

**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

Title	i
Certificate	ii
Acknowledgments	iii
Publications arising from this thesis	iv
Table of contents	v
List of figures	x
List of tables	xii
Abbreviations	xiii
Abstract	xiv

Section	Title	Page
CHAPTER ONE	Introduction	1
1.1	Basic structure of immunoglobulins	1
1.1.1	Genes encoding variable domains of immunoglobulins	2
1.1.2	Characteristics of an intact immunoglobulin molecule	2
1.1.3	The immunoglobulin fold and interactions between the domains	4
1.2	Structure of the immunoglobulin combining site	6
1.2.1	Association of VL and VH domains	8
1.2.2	Framework regions and CDR conformation	10
1.2.3	Shape (topology) of the immunoglobulin combining site	12
1.2.4	Common residues located within immunoglobulin combining sites	13
1.3	Antibody/antigen complexation	14
1.3.1	Relationship of thermodynamic parameters and structure of antibody/antigen complexation	14
1.3.2	The role of hydrogen bonds, van der Waal's interactions and neutralisation of charged side-chains in antibody/antigen complexation	15
1.3.3	Contribution of aromatic residues to antibody/antigen complexes	18

1.3.4	Buried surface area at the interface of antibody and antigen	18
1.3.5	Role of water at the antibody/antigen interface	19
1.3.6	Conformational changes in antibody and antigen due to complex formation	20
1.4	Polyreactive immunoglobulins	22
1.4.1	Basic immunology of polyreactive antibodies	22
1.4.2	Antigen arrays bound by polyreactive antibodies	23
1.4.3	Gene utilisation and somatic mutation of polyreactive immunoglobulins	24
1.4.4	Proposed roles of polyreactive antibodies in the immune system	24
1.4.5	Hypothesised structure of polyreactive immunoglobulin combining sites	26
1.5	The chronic B lymphocytic leukaemia model of polyreactivity	26
1.6	The experimental system	27
CHAPTER TWO	Variable regions of immunoglobulins expressed in human chronic B lymphocytic leukaemia	28
2.1	Introduction	28
2.1.1	Genes encoding human variable domains	28
2.1.2	Junctional diversity within variable regions	30
2.1.3	Somatic hypermutation	30
2.1.4	Characteristics of variable region genes utilised by polyreactive antibodies	31
2.2	Materials and Methods	34
2.2.1	General reagents	34
2.2.2	The patients	34
2.2.3	Isolation of total cellular RNA	34
2.2.4	Purification of messenger RNA and first strand cDNA synthesis	35
2.2.5	Oligonucleotide primers	36
2.2.6	Polymerase chain reaction (PCR)	36
2.2.7	Cloning of PCR products for DNA sequencing	36
2.2.8	Dideoxy DNA sequencing of cloned genes	37
2.2.9	Analysis of variable region gene expression	37

2.2.10	Phenotyping of B CLL cells	38
2.2.11	Binding of mouse IgG by B CLL cells	38
2.3	Results	39
2.3.1	Amplification, cloning and sequencing of B CLL variable region genes	39
2.3.2	Analysis of variable region gene expression	45
2.3.3	Primary structure of B CLL variable domains	45
2.3.4	Phenotyping of B CLL lymphocytes	52
2.3.5	Low affinity binding to mouse IgG1 (K.1.21)	52
2.4	Discussion	59
2.4.1	Monoclonality of PBLs from B CLL patients	59
2.4.2	Immunoglobulin gene usage in B CLL	60
2.4.3	Incidence of somatic mutation in variable region genes	63
2.4.4	Junctional diversity in variable regions	64
2.4.5	Primary structures of variable domains	65
2.4.6	Low affinity binding to mouse IgG1 by B CLL cells	65
2.4.7	Conclusion	66
CHAPTER THREE	Homology modelling of B CLL variable regions (Fv): Structural diversity of human polyreactive Ig combining sites	68
3.1	Introduction	68
3.1.1	Homology-based protein modelling	68
3.1.2	Homology modelling of immunoglobulin variable domains	70
3.2	Materials and Methods	71
3.2.1	Immunoglobulins	71
3.2.2	Computer-aided modelling	71
3.2.3	Modelling strategy	71
3.2.4	Comparison of structures	72
3.3	Results	73
3.3.1	Validation of modelling strategy	73
3.3.2	Template models of polyreactive immunoglobulins	76
3.3.3	Refined Fv models	83
3.3.4	Electrostatic surface models and location of aromatic side-chains in polyreactive binding sites	83
3.4	Discussion	93
3.4.1	Validation of modelling strategy	93

