Solution structure of Interleukin-1β bound to a potential therapeutic antibody

Ian Craig Wilkinson BSc, MSc (Nottingham)

Thesis submitted for the Degree of Doctor of Philosophy

Department of Biochemistry

University of Leicester

March 2009

Acknowledgements

I would first like to extend my immense gratitude to my supervisor Professor Mark Carr for giving me the opportunity to conduct this research and for his continual support and guidance. I am indebted to Dr. Alistair Henry and Dr. Richard Taylor for their advice and assistance during my PhD, and to Professor Gordon Roberts for his insightful input during my committee meetings. I am also grateful to the Biotechnology and Biological Sciences Research Council and UCB-Celltech for funding this studentship.

I would like to acknowledge those who have contributed to the work reported in this thesis. In particular I thank Dr. Vaclav Veverka, who always gave up his time to help me and made an excellent golf partner, Catherine Hall for her thoughtful opinions and friendship, and to Dr. Paul Stephens, who passed on a large amount of his molecular biology experience to me during the initial stages of my project. I offer my thanks to all members of the department both past and present. In particular I would like to thank Dr. Philip Renshaw, Dr. Kirsty Lightbody, Dr. Lorna Waters and Dr. Fred Muskett for their scientific input and friendship, and to Jon, Ali, Ojay, Dariush, Sid, Abbas, Sarah, Sami, Phil and Alice for helping to make the lab a truly enjoyable environment to work in. To all my friends who I have not been able to mention, thank you for your friendship.

My sincerest thanks go to my mum, dad, brother Gareth and grandparents for their continuing support and love that they have shown me throughout my life. Without them I would not have reached this point. Finally, from the bottom of my heart I thank Stacey for her immeasurable love and support over the last three years.

Ian Craig Wilkinson, Leicester, March 2009.

Abstract

The ability of antibodies to bind to target proteins with high specificity makes them an important class of biotherapeutic. The minimal functional form of an antibody, a single chain Fv (scFv), consists of two variable domains joined by a linker. Many therapeutics are based on the larger Fab fragments, which consist of two variable and two constant domains. Detailed structural information for potential therapeutic antibody-target protein complexes could clearly provide valuable insights into their mechanisms of action and mode of binding, however, currently there are very few structures available and none determined by NMR-based methods. To date, the collection of detailed structural information for scFvs has been limited by the formation of domain-swapped dimers. This equilibrium results in the formation of almost 50% dimer at the concentrations required for NMR. However, the work reported here shows that the monomeric form can be stabilised when bound to its target protein, which has enabled NMR-based structural analysis. This has allowed triple resonance data to be collected for the first time for a scFv-target protein complex, enabling over 90% of the backbone resonances to be assigned. Collection of chemical shift perturbation data, intraresidue NOEs and backbone amide RDCs has enabled a homology model for the scFv IC8 to be refined and docked to the structure of its target protein IL-1β. This has produced a reliable and well defined structure for the scFv-IL-1β complex. Evidence has also been obtained for a reorientation of the two variable domains induced by IL-1β binding, which could reflect conformational changes associated with antibody signalling. This study demonstrates that NMR spectroscopy can be used to obtain detailed structural information for antibody-target protein complexes, providing valuable information that could be used in the design and selection of future therapeutic antibodies.

Contents

Abbreviations

Single and three letter codes for amino acids

Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparagine	N	Asn	Methionine	M	Met
Aspartic acid	D	Asp	Phenylalanine	F	Phe
Cysteine	C	Cys	Proline	P	Pro
Glutamine	Q	Gln	Serine	S	Ser
Glutamic acid	E	Glu	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile	Valine	V	Val

DNA bases

Adenine	A	Guanine	G
Cytosine	C	Thymine	T

Abbreviations

$\Delta\delta$	Combined chemical shift difference
ϕ	Phi, the amide nitrogen alpha carbon dihedral angle
ψ	Psi, the alpha carbon carboxyl carbon dihedral angle
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
AIR	Ambiguous interaction restraint
Ag	Antigen
CAPRI	Critical assessment of predicted interactions

CD	Circular dichroism
CDR	Complementarity determining region
C_H	Constant heavy domain
C_L	Constant light domain
CSI	Chemical shift index
Da	Axial tensor component
Dr	Rhombic tensor component
DNA	Deoxyribonucleic acid
DTT	1,4-dithiothreitol
ECL	Electrochemical luminescence
EDTA	Ethylenediaminetetraacetic acid
EIN	Amino-terminal domain of enzyme 1 from *Escherichia coli*
Fab	Antigen binding fragment
Fc	Crystallisable fragment
FcRn	Neonatal crystallisable fragment receptor
FDA	United States food and drug administration
FPLC	Fast protein liquid chromatograpy
FR	Framework region
Fv	Variable fragment
GlcNAc	*N*-acetyl-glucosamine
HADDOCK	High ambiguity driven protein-protein docking
HEL	Hen egg white lysozyme
HPLC	High performance liquid chromatography
HPr	Histidine containing phosphocarrier protein
HRP	Horse radish peroxidase

HSQC	Heteronuclear single quantum coherence
Ig	Immunoglobulin
IL-1	Interluekin-1
IL-1Ra	Interleukin-1 receptor antagonist
IL-1RI	Interleukin-1 receptor type I
IMGT	International immunogenetics information system
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
ITAM	Immunoreceptor tyrosine-based activation motif
Kav	Average participation coefficient
LB	Luria-Bertani broth
LDS	Lithium dodecyl sulphate
MES	2-(N-morpholino)ethanesulphonic acid
MHC	Major histocompatibility complex
MW	Molecular weight
Ni-NTA	Nickel-nitriloacetic acid
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PALES	Prediction of alignment from structure
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulphnoyl fluoride

PVDF	Polyvinylidene fluoride
R	Rhombicity
RDC	Residual dipolar coupling
RMSD	Root mean square deviation
ScFv	Single chain variable fragment
ScFv-M	Alanine to Aspartate mutant single chain variable fragment
ScFv-X	Extended linker single chain variable fragment
ScFv-XM	Extended linker mutant single chain variable fragment
SDS	Sodium dodecyl sulphate
SLAM	Selected lymphocyte antibody method
TALOS	Torsion angle likelihood obtained from shift and sequence similarity
TBST	Tris buffered saline with tween-20
TEV	Tobacco etch virus
TNF-α	Tumour necrosis factor α
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
TROSY	Transverse relaxation-optimised spectroscopy
UV	Ultraviolet
Vc	Column volume
V(D)J	Variable, diversity and joining immunoglobulin gene segments
Ve	Elution volume
V_H	Variable heavy domain
VHH	Camelid variable heavy domain
V_L	Variable light domain
V-NAR	Shark variable heavy domain

Vo	Void volume
WATERGATE	Water suppression by gradient-tailored excitation
WCL	Whole cell lysate

Chapter 1

General Introduction

1.1 Antibodies

The recognition of foreign material is the hallmark of the specific adaptive immune response in mammals, of which immunoglobulins are an integral part. Immunoglobulins are a group of glycoproteins present in the tissue and fluids of all mammals. Some are on the surface of B lymphocytes (or B-cells) and others, known as antibodies, are free in the blood or lymph. Over 100 years ago Paul Ehrlich was the first to propose a model for an antibody molecule in which the antibody was branched and consisted of multiple sites for binding to foreign material, known as antigen, and for the activation of the complement pathway (Davies & Chacko, 1993). This model agreed with the 'lock and key' hypothesis for enzymes proposed by Emil Fischer (Fischer, 1894; Lemieux & Spohr, 1994) and still in general terms holds true today. While one part of the antibody, the antigen binding fragment (Fab), recognises the antigen, the other part of the antibody, known as the crystallisable fragment (Fc), interacts with other elements of the immune system such as phagocytes or components of the complement pathway to promote removal of the antigen.

