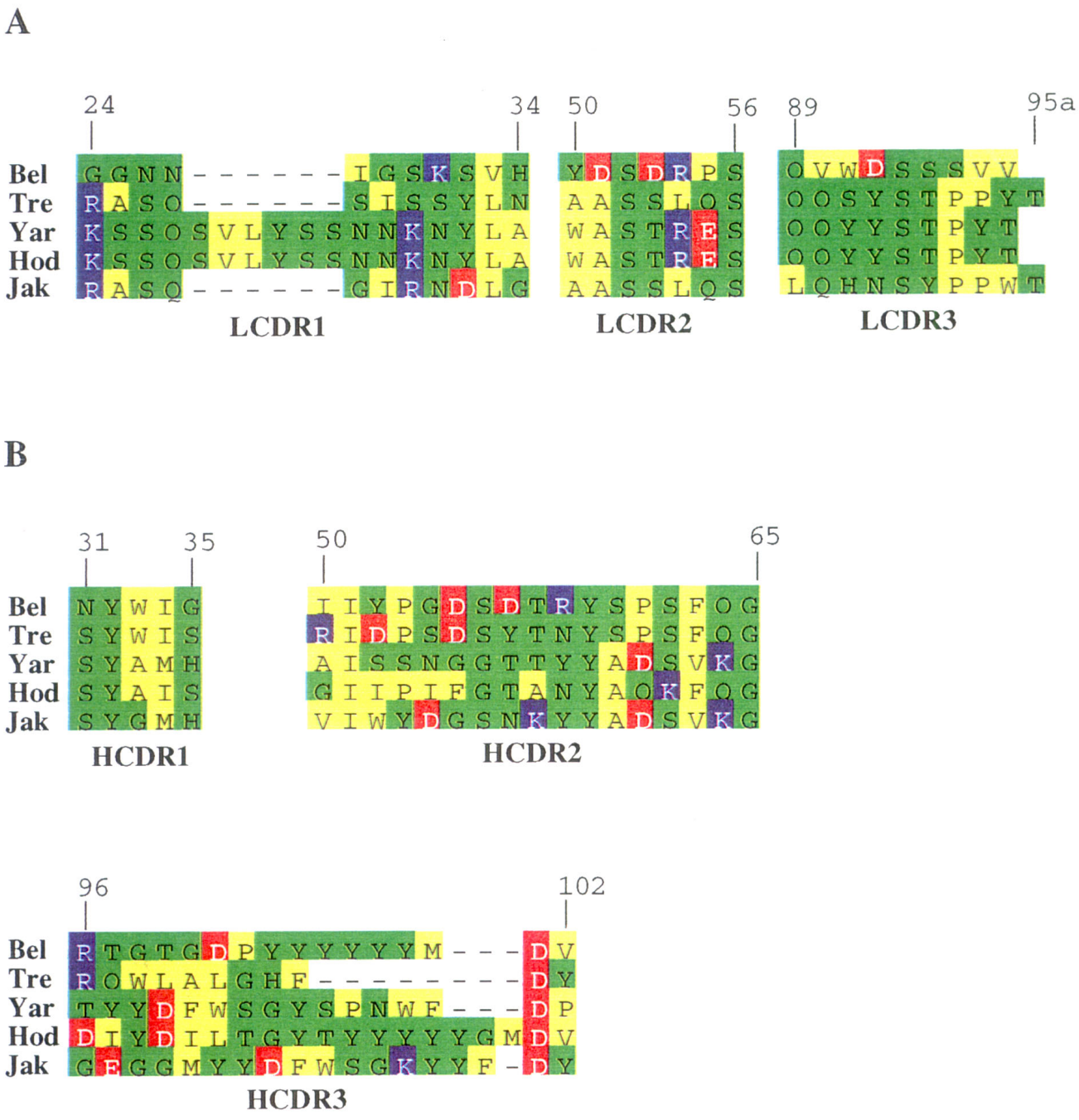


Figure 2.8: Nature of residues in B CLL CDR loops. **A:** residues of light chain CDRs. **B:** residues of heavy chain CDRs. Residues are grouped into non-polar (yellow); polar - uncharged (green), basic (blue) and acidic (red). Figure was produced by SeqVu v1.1 (Garvan Institute, Sydney, NSW, Australia).

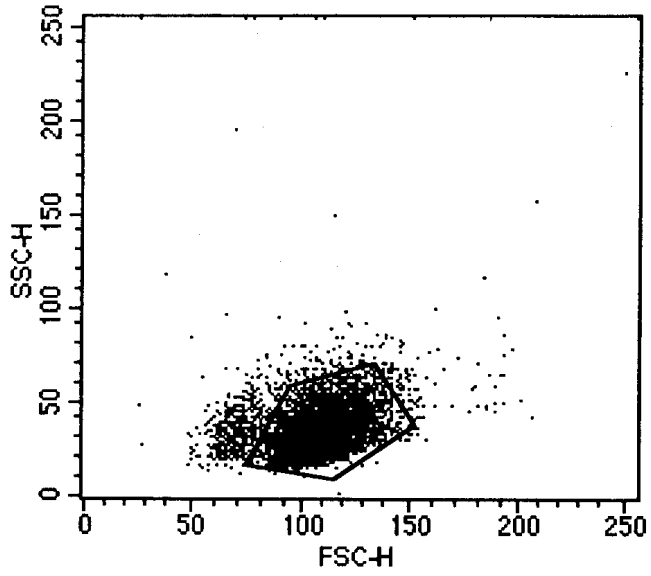
2.3.4: Phenotyping of B CLL lymphocytes

The PBLs obtained from the five B CLL patients were phenotyped for surface expression of light chains (κ or λ) and heavy chains (γ or μ) using FITC conjugated isotyping antibodies (Figure 2.9 and Appendix A). The isotypes of light chains correspond with VL region genes which were sequenced (Section 2.3.1). The majority of B lymphocytes (CD19 positive) were shown to express the CD5 antigen using two-colour staining (Figure 2.10 and Appendix A). Staining the B CLL cells with the anti-CD19-FITC reagent resulted in a small magnitude unimodal shift in FL1 flouresence compared with an isotype matched control. This low level expression of CD19 was consistent for all B CLL patient's PBLs examined here. However, the assignment of the majority of the PBL as B lymphocytes was made on the basis of the $CD19^{Lo}$ phenotype being corroborated with surface IgM expression by the B CLL cells. The results of the phenotyping are summarised (Table 2.4). All positive histograms were observed as a unimodal shift in fluorescence intensity, indicating that the majority of lymphocytes were positive for the assigned antigens.

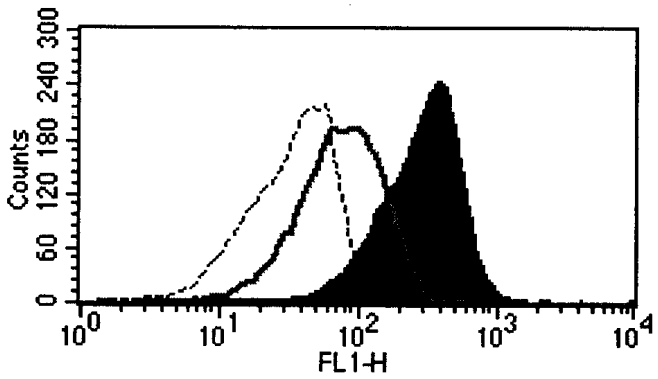
2.3.5: Low affinity binding to mouse IgG1 (K.1.21)

The surface IgM expressed by the B CLL cells was tested for binding to mouse IgG1 (K.1.21) using a flow cytometric assay (Weston & Raison, 1991). The phenomenon of binding to mouse IgG1/κ was demonstrated for the malignant B cells from all five patients (Figures 2.11 and 2.12; Table 2.5). The binding was titratable with a maximal occupation of the surface IgM binding sites shown when 100 μg/mL of K.1.21 was used. No binding was observed at the lowest concentration of K.1.21 tested (1 μg/mL) for patients Bel, Tre, Yar and Jak. However, binding was observed in all cases when K.1.21 was at 10 μg/mL, although the effect was marginal for Bel B CLL cells. Furthermore, the low affinity nature of the interaction was demonstrated by the reduction of binding to background levels after three washes prior to the addition of the sheep α-mouse Ig FITC conjugated antibody.

A



B



C

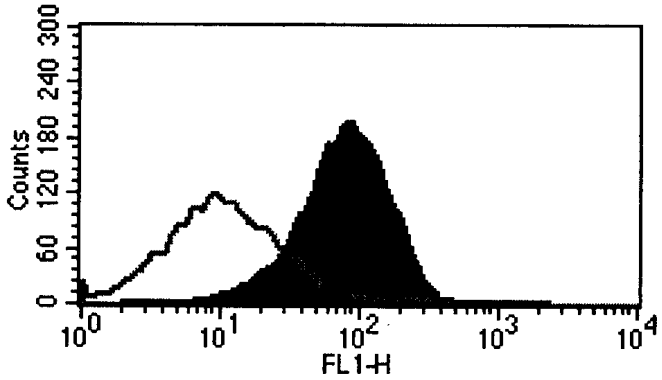


Figure 2.9: Isotyping of B CLL lymphocytes (representative data from patient Bel). **A:** Dot plot of 90° side scatter (SSC-H) *versus* forward scatter (FSC-H). The viable small lymphocytes are gated (solid line). **B:** L chain isotyping histograms: α-λ-FITC (solid histogram); α-κ-FITC (grey line) and sheep polyclonal-FITC (dashed line). **C:** H chain isotyping histograms: α-μ-FITC (solid histogram) and α-γ-FITC (grey lines).

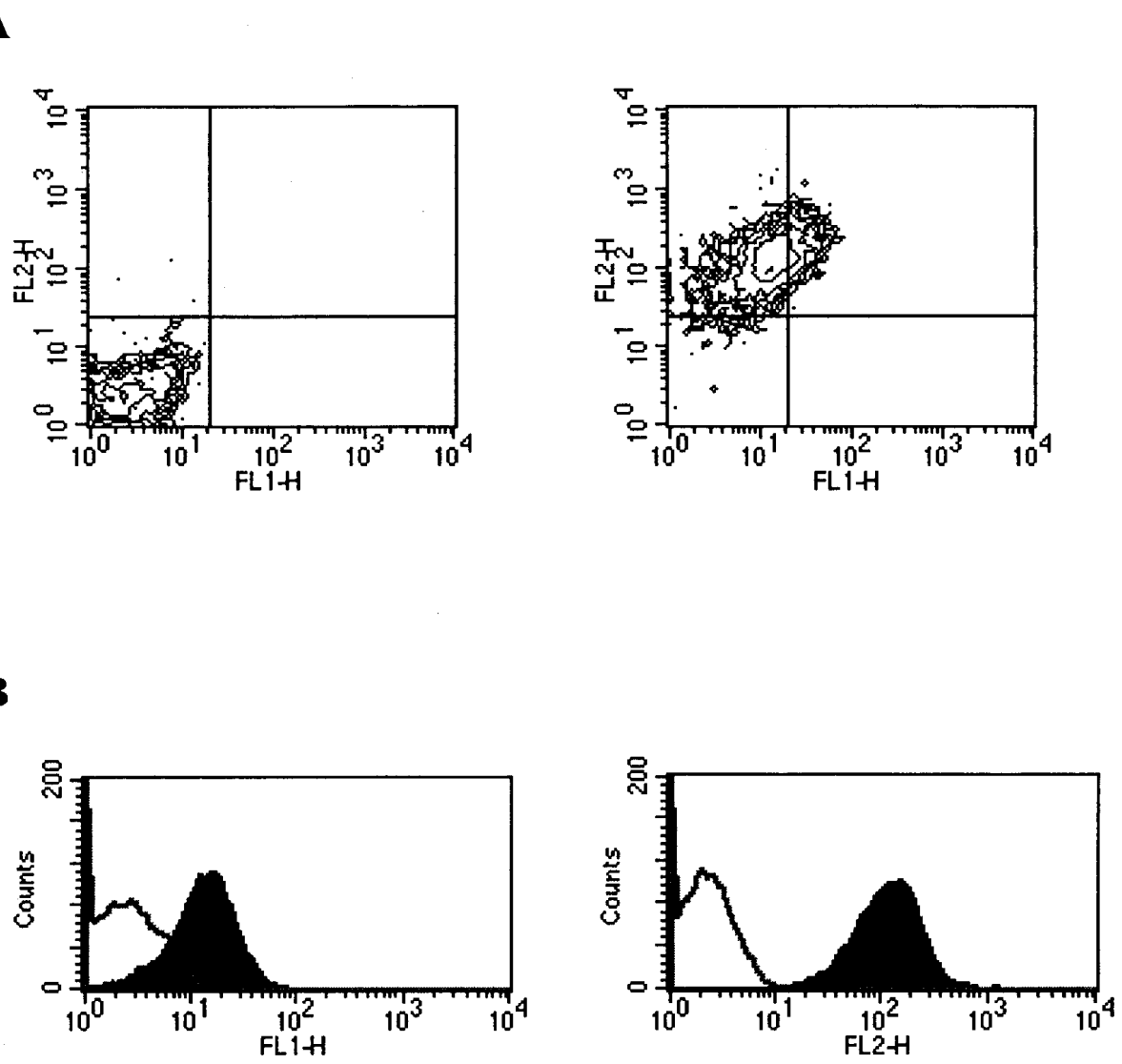


Figure 2.10: Two-colour immunofluorescence of B CLL cells (representative data from patient Jak). **A:** Contour plots of isotyped matched control antibodies (left) and α-CD19-FITC (FL1-H) *versus* α-CD5-PE (FL2-H) fluorescence (right). **B:** Fluorescence histograms of cells stained for CD19 (left) and CD5 (right). Solid histograms represent cells stained with specific antibodies; control histograms are shown as grey lines.

Table 2.4

Phenotypes of B CLL lymphocytes[a]

Patient	Light Chain Isotype (κ or λ)	Heavy chain Isotype (γ or μ)	CD5/CD19 double positive (y/n)	CD5 intensity (Lo or Hi)[b]
Bel	λ	μ	y	Lo
Tre	κ	μ	y	Hi
Yar	κ	μ	y	Hi
Hod	κ	μ	y	Hi
Jak	κ	μ	y	Hi

[a] Histograms for phenotyping are shown in Figures 2.9 and 2.10 and Appendix A. All positive antigens demonstrated a unimodal shift in fluorescent intensity.

[b] Surface expression of CD5 antigen is described as either Lo or Hi if the mean fluorescence intensity shift was in either the second or third decade for FL2-H respectively.

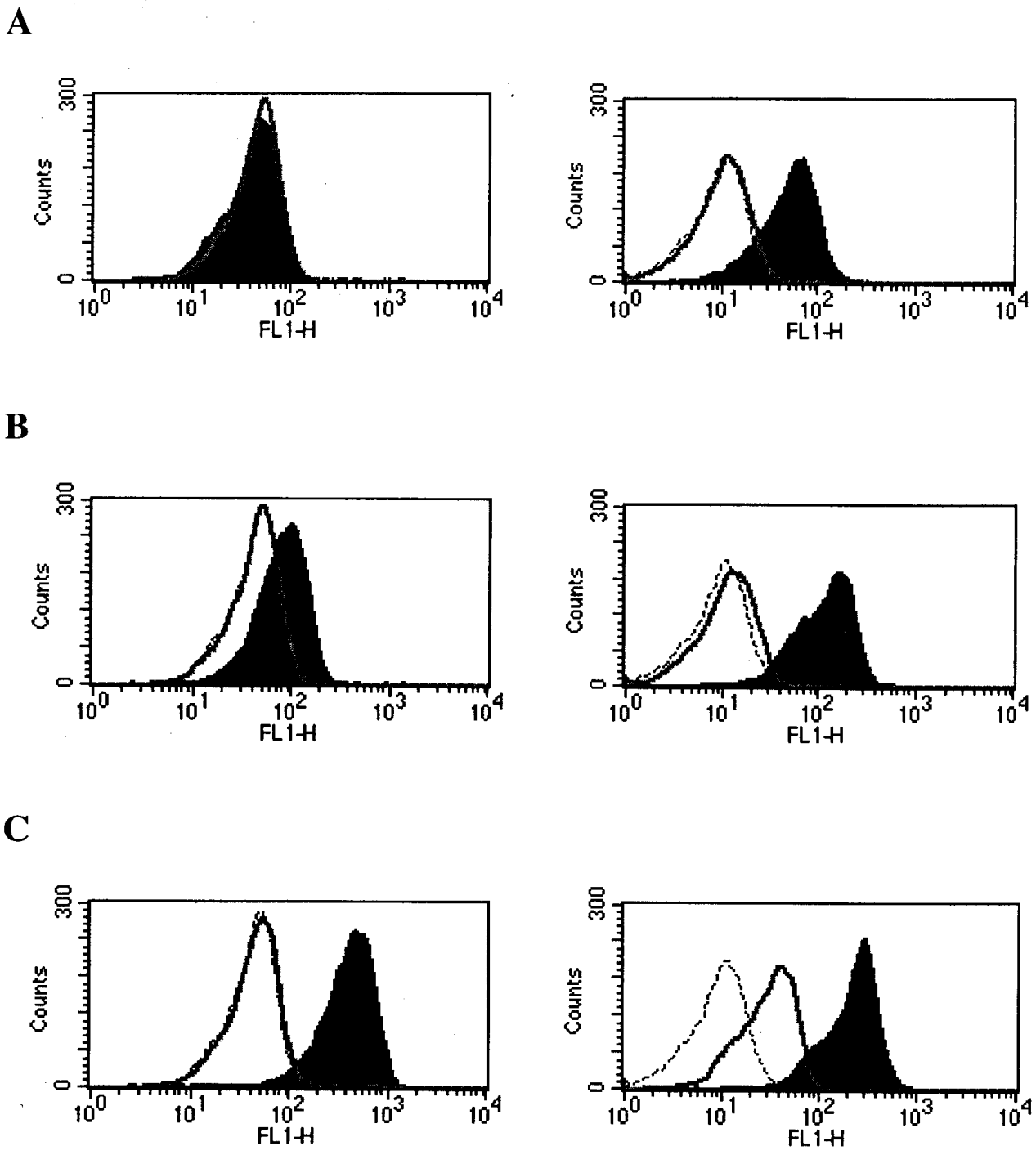


Figure 2.11: Low affinity binding of mouse IgG1/κ (K.1.21) by B CLL cells: Bel (left panels) and Hod (right panels). **A:** K.1.21 at 1 μg/mL. **B:** K.1.21 at 10 μg/mL. **C:** K.1.21 at 100 μg/mL. Cells were incubated with K.1.21 for 20 min at 4°C and washed once (solid histograms) or three times (grey lines) before detection of bound mouse IgG with a sheep α-mouse Ig FITC conjugated polyclonal antibody. Background fluorescence of the sheep α-mouse Ig FITC is shown (dashed lines).

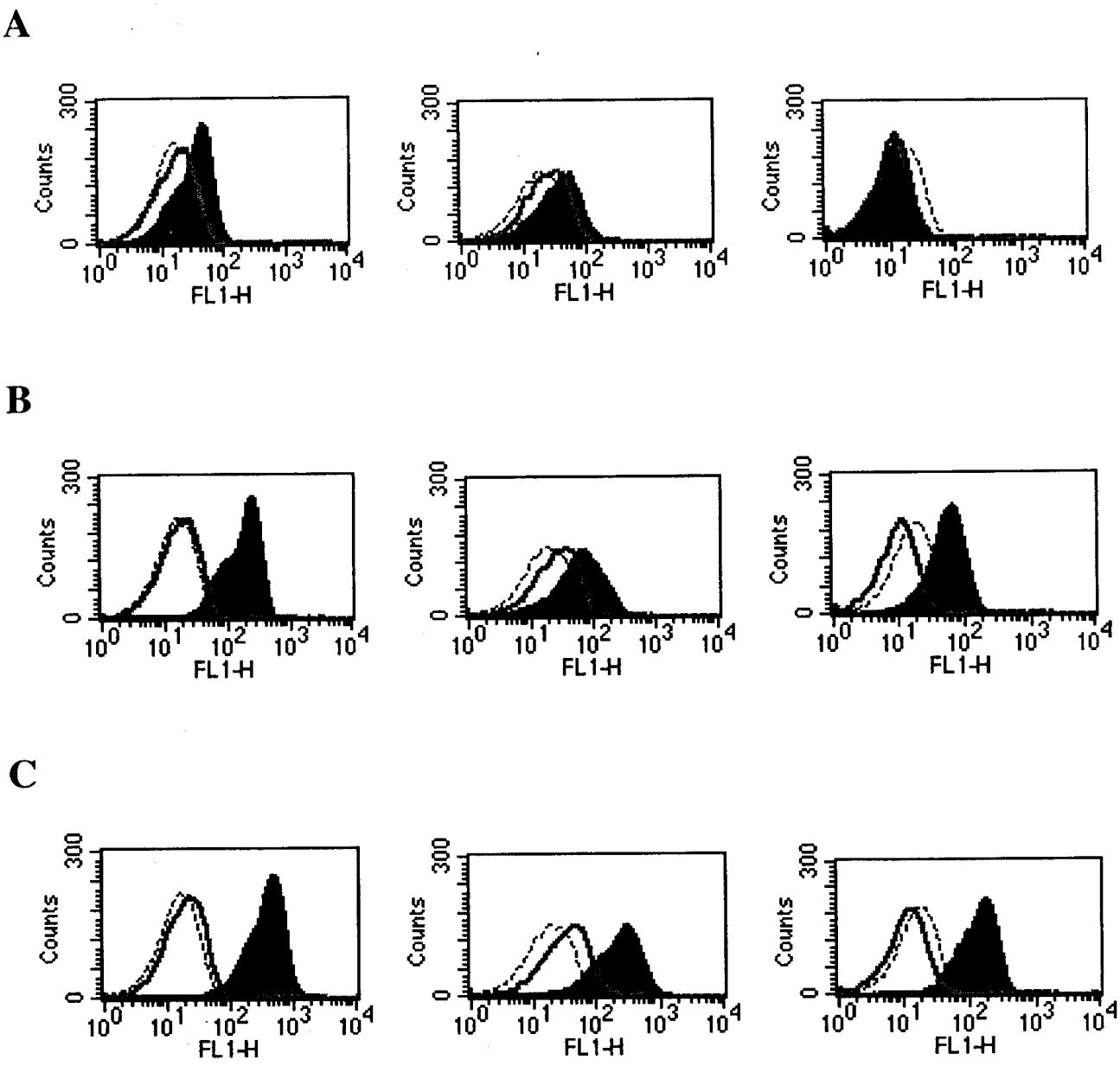


Figure 2.12: Low affinity binding of mouse IgG1/κ (K.1.21) by B CLL cells: Tre (left panels), Yar (middle panels) and Jak (right panels). **A:** K.1.21 at 1 μg/mL. **B:** K.1.21 at 10 μg/mL. **C:** K.1.21 at 100 μg/mL. Cells were incubated with K.1.21 for 20 min at 4°C and washed once (solid histograms) or three times (grey lines) before detection of bound mouse IgG with a sheep α-mouse Ig FITC conjugated polyclonal antibody. Background fluorescence of the sheep α-mouse Ig FITC is shown (dashed lines).

Table 2.5

Low affinity binding to mouse IgG1 by B CLL cells

Patient	One wash			Three washes		
	Concentration K.1.21 (μg/mL)			Concentration K.1.21 (μg/mL)		
	1	10	100	1	10	100
Bel	0.87	1.74	8.38	0.95	0.92	0.95
Tre	1.77	7.81	16.32	1.19	0.96	1.22
Yar	1.97	3.38	10.92	1.27	1.47	1.81
Hod	3.70	8.36	14.91	0.81	1.14	2.14
Jak	0.60	2.52	5.86	0.53	0.73	0.84

The mean fluorescence intensity (MFI) of the B CLL cells in the presence of mouse IgG1 (K.1.21) was normalised with background MFI given by the detection antibody (sheep α-mouse Ig FITC conjugate). Values of MFI greater than two times the background MFI were considered positive for mouse IgG1/κ binding and are boxed.

2.4 Discussion

No generalisation regarding the primary structure of the V regions of polyreactive Ig is completely without exception. However, some observations such as the biased usage of VH and VL genes and the complete lack or minimal amount of somatic mutation within V domains are true of a large proportion of murine and human polyreactive Ig. These features are to some extent recapitulated in the B CLL V domain sequences presented here (Figures 2.1 to 2.5 and Appendix A). A comparison of the VL and VH region genes expressed by the B CLL malignant B cells with highly related human germline gene sequences revealed that most of the sequences were basically unmutated from germline (all segments analysed having >98% identity to germline). Low levels of mutation, most likely somatic in origin, were observed for the VH genes of Tre, Yar and Jak immunoglobulins (Tables 2.1 and 2.3). The few observed somatic mutations may be the result of stimulation of the B cell receptor by an autoantigen during the course of the malignancy, although this phenomenon has not been tested. Furthermore, the diversity of the VL sequences is reduced through a remarkable fidelity in VL–JL joining with the absence of N-additions at the junction of VL and JL (Figures 2.1-2.5). This is consistent with the fact that around eighty percent of all normal human light chain recombinations show no evidence of N-addition (Victor & Capra, 1994). In comparison, the VH regions show both imprecise joining of VH, D and JH genes (mainly deletions to the germline genes) and sequences potentially resulting from N-addition were seen in all five sequences analysed (Appendix A). Comparison of the primary amino acid sequence of the five Fv failed to identify an obvious structural feature within the CDR which could be related to the polyreactive nature of the antibodies expressed by B CLL cells. However, the binding of mouse IgG1 by all patients B CLL cells studied suggests that a conserved specificity exists amongst these polyreactive/autoreactive IgM.

2.4.1: Monoclonality of PBLs from B CLL patients

The five Igs expressed by $CD5^+$ B CLL cells studied here were of IgM isotype for the heavy chain and the light chains were λ for patient Bel and κ for patients Tre, Yar, Hod and Jak (Table 2.4). A vast majority of the PBLs were CD19/CD5 double positive and the lymphocytes displayed a unimodal shift in fluorescence intensity when stained for IgM and λ or κ light chains (Figures 2.9 and 2.10 and Appendix A). These

results suggested that most of the PBLs were of the phenotype of the malignant B CLL cells. The light chain V genes were isolated using λ or κ specific primers and the PCR product was cloned into appropriate plasmids for sequencing. A unique VL-JL rearrangement was isolated from cDNA obtained from the PBLs of each B CLL patient. In all cases only one rearrangement was sequenced from the multiple independent clones analysed; indicating that the monoclonal rearrangement of the VL domain had been sequenced. The VH genes have been previously sequenced using an N-terminal consensus and CH1μ specific primer set (Brock, 1995). The VH genes of Bel, Tre and Yar were originally sequenced from only one clone containing a VH sized insert while the VH sequences of Hod and Jak were obtained through direct sequencing of the PCR product (Brock, 1995). The VH sequences of Bel, Tre and Yar have been confirmed through automated DNA sequencing of a further 6 to 8 independent clones obtained from cloned PCR product using an N-terminus VH and a JH specific primer set (Chapter Four). All the VH clones sequenced confirmed the sequence originally obtained by Brock (1995). The correspondence of the ten V genes to unique rearrangements indicates that there was no observable cDNA contamination from normal B Cells in the PBLs of the five B CLL patients. These results give confidence to the assumption that the correct VL–VH pairing expressed by the malignant B CLL cells as surface IgM has been obtained.

2.4.2: Immunoglobulin gene usage in B CLL

Light chain variable region genes

The λ light chain V domain of Bel is the product of a rearrangement of a Vλ3 family gene (IGLV3S2) to the Jλ2 germline gene. The four κ light chains of this study utilise Vκ genes from families 1 and 4. While the Jκ2 gene is used three times (Tre, Yar and Hod) the Jκ1 segment is used once (Jak). Two rearrangements involved Vκ1 family genes. Tre used the germline gene DPK9 and Jak used a diffierent Vκ1 family gene designated as A30. Yar and Hod utilised the same Vκ4 family germline gene (DPK24) rearranged onto the Jκ2 germline gene. In fact the only differences observed between the DNA sequences of Hod and Yar are attributed to somatic mutations in the VL gene expressed by Hod which will be discussed separately (Section 2.4.3). The finding of an almost identical VL rearrangement in two of the five sequences and the use of the same Jκ2 gene segment by three of the Ig supports the idea of a restricted V gene usage by the light chains of the autoreactive/polyreactive IgM expressed by B CLL cells.

While around 40% of all B CLL malignancies express λ light chains (Ikematsu *et al.*, 1994) there are only a few sequences of the Vλ genes expressed by these B cells (Pritsch *et al.*, 1993). Most investigations of VL genes expressed in B CLL have reported on Vκ gene usage (Kipps & Carson, 1993; Friedman *et al.*, 1992). Immunoglobulin Bel is the product of the Vλ gene IGLV3S2. Since Bel IgM was the only λ light chain antibody included in this study no conclusions can be made regarding the significance of the expression of a particular Vλ germline gene in cases of B CLL where λ light chains are expressed.

None of the κ light chains were products of the Vκ3 family gene *Humkv325* which is frequently rearranged in B CLL (Kipps & Carson, 1993). Around 25% of κ light chain expressing B CLL cells display the 17.109 CRI (Kipps, 1993). It is not surprising that the Vκ regions sequenced in this study did not result from the rearrangement of the *Humkv325* gene since on a statistical basis only one of the sequences would be expected to display the associated 17.109 CRI. Two of the B CLLs (Tre and Jak) were found to express different Vκ1 family genes, which is the largest and most frequently rearranged of the Vκ gene families (Cox *et al.*, 1994). Tre and Jak were found to utilise Jκ2 and Jκ1 germline genes respectively. In combination, with the use of two different Vκ1 and Jκ genes, the light chains of Tre and Jak suggest a random use of Ig genes. However, the Vκ regions of Yar and Hod are products of the same Vκ4 germline gene (DPK24) and the Jκ2 gene segment. The Vκ4 family has only one member which has been found in functionally rearranged antibodies (Tomlinson *et al.*, 1995; Cox *et al.*, 1994) so it is not surprising to find the same germline gene in two separate rearrangements involving the Vκ4 gene family. However, it is remarkable that in a group of only four κ light chain B CLLs that two of the rearrangements utilise the single Vκ4 functional germline gene and the Jκ2 gene segment. Together with *Humkv325* the Vκ4 segment designated DPK24 in combination with Jκ2 may be found in a large number of B CLL cases where a κ light chain is expressed.

Heavy chain variable region genes

Heavy chain V regions of the B CLL Ig utilise VH genes from families 1 (Hod), 3 (Jak and Yar) and 5 (Bel and Tre). Yar and Jak VH sequences contained a stretch of sequence assigned to the DXP4 minigene. Bel used the DXP'1 minigene, Tre used the D6-19 minigene and Hod used the DXP1 minigene. Both Bel and Hod utilised the

JH6 gene (Bel: JH6c; Hod: JH6a), while the JH4b gene was used by Tre and Jak and the JH5b gene was used by Yar (Table 2.3 and Appendix A).

Human polyreactive immunoglobulins frequently utilise members of the small VH 4, 5 and 6 gene families. The VH regions of Bel (DP-73) and Tre (VH32) are encoded by members of the VH5 gene family. The VH251 gene has previously been noted in heavy chains expressed in up to 30% of VH5 family expressing B CLL tumours (Humphries *et al.*, 1988). While VH251 appears not to have been expressed by the two B CLLs expressing VH5 family genes in this particular study, the genes DP-73 and VH32 are highly similar to VH251. This similarity resulted in the assignment of VH251 to both Bel and Tre VH region gene sequences in the original study of Brock (1995). It is interesting that historically the VH gene sequences of certain antibodies may have been erroneously assigned due to the lack of an exhaustive sequence database of human H chain germline genes, a subject which has more recently been addressed (reviewed in: Cook and Tomlinson, 1995). The VH regions of Bel and Tre are products of different D minigenes and JH gene segments (Table 2.3), an indication that the primary structures of these antibodies are most different within the HCDR3 regions. Moreover, Bel expresses a Vλ while Tre expresses a Vκ region which further emphasises a difference in the primary and corresponding three-dimensional structures of these two Igs within their V domains. The VL and VH region genes of Bel and Tre are good examples of where the expression of highly related VH genes should not be taken in isolation as an indicator of a potential conservation of binding site structure within a group of immunoglobulins.

Of all the VH gene families, the VH3 gene family is the largest and most frequently rearranged in normal humans (Cook & Tomlinson, 1995). The p1 gene expressed by Yar and the DP-49 gene expressed by Jak B CLL cells are members of the VH3 gene family and have been found expressed in B cells producing both monoreactive and polyreactive immunoglobulins (Deane & Norton, 1990; Cook & Tomlinson, 1995). The use of two different VH3 family genes by two of the five B CLL tumours studied here supports evidence that VH3 genes are rearranged as frequently in B CLL as normal B cells. While the VH1 family of genes is the second largest family, these germline genes are rearranged less frequently in B CLL compared to normal B cells. There is also an increased frequency of rearrangements in B CLL which involve the small VH4, 5 and 6 gene families (Deane & Norton, 1990). Only one B CLL (Hod) studied here expressed a VH1 family gene. Hod expresses the DP-10 (51p1) gene which, in its unmutated form, is related to the G6 CRI. The G6 CRI has been found in many VH1 expressing B CLL cells (Kipps, 1993). This overexpression of DP-10 in

B CLL provides evidence that even though all gene families are expressed in these B cells there is a propensity for particular VH genes. Furthermore, as more VH gene sequences from human polyreactive Ig are published, specific members of the VH3 gene family such as p1 (Yar) and DP-49 (Jak) may emerge as genes which predispose an Ab towards binding in a polyreactive manner.

Two of the VH sequences (Yar and Jak) contained gene segments assigned to the DXP4 germline D minigene. The remaining three sequences contained unique gene segments encoded by DXP'1 (Bel), D6-19 (Tre) and DXP1 (Hod) D minigene segments. This has resulted in significant differences in the VH-D-JH junctions of all the B CLL derived immunoglobulins.

Selective use of JH genes may occur in autoreactive/polyreactive Ig expressing B CLL tumours. In B CLL patients with autoimmune haemolytic anaemia, one third (4/12) of the VH regions contain a tyrosine rich HCDR3 encoded by a JH6 gene (Efremov *et al.*, 1996). Interestingly, two of the B CLL studied here utilise JH6 gene segments (Bel: JH6c; Hod: JH6a). Tre and Jak VH rearrangements involve the JH4b gene and Yar the JH5b gene which, unlike JH6, does not result in the HCDR3 of these Ab having long stretches of tyrosine residues.

2.4.3: Incidence of somatic mutation in variable region genes

The light and heavy chain V domain sequences were found to contain either none or very few mutations from germline Ig gene sequences (Tables 2.1 to 2.3). Low levels of mutations, presumably somatic mutations, were observed in the VL gene of Hod (4 mutations) and the VH genes of Tre (4 mutations), Yar (6 mutations) and Jak (4 mutations). Otherwise the genes contained either a single or zero differences from germline sequence. A complete absence of somatic mutations was observed for the JL genes. The D and JH genes were not analysed for somatic mutations since significant junctional diversity (Section 2.4.4) makes it difficult to assign somatic mutations in these genes.

While the majority of the V region sequences appear to correlate with these Ab falling into a naive or germline population, the low level of somatic mutation observed could be significant. The presence of somatic mutation in some but not all of the VL and VH genes could be the result of the polyreactive IgM expressed by the B CLL cells

recognising an autoantigen which provides some stimulus for somatic hypermutation. The incidence of somatic mutation in the V region genes reported here may support the theory that an Ag which is most likely an autoantigen is involved in aetiology or pathogenesis of B CLL (Schroeder & Dighiero, 1994). Furthermore, the low numbers of mutations seen in most of the VL and VH sequences reflects the frequent germline nature of the 'natural' or polyreactive Ig repertoires of humans and mice (Baccala *et al.*, 1989; van der Heijden *et al.*, 1991; Pascual *et al.*, 1992).

2.4.4: Junctional diversity in variable regions

A remarkable precision of joining of VL to JL genes was seen for all five B CLL light chain V region sequences (Figures 2.1-5). The same splicing of Vκ to Jκ genes occurred in the four κ light chains and a similar splice site was seen for the VL–JL of Bel, indicating that junctional diversity is reduced to negligible levels for the light chain V regions of immunoglobulins expressed in cases of B CLL. No N-additions were found at the VL–JL gene junctions as expected from previous studies of normal and autoantibody VL region gene sequences (Victor & Capra, 1994). The partner VH domain Jak contained a 14 nucleotide stretch at the junction of the VH and D gene elements attributed to N-additions. Short sequences were also seen at the VH and D junctions of Yar and Hod, although it is unclear whether these nucleotides were the result of N-addition. Sequences which are potentially the result of N-addition were also found at the D–JH junctions of Bel, Tre and Yar immunoglobulins. Furthermore, for Bel, Tre, Hod and Jak the D genes were truncated compared to the possible coding regions of the germline D minigenes (Brock, 1995 and Appendix A). These findings confirm that the HCDR3 of the Ig expressed in B CLL display extensive variability in sequence and length. Human polyreactive Ab have previously been shown to display high levels of variability in the sequence and length of HCDR3 (Pascual *et al.*, 1992; Ditzel *et al.*, 1996). In contrast, the high fidelity of VL–JL joining described here suggests that the variability in sequence and length of LCDR3 loops is not accentuated in polyreactive compared to monoreactive immunoglobulins.

2.4.5: Primary structures of variable domains

In this thesis the question of whether the combining sites of polyreactive antibodies are somehow conserved in structure compared to monoreactive antibodies is posed. To address this, the primary structures of the five VL and VH domains were compared to determine if any motifs could be identified within CDR sequences which may correlate with polyreactive Ag binding. If the highly similar VL regions of Yar and Hod are excluded from the discussion, the remaining V domains display distinct differences at the primary sequence level (Figures 2.6 and 2.7). No particular sequence motif was conserved for all five Fv within CDR sequences. The high level of variability within CDR sequences observed for the B CLL sequences is reminiscent the hypervariability of these sequences in normal immunoglobulins (Kabat *et al.*, 1991; Kabat *et al.*, 1977). Moreover, comparison of the amino acids within the CDR sequences (Figure 2.8) of the B CLL immunoglobulins revealed no striking conservation of chemical groups which could be assigned to a particular structural motif for polyreactive Ag binding. This type of analysis suggested that the key to polyreactivity may be found upon examination of three-dimensional rather than primary structures of antibodies of this type.

2.4.6: Low affinity binding to mouse IgG1 by B CLL cells

The binding of mouse IgG1/κ (K.1.21) to the surface of the five CD5 positive B CLL cells was demonstrated using indirect immunofluorescence (Figures 2.11 and 2.12; Table 2.5). This binding to mouse Ig has been previously shown to be isotype independent since the B CLL cells bind mouse antibodies of IgM, IgG1, IgG2a, IgG3 and IgA isotypes (Weston & Raison, 1991). These experiments suggest that the reactivity of the surface IgM of B CLL cells with mouse Ig is not a classical IgG Fc binding as occurs with rheumatoid factor (RF) activity and that the B CLL derived IgM may in fact bind to the light chain rather than heavy chain.

Polyreactive antibodies produced by B CLL cells are potentially stimulated in a similar manner by an endogenous or foreign Ag during the initial onset or course of the malignancy (Schroeder & Dighiero, 1994). The binding by all five B CLL cells to mouse IgG1/κ (Figures 2.11 and 2.12; Table 2.5) supports a common reactivity for

the polyreactive IgM, although mouse Ig is obviously not a selecting Ag in the malignancy. However, this conserved reactivity towards mouse immunoglobulins does suggest that a common structural mechanism is involved in the binding. The relative abilities of different B CLL cells in binding mouse IgG1/κ was investigated by incubating the cells with 1, 10 or 100 μg/mL of K.1.21 (Table 2.5). Binding was only detected at the highest concentration of mouse IgG/κ for patient Bel. Patients Tre, Yar and Jak were shown to bind mouse IgG/κ at 10 and 100 μg/mL and Hod at all concentrations of mouse IgG/κ tested. The binding was shown to be removed by three washing steps, indicating the low affinity nature of the interaction. While it is possible the surface IgM of Bel was of particularly low and Hod of higher affinity compared to the other B CLL IgM molecules, an alternative explanation is that the total surface Ig concentration varied between the B CLL tumours.

2.4.7: Conclusion

Variable region genes expressed in five cases of B CLL have been sequenced and the partner VL–VH domains have been assessed for gene usage, somatic mutation, junctional diversity and primary amino acid structural similarity. The expression patterns of VL and VH region genes exhibited by the B CLL cells strengthens the idea that restricted genes are utilised by polyreactive Ig as compared with monoreactive Igs. Evidence for a restricted use of Ig genes is supported by the use of the same germline VL and JL genes in Yar and Hod and the expression of VH5 family genes in Bel (DP-73) and Tre (VH32) immunoglobulins. However, no apparent restricted gene usage was noted for D minigenes or JH gene segments. The VL and VH regions showed either an absence or low levels of mutations from germline gene sequences reflecting the naive or germline nature of polyreactive Ab in general. Junctional diversity does not appear to be increased, when compared to normal B cells, for the VL–JL joining since all five VL regions displayed similar recombination splice sites. In contrast, the VH–D–JH junctions showed a high frequency of N-additions and deletions to the D genes. This resulted in the primary structures of HCDR3 being highly variable in sequence and length. Inspection of the primary structures of the CDRs revealed no striking features which could be related to polyreactive Ag binding. However, the conserved binding specificity for mouse Ig by the polyreactive surface Ig of B CLL cells suggests that a common structural feature is likely to be involved in polyreactive Ag recognition. These observations suggested that determining the 3D structure, of the binding sites of the polyreactive immunoglobulins, is an essential step in elucidating

the structural mechanisms involved in polyreactivity. The potential conservation of 3D structure of human polyreactive Ig combining sites is examined in the next chapter (Chapter Three) through the use of homology models of Fv constructed from the VL and VH domain sequences derived from the B CLL cells.