3.4.2	Modelling the binding sites of polyreactive immunoglobulins	94
3.4.3	Refinement of variable domain models	96
3.4.4	Diversity of human polyreactive immunoglobulin combining sites	97
3.4.5	Conclusion	99
CHAPTER FOUR	Cloning and bacterial expression of B CLL derived immunoglobulins as Fv molecules	100
4.1	Introduction	100
4.1.1	Architecture of bacterial expression vectors	100
4.1.2	Denaturation and refolding of insoluble antibody fragments	101
4.1.3	Periplasmic expression of soluble antibody fragments	101
4.1.4	Structural studies using bacterially expressed antibody fragments	102
4.2	Materials and Methods	105
4.2.1	General reagents	105
4.2.2	Oligonucleotide primers	105
4.2.3	Polymerase chain reaction	106
4.2.4	Automated DNA sequencing	106
4.2.5	Cloning of VL and VH genes and construction of a dicistronic operon in pFLAG-CTS	107
4.2.6	PCR site-directed mutagenesis	109
4.2.7	Analysis of expressed proteins	110
4.2.8	Optimisation of protein expression	110
4.2.9	Affinity and size-exclusion chromatography of bacterially expressed Fv	111
4.3	Results	112
4.3.1	Cloning of light and heavy chain variable region genes derived from B CLL cells into bacterial expression vectors	112
4.3.2	Construction of a dicistronic operon for the soluble expression of VL and VH domains	114
4.3.3	Mutagenesis of the Tre dicistronic construct and sequencing of dicistronic operons of Bel and Tre Fv expression vectors	114

4.3.4	Optimisation of protein expression from Bel and Tre pFLAG-CTS (VL–VH) vectors	119
4.3.5	Affinity purification of expressed proteins	123
4.3.6	Size-exclusion chromatography of Bel and Tre bacterial expression cultures	125
4.4	Discussion	127
4.4.1	Development of dicistronic vectors for soluble expression of B CLL variable region fragments	127
4.4.2	Characterisation of protein expressed by Bel and Tre pFLAG-CTS (VL–VH) expression systems	128
4.4.3	Affinity and size-exclusion chromatography of Bel and Tre expressed protein	129
4.4.4	Conclusion	131
CHAPTER FIVE	Conclusions and perspective	132
Appendix A		137
Appendix B		145
Appendix C		150
References		152

List of figures

Figure	Title	Page
1.1	Schematic representation of stages in the production of intact antibodies by B cells	3
1.2	The immunoglobulin fold	5
1.3	Representation of an Fab molecule	7
1.4	Nature of the VL–VH (Fv) association	9
2.1	Consensus nucleotide sequence of Bel VL region	40
2.2	Consensus nucleotide sequence of Tre VL region	41
2.3	Consensus nucleotide sequence of Yar VL region	42
2.4	Consensus nucleotide sequence of Hod VL region	43
2.5	Consensus nucleotide sequence of Jak VL region	44
2.6	Translated amino acid sequences of five B CLL light chain V domains	49
2.7	Translated amino acid sequences of five B CLL heavy chain V domains	50
2.8	Nature of residues in B CLL CDR loops	51
2.9	Isotyping of B CLL lymphocytes	53
2.10	Two-colour immunofluorescence of B CLL cells	54
2.11	Low affinity binding of mouse IgG1/κ (K.1.21) by B CLL cells	56
2.12	Low affinity binding of mouse IgG1/κ (K.1.21) by B CLL cells	57
3.1	Comparison of known and predicted conformations of 3D6 CDR loops	75
3.2	Cα representations of the HCDR3 conformations	80
3.3	Side views of the two Yar Fv models	81
3.4a,b	Stereo Cα diagrams of the side views of the polyreactive Fv homology models	84,85
3.5a,b	Stereo Cα diagrams of the end-on views of the polyreactive Fv homology models	86,87
3.6	Ramachandran plot of the ψ, ϕ angles for all Cα dihedral angles in the Bel Fv homology model	89
3.7	Residue by residue comparison of template with refined Fv models	90
3.8	Electrostatic surface representations of the Fv homology models (end-on views)	91

3.9	Aromatic residues potentially contacting antigen	92
4.1	Dicistronic expression vector for soluble Fv expression	108
4.2	Cloning of the VL and VH region genes of B CLL immunoglobulins into bacterial expression vectors	113
4.3	Ligation-PCR construction of VL–VH dicistronic operons	115
4.4	Site-directed mutagenesis of the first codon of Tre VL region gene	116
4.5	DNA and translated amino acid sequence of the Bel Fv pFLAG-CTS (VL–VH) dicistronic construct	117
4.6	DNA and translated amino acid sequence of the Tre Fv pFLAG-CTS (VL–VH) dicistronic construct	118
4.7	Time course of Bel and Tre Fv expression in pFLAG-CTS	120
4.8	Optimisation of IPTG concentration	121
4.9	Host <i>E.coli</i> cell strains	122
4.10	Affinity purification of expressed protein	124
4.11	Size-exclusion chromatography of bacterial expression culture supernatants	126

List of tables

Table	Title	Page
2.1	Analysis of light chain variable region gene expression in B CLL patients	46
2.2	Mutations from germline Ig light chain gene sequences	47
2.3	Analysis of heavy chain variable region gene expression in B CLL patients	48
2.4	Phenotypes of B CLL lymphocytes	55
2.5	Low affinity binding to mouse IgG1 by B CLL cells	58
3.1	Comparison of the 3D6 crystal structure with the homology Fv model of 3D6	74
3.2	Variable domains used to construct Fv model framework regions	78
3.3	CDR sequences of the polyreactive Fv molecules and the CDR templates selected for modelling	79
3.4	Interchain interaction energies of two alternative Yar Fv models	82
3.5	Fv model stereochemistry after positional refinement	88

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	<i>alpha</i> carbon atom	PAGE	polyacrylamide gel electrophoresis
CDR	complementarity determining region	PBL	peripheral blood lymphocytes
CIAA	24:1 Chloroform:isoamyl alcohol	PBS	phosphate buffered saline
CRI	cross reactive idio type	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.