Antibodies all have the same basic structure consisting of two heavy and two light chains forming two Fab arms containing identical domains at either end attached by a flexible hinge region to the stem of the antibody, the Fc domain, giving the classical 'Y' shape. The chains fold into repeated immunoglobulin folds (Poljak *et al.*, 1973), which form either constant or variable domains. The Fab domains consist of two variable and two constant domains, with the two variable domains making up the variable fragment

(Fv), which provides the antigen specificity of the antibody (Inbar *et al.*, 1972) with the constant domains acting as a structural framework. Each variable domain contains three hypervariable loops, known as complementarity determining regions (CDRs), evenly distributed between four less variable framework (FR) regions. It is the CDRs that provide a specific antigen recognition site on the surface of the antibody and the hypervariability of these regions enables antibodies to recognise an almost unlimited number of antigens (Morea *et al.*, 2000).

Antibodies fall into five main classes (IgA, IgD, IgE, IgG and IgM), which differ in the heavy chains, hinge structure and the valency of the antibody. Figure 1.1 shows a diagram of the IgG antibody and comparison of this structure with the other antibody isotypes is shown in figure 1.2.

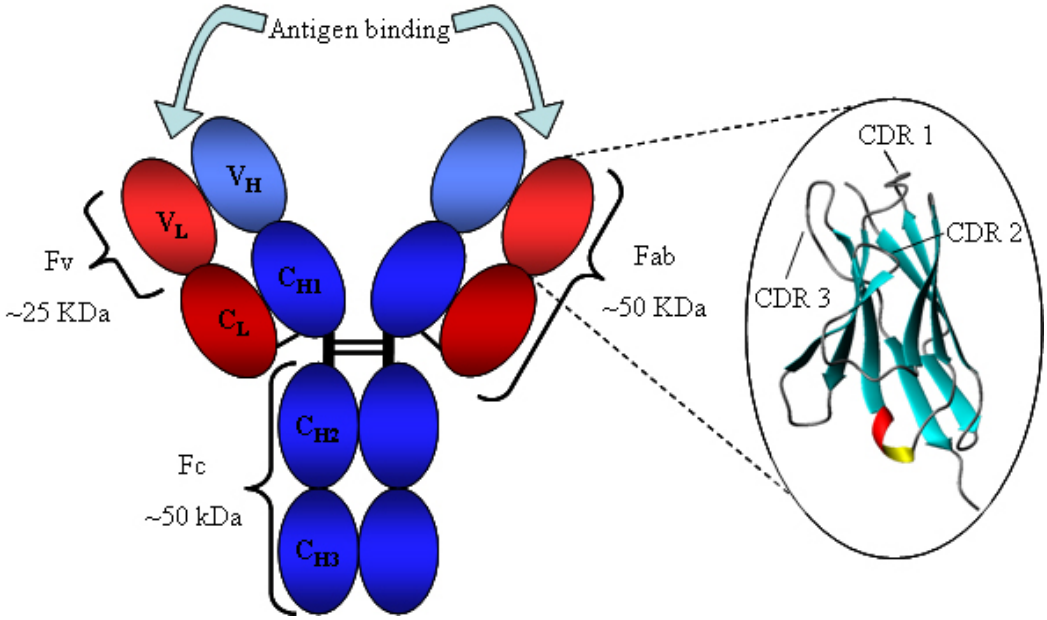

Figure 1.1 A schematic diagram of an IgG antibody consisting of two heavy chains (blue) and two light chains (red) folded into constant (C) and variable (V) domains. The enlargement of the variable domain shows a ribbon diagram of the β-sheet framework and CDR loops.

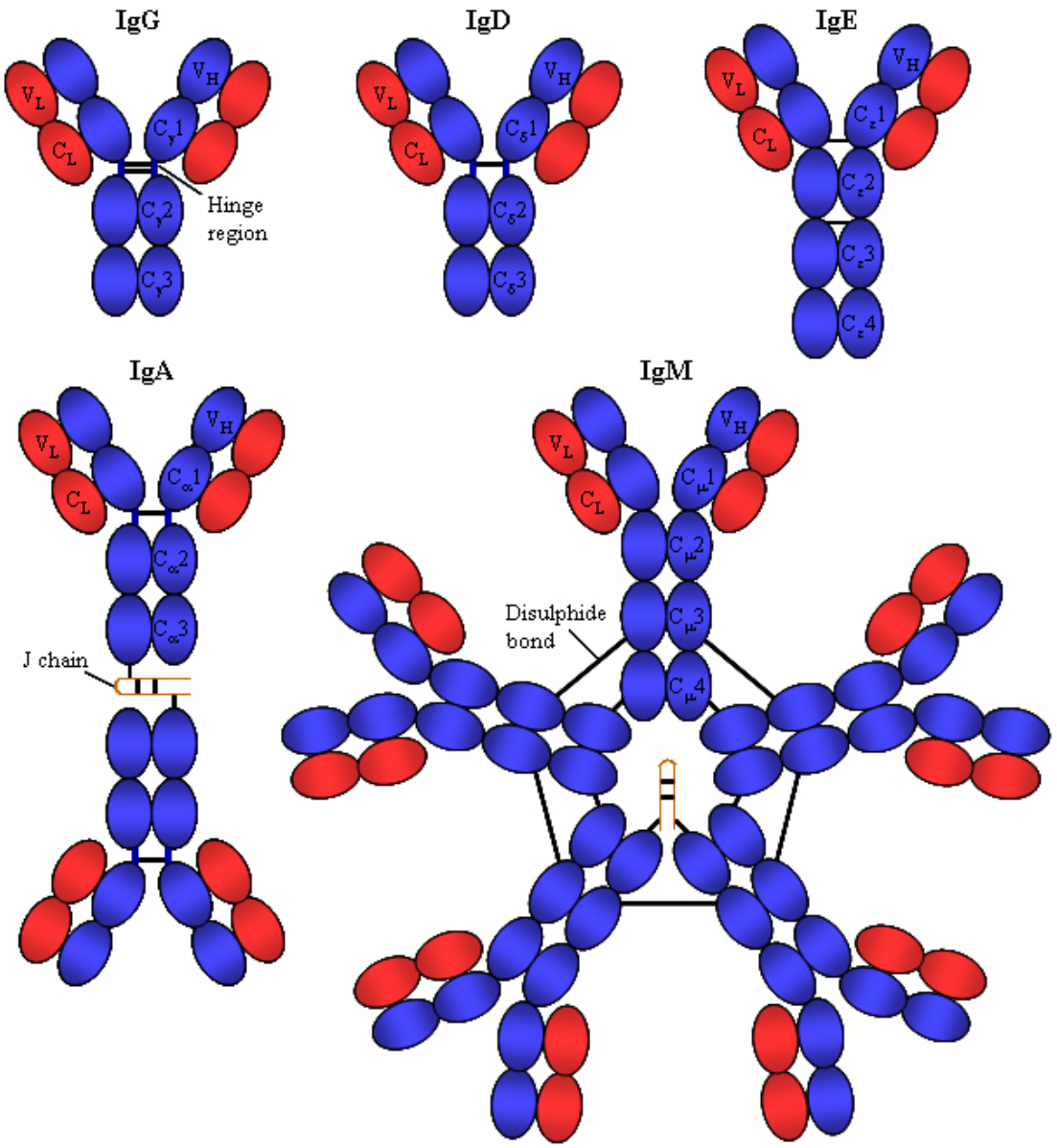

Figure 1.2 A representation of the five classes of human antibodies. Each consists of heavy chains, shown in blue, and light chains, shown in red, forming constant and variable domains. The antibody isotype is determined by the type of constant heavy domain.

IgG is the most abundant antibody in normal human serum, accounting for 70-75% of the total immunoglobulin pool. It is monomeric with a molecular weight of approximately 150 kDa and is the major antibody of the secondary immune response.