CHAPTER THREE
Homology modelling of B CLL variable regions (Fv): Structural diversity of human polyreactive Ig combining sites

3.1 Introduction

While the majority of immunoglobulins exhibit restricted antigen binding specificity, polyreactive immunoglobulins can recognise multiple structurally dissimilar antigens. Malignant cells from patients with chronic B lymphocytic leukaemia frequently express polyreactive Ig, thus providing a system for investigations into the structural features important for polyreactive Ag binding. Unlike the situation with monospecific antibodies where structural correlation of the binding site with antigen specificity has been elucidated for a number of antigen/antibody systems (Amit *et al.*, 1986; Arevalo *et al.*, 1994; Colman *et al.*, 1987; Herron *et al.*, 1989; Herron *et al.*, 1991), the structural basis of polyreactivity of certain antibody combining sites remains unclear. Clarification of this problem requires detailed structural analysis of a large number of polyreactive binding sites, a process that will be greatly assisted by the use of predicted structures in addition to the essential, but limited, number of experimentally determined structures for binding sites of this type. The structural diversity inherent in the polyreactive Ig combining site is investigated here by exploiting the V region sequences (Chapter Two) to obtain 3D structures, through homology-based modelling procedures, of five Fv fragments expressed as surface Ig by human B CLL cells. The homology modelling of Bel, Tre and Yar Fvs has been published and a copy of this paper is bound at the back of this thesis (Ramsland *et al.*, 1997).

3.1.1: Homology-based protein modelling

Homology-based or comparative protein modelling techniques rely on the use of structural features extracted from X-ray 'template' structures of proteins which have some sequence similarity with the protein being predicted. The structural features extracted from the template molecule may include the Cartesian (x,y,z) coordinates of atoms and spacial features or restraints such as pair-wise interaction distances and dihedral angles (reviewed in: Blundell *et al.*, 1987). Features which are similar between the template and modelled proteins are incorporated into the model and sequences which are unique to the modelled protein are generated using other

techniques. Construction of the template model can be carried out manually through the careful choice of a conserved protein core onto which loops and deletions can be modelled from further templates selected on the basis of structural and sequence similarities (reviewed in: Bajorath *et al.*, 1993; Sanchez & Sali, 1997). The construction and optimisation of template-based models has been automated successfully (Sali & Blundell, 1993; Srinivasan & Blundell, 1993). However, the most accurate results are still obtained through the careful choice of template structures and the best alignment is made of the template and unknown sequences using both sequence and structural comparisons. These processes require the operator to make decisions which are occasionally intuitive and subjective. Errors are introduced into the protein model in regions where the sequence similarity between the template and target sequence is low and where insertions or deletions to the template structure are required. However, errors such as those introduced during the splicing of a template loop onto the core template are minimal if the distances are small between atoms of the loop stems and similar regions on the core template (reviewed in: Bajorath *et al.*, 1993; Sanchez & Sali, 1997).

The protein model based upon structural templates is frequently refined using molecular mechanics simulations including energy minimisation and molecular dynamics within the environment provided by protein and solution force fields. The goal of model refinement is an overall improvement in intramolecular contacts (mainly a relaxation of steric strain) and the optimisation of the stereochemistry and geometry of the modelled protein (Bajorath *et al.*, 1993). Refinement generally does not result in large deviations between protein backbone atoms of the template structures and the optimised model. Since the original templates are usually highly resolved and refined X-ray structures, any large deviations resulting from refinement without the original crystallographic restraints, using conventional force-fields, may be caused by an inappropriate treatment of parameters such as protein or solvent electrostatics.

Although homology-based protein modelling has certain limitations, the predictions of protein 3D structure using such techniques can be highly accurate. Structural predictions can result in mainchain root mean square (rms) errors as low as 1 Å for 90% of residues in proteins where the identity is 40% or higher between the template and target proteins (reviewed in: Sanchez & Sali, 1997). Furthermore, comparative modelling has been used to generate useful models of proteins where the sequence identity, between the target and template structures, was as low as 27 to 37 percent. These predictions were independently assessed against the X-ray structures of the target proteins and the modelled proteins were shown, in some group's hands, to have

large regions which were accurately predicted, even where the sequence identity with know 3D structures was not obvious (reviewed in: Dunbrack *et al.*, 1997).

3.1.2: Homology modelling of immunoglobulin variable domains

The antigen binding properties of Ig are largely dictated by six (three on light chains and three on heavy chains) loops referred to as complementarity determining regions (CDR) in the variable (V) domains which provide a unique molecular surface for specific antigen recognition. The CDR are supported by a highly conserved molecular scaffolding provided by the framework regions (Fr). Various methods have been described to predict the 3D structure of antibody Fv (VL–VH) molecules (Bajorath, 1994; Barry *et al.*, 1994; Bassolino-Klimas *et al.*, 1992; Chothia *et al.*, 1986; de la Paz *et al.*, 1986; Martin *et al.*, 1991). Usually the structurally conserved framework regions of the V domains are constructed from atomic coordinates taken directly from an Ig crystal structure. The conformations of the six CDR loops are predicted by scanning the sequence for key 'canonical' structure determining residues to select a template loop (Chothia *et al.*, 1992; Chothia *et al.*, 1989) or by randomly sampling conformational space for loops with low potential energy values and minimal solvent accessible surface areas (Bajorath & Fine, 1992; Bruccoleri & Karplus, 1987). These approaches are usually adequate for five of the six loops, but the third loop on the heavy chain (HCDR3) shows the greatest variability in length and sequence, resulting in an inability to accurately predict its structure. In view of this, there is a need for a combined modelling and crystallographic approach to improve the quality of HCDR3 predictions. In comparison, framework regions and CDR loops other than HCDR3 are well predicted by homology modelling. Homology-based models of variable domain structures, when compared with experimentally determined structures, show rms deviations of less than 1.0 Å for framework and canonically predicted CDR other than HCDR3 (Chothia *et al.*, 1992; Chothia *et al.*, 1989; Martin *et al.*, 1991). Thus, homology modelling can provide a useful structural description of the binding site, comparable to the quality of a low-resolution X-ray structure (Sali, 1995).

In this study, homology-based modelling has been employed to determine the structures of the variable domains of five polyreactive Ig expressed on the surface of $CD5^+$ human B CLL cells. Examination of the basic binding site topologies and electrostatic surface models reveal an unexpectedly high level of structural diversity within the binding sites of polyreactive immunoglobulins.

3.2 Materials and Methods

3.2.1: Immunoglobulins

The nucleotide and derived amino acid sequence of the variable domains of the Ig is described in Chapter Two. The amino acids comprising the CDR loops are presented in Table 3.3. The human IgM molecules bind mouse Ig (Chapter Two) and are expressed on the surface of the monoclonal $CD5^+$ B CLL cells; characteristic features of polyreactive immunoglobulins. The Fv models of the B CLL IgM are designated as Bel, Tre, Yar, Hod and Jak.

3.2.2: Computer-aided modelling

All computational work was performed on an INDY workstation running IRIX, v.5.2 (Silicon Graphics Inc., Mountain View, CA, USA). Template models were constructed using the software TURBO-FRODO, v.5 (Biographics, Marseille, France). Minimisation and molecular dynamics were performed using X-PLOR v.3.1 (Brunger, 1992). Molecular surfaces and electrostatic potentials were calculated and visualised with GRASP v.1.2.5 (Nicholls *et al.*, 1993).

3.2.3: Modelling strategy

The B CLL Fv molecules were constructed by starting with known high-resolution crystal structures (1BAF, 1BBD, 1IGM, 2FB4, 8FAB, 1DFB, 1REI, 1HIL, 1MCP) taken from the Brookhaven Protein Databank (Bernstein *et al.*, 1977). Variable domain templates were selected from human Ig fragments on the basis of highest amino acid identity within framework regions. Where necessary, V domain templates were aligned, using rigid-body alignment of the most regular parts of the Ig β-sheets (residues 3-7, 22-26, 32-38, 63-67, 70-74, 85-91,91-104 for light chains and 3-7, 21-25, 33-39, 68-72, 77-81, 88-95, 104-109 for heavy chains (Guddat *et al.*, 1994)) over the corresponding domain of the Pot (1IGM) structure. Thus, all Fv models have the VL–VH domain interface of Pot IgM Fv (Fan *et al.*, 1992). CDR loop templates were grafted onto the frameworks using rigid body alignments of the three framework residues on either side of the CDR sequence. Structural alignments of CDR loop templates with the framework model were only accepted if the rmsd values were less than 0.3 Å for the six Cα atoms. CDR templates were selected initially on the basis of

an identical number of residues comprising the loop. Sequence homology was used as a secondary criterion. Furthermore, only residues within the CDR as defined by Kabat *et al.* (1991) were considered. The CDR were classified, where possible, into canonical classes (Chothia & Lesk, 1987; Chothia *et al.*, 1989). Loops which were found to be of a different length to the available loops of known conformation were constructed using a template-based approach, where the loop stems were built using a CDR of known conformation. Additional residues were placed onto these stems and chain closure was accomplished by grafting a 'classical' β-turn onto the CDR loop stems. The type of β-turn assigned was based on the criteria detailed by Wilmot and Thornton (1988).

Structures were optimised using the CHARMM22 topology and parameter set implemented within X-PLOR v.3.1 (Brunger, 1992). The template-based models were subjected to an initial 100 cycles of conjugate gradient energy minimisation (Powell, 1977). After this initial minimisation step, harmonic coordinate restraints of 20.0 kcal/(mole Å^2) were set for all main-chain α carbon atoms. A further 200 cycles of energy minimisation were then carried out. The Fv models were subjected to molecular dynamics and simulated annealing where the structures were initially heated to 1000 K followed by slow cooling to a bath temperature of 300 K. The time step for the simulation was set to 0.5 fs. A final step of 300 cycles of energy minimisation was carried out.

3.2.4: Comparison of structures

The stereochemistry and geometry of Fv models were assessed by calculating the deviations from ideality of bond lengths, bond angles, dihedrals and improper angles using the parameter set of Engh and Huber (Engh & Huber, 1991). All ψ,ϕ torsional angles were analysed by the method of Ramachandran and Sasisekharan (1968) using Procheck (Laskowski *et al.*, 1993). Optimal rigid-body structural alignments and root mean square deviations (rmsd) between structures were performed with TURBO-FRODO. Rmsd for backbone and side-chain distances between template and optimised structures of the Fv models were calculated using TURBO-FRODO. Interchain energies were calculated from the van der Waals (vdW) and electrostatic energy terms in X-PLOR (Brunger, 1992). For these calculations, the protein dielectric was set to 2.0 without a nonbonded cutoff using the CHARMM22 force field.

3.3: Results

3.3.1: Validation of modelling strategy

To evaluate the template-based modelling strategy, a homology model of the variable domains of the human antibody 3D6 was constructed. The 3D6 Fab structure (He *et al.*, 1992) was not included in the template database during the preparation of a 3D6 Fv model. The coordinates from 1IGM were used to determine the V domain framework regions. Residue identities in these regions of 3D6 and 1IGM were 86% and 90% for VH and VL, respectively. Template CDR structures of the same residue length as the 3D6 loops were selected on the basis of maximum sequence homology with the target CDR. LCDR1 and LCDR2 were taken from 1IGM, and HCDR1 and HCDR2 from 2FB4. Template loops of the same length were not available for LCDR3 (7 residues) and HCDR3 (17 residues). CDR3 loops were built using loop stems of known conformation, taken from corresponding loops of 1REI for LCDR3 and NC10.14 (coordinates generously supplied by Drs L.W. Guddat and A.B. Edmundson) for HCDR3, followed by chain closure with a classical β–turn (Wilmot & Thornton, 1988). The rms deviations (Cα) between the model and the crystal structure were determined after optimal rigid-body alignments of various regions of the two structures (Table 3.1). The modelled CDR loops, the individual V domains and the entire Fv fragment were compared to the published 3D6 structure, with rms deviations between the crystal and model structures of ≤1.0 Å for all regions except for HCDR3, which has an rms deviation of 1.30 Å. Figure 3.1 depicts the modelled CDR loops compared to the X-ray structure of the CDR loops of 3D6 Fab after rigid-body alignments using all CDR Cα atoms.

Table 3.1

Comparison of the 3D6 crystal structure with the homology Fv model of 3D6

Region of the variable domain	Rmsd for all Cα atoms (Å)[a]
LCDR1	0.67
LCDR2	0.35
LCDR3	0.84
HCDR1	0.23
HCDR2	0.70
HCDR3	1.30
VL[b]	0.74
VH[b]	0.49
VL-VH[b]	1.0

[a] Rmsd values shown are for the α carbon atoms used for rigid-body alignments.
[b] V domains and Fv (VL-VH) were aligned using the conserved β-sheet framework residues (ie. framework residues are 1-23, 35-49, 57-88, 98-107 for light chains and 1-30, 36-49, 66-94, 103-113 for heavy chains)

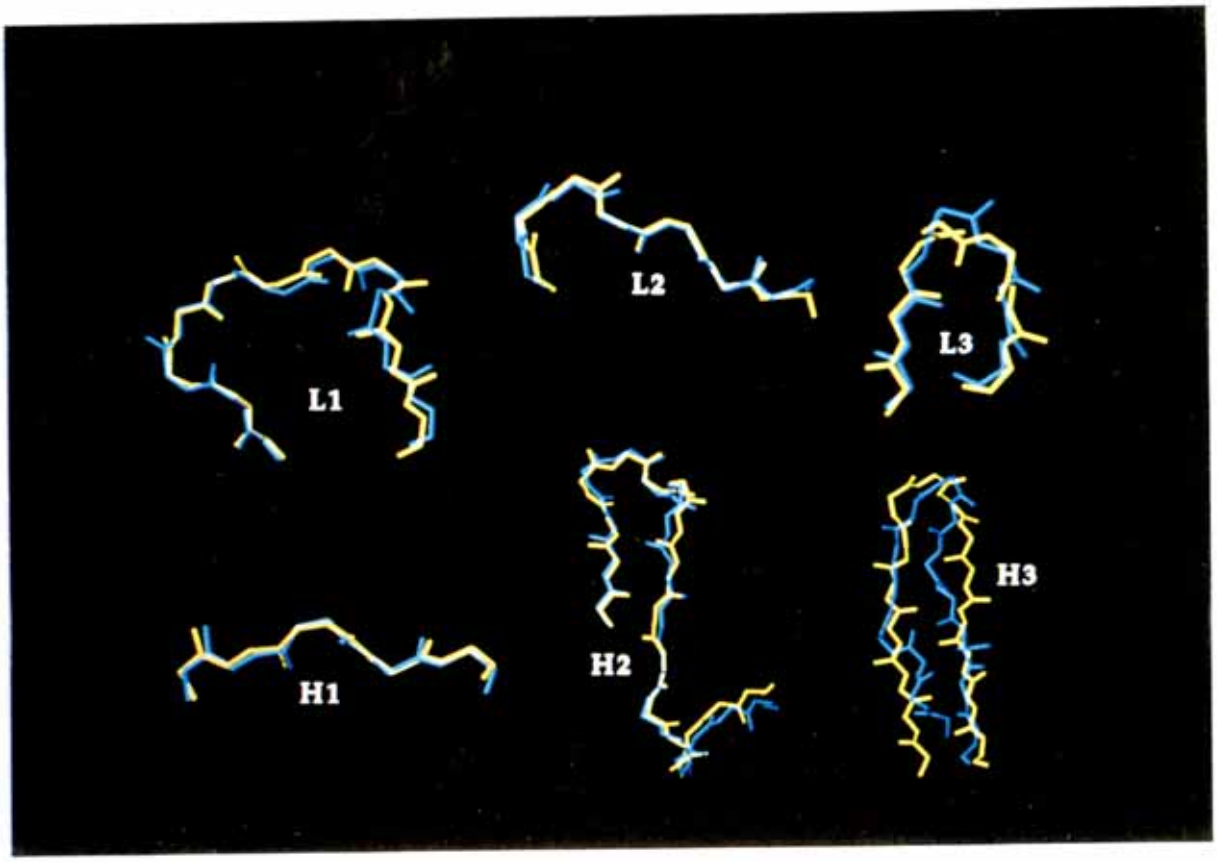


Figure 3.1: Comparison of known and predicted conformations of 3D6 CDR loops. Individual loops were aligned using all Cα atoms within the CDR of the X-ray structure (blue) and the homology model (yellow) of the 3D6 antibody.

3.3.2: Template models of polyreactive immunoglobulins

The template V domains used to model the framework regions of the polyreactive Fv are shown in Table 3.2. In each case, they represent the best identity match from the database with the respective CLL IgM V domains. Residue identities ranged from 93% for the Jak/1DFB VL pairing to 60% for the VH of Bel and 1IGM. Residues were replaced within the program TURBO-FRODO (Biographics) so that the framework sequences of the B CLL V domains were accounted for in the models. The side-chains of the mutated residues were manually positioned into a conformation which minimised any nonbonded contacts and the bond geometries were refined against a rotamer library (TURBO-FRODO). The templates used for constructing the CDR of the Fv models are summarised in Table 3.3. Where a canonical class could be assigned to a CDR loop, it is indicated. However, loops were initially chosen as templates on the basis of residue identities and similarities within the CDR as defined by Kabat *et al.* (1991).

Since immunoglobulins Bel, Tre and Yar, were used to develop the modelling methods (Ramsland *et al.*, 1997), the construction of the HCDR3s of these Fv models will be discussed in more detail than the corresponding loops in Hod and Jak. In an attempt to develop a template-based strategy for predicting the structure of HCDR3, the loop coordinates were extracted from the 30 available unliganded Ig crystal structures (29 from the Protein Databank (Bernstein *et al.*, 1977) and one from the NC10.14 Fab (coordinates provided by Drs L.W. Guddat and A.B. Edmundson)). Sequences of the HCDR3 loops used are shown in Appendix B. The length of the HCDR3 loops ranged from 5 to 17 residues but there were no loops of 6 or 13 residues. Bel and Yar antibodies both have 16 residue HCDR3 loops, the same length as a murine Ig, NC10.14. For the 11 residue HCDR3 of Tre, there were three structures (1FVC, 1HIL and 1MCP) in the database which had a unique sequence and conformation (rmsd all Cα atoms >0.5 Å).

The predicted HCDR3 conformations (Bel, Tre and Yar) and a comparison with their starting models are presented in Figure 3.2. The Bel HCDR3 was constructed from the HCDR3 of the NC10.14 Fab structure. This loop conformation fitted the Bel template model and did not result in unfavourable interactions with the other five CDR in the Fv model. The Yar HCDR3 was initially constructed using the NC10.14 conformation (Yar 1). However, when this structure was used in the Yar Fv model, the backbone of the LCDR1 loop came into close contact with the HCDR3. An alternative model for the Yar HCDR3 (Yar 2) was then constructed to incorporate the fold seen in the Pot (1IGM) HCDR3 (Fan *et al.*, 1992). Additional residues were added and the chain closed using a β–turn taken from the trypsin (4PTP) structure.

The two models of Yar Fv are presented (Figure 3.3) showing the vdW surfaces of the LCDR1 and HCDR3 loops. It was noted that the vdW radii of the HCDR3 and LCDR1 of Yar 1 exhibited many close contacts, whereas the HCDR3 vdW radii in Yar 2 made no close contacts with LCDR1. The interchain energies between the two V domains and between HCDR3 and VL were calculated for the two proposed structures of Yar Fv using X-PLOR (Table 3.4). Although the vdW energy terms were similar between the two loop structures with the VL domain, unfavourable electrostatic energy was calculated for the HCDR3 structure based on an NC10.14 template (Yar 1). In comparison, modelling based on Pot HCDR3 (Yar 2) resulted in a significantly improved electrostatic energy term. The template for the Tre HCDR3 was the 11 residue loop of 1MCP. This loop showed the greatest similarity in sequence to the Tre loop compared with the other known loops of identical length and unique conformations. A comparison of the conformations of the three 11 residue loops, of different conformation, is shown (Figure 3.2) against the final conformation predicted for the Tre HCDR3.

The immunoglobulins Hod and Jak have HCDR3 which are longer, 19 and 18 residues respectively, than all X-ray structures of HCDR3 so far determined in unliganded form (Appendix B). Consequently, the HCDR3 of these Ig were initially predicted using loop stems taken from the 17 residue HCDR3 loop of 3D6 (1DFB). The turn of the template HCDR3 was initially deleted and additional residues were placed onto the loop stems and a β–turn was used to close the modelled loops. While this approach to modelling HCDR3 of Jak was suitable (ie. it was structurally compatible with the remaining five CDR loops), such an approach was not appropriate for Hod Fv. The light chain V domains of Yar and Hod differ by only two amino acids, within Fr3, due to their use of identical germline VL region genes (Chapter Two). Similar to the HCDR3 of Yar, a HCDR3 conformation for Hod which incorporates the HCDR3 fold of Pot (1IGM) is more energetically favourable with the long LCDR1 loop than a loop based on the 3D6 HCDR3 conformation (Appendix B). Thus, the Hod Fv has a HCDR3 based upon the loop stems of 1IGM with additional residues placed onto the stems and chain-closure with a β–turn.

Table 3.2

Variable domains used to construct Fv model framework regions

B-CLL Fv model (isotype)	VL framework residue identity (%) (PDB code)	VH framework residue identity (%) (PDB code)
Bel (IgM λ)	75 (8FAB)	60 (1IGM)
Tre (IgM κ)	82 (1IGM)	62 (1IGM)
Yar (IgM κ)	69 (1IGM)	87 (1IGM)
Hod (IgM κ)	69 (1IGM)	70 (1IGM)
Jak (IgM κ)	93 (1DFB)	87 (8FAB)

Table 3.3

CDR sequences of the polyreactive Fv molecules and the CDR templates selected for modelling

CDR[a]	Bel (PDB codes)[b,c]	Tre (PDB codes)[b,c]
L1	GGNNIGSDSVH (8FAB)	RASQSISSYLN (1DFB), **2**
L2	YDSDRPS (2FB4)	AASSLQS (1DFB), **1**
L3	QVWDSSSVV (8FAB)	QQSYSTPPYT (1BAF), **1**
H1	NYWIG (8FAB), **1**	SYWIS (1HIL), **1**
H2	IIYPGDSDTRYSPSFQG (2FB4), **2**	RIDPSDSYTNYSPSFQG (1BBD), **2**
H3	RTGTGDPYYYYYYMDV (NC10.14)	RQWLALGHFDY (1MCP)

CDR[a]	Yar 2 (PDB codes)[b,c]	Hod (PDB codes)[b,c]
L1	KSSQSVLYSSNNKNYLA (1BBD), **3**	KSSQSVLYSSNNKNYLA (1BBD), **3**
L2	WASTRES (1DFB), **1**	WASTRES (1DFB), **1**
L3	QQYYSTPYT (1REI), **1**	QQYYSTPYT (1REI), **1**
H1	SYAMH (1DFB), **1**	SYAIS (2FB4), **1**
H2	AISSNGGTTYYADSVKG (1HIL), **3**	GIIPIFGTANYAQKFQG (1BBD), **2**
H3	TYYDFWSGYSPNWFDP (1IGM, 4PTP)	DIYDILTGYTYYYYYGMDV (1IGM)

CDR[a]	Jak (PDB codes)[b,c]
L1	RASQGIRNDLG (1DFB), **2**
L2	AASSLQS (1DFB), **1**
L3	LQHNSYPPWT (1BAF), **1**
H1	SYGMH (8FAB), **1**
H2	VIWYDGSNKYYADS (8FAB), **3**
H3	GEGGMYYDFWSGKYYFDY (1DFB)

[a] CDR sequences were assigned according to Kabat *et al.* (1991).

[b] Residue identities between the template loop (PDB code indicated) and the polyreactive Fv sequences have been underlined.

[c] Where a 'canonical' class could be assigned to a CDR, the class is indicated in bold (Chothia & Lesk, 1987; Chothia *et al.*, 1989).

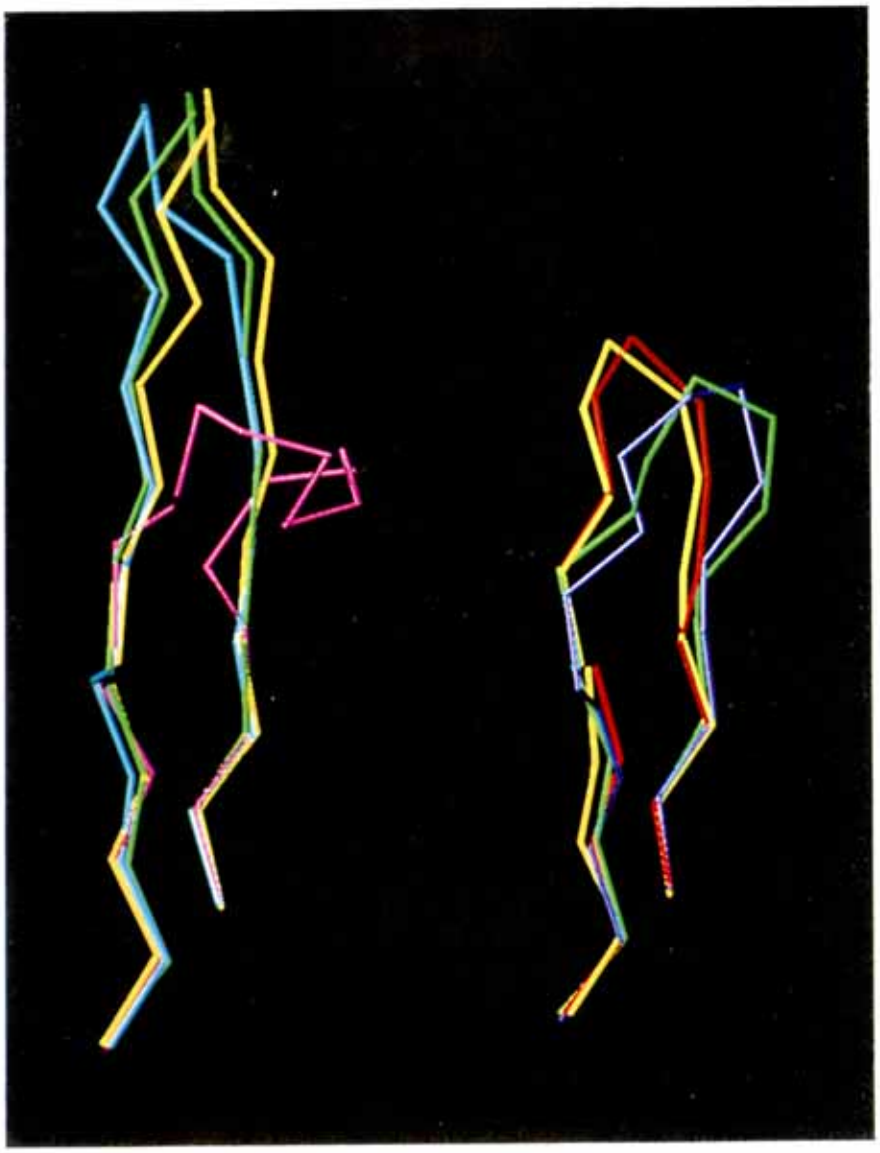


Figure 3.2: Cα representations of the HCDR3 conformations. Left: the 16 residue HCDR3 loops. Bel is coloured blue, Yar 1 is in yellow and Yar 2 in magenta. The crystal structure of the HCDR3 of NC10.14 is in green. Right: the 11 residue HCDR3 loops. Tre is yellow, and the crystal structures of 1FVC, 1HIL and 1MCP are in green, purple and red, respectively.

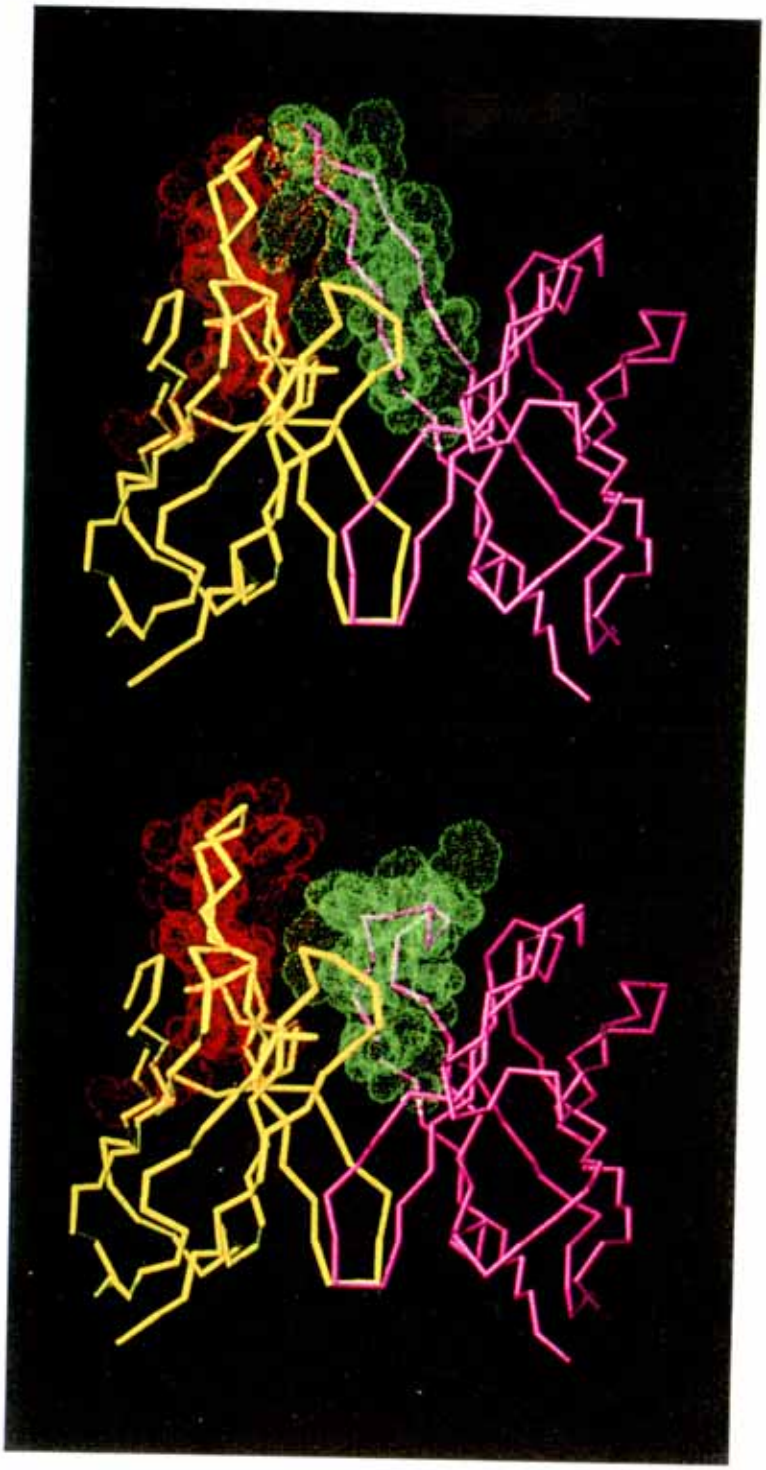


Figure 3.3: Side views of the two Yar Fv models. The heavy chain V domains are represented in magenta and the light chains in yellow. The vdW radii are in red for LCDR1 and in green for HCDR3. Top: Yar 1 based on the HCDR3 conformation of NC10.14; bottom: Yar 2 based on the HCDR3 conformation of Pot (1IGM).

Table 3.4

Interchain interaction energies of two alternative Yar Fv models

Interaction	E-total[a]	E-vdW[b]	E-ELEC[c]
Yar 1			
VL-VH	-58.3	-84.1	25.7
VL-HCDR3	-7.5	-45.9	38.4
Yar 2			
VL-VH	-103.4	-86.5	-16.9
VL-HCDR3	-50.5	-45.8	-4.7

[a] All energy values are calculated for interchain interactions only using X-PLOR (Brunger, 1992).

[b] van der Waals energy term.

[c] Electrostatic energy term.

3.3.3: Refined Fv models

The Fv models after energy minimisation and molecular dynamics refinement using the program X-PLOR (Brunger, 1992) are presented in Figures 3.4 and 3.5. The rms deviations in bond lengths, bond angles and dihedral angles are presented in Table 3.5. Ramachandran plots (Ramachandran & Sasisekharan, 1968) were calculated and the majority of ψ,ϕ torsion angles around Cα atoms are within the allowed regions. These plots are shown for the Fv models after the full positional refinement protocol (Figure 3.6 and Appendix B) and are a further indication that appropriate stereochemistry and geometry have been achieved in the models. Figure 3.7 depicts rmsd comparisons between the template (initial) and refined (final) models. The majority of backbone atoms were not displaced by large distances during model refinement with the largest movements observed in CDR residues, where most of the structural variation occurs in immunoglobulins. Side-chain positions were affected more noticeably by model refinement; appropriate with these groups adopting minimal energy conformations during the simulation.

3.3.4: Electrostatic surface models and location of aromatic side-chains in polyreactive binding sites

Binding site topology is shown as a surface representation (Nicholls *et al.*, 1993) in Figure 3.8. The solvent-accessible surfaces and the electrostatic potentials are shown for the polyreactive Fv structures. A deep binding pocket is seen at the VL–VH domain interface of Bel, Tre and Jak Fv models. The two alternative Yar Fv models (Yar 1 and Yar 2) are also represented in Figure 3.8. The energetically favoured Yar model (Yar 2) and Hod Fv have a relatively flat binding surface. The electrostatic surfaces of the Fv molecules show considerable variation in the location of surface charges or hydrophobic regions. Residues with aromatic side-chains occur in CDR loops at a higher frequency compared with similar residues in framework regions suggesting the importance of such residues in Ag binding (Padlan, 1990). The location of aromatic side-chains within the binding sites of the polyreactive Fv models was examined (Figure 3.9). The polyreactive Fv contain a high number of aromatic residues with side-chains which contribute to the molecular surface. The aromatic side-chains were distributed across most of the potential binding sites of all the B CLL Fv molecules.

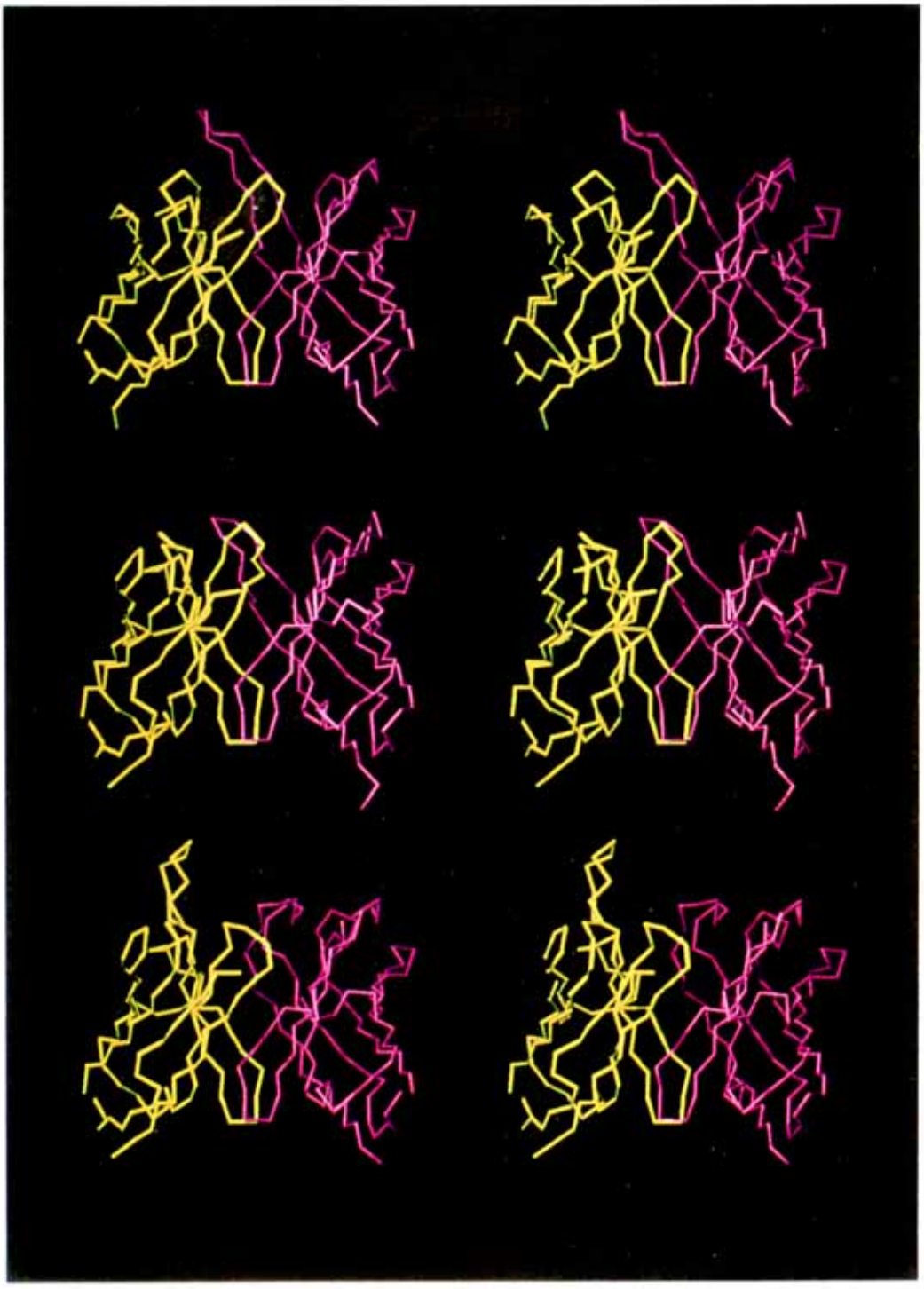


Figure 3.4a: Stereo Cα diagrams of the side views of the polyreactive Fv homology models. The heavy chains are in magenta and the light chains are in yellow. Top: Bel; middle: Tre; bottom: Yar.

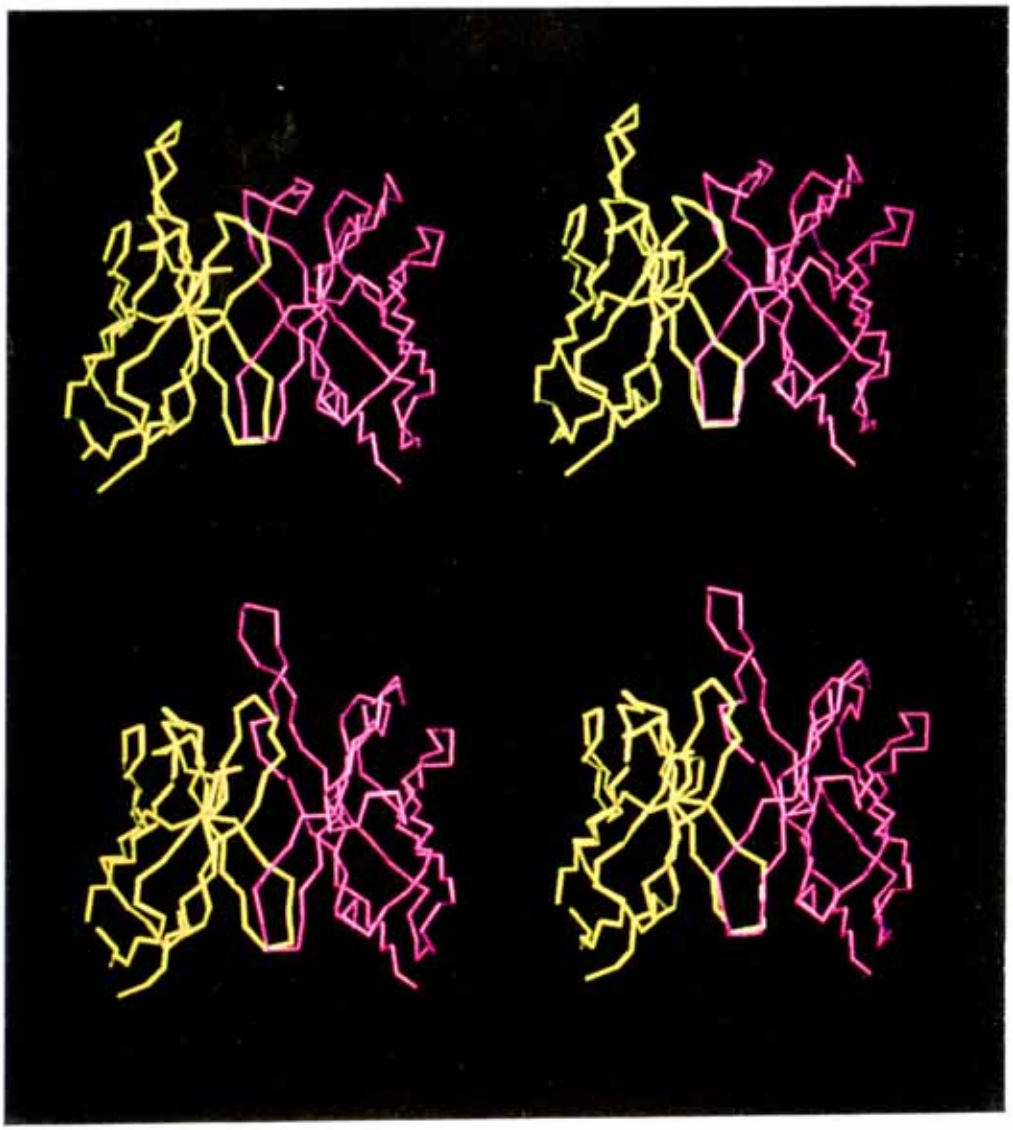


Figure 3.4b: Stereo Cα diagrams of the side views of the polyreactive Fv homology models. The heavy chains are in magenta and the light chains are in yellow. Top: Hod and bottom: Jak.

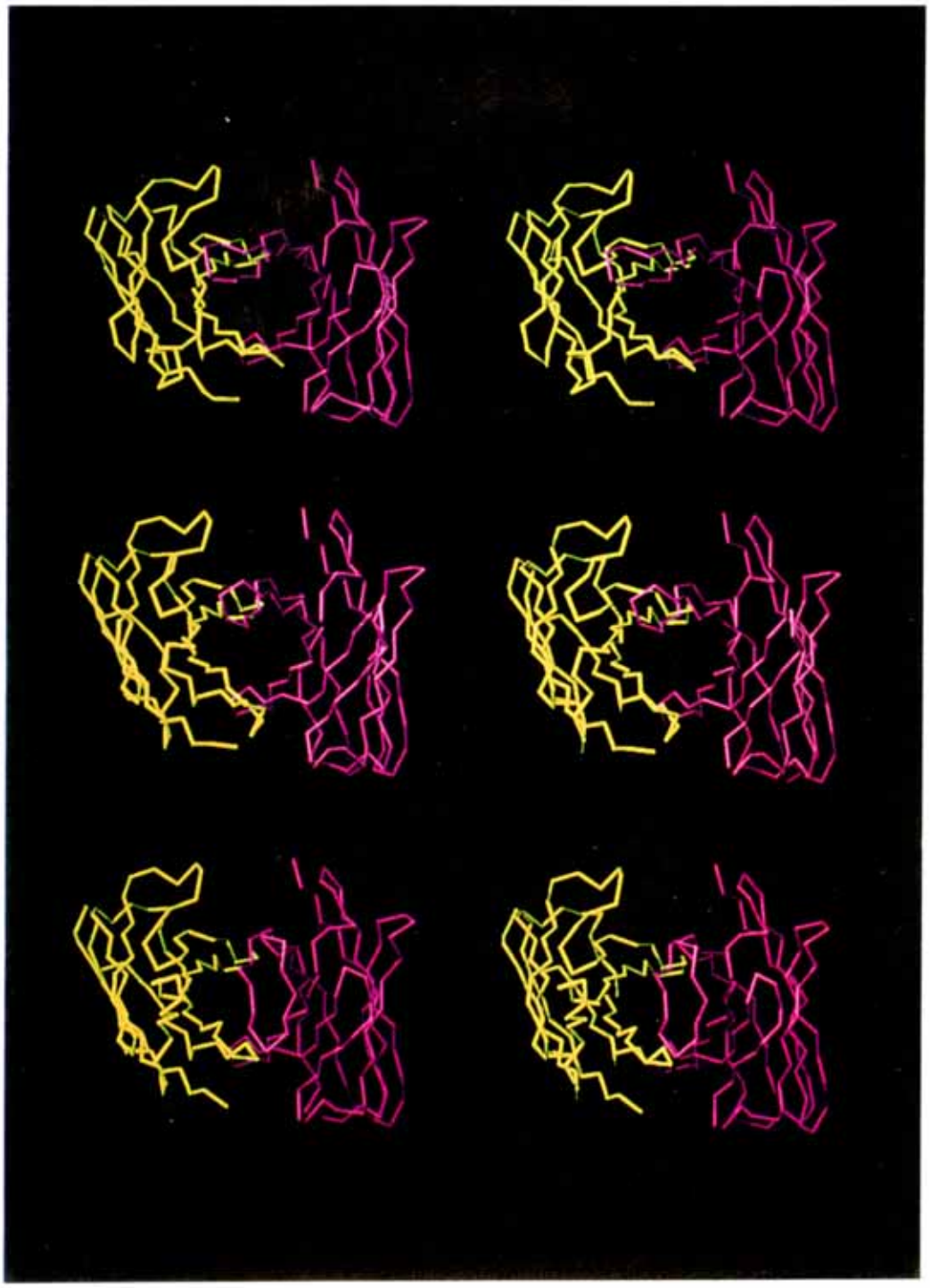


Figure 3.5a: Stereo Cα diagrams of the end-on views of the polyreactive Fv homology models. The heavy chains are in magenta and the light chains are in yellow. Top: Bel; middle: Tre; bottom: Yar.

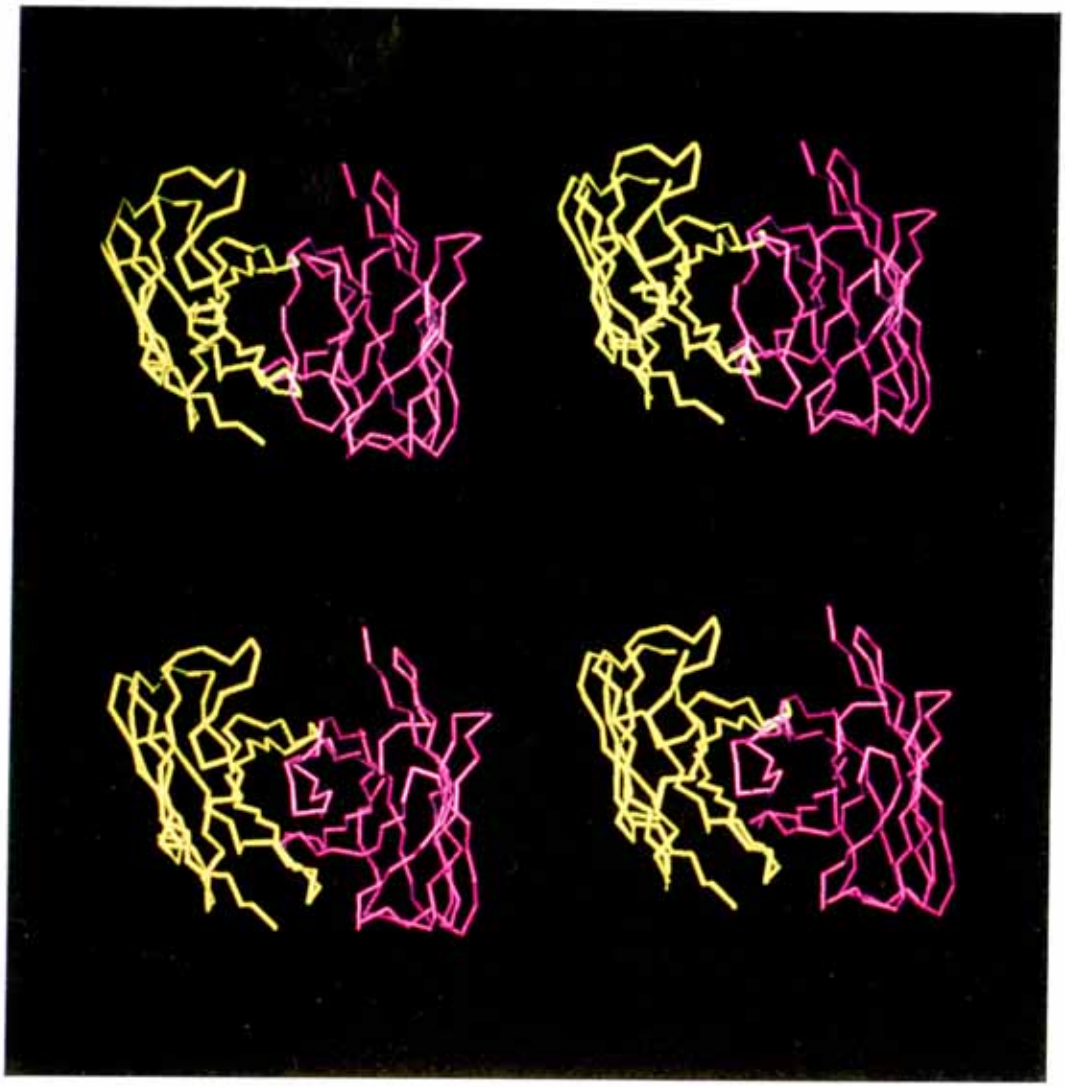


Figure 3.5b: Stereo Cα diagrams of the end-on views of the polyreactive Fv homology models. The heavy chains are in magenta and the light chains are in yellow. Top: Hod and bottom: Jak.

Table 3.5

Fv model stereochemistry after positional refinement

rms. deviations[a]	Bel	Tre	Yar 2
bond distances (Å)	0.004	0.004	0.004
bond angles (°)	0.9	1.0	1.0
dihedral angles (°)	26.0	26.1	26.0
improper angles (°)	0.8	0.9	0.9

rms. deviations[a]	Hod	Jak
bond distances (Å)	0.004	0.004
bond angles (°)	0.9	1.0
dihedral angles (°)	25.5	25.6
improper angles (°)	0.9	0.9

[a] Rmsd from ideality were calculated for all residues in the given Fv model structure, compared against Engh and Huber parameters (Engh & Huber, 1991).

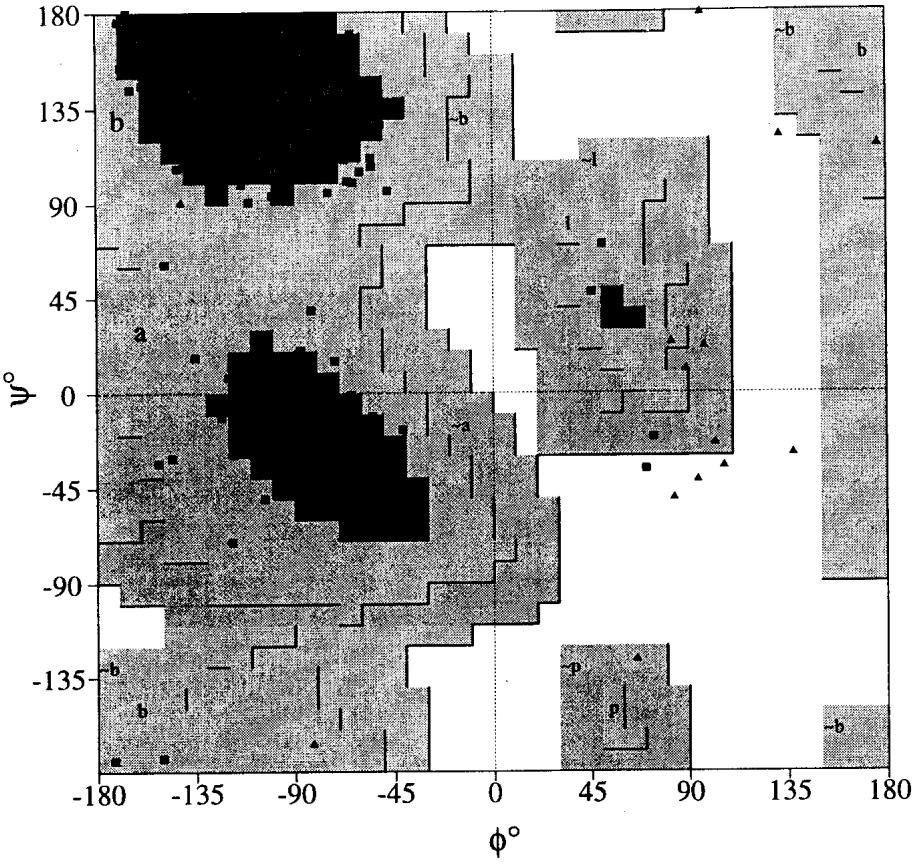


Figure 3.6: Ramachandran plot (Ramachandran & Sasisekharan, 1968) of the ψ,ϕ angles for all Cα dihedral angles in the Bel Fv homology model. Squares represent all the residue types excluding glycines which are described by triangles. Regions are shaded: A, B, L for core regions; a, b, l and p for allowed regions; and ~a, ~b, ~l and ~p for generously allowed regions. Plot was calculated using PROCHECK (Laskowski *et al.*, 1993)

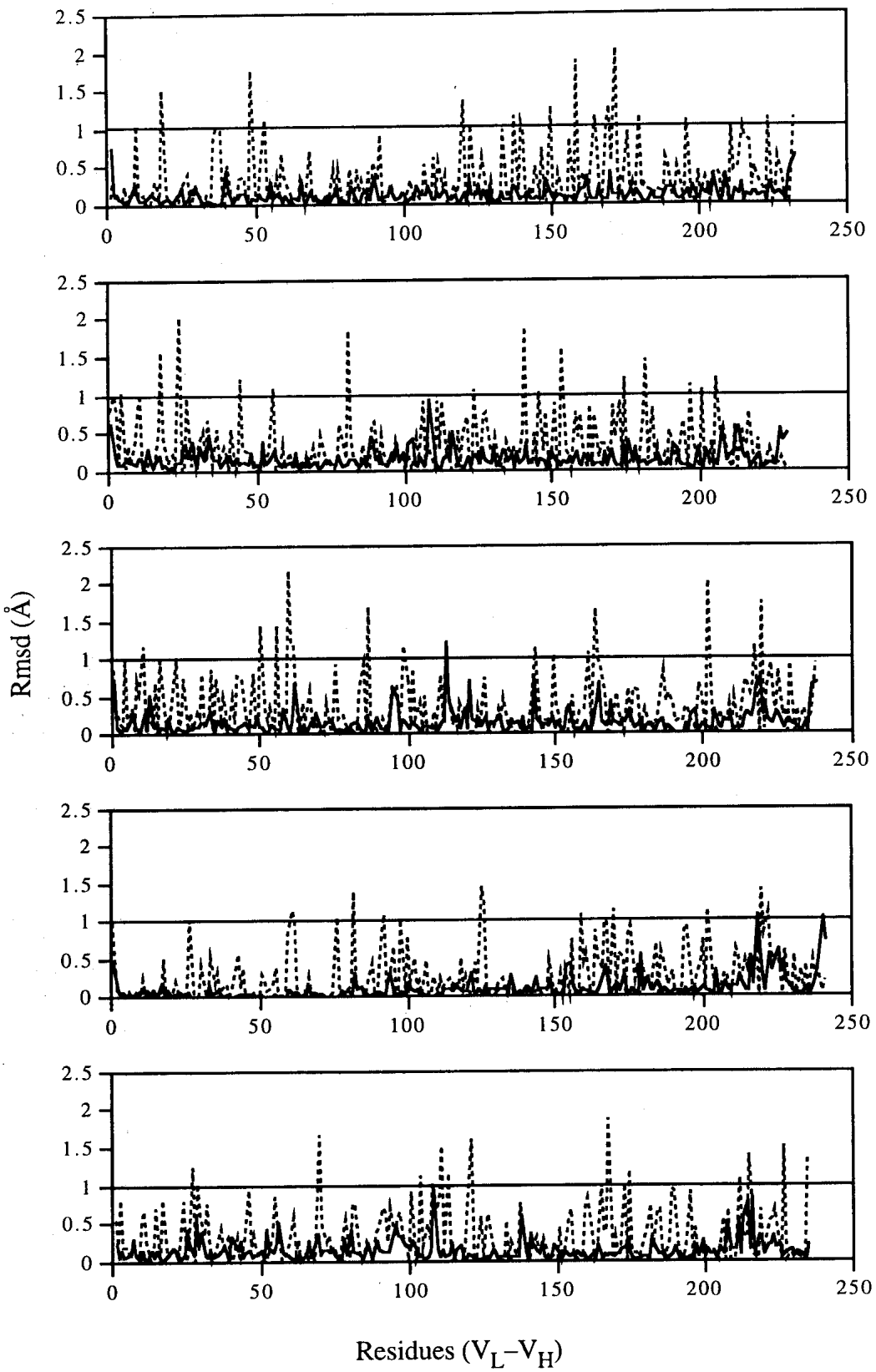


Figure 3.7: Residue by residue comparison of template with refined Fv models. Rmsd were calculated using TURBO-FRODO. Solid line delineates rmsd of 1 Å. Plots of mainchain (solid lines) and side-chain (dashed) rmsd are shown. Top to bottom: Bel, Tre, Yar 2, Hod and Jak.

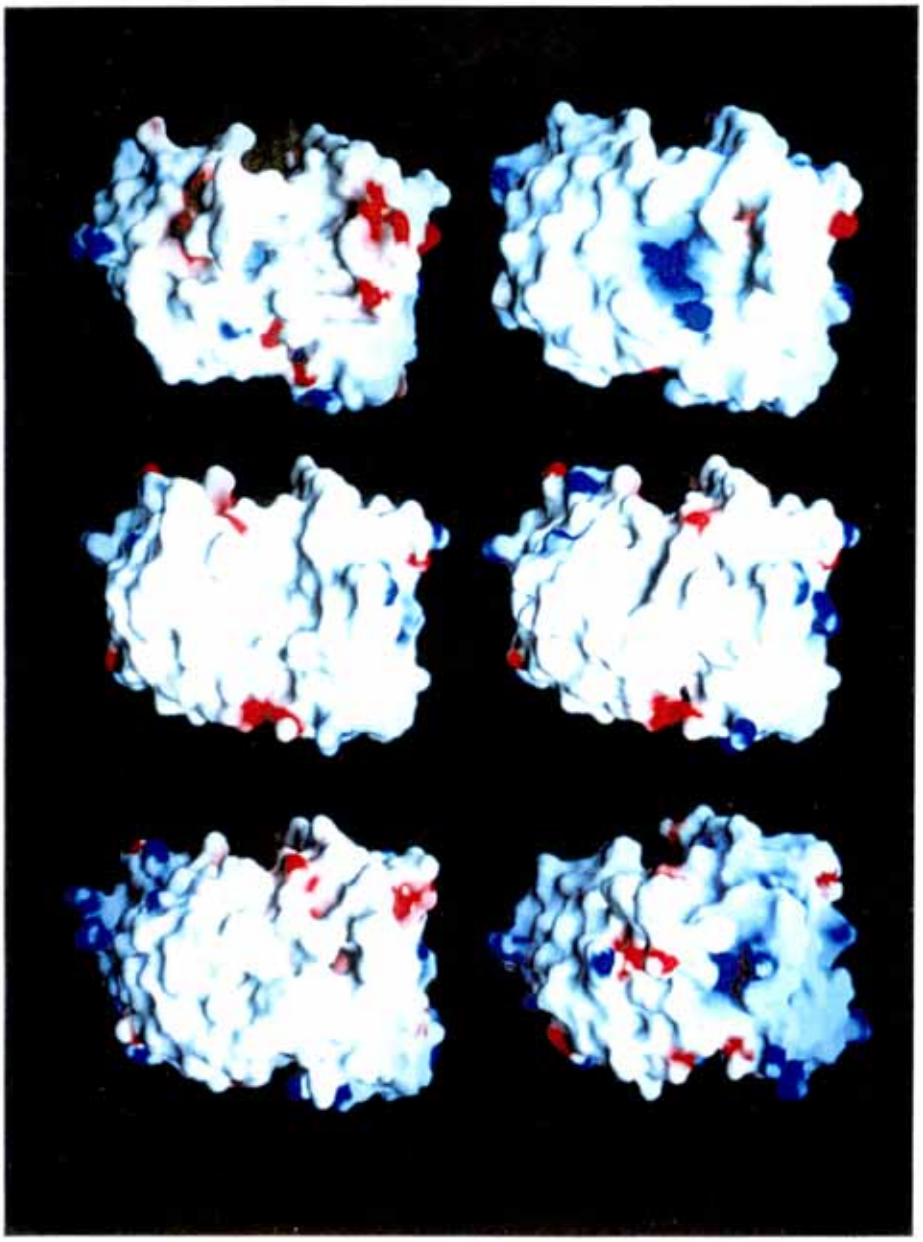


Figure 3.8: Electrostatic surface representations of the Fv homology models (end-on views). Top left: Bel; top right: Tre; middle left: Yar 1; middle right: Yar 2; bottom left: Hod and bottom right: Jak. The molecular surface and the electrostatic potential across this surface have been represented using the program GRASP (Nicholls *et al.*, 1993). White represents areas of zero potential, blue represents positive potential and red represents negative surface potential.

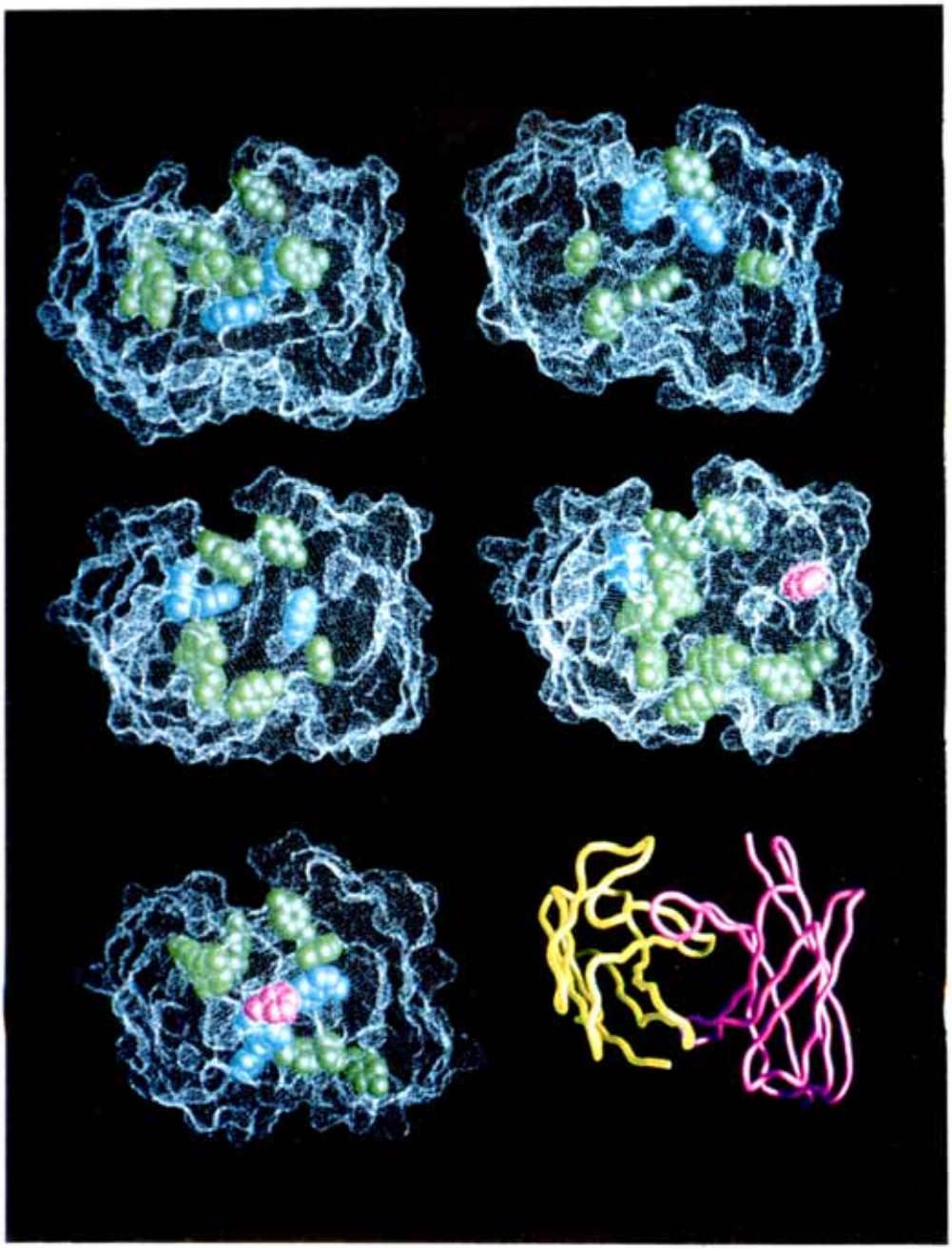


Figure 3.9: Aromatic residues potentially contacting antigen. The molecular surface has been represented (white dots) with the Fv molecules shown as end-on views. Top left: Bel; top right: Tre; middle left: Yar 2; middle right: Hod; bottom left: Jak and bottom right: Cα spline of Tre Fv showing the orientation of the Fv models (VL is yellow and VH is magenta). Aromatic side-chain atoms are displayed as spheres coloured green: tyrosines; cyan: tryptophans and magenta: phenylalanine residues. Only those aromatic residues within CDR sequences which are not buried in the interior of the protein have been displayed.

3.4: Discussion

Computational techniques for predicting antibody combining site structures have generally approached the problem by utilising the conserved Ig variable domain scaffolding to construct good approximations of antibody variable region fragments (Martin *et al.*, 1991). These models provide important information regarding molecular recognition. The homology models of polyreactive Fv molecules presented here were constructed in an attempt to identify structural features of polyreactive Ig antigen binding sites.

3.4.1: Validation of modelling strategy

The test case for the modelling protocol was to model the variable domains of the human antibody 3D6 without regard to its experimental structure (He *et al.*, 1992) during the construction of the model. Examination of this model compared to the X-ray structure (Table 3.1 and Figure 3.1) suggests some strengths and potential weaknesses of using a template-based approach for predicting the structure of other human Fv molecules. The framework regions of both V domains are well described by the model with an rmsd of 0.74 Å for VL and 0.49 Å for VH domains when the structures were aligned using all framework Cα atoms. The rmsd for the framework regions was increased to 1.0 Å when the entire Fv was aligned with the V domains of the X-ray structure of 3D6. This indicates that the VL–VH domain interface has not been predicted accurately by using the quaternary association of the Pot IgM variable domains (Fan *et al.*, 1992). The use of the Pot Fv (VL–VH) domain interface for the polyreactive Fv models standardises the quaternary associations of the V domains for these structures. Thus, any structural differences observed for the binding sites of the B CLL Fv should be the result of variation in CDR and framework conformations rather than the V domain interactions. Obviously, the Pot Fv quaternary association may result in an inadequate description of the quaternary association of the V domains, but, currently there is no preferable method to resolve this problem (Bajorath & Sheriff, 1996). Confidence is given by the 3D6 Fv model for the prediction of CDR loop conformations when starting with template loops of the same length and canonical class compared with the CDR being predicted. Although the LCDR3 was predicted using the loop stems of one CDR and capped with a β-turn taken from another structure, the rmsd was only 0.84 Å when comparing the modelled loop with the X-ray structure of LCDR3. However, a similar approach to modelling the 17 residue HCDR3 of 3D6 resulted in an rmsd of 1.3 Å when comparing the predicted and experimental conformations. Compared to alternative techniques for predicting

HCDR3, which also yield rmsd values significantly larger than one (Bajorath & Fine, 1992; Bruccoleri & Karplus, 1987), the result of using a template-based prediction of the loop stems is acceptable. Clearly a better understanding of the sequence–structure relationships determining HCDR3 conformation is required to improve the accuracy of modelling these loops in antibodies.

3.4.2: Modelling the binding sites of polyreactive immunoglobulins

The topology of immunoglobulin combining sites is predominantly determined by the primary sequence and the folded conformation of the six CDRs. The conformation adopted by CDR loops is largely dependent on the number of residues therein. Additionally, CDR conformation is affected by 'key' structure determining residues both within the loop or in neighbouring framework regions. Chothia *et al.* (1989; 1992) identified this dependence of loop conformation on length and sequence and produced an elegant theory of canonical CDR loop structures. In the Fv models described here, the templates were not initially analysed for the presence of 'key' framework residues. Only the residues assigned to CDR loops by Kabat *et al.* (1991) were considered in the choice of the templates. Subsequent analysis revealed that the CDR loop templates, chosen by this approach, for all the κ light-chain V domains (Tre, Yar, Hod and Jak) and all heavy-chain CDR1 and CDR2 structures fall into previously described canonical classes (Table 3.3). The antibody Bel expresses a λ light chain. In this case, none of the light-chain CDR loops could be assigned to a canonical class. This can be attributed to a disproportionate number of κ compared to λ light-chain structures available at high atomic resolution. At least five out of the six CDR loops within each of the human Fv homology models were predicted using template loops of the same length and in most cases the same canonical class. Previous studies have demonstrated (Barry *et al.*, 1994; Chothia *et al.*, 1992; Chothia *et al.*, 1989) that these loops are predicted accurately by homology modelling and so provide a good approximation of the basic binding site topology.

No current modelling procedure has consistently predicted the conformation of heavy chain CDR3 loops with a high level of accuracy. This is mainly due to the heavy chain CDR3 loop being the most variable with respect to length and sequence. Human HCDR3 loops have lengths varying from 2 to 26 residues (Wu *et al.*, 1993). In order to address this difficulty, the approach taken here for the prediction of HCDR3 loops has been to construct these loops using a template-based loop building strategy. A 3D database of HCDR3 from native X-ray antibody structures was compiled in order to

assess the amount of structural data available for HCDR3 and to aid in the selection of templates for these loops. Although the loops range from 5 to 17 residues, not all HCDR3 lengths are represented in the database. In some cases, only one structure is known for a given HCDR3 length. As the number of Ig crystal structures increases, the ability of an empirical approach to predict HCDR3 conformations will become increasingly more accurate. In the present study, Bel and Yar contain 16-residue HCDR3s while Tre Fv contains an 11-residue HCDR3. The HCDR3 of Hod and Jak are longer than any previously crystallised Ab at 19 and 18 residues in length respectively. The modelling strategy was developed using the Fv of Bel, Tre and Yar since for all of these Ab a HCDR3 of the same length was present in the database of uncomplexed Ig crystal structures.

The Bel Fv HCDR3 was constructed from the coordinates of a murine monoclonal antibody NC10.14 which has a well-ordered 16-residue β-loop which protrudes out from the binding pocket into the solvent (Drs L.W. Guddat and A.B. Edmundson, unpublished structure). The HCDR3 of Yar was constructed using either the conformation of this loop found in the NC10.14 Fab (Yar 1) or that of the 1IGM Fv (Yar 2). A comparison of two alternative conformations for the Yar HCDR3 suggested that the more energetically favourable conformation was for the loop to fold across the binding site, resulting in a relatively flat binding surface (Yar 2). This fold was based on the crystallographic structure for the HCDR3 of Pot Fv (1IGM; Fan *et al.*, 1992). Pot IgM is a paraprotein with cryoglobulin and polyreactive properties (Ramsland, 1993) which was originally isolated from the plasma of a patient with Waldenströms Macroglobulinaemia, a malignancy of B lineage cells (R.L. Raison, personal communication). While it is a distinct possibility that the HCDR3 of Yar could adopt a conformation somewhere intermediate between that of NC10.14 and Pot HCDR3s, the current database of HCDR3 loop folds does not provide a more energetically favourable alternative. Furthermore, it may become apparent that the combining site structure of Pot Fv (Fan *et al.*, 1992) provides one of the archetypal binding sites for polyreactive Ig, where a relatively flat binding surface is responsible for multiple Ag recognition. It will be of great interest to determine the experimental structure of Yar Fv to examine how influential the relatively long LCDR1 is on the conformation of HCDR3. The light chain variable domain of Hod IgM is extremely similar to that of Yar. Again the long LCDR1 loop determined the selection of a HCDR3 constructed using the fold of the Pot HCDR3 loop. The Hod Fv model strengthens an argument that some polyreactive antibodies have a combining site which does not possess a cavity or groove at the interface of the variable domains.

The template for the 11-residue HCDR3 of Tre was 1MCP. The 1MCP loop was selected from the three unique HCDR3 sequences of known conformation (Figure 3.2). Perhaps shorter HCDR3 loops are more similar in structure to each other than longer loops, a feature which is at least reflected by the 11-residue HCDR3 loops examined. Although there are differences between the three 11-residue HCDR loops of known conformation, these are restricted to small movements towards the top of the loops. These differences are minimal considering that the top of HCDR3 is exposed to solvent and is expected to have more dynamic properties in solution compared with other regions of the variable domain. It appears that large differences in loop conformation are not common for these shorter loops. The apparent conformational restriction of HCDR3 of 11-residues in length encourages the conclusion that the Tre HCDR3 loop is accurately predicted using the HCDR3 of 1MCP as a template. Another indication that this loop was an appropriate template for Tre HCDR3 is that the conformation of 1MCP HCDR3 is closely maintained after refinement of the Tre Fv model (Figure 3.2 and 3.7) and so this loop fitted well with the other five CDR loops. Especially in cases where several experimental structures of HCDR3 of the same length are available, a template-based approach to modelling HCDR3 should equal or surpass the predictive capabilities of methods based on systematic conformational searching or the random generation of loop structures (Bajorath & Fine, 1992; Bruccoleri & Karplus, 1987).

The modelling of the Jak HCDR3 (18 residues) was based on construction of the loop stems using the 17-residue HCDR3 of 1DFB. Chain-closure was accomplished by joining the loop stems with a β-turn; an approach which yielded an rmsd of 1.3 Å between the predicted and actual structures when the HCDR3 of 3D6 was predicted using the 16-residue loop of NC10.14. Thus, the conformation obtained was a HCDR3 which projected into solvent similar to the HCDR3 of Bel. However, unlike the Bel HCDR3, the Jak loop is twisted across the VL–VH interface and so may partially hinder access of Ag to an actual binding cavity (Figure 3.4b and 3.5b).

3.4.3: Refinement of variable domain models

Refinement of the Fv models using conjugate-gradient energy minimisation, simulated annealing and a further energy minimisation stage produced structures which have good stereochemistry and geometry comparable to that of a high quality Ab crystal structure (Table 3.5). The majority of ψ,ϕ torsion angles were in favourable regions of a Ramachandran plot (Figure 3.6 and Appendix B). One residue on the light-chain variable domains of antibodies frequently assumes an unfavourable ψ,ϕ relationship

(Bossart-Whitaker *et al.*, 1995; Guddat *et al.*, 1994) and this has been reproduced in the modelled VL domains (residue 50L in Bel and 51L (Kabat numbering) for other structures). The presence of this unusual feature of Ab light-chain V domains and the visual examination of the structure of the refined models (Figures 3.4 and 3.5) supports a conclusion that the basic structural features of the Ig fold have been maintained during model refinement. Structural differences between the template and refined models were minimal for all mainchain atoms and most side-chain atoms (Figure 3.7) reflecting the high quality of the starting template models. These differences were slightly higher within some CDRs and regions where two template structures were spliced together during the model building process. However, even after energy refinement had adjusted for such imperfectly spliced sections the structural movements in the mainchain atoms were minimal, suggesting that model refinement retained the pertinent features of antibody variable domains and did not deviate too much from the template-based starting models. Thus, the target of model refinement was fulfilled, which was to optimise side-chain positions and alleviate minimal energy strain within the protein backbone while maintaining the structure of an Fv molecule.

3.4.4: Diversity of human polyreactive immunoglobulin combining sites

One of the strongly held precepts regarding the molecular mechanisms of polyreactive antibody binding is that these immunoglobulins would have a conserved combining site structure (Cheung *et al.*, 1995; Padlan, 1994). Padlan (1994) argues in a comprehensive review of antibody structure that the polyreactive antibody combining site is relatively large, plastic or 'sticky' (high number of aromatic side-chains within CDR) in character. The present modelling contradicts the concept of a conservation of binding site structure and points towards a more diverse pattern of binding site 3D structure for polyreactive immunoglobulins. However, an argument that side-chains with aromatic character are present in high numbers within polyreactive Ab combining sites may still hold true. The polyreactive Fv homology models (Figures 3.4 and 3.5) suggest that at least two basic binding site topologies are involved in polyreactive antigen recognition: (i) a pocket-type binding site is predicted for Bel, Tre and to a lesser extent Jak antibodies; and (ii) a relatively flat binding surface is described for Yar and Hod immunoglobulins. Of the pocket-type binding sites, Tre Fv presents the largest cavity for Ag binding and Jak Fv potentially has its cavity partially blocked to large antigens by the long HCDR3 loop. Furthermore, the HCDR3 of Bel and Jak as well as the LCDR1 of Yar and Hod project out from the compact variable domains into solvent. These solvent-exposed loops are potentially buried during antigen binding.

In particular, the highly aromatic nature of the Bel HCDR3, resulting mainly from a stretch of six tyrosine residues, suggests that this loop is buried at the interface between antigen and antibody. The protrusion of LCDR1 into solvent in this manner results in a large bulge in the otherwise relatively flat binding surfaces presented by Yar and Hod immunoglobulins. Clearly there are some similarities and more obvious differences in the basic architecture of the combining sites of the human polyreactive Ig examined here.

The electrostatic surface representations of the Fv models (Figure 3.8) reveal further differences in the binding sites of the B CLL immunoglobulins. The binding site of Bel is predominantly hydrophobic in character with one region of negative potential running along the VH rim of the binding cavity. The combining surfaces of Yar and Hod are also largely hydrophobic in character. Somewhat similar to Bel Ig a few negative charges are located towards the VH side of the binding surface of Hod and Yar Fv molecules. The binding pocket of Tre immunoglobulin has a strongly positive electrostatic potential. The combining site of Jak Ig has substantial regions where hydrophobic residues are forming the molecular surface. The partially occluded binding cavity of Jak Ig is similar to the large cavity of Tre in that it has a strong positive electrostatic potential. Finally, with respect to the electrostatics, an examination of the two Yar models (Figure 3.8) shows that the electrostatic nature of the binding site is relatively similar even though the HCDR3 is shifted dramatically from a loop which projects upwards from the V domains (Yar 1) to one which folds across the VL–VH domain interface (Yar 2). This indicates that an electrostatic surface representation is a robust technique for investigating the binding sites of antibodies even in cases where the confidence of prediction of the HCDR3 may be low.

Residues with aromatic side-chains have been observed with higher than expected frequencies within CDR sequences when compared to framework regions of antibody variable domains (Padlan, 1990). The contribution of these aromatic side-chains to Ag binding has been observed in X-ray structures of Fab molecules complexed to hen egg lysozyme (Amit *et al.*, 1986; Padlan *et al.*, 1989). Tyrosine side-chains participate in Ag binding through hydrogen bonding and a significant van der Waal's contribution. Furthermore, compared to other proteins, the side-chains of tyrosine and tryptophan residues are more frequently exposed to solvent within CDR rather than being buried in the interior of the protein (Padlan, 1994). To examine the potential contribution to Ag binding of aromatic residues within the binding sites of polyreactive Ig the aromatic groups within CDR which were partially contributing to the molecular surface were displayed (Figure 3.9). As with monoreactive Ig, the polyreactive Fv contain a high number of solvent-exposed side-chains within CDRs which are aromatic in character,

particularly tyrosines. Interestingly, these residues appear to be distributed over a substantial region of the combining sites of the B CLL immunoglobulins. This observation supports Padlan's (1994) argument that polyreactive Ab have 'sticky' binding sites. Polyreactive IgM expressed by human B CLL cells display a highly conserved binding to mouse immunoglobulins (Weston & Raison, 1991). The binding of mouse IgG1/κ was demonstrated for the five B CLL tumours utilised in this modelling study (Chapter Two). Perhaps this binding phenomenon is conferred by the high numbers of aromatic side-chains which contribute to the combining site surface of the B CLL Fv molecules. Furthermore, if aromatic residues hold the key to understanding the polyreactive phenomenon, the antigens themselves may contain structural similarity based upon the clustering of aromatic groups on their surfaces.

3.4.5: Conclusion

Elucidation of the structural basis of polyreactive antigen binding will require analysis and comparison of the binding sites of numerous polyreactive Igs. While X-ray crystallography is arguably the definitive method for achieving this goal, it is time consuming, costly and dependent on the ability to obtain suitable protein crystals. Thus, the development of improved antibody modelling techniques will make a significant contribution towards the solution of this problem. The structures of five variable region fragments, of polyreactive Ig expressed by human B CLL cells, have been predicted using a homology-based modelling approach. Comparison with the X-ray structures of these same Fv molecules would result in the refinement of the modelling process. This will require the VL and VH domains to be expressed in bacteria in a form which enables the purification of large quantities of Fv which can then be utilised in crystallisation experiments. The examination of the predicted structures of the Fv molecules has revealed the structural diversity of polyreactive binding sites is significant. The basic differences in topology and surface electrostatics of the Fv is potentially compensated for by a large number of solvent-exposed side-chains with an aromatic nature localised to the combining sites of these human polyreactive immunoglobulins. The expression of two of the B CLL derived Fv was attempted and the results are presented in the next chapter (Chapter Four). The successful purification of large quantities of a few human polyreactive Ig fragments may result in the first experimental structures to be determined for these interesting immunoglobulins.

CHAPTER FOUR
Cloning and bacterial expression of B CLL derived immunoglobulins as Fv molecules

4.1: Introduction

Understanding the structural basis of Ig polyreactivity will require the solution of the X-ray structures of at least one such Ab fragment in complex with several different antigens. The cloning for bacterial expression of two Fv portions derived from B CLLs expressing polyreactive surface IgMs is described here. The decision to use a non-covalently linked Fv was based upon the fact that an Fv portion is the smallest fragment of an Ig which associates in a manner similar to its quaternary structure in an intact Ig and still contains the Ab combining site.

Strategies for the *in vitro* production of antibodies and antibody fragments have utilised recombinant DNA technology to exploit the protein synthesising machinery of bacterial, yeast, insect and mammalian cells (reviewed in: Morrison, 1992). The Ab genes are introduced into the cells within a DNA vector which usually contains a strongly selectable genetic marker and the elements required for protein expression. Protein expression can be achieved through harnessing the host cells promoter and transcriptional elements, a process requiring chromosomal integration of the vector DNA. More commonly the vector DNA contains all the gene elements required for replication and mRNA transcription; such as a strongly inducible and regulated promoter/repressor system. Antibody fragments are frequently produced in bacterial systems due to the ease of DNA manipulations and the high protein yields obtainable (reviewed in: Skerra, 1993). The antibody fragments commonly expressed include Fab, Fv and single chain Fv (scFv) molecules. Through judicious design of the expression vectors, containing the Ab genes, the protein can be produced as insoluble cytoplasmic inclusion bodies or as soluble protein within the periplasm of Gram-negative bacteria (Wulfing & Pluckthun, 1994; Pluckthun & Skerra, 1989; Skerra *et al.*, 1991).

4.1.1: Architecture of bacterial expression vectors

Bacterial expression vectors are plasmids which contain a functional operon for foreign gene expression. The vector DNA contains an origin of replication, a strong

genetic selectable marker such as the ampicilin resistance gene, a promoter/repressor system and a cloning site which enables the foreign gene to be introduced within the operon so that it is in frame with an upstream ribosome binding site (RBS) (reviewed in: Pluckthun & Skerra, 1989). Expression of the VL and VH domains as functional Fv molecules requires the operon to contain two functional cistrons. Such dicistronic expression vectors contain two RBSs followed by cloning sites where the VL and VH genes can be introduced, with or without a secretion peptide, so that both genes are under control of a single inducible promoter (Skerra *et al.*, 1991).