IgG consists of four human subclasses (IgG1, IgG2, IgG3 and IgG4) each containing a different heavy chain. They are highly homologous and differ mainly in the hinge region and the extent to which they activate the complement pathway. IgM accounts for approximately 10% of the immunoglobulin pool and is the predominant antibody in the primary immune response. It is a pentamer of the basic four chain structure with a molecular weight of approximately 1 MDa and thus due to its size is largely confined to the intravascular pool. IgA represents approximately 15-20% of the antibody pool and either exists as a monomer or a dimer. IgA is the predominant antibody in mucous secretions such as saliva, tears, milk and intestinal juice. IgD accounts for less than 1% of the total plasma immunoglobulin but is present in large quantities on the membrane of B-cells. The precise function of this class of antibody is still unknown. IgE is very scarce in the serum but is found on the basophils and mast-cells of all individuals. In a similar manner to IgM, IgE has two additional constant domains in place of the hinge region. This class may play a role in immunity to parasites but is more commonly associated with type I immediate hypersensitivity, where an IgE immune response occurs to innocuous environmental antigens such as pollen and peanuts.

The immense diversity of antibodies produced by the mammalian immune system depends on the ability of B-cells to somatically alter their genomes by several distinct methods. Developing B-cells in the bone marrow assemble the exons that encode the immunoglobulin heavy and light chain variable domains from component variable (V), diversity (D) and joining (J) segments through a process known as V(D)J recombination (Tonegawa, 1983). This site-directed antigen independent recombination event is initiated by double strand cleavage of the DNA and is completed by non-homologous end joining, resulting in the random joining of V, D and J segments to create a

functional heavy chain gene and joining of a V and J gene segment to produce a light chain gene (Oka & Kawaichi, 1995). Due to the number of potential V, D and J segments and the different pairing of V_H and V_L domains this can lead to over 10^6 different combinations, which is further increased by the phenomenon of junctional flexibility. This is brought about by the imprecise joining of the coding sequences, which can result in non-productive rearrangements due to the segments being joined out of phase but when the reading frame is maintained productive combinations can encode for alternative amino acids at the joining site. The joining site falls within the third CDR of both the V_H and V_L sequences and thus any amino acid changes generated by junctional flexibility can have a major impact in generating antibody diversity. This process leads to the generation of mature immunocompetent B-cells that are antigenically committed to a single epitope and express membrane-bound immunoglobulin on their surface. This leads to the generation of antibody diversity in the mammalian immune system estimated to be in the range of 10^8 to 10^{11} (Goldsby *et al.*, 2000).

The membrane-bound immunoglobulin (IgM or IgD) acts as a B-cell antigen receptor and plays two roles once bound to an antigen. Firstly, it sends an activating signal to the interior of the cell when antigen binds. Secondly, it delivers the antigen to intracellular sites where the antigen is degraded into peptides and these are presented on the surface of the B-cell bound to major histocompatibility (MHC) class II molecules. These surface complexes are recognised by helper T-cells, which then secrete cytokines causing the B-cells to proliferate and differentiate into antibody secreting cells and memory B-cells. Some of the activated B-cells migrate into a primary lymphoid follicle and form a germinal centre. It is here that the process known as antibody affinity

maturation occurs. Somatic hypermutation specifically targets a region that includes the V(D)J segments but not the constant domains resulting in nucleotides in this region being replaced with alternative bases (Diaz & Casali, 2002). This creates a series of clones that differ subtly in their specificity and affinity for the antigen. Somatic hypermutation occurs at a rate of approximately 10^{-3} per base pair per generation, which is a million fold higher than the spontaneous mutation rate in other genes, and thus on average every second B-cell acquires a mutation at each division. Most mutations have a negative affect and these cells are removed by apoptosis but those mutations that cause an increase in antigen affinity are efficiently selected and expanded (Murphy *et al.*, 2007).

The first antibody secreted by the B-cells is IgM but in later stages other classes of antibody such as IgG, IgA and IgE are much more abundant. This is due to class-switch recombination where the heavy chain DNA undergoes a further irreversible rearrangement in which the V(D)J region combines with any C_H gene segment, with the type of constant gene determining the antibody isotype that the B-cell secretes (Wabl & Steinberg, 1996). The secreted affinity maturated antibodies can then contribute to immunity in three main ways with the exact process being determined by the class of antibody. Antibodies can neutralise viruses, intracellular bacteria and toxins by binding and preventing them from targeting the cell surface. Second, antibody constant domains are recognised by Fc receptors on phagocytes thus facilitating the phagocytosis of pathogens. In a similar manner the Fc domain of antibodies also binds proteins in the complement pathway, which recruits phagocytic cells and can lyse certain microorganisms directly.

Signalling within the immune response is still not completely understood and, of particular relevance to the work undertaken in this thesis, the exact mechanisms of antigen induced activation of B and T-cells are unknown. The antigen receptor complexes are associated with invariant accessory signalling proteins (figure 1.3), which contain immunoreceptor tyrosine-based activation motifs (ITAM). ITAMs contain two YXXL/I motifs separated by six to nine amino acids with antigen binding inducing phosphorylation of the tyrosines and providing sites for recruitment of downstream signalling proteins (Dal Porto *et al.*, 2004). It has been hypothesised that a conformational change induced upon antigen binding could pass through the receptor to the transmembrane signalling proteins but there is currently no structural data to support this signal propagation theory (Tolar *et al.*, 2008). Current theories propose that signalling is initiated when two or more receptors are cross-linked by a multivalent antigen. This idea, however, does not explain how binding of a monovalent antigen activates an antibody response.

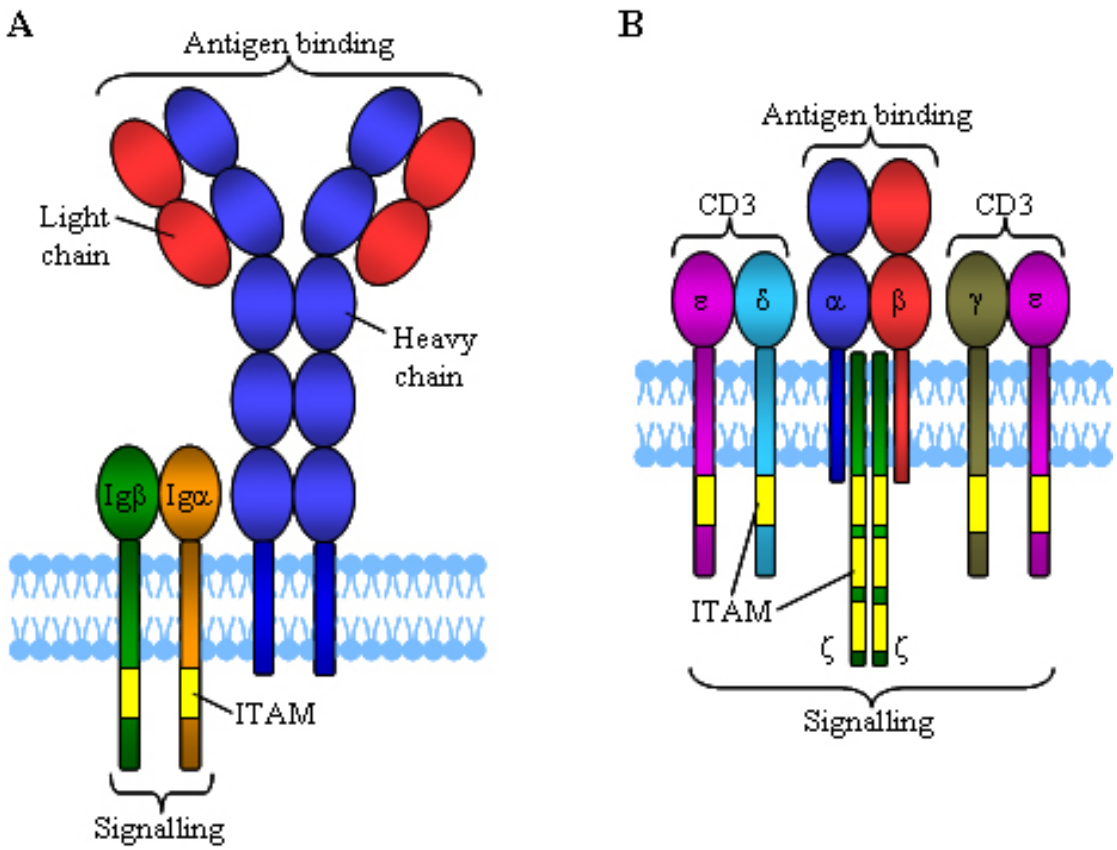

Figure 1.3 The B-cell receptor (A) is made up of a cell surface IgM or IgD (IgM is shown in this case) bound to the signalling molecules Igα and Igβ. These each have a single ITAM motif, shown in yellow, in their cytoplasmic tails that enables them to signal when the B-cell receptor binds antigen. The T-cell receptor (B) consists of an antigen recognition heterodimer of α and β chains, similar to a Fab, associated with a complex of four other signalling proteins (two ε, one δ and one γ) collectively called CD3. The receptor is also associated with a homodimer of ζ chains. The ITAM motifs in the cytoplasmic tails of the signalling proteins act in a similar manner to those of the B-cell receptor.

1.2 Therapeutic antibodies

The acquisition of immunity to a disease that a patient has already encountered has been documented for many centuries but it wasn't until the work of Edward Jenner in 1798 that it was most notably shown that this immunity could be transferred. This was demonstrated by inoculating a boy with the fluid from a cowpox pustule giving him immunity to the very similar but much more serious disease smallpox (Riedel, 2005). However, the true potential of antibodies as therapeutic agents or 'magic bullets', a term first coined by Paul Ehrlich over 100 years ago, was not able to be realised until the groundbreaking work by Kohler and Milstein on the production of monoclonal antibodies from immunised mouse spleen cells fused with mouse myeloma cells (Kohler & Milstein, 1975).

The early success of murine monoclonal antibodies led to the U.S. Food and Drug Administration (FDA) approval of the first therapeutic antibody OKT3 (muromonab) in 1986, which is still in use today for the treatment of kidney transplant rejection. However, most murine antibodies were shown to have limited use as therapeutic agents because of a short serum half-life, an inability to trigger human effector functions and in particular they were recognised by the patients' immune systems as foreign proteins resulting in a human anti-mouse antibody response. In an attempt to reduce the immunogenicity of the mouse antibodies, genetic engineering was used to generate chimeric antibodies containing human constant domains and the mouse variable domains to retain the specificity (Boulianne *et al.*, 1984; Morrison *et al.*, 1984). This was then taken a step further by grafting of the CDRs from a murine antibody onto a human variable region framework, creating humanised antibodies (Jones *et al.*, 1986). The crystal structures of rat and humanised antibodies later showed that an

appropriately selected human scaffold correctly supports the orientation of the CDRs (Cheetham *et al.*, 1998). The speed of approval of this new wave of therapeutic antibodies was slower than expected due to the increase in regulatory burdens and it was not until 1996 that the FDA approved the first chimeric antibody ReoPro (abciximab), which lessons the risk of blood clots in patients with cardiovascular disease by binding to a receptor on platelets (Lefkovits *et al.*, 1996), and the first humanised antibody Zenapax (daclizumab) in 1997, which is an anti-CD25 antibody used to prevent organ transplant rejection (Przepiorka *et al.*, 2000). The drive to reduce immunogenicity by decreasing the mouse content of monoclonal antibodies inevitably culminated in the creation of so-called fully human antibodies, which was made possible by the expression of isolated human variable domain genes in *E. coli* (Better *et al.*, 1988; Skerra & Pluckthun, 1988). Two of the most widely used techniques developed for the production of fully human monoclonal antibodies are phage display, where a library of human antibodies is expressed on the surface of phage and subsequently selected and amplified in *E. coli* (Mccafferty *et al.*, 1990), and transgenic mice expressing a human antibody repertoire (Green *et al.*, 1994).

The chimeric, humanised and fully human monoclonal antibodies are much less immunogenic than the original murine antibodies but human anti-antibody responses have still been noted in patients. Removal of the mouse constant domains clearly removed the most immunogenic portion of the antibody but it is argued by some that subsequent humanisation has had little or no affect on the immunogenicity of therapeutic antibodies (Clark, 2000). More recent evidence suggests that while the effects of humanising the variable domains are not as dramatic as the removal of murine

constant domains there is some reduction in the severity of the human anti-antibody response to humanised or fully human antibodies (Hwang & Foote, 2005).

The majority of therapeutic antibodies currently on the market are chimeric or humanised and all but one of the antibodies are of the IgG class for several practical reasons. IgG antibodies are very stable and easily purified and stored. *In vivo* they have a long biological half-life both due to their size and their interaction with the FcRn receptor (Junghans & Anderson, 1996). This salvage receptor protects IgG from the default degradative pathway within cells and recycles it back into the plasma. The four IgG subclasses show more than 95% homology in the amino acid sequences of the constant domains but show their most conspicuous differences in the composition and structure of the hinge region. IgG1 has a hinge length of 15 amino acids, which provides a large amount of flexibility and enables activation of the complement pathway via C1q binding. The hinge regions of IgG2 and 4 are shorter, with a length of 12 residues, which reduces the flexibility decreasing the complement activation of IgG2 and in the case of IgG4 abolishing complement activation completely. IgG3 has an elongated hinge of 62 amino acids, which results in this subclass being the most potent activator of the complement pathway but also leads to a much shorter half-life due to increased proteolysis in the hinge region (Murphy *et al.*, 2007). These factors must, therefore, be taken into consideration when selecting the therapeutic antibody subclass, with IgG1 commonly chosen if complement activation is desired and IgG4 selected if it is not.

Differential IgG glycoslyation also plays a pivotal role in the activity of therapeutic antibodies but its importance is often neglected (Wuhrer *et al.*, 2007). The Fc region of an IgG molecule consists of two paired CH3 domains and, in contrast, two CH2

domains that are separated and do not interact but have two oligosaccharide chains interposed between them. These chains cover the hydrophobic faces that would normally lead to domain pairing. The *N*-glycans contain a common core region of two *N*-acetyl-glucosamine residues (GlcNAc) linked to asparagine via an amide bond and three mannose residues. This core structure may contain additional terminal sugars, such as mannose, GlcNac, galactose, fucose and sialic acid, generating a large amount of heterogeneity (Wright & Morrison, 1997). The first indication that differences in glycosylation affect antibody function came from a study reporting an increase in G0 glycoforms (no galactose residues) in rheumatoid arthritis with disease progression, which returned to normal when patients went into remission (Parekh *et al.*, 1985). Other studies have subsequently reported that the presence of fucose can influence binding of IgG to Fc receptor molecules enhancing antibody-dependent cell mediated cytotoxicity (Shields *et al.*, 2002; Shinkawa *et al.*, 2003), whereas the lack of galactose lowers the *in vitro* cell dependent cytotoxicity capacity of antibodies (Wright & Morrison, 1998), and sialylation has been shown to have both positive and negative effects on antibody functions (Anthony *et al.*, 2008; Kaneko *et al.*, 2006; Scallon *et al.*, 2007). The described heterogeneity in glycan structure is also dependent on species, age and gender (Raju, 2008). All together selection of a cell-line for production of therapeutic antibodies must be driven by the desired antibody effector mechanisms and their dependence on fucosylation, galactoslyation and/or sialylation (van Berkel *et al.*, 2009).

Therapeutic antibodies function by three principal modes of action: (i) by blocking the actions of specific molecules; (ii) by targeting specific cells; (iii) by functioning as signalling molecules. The blocking activity is achieved by preventing growth factors, cytokines or other soluble mediators reaching their target receptors, which is

accomplished by antibody binding to either the factor itself or its receptor. Targeting involves directing antibodies against a specific population of cells and typically antibodies are engineered to carry effector moieties such as enzymes, toxins or radionuclides to the target cells, where the attached moiety can then exert its effect. This approach is becoming increasingly common in the anti-cancer therapeutic antibodies currently undergoing clinical trials (Reichert & Valge-Archer, 2007). Therapeutic antibodies can also utilise their natural effector functions to bind to complement proteins and induce complement dependent cytotoxicity.