4.1.2: Denaturation and refolding of insoluble antibody fragments

Immunoglobulin variable domains may be expressed in bacterial systems in the absence of a secretion peptide. In this system the majority of the polypeptides produced aggregate in the cytoplasm as bacterial inclusion bodies (Buchner *et al.*, 1992). The cytoplasm of bacteria does not favour the folding of the nascent polypeptide and disulfide bonds are not formed due to the reducing environment of the cytoplasm (Wulfing & Pluckthun, 1994). Thus, V domains which are expressed as bacterial inclusion bodies require purification and the denaturation and refolding of the proteins *in vitro* in order to obtain correctly folded protein. Isolated inclusion bodies are subjected to strong denaturants such as 6 M guanidine and reducing agents like 0.3 M dithioerythritol to fully denature and reduce the variable domains (Tsunenaga *et al.*, 1987; Goto & Hamaguchi, 1982). Refolding buffers frequently contain L-arginine which tends to minimise aggregation of unfolded polypeptides and low molecular weight thiol redox reagents such as GSH and GSSG which favour correct disulfide bond formation (Buchner *et al.*, 1992). Increased yields of properly refolded Ab fragments have been achieved through the addition of chaperones and protein disulfide isomerase (PDI) which catalyses disulfide bond formation (Lilie *et al.*, 1994). However, the recovery of correctly refolded Ab fragments requires further purification and frequently the yield of functional Fv molecules is low and is highly dependent on the affinity of the VL–VH domain interaction (Buchner *et al.*, 1992; Glockshuber *et al.*, 1990).

4.1.3: Periplasmic expression of soluble antibody fragments

The periplasm is a bacterial compartment which is located between the inner and the outer membranes of Gram-negative bacteria and can constitute around thirty percent of

the total cell volume. The contents of the periplasm are predominantly proteins, peptidoglycans and bound water. In comparison to the cytoplasm, the periplasm is considered to be a non-reducing environment which facilitates the formation of disulfide bonds (Wulfing & Pluckthun, 1994). Furthermore, this compartment contains proteins involved in protein folding and the disulfide isomerases such as DsbA (Knappik *et al.*, 1993). Proteins are directed to the periplasm by specific secretion peptides which mediate transport across the inner membrane and are subsequently cleaved to release the mature polypeptide. Antibody fragments have been successfully expressed as soluble protein in the periplasmic space by fusing the secretion peptides such as OmpA, PelB and PhoA to the amino-terminus of the variable domains (Sawyer *et al.*, 1994; Pluckthun & Skerra, 1989). Soluble protein expression within the periplasm varies significantly with different antibody fragments. Addition of the secretion peptide PelB to the NC10 single chain Fv did not result in high yield of correctly folded protein within the periplasm. Sufficient quantities of correctly folded protein was only obtained after the NC10 scFv was denatured with 6 M guanidine hydrochloride and refolded from the periplasmic preparation (Kortt *et al.*, 1994). Although not fully understood, there appears to be a large dependence on the primary amino acid sequence of the V domains with respect to the ability to express them as soluble and functional protein within the periplasm. Kipriyanov *et al.*, (1997) demonstrated that the change of a framework residue at position six of VH from a glutamic acid to glutamine resulted in an astounding thirty-fold increase in soluble scFv production. While this substitution may not act universally in the enhancement of soluble Ab fragment expression, it highlights the severe dependence of soluble periplasmic expression of foreign proteins in bacteria on a protein's primary amino acid sequence.

4.1.4: Structural studies using bacterially expressed antibody fragments

Bacterially expressed Ab fragments have similar or identical immunological properties when compared to the analogous regions of the parent murine or human immunoglobulins (Bhat *et al.*, 1990). Structural studies are revealing that bacterially derived Ab fragments are also indistinguishable from the parent Ab fragments at the level of tertiary and quaternary structure. Comparison of the structure of an enzymatically derived Fab variable regions with the bacterially expressed Fv of the hen egg lysozome antibody D1.3 demonstrated that the two structures were highly similar in 3D structure. Superimposition of the variable regions of the two D1.3 structures

gave an rmsd of only 0.39 Å for all Cα atoms. Furthermore, the contacts made with HEL by the mouse Fab and the bacterially expressed Fv of D1.3 were highly similar, indicative of the maintenance of structure and function of Ab fragments when expressed in bacteria (Bhat *et al.*, 1990). Artificial Ab fragments, such as scFv molecules, contain an oligopeptide linker which tethers the carboxyl terminus of VH to the amino terminus of VL. A very similar association of VL–VH domains occurs in scFv molecules as found for Fv or Fab molecules. Examination of the X-ray structure of the fifteen-residue scFv of NC10 in complex with neuraminidase demonstrated that the VL–VH domain interface on individually refined Fv molecules were very similar to that of an ezymatically derived Fab complexed to Ag (Kortt *et al.*, 1994; Colman *et al.*, 1987). The two complexes within the unit cell were arranged back to back with the possibility that the Fv were associated as a dimer with the VL of one scFv associated with the VH on a second scFv molecule. Since the linker sequence could not be located within the electron density it was impossible to determine if the NC10 was present as a monomer or a dimer of scFv molecules (Kortt *et al.*, 1994; Malby *et al.*, 1993). However, this complex of NC10 scFv with neuraminidase demonstrated that a bacterially expressed scFv was capable of both associating as Fv molecules and binding Ag in a manner which is structurally very similar to the Fab of the parent mouse Ab.

The successful purification of Ab fragments for X-ray diffraction studies is dependent on obtaining high yields of protein purified to homogeneity and on the ability to grow diffraction quality crystals of that protein. The structures of the bacterially expressed D1.3 Fv (native and mutants) required the purification of the Fv using affinity chromatography with HEL as the ligand (Eisele *et al.*, 1992). The X-ray structure of the D1.3 Fv bound to HEL generally has lower temperature factors (B) for Fv residues when compared to the uncomplexed Fv. This suggests that the interaction with HEL results in an overall conformational stabilisation of the Fv, including a stabilisation of the VL–VH domain interaction (Bhat *et al.*, 1994). Affinity purification using an Ag column was used to obtain large quantities of a scFv specific for a trisaccharide. The crystals of the complex of this scFv with Ag diffracted to a remarkable 1.7 Å resolution allowing for a high quality structure to be solved (Zdanov *et al.*, 1994). However, diffraction quality crystals of the scFv of NC10 were obtained from protein purified by denaturation and refolding followed by size-exclusion and anion exchange chromatographic methods (Malby *et al.*, 1993). The harsher treatment of the NC10 scFv is potentially not transferable to a non-covalently associated Fv since in the scFv of NC10 the VL and VH were joined by a $(Gly_4Ser)_3$ linker which guaranteed that an equimolar amount of the two variable domains was present in the purified protein.

For this study, two B CLL derived Fv were cloned into the bacterial expression vector pFLAG–CTS as dicistronic constructs which contain the periplasmic secretion peptides OmpA and PelB fused to VL and VH region genes respectively. The expression of these Fv (Bel and Tre) was optimised and affinity purification was attempted using columns specific for the VH (FLAG octapeptide fusion) or VLκ domains (for Tre Fv). The purification did not yield Fv fragments in a form suitable for X-ray crystallography and so the possible use of a scFv product is discussed. Determination of the X-ray structures of the B CLL derived Fv or scFv in complex with several Ag may provide an understanding of the molecular basis of immunoglobulin polyreactivity.

4.2: Materials and Methods

4.2.1: General reagents

All reagents were of analytical or molecular biology grade and the water (MilliQ™) was purified using a Modulab™ system. General molecular biology reagents, restriction enzymes and sequencing vectors have been detailed (Section 2.2.1). Automated sequencing reagents; Sequitherm long read dideoxy-termination kits and M13-IRD41 (forward and reverse) labelled primers, were purchased from Epicentre technologies (Madison, WI, USA). The bacterial expression vector pFLAG–CTS, M2 monoclonal antibody and M2-Sepharose affinity matrix (Kodak, New Haven, CT, USA) were purchased from Integrated Sciences (Sydney, NSW, Australia). The vector pHUD was generously provided by Dr Peter J. Hudson (Biomolecular Research Institute, Parkville, Australia). KappaLock™ affinity matrix was purchased from Zymed (San Francisco, CA, USA). Anti-mouse Ig (H & L chains) affinity purified polyclonal sheep antibody conjugated to alkaline phosphatase was purchased from Promega (Madison, WI, USA). Nitrocellulose (Biorad, Hercules, CA, USA), YM10 (Amicon, Beverly, MA, USA), benzolated-cellulose acetate 2 kDa cutoff and cellulose acetate 10 kDa cuttoff dialysis (Sigma, St Louis, Mo, USA) membranes were purchased from the respective companies. ImmobilonP^{SQ} PVDF membrane (Millipore, Bedford, MA, USA) was used for N-terminal protein sequencing (kindly carried out by Dr Denis Shaw, Australian National University, Canberra, ACT, Australia).

4.2.2: Oligonucleotide primers

Oligonucleotide primers were synthesised and purified using HPLC by Beckman (Sydney, NSW, Australia). Oligonucleotides for amplification of VL region genes were designed to bind to the 5' of the VL (Bel1 and Tre1) and 3' of JL (Bel2 and Tre2) genes with *HindIII* and *XhoI* restriction sites for forward and reverse primers respectively. Oligonucleotides for amplification of VH region genes were designed to bind to the 5' of VH (VHUD) and 3' of JH (Bel4 and Tre4) genes with *NcoI* and *SalI* restriction sites for forward and reverse primers respectively. The sequences of the primers are shown:

Bel1	5'–TAACCCAAGCTTCCTATGTGCTGACTCAGCCACC-3'
Bel2	5'–ATTCCGCTCGAGTCATCATAGGACGGTCAGCTTGGTCC-3'

VHUD[a]	5'–TGCATGCCATGGCCGAGGTGCAGCTGCAGGAGTCTGG-3'
Bel4	5'–GCCGACGTCGACTGAGGACACGGTGACCGTGGTCC-3'
Tre1	5'–TAACCCAAGCTTACATCCAGATGACCCAGTCTCC-3'
Tre2	5'–ATTCCGCTCGAGTCATCATTTGATCTCCAGCTTGGTCC-3'
Tre4	5'–GCCGACGTCGACTGAGGAGACGGTGACCAGGG-3'

[a] The oligonucleotide, VHUD, is the VH forward primer (ie. Bel3 and Tre3) for ligation into the pHUD expression vector.

4.2.3: Polymerase chain reaction

Variable region genes were amplified from 0.1 ng of vector DNA (Section 2.2.7) for VL region genes and 1-10 ng of cDNA (Section 2.2.4) for VH region genes. Polymerase chain reaction consisted of 1 cycle of denaturation at 95°C for 90 s followed by 30 amplification cycles: 30 s at 69°C, 30 s at 72°C and 30 s at 95°C. A final cycle of 30 s at 69°C and 5 min at 72°C was carried out (Omnigene thermocycler, Hybaid). Typical 50 µL reactions contained 1xPCR reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 5 pmol each primer (Section 4.2.2) and 2.5 units Taq DNA polymerase (Perkin Elmer, Branchburg, NJ, USA). Amplification products were analysed by 10% (v/v) PAGE using 5 µL of each reaction.

4.2.4: Automated DNA sequencing

Plasmid DNA or PCR product was sequenced using a Li-Cor 4000L automated DNA sequencer system (Li-Cor, Lincoln, NE, USA). Sequencing reactions were performed using a modification of the SequiTherm™ Long Read kit protocol (Epicentre). For plasmid DNA (pGem or pFLAG) 0.2 pmol was used per sequencing reaction in the presence of 1x sequencing buffer, 2.0 pmol IRD41 labelled primer (M13f or M13r for pGem and N26 or C24 for pFLAG), 5 units Sequitherm™ thermostable DNA polymerase in a total of 17 µL sterile water. Constructs in the pHUD plasmid were sequenced using a PCR method which incorporated M13 primer sites into vector specific primers. The PCR product was then purified using the Advantage™ PCRpure kit (Clontech, Palo Alto, CA, USA) and 0.1-0.2 pmol of this DNA was sequenced, as described for plasmid DNA, using M13–IRD41 labelled primers (Joshua Moses and P.A.R., unpublished method). The DNA/primer/enzyme mix (4 µL) was mixed with 2.0 µL of each termination mix in a 96 well (well volume of 0.2 mL) polycarbonate plate (Hybaid). Thermocycling (Omnigene thermocycler,

Hybaid) consisted of 1 cycle of 95°C for 5 min followed by 30 amplification cycles: 95°C for 30 s, 60°C for 30 s and 70°C for 60 s with the reactions stopped by the addition of 4 μL of SequiTherm™ stop solution. Samples were denatured at 95°C for 3 min and chilled on ice prior to loading 1.2 μL on a 4% (w/v) LongRanger™ polyacrylamide gel containing 1x LR-TBE (1.34 M tris, 45 mM boric acid, 25 mM EDTA pH 8.0) and 7 M urea in a total volume of 75 mL sterile water. Electrophoresis conditions were at 2000 V, 45 W and 25 mA at 45°C. Sequence analysed using the program BaseImager™ v2.30 (Li-Cor).

4.2.5: Cloning of VL and VH genes and construction of a dicistronic operon in pFLAG–CTS

The VL and VH genes were amplified and cloned for Bel and Tre in the form of a dicistronic operon in pFLAG–CTS as shown by Figure 4.1. The light chain V region genes were amplified from pBluescript or pGem vectors containing the B CLL VL region genes (Chapter Two) by PCR (Bel1 and Bel2 or Tre1 and Tre2 primers), digested with *HindIII* and *XhoI* restriction enzymes and purified from 2.0% (w/v) low-melting point agarose gels (PCRpure kit, Clontech). The vector pFLAG-CTS was prepared for ligation by digesting with *HindIII* and *XhoI* restriction enzymes and the cut vector was purified from a 0.8% (w/v) low-melting point agarose gel (PCRpure kit, Clontech). The VL region gene was ligated using 200 units DNA T4 ligase (NEB) into the pFLAG-CTS vector and electrotransfected into electrocompetent DH5α *E.coli* cells. The pFLAG-CTS (VL) constructs were sequenced with the N26 and C24 IRD41 labelled primers, Li-Cor (Section 4.2.4). The VH region genes were isolated using PCR from cDNA prepared from the B CLL cells (Section 2.2.4) and cloned into T-tailed pGem (Promega). Plasmid DNA was prepared by alkaline lysis (Sambrook *et al.*, 1989). Multiple independent clones were sequenced using M13f and M13r IRD41 labelled primers (Li-Cor). The pGem constructs containing the VH region genes were used as template for PCR amplification of the VH genes for ligation into the vector pHUD into the *NcoI* and *SalI* restriction sites following a similar procedure for the ligation of the VL region genes. The pHUD (VH) constructs were sequenced directly from bacterial colonies using oligonucleotides (containing M13 primer binding sites) specific for the pHUD vector backbone (Joshua Moses and P.A.R., unpublished method). Constructs which contained the correct sequence for VH region genes and were suitably ligated into the pHUD vector were isolated from bacterial cultures using Wizard plasmid preps (Promega).

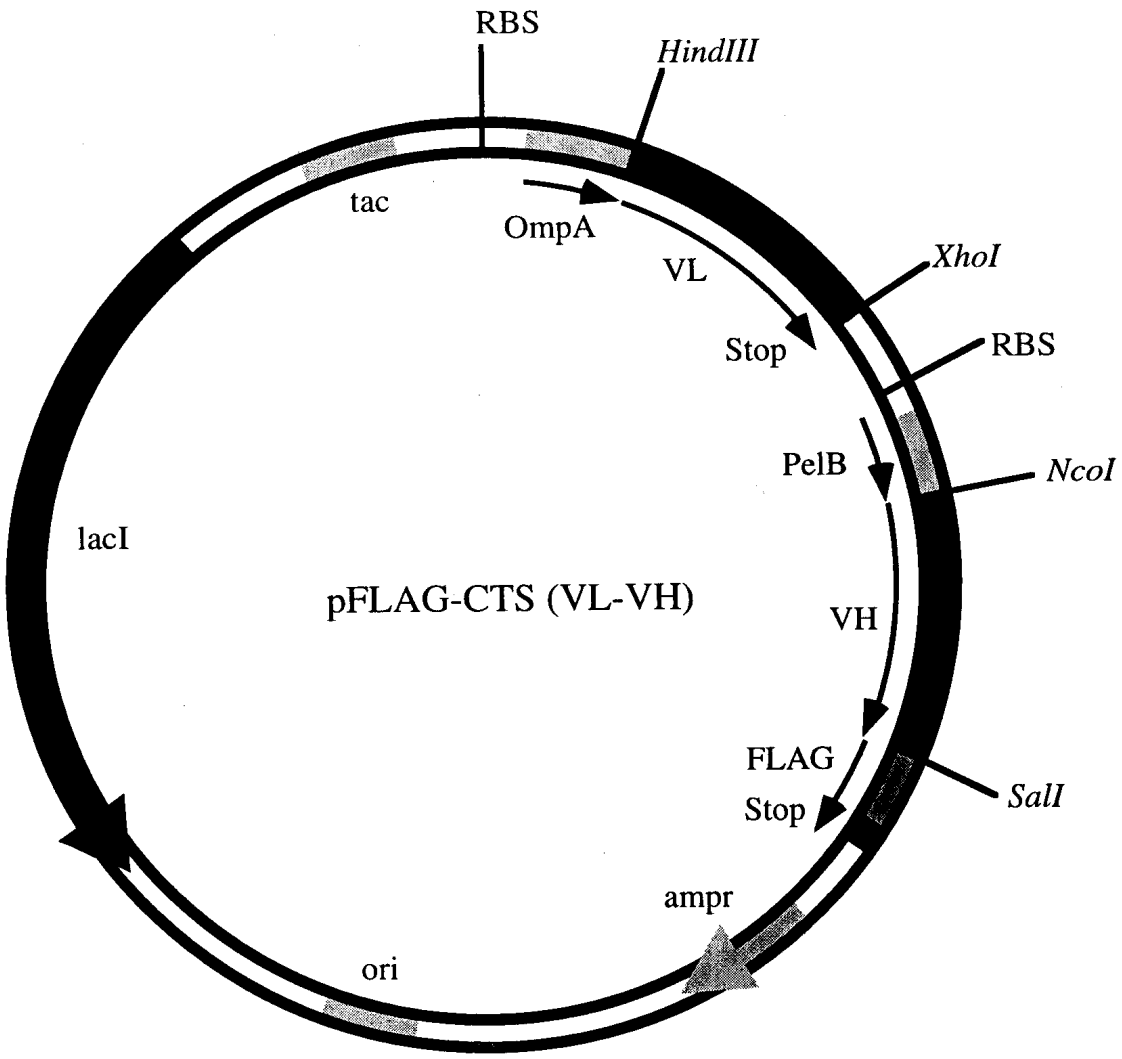


Figure 4.1: Dicistronic expression vector for soluble Fv expression. Bacterial secretion signals OmpA and PelB are fused to VL and VH region genes respectively. Production of Fv is under control of the IPTG inducible promoter Tac. The FLAG octapeptide is fused to the carboxyl-terminus of VH. Relevant restriction sites and general elements of the pFLAG-CTS (VL–VH) construct are indicated on the diagrammatic representation.

Construction of the dicistronic operon was achieved by a heteroligation of the two V region gene containing vectors (pFLAG-CTS (VL) and pHUD (VH)) and the amplification of the VL–VH region genes using PCR. The VL–VH recombinant gene was digested with *HindIII* and *SalI* restriction enzymes (NEB) and ligated into pFLAG–CTS to obtain constructs which contained a dicistronic operon which was suitable for the simultaneous expression of VL and VH genes under control of a single IPTG inducible Tac promoter element. The reaction conditions for the construction of the Bel and Tre dicistronic vectors pFLAG-CTS (VL–VH) are detailed in the results section (4.3.2). Procedures for the rapid construction of B CLL Fv expression systems will facilitate structural studies on the polyreactive Ig expressed by these malignant cells.

4.2.6: PCR site-directed mutagenesis

The first codon of the Tre VL in the dicistronic pFLAG–CTS (VL–VH) construct was mutagenised from TAC (Y) to GAC (D) using the Quickchange™ mutagenesis kit (Stratagene, La Jolla, CA, USA). The mutagenic oligonucleotides were designed to bind to the sense and anti-sense strands with 14 bases either side of the mutagenesis site, in brackets:

1TD; 5'–CCGTTGCGCAAGCT(G)ACATCCAGATGACC–'3

2TD; 5'–GGTCATCTGGATGT(C)AGCTTGCGCAACGG–'3

For the mutagenic PCR 50 ng of vector DNA (Tre VL-VH in pFLAG-CTS) and 125 ng of each primer was used. Thermocycling consisted of 1 cycle of 94°C for 90 s followed by 12 amplification cycles: 60 s at 60°C, 14 min at 72°C and 30 s at 95°C after which the reactions were cooled to 4°C. A 50 µL reaction contained 1x reaction buffer, 1 µL dNTP mix, 2.5 units PFU DNA polymerase. The methylated parental plasmid was digested for 2 h at 37°C with 10 units of *DpnI* restriction enzyme. The reaction was extracted with equal volumes of phenol (pH 8) and CIAA (24:1), precipitated with 2.5 volumes absolute ethanol (-20°C) and washed once with 70% ethanol. DNA was resuspended in 5 µL of sterile MilliQ™ water and electrotransformed into DH5α *E.coli*. Mutagenised plasmids were selected with *HindIII*, with uncut plasmids representing the mutant vectors (the mutagenesis of TAC to GAC destroyed a unique *HindIII* site).

4.2.7: Analysis of expressed proteins

Bacterial fractions from expression cultures were analysed using the SDS-PAGE method of Laemmli (1970) and Western blot immunoassay using the anti-FLAG M2 antibody (Kodak). Whole cells, osmotic shock periplasmic fractions and supernatants were prepared following the NEB expression protocols (NEB, Beverly, MA, USA). Samples were prepared in 1x loading buffer (0.01% (w/v) bromophenol blue, 0.1% (w/v) SDS, 20% (v/v) glycerol, 0.1 M Tris-HCl pH 6.8) and boiled for 5 min prior to electrophoresis using a MiniProtean II gel system (Biorad) through 4% (w/v) polyacrylamide stacking (0.1 M Tris-HCl pH 6.8) and 15% (w/v) polyacrylamide resolving (1 M Tris-HCl pH 8.8) gels in the presence of 0.1% sodium dodecyl sulfate (SDS). Polyacrylamide gels were either stained for proteins (Coomassie brilliant blue or silver staining methods) or the proteins were transferred to nitrocellulose membrane (electrotransfer system, Biorad) for Western blot immunoassay.

For Western blots the nitrocellulose membranes were blocked with 3% (w/v) BSA in TBS/Az overnight at 4°C. Membranes were probed with a 1:15000 dilution of the M2 anti-FLAG monoclonal Ab for 1 h at 21°C. The secondary Ab (anti-mouse Ig alkaline phosphatase conjugated sheep polyclonal Ab, Promega) was prepared at a 1:15000 dilution and incubated with the NC membranes for 1 h at 21°C. The membranes were developed with 1.65 µg BCIP (bromochloroindoyl-β-D-galactopyranoside) and 3.3 µg NBT (nitroblue tetrazolium) substrates prepared in 10 mL alkaline phosphatase buffer (100 mM NaCl, 5 mM $MgCl_2$, 100 mM diethanolamine pH 9.5) for 10 to 20 min. Between each step the membranes were washed 3x for 5 min with TBS/Az containing 0.05% (v/v) Tween 20 (polyoxyethylene-20-sorbitan monolaurate) and once with TBS/Az before the addition of substrate.

4.2.8: Optimisation of protein expression

Protein expression from the pFLAG-CTS (VL–VH) dicistronic constructs was induced with IPTG and monitored by Western blot immunoassay for the VH-FLAG fusion protein with the M2 antibody (Section 4.2.7). Single bacterial colonies were selected from *E.coli* strains which were electrotransfected with purified plasmid DNA for Bel or Tre Fv vector constructs. Bacterial cultures were grown overnight at 37°C (230 rpm) in 2xYT broth (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, pH 7.0) containing 50 µg/mL of ampicilin. Starter cultures were prepared in Superbroth (containing: trace elements; biotin; 0.4% (w/v) glucose and ampicilin) were inoculated with 1:100 dilution of an overnight culture and were grown to an optical density (600 nm) of 0.8 to 1.0, chilled on ice for 20 min and centrifuged at 1500g for 15 min. The

bacterial pellet was resuspended in twice the volume of Superbroth (tryptone 20 g/L, yeast extract 10 g/L, NaCl 5 g/L, K_2HPO_4 2.5g/L and MgSO4.7H2O 1.0 g/L, trace elements, biotin and ampicilin) and incubated at 37°C (230 rpm) for 1 h before induction of protein expression by the addition of IPTG. During protein expression the bacterial cultures were incubated at 25°C. Conditions for Fv expression were optimised in small-scale expression cultures (5 or 10 mL) by varying the induction times, IPTG concentration and bacterial host strains.

4.2.9: Affinity and size-exclusion chromatography of bacterially expressed Fv

Proteins expressed in the periplasmic fraction or the culture supernatant were subjected to affinity chromatography using the M2 anti-FLAG monoclonal linked to Sepharose4B (Kodak) according to the manufacturers instructions. Briefly, 10 mL of M2 matrix was equilibrated with 50 mL of TBS. Generally, 30 mL of concentrated (YM10, amicon) periplasmic fraction or culture supernatant was passed over the matrix and the column was washed with 100 mL TBS prior to eluting the bound protein with six 2.5 mL fractions of 0.1 M glycine (pH 3.0) which was neutralised immediately (200 µL of 1 M Tris-amine, pH 8.0). The eluted proteins were analysed by SDS-PAGE and Western blot with the M2 antibody (Section 4.2.7). Eluted protein was dialysed for at least 24 h against large volumes of TBS containing 0.02% (w/v) sodium azide. Protein expressed in the supernatant of Tre Fv expression was purified using the KappaLock™ affinity matrix according to the manufacturers instructions (Zymed). The eluted protein was dialysed for at least 24 h against TBS/Az.

The periplasmic fractions and supernatants of 10 mL expression cultures were analysed using size-exclusion FPLC. A Superdex 75 column, MWCO 100 kDa, bed volume 2.4 mL (Pharmacia, Uppsala, Sweden) was attached to a SMART HPLC and equilibrated with 5 bed volumes of 0.1 M ammonium acetate. Samples (50 µL) of periplasmic fractions or supernatants were loaded and chromatographed at a flow rate of 40 µL/min for a run time of 1 h with 100 µL fractions collected. The fractions were analysed by SDS-PAGE and western bloting to detect where the VH-FLAG fusion protein was eluting. Molecular weights corresponding to retention times were determined from a curve fitted to the elution profiles of the following proteins: aldolase (158 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), chymotrypsin (25 kDa) and lysozyme (14.4 kDa) (Pharmacia).

4.3: Results

4.3.1: Cloning of light and heavy chain variable region genes derived from B CLL cells into bacterial expression vectors

The genes for the VL and VH regions of Bel and Tre were isolated and cloned into a dicistronic operon constructed in the IPTG inducible bacterial expression vector pFLAG-CTS in the form shown in Figure 4.1. The light chain V region genes were amplified from 0.1 ng of recombinant pBluescript–VL constructs (Chapter Two) using 5 pmol of each primer of the Bel1/Bel2 and Tre1/Tre2 primer sets. The PCR product was digested for 1 h at 37°C with 1 unit/μg DNA of *HindIII* and *XhoI* restriction endonucleases. The VL PCR product was resolved on 2% (w/v) low melting point agarose and purified from the gel using a silica matrix (PCRpure, Clontech). This approach yielded DNA which was subsequently ligated into pFLAG-CTS *HindIII/XhoI* digested and agarose gel purified vector. The heavy chain V region genes were amplified from 1 ng of cDNA, obtained from the B CLL cells, using 5 pmol of each primer of the VHUD/Bel4 and VHUD/Tre4 primer sets. The VH region gene of Yar B CLL was also amplified using similar primers (Joshua Moses, unpublished data). PCR products were either cloned into T-tailed pGem for sequencing or digested for 1 h with 1 unit/μg DNA of *NcoI* and *SalI* restriction endonucleases, purified from agarose gels and ligated into the *NcoI/SalI* restriction digested and gel purified pHUD expression vector. The VH genes were sequenced using a procedure developed for direct automated sequencing of any template using PCR products obtained from bacterial colonies using pHUD vector specific primers containing M13 primer binding sequences (see Section 4.2.4; Joshua Moses and P.A.R., manuscript in preparation). A total of 8, 8 and 6 independent clones of Bel, Tre and Yar VH region genes were sequenced in both directions to confirm the homogeneity of the VH region genes isolated using PCR from the patients PBLs (Section 2.4.1). Schematic diagrams of the expression vectors containing VL and VH region genes and representative electrophoretic gels showing the cloned the genes are shown (Figure 4.2).

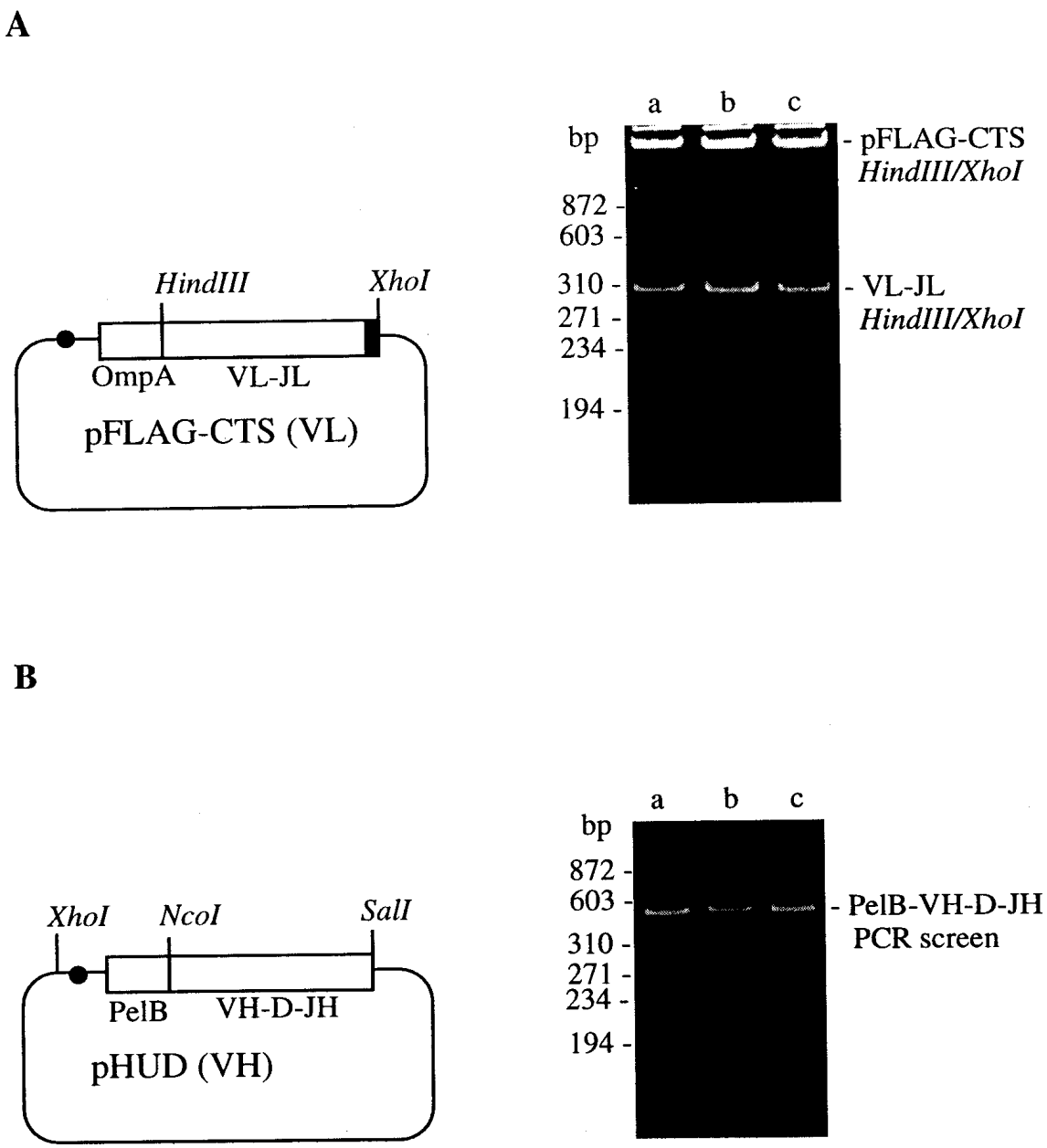


Figure 4.2: Cloning of the VL and VH region genes of B CLL immunoglobulins into bacterial expression vectors. A: schematic diagram of the VL constructs in pFLAG-CTS and electrophoretic gel showing representative constructs (Bel) cut with *HindIII* and *XhoI*. B: schematic diagram of the VH constructs in pHUD and electrophoretic gel showing representative PCR screen (Bel) with pHUD vector primers. The schematic vectors are labelled with RBSs shown as circles and stop codons as solid boxes.

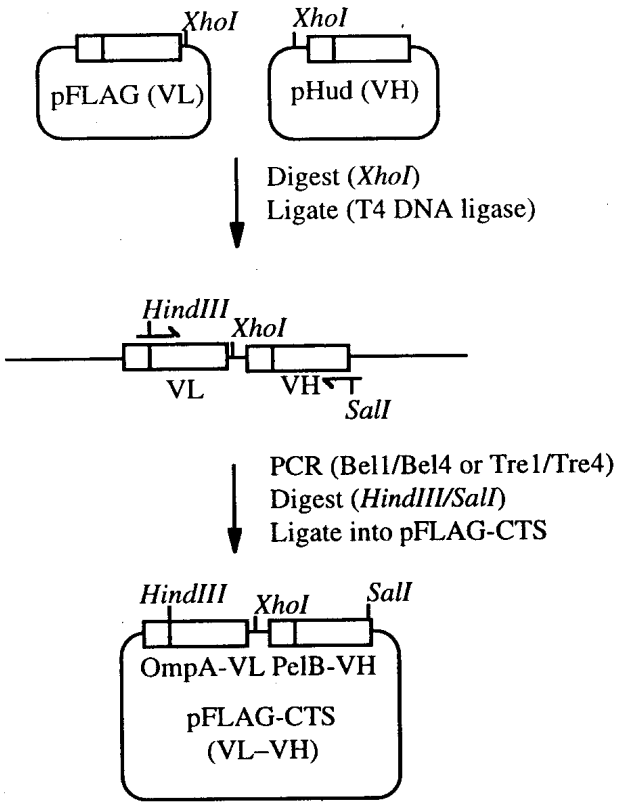
4.3.2: Construction of a dicistronic operon for the soluble expression of VL and VH domains

The cistron containing pelB fused to VH in pHUD was inserted into the *XhoI* and *SalI* restriction enzyme sites of the pFLAG–CTS vector (Figure 4.1). However, the cohesive ends of *XhoI* and *SalI* digested DNA are compatible and religate to form a unique restriction site recognised by the *TaqI* restriction enzyme. A PCR-based method to rapidly construct a dicistronic operon for VL and VH genes was developed to overcome the reduced efficiency of ligating DNA into *XhoI* and *SalI* restriction sites (Figure 4.3). The final conditions involved the incubation of equal quantities of pFLAG–CTS (VL) and pHUD (VH) plasmids for 1 h at 37°C in the presence of 2 units of *XhoI* restriction enzyme (NEB). The restriction enzyme was inactivated by heating to 65°C for 15 min and the DNA was precipitated with 2.5 volumes of ethanol at -20°C for 1 h and washed once with 1 mL of 70% (v/v) ethanol. The *XhoI* cut DNA was ligated for 30 min with 400 units of T4 DNA ligase (NEB) at 16°C and the ligation mix used as template for the amplification by PFU DNA polymerase of the VL–VH construct using PCR (Section 4.2.3) and primers specific for VL (Tre1 or Bel1) and the JH genes (Tre4 or Bel4). The VL–VH amplified DNA (Figure 4.3) was precipitated with ethanol and digested with *HindIII* and *SalI* restriction endonucleases, gel purified and ligated into the *HindIII* and *SalI* sites of the vector pFLAG-CTS. This method resulted in between 80 and 100 percent efficiency in obtaining the pFLAG–CTS (VL–VH) dicistronic vector constructs as determined through purification of plasmid DNA from multiple bacterial colonies (data not shown).

4.3.3: Mutagenesis of the Tre dicistronic construct and sequencing of dicistronic operons of Bel and Tre Fv expression vectors

To clone the VL region gene of Tre into pFLAG-CTS, the N-terminal residue was modified from an aspartic acid to a tyrosine in order to maintain compatibility with the *HindIII* restriction site. This codon was then mutagenised using a PCR site-directed mutagenesis method resulting in the change of TAC (Y) → GAC (D) so that the native VL and VH region sequences could be expressed in the Tre VL–VH dicistronic pFLAG-CTS construct. Selection of mutated plasmids was based on the destruction of the *HindIII* site as a result of the correct mutation. All selected clones contained the GAC mutation (Figure 4.4). Figures 4.5 and 4.6 show the sequences for the dicistronic constructs of Bel and Tre Fv obtained through automated DNA sequencing (Section 4.2.4) with N26 and C24 labelled (IRD41, Li-Cor) primers.

A



B

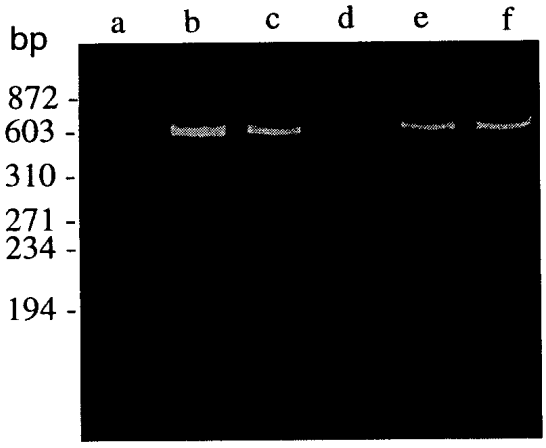


Figure 4.3: Ligation-PCR construction of VL–VH dicistronic operons. **A:** Schematic representation of the strategy used for constructing the VL–VH constructs in pFLAG-CTS. **B:** Polyacrylamide gel (10%) showing the PCR amplification the VL–VH construct from the *XhoI* cut and ligated vectors. Lanes: Bel no template control (a); Bel 1 μL of ligation mix (b); Bel 1 μL of 1:10 dilution of ligation mix (c); Tre no template control (d); Tre 1 μL of ligation mix (e); Tre 1 μL of 1:10 dilution of ligation mix (f).

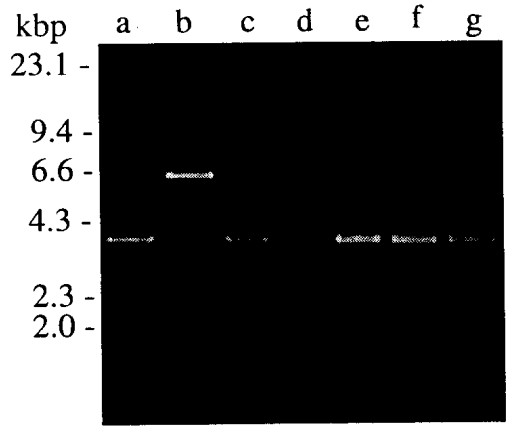


Figure 4.4: Site-directed mutagenesis of the first codon of Tre VL region gene. Agarose gel (0.8%) of parental Tre Fv plasmid with tyrosine (TAC) at position one and mutated plasmids with aspartic acid (GAC) at position one of the VL region gene. Lanes: uncut parent Tre plasmid (a); *HindIII* digested parent plasmid (b); *HindIII* digested mutated plasmids (c-g).

ATGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTTGCG
M K K T A I A I A V A L A G F A T V A

*CAAGCTT*CCTATGTGCTGACTCAGCCACCCTCAGTGTCAGTGGCCCCAGGAAAGACG
Q A S Y V L T Q P P S V S V A P G K T

GCCAGGATTACCTGTGGGGGAAACAACATTGGAAGTAAAAGTGTGCACTGGTACCAG
A R I T C G G N N I G S K S V H W Y Q

CAGAAGCCAGGCCAGGCCCCTGTGCTGGTCATCTATTATGATAGCGACCGGCCCTCA
Q K P G Q A P V L V I Y Y D S D R P S

GGGATCCCTGAGCGATTCTCTGGCTCCAACTCTGGGAACACGGCCACCCTGACCATC
G I P E R F S G S N S G N T A T L T I

AGCAGGGTCGAAGCCGGGGATGAGGCCGACTATTACTGTCAGGTGTGGGATAGTAGT
S R V E A G D E A D Y Y C Q V W D S S

AGTGTAGTATTCGGCGGAGGGACCAAGCTGACCGTCCTATGATGA<u>CTCGAG</u>TAATT
S V V F G G G T K L T V L * *

TACCAACACTACTACGTTTTAACTGAAACAAACTGGAGACTCAT*ATGAAATACCTAT*
M K Y L

*TGCCTACGGCAGCCGCTGGATTGTTATTACTCGCGGCCCAGCCGG<u>CCATGG</u>CC*GAGG
L P T A A A G L L L L A A Q P A M A E

TGCAGCTGCAGGAGTCTGGAGCAGAGGTGAAAAAGCCCGGGGAGTCTCTGAAGATCT
V Q L Q E S G A E V K K P G E S L K I

CCTGTAAGGGTTCTGGATACAGCTTTACCAACTACTGGATCGGCTGGGTGCGCCAGA
S C K G S G Y S F T N Y W I G W V R Q

TGCCCGGGAAAGGCCTGGAGTGGATGGGGATCATCTATCCTGGTGACTCTGATACCA
M P G K G L E W M G I I Y P G D S D T

GATACAGCCCGTCCTTCCAAGGCCAGGTCACCATCTCAGCCGACAAGTCCATCAGCA
R Y S P S F Q G Q V T I S A D K S I S

CCGCCTACCTGCAGTGGAGCAGCCTGAAGGCCTCGGACACCGCCATGTATTACTGTG
T A Y L Q W S S L K A S D T A M Y Y C

CGAGACGGACGGGGACCGGGGATCCTTACTACTACTACTACATGGACGTCTGGG
A R R T G T G D P Y Y Y Y Y Y M D V W

GCAAAGGGACCACGGTCACCGTGTCCTCA<u>GTCGac</u>tacaaggacgacgatgacaagTGA
G K G T T V T V S S V D Y K D D D D K *

Figure 4.5: DNA and amino acid sequence of the Bel pFLAG-CTS (VL–VH) dicistronic construct. Periplasmic secretion sequences OmpA/PelB are in italics. The DNA sequence of the FLAG peptide is in lowercase. Stop codons are indicated (*). Restriction enzyme sites are underlined. The sequence was obtained with N26 and C24 IRD41 labelled primers (Li-Cor).

ATGAAAAAGACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTTGCG
M K K T A I A I A V A L A G F A T V A

*CAAGCT*GACATCCAGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTAGGAGAC
Q A D I Q M T Q S P S S L S A S V G D

AGAGTCACCATCACTTGCCGGGCAAGTCAGAGCATTAGCAGCTATTTAAATTGGTAT
R V T I T C R A S Q S I S S Y L N W Y

CAGCAGAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAA
Q Q K P G K A P K L L I Y A A S S L Q

AGTGGGGTCCCATCAAGGTTCAGTGGCAGTGGATCTGGGACAGATTTCACTCTCACC
S G V P S R F S G S G S G T D F T L T

ATCAGCAGTCTGCAACCTGAAGATTTTGCAACTTACTACTGTCAACAGAGTTACAGT
I S S L Q P E D F A T Y Y C Q Q S Y S

ACCCCTCCGTACACTTTTGGCCAGGGGACCAAGCTGGAGATCAAATGATGACTCGA
T P P Y T F G Q G T K L E I K * *

GTAATTTACCAACACTACTACGTTTTAACTGAAACAAACTGGAGACTCAT*ATGAAAT*
M K

ACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCGGCCCAGCCGGCCATGG
Y L L P T A A A G L L L L A A Q P A M

*CC*GAGGTGCAGCTGCAGGAGTCTGGAGCAGAGGTGAAAAAGCCCGGGGAGTCTCTGA
A E V Q L Q E S G A E V K K P G E S L

GGATCTCCTGTAAGGGTTCTGGATACAGCTTTACCAGCTACTGGATCAGCTGGGTGC
R I S C K G S G Y S F T S Y W I S W V

GCCAGATGCCCGGGAAAGGCCTGGAGTGGATGGGGAGGATTTATCCTGGTGACTCTT
R Q M P G K G L E W M G R I D P S D S

ATACCAACTACAGCCCGTCCTTCCAAGGCGTCGTCACCATCTCAGCTGACAAGTCCA
Y T N Y S P S F Q G V V T I S S D K S

TCAGCACTGCCTACCTGCAGTGGAGCAGCCTGAAGGCCTCGGACACCGCCATGTATT
I S S A Y L Q W S S L K A S D T A M Y

ACTGTGCGAGAAGGCAGTGGCTGGCCCTAGGCCACTTTGACTACTGGGGCCAGGGAA
Y C A R R Q W L A L G H F D Y W G Q G

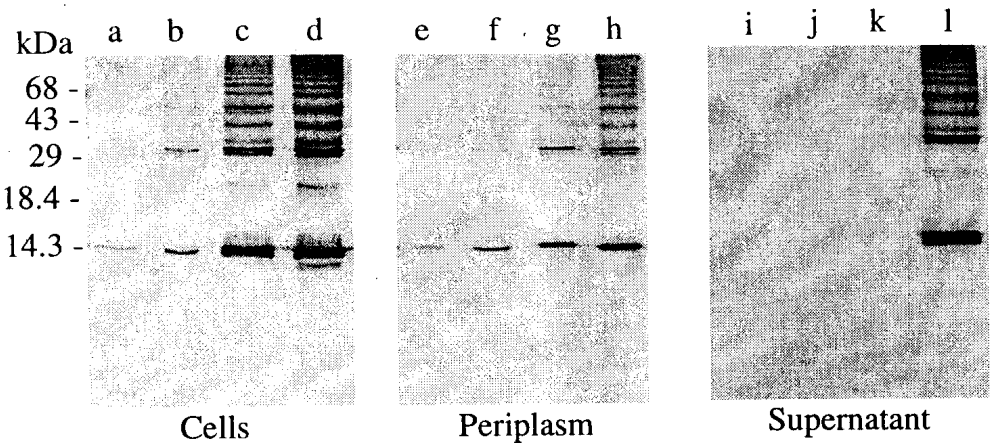
CCCTGGTCACCGTCTCCTCAGTCgactacaaggacgacgatgacaagTGA
T L V T V S S V D Y K D D D D K *

Figure 4.6: DNA and translated amino acid sequence of the Tre Fv pFLAG-CTS (VL–VH) dicistronic construct. Periplasmic secretion sequences OmpA/PelB are in italics. The DNA sequence of the FLAG peptide is in lowercase. Stop codons are indicated (*). Restriction enzyme sites are underlined. The sequence was obtained with N26 and C24 IRD41 labelled primers (Li-Cor).

4.3.4: Optimisation of protein expression from Bel and Tre pFLAG-CTS (VL–VH) vectors

Protein expression was monitored by Western blot immunoassays with the primary antibody (M2) detecting the FLAG octapeptide which was fused to the carboxyl terminus of the VH regions of Bel and Tre. Variables which affected VH expression were induction time (Figure 4.7), molar concentration of IPTG (Figure 4.8) and the strain of *E.coli* (Figure 4.9). The VH protein ran at approximately 14 kDa and occasionally aggregated on non-reducing gels as high molecular weight species. The temperature of induction (37°C or 25°C) did not appear to result in different amounts of VH within the soluble fractions obtained from the periplasm or culture supernatant (data not shown). The expression of Bel VH in the whole cell fraction reached maximal levels after 3 h post-induction with IPTG, although the VH was detected at all time points tested (Figure 4.7). The VH of Bel was detectable in the periplasm from 0 h post-induction and obtained maximal levels between 3 h and 16 h post-induction. Bel VH was detected in the culture supernatant only after 16 h post-induction. Expression of Tre VH in the whole cell culture was detected after 1 h and attained a maximal level by 3 h post-induction. Similar to Bel, the Tre VH was detected in the periplasm at 1 h post-induction and obtained maximal levels between 3 and 16 h post-induction. Tre VH was detected in the culture supernatant only after 16 h post-induction (Figure 4.7). A concentration of 0.02 mM IPTG was determined to result in optimal expression of Bel and Tre VH regions in the soluble periplasmic fraction. The lower IPTG concentration (0.004 mM) resulted in substantially lower yields of VH. Bel VH was expressed at low levels in the absence of IPTG induction. Higher IPTG concentrations (0.1 and 0.5 mM) favoured the formation of aggregates, decreased the amount of soluble product in the case of Bel Fv and increased the levels of unprocessed PelB–VH (ie. a second band at around 14.5 kDa) in the periplasm for Tre Fv (Figure 4.8). Similar levels of protein expression were observed in the whole cell fraction and periplasmic fraction for DH5α, TOPP2 and LE392 strains of *E.coli* for Bel Fv. Tre Fv expressed optimally in TOPP2 cells, although DH5α and LE392 cells also contained considerable levels of VH in the whole cell and periplasmic fractions. The BL21 strain of *E.coli* expressed low levels of Bel VH, however, the same cells containing the Tre Fv vector did not grow to sufficient density for testing protein expression. When tested in the Y1090 strain of *E.coli* neither Bel or Tre Fv vectors expressed detectable levels of protein (Figure 4.9).

A



B

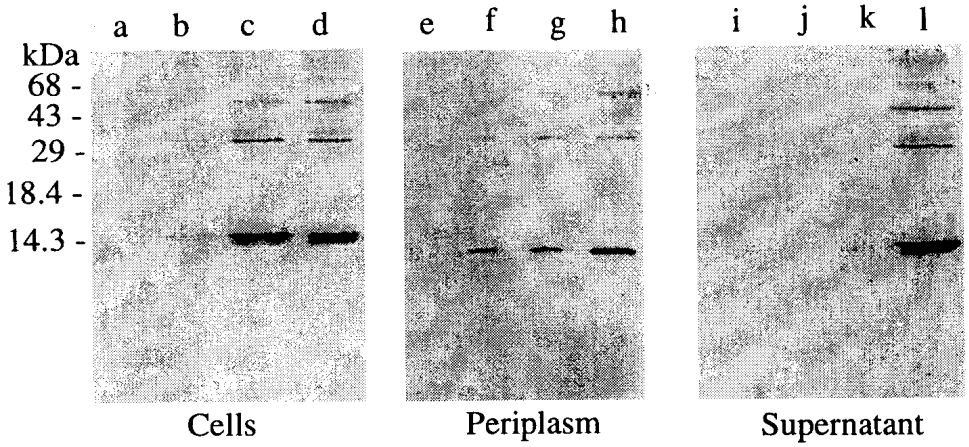
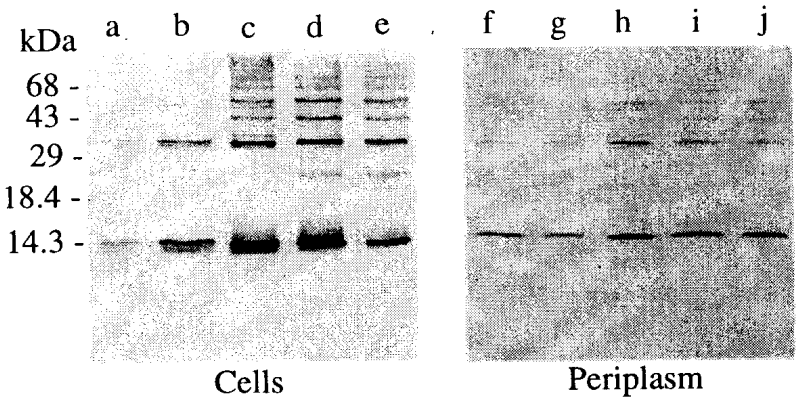


Figure 4.7: Time course of Bel and Tre Fv expression in pFLAG–CTS. **A:** Bel Fv expressions. **B:** Tre Fv expressions. Western blots probed with 1:15000 dilutions of anti-FLAG M2 and an anti-mouse Ig (H + L) AP conjugate. Samples of cellular (5 µL), periplasmic (15 µL) and supernatant (15 µL) fractions were prepared at 0 h (a,e,i); 1 h (b,f,j); 3 h (c,g,k) and 16 h (d,h,l) post-induction time points.

A



B

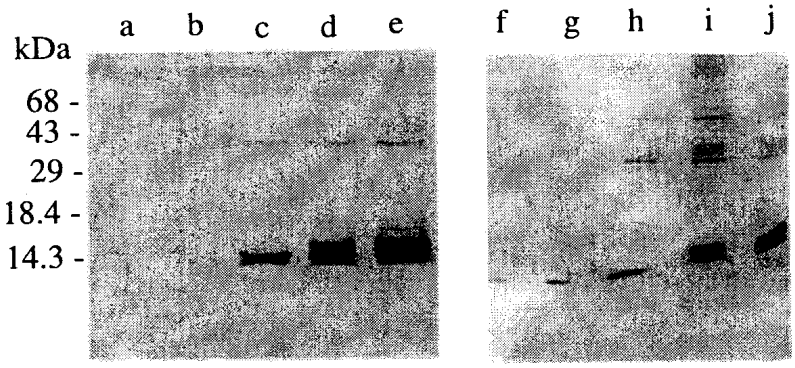
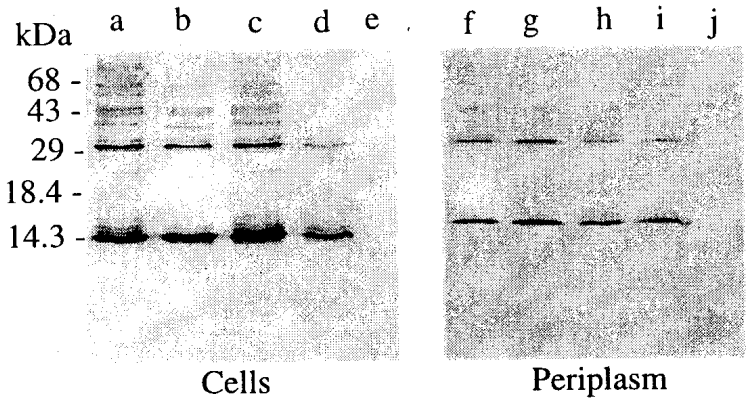


Figure 4.8: Optimisation of IPTG concentration. **A:** Bel Fv expressions. **B:** Tre Fv expressions. Cultures (10 mL) were incubated with various concentrations of IPTG for 3 h. The whole cells (5 μL) and periplasmic fractions (15 μL) were analysed by Western blot immunoassay using 1:15000 dilutions of anti-FLAG M2 and an anti-mouse Ig (H + L) AP conjugate. IPTG concentrations: 0 mM (a,f); 0.004 mM (b,g); 0.02 mM (c,h); 0.1 mM (d,i); 0.5 mM (e,j)

A



B

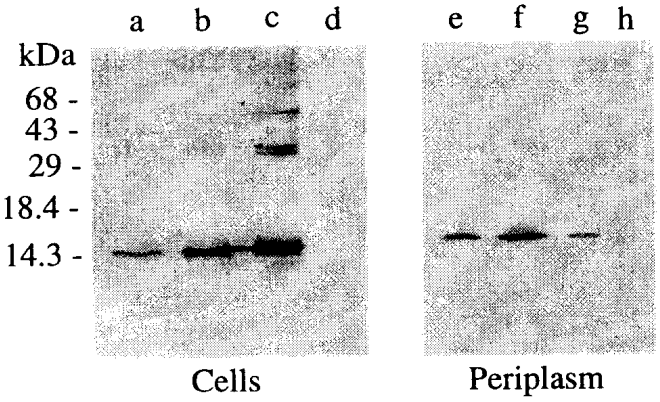


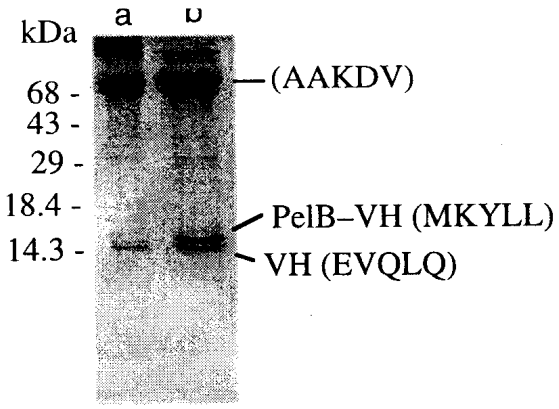
Figure 4.9: Host *E.coli* cell strains. A: Bel Fv expressions. B: Tre Fv expressions. Host strains used for Bel: DH5α (a,f); TOPP2 (b,g); LE392 (c,h); BL21 (d,i) and Y1090 (e,j). Host strains used for Tre: DH5α (a,e); TOPP2 (b,f); LE392 (c,g) and Y1090 (d,h). The expressions were in the presence of 0.02 mM IPTG and for 3 h post-induction.

4.3.5: Affinity purification of expressed proteins

Proteins expressed into the supernatant of 16 h post-induction bacterial cultures (Bel or Tre) were purified using the anti-FLAG (M2) affinity matrix (Kodak). The yield of protein which bound to the column was between 0.5 and 2 mg/L of culture supernatant. However, the affinity purified protein consisted of three main components when analysed under non-reducing and reducing conditions using SDS-PAGE (Figure 4.10). The M2 antibody purified peptides of Bel and Tre Fv expression cultures had approximate molecular masses of 150 kDa, 15 kDa and 14 kDa. The 15 and 14 kDa proteins stained positive for the FLAG octapeptide while the 150 kDa protein was negative when analysed by Western blot immunoassay (data not shown). The affinity purified proteins for Bel Fv were transferred to PVDF membrane and the N-terminal sequence was obtained from each peptide. The 150 kDa component had an N-terminal sequence (AAKDV) which did not correspond to VL, VH, OmpA or PelB protein sequences. The 15 kDa peptide had sequence of MKYLL which corresponds to the N-terminus of the PelB secretion signal which was fused to the VH region. The 14 kDa peptide had a sequence of EVQLQ which is the N-terminus of the VH domain. Furthermore, no protein was obtained from M2 antibody purified supernatant which corresponded to the VL regions of Bel or Tre. Periplasmic fractions gave the same result as culture supernatant for M2 antibody purification of Bel and Tre expressions (data not shown).

Streptococcal protein L specifically binds to VLκ domains (Murphy *et al.*, 1994). Thus, the protein expressed into the supernatant of Tre (κ L chain) but not Bel (λ L chain) bacterial expression cultures was affinity purified using a protein L column (KappaLock™, Zymed). The yield of KappaLock™ purified protein was only 100 to 200 ng/L from 16 h post-induction culture supernatants. No protein could be purified from the periplasmic fraction. The bound protein resolved as a single band at approximately 12 kDa on 15 % (w/v) PAGE in the presence of 0.1% SDS and under reducing and non-reducing conditions (Figure 4.10). This protein was not positive for the FLAG peptide when analysed by Western blot immunoassay (data not shown). The protein was transferred to PVDF membrane and N-terminal sequencing (DIQMT) confirmed that the 12 kDa protein was the Tre light chain variable domain.

A



B

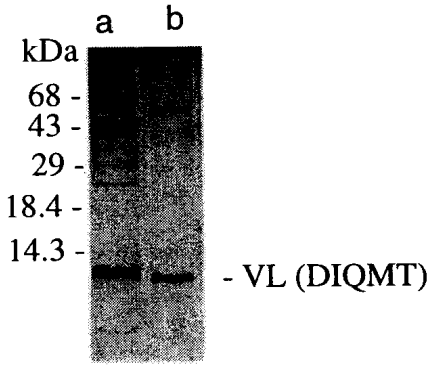
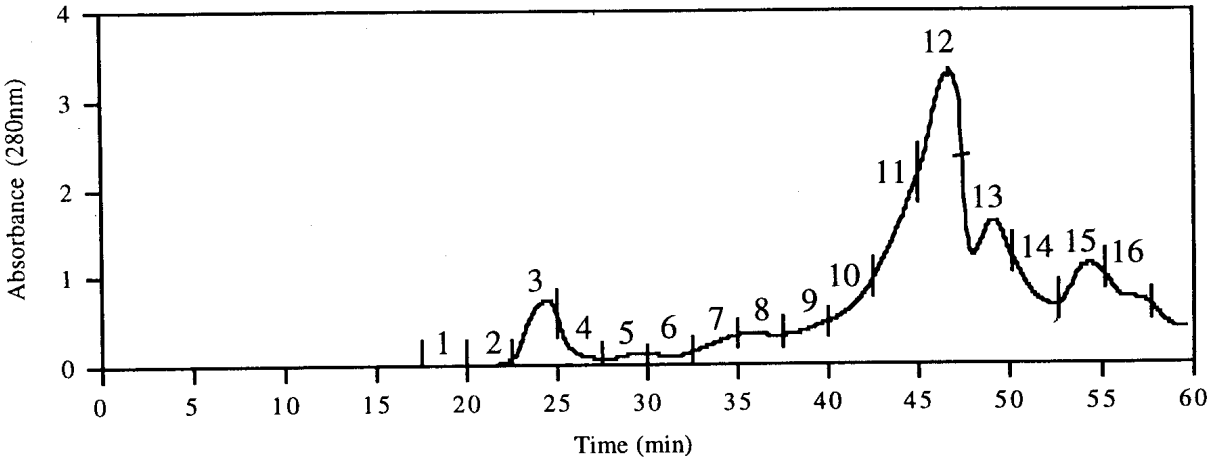


Figure 4.10: Affinity purification of expressed protein. **A:** M2 affinity purified Bel Fv expression. **B:** KappaLock™ purified Tre Fv expression. Lanes: non-reduced (a) and reduced (b). The N-terminal sequence of each peptide is indicated.

4.3.6: Size-exclusion chromatography of Bel and Tre bacterial expression cultures

To determine the species present in solution for the expressed Fv of Bel and Tre the culture supernatants or periplasmic fraction (16 h post-induction) were subjected to size-exclusion chromatography using a Superdex 75 column (Pharmacia). Figure 4.11 shows a typical separation profile of bacterial expression culture supernatant and the corresponding Western blot immunoassay of fractions collected during the chromatography. The column was calibrated (Appendix C) and the FLAG positive proteins were determined to have molecular weight ranges in solution of 52 to >100 kDa for Bel and 41 to >100 kDa for Tre cultures. Similar results were obtained from the periplasmic fractions (data not shown). No FLAG fusion protein was eluted at time points consistent with the presence of monomers (VH), homo-dimers (VH–VH) or hetero-dimers (VL–VH).

A



B

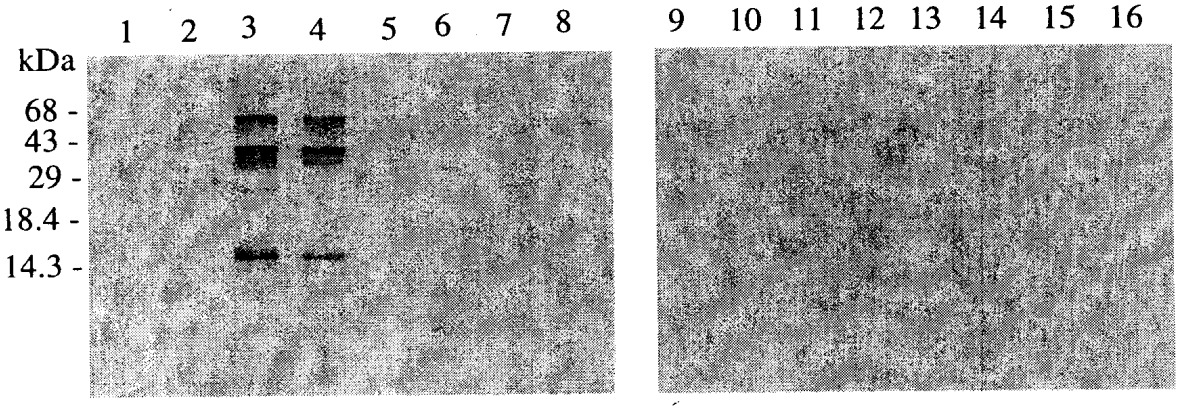


Figure 4.11: Size-exclusion chromatography of bacterial expression culture supernatants. **A:** a representative elution profile (Bel). Fractions (100 µL) have been indicated and correspond to lanes on the Western blot. **B:** fractions analysed by Western blot immunoassay using 1:15000 dilutions of anti-FLAG M2 and an anti-mouse Ig (H + L) AP conjugate. Fractions are labelled 1 to 16 and were resolved on a 15% polyacrylamide gel containing 0.1% SDS.

4.4: Discussion

Elucidation of the structural mechanisms underpinning immunoglobulin polyreactivity requires the solution of structures of at least one such Ig fragment to be determined in complex with several heterologous antigens. The smallest fragment of an antibody which retains the native structure and function of the Ag binding site is the Fv (VL–VH). An Fv was the fragment of choice for bacterial expression of the B CLL immunoglobulins which are the subject of this study. The advantage of an Fv molecule is that the small size of an Fv should enable higher resolution data to be collected from crystals of Fv compared to Fab molecules. Furthermore, an oligopeptide linker of a scFv potentially may interfere with the association of the V domains and partially block the Ab combining site. However, bacterial expression of Fv molecules derived from Bel and Tre B CLL immunoglobulins demonstrated that non-covalently linked VL and VH domains of these Ig are not suitable for large scale purification required for X-ray crystallography. This conclusion was based upon the inability to co-purify VL–VH using affinity for a single V domain. Also a propensity of expressed VH domain to aggregate in solution was observed and this was not consistent with the association of VL and VH domains as Fv molecules. Further attempts at structural studies of polyreactive Ig fragments should focus on the expression of Fab or scFv molecules without forgetting the potential disadvantages of such fragments.

4.4.1: Development of dicistronic vectors for soluble expression of B CLL variable region fragments

The variable region genes of the light and heavy chains of Bel and Tre immunoglobulins were cloned into a bacterial expression vector (pFLAG-CTS) as a dicistronic operon (Figure 4.1). The bacterial secretion sequences OmpA and PelB were fused to the VL and VH region genes respectively. A method based upon heteroligation of *XhoI* digested constructs of pFLAG-CTS (VL) and pHUD (VH) and the amplification of the artificial operon using PCR (Figure 4.3) was successful in overcoming difficulties involved in the cloning of fragments into *XhoI* and *SalI* restriction sites. The amplified construct was then cloned into pFLAG-CTS vector with between 80 and 100 percent recombinant plasmids obtained. This approach is both rapid and highly efficient suggesting its suitability for the efficient cloning of VL and VH region genes for soluble expression of Fv from various antibodies.

4.4.2: Characterisation of protein expressed by Bel and Tre pFLAG-CTS (VL–VH) expression systems

The protein expression of Bel and Tre pFLAG-CTS (VL–VH) vectors was monitored using an antibody (M2) specific for the FLAG octapeptide which was fused to the carboxyl-terminus of the VH regions. Unfortunately, the expression of VL regions of Tre and Bel could not be similarly detected since an antibody specific for the VL was not available. However, since both VL and VH region genes are cloned as a dicistronic operon under control of a single promoter, the proteins should be produced in roughly equimolar amounts (Skerra & Pluckthun, 1988). Affinity purification of protein expressed into the bacterial culture supernatant using a protein L column directly proved (ie. N-terminal amino acid sequencing) that the Tre VL domain was synthesised. However, the Tre VL appeared to be produced at lower levels (100-200 ng/L) compared to VH. Although it is possible that protein L only binds to correctly folded VL domains and so the lower yield may indicate that this domain is not folding appropriately in the bacteria. A similar approach was not possible for Bel VL since protein L does not bind to λ isotype VL domains (Murphy *et al.*, 1994), although given the sequences of the dicistronic constructs (Figures 4.5 and 4.6) it is likely that both VL and VH domains are produced in Bel and Tre Fv expression systems.

Expression of the VH domains reached maximal levels in the total cultures at 3 h post-induction with IPTG for Bel and Tre expression cultures (Figure 4.7). The accumulation of VH within the periplasmic space followed a similar pattern to that of the total culture and VH was only present in the culture supernatant after 16 h. Froyen *et al.* (1993) observed similar patterns of Fv expression when comparing the bacterial expression of an Fv and scFv derived from an anti-human interferon-γ (anti-HuINF-γ) Ab. The kinetics of the expression of the anti-HuINF-γ Fv were more rapid than observed for the B CLL fragments which may be a result of the higher concentration of IPTG used to induce protein expression (0.1 mM for the anti-HuINF-γ compared with 0.02 mM for the B CLL Fv). Similar to the B CLL Fv the accumulation of the anti-HuINF-γ fragments within the bacterial culture media was much slower than within the periplasmic space and reached maximal levels at 20 h post-induction (Froyen *et al.*, 1993).

A variety of factors affect the level and quality of Ab fragment expression in bacteria including the concentration of induction agent, host cell strain, temperature of induction and the primary amino acid sequence (Sawyer *et al.*, 1994). An IPTG

concentration of 0.02 mM was determined to be optimal for the soluble expression of Bel and Tre Fv (Figure 4.8). At this concentration of IPTG the level of expression of VH within the periplasmic space was maximised and the PelB secretion signal was correctly processed. At higher IPTG concentrations (0.1 and 0.5 mM) a band corresponding to unprocessed PelB–VH was present in the periplasmic protein. Unprocessed PelB–VH was a minor problem in expression of the Bel construct with the main feature of the higher IPTG concentrations being a reduction in the amount of VH secreted into the periplasmic space. However, the amount of unprocessed polypeptide was greater than processed VH for expression of Tre Fv at concentrations of IPTG higher than 0.02 mM. Expression of VH occurred for Bel Fv without adding IPTG which is attributed to the leakiness of the Tac/Lac promoter/repressor system (Pluckthun & Skerra, 1989). The host stain of *E.coli* used for the expression of Bel and Tre Fv molecules was important (Figure 4.9). Bel Fv was expressed at similar levels in DH5α, TOPP2 and LE392 strains of *E.coli*, with lower levels of VH in BL21. In contrast, VH was not expressed in Y1090 cells. For Tre Fv maximal quantities of VH were expressed in the periplasm of TOPP2 cells. Although Tre Fv in DH5α cells expressed lower amounts of VH in the total culture compared to LE392 cells, the levels attained in the periplasm were similar if not reversed for these two host cell strains. Tre Fv did not express in Y1090 cells and expression could not be tested for BL21 cells which did not grow to sufficient density in liquid culture. The temperature of induction did not appear to affect the level of VH expression in either system. Primary amino-acid sequence has been shown to dramatically influence the soluble expression of Ab fragments (Kipriyanov *et al.*, 1997). However, it is clearly not desirable to modify the primary sequences of Bel and Tre Fv molecules for the purposes of this study since the focus is understanding the function of the native proteins.

4.4.3: Affinity and size-exclusion chromatography of Bel and Tre expressed protein

Purification of Bel or Tre expressed protein using an anti-FLAG (M2) column did not co-purify VL with the VH-FLAG fusion protein. Affinity purification of the Tre VLκ domain also did not co-purify the partner VH domain (Figure 4.10). These results suggest that the Fv heterodimer is either not present in the supernatant or periplasmic space or is dissociating during affinity purification. To assess the nature of the B CLL Fv molecules in the culture supernatant and periplasmic fractions size-exclusion chromatography was performed immediately after 16 h of protein expression. The VH

domains of Bel and Tre eluted in a size range attributable to the formation of aggregates (Figure 4.11). Furthermore, no VH was eluted at retention times consistent with the presence of the VH as a heterodimer (VL–VH), homodimer (VH–VH) or monomer (VH). Glockshuber *et al.* (1992) demonstrated that some Fv molecules dissociate into monomers of VL and VH at low concentrations and that VH domains have a propensity to aggregate irreversibly in the absence of the VL–VH interaction. An approach to obtain the correctly folded Fv may be to denature and refold the expressed VL and VH domains *in vitro*. To facilitate this strategy, the VL and VH regions of the B CLLs should be cloned without the use of periplasmic secretion peptides so that large quantities of protein could be obtained from bacterial inclusion bodies. However, the B CLL Fv molecules potentially have a very low affinity of VL for VH which resulted in the inability to produce soluble Fv heterodimers in the Bel and Tre dicistronic expression systems. Further studies are required to assess the stability of the VL–VH association for B CLL derived antibody fragments. Furthermore, expression of a covalently linked fragment such as an Fab or scFv may overcome the difficulties seen with the expression of the non-covalently linked Fv molecules described here.

The potential use of Fv molecules, for X-ray crystallography, was investigated in this study since these fragments are the smallest portion of an Ab which contains the combining site. If large quantities of the Fv were purified and grown into crystals the potential of obtaining high resolution X-ray structures was likely, due to the small size of the Fv molecules. Furthermore, it was thought desirable to avoid any possible interference with the combining site which may result from the inclusion of a peptide linker as found in a scFv fragment. However, the X-ray structures available for scFv in complex with Ag suggest that the VL–VH domain interface is maintained and the interaction with Ag is similar in these recombinant fragments when compared to structures of the parent Ab fragments (Kortt *et al.*, 1994; Zdanov *et al.*, 1994). No electron density has been observed for the linker sequence in any scFv crystallised. While the inability to determine the structure of the linker oligopeptide has been assumed to reflect flexibility of this sequence, it is possible that the residues of the linker are capable of affecting the structure of the binding site through transient interactions with the residues of the VL or VH domains. An alternative to an Fv is an Fab which consists of the light chain covalently joined to the Fd (VH and CH1) through a single disulfide bond. Expression of an Fab in bacteria is more difficult than an Fv or scFv since the formation of the interchain disulfide bond is required (Glockshuber *et al.*, 1992).

4.4.4: Conclusion

Determining the X-ray structures of at least one human polyreactive Ig in complex with several of its Ag is an essential step in elucidating the structural basis of the polyreactive binding phenomenon. The use of bacteria to express Fv of human Ig synthesised by B CLL cells was investigated here. Protein expression was determined to be dependent upon several factors including the time of induction, concentration of induction agent and the host cell strain. However, purification failed to yield Fv heterodimers which were not formed during protein expression in sufficient quantities to use. Future structural studies of human polyreactive Ig should concentrate on the expression of covalently linked fragments (scFv or Fab) in an attempt to overcome the dissociation of heterodimers (VL–VH) and the aggregation (VH) of the expressed protein. However, the caveats of using such fragments in structural studies should be realised.

CHAPTER FIVE
Conclusions and perspective

This thesis has examined the structural basis of human immunoglobulin polyreactivity through the systematic analysis of the variable region genes, amino acid sequence and the three-dimensional structure of five B CLL derived polyreactive Fv molecules. The steps taken to characterise the Ig have provided some important insights into the molecular mechanisms of the polyreactive binding phenomenon exhibited by a significant proportion of human immunoglobulins. The experiments addressed three specific aims: (1) to sequence and analyse the gene segments used in the VL and VH rearrangements within the CD5 positive B CLL cells (Chapter Two); (2) to examine the 3D structure of the binding sites of the polyreactive IgM molecules through homology modelling of the Fv (VL–VH) molecules (Chapter Three), and (3) to clone and express two Fv in a form which would enable X-ray crystallographic studies of native and complexed polyreactive Ig (Chapter Four). The first two objectives were thoroughly explored for the five Igs involved in this study and some interesting aspects regarding the structural nature of Ig polyreactivity were exposed. The non-covalently associated VL and VH were not successfully purified from bacterial expression cultures. This led to a conclusion that the expression of a stabilised fragment is necessary for the successful large scale purification of B CLL derived polyreactive Ig fragments.

The demonstration that all five B CLL cells were capable of binding to mouse immunoglobulins (Chapter Two) is interesting since it suggests that the structural mechanism of polyreactive binding is somehow conserved. Similar to previous attempts at defining a primary structural correlate related to polyreactivity (Ditzel *et al.*, 1996), no conserved primary amino acid motifs were obvious within the CDR sequences of the five IgMs compared in this study. It appears that primary structural diversity is maintained even though the Ig gene repertoire of the V regions is likely to be restricted for polyreactive Igs as opposed to a larger gene repertoire used in monoreactive Igs (Sanz *et al.*, 1989; Schutte *et al.*, 1991; Brown *et al.*, 1992). The genes utilised in the recombination of the V regions of the B CLL immunoglobulins (Chapter Two) partially recapitulates an argument that the structural diversity of the V region genes of polyreactive Igs is reduced compared with that of the monoreactive Ig gene repertoire. Adding to the relatively few λ light chain V region genes sequenced from B CLL cells (Pritsch *et al.*, 1993), Bel expresses a Vλ3 family germline gene (IGLV3S2). At the present stage the number of Vλ region sequences derived from B

CLL cells is too few to decide if gene restriction occurs in rearrangements of λ light chains. The κ light chain expressing B CLLs in this study highlight the restricted gene usage pattern through the expression of the same Vκ4 gene twice (Yar and Hod) and two different Vκ1 family genes (Tre and Jak). A remarkable similarity was seen for the VL regions of Yar and Hod which differed by only two amino acids due to the use of the same Vκ4 and Jκ2 gene segments in the rearrangement. This gene usage pattern supports the non-stochastic use of particular Ig genes in rearrangements of B CLL and polyreactive immunoglobulins. Interestingly, Bel and Tre B CLLs utilise VH5 family genes in their heavy chain rearrangement. The VH5 family is a small gene family which, like the VH4 and VH6 families, tends to be over represented in cases of B CLL (Dean & Norton, 1990). A similar frequent usage of the VH1 family gene DP-10 (51p1) in its unmutated form occurs in B CLL (Kipps, 1993) and this gene is rearranged in the Hod VH region. No restricted gene usage pattern was observed for D minigenes or JH genes in the rearrangements of the B CLL heavy chains, indicating that the sequences of the HCDR3 regions of these Ab is diverse.

Somatic hypermutation results in the V region genes becoming mutated from their germline sequences. The process is initiated by the interaction of the membrane Ig with antigen and results in affinity maturation towards the selecting Ag (reviewed in: Neuberger & Milstein, 1995). An absence (Tre, Yar and Jak) or minimal levels (Bel and Hod) of somatic mutation was found in the VL region genes which is consistent with previous findings that somatic mutations are usually absent from the V region genes expressed by $CD5^+$ B cells in healthy and diseased situations (Baccala *et al.*, 1989; Sanz *et al.*, 1989). Furthermore, an absence (Hod) or low levels of somatic mutation were seen in the VH genes of the partner heavy chains. The low frequency of mutations within VH region sequences may indicate that somatic hypermutation had been initiated by interaction with foreign Ag, in the original B cells prior to, malignant transformation. Alternatively, an auto-antigen recognised by the polyreactive surface IgM of the malignant B cells may have stimulated the somatic mutation process and conceivably the persistence or progression of the disease. These findings support the hypothesis of Schroeder and Dighiero (1994) that, "*antigen-driven stimulation and expansion may play a significant role in subsets of CLL*". This theory which was developed from observations of somatic mutations occurring, consistent with Ag selection, in some cases of B CLL, especially when certain VH genes were expressed (ie. VH251 and VH4.21; Schroeder & Dighiero, 1994). Although the mutation frequency observed here was relatively low, it is still possible that Ag is involved in B CLL.

The three-dimensional structure of polyreactive immunoglobulin combining sites were investigated for the B CLL derived Fv molecules by homology protein modelling techniques. Unexpectedly, the predicted structures of the Fv portions of the B CLL derived Igs revealed a remarkable diversity in binding site structures for these human polyreactive immunoglobulins (Ramsland *et al.*, 1997). Polyreactive Ig combining sites are divided into two basic topologies: (i) a pocket-type binding site is predicted for Bel, Tre and Jak antibodies; and (ii) a relatively flat binding surface is described for Yar and Hod immunoglobulins. Specific differences between the binding surfaces were described (Chapter Three) and these are mainly the result of certain CDRs projecting out from the otherwise compact V domains into solvent. Furthermore, the clear differences in topologies and combining site surface-electrostatics did not agree with the conserved binding to mouse Ig exhibited by the polyreactive surface IgM of the B CLL cells. These observations prompted an analysis of the potential contribution to Ag binding by residues with aromatic side-chains. It was shown that the aromatic groups of these residues, particularly tyrosines, were contributing to the molecular surface over large regions of the combining sites of all five polyreactive Fv molecules (Figure 3.9). Thus, aromatic residues may be heavily involved in the conserved binding to mouse Ig displayed by the B CLL cells and the dominant determinants of human immunoglobulin polyreactivity. The diverse binding site topologies and electrostatics may allow a particular polyreactive Ab to specifically bind to a series of Ags, some or all of which not recognised by another polyreactive Ab. However, the main contribution to the affinity of binding appears to arise from the aromatic side-chains clustered within the Ag binding sites of these immunoglobulins.

Some recently solved X-ray structures of antibodies which are not polyreactive may shed some light on the polyreactive phenomenon since they display some interesting characteristics. Corper *et al.* (1997) have described the first human IgM derived Fab (RF-AN) in complex with its autoantigen (Fcγ). Although this rheumatoid factor activity of the IgM is not associated with a polyreactive binding in RF-AN, the structure of the complex provides some insights into a potential structural mechanism for polyreactivity. The interaction of Fcγ with the binding site of RF-AN occurs along the periphery of a conventional combining site. The only CDRs involved in the binding of Fcγ are LCDR2, HCDR1 and HCDR3. This leaves a major proportion of the combining site unoccupied by Ag which, the authors suggest, is then potentially available for the recognition of a different Ag (Corper *et al.*, 1997). It is therefore possible that the structural diversity observed in the B CLL polyreactive Ig combining sites enables multiple Ag recognition through the recruitment of different portions of these surfaces for binding individual antigens. Further insights into a possible

mechanism of polyreactive Ag binding are provided by Wedemayer and colleagues (1997) in a crystallographic investigation of two hapten binding antibodies which are described as germline (unmutated) and affinity-matured (mutated). These antibodies express the same VH and VL region genes with the affinity-matured antibody having a 30,000 times higher affinity compared to the germline Ab for the nitrophenol phosphonate hapten as a result of several somatic mutations. The important observation relating to these complexes was that in the germline encoded Ab the VL–VH interface and CDR residues were substantially reorganised upon Ag binding compared to the uncomplexed Fab. The movements occurring in the germline Fab, consistent with an induced fit, were not seen for the affinity-matured Ab. Instead the uncomplexed form of the affinity-matured Fab already contained the structural features of the complexed structure of the germline Fab. Clearly the combining sites of germline encoded antibodies have the ability to accommodate Ag by adopting more than one combining site structure and so potentially allowing for multiple Ag recognition. This flexibility of the combining site may be lost after affinity maturation for a particular antigen (Wedemayer *et al.*, 1997). Reflecting on the frequent germline nature of the polyreactive repertoire (Baccala *et al.*, 1989; Sanz *et al.*, 1989) it is easier to reconcile the multiple antigen binding phenomenon. The few somatic mutations found within most of the VH regions of the B CLL derived IgMs may limit the type and number of antigens which these polyreactive immunoglobulins recognise. However, flexibility of the combining sites could be maintained by the remarkably low incidence of somatic mutations within VL regions, allowing for these domains to undergo dramatic movements during antigen binding.