In addition to the immunogenicity of therapeutic antibodies discussed earlier the possible side effects of antibody therapy include cytokine release syndrome, which is usually mediated through recruitment of immune effector cells. This was tragically the case in the phase 1 trial of the CD28 antibody TGN1412, which lead to a cytokine storm and multiple organ failure in six healthy volunteers (Suntharalingam *et al.*, 2006). In this unusual case the effects were due to the antibody acting as a superagonist by binding to and stimulating T-cells without the need for antigen-receptor activation. Use of the IgG2 or IgG4 subclasses can alleviate some of the problems caused by activation of antibody effector functions as these isoforms lack some of the effector functions (Clark, 1997) but a more common solution is the use of antibody fragments that lack the Fc domain (Holliger & Hudson, 2005). This also helps to reduce the other main failure of therapeutic antibodies, namely the lack of delivery, which is especially true for anti-cancer antibodies. Solid tumours have substantial physical barriers often preventing the penetration of antibodies to the centre and resulting in reduced therapeutic effects (Christiansen & Rajasekaran, 2004). The use of smaller fragments enables deeper

penetration with the affinity of the antibody also being critical and if it is too high this will restrict its ability to penetrate a tumour (Adams *et al.*, 2001).

Over the past two decades antibodies have been dissected into smaller antigen binding fragments, initially by proteolysis and later by genetic engineering to produce mono or multivalent fragments. Figure 1.4 shows a selection of the antibody fragments that have been engineered. The most common forms are the Fab fragment, the single chain Fv (scFv), which will be discussed in more detail in chapter 2, and multivalent variants of these. There are a range of applications in which Fc mediated effects are not required and are even undesirable. Antibody fragments are ideal for these applications where the antibody is only being used to block a signalling molecule or receptor and thus effector functions are not necessary. In addition they are useful in imaging and cancer therapy, where a long serum half-life mediated by Fc interaction with the FcRn receptor results in poor contrast and in the case of radiolabelled antibodies fast clearance from the circulation via the kidney is also advantageous to reduce prolonged exposure (Brekke & Loset, 2003).

When mouse single V_H domains were shown to be functional (Ward *et al.*, 1989) it was proposed that because of their small size they could potentially target cryptic epitopes that have evolved in many pathogenic viruses to contain narrow cavities that bind to target receptors but are poorly accessible to intact antibodies. However, they rarely retain the affinity of their parent antibody and are poorly soluble and prone to aggregation (Ward *et al.*, 1989). Discovery of two types of organisms, the camelids and cartilaginous fish, that have evolved high affinity single variable domains (called the VHH domain in camelids (Muyldermans *et al.*, 1994) and the V-NAR domain in sharks

(Greenberg *et al.*, 1995)) mounted on an Fc equivalent constant domain framework resparked interest in this area. Both of these domains contain long CDR loops, often larger than those observed in conventional murine and human antibodies, and as such these are able to penetrate cavities in target antigens such as active sites (Stanfield *et al.*, 2004) and canyons in viral and infectious disease biomarkers (Nuttall *et al.*, 2004; Streltsov *et al.*, 2004). Although it is claimed that VHH domains are only minimally immunogenic (Cortez-Retamozo *et al.*, 2004), humanisation of these would presumably be necessary for *in vivo* administration and the use of human single domains would be preferable.

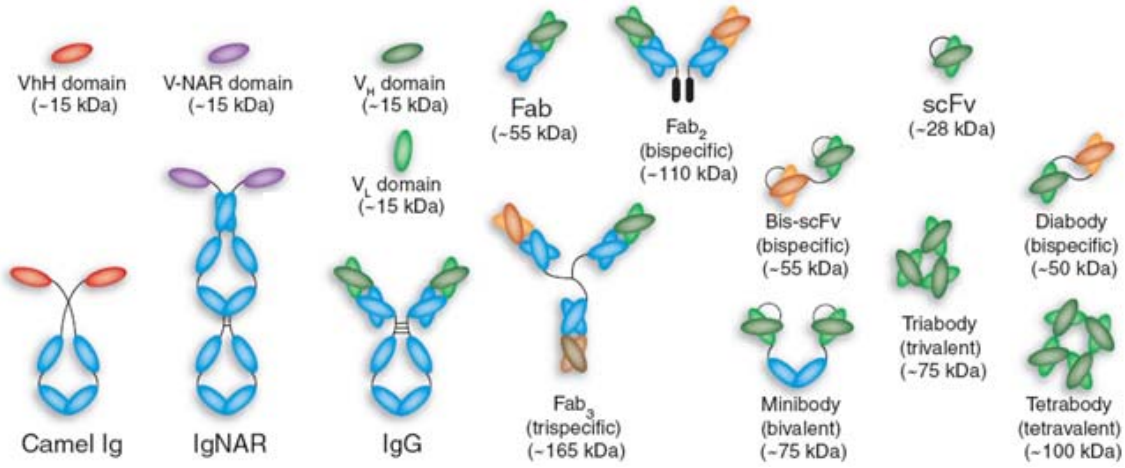

Figure 1.4 A schematic representation of different antibody formats showing the classical IgG molecule, camelid Ig and shark IgNar alongside antibody fragments generated from these (Holliger & Hudson, 2005).

Recent advances in this area have resulted in the engineering of 3 kDa antibody mimetics, consisting of only one CDR from the heavy chain and one from the light chain separated by a framework region (Qiu *et al.*, 2007). It was proposed that

separation of two CDRs by a framework region would allow the loops to assume a conformation similar to that of the parental antibody. The mimetics produced bound their antigens with 1-10% of the affinity of the parent antibodies. These were then coupled to a 75 kDa cytotoxin and targeted killing of tumour cells was observed. Far greater accumulation of the antibody mimetic was seen in the core of the tumour than for the full antibody, which can only in part be attributed to the reduction in size, once bound to the toxin the product is half the size of an IgG, but must also be due to the reduced affinity allowing for greater penetration.

For some applications it is desirable to engineer multivalent antibody fragments, which show a significantly increased functional affinity (avidity). Multivalent engagement of cell surface receptors, such as CD20 (Teeling *et al.*, 2004), can result in a desired activation and/or apoptosis through transmembrane signalling pathways. Multispecific fragments allow the direct association of two or more different targets with important applications in recruitment of cytotoxic T and natural killer cells (Kipriyanov & Le Gall, 2004).

Another area of antibody engineering that has shown promise is the covalent attachment of antibody fragments to polyethylene glycol (PEG), which is known as pegylation. This results in the size of the molecule increasing above the glomerular filtration limit and has also been shown to reduce immunogenicity and improve solubility (Chapman, 2002). Cimzia (certolizumab pegol) is an anti-TNF-α humanised Fab fragment covalently attached to PEG. It acts by blocking the signalling of TNF-α and is currently approved for use in Crohn's disease and may soon be approved for rheumatoid arthritis. The site specific pegylation in the hinge region has been shown to prolong the

circulating half-life of the Fab fragment to 14 days (Choy *et al.*, 2002). Unlike the two other anti-TNF-α monoclonal antibodies currently on the market, Remicade (infliximab) and Humira (adalimumab), and the fusion protein Enbrel (etanercept) consisting of two naturally occurring TNF-α receptors linked to the Fc portion of an IgG1, Cimzia does not suffer from the adverse effects associated with Fc mediated complement activation and cellular toxicity (Sandborn *et al.*, 2007).

All of these advances in the antibody field over the last 30 years have resulted in antibodies becoming the fastest growing area of the biopharmaceuticals market accounting for US sales worth $22 billion in 2007 (Therapeutic Monoclonal Antibody Report 2008-2023). The 21 FDA approved antibodies, which includes 8 'blockbuster' drugs that generate over $1 billion in revenue each year, are heavily focused in oncology, arthritis, immune and inflammatory disorders but also cover areas as diverse as viral infections, transplant rejection and allergies. However, there is a relatively high degree of duplication of specificities between the products under development. A recent review identified 23 antibody fragments in clinical or pre-clinical development (Holliger & Hudson, 2005). Twelve are targeted against four antigens already recognised by marketed antibodies (CEA, Her2/Neu, TNF-α and VEGF) and the other eleven are targeted against common antigens. This is because it is very rare to find truly disease specific antigens that have the appropriate properties to be effectively targeted by monoclonal antibodies. Successful antibody therapies are generally targeted against common molecules where there is sufficient operational specificity to achieve a clinical benefit with tolerable side effects. For these reasons cell surface receptors and cytokines, such as the interleukin family of proteins, are ideal targets for therapeutic antibodies.