Unpredicted behaviour of the B CLL derived VH and VL domains resulted in the inability to obtain Fv molecules for the growth of protein crystals. Expression of an Fv relied upon the correct folding of the V domains and their non-covalent association as VL–VH heterodimers within the periplasmic space of the bacterial cells. The purification of the expressed proteins (VL–VH) of Bel and Tre cultures using affinity and size-exclusion techniques demonstrated that the V domains were not associating in a form consistent with the correct non-covalent association of an Fv molecule (Chapter Four). It appears that VH and possibly VL are not folding into a native form, which in the case of VH led to its aggregation in solution immediately after protein expression. Although there have been some reports of the successful expression and purification of bacterially expressed Fv (Skerra & Pluckthun, 1988; Eisele *et al.*, 1992), in some cases the affinity of VL for VH appears to be low. In cases where the VL–VH domain interaction is relatively weak the V domains display a propensity to dissociate with the VH, in particular, generally aggregating (Glockshuber *et al.*, 1992). Schiweck and

Skerra (1997) recently noted that the expression of a rationally 'designed' antibody Fv fragment did not yield sufficient protein for crystallisation of the VL–VH heterodimer. The success finally came after an Fab of the same antibody was expressed (Schiweck & Skerra, 1997). However, expression of the Fv portion was desirable, for the present investigation of the structural basis of Ig polyreactivity. It is known definitively that the native conformation of the combining site is maintained in Fv fragments (Bhat *et al.*, 1990; Fan *et al.*, 1992). Furthermore, the relatively small size of an Fv portion of an immunoglobulin allows for the solution of the X-ray structures of Fv molecules to very high atomic resolutions compared with the disulfide linked Fab fragments (Eisele *et al.*, 1992). Because of the difficulty of purifying the non-covalently associated Fv portions of the polyreactive IgMs, a more productive approach to achieve the soluble expression of polyreactive Ig fragments, in future investigations, may be to express stabilised fragments of these polyreactive Igs such as Fab or scFv molecules. Elucidation of the structural mechanisms of human immunoglobulin polyreactivity will require X-ray structures to be solved for at least one polyreactive Ab in complex with several different antigens.

In investigating the structural basis of human immunoglobulin polyreactivity, several insights and lessons have been learnt. Analysis of the variable regions of five B CLL derived polyreactive IgM molecules suggested a restricted pattern of gene usage and the somatic mutation patterns in VL and VH region genes, did not result in a conservation of primary structures in these immunoglobulins. This finding prompted the investigation of the three-dimensional structures of the combining sites for the same antibodies. Homology models of the polyreactive Fvs revealed a diversity of binding site topologies and electrostatics. A dominant role is proposed for aromatic residues in both the conserved binding to mouse Ig and the inherent polyreactivity of the IgM expressed by the $CD5^+$ malignant B cells. Further structural studies should concentrate on the expression of stabilised fragments of these particular immunoglobulins since the non-covalently associated Fv molecules were resistant to purification.

Appendix A

VH region gene sequences and phenotyping of B CLL cells

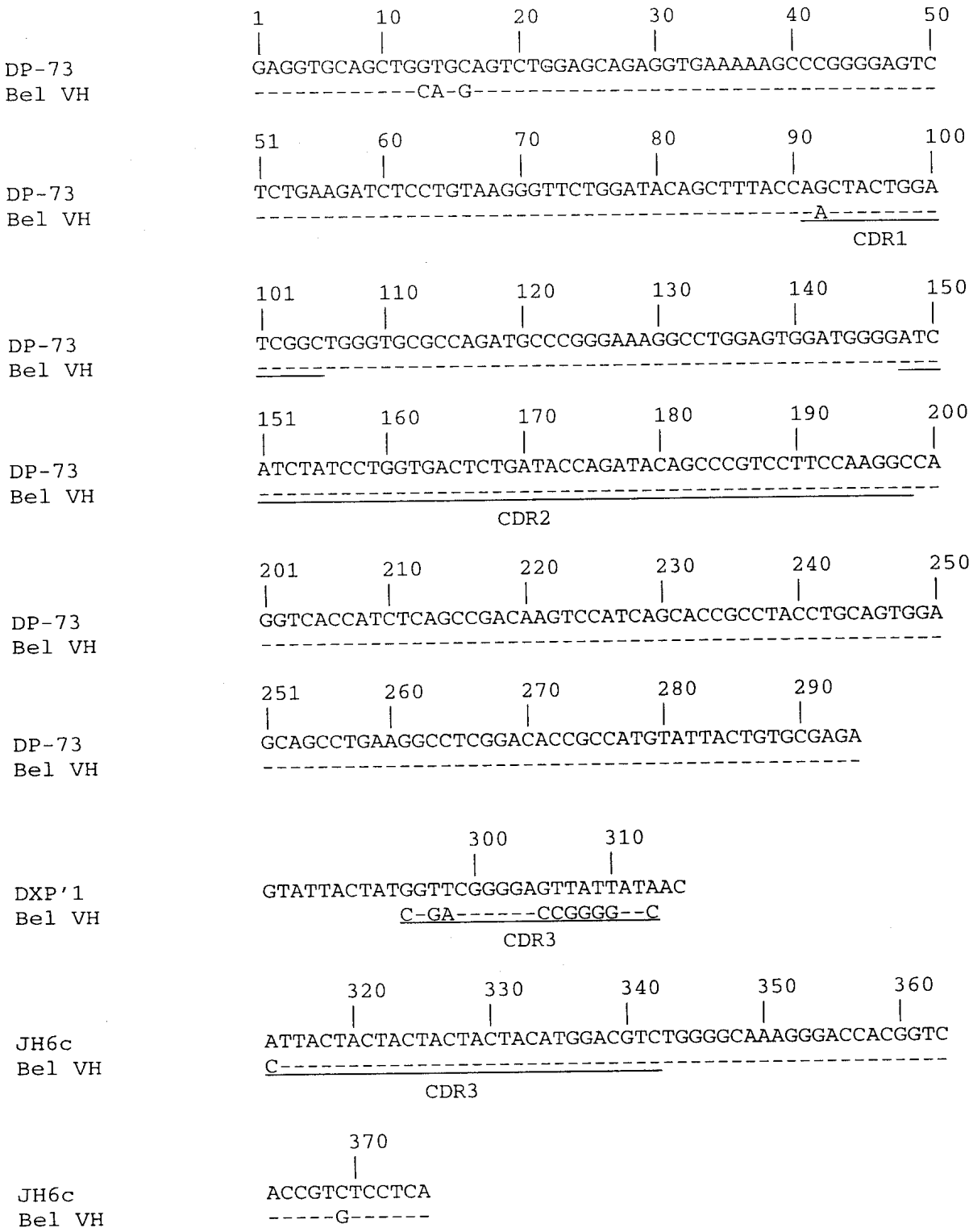


Figure A.1: Consensus nucleotide sequence of Bel VH region. The closest germline gene elements are aligned with the Bel DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Gene assignments are different from the study of Brock (1995) and were made using V BASE. An alternative name for DXP'1 is D3-10. DP-73 is also known as V5-51.

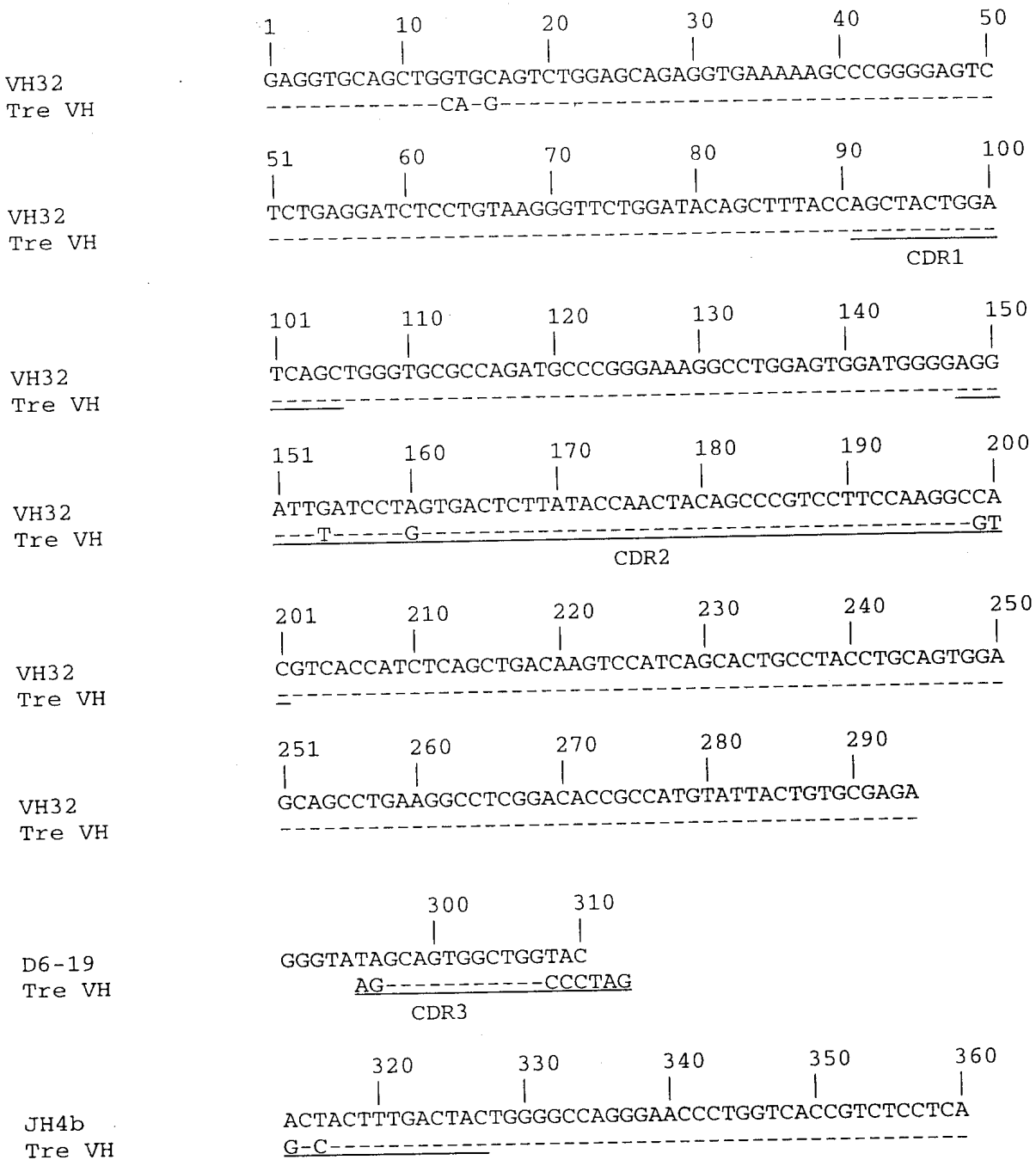


Figure A.2: Consensus nucleotide sequence of Tre VH region. The closest germline gene elements are aligned with the Tre DNA sequence. Identity with the germline gene elements is indicated by dashes. Gene assignments are different from the study of Brock (1995) and were made using V BASE.

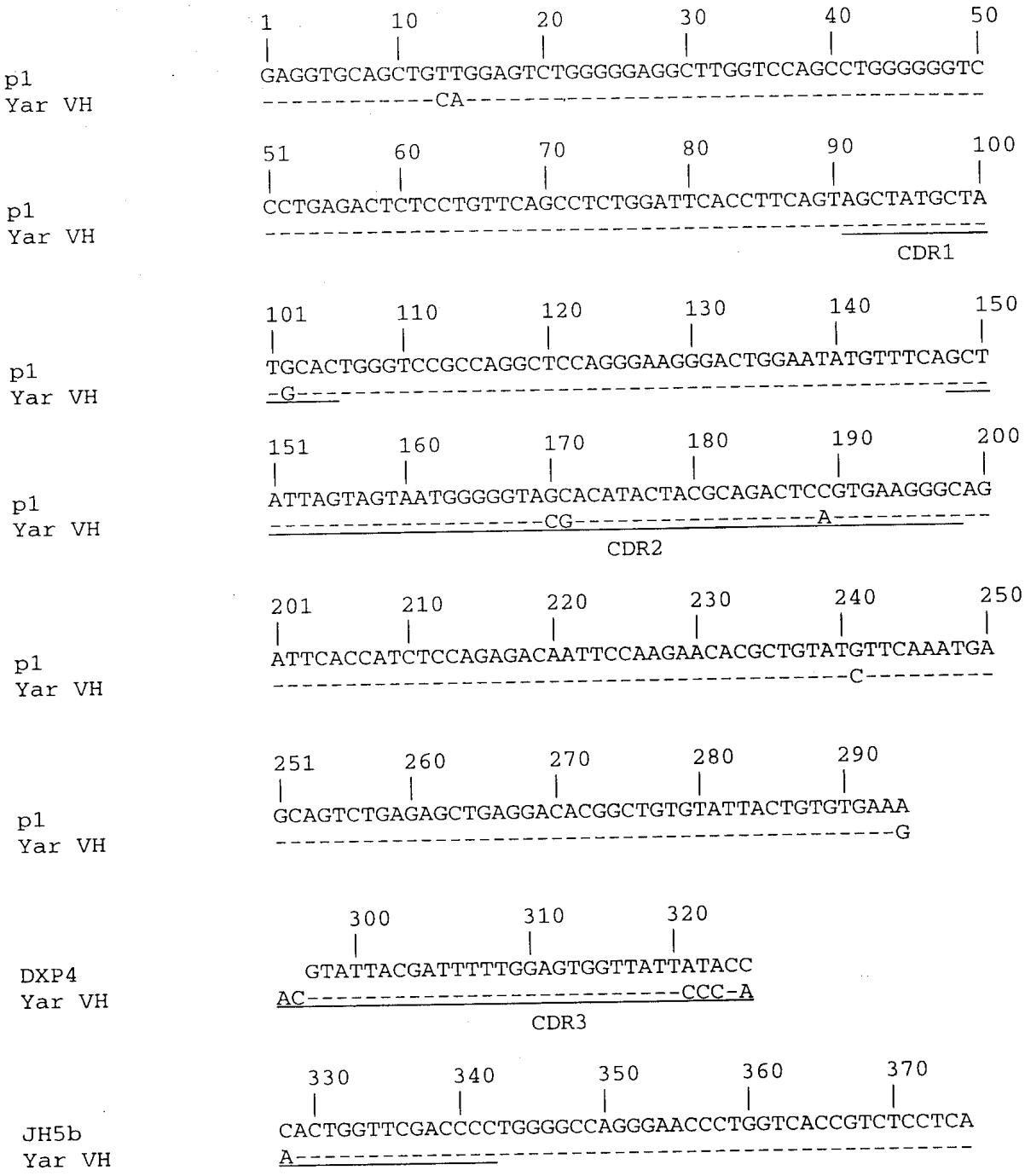


Figure A.3: Consensus nucleotide sequence of Yar VH region. The closest germline gene elements are aligned with the Yar DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Gene assignments are different from the study of Brock (1995) and were made using V BASE. An alternative name for DXP4 is D3-3.

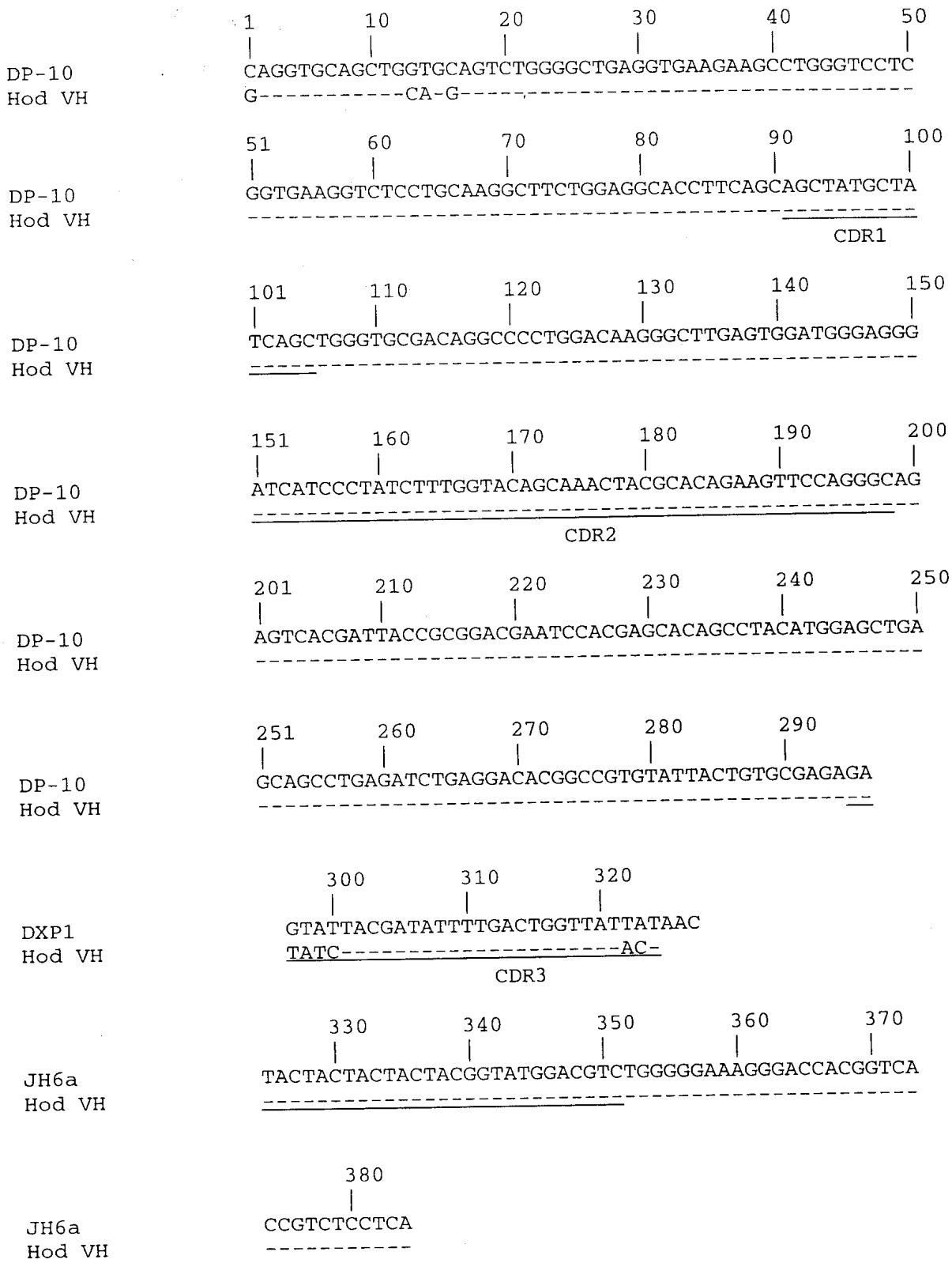


Figure A.4: Consensus nucleotide sequence of Hod VH region. The closest germline gene elements are aligned with the Hod DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Gene assignments are different from the study of Brock (1995) and were made using V BASE. DXP1 is also known as D3-9. DP-10 is also known as 51p1, hv1051 and 1M27.

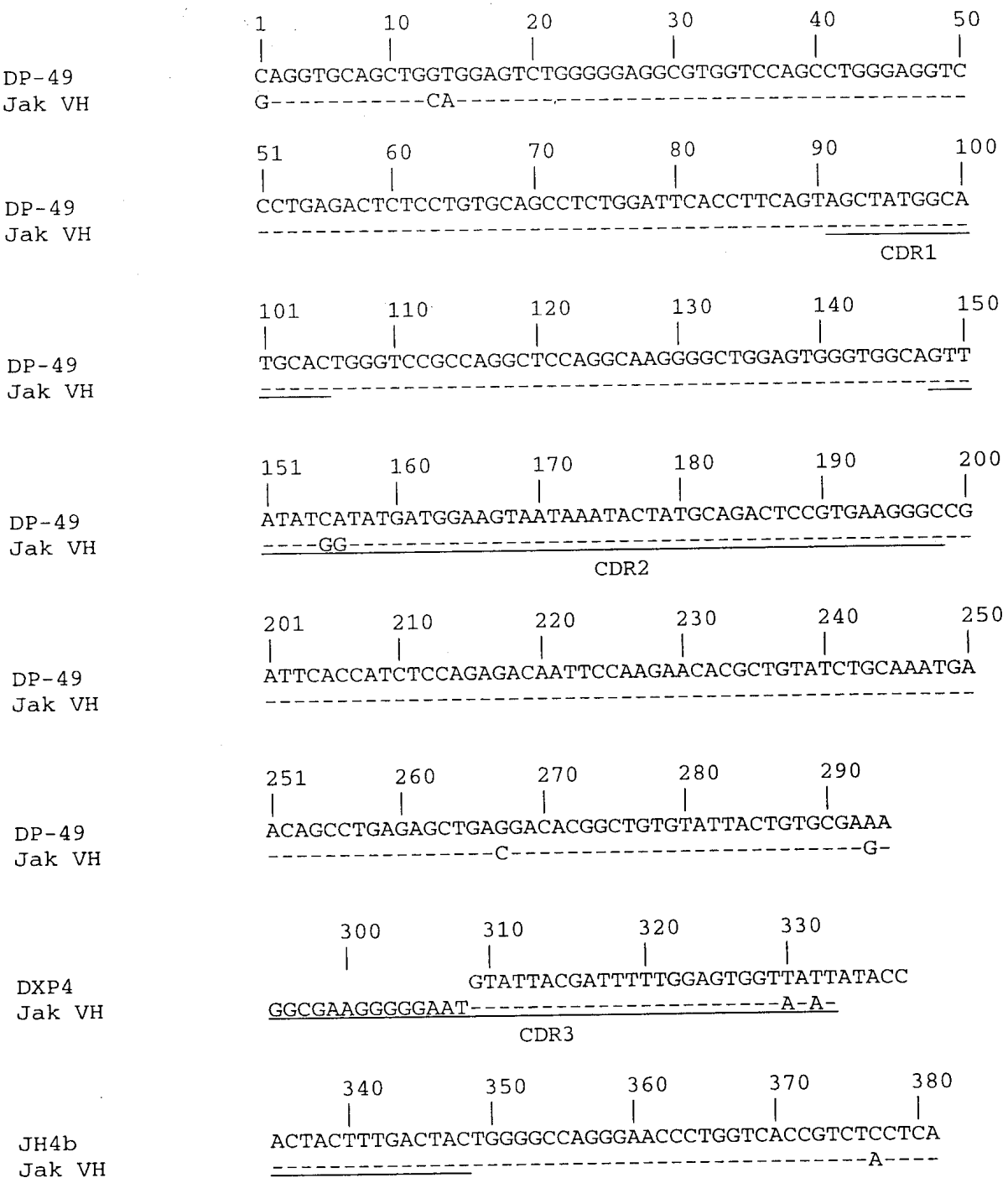


Figure A.5: Consensus nucleotide sequence of Jak VH region. The closest germline gene elements are aligned with the Jak DNA sequence. Identity with the germline gene elements is indicated by dashes. Complementarity determining regions are underlined. Gene assignments are different from the study of Brock (1995) and were made using V BASE. An alternative name for DXP4 is D3-3. DP-49 is also known as VH1.9III.

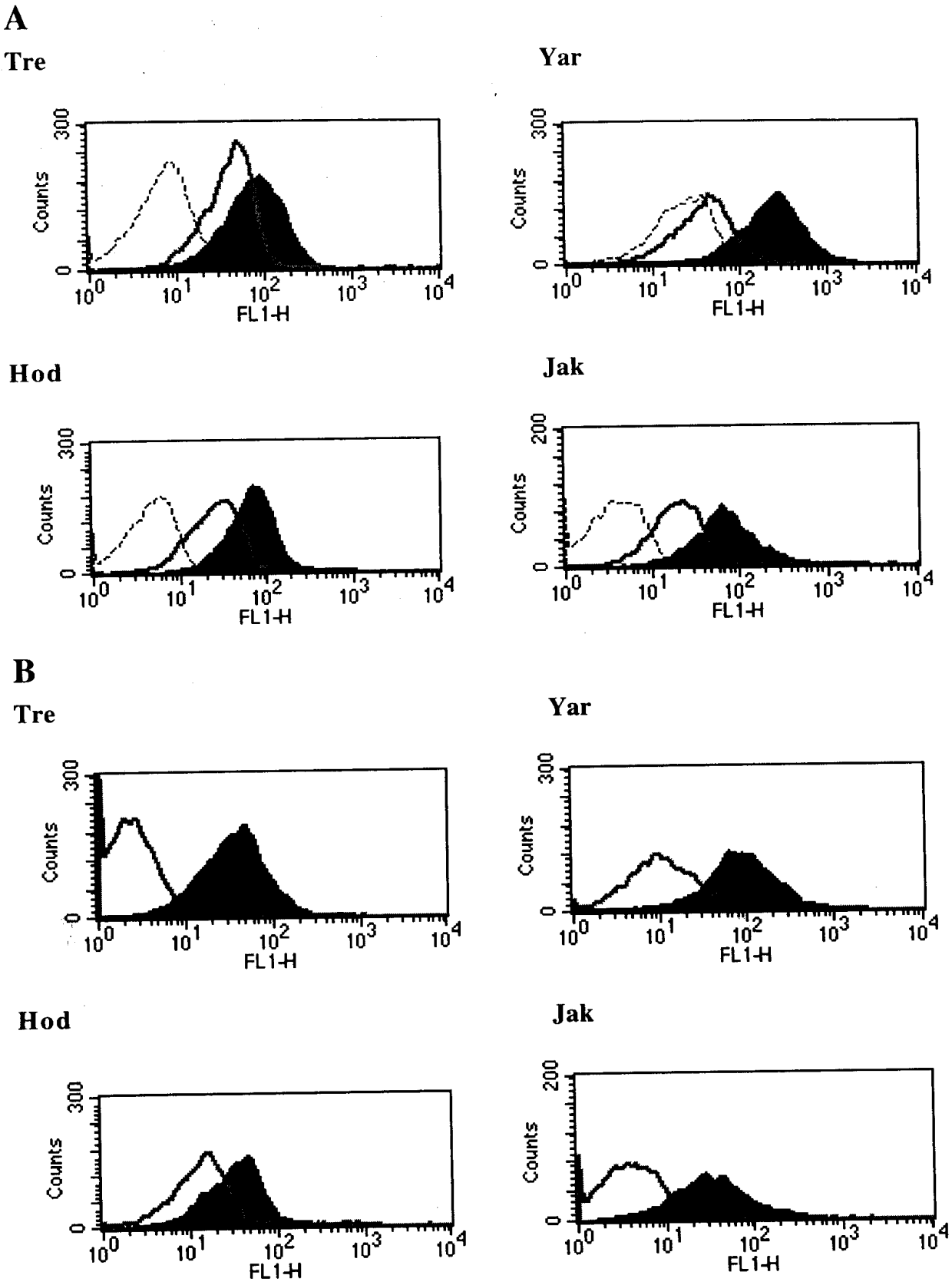


Figure A.6: Isotyping of membrane bound Ig of B CLL cells. Patients are indicated in top left hand corner of histograms. A: L chain isotyping histograms: sheep polyclonal-FITC control (dashed lines); α-λ-FITC (grey lines) and α-κ-FITC (solid grey). B: H chain isotyping histograms. α-γ-FITC (grey lines); α-μ-FITC (solid grey).

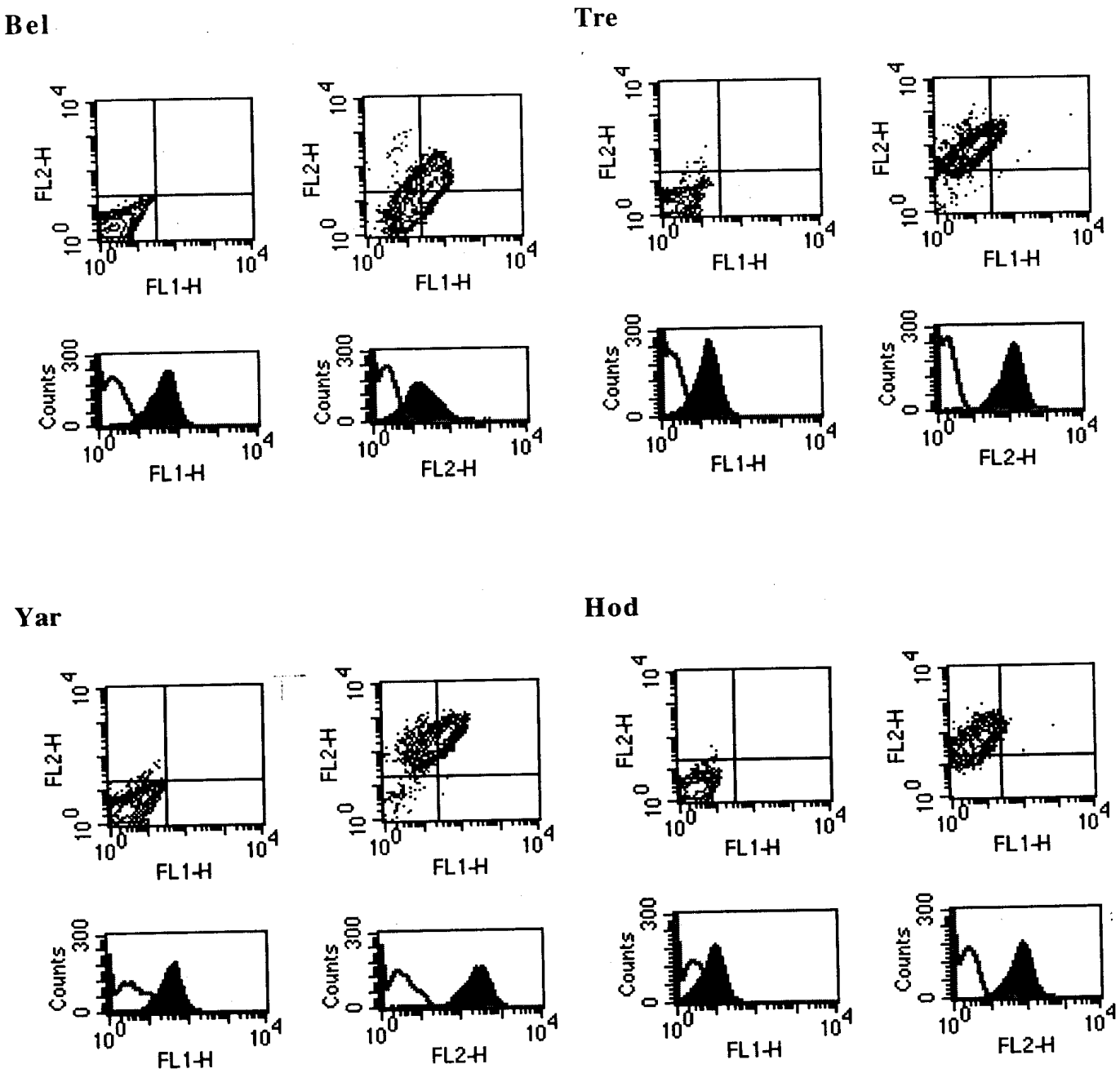


Figure A.7: Two-colour fluorescence staining of B CLL cells. Patients are indicated at top left corner of panels. Contour plots: left, isotype matched controls and right, CD19-FITC (FL1) *versus* CD5-PE (FL2) antibodies. Histograms show isotype matched controls (grey lines) and CD19 or CD5 antibodies (solid grey)

Appendix B

HCDR3 templates, Hod 1 and Hod 2 models and Ramachandran plots of refined models

Table B.1

Heavy chain CDR3 from X-ray structures of unliganded immunoglobulin fragments

PDB code	HCDR3 sequence	Length
1BBD	YYSYYDMDY	9
1BBJ	SYYGH	5
1CGS	GYSSMDY	7
1DBA	GDYVNWYFDV	10
1DFB	GRDYYDSGGYFTVAFDI	17
1FAI	SFYGSDLAVYYFDS	14
1FGV	WRGLDVRYFDV	11
1FIG	RRDGNYGFTY	10
1FOR	SGNYPYAMDY	10
1FVC	WGGDGFYAMDY	11
1GGB	EGYIY	5
1GIG	DFYDYDVFYYAMDY	14
1HIL	RERYDENGFAY	11
1IGF	YSSDPFYFDY	10
1IGI	SSGNKWAMDY	10
1IGM	HRVSYVLTGFDS	12
1MAM	DPYGPAAY	8
1MCP	NYYGSTWYFDV	11
1MRC	LRGYFDY	7
1NBV	DQTGTAWFAY	10
1VFA	ERDYRLDY	8
2F19	SFYGGSDLAVYYFDS	15
2FB4	DGGHGFCSSASCFGPDY	17
2FGW	WRRYFDV	7
2IG2	DGGHGFCSSASCFGPDY	17
6FAB	SEYYGGSYKFDY	12
7FAB	NLIAGGIDV	9
8FAB	DPDILTAFSFDY	12
1FVE	WGGDGFYAMDY	11
NC10.14	RTFSYYYGSSFYYFDN	16

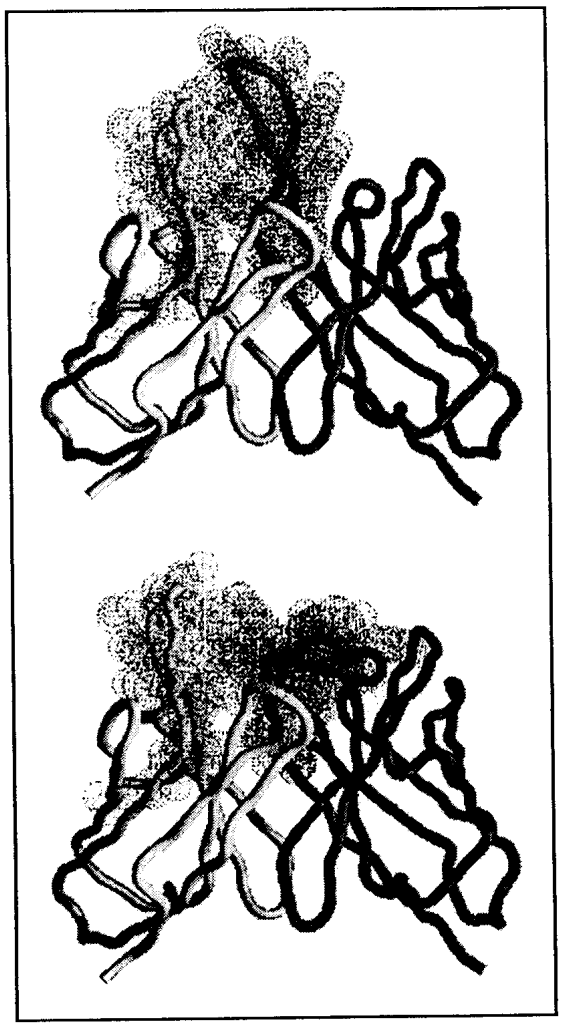


Figure B.1: Alternative HCDR3 models of Hod Fv. The VH regions are shown in dark grey and the VL regions in light grey as Cα traces. The vdW radii are shown as dots for residues of HCDR3 (dark grey) and LCDR1 (light grey). Hod 1 (top) used the 1DFB HCDR3 and Hod 2 (bottom) used the 1IGM HCDR3 as templates to predict the conformation of the Hod HCDR3 loop.

Table B.2

Interchain interaction energies of two alternative Hod Fv models

Interaction	E-total[a]	E-vdW[b]	E-ELEC[c]
Hod 1[d]			
V_L-V_H	-87.8	-85.4	-2.4
V_L-HCDR3	-22.0	-50.2	28.1
Hod 2[d]			
V_L-V_H	-89.5	-72.8	-16.7
V_L-HCDR3	-36.7	-35.8	-0.919

[a] All energy values are calculated for interchain interactions only using X-PLOR (Brunger, 1992).

[b] van der Waals energy term.

[c] Electrostatic energy term.

[d] The HCDR3 loops were constructed using either the 17-residue loop from 1DFB (Hod 1) or the 12-residue loop from 1IGM (Hod 2) to predict the conformation of the loop stems. These stems were closed with a β-turn after sufficient residues had been added to complete the HCDR3 of Hod.

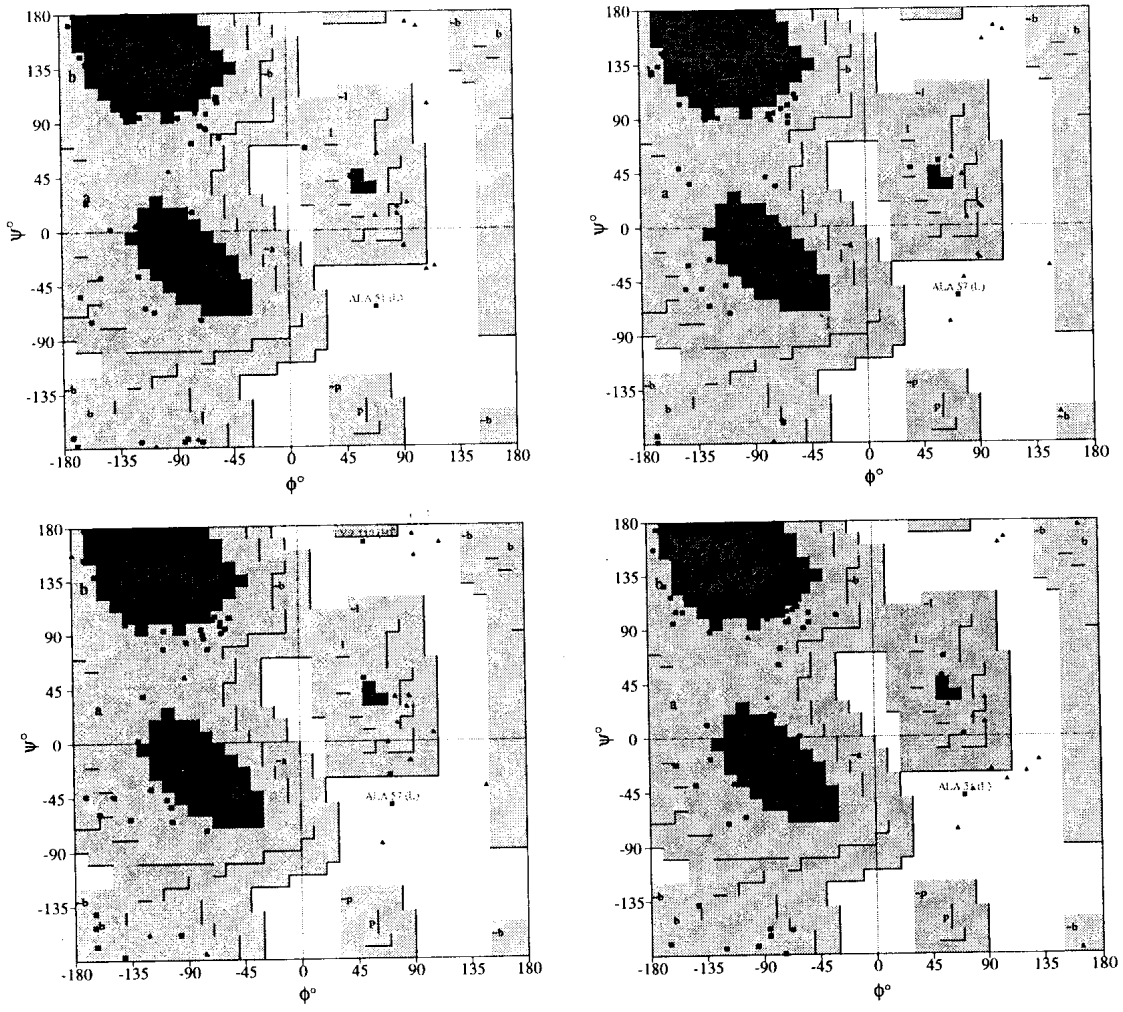


Figure B.2: Ramachandran plots of ψ,φ angles for all Cα atoms in refined Fv models. Residues are shown as squares except for glycines which are shown as triangles. Regions are labelled: A, B, L for core; a,b,l,p for allowed and ~a,~b,~l,~p for generously allowed. Plots: top left (Tre); top right (Yar 2); bottom left (Hod)

Appendix C
Calibration of size-exclusion column

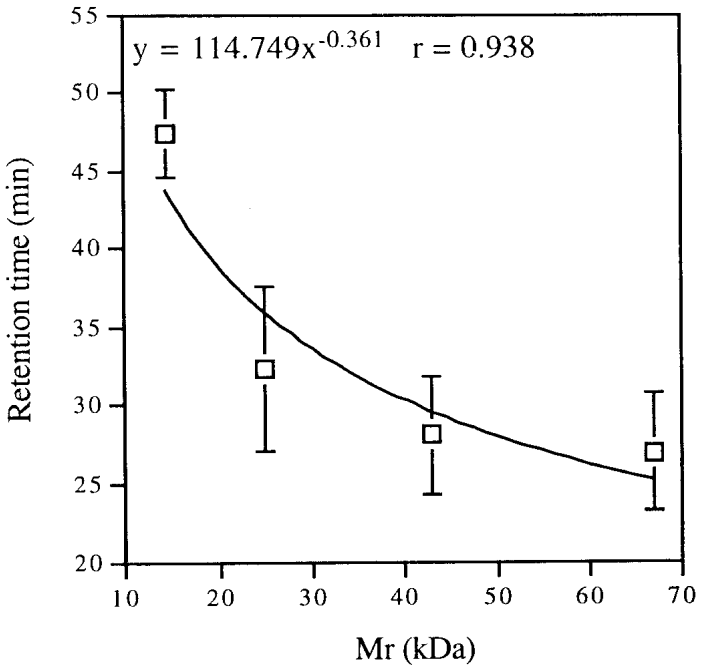


Figure C.1: Calibration of superdex 75 size-exclusion column. Geometric mean retention times and ranges are plotted for proteins: bovine serum albumin (67 kDa), chymotrypsin (25 kDa), ovalbumin (43 kDa) and lysozyme (14.4 kDa). Aldolase (158 kDa) eluted in a peak with a geometric mean of 24.59 min (Vo) and represents the Mr cutoff limit of the column. The curve fitted to the data is indicated.

References

Adib, M., Ragimbeau, J., Avrameas, S. and Ternynck, T. (1990). IgG autoantibody activity in normal mouse serum is controlled by IgM. *J. Immunol.* **145**, 3807-3813.

Adib-Conquy, M., Gilbert, M., Christodoulou, C. and Avrameas, S. (1994). Reactivity and structure of a mouse anti-F(ab')2 IgM. Comparison of its variable region sequences with those of a structurally close polyreactive natural IgM. *Mol. Immunol.* **31**, 555-562.

Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J. (1986). Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* **233**, 747-753.

Arevalo, J.H., Hassig, C.A., Stura, E.A., Sims, M.J., Taussig, M.J. and Wilson, I.A. (1994). Structural analysis of antibody specificity: Detailed comparison of five Fab'-steroid complexes. *J. Mol. Biol.* **241**, 663-690.

Arevalo, J.H., Taussig, M.J. and Wilson, I.A. (1993). Molecular basis of crossreactivity and the limits of antibody-antigen complementarity. *Nature* **365**, 859-863.

Avrameas, S. and Ternynck, T. (1993). The natural autoantibodies system: Between hypotheses and facts. *Mol. Immunol.* **30**, 1133-1142.

Axelrod, O., Silverman, G.J., Dev, V., Kyle, R., Carson, D.A. and Kipps, T.J. (1991). Idiotypic cross-reactivity of immunoglobulins expressed in Waldenström's macroglobulinemia, chronic lymphocytic leukemia, and mantle zone lymphocytes of secondary B-cell follicles. *Blood* **77**, 1484-1490.

Baccala, R., Quang, T.V., Gilbert, M., Ternynck, T. and Avrameas, S. (1989). Two murine natural polyreactive autoantibodies are encoded by nonmutated germ-line genes. *PNAS* **86**, 4624-4628.

Bajorath, J. (1994). Three-dimensional model of the BR96 monoclonal antibody variable fragment. *Bioconj. Chem.* **5**, 213-219.

Bajorath, J. and Fine, R.M. (1992). On the use of minimization from many randomly generated loop structures in modeling antibody combining sites. *Immunomethods* **1**, 137-146.

Bajorath, J. and Sheriff, S. (1996). Comparison of an antibody model with an X-ray structure: The variable fragment of BR96. *Proteins: Struct. Func. and Genet.* **24**, 152-157.

Bajorath, J., Stenkamp, R. and Aruffo, A. (1993). Knowledge-based model building of proteins: Concepts and examples. *Protein Sci.* **2**, 1798-1810.

Barbouche, M.R., Guilbert, B., Makni, S., Gorgi, Y., Ayed, K. and Avrameas, S. (1993). Common idiotypes expressed on human, monoclonal, abnormal immunoglobulins and cryoglobulins with polyreactive autoantibody activities. *Clin. Exp. Immunol.* **91**, 196-201.

Barbouche, R., Forveille, M., Fischer, A., Avrameas, S. and Durandy, A. (1992). Spontaneous IgM autoantibody production in vitro by B lymphocytes of normal human neonates. *Scand. J. Immunol.* **35**, 659-667.

Barry, M.M., Mol, C.D., Anderson, W.F. and Lee, J.S. (1994). Sequencing and modeling of anti-DNA immunoglobulin Fv domains: Comparison with crystal structures. *J. Biol. Chem.* **269**, 3623-3632.

Bassolinò-Klimas, D., Bruccoleri, R.E. and Subramaniam, S. (1992). Modeling the antigen combining site of an anti-dinitrophenyl antibody, ANO2. *Protein Sci.* **1**, 1465-1476.

Benjamin, D.C., Williams, D.C., Smith-Gill, S.J. and Rule, G.S. (1992). Long-range changes in a protein antigen due to antigen-antibody interaction. *Biochemistry* **31**, 9539-9545.

Berneman, A., Ternynck, T. and Avrameas, S. (1992). Natural mouse IgG reacts with self antigens including molecules involved in the immune response. *Eur. J. Immunol.* **22**, 625-633.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Betz, A.G., Rada, C., Pannell, R., Milstein, C. and Neuberger, M.S. (1993). Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: Clustering, polarity, and specific hot spots. *PNAS* **90**, 2385-2388.

Bhat, T.N., Bentley, G.A., Boulot, G., Greene, M.I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F.P., Mariuzza, R.A. and Poljak, R.J. (1994). Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *PNAS* **91**, 1089-1093.

Bhat, T.N., Bentley, G.A., Fischmann, T.O., Boulot, G. and Poljak, R.J. (1990). Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. *Nature* **347**, 483-485.

Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E. and Thornton, J.M. (1987). Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326**, 347-352.

Bork, P., Holm, L. and Sander, C. (1994). The Immunoglobulin Fold: Structural Classification, Sequence Patterns and Common Core. *J. Mol. Biol.* **242**, 309-320.

Bossart-Whitaker, P., Chang, C.Y., Novotny, J., Benjamin, D.C. and Sheriff, S. (1995). The crystal structure of the antibody N10-staphylococcal nuclease complex at 2.9 Å resolution. *J. Mol. Biol.* **253**, 559-575.

Boux, H.A., Raison, R.L., Walker, K.Z., Hayden, G.E. and Basten, A. (1983). A tumor-associated antigen specific for human kappa myeloma cells. *J. Exp. Med.* **158**, 1769-1774.

Braden, B.C., Souchon, H., Eisele, J.-L., Bentley, G.A., Bhat, T.N., Navaza, J. and Poljak, R.J. (1994). Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *J. Mol. Biol.* **243**, 767-781.

Brock, C.R. (1995) The cloning and sequencing of the polyspecific IgM molecule on the surface of CLL cells. MSc Thesis, University of Technology, Sydney.

Brodeur, P.H., Osman, G.E., Mackle, J.J. and Lalor, T.M. (1988). The organization of the mouse IgH-V locus. *J. Exp. Med.* **168**, 2261-2278.

Broker, B.M., Klajman, A., Youinou, P., Jouquan, J., Worman, C.P., Murphy, J., Mackenzie, L., Quartey-Papafio, R., Blaschek, M., Collins, P., Lal, S. and Lydyard, P.M. (1988). Chronic lymphocytic leukemia (CLL) cells secrete multispecific autoantibodies. *J. Autoimmun.* **1**, 469-481.

Brown, C.M.S., Longhurst, C., Haynes, G., Plater-Zyberk, C., Malcolm, A. and Maini, R.N. (1992). Immunoglobulin heavy chain variable region gene utilization by B cell hybridomas derived from rheumatoid synovial tissue. *Clin. Exp. Immunol.* **89**, 230-238.

Bruccoleri, R.E. and Karplus, M. (1987). Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* **26**, 137-168.

Brunger, A.T. (1992). X-PLOR v.3.1. A System for X-ray Crystallography and NMR. New Haven, CT, U.S.A.: Yale University Press.

Buchner, J., Pastan, I. and Brinkmann, U. (1992). A method for increasing the yield of properly folded recombinant fusion proteins: Single-chain immunotoxins from renaturation of bacterial inclusion bodies. *Anal. Biochem.* **205**, 263-270.

Burastero, S.E., Casali, P., Wilder, R.L. and Notkins, A.L. (1988). Monoreactive high affinity and polyreactive low affinity rheumatoid factors are produced by $CD5^+$ B cells from patients with rheumatoid arthritis. *J. Exp. Med.* **168**, 1979-1992.

Burastero, S.E., Cutolo, M., Dessi, V. and Celada, F. (1990). Monoreactive and polyreactive rheumatoid factors produced by in vitro epstein-barr virus-transformed peripheral blood and synovial B lymphocytes from rheumatoid arthritis patients. *Scan. J. Immunol.* **32**, 347-357.

Caligaris-Cappio, F. (1996). B-chronic lymphocytic leukemia: a malignancy of anti-self B cells. *Blood* **87**, 2615-2620.

Caligaris-Cappio, F., Gottardi, D., Alfarano, A., Stacchini, A., Gregoretti, M.G., Ghia, P., Bertero, M.T., Novarino, A. and Bergui, L. (1993). The nature of the B lymphocyte in B-chronic lymphocytic leukemia. *Blood Cells* **19**, 601-613.

Chacko, S., Silverton, E., Kam-Morgan, L., Smith-Gill, S., Cohen, G. and Davies, D. (1995). Structure of an antibody-lysozyme complex unexpected effect of a conservative mutation. *J. Mol. Biol.* **245**, 261-274.

Chen, Z.J., Wheeler, J. and Notkins, A.L. (1995). Antigen-binding B cells and polyreactive antibodies. *Eur. J. Immunol.* **25**, 579-586.

Cheung, S.C., Takeda, S. and Notkins, A.L. (1995). Both VH and VL chains of polyreactive IgM antibody are required for polyreactivity: expression of Fab in *Escherichia Coli*. *Clin. Exp. Immunol.* **101**, 383-386.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Chothia, C. and Lesk, A.M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901-917.

Chothia, C., Lesk, A.M., Gherardi, E., Tomlinson, I.M., Walter, G., Marks, J.D., Llewelyn, M.B. and Winter, G. (1992). Structural repertoire of the human VH segments. *J. Mol. Biol.* **227**, 799-817.

Chothia, C., Lesk, A.M., Levitt, M., Amit, A.G., Mariuzza, R.A., Phillips, E.V. and Poljak, R.J. (1986). The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. *Science* **233**, 755-758.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M. and Poljak, R.J. (1989). Conformations of immunoglobulin hypervariable regions. *Nature* **342**, 877-883.

Chothia, C., Novotny, J., Bruccoleri, R. and Karplus, M. (1985). Domain association in immunoglobulin molecules: the packing of variable domains. *J. Mol. Biol.* **186**, 651-663.

Colman, P.M., Laver, W.G., Varghese, J.N., Baker, A.T., Tulloch, P.A., Air, G.M. and Webster, R.G. (1987). Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. *Nature* **326**, 358-362.

Colman, P.M., Schramm, H.J. and Guss, J.M. (1977). Crystal and molecular structure of the dimer of variable domains of the Bence-Jones protein ROY. *J. Mol. Biol.* **116**, 73-79.

Cook, G.P. and Tomlinson, I.M. (1995). The human immunoglobulin VH repertoire. *Immunol. Today* **16**, 237-242.

Corbett, S.J., Tomlinson, I.M., Sonnhammer, E.L.L., Buck, D. and Winter, G. (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.

Corper, A.L., Sohi, M.K., Bonagura, V.R., Steinitz, M., Jefferis, R., Feinstein, A., Beale, D., Taussig, M.J. and Sutton, B.J. (1997). Structure of human IgM rheumatoid factor Fab bound to its autoantigen IgG Fc reveals a novel topology of antibody-antigen interaction. *Nature: Struc. Biol.* **4**, 374-381.

Coutinho, A., Kazatchkine, M.D. and Avrameas, S. (1995). Natural autoantibodies. *Curr. Opin. Immunol.* **7**, 812-818.

Cox, D.W., Markovic, V.D. and Teshima, I.E. (1982). Genes for immunoglobulin heavy chains and for α1-antitrypsin are localized to specific regions of chromosome 14q. *Nature* **297**, 428-430.

Cox, J.P.L., Tomlinson, I.M. and Winter, G. (1994). A directory of human germ-line Vκ segments reveals a strong bias in their usage. *Eur. J. Immunol.* **24**, 827-836.

Davies, D.R. and Chacko, S. (1993). Antibody Structure. *Acc. Chem. Res.* **26**, 421-427.

de la Chapelle, A., Lenoir, A., Boue, G., Gallano, P., Huerre, C., Szajnert, M.F., Jeanpierre, M., Lalouel, M. and Kaplan, J.C. (1983). Lambda Ig constant region genes are translocated to chromosome 8 in a Burkitt's lymphoma (8,22). *Nucl. Acids Res.* **11**, 1133-1142.

de la Paz, P., Sutton, B.J., Darsley, M.J. and Rees, A.R. (1986). Modelling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope. *EMBO J.* **5**, 415-425.

Deane, M. and Norton, J.D. (1990). Immunoglobulin heavy chain variable region family usage is independent of tumor cell phenotype in human B lineage leukemias. *Eur. J. Immunol.* **20**, 2209-2217.

Ditzel, H.J., Itoh, K. and Burton, D.R. (1996). Determinants of polyreactivity in a large panel of recombinant human antibodies from HIV-1 infection. *J. Immunol.* **157**, 739-749.

Dunbrack, R.L., Gerloff, D.L., Bower, M., Chen, X., Lichtarge, O. and Cohen, F.E. (1997). Meeting review: the second meeting of the critical assessment of techniques for protein structure prediction (CASP2), Asilomar, California, December 13-16, 1996. *Folding and Design* **1**, R27-42.

Edmundson, A.B., Ely, K.R., Abola, E.E., Schiffer, M. and Panagiotopoulos, N. (1975). Rotational allomerism and divergent evolution of domains in immunoglobulin light chains. *Biochemistry* **14**, 3953-3961.

Edmundson, A.B., Ely, K.R., Girling, R.L., Abola, E.E., Schiffer, M. and Westholm, F.A. (1974) Structure and binding properties of a λ-type Bence-Jones dimer. In *Progress in Immunology II* (Edited by L. Brent & J. Holborow), p. 103-113. North-Holland Publishing Company, Amsterdam.

Edmundson, A.B., Guddat, L.W. and Andersen, K.N. (1993). Crystal structures of intact IgG antibodies. *Immunomethods* **3**, 197-210.

Efremov, D.G., Ivanovski, M., Siljanovski, N., Pozzato, G., Cevreska, L., Fais, F., Chiorazzi, N., Batista, F.D. and Burrone, O.R. (1996). Restricted immunoglobulin VH region repertoire in chronic lymphocytic leukemia patients with autoimmune hemolytic anemia. *Blood* **87**, 3869-3876.

Eisele, J.-L., Boulot, G., Chitarra, V., Riottot, M.-M., Souchon, H., Houdusse, A., Bentley, G.A., Bhat, T.N., Spinelli, S. and Poljak, R.J. (1992). Cloning, bacterial expression and crystallization of Fv antibody fragments. *J. Cryst. growth* **122**, 337-343.

Engh, R.A. and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Cryst.* **A47**, 392-400.

Fan, Z.-C., Shan, L., Guddat, L.W., He, X.-M., Gray, W.R., Raison, R.L. and Edmundson, A.B. (1992). Three-dimensional structure of and Fv from a human IgM immunoglobulin. *J. Mol. Biol.* **228**, 188-207.

Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Wayne, M.M. and Winter, G. (1985). Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**, 235-238.

Fields, B.A., Goldbaum, F.A., Dall'Acqua, W., Malchiodi, E.L., Cauerhff, A., Schwarz, F.P., Ysern, X., Poljak, R.J. and Mariuzza, R.A. (1996). Hydrogen bonding and solvent structure in an antigen-antibody interface. Crystal structures and thermodynamic characterization of three Fv mutants complexed with lysozyme. *Biochemistry* **35**, 15494-15503.

Foote, J. and Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.* **224**, 487-499.

Friedman, D.F., Moore, J.S., Erikson, J., Manz, J., Goldman, J., Nowell, P.C. and Silberstein, L.E. (1992). Variable region gene analysis of an isotype-switched (IgA) variant of chronic lymphocytic leukemia. *Blood* **80**, 2287-2297.

Froyen, G., Ronsse, I. and Billiau, A. (1993). Bacterial expression of a single-chain antibody fragment (scFV) that neutralizes the biological activity of human interferon-γ. *Mol. Immunol.* **30**, 805-812.

Gilson, M.K. (1995). Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.* **5**, 216-223.

Glockshuber, R., Malia, M., Pfitzinger, I. and Pluckthun, A. (1990). A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry* **29**, 1362-1367.

Glockshuber, R., Schmidt, T. and Pluckthun, A. (1992). The disulfide bonds in antibody variable domains: Effects on stability, folding in vitro, and functional expression in Escherichia coli. *Biochemistry* **31**, 1270-1279.

Goodnow, C.C., Crosbie, J., Adelstein, S., Lavoie, T.B., Smith-Gill, S.J., Brink, R.A., Pritchard-Briscoe, H., Wotherspoon, J.S., Loblay, R.H., Raphael, K., Trent, R.J. and Basten, A. (1988). Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature* **334**, 676.

Goodnow, C.C., Crosbie, J., Jorgensen, H., Brink, R.A. and Basten, A. (1989). Induction of self-tolerance in mature peripheral B lymphocytes. *Nature* **342**, 385.

Goto, Y. and Hamaguchi, K. (1982). Unfolding and refolding of the constant fragment of the immunoglobulin light chain. *J. Mol. Biol.* **156**, 891-910.

Guddat, L.W., Shan, L., Anchin, J.M., Linthicum, D.S. and Edmundson, A.B. (1994). Local and transmitted conformational changes on complexation of an anti-sweetener Fab. *J. Mol. Biol.* **236**, 247-274.

Hansen, A., Jahn, S., Lukowsky, A., Grutz, G., Bohn, J., von Baehr, R. and Settmacher, U. (1994). VH/VL gene expression in polyreactive-antibody-producing human hybridomas from the fetal B cell repertoire. *Exp. Clin. Immunogenet.* **11**, 1-16.

Hartman, A.B., Mallett, C.P., Srinivasappa, J., Prabhakar, B.S., Notkins, A.L. and Smith-Gill, S.J. (1989). Organ reactive autoantibodies from non-immunized adult BALB/c mice are polyreactive and express non-biased VH gene usage. *Mol. Immunol.* **26**, 359-370.

Hartman, A.B., Mallett, C.P., Srinivasappa, J., Prabhakar, B.S., Notkins, A.L. and Smith-Gill, S.J. (1988). VH family usage and binding analyses of polyreactive monoclonal autoantibodies derived from nonimmunized adult BALB/c mice. *Autoimmunity* **137**, 211-215.

He, X.M., Ruker, F., Casale, E. and Carter, D.C. (1992). Structure of a human monoclonal antibody Fab fragment against gp41 of human immunodeficiency virus type 1. *PNAS* **89**, 7154-7158.

Herron, J.N., He, X.-M., Mason, M.L., Voss, E.W. and Edmundson, A.B. (1989). Three-dimensional structure of a fluorescein-Fab complex crystallized in 2-methyl-2,4-pentanediol. *Proteins: Stuct. Func. Genet.* **5**, 271-280.

Herron, J.N., He, X.M., Ballard, D.W., Blier, P.R., Pace, P.E., Bothwell, A.L.M., Voss, E.W. and Edmundson, A.B. (1991). An autoantibody to single-stranded DNA: Comparison of the three-dimensional structures of the unliganded Fab and a deoxynucleotide-Fab complex. *Proteins: Stuct. Func. Genet.* **11**, 159-175.

Herzenberg, L.A. and Kantor, A.B. (1993). B-cell lineages exist in the mouse. *Immunol.Today* **14**, 79-91.

Houdayer, M., Bouvet, J.-P., Wolff, A., Magnac, C., Guillemot, J.-C., Borche, L. and Dighiero, G. (1993). Simultaneous presence, in one serum, of four monoclonal

antibodies that might correspond to different steps in a clonal evolution from polyreactive to monoreactive antibodies. *J. Immunol.* **150**, 311-319.

Humphries, C.G., Shen, A., Kuziel, W.A., Capra, J.D., Blattner, F.R. and Tucker, P.W. (1988). A new human immunoglobulin VH family preferentially rearranged in immature B-cell tumours. *Nature* **331**, 446-449.

Ichiyoshi, Y. and Casali, P. (1994). Analysis of the structural correlates for antibody polyreactivity by multiple reassortments of chimeric human immunoglobulin heavy and light chain V segments. *J. Exp. Med.* **180**, 885-895.

Ikematsu, H., Kasaian, M.T., Schettino, E.W. and Casali, P. (1993). Structural analysis of the VH-D-JH segments of human polyreactive IgG mAb. *J. Immunol.* **151**, 3604-3616.

Ikematsu, W., Ikematsu, H., Okamura, S., Otsuka, T., Harada, M. and Niho, Y. (1994). Surface phenotype and Ig heavy-chain usage in chronic B-cell leukemias: Expression of myelomonocytic surface markers in $CD5^-$ chronic B-cell leukemia. *Blood* **83**, 2602-2610.

Kabat, E.A. and Wu, T.T. (1971). Attempts to locate complementarity determining residues in the variable positions of light and heavy chains. *Ann. N. Y. Acad. Sci.* **190**, 382-393.

Kabat, E.A., Wu, T.T. and Bilofsky, H. (1977). Unusual distribution of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody combining sites. *J. Biol. Chem.* **252**, 6609-6616.

Kabat, E.A., Wu, T.T. and Bilofsky, H. (1978). Variable region genes for the immunoglobulin framework are assembled from small segments of DNA-a hypothesis. *PNAS* **75**, 2429-2433.

Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. and Foeller, C. (1991). Sequences of proteins of immunological interest (5th ed.). Bethesda, MD: U.S Department of Health and Human Services, Public Health Service, National Institutes of Health.

Kahn, S., Kahn, M. and Eisen, H. (1992). Polyreactive autoantibodies to negatively charged epitopes following *Trypanosoma cruzi* infection. *Eur. J. Immunol.* **22**, 3051-3056.

Kipps, T.J. (1989). The CD5 B cell. *Adv. Immunol.* **47**, 117-163.

Kipps, T.J. (1993). Immunoglobulin genes in chronic lymphocytic leukemia. *Blood Cells* **19**, 615-625.

Kipps, T.J. and Carson, D.A. (1993). Autoantibodies in chronic lymphocytic leukemia and related systemic autoimmune diseases. *Blood* **81**, 2475-2487.

Kipriyanov, S.M., Moldenhauer, G., Martin, A.C.R., Kupriyanova, O.A. and Little, M. (1997). Two amino acid mutations in an anti-human CD3 single chain Fv antibody fragment that affect the yield on bacterial secretion but not affinity. *Protein Eng.* **10**, 445-453.

Knappik, A., Krebber, C. and Pluckthun, A. (1993). The effect of folding catalysts on the in vivo folding process of different antibody fragments expressed in Escherichia coli. *Bio/Technology* **11**, 77-83.

Kortt, A.A., Lah, M., Oddie, G.W., Gruen, C.L., Burns, J.E., Pearce, L.A., Atwell, J.L., McCoy, A.J., Howlett, G.J., Metzger, D.W., Webster, R.G. and Hudson, P.J. (1997). Single-chain Fv fragments of anti-neuraminidase antibody NC10 containing five- and ten-residue linkers form dimers and with zero-residue linker a trimer. *Protein Eng.* **10**, 423-433.

Kortt, A.A., Malby, R.L., Caldwell, B., Gruen, L.C., Ivancic, N., Lawrence, M.C., Howlett, J.J., Webster, R.G., J., H.P. and Colman, P.M. (1994). Recominant anti-sialidase single-chain variable fragment antibody. *J. Biochem.* **221**, 151-157.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lascombe, M.-B., Alzari, P.M., Poljak, R.J. and Nisonoff, A. (1992). Three-dimensional structure of two crystal forms of FabR19.9 from a monoclonal anti-arsonate antibody. *PNAS* **89**, 9429-9433.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

Levitt, M. and Perutz, M.F. (1988). Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* **201**, 751-754.

Lilie, H., McLaughlin, S., Freedman, R. and Buchner, J. (1994). Influence of protein disulfide isomerase (PDI) on antibody folding in vitro. *J. Biol. Chem.* **269**, 14290-14296.

Locniskar, M., Zumla, A., Mudd, D.W., Isenberg, D.A., Williams, W. and Mcadam, K.P.W.J. (1988). Human monoclonal antibodies to phenolic glycolipid-I derived from patients with leprosy, and production of specific anti-idiotypes. *Immunology* **64**, 245-251.

Logtenberg, T., Kroon, A., Gmelig-Meyling, F.H.J. and Ballieux, R.E. (1987). Analysis of the human tonsil B cell repertoire by somatic hybridization: occurrence of both "monospecific" and "multispecific" (auto)antibody-secreting cells. *Eur. J. Immunol.* **17**, 855-859.

Lydyard, P.M., Quartey-Papafio, R., Broker, B., Mackenzie, L., Jouquan, J., Blaschek, M.A., Steele, J., Petrou, M., Collins, P., Isenberg, D. and Youinou, P.Y. (1990). The antibody repertoire of early human B cells. I: High frequency of autoreactivity and polyreactivity. *Scand. J. Immunol.* **31**, 33-43.

Mackenzie, L.E., Youinou, P.Y., Hicks, R., Yuksel, B., Mageed, R.A. and Lydyard, P.M. (1991). Auto- and polyreactivity of IgM from CD5+ and CD5- cord blood B cells. *Scand. J. Immunol.* **33**, 329-335.

Malby, R.L., Caldwell, J.B., Gruen, L.C., Harley, V.R., Ivancic, N., Kortt, A.A., Lilley, G.G., Power, B.E., Webster, R.G., Colman, P.M. and Hudson, P.J. (1993). Recombinant antineuraminidase single chain antibody: Expression, characterization, and crystallization in complex with antigen. *Proteins: Stuct. Func. Genet.* **16**, 57-63.

Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M.A., Bentley, D.L. and Rabbitts, T.H. (1982). Localization of human immunoglobulin κ light chain variable region genes to the short arm of chromosome 2 by in situ hybridization. *PNAS* **79**, 4957-4961.

Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1991). Molecular modeling of antibody combining sites. *Meth. Enzymol.* **203**, 121-153.

Martin, T., Crouzier, R., Weber, J.-C., Kipps, T.J. and Pasquali, J.-L. (1994). Structure-function studies of a polyreactive (natural) autoantibody. *J. Immunol.* **152**, 5988-5996.

Morrison, S.L. (1992). IN VITRO ANTIBODIES: Strategies for production and application. *Ann. Rev. Immunol.* **10**, 239-265.

Murphy, J.P., Duggleby, C.J., Atkinson, M.A., Trowern, A.R., Atkinson, T. and Goward, C.R. (1994). The functional units of a peptostreptococcal protein L. *Mol. Micro.* **12**, 911-920.

Neuberger, M.S. and Milstein, C. (1995). Somatic hypermutation. *Curr. Opin. Immunol.* **7**, 248-254.

Newkirk, M.M., Rauch, J., Mageed, R.A.K., Jefferis, R., Posnett, D.N. and Silverman, G.J. (1993). Restricted immunoglobulin variable region gene usage by hybridoma rheumatoid factors from patients with systemic lupus erythematosus and rheumatoid arthritis. *Mol. Immunol.* **30**, 255-263.

Nicholls, A., Bharadwaj, R. and Honig, B. (1993). GRASP: graphical representation and analysis of surface properties. *Biophys. J.* **64**, 166-170.

Novotny, J., Bruccoleri, R.E. and Saul, F.A. (1989). On the attribution of binding energy in antigen-antibody complexes McPc603, D1.3 and HyHEL-5. *Biochemistry* **28**, 4735-4749.

Okawa-Takatsuji, M., Aotsuka, S., Uwatoko, S., Sumiya, M. and Yokohari, R. (1992). The B cell repertoire in patients with systemic autoimmune diseases: analysis of Epstein-Barr virus (EBV)-inducible circulating precursors that produce autoantibodies against nuclear ribonucleoprotein (nRNP). *Clin. Exp. Immunol.* **90**, 415-421.

Padlan, E.A. (1994). Anatomy of the antibody molecule. *Mol. Immunol.* **31**, 169-217.

Padlan, E.A. (1990). On the nature of antibody combining sites: Unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins: Stuct. Func. Genet.* **7**, 112-124.

Padlan, E.A. and Davies, D.R. (1975). Variability of three-dimensional structure in immunoglobulins. *PNAS* **72**, 819-823.

Padlan, E.A., Silverton, E.W., Sheriff, S., Cohen, G.H., Smith-Gill, S.J. and Davies, D.R. (1989). Structure of an antibody-antigen complex: crystal structure of the HyHEL-10 Fab-lysozyme complex. *PNAS* **86**, 5938-5942.

Pascual, V., Victor, K., Randen, I., Thompson, K., Steinitz, M., Forre, O., Fu, S.-M., Natvig, J.B. and Capra, J.D. (1992). Nucleotide sequence analysis of rheumatoid factors and polyreactive antibodies derived from patients with rheumatoid arthritis reveals diverse use of VH and VL gene segments and extensive variability in CDR-3. *Scand. J. Immunol.* **36**, 349-362.

Pluckthun, A. and Skerra, A. (1989). Expression of functional antibody Fv and Fab fragments in Escherichia coli. *Meth. Enzymol.* **178**, 497-515.

Poljak, R.J., Amzel, L.M., Avey, H.P., Chen, B.L., Phizackerley, R.P. and Saul, F. (1973). Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8-Å resolution. *PNAS* **70**, 3305-3310.

Powell, M.J.D. (1977). Restart procedures for the conjugate gradient method. *Math. Prog.* **12**, 241-254.

Pritsch, O., Magnac, C., Dumas, G., Egile, C. and Dighiero, G. (1993). V gene usage by seven hybrids derived from CD5+ B-cell chronic lymphocytic leukemia and displaying autoantibody activity. *Blood* **82**, 3103-3112.

Ramachandran, G.N. and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23**, 283-437.

Ramsland, P.A. (1993) Molecular characterisation of polyreactive IgM binding. BSc (Hon) Thesis, University of Technology, Sydney.

Ramsland, P.A., Guddat, L.W., Edmundson, A.B. and Raison, R.L. (1997). Diverse binding site structures revealed in homology models of polyreactive immunoglobulins. *JCAMD* **11**, 453-461.

Sali, A. (1995). Modelling mutations and homologous proteins. *Curr. Opin. Biotech.* **6**, 437-451.

Sali, A. and Blundell, T.L. (1993). Comparative modelling by satisfaction of spacial restraints. *J. Mol. Biol.* **234**, 779-815.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning: A laboratory Manual (2nd Ed.). Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Sanchez, R. and Sali, A. (1997). Advances in comparative protein-structure modelling. *Curr. Opin. Struct. Biol.* **7**, 206-214.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *PNAS* **74**, 5463-5467.

Sanz, I., Casali, P., Thomas, J.W., Notkins, A.L. and Capra, J.D. (1989). Nucleotide sequences of eight human natural autoantibody VH regions reveals apparent restricted use of VH families. *J. Immunol.* **142**, 4054-4061.

Sawyer, J.R., Schlom, J. and Kashmiri, S.V.S. (1994). The effects of induction conditions on production of a soluble anti-tumor sFv in Escherichia coli. *Protein Eng.* **11**, 1401-1406.

Schatz, D.G., Oettinger, M.A. and Schlissel, M.S. (1992). V(D)J RECOMBINATION: Molecular biology and regulation. *Ann. Rev. immunol.* **10**, 359-383.

Schiffer, M., Girling, R.L., Ely, K.R. and Edmundson, A.B. (1973). Structure of a λ-type Bence-Jones protein at 3.5-Å resolution. *Biochemistry* **12**, 4620-4631.

Schiweck, W. and Skerra, A. (1997). The rational construction of an antibody against cystatin: Lessons from the crystal structure of an artificial Fab fragment. *J. Mol. Biol.* **268**, 934-951.

Schroeder, H.W. and Dighiero, G. (1994). The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire. *Immunol. Today* **15**, 288-294.

Schutte, M.E.M., Ebeling, S.B., Akkermans, K.E., Gmelig-Meyling, F.H.J. and Logtenberg, T. (1991). Antibody specificity and immunoglobulin VH gene utilization of human monoclonal $CD5^+$ B cell lines. *Eur. J. Immunol.* **21**, 1115-1121.

Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C. and Davies, D.R. (1987). Three-dimensional structure of an antibody-antigen complex. *PNAS* **84**, 8075-8079.

Shirai, H., Kidera, A. and Nakamura, H. (1996). Structural classification of CDR-H3 in antibodies. *FEBS Lett.* **399**, 1-8.

Singha, N.C., Surolia, N. and Surolia, A. (1996). On the relationship of thermodynamic parameters with the buried surface area in protein-ligand complex formation. *Biosci. Rep.* **16**, 1-10.

Skerra, A. (1993). Bacterial expression of immunoglobulin fragments. *Curr. Opin. Immunol.* **5**, 256-262.

Skerra, A., Pfitzinger, I. and Pluckthun, A. (1991). The functional expression of antibody Fv fragments in Escherichia coli: Improved vectors and a generally applicable purification technique. *Bio/Technology* **9**, 273-278.

Skerra, A. and Pluckthun, A. (1988). Assembly of functional immunoglobulin Fv fragment in Escherichia coli. *Science* **240**, 1038-1040.

Srinivasan, N. and Blundell, T.L. (1993). An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Eng.* **6**, 501-512.

Stanfield, R.L., Fieser, T.M., Lerner, R.A. and Wilson, I.A. (1990). Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 Å. *Science* **248**, 712-719.

Steele, E.J., Rothenfluh, H.S. and Blanden, R.V. (1997). Mechanism of antigen-driven somatic hypermutation of rearranged immunoglobulin V(D)J genes in the mouse. *Immunol. Cell Biol.* **75**, 82-95.

Tomlinson, I.M., Cox, J.P.L., Gherardi, E., Lesk, A.M. and Chothia, C. (1995). The structural repertoire of the human Vκ domain. *EMBO J.* **14**, 4628-4638.

Tomlinson, I.M., Walter, G., Jones, P.T., Dear, P.H., Sonnhammer, E.L.L. and Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. *J. Mol. Biol.* **256**, 813-817.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302**, 575-581.

Tramontano, A., Chothia, C. and Lesk, A.M. (1990). Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J. Mol. Biol.* **215**, 175-182.

Tsunenaga, M., Goto, Y., Kawata, Y. and Hamaguchi, K. (1987). Unfolding and refolding of a type κ immunoglobulin light chain and its variable and constant fragments. *Biochemistry* **26**, 6044-6051.

Ueki, Y., Goldfarb, I.S., Harindranath, N., Gore, M., Koprowski, H., Notkins, A.L. and Casali, P. (1990). Clonal analysis of a human antibody response. *J. Exp. Med.* **171**, 19-34.

van der Heijden, R.W.J., Bunschoten, H., Hoek, A., van Es, J., Punter, M., Osterhaus, A.D.M.E. and Uytdehaag, F.G.C.M. (1991). A human CD5+ B cell clone that secretes an idiotype-specific high affinity IgM monoclonal antibody. *J. Immunol.* **146**, 1503-1508.

Vargas-Mardrazo, E., Lara-Ochoa, F. and Almagro, J.C. (1995). Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition. *J. Mol. Biol.* **254**, 497-504.

Victor, K.D. and Capra, J.D. (1994). An apparently common mechanism of generating antibody diversity: Length variation of the VL-JL junction. *Mol. Immunol.* **31**, 39-46.

Walls, P.H. and Sternberg, J.E. (1992). New algorithm to model protein-protein recognition based on surface complementarity: Applications to antibody-antigen docking. *J. Mol. Biol.* **228**, 277-297.

Wedemayer, G.J., Patten, P.A., Wang, L.H., Schultz, P.G. and Stevens, R.C. (1997). Structural insights into the evolution of an antibody combining site. *Science* **276**, 1665-1669.

Weston, K.M. and Raison, R.L. (1991). Low affinity binding of mouse immunoglobulin to human $CD5^+$ B cells. *Immunol. Cell Biol.* **69**, 261-271.

Williams, D.C., Benjamin, D.C., Poljak, R.J. and Rule, G.S. (1996). Global changes in amide hydrogen exchange rates for a protein antigen in complex with three different antibodies. *J. Mol. Biol.* **257**, 866-876.

Williams, S.C., Frippiat, J.-P., Tomlinson, I.M., Ignatovich, O., Lefranc, M.-P. and Winter, G. (1996). Sequence and evolution of the human germline Vλ repertoire. *J. Mol. Biol.* **264**, 220-232.

Wilmot, C.M. and Thornton, J.M. (1988). Analysis and prediction of the different types of β-turn in proteins. *J. Mol. Biol.* **203**, 221-232.

Wu, S. and Cygler, M. (1993). Conformation of complementarity determining region L1 loop in murine IgG λ light chain extends the repertoire of canonical forms. *J. Mol. Biol.* **229**, 597-601.

Wu, T.T., Johnson, G. and Kabat, E.A. (1993). Length distribution of CDRH3 in antibodies. *Proteins: Stuct. Func. Genet.* **16**, 1-7.

Wu, T.T. and Kabat, E.A. (1970). An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma light chains and their implication for antibody complementarity. *J. Exp. Med.* **132**, 211-249.

Wulfing, C. and Pluckthun, A. (1994). Protein folding in the periplasm of Escherichia coli. *Mol. Micro.* **12**, 685-692.

Ysern, X., Fields, B.A., Bhat, T.N., Goldbaum, F.A., Dall'Acqua, W., Schwarz, F.P., Poljak, R.J. and Mariuzza, R.A. (1994). Solvent rearrangement in an antigen-antibody interface introduced by site-directed mutagenesis of the antibody combining site. *J. Mol. Biol.* **238**, 496-500.

Zachau, H.G. (1993). The immunoglobulin κ locus - or - what has been learned from looking closely at one-tenth of a percent of the human genome. *Gene* **135**, 167-173.

Zdanov, A., Li, Y., Bundle, D.R., Deng, S.-J., MacKenzie, C.R., Narang, S.A., Young, N.M. and Cygler, M. (1994). Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-Å resolution. *PNAS* **91**, 6423-6427.

Journal of Computer-Aided Molecular Design, 11 (1997) 453–461.
KLUWER/ESCOM

J-CAMD 407

Diverse binding site structures revealed in homology models of polyreactive immunoglobulins

Paul A. Ramsland[a], Luke W. Guddat[b], Allen B. Edmundson[c] and Robert L. Raison[a,*]

[a]*Immunobiology Unit, University of Technology, Sydney, Westbourne Street, Gore Hill, NSW 2065, Australia*
[b]*Centre for Drug Design and Development, The University of Queensland, St. Lucia, QLD 4072, Australia*
[c]*Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, OK 73104, U.S.A.*

Received 15 April 1996
Accepted 22 April 1997

Keywords: Polyreactive IgM; Natural (auto)antibodies; Template-based models; Chronic B lymphocytic leukaemia

Summary

We describe here computer-assisted homology models of the combining site structure of three polyreactive immunoglobulins. Template-based models of Fv (V_L–V_H) fragments were derived for the surface IgM expressed by the malignant CD5 positive B cells from three patients with chronic lymphocytic leukaemia (CLL). The conserved framework regions were constructed using crystal coordinates taken from highly homologous human variable domain structures (Pot and Hil). Complementarity determining regions (CDRs) were predicted by grafting loops, taken from known immunoglobulin structures, onto the Fv framework models. The CDR templates were chosen, where possible, to be of the same length and of high residue identity or similarity. LCDR1, 2 and 3 as well as HCDR1 and 2 for the Fv were constructed using this strategy. For HCDR3 prediction, a database containing the Cartesian coordinates of 30 of these loops was compiled from unliganded antibody X-ray crystallographic structures and an HCDR3 of the same length as that of the B CLL Fv was selected as a template. In one case (Yar), the resulting HCDR3 model gave unfavourable interactions when incorporated into the Fv model. This HCDR3 was therefore modelled using an alternative strategy of construction of the loop stems, using a previously described HCDR3 conformation (Pot), followed by chain closure with a β-turn. The template models were subjected to positional refinement using energy minimisation and molecular dynamics simulations (X-PLOR). An electrostatic surface description (GRASP) did not reveal a common structural feature within the binding sites of the three polyreactive Fv. Thus, polyreactive immunoglobulins may recognise similar and multiple antigens through a diverse array of binding site structures.

Introduction

While the majority of immunoglobulins (Igs) exhibit restricted antigen binding specificity, polyreactive immunoglobulins can recognise multiple structurally dissimilar antigens. Malignant cells from patients with chronic B lymphocytic leukaemia (B CLL) frequently express polyreactive antibodies, thus providing a system for investigations into the structural features important for polyreactive antigen binding. Unlike the situation with monospecific antibodies where structural correlation of the binding site with antigen specificity has been elucidated for a number of antigen/antibody systems [1–5], the structural basis of the polyreactivity of certain antibody combining sites remains unclear. Clarification of this problem requires detailed structural analysis of a large number of polyreactive binding sites, a process that will be greatly assisted by the use of predicted structures in addition to the essential, but limited, number of experimentally determined structures for binding sites of this type.

The antigen binding properties of Ig are largely dictated by six (three on light chains and three on heavy chains) complementarity determining regions (CDRs) in the variable (V) domains which provide a unique molecular surface for specific antigen recognition. The CDRs are supported by a highly conserved molecular scaffolding provided by the framework regions. Various methods have been described to predict the three-dimensional (3D)

*To whom correspondence should be addressed.

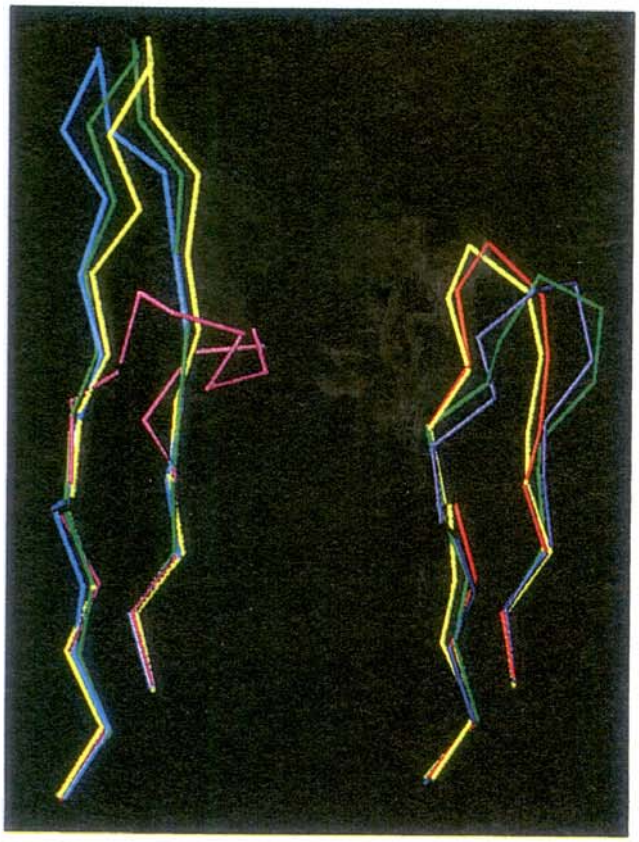


Fig. 1. C^{α} representations of the HCDR3 conformations. Left: the 16-residue HCDR3 loops. Bel is coloured blue, Yar 1 is in yellow and Yar 2 in magenta. The crystal structure of the HCDR3 of NC10.14 is in green. Right: the 11-residue HCDR3 loops. Tre is coloured yellow, and the crystal structures of 1FVC, 1HIL and 1MCP are in green, purple and red, respectively.

structure of antibody Fv (V_L–V_H) molecules [6–11]. Usually the structurally conserved framework regions of the V domains are constructed from atomic coordinates taken directly from an Ig crystal structure. The conformations of the six CDR loops are predicted by scanning the sequence for key 'canonical' structure determining residues to select a template loop [12,13] or by randomly sampling conformational space for loops with low potential energy values [14,15]. This approach is usually adequate for five of the six loops, but the third loop on the heavy chain (HCDR3) shows the greatest variability in length and sequence, resulting in an inability to accurately predict its structure. In view of this, there is a need for a combined modelling and crystallographic approach to improve the quality of HCDR3 predictions. In comparison, framework regions and CDR loops other than HCDR3 are well predicted by homology modelling. Homology-based models of variable domain structures, when compared with experimentally determined structures, show root

TABLE 1
COMPARISON OF THE 3D6 CRYSTAL STRUCTURE WITH THE HOMOLOGY Fv MODEL

Region of the variable domain	Rmsd for all C^{α} atoms (Å)[a]
LCDR1	0.67
LCDR2	0.35
LCDR3	0.84
HCDR1	0.23
HCDR2	0.70
HCDR3	1.30
V_L[b]	0.74
V_H[b]	0.49
V_L–V_H[b]	1.0

[a] Rms values shown are for the α carbon atoms used for rigid-body alignments.