1.3 Interleukin-1β

Interleukins are small, highly active proteins that are part of the cytokine family of secreted signalling molecules that are produced by a variety of cells and generally induce changes in cellular metabolism. Interleukin-1 (IL-1) was one of the first cytokines described and has been shown to trigger a wide range of effects including fever, damage to joints and an increase in the numbers of bone marrow cells. The IL-1 gene family is composed of IL-1α, IL-1β and IL-1Ra. IL-1α and IL-1β are both potent agonists, stimulating a wide range of biological responses in the pico and femtomolar ranges, while IL-1Ra is a naturally occurring antagonist that does not induce a signal. The sequence homology of the three proteins is 20-30%, with each of them forming a similar structure consisting primarily of a β-barrel composed of 12-14 β-sheets (Graves *et al.*, 1990; Priestle *et al.*, 1989; Vigers *et al.*, 1994). All three proteins bind to IL-1 receptors with similar affinities and during infections and inflammation they compete for IL-1 receptor occupancy with the host response being dependent on the net occupancy of IL-1α, IL-1β or IL-1Ra on the IL-1 receptors (Dinarello, 1994). In general IL-1α and IL-1β induce the same biological responses but a large amount of IL-1α remains in the cytosol or membrane bound, while the majority of IL-1β is released into the extracellular space and circulation (Dinarello & Wolff, 1993). The remainder of this section will concentrate on IL-1β and its therapeutic relevance as this is the antigen targeted by the antibodies in this thesis.

IL-1β is predominantly produced by activated macrophages and monocytes and its main role is as a pro-inflammatory cytokine. It activates local and systemic responses to infection and injury by generating fever, activating lymphocytes and promoting infusion of leukocytes into the sites of injury or infection (Dinarello, 1998). It is the primary

cause of chronic and acute inflammation and plays a key role in the febrile response as an endogenous pyrogen (Dinarello, 2000). IL-1β affects most cell types and exerts its activities extracellularly by binding to a high affinity receptor, IL-1RI. Upon IL-1β binding, the IL-1RI receptor associates with a related protein, IL-1RAcP, triggering the formation of a membrane signalling complex that activates the IκB kinase (IKK) and mitogen-activated protein kinase (MAPK) pathways (Sims, 2002), as shown in figure 1.5. This results in the activation of the transcription factors nuclear factor κB (NF-κB) and activator protein 1 (AP-1), causing the induction of genes encoding cytokines, actute-phase proteins, cell adhesion molecules and enzymes involved in the production of small pro-inflammatory substances (Luheshi, 1998).

IL-1β is implicated in many diseases including rheumatoid arthritis, inflammatory bowel disease and insulin dependent diabetes, and inhibitors have been shown to ease symptoms. IL-1's closest biological relative is the inflammation inducing protein tumour necrosis factor alpha (TNF-α). TNF-α inhibitors are prescribed to treat chronic inflammatory conditions and sales of the top three TNF-α inhibitors reached $10 billion in 2006, which for many years took the focus away from IL-1 research (Ledford, 2007). However, TNF-α inhibitors also heighten the risk of opportunistic infections and do not work in some patients. This has led to many pharmaceutical companies starting or restarting research projects on inhibitors of IL-1 and/or IL-1 receptors. Currently the only approved IL-1 drugs are Kineret (Anakinra) from Amgen, which is a recombinant form of IL-1Ra that is approved for rheumatoid arthritis treatment, and Arcalyst (Rilonacept) from Regeneron, which is a fusion protein that binds to and neutralises IL-1 and is approved for the treatment of Cryopyrin-Associated Periodic Syndrome.

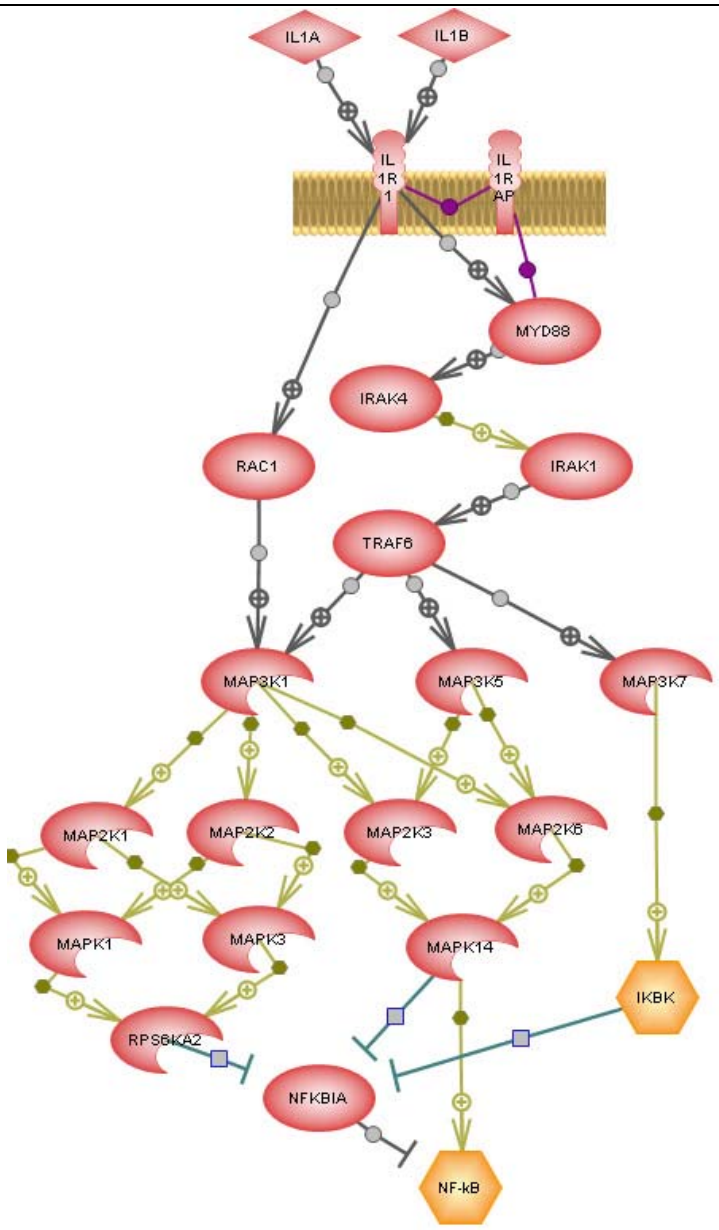

Figure 1.5 A schematic representation of part of the signalling pathway induced by IL-1α or IL-1β binding to the IL-1 receptor. This results in the activation of the transcription factor NF-κB, causing activation of pro-inflammatory genes.

As stated previously, therapeutic antibodies are ideal for specifically blocking a cytokine without severe adverse effects. Several companies have anti-IL-1β therapeutic antibodies currently undergoing clinical trials. Amongst the most promising candidates are AMG 108 (Amgen Inc.), a monoclonal antibody that inhibits IL-1β and is currently

undergoing phase two clinical trials for osteoarthritis, and Xoma 052 (XOMA Ltd.) which is an anti-IL-1β IgG2 undergoing phase two clinical trials for type II diabetes.

In addition to its clear therapeutic importance, IL-1β is a relatively robust small protein that is easily expressed and purified in large amounts in *E. coli*. The crystal and solution structures for free IL-1β have also been determined (Driscoll *et al.*, 1990; Priestle *et al.*, 1989), which makes it an ideal candidate as a model system for the determination of the structure of a therapeutic antibody-target protein complex.

1.4 Mapping antibody-antigen binding sites

An epitope is defined as the part of an antigen that is involved in its recognition by an antibody. Epitopes are categorised into two major classes. Discontinuous epitopes (also known as conformational or assembled epitopes) consist of amino acids far apart in the protein primary sequence but are brought into close proximity by protein folding. Continuous epitopes (also known as linear or sequential epitopes) contain at least four to six adjacent amino acids of the primary sequence and can be identified on denatured as well as native proteins. Given the nature of protein structures most epitopes on native proteins are likely to be discontinuous (Barlow *et al.*, 1986) and, as a consequence of this, most antibodies in polyclonal antisera raised against native proteins do not recognise short peptides (Van Regenmortel, 1989). Information about the epitope on an antigen is useful in a wide range of disciplines in which antibodies are used as molecular reagents. For example, epitope mapping for an antibody that blocks the interaction of a protein with its receptor can provide valuable information about the positioning and characteristics of the protein-receptor interaction, and mapping of all the epitopes on a viral antigen could be useful to someone wishing to produce antiviral

vaccines (Morris, 1996). The most widely used techniques for mapping antibody-antigen binding sites include mutagenesis, peptide scanning and mass spectrometry (Morris, 1996).

Natural variants of the antigen of interest, normally from different species or different isoforms of the same antigen, can in some cases provide valuable information on the location of an epitope. Since natural variants usually retain their function, the amino acid changes are unlikely to affect antibody binding by causing global changes in protein conformation. Antibody binding assays can therefore reveal information about the mutations that disrupt binding and thus positional information about the binding site can be determined. An extension of this procedure is the use of molecular biology to engineer mutants. Deletion mutants creating fragments from which antibody binding assays can determine the region associated with antibody binding can be useful, particularly in the case of multidomain proteins that can be broken down into smaller domains (Zhang *et al.*, 2008), but the benefits of this technique are limited. Well designed site-directed mutagenesis experiments can be used to map an epitope with a reasonable degree of accuracy. It is important to ensure that the mutant proteins do not lose binding activity due to a destabilising mutant producing a protein with a significantly different tertiary structure. Methods such as circular dichroism (Greenfield & Fasman, 1969) or fluorescence emission spectroscopy (Royer, 2006) can be used to determine whether the protein is folded or not. Work on mapping the antibody binding site on HPr provides an elegant example of how site-directed mutagenesis can be successfully used. The initial mutagenesis work (Sharma *et al.*, 1991) was later confirmed as correlating well with the binding site revealed by the antibody-HPr crystal structure, with 9 of the 14 residues at the binding site being identified by mutagenesis

(Prasad *et al.*, 1993). However, in general mutagenesis data must be interpreted with caution when no supplementary epitope mapping data is available.

In addition to mutagenesis studies one of the early techniques developed to identify the recognition sequence bound by an antibody was peptide scanning. This involves the screening of a collection of peptide fragments to determine which fragments bind to the antibody and thus localise the epitope to a specific region on the antigen. Initially this procedure used antigen-derived fragments (Sela, 1969) but more commonly now random peptide libraries are displayed on the surface of filamentous bacteriophage (Cwirla *et al.*, 1990; Devlin *et al.*, 1990; Scott & Smith, 1990). Large numbers of unique peptide sequences ($>10^8$) can be generated and screened simultaneously by biopanning. Each round of competitive selection of the displayed peptides is followed by amplification of the binders. This results in the isolation of one of more subsets of related peptides with each subset potentially defining a consensus sequence. The degree of homology between the consensus sequence and continuous stretches of ligand enables localisation of the epitope, determination of the amino acids most important for binding and identification of the allowable substitutions. A collection of sequences that lack consensus may indicate that the antibody recognises a discontinuous epitope (Stephen & Lane, 1992). The most clear-cut results are obtained from this technique when each residue in the epitope makes a major contribution to antibody binding and is therefore specified in the isolated peptide sequences. There are, however, several limitations to this technique: (i) the localisation of residues that make a minor contribution to binding requires careful optimisation to select for lower affinity ligands and is not always possible; (ii) the minimal epitope recognised by many antibodies is longer than the peptides represented in the library; (iii) discontinuous epitopes may not

be detectable using peptide probes; (iv) sequences bound by an antibody but unrelated in sequence to the original immunising antigen can sometimes be isolated (Lane & Stephen, 1993).

The development of time-of-flight mass spectrometry applicable to intact proteins in the early 1990s allowed the development of more direct approaches for epitope mapping based on the use of mass spectrometry to obtain accurate measurements of molecular weight for intact proteins or peptide fragments. Studies have shown that antibodies exhibit resistance towards proteolytic cleavage, that antigens bound to an antibody show a decreased rate of proteolysis with the greatest effect in the region in contact with the antibody, and that proteolysis does not lead to the dissociation of an antibody-antigen complex (Eisenberg *et al.*, 1982; Moelling *et al.*, 1980; Schwyzer *et al.*, 1980). Limited proteolytic cleavage of antibody-target protein complexes and characterisation of the epitopes was originally done by means of HPLC (Jemmerson & Paterson, 1986) and PAGE (Sheshberadaran & Payne, 1988) but these did not always allow for unambiguous peptide identification due to lack of resolution and thus the protected antigen fragments are now dissociated from the antibody and analysed with a very high degree of accuracy by mass spectrometry. An additional approach is the chemical modification and comparison of free and antibody bound antigen, yielding identification of residues which are protected from modification by the antibody and are hence considered to be part of the epitope (Burnens *et al.*, 1987). In principle amino (Lys), thiol (Cys), carboxyl (Asp, Glu), imidazole (His), guanidine (Arg) and hydroxyl (Ser, Thr, Tyr) groups can be modified (Morris, 1996) but in practice modification of Lys, Arg and Tyr residues is most common (Van de Water *et al.*, 1997). While not providing

a true representation of the epitope this approach can provide additional supporting information on its location.

Perhaps the most commonly used analytical approach for mapping epitopes is hydrogen-deuterium exchange mass spectrometry (Ehring, 1999). The exchange rate of amide hydrogens in protein-protein complexes is highly dependent on whether the amide group is solvent exposed. Those backbone amide groups that are at the protein-protein interface are protected from exchange and remain protonated. These regions can then be analysed by peptic digest and mass spectrometry and compared with the fragments from the free antigen to identify residues at the interface. During the peptic digest and mass spectrometry analysis back exchange can occur, particularly for very small peptides, leading to the loss of information (Lu *et al.*, 2005).

All of the methods discussed can provide useful information about the position and characteristics of an epitope, especially when used in combination, but each has drawbacks preventing the determination of a true spatial representation of the epitope. High resolution structural techniques, such as X-ray crystallography and NMR, are the only ways of definitively determining the epitope in an antibody-target protein complex.

1.5 Antibody-antigen structures

Crystal structures of free antibodies and their complexes have revealed a large amount about the basic structure of antibodies and the characteristics of antibody-antigen interfaces. The landmark publications of the first two immunoglobulin protein structures, a light chain dimer (Schiffer *et al.*, 1973) and an IgG1 Fab (Poljak *et al.*, 1973), revealed the now well recognised immunoglobulin fold supporting three CDRs

from each variable domain. Subsequent studies of antibody complexes demonstrated that the antibody binding site is almost exclusively located within the CDRs (Edmundson *et al.*, 1974; Segal *et al.*, 1974). To date the vast majority of antibody structures, which totals over 400, are for free or peptide bound antibody fragments, with Fabs being the most common. There are less than 50 unique Fab-protein structures, only three scFv-protein structures and currently no free scFv crystal structures, with the majority of target proteins coming from a small selection of proteins dominated by lysozyme and neuraminidase.

The crystal structures of antibody fragment-target protein complexes reveal several key features: (i) both the light and heavy chains of antibodies make extensive contacts with the target protein, with the heavy chain frequently making the greater proportion of contacts; (ii) the binding specificity is determined mostly but not always completely by the CDRs; (iii) the epitope on the antigen is discontinuous; (iv) van der Waals interactions, hydrogen bonds and to a lesser extent salt bridges mediate the binding of antibody to antigen (Braden & Poljak, 1995). Extensive studies of antibody-target protein interfaces have revealed that antibody binding sites consist of between 12 and 20 interacting residues from each partner in the complex giving a total buried surface area of approximately 1700 ±260 $Å^2$ (Lo Conte *et al.*, 1999). The interface on an antibody is on average more polar than other protein-protein interfaces, with Tyr contributing most to the interface (Jones & Thornton, 1996). In the complexes determined to date, CDR L2 only rarely contributes to binding (Wilson & Stanfield, 1993), while CDR H3 often makes the largest contribution to binding (MacCallum *et al.*, 1996). Perhaps surprising given antibodies abilities to bind a limitless number of targets, the CDRs appear to adopt only a limited number of canonical backbone

conformations determined by their loop length and certain key structurally determining residues (Chothia *et al.*, 1989).