[b] V domains and Fv (V_L–V_H) were aligned using the conserved β-sheet framework residues (i.e. framework residues are 1–23, 35–49, 57–88, 98–107 for light chains and 1–30, 36–49, 66–94, 103–113 for heavy chains) [20].

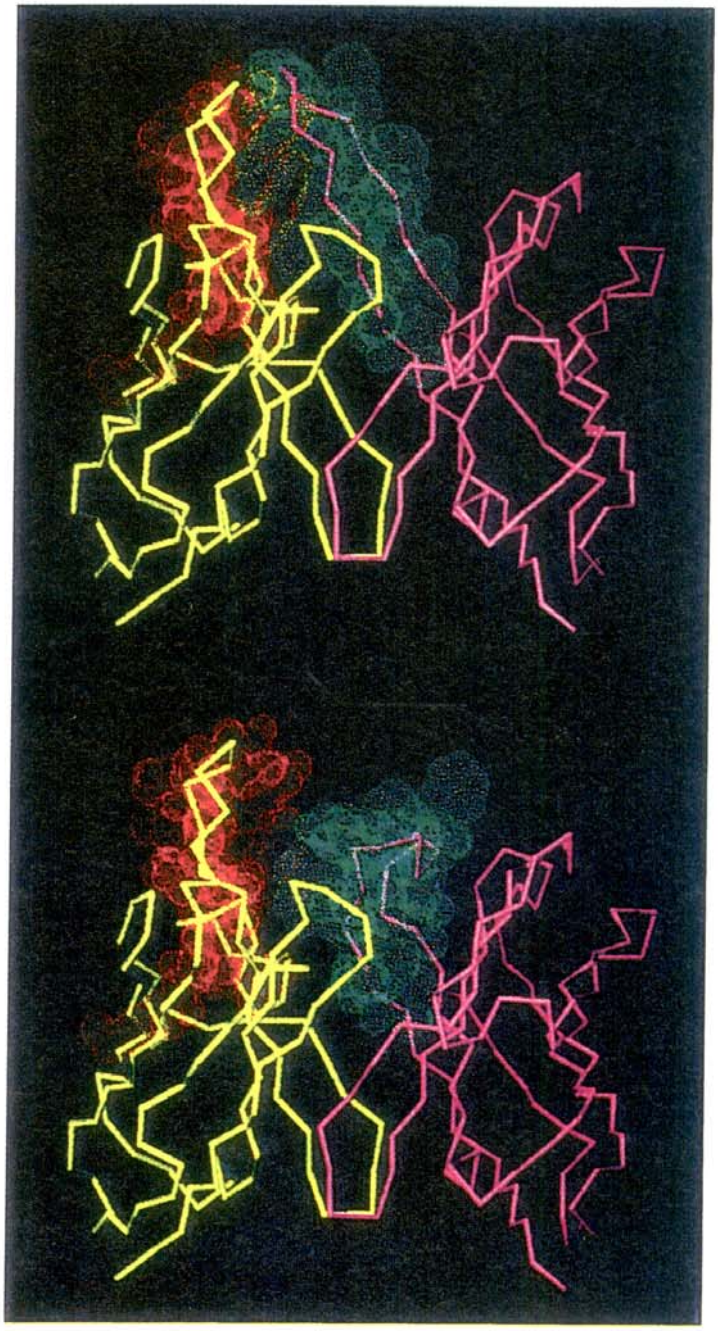


Fig. 2. Side views of the two Yar Fv models. The heavy-chain V domains are represented in magenta and the light chains in yellow. The van der Waals radii are in red for LCDR1 and in green for HCDR3. Top: Yar 1 based on the HCDR3 conformation of NC10.14; bottom: Yar 2 based on the HCDR3 conformation of 1IGM.

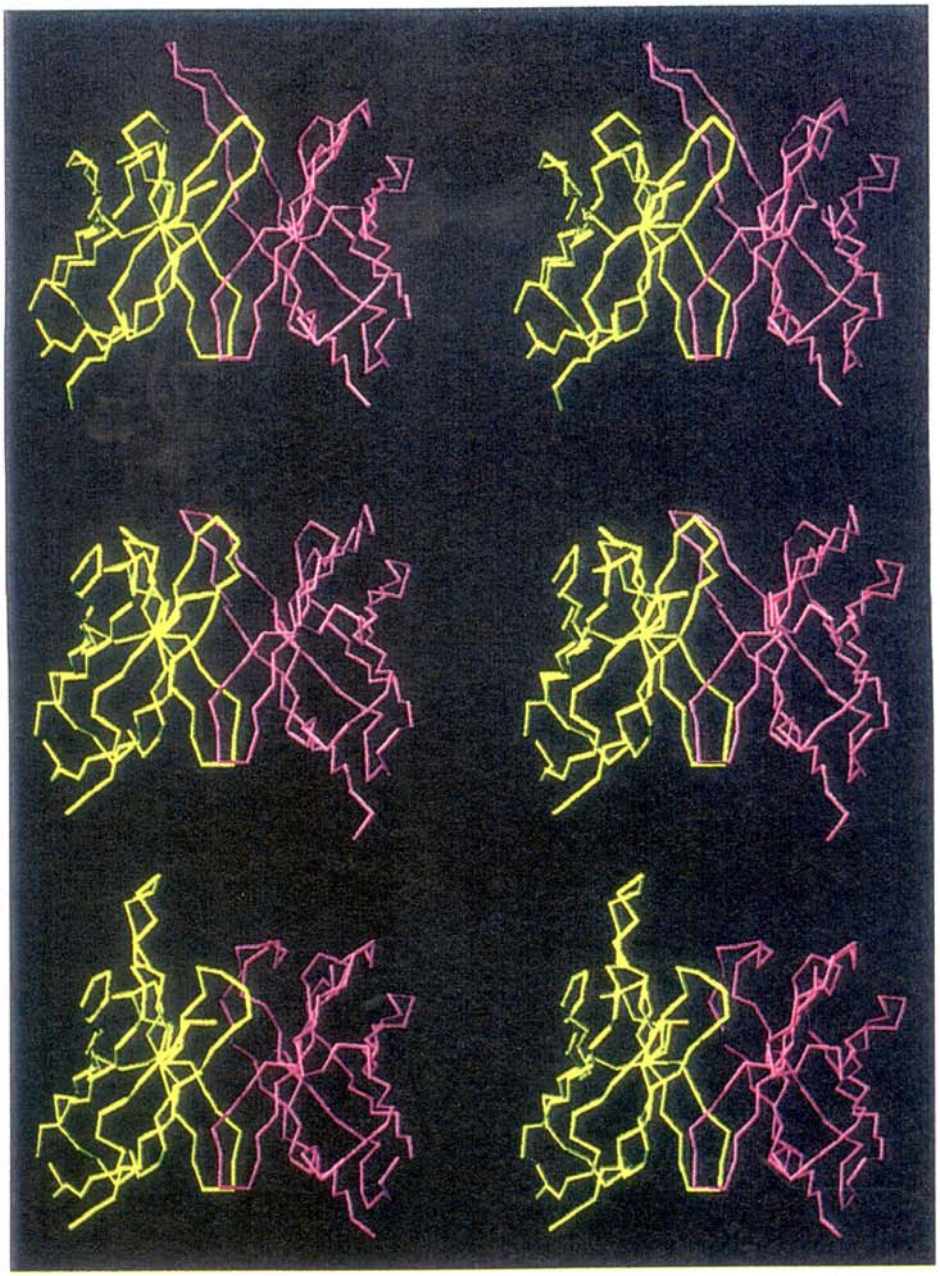


Fig. 3. Stereo C^{α} diagrams of the side views of the polyreactive Fv homology models. The heavy chains are in magenta and the light chains are in yellow. Top: Bel; middle: Tre; bottom: Yar.

mean square (rms) deviations of less than 1.0 Å for framework and canonically predicted CDRs other than HCDR3 [11–13]. Thus, homology modelling can provide a useful structural description of the binding site, comparable to the quality of a low-resolution X-ray structure [16].

In this study, we have applied homology modelling to determine the structures of the variable domains of three polyreactive antibodies. Electrostatic molecular surface models [17] reveal the level of binding site diversity exhibited by this group of immunoglobulins.

TABLE 2
VARIABLE DOMAINS USED TO CONSTRUCT Fv MODEL FRAMEWORK REGIONS

B-CLL Fv model (isotype)	V_L framework residue identity (%) (PDB code)	V_H framework residue identity (%) (PDB code)
Bel (IgM λ)	75 (8FAB)	60 (1IGM)
Tre (IgM κ)	82 (1IGM)	62 (1IGM)
Yar (IgM κ)	69 (1IGM)	87 (1IGM)

Materials and Methods

Immunoglobulins

The complete nucleotide sequences of the variable domains of the Ig described here will be published elsewhere (manuscript in preparation). Here we only show the translated amino acid sequences of the light- and heavy-chain CDR loops (Table 3). The IgM molecules are monoclonal, exhibit polyreactive binding properties and are expressed on the surface of malignant $CD5^+$ B cells from patients with B CLL. The IgM were designated as Bel (IgM λ), Tre (IgM κ) and Yar (IgM κ).

Computer-aided modelling

All computational work was performed on an INDY workstation running IRIX, v. 5.2 (Silicon Graphics Inc.). Template models were constructed using the software TURBO-FRODO, v. 5 (Biographics, France). Minimisation and molecular dynamics were performed using X-PLOR, v. 3.1 [18]. Molecular surfaces and electrostatic potentials were calculated and visualised with GRASP, v. 1.2.5 [17].

Modelling strategy

The B CLL Fv molecules were constructed by starting with known high-resolution crystal structures (1BAF, 1BBD, 1IGM, 2FB4, 8FAB, 1DFB, 1REI, 1HIL, 1MCP) taken from the Brookhaven Protein Databank [19]. Variable domain templates were selected from human Ig fragments on the basis of highest amino acid identity within framework regions. Where necessary, V domain templates were aligned, using rigid-body alignment of conserved Ig β-sheets [20] over the corresponding domain of the Pot (1IGM) structure. Thus, all Fv models have the V_L–V_H domain interface of Pot IgM Fv [21]. CDR templates were selected initially on the basis of an identical number of residues comprising the loop. Sequence homology was used as a secondary criterion. Furthermore, only residues within the CDRs as defined by Kabat et al. [22] were considered. The CDRs were classified, where possible, into canonical classes [12,13]. Loops which were found to be of a different length to the available loops of known conformation were constructed using a template-based approach, where the loop stems were built using CDRs of known conformation. Additional residues were placed onto these stems and chain closure was accomplished by grafting a 'classical' β-turn onto the CDR loop stems. The type of β-turn assigned was based on the criteria detailed by Wilmot and Thornton [23].

Structures were optimised using the CHARMM22 topology and parameter set implemented within X-PLOR, v. 3.1 [18]. The template-based models were subjected to an initial 100 cycles of conjugate gradient energy minimisation [24]. After this initial minimisation step, harmonic coordinate restraints of 20.0 kcal/(mole $Å^2$) were set for all main-chain α carbon atoms. A further 200 cycles of energy minimisation were then carried out. The Fv models were subjected to molecular dynamics and simulated annealing where the structures were initially heated to 1000 K followed by slow cooling to a bath temperature of 300 K. The time step for the simulation was set to 0.5 fs. A final step of 300 cycles of energy minimisation was carried out.

Comparison of structures

The stereochemistry and geometry of Fv models were assessed by calculating the deviations from ideality of bond lengths, bond angles, dihedrals and improper angles using the parameter set of Engh and Huber [25]. All φ, ψ torsional angles were analysed by the method of Ramachandran and Sasisekharan [26]. Optimal rigid-body structural alignments and rms deviations between structures were calculated within TURBO-FRODO. Interchain energies were calculated from the van der Waals (vdW) and electrostatic energy terms in X-PLOR [18]. For these calculations, the protein dielectric was set to 2.0 without a nonbonded cutoff using the CHARMM22 force field.

Results

Validation of modelling strategy

To evaluate the template-based modelling strategy, a homology model of the variable domains of the human antibody 3D6 was constructed. The 3D6 structure [27]

TABLE 3
CDR SEQUENCES OF THE POLYREACTIVE Fv MOLECULES AND THE CDR TEMPLATES SELECTED FOR MODELLING

CDR[a]	Bel Fv model (PDB codes)[b,c]	Tre Fv model (PDB codes)[b,c]	Yar 2 Fv model (PDB codes)[b,c]
L1	GGNNIGSDSVH (8FAB)	RASQSISSYLN (1DFB), **2**	KSSQSVLYSSNNKNYLA (1BBD), **3**
L2	YDSDRPS (2FB4)	AASSLQS (1DFB), **1**	WASTRES (1DFB), **1**
L3	QVWDSSSVV (8FAB)	QQSYSTPPYT (1BAF), **1**	QQYYSTPYT (1REI), **1**
H1	NYWIG (8FAB), **1**	SYWIS (1HIL), **1**	SYAMH (1DFB), **1**
H2	IIYPGDSDTRYSPSFQG (2FB4), **2**	RIDPSDSYTNYSPSFQG (1BBD), **2**	AISSNGGTTYYADSVKG (1HIL), **3**
H3	RTGTGDPYYYYYYMDV (NC10.14)	RQWLALGHFDY (1MCP)	TYYDFWSGYSPNWFDP (1IGM, 4PTP)

[a] CDRs were assigned according to Kabat et al. [22].
[b] Residue identities between the template loop (PDB code indicated) and the polyreactive Fv sequences have been underlined.
[c] Where a 'canonical' class could be assigned to a CDR, the class is indicated in bold [12,13].

TABLE 4
INTERCHAIN INTERACTION ENERGIES OF TWO ALTERNATIVE YAR Fv MODELS

Interaction	E-total[a] (kcal/mol)	E-vdW[b] (kcal/mol)	E-ELEC[c] (kcal/mol)
Yar 1			
V_L–V_H	−58.3	−84.1	25.7
V_L–HCDR3	−7.5	−45.9	38.4
Yar 2			
V_L–V_H	−103.4	−86.5	−16.9
V_L–HCDR3	−50.5	−45.8	−4.7

[a] All energy values are calculated for interchain interactions only using X-PLOR [18].
[b] van der Waals energy term.
[c] Electrostatic energy contribution term.

was not in our template database. The coordinates from 1IGM were used to determine the V domain framework regions. Residue identities in these regions of 3D6 and 1IGM were 86% and 90% for V_H and V_L, respectively. Template CDR structures of the same residue length as the 3D6 CDR loops were selected on the basis of maximum sequence homology with the target CDR. LCDR1 and LCDR2 were taken from 1IGM, and HCDR1 and HCDR2 from 2FB4. Template loops of the same length were not available for LCDR3 (7 residues) and HCDR3 (17 residues). CDR3 loops were built using loop stems of known conformation, taken from corresponding loops of 1REI for LCDR3 and NC10.14 (Guddat and Edmundson, unpublished structure) for HCDR3, followed by chain closure with a classical β-turn [23]. The rms deviations (α carbon atoms) between the model and the crystal structure were determined after optimal rigid-body alignments of various regions of the two structures (Table 1). The modelled CDR loops, the individual V domains and the modelled Fv fragment were compared to the published 3D6 structure, with rms deviations between the crystal and model structures of <1.0 Å for all regions except for HCDR3, which has an rms deviation of 1.30 Å.

Template models of polyreactive immunoglobulins

The template V domains used to model the framework regions of the polyreactive Fv are shown in Table 2. In each case, they represent the best identity match from the database with the respective CLL IgM V domains. Residue identities ranged from 87% for the Yar/1IGM V_H pairing to 60% for the V_H of Bel and 1IGM. Residues were replaced within the program TURBO-FRODO (Biographics). The side chains of the mutated residues were manually positioned into a conformation which minimised any nonbonded contacts and the bond geometries were refined against a rotamer library (TURBO-FRODO). The templates used for constructing the CDRs of the three Fv models are summarised in Table 3. Where a canonical class could be assigned to a CDR loop, it is indicated. However, loops were initially chosen as templates on the basis of residue identities and similarities within the CDRs as defined by Kabat et al. [22].

In an attempt to develop a template-based strategy for predicting the structure of HCDR3, we extracted these loop coordinates from the 30 available unliganded Ig crystal structures (29 from the Protein Databank [19] and one from the NC10.14 structure (Guddat and Edmundson, unpublished results)). The length of the HCDR3 loops ranged from 5 to 17 residues but there were no loops of 6 or 13 residues long. Bel and Yar antibodies both have 16-residue HCDR3 loops, the same length as a murine Ig, NC10.14. For the 11-residue HCDR3 of Tre, there were three structures in the database.

The predicted HCDR3 conformations and a comparison with their starting models are presented in Fig. 1. The Bel HCDR3 was constructed from the HCDR3 of the NC10.14 Fab structure. This loop conformation fitted the Bel template model and did not result in unfavourable interactions with the other five CDRs in the Fv model. The Yar HCDR3 was initially constructed using the NC10.14 conformation (Yar 1). However, when this structure was used in the Yar Fv model, the backbone of the LCDR1 loop came into close contact with the HCDR3. An alternative model for the Yar HCDR3 (Yar 2) was then constructed to incorporate the fold seen in the Pot (1IGM) HCDR3 [21]. Additional residues were added and the chain closed using a β-turn taken from the trypsin (4PTP) structure. The two models of Yar Fv are presented (Fig. 2) showing the vdW surfaces of the LCDR1 and HCDR3 loops. It was noted that the vdW radii of the HCDR3 and LCDR1 of Yar 1 exhibited many close contacts, whereas the HCDR3 vdW radii in Yar 2 made fewer close contacts with LCDR1. The interchain energies between the two V domains and between HCDR3 and V_L were calculated for the two proposed structures of Yar Fv using X-PLOR (Table 4). Although the vdW contacts were similar between the two loop structures, unfavourable electrostatic energy was calculated for the HCDR3 structure based on an NC10.14 template (Yar 1). In comparison, modelling based on Pot HCDR3 (Yar 2) resulted in a significantly improved electrostatic energy term. The template for the Tre

TABLE 5
Fv MODEL STEREOCHEMISTRY

Rms deviations[a]	Bel Fv	Tre Fv	Yar 2 Fv
Bond distances (Å)	0.004	0.004	0.004
Bond angles (°)	0.9	1.0	1.0
Dihedral angles (°)	26.0	26.1	26.0
Improper angles (°)	0.8	0.9	0.9

[a] Rms deviations from ideality were calculated for all residues in the given Fv model structure, compared against Engh and Huber parameters [25].

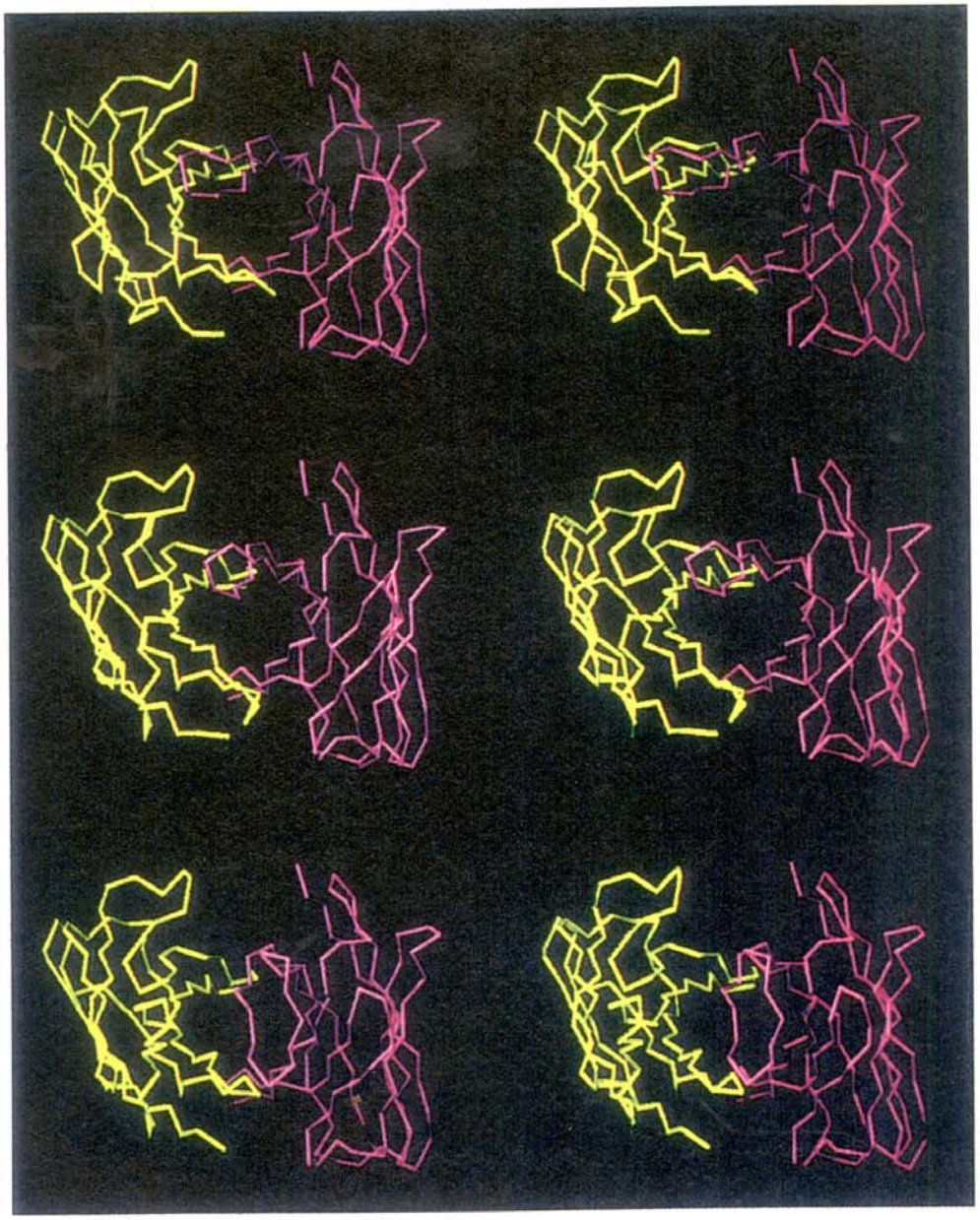


Fig. 4. Stereo C^{α} diagrams of the end-on views of the polyreactive Fv homology models. The heavy chains are in magenta and the light chains are in yellow. Top: Bel; middle: Tre; bottom: Yar.

HCDR3 was the 11-residue loop of 1MCP. This loop showed the greatest similarity in sequence to the Tre loop compared with the other known loops of identical length.

Refined Fv models

The Fv models after energy minimisation and molecular dynamics refinement using the program X-PLOR [18] are presented in Figs. 3 and 4. The rms deviations in bond lengths, bond angles and dihedral angles are presented in Table 5. A Ramachandran plot [26] was calculated and the majority of ϕ,ψ torsion angles around C^{α} atoms are within the allowed regions. These plots are shown for the Fv models after the full positional refinement protocol (Fig. 5) and are a further indication that appropriate stereochemistry and geometry have been achieved in the models.

Electrostatic surface models of polyreactive Fv

Binding site topology is shown as a surface representation [17] in Fig. 6. The solvent-accessible surfaces and the electrostatic potentials are depicted for the polyreactive Fv structures. A deep binding pocket is seen at the V_L–V_H domain interface of the Bel and Tre Fv models. The two alternative Yar Fv models (Yar 1 and Yar 2) are also presented in Fig. 6. The electrostatic surface potential of the Bel binding site is predominantly hydrophobic in nature, with regions of negative potential restricted to the edges of the binding cavity. The Tre binding site has a strong positive electrostatic potential lining the floor of the binding pocket. The binding site of Yar Fv exhibits an essentially hydrophobic electrostatic surface potential which is not significantly altered between the two models presented for this Fv.

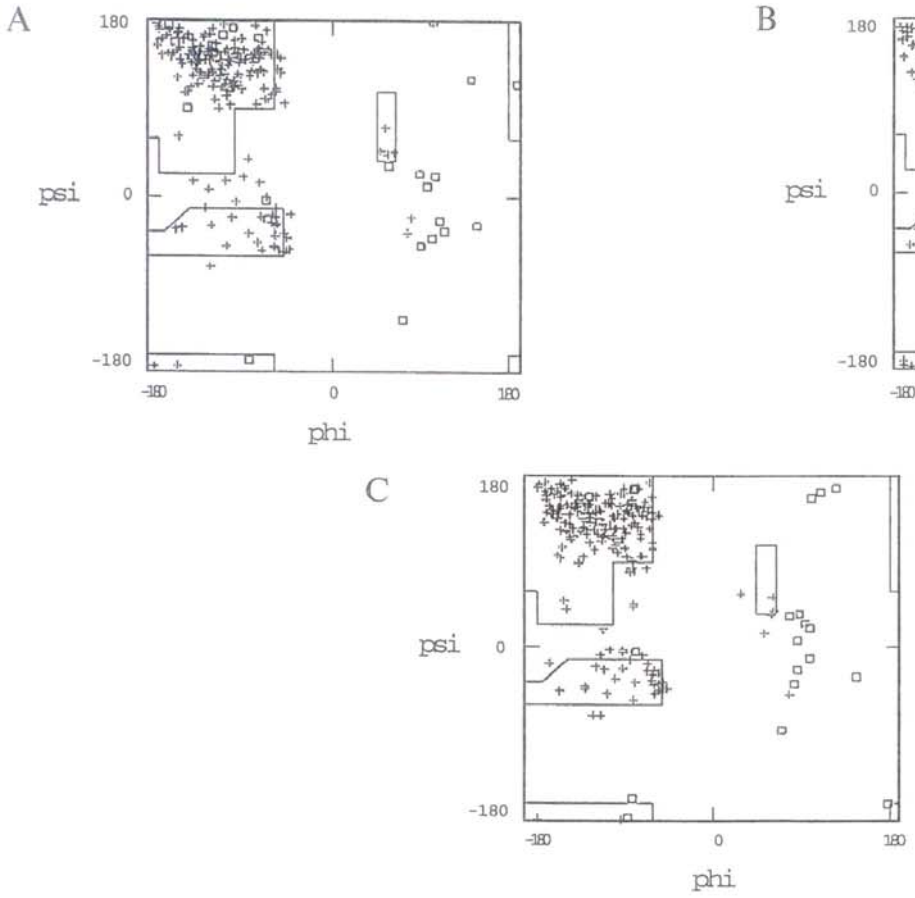


Fig. 5. Ramachandran plots [26] of the ϕ,ψ angles for all α carbon dihedral angles in the variable region fragment homology models. Crosses represent all the residue types excluding glycines which are described by the boxes. Lines delimit the favoured dihedral conformations. Ramachandran plots are labelled: (A) Bel; (B) Tre; (C) Yar 2.

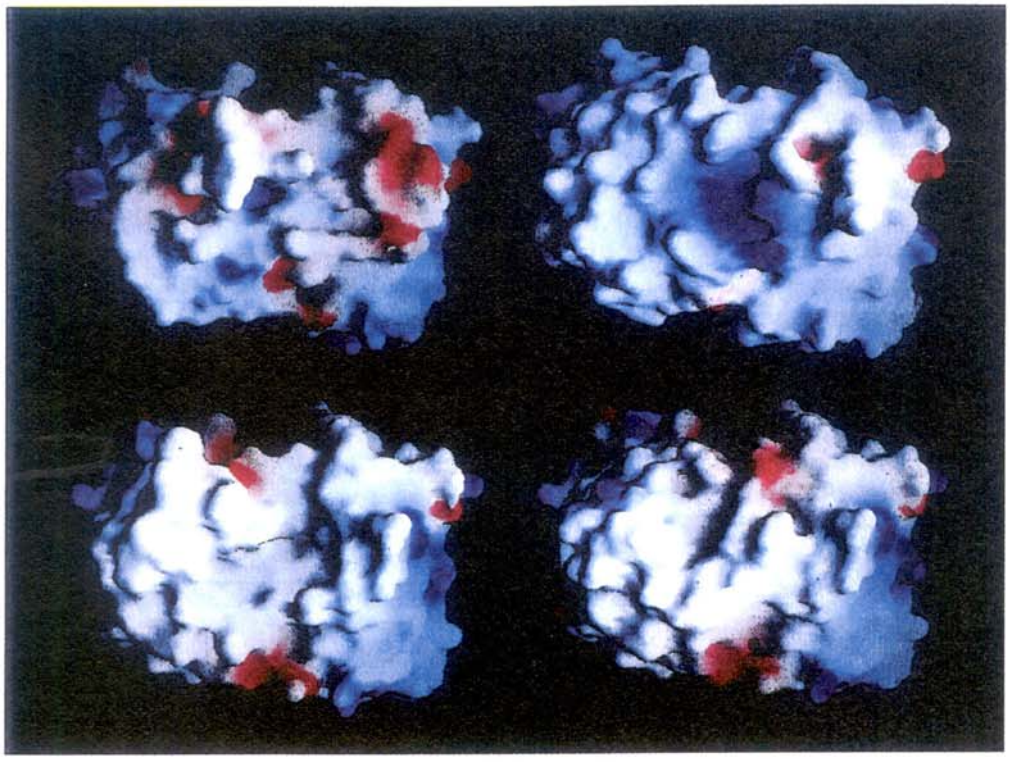


Fig. 6. Electrostatic surface representations of the Fv homology models (end-on views). Top left: Bel; top right: Tre; bottom left: Yar 1; bottom right: Yar 2. The solvent-accessible surface and the electrostatic potential across this surface have been represented using the program GRASP [17]. White represents areas of zero surface potential, blue represents positive surface potential and red represents negative surface potential.

Discussion and Conclusions

Computational techniques for predicting antibody combining site structures have generally approached the problem by utilising the conserved Ig variable domain scaffolding to construct good approximations of antibody variable region fragments [11]. These models provide important information regarding molecular recognition. The homology models of polyreactive Fv molecules presented here were constructed in an attempt to identify structural features of polyreactive Ig antigen binding sites.

The topology of immunoglobulin combining sites is predominantly determined by the primary sequence and folded conformation of the six CDRs. The conformation adopted by CDR loops is largely dependent on the number of residues therein. Additionally, CDR conformation is affected by 'key' structure determining residues both within the loop or in neighbouring framework regions. Chothia et al. [12,13] identified this dependence of loop conformation on length and sequence and produced an elegant theory of canonical CDR loop structures. In our homology models, we did not initially analyse the templates for the presence of 'key' framework residues. Only the residues assigned to CDR loops by Kabat et al. [22] were considered in the choice of the templates. Subsequent analysis has revealed that the CDR loops of the two κ light-chain V domains (Tre and Yar) and all three heavy-chain CDR1 and CDR2 structures fall into previously described canonical classes (Table 3). The antibody Bel expresses a λ light chain. In this case, none of the light-chain CDR loops could be assigned to a canonical class. This is mainly due to a disproportionate number of κ compared to λ light-chain structures available at high atomic resolution. At least five out of the six CDR loops within each of our polyreactive Fv homology models were predicted using template loops of the same length. Previous studies have demonstrated [6,12,13] that these loops are predicted accurately by homology modelling and so provide a good approximation of the basic binding site topology.

No current modelling procedure has consistently predicted the conformation of heavy-chain CDR3 loops with a high level of accuracy. This is mainly due to the heavy-chain CDR3 loop being the most variable with respect to length and sequence. Human HCDR3 loops have lengths varying from 2 to 26 residues [28]. In order to address this difficulty, our approach to the prediction of HCDR3 loops has been to construct these loops using a template-based loop building strategy. We constructed a 3D database of HCDR3 from native X-ray structures. Although the loops range from 5 to 17 residues, not all HCDR3 lengths are represented in the database. In some cases, only one structure is known for a given HCDR3 length. As the number of Ig crystal structures increases, the ability of an empirical approach to predict HCDR3 conformations will become increasingly more accurate. In the present study, Bel and Yar contain 16-residue HCDR3 while Tre Fv contains an 11-residue HCDR3 (Fig. 1). The Bel Fv HCDR3 was constructed from the coordinates of a murine monoclonal antibody NC10.14 which has a well-ordered 16-residue β-loop that protrudes out from the binding pocket into the solvent. The HCDR3 of Yar was constructed using either the conformation of this loop found in the NC10.14 Fab (Yar 1) or that of the 1IGM Fv (Yar 2). A comparison of two alternative conformations for the Yar HCDR3 suggested that the more energetically favourable conformation was for the loop to fold across the binding site, resulting in a relatively flat binding surface (Yar 2). This fold was based on the crystallographic structure for the HCDR3 of Pot Fv described previously by Edmundson's group [21]. While it is a distinct possibility that the HCDR3 of Yar could adopt a conformation somewhere intermediate between that of NC10.14 and Pot HCDR3, the current database of HCDR3 loop folds does not provide a more energetically favourable alternative. It will be of great interest to determine the experimental structure of Yar Fv to examine how influential the relatively long LCDR1 is on the conformation of the HCDR3.

The template for the 11-residue HCDR3 of Tre was 1MCP. The 1MCP loop was selected from the three unique HCDR3 sequences of known conformation (Fig. 1). Perhaps shorter HCDR3 loops are more similar in structure to each other than longer loops, a feature which is at least reflected by the 11-residue HCDR3 loops examined. Although there are differences between the three 11-residue HCDR3 loops of known conformation, these are restricted to small movements towards the top of the loops. It appears that large differences in loop conformation are not common for these shorter loops. In view of the apparent conformational restriction of 11-residue HCDR3, we were encouraged to conclude that the fold of the Tre HCDR3 loop is accurately predicted using 1MCP HCDR3 as a template. Another indication that this loop was an appropriate template for Tre HCDR3 is that the conformation of 1MCP HCDR3 is closely maintained after refinement of the Tre Fv model and so this loop fitted well with the other five CDR loops. Especially in cases where several experimental structures of HCDR3 of the same length are available, a template-based approach to modelling HCDR3 should equal or surpass the predictive capabilities of methods based on systematic conformational searching or the random generation of loop structures [14,15].

One of the strongly held precepts regarding the molecular mechanisms of polyreactive antibody binding is that these immunoglobulins would have a conserved combining site structure [29,30]. Padlan [31] argues in a comprehensive review of antibody structure that the poly-

reactive antibody combining site is relatively large, plastic or 'sticky' (high number of aromatic side chains within CDR) in character. Our current data would appear to contradict this concept and point towards a more diverse pattern of binding site structure for polyreactive immunoglobulins. The polyreactive Fv homology models (Figs. 3 and 4) suggest that at least two basic binding site topologies are involved in polyreactive antigen recognition: (i) a pocket-type binding site is predicted for Bel and Tre antibodies; and (ii) a relatively flat binding surface is described for Yar Ig. Furthermore, the HCDR3 of Bel and the LCDR1 of Yar project out from the compact variable domains into the solvent. These solvent-exposed loops are potentially buried during antigen binding. In particular, the highly aromatic nature of the Bel HCDR3, resulting mainly from a stretch of six tyrosine residues, suggests that this loop is buried at the interface between antigen and antibody. The electrostatic surface representations of the Fv models (Fig. 6) indicate that the binding site of Bel is predominantly hydrophobic in character with one region of negative potential. Yar IgM also has a hydrophobic binding surface. However, the binding 18pocket of Tre immunoglobulin has a strongly positive electrostatic surface potential. Finally, with respect to the electrostatics, an examination of the two Yar models (Fig. 6) shows that the electrostatic nature of the binding site is relatively similar even though the HCDR3 is shifted dramatically between a loop which projects upwards from the V domains (Yar 1) to one which folds across the V_L–V_H domain interface (Yar 2). This indicates that an electrostatic surface representation is a robust technique for investigating the binding sites of antibodies even in cases where the confidence of prediction of the HCDR3 may be low.

Elucidation of the structural basis of polyreactive antigen binding will require analysis and comparison of the binding sites of numerous polyreactive Igs. While X-ray crystallography is arguably the definitive method for achieving this goal, it is time consuming, costly and dependent on the ability to obtain suitable protein crystals. Thus, the development of improved antibody modelling techniques will make a significant contribution towards the solution of this problem. We will attempt to establish X-ray structures of Bel, Tre and Yar Fv from protein obtained through the bacterial expression of the V_L and V_H domains. The model structures (published here in advance) will then be compared with the X-ray structures. Such comparisons will result in refinement of the modelling process and will increase the size of the database of relevant antibody structures.

Acknowledgements

We are grateful to Dr. Leif Hanson for pertinent comments on the manuscript and to Kim Andersen for her expert computing skills. This work was supported by a grant from the Australian National Health and Medical Research Council (NHMRC). P.A.R. is a recipient of an Australian Postgraduate Award (APA).

References

1 Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J., Science, 233 (1986) 747.
2 Colman, P.M., Laver, W.G., Varghese, J.N., Baker, A.T., Tulloch, P.A., Air, G.M. and Webster, R.G., Nature, 326 (1987) 358.
3 Herron, J.N., He, X.-M., Mason, M.L., Voss Jr., E.W. and Edmundson, A.B., Proteins Struct. Funct. Genet., 5 (1989) 271.
4 Herron, J.N., He, X.-M., Ballard, D.W., Blier, P.R., Pace, P.E., Bothwell, A.L.M., Voss Jr., E.W. and Edmundson, A.B., Proteins Struct. Funct. Genet., 11 (1991) 159.
5 Arevalo, J.H., Hassig, C.A., Stura, E.A., Sims, M.J., Taussig, M.J. and Wilson, I.A., J. Mol. Biol., 241 (1994) 663.
6 Barry, M.M., Mol, C.D., Anderson, W.F. and Lee, J.S., J. Biol. Chem., 269 (1994) 3623.
7 Bassolino-Klimas, D., Bruccoleri, R.E. and Subramaniam, S., Protein Sci., 1 (1992) 1465.
8 Bajorath, J., Bioconj. Chem., 5 (1994) 213.
9 Chothia, C., Lesk, A.M., Levitt, M., Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J., Science, 233 (1986) 755.
10 De la Paz, P., Sutton, B.J., Darsley, M.J. and Rees, A.R., EMBO J., 5 (1986) 415.
11 Martin, A.C.R., Cheetham, J.C. and Rees, A.R., Methods Enzymol., 203 (1991) 121.
12 Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M. and Poljak, R.J., Nature, 342 (1989) 877.
13 Chothia, C., Lesk, A.M., Gherardi, E., Tomlinson, I.M., Walter, G., Marks, J.D., Meirion, B.L. and Winter, G., J. Mol. Biol., 227 (1992) 799.
14 Bruccoleri, R.E. and Karplus, M., Biopolymers, 26 (1987) 137.
15 Bajorath, J. and Fine, R.M., Immunomethods, 1 (1992) 137.
16 Sali, A., Curr. Opin. Biotechnol., 6 (1995) 437.
17 Nicholls, A. and Honig, B., J. Comput. Chem., 12 (1991) 435.
18 Brunger, A.T., X-PLOR v. 3.1. A System for X-ray Crystallography and NMR, Yale University Press, New Haven, CT, U.S.A., 1992.
19 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
20 Guddat, L.W., Shan, L., Anchin, J.M., Linthicum, D.S. and Edmundson, A.B., J. Mol. Biol., 236 (1994) 247.
21 Fan, Z.C., Shan, L., Guddat, L.W., He, X.-M., Gray, W.R., Raison, R.L. and Edmundson, A.B., J. Mol. Biol., 228 (1992) 188.
22 Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. and Foeller, C., Sequences of Proteins of Immunological Interest, 5th ed., National Institutes of Health, Bethesda, MD, U.S.A., 1991.
23 Wilmot, C.M. and Thornton, J.M., J. Mol. Biol., 203 (1988) 221.
24 Powell, M.J.D., Math. Programming, 12 (1977) 241.
25 Engh, R.A. and Huber, R., Acta Crystallogr., A47 (1991) 392.
26 Ramachandran, G.N. and Sasisekharan, V., Adv. Protein Chem., 23 (1968) 283.
27 He, X.-M., Ruker, F., Casale, E. and Carter, D.C., Proc. Natl. Acad. Sci. USA, 89 (1992) 7154.
28 Wu, T.T., Johnson, G. and Kabat, E.A., Proteins Struct. Funct. Genet., 16 (1993) 1.
29 Avrameas, S. and Ternynck, T., Mol. Immunol., 30 (1993) 1133.
30 Cheung, S.C., Takeda, S. and Notkins, A.L., Clin. Exp. Immunol., 101 (1995) 383.
31 Padlan, E.A., Mol. Immunol., 31 (1994) 231.