There are only a small number of structures that have been determined for antibodies in their free and complexed forms but these have resulted in the reporting of a variety of conformational changes in both the antibody and antigen (Bhat *et al.*, 1990; Herron *et al.*, 1991; Rini *et al.*, 1992; Stanfield *et al.*, 1990). The changes that occur in the antibody can be divided into three categories. First, there are minor adjustments of side chains at the binding interface, which is a common feature for all protein-protein interactions. Second, there are rearrangements of individual CDRs. Finally there are changes resulting from the rigid body movement of the V_H and V_L domains. The latter two rearrangements occur to a varying degree and in some cases no major conformational rearrangements have been observed (Bentley *et al.*, 1990; Rini *et al.*, 1992), suggesting that some antibodies are more flexible or 'inducible' than others (Stanfield *et al.*, 1993).

The most complete structural analysis of a specific antibody has been carried out on the mouse anti-hen egg white lysozyme (HEL) antibody D1.3. The crystal structure of the Fab D1.3-HEL complex (Amit *et al.*, 1986) has been followed by high resolution structures of the free Fv D1.3 fragment and Fv D1.3-HEL complex (Bhat *et al.*, 1990). The structure of the Fv D1.3-HEL complex closely superimposes with the variable domains from the Fab D1.3-HEL complex, giving an RMSD difference in Cα positions of 0.39 Å. In particular there is a very close correspondence between the position and orientation of the HEL contacting residues. Similarly the structure of HEL is conserved in both complexes. Superimposition of the Cα coordinates of the Fv D1.3 V_L domain

with the V_L domain from the Fv D1.3-HEL complex gives a deviation of 0.37 Å, which is comparable to the estimated error in atomic coordinates of 0.33 Å. A similar observation is made upon superimposition of the V_H domains, indicating that antigen binding does not induce major conformational changes in the individual variable domains. However, superimposition of the V_L domains for the free and HEL bound Fv D1.3 causes a subtle displacement of the V_H domains (RMSD 0.99 Å) and the same is equally true upon overlaying the V_H domains, as shown in figure 1.6. This results in the contact residues from the antibody being brought closer to the antigen, causing an increased separation of the variable domains from each other at their C termini. Significant conformational changes in CDRs were first observed for the Fab BV04-01-single stranded DNA complex (Herron *et al.*, 1991). Large changes at the antibody binding site were seen due to the significant motion of CDRs L1, L2 and H3. A similar sized subtle reorientation of the two variable domains upon binding was also reported.

Comparison of the few antibody structures that are available in both the free and complexed state shows no evidence that any signal is being transmitted to the constant domains by this variable domain rearrangement (Stanfield *et al.*, 1993). It has been suggested that the plasticity of the antibody binding site that allows induced fit provides a means of generating diversity of antibodies in addition to those that involve changing the amino acid sequence (Davies & Padlan, 1992). The idea of an induced fit may represent the 'relaxing' of the antibody binding site in the absence of antigen (MacCallum *et al.*, 1996) and the selection or stabilisation of different pre-existing conformational states in antigen and/or antibody conformations (Braden & Poljak, 1995). Little is known about the dynamic nature of the CDRs and whether they have a preformed structure prior to binding (lock and key theory), or are fully flexible and form

a defined structure upon association with the antigen (induced fit theory) or perhaps more likely a compromise between the two extremes.

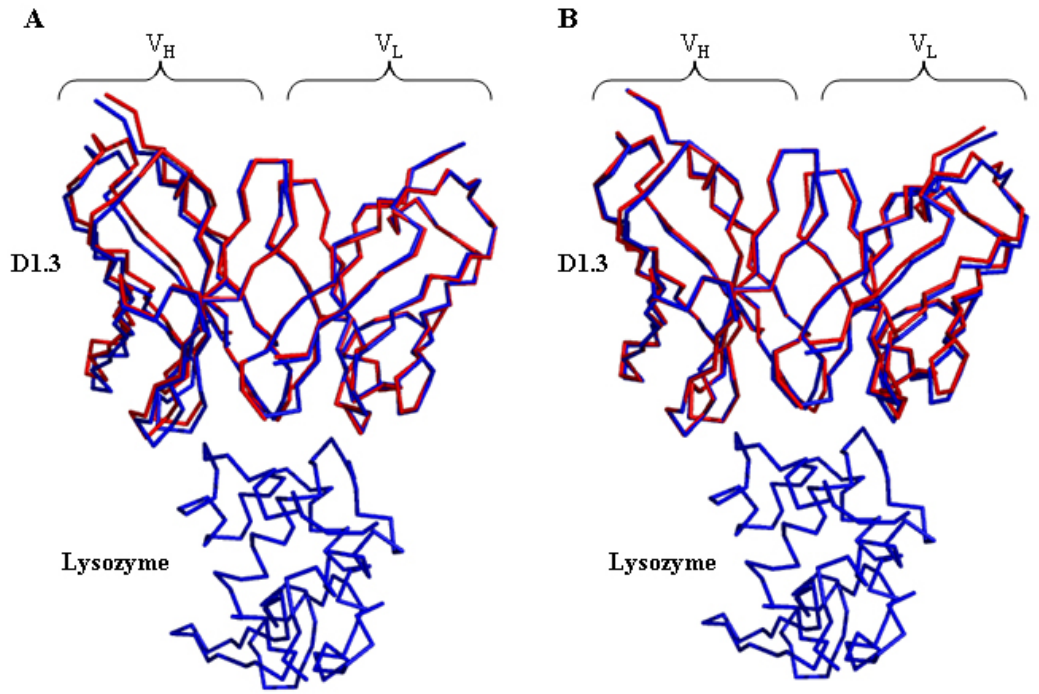

Figure 1.6 Backbone line representations of the crystal structures for the free Fv D1.3 antibody fragment (red) and the D1.3-HEL complex (blue). The V_L domains of the free and bound antibody fragments are superimposed in panel A, with panel B showing the superimposition of the V_H domains. The two domains clearly overlay very well, representing no change in structure of the individual domain but their orientations with respect to each other are altered slightly upon formation of the complex with the C terminal domains moving apart.

To obtain a better understanding of antibody binding further structural characterisation of a wider variety of antibody-target protein complexes is required. Unfortunately X-ray crystallography is restricted by the necessity of obtaining good crystals. It is clear from

the limited number of crystal structures over the last 30 years that the crystallisation of antibody fragment-target protein complexes is problematic, which probably reflects the inherent flexibility and solubility of antibodies. There is also the added complication that the proteins are in a non-natural crystal state and differences observed between structures could potentially be due to distortions from crystal contacts.

1.6 Nuclear magnetic resonance spectroscopy

NMR is a high resolution structural technique that can determine the structure of proteins and protein-protein complexes, as well as mapping their interactions. Isotopes of certain nuclei, such as ^{1}H, ^{13}C and ^{15}N, absorb radiofrequency radiation of an appropriate frequency when placed in a constant magnetic field. So-called NMR active nuclei of the same element in different chemical environments absorb at slightly different frequencies and it is this that provides the possibility to assign individual NMR signals to specific atoms within a molecule. Through bond and through space mediated NMR interactions can then be used to obtain structural restraints for proteins and complexes, which allow determination of a three-dimensional structure. NMR has several advantages over crystal diffraction methods: (i) NMR is a solution state method and so proteins are being analysed in a more natural state than with crystallography; (ii) the effects of protein dynamics, ligand binding and environmental effects such as pH, salt and temperature can be measured relatively easily (Campbell & Downing, 1998); (iii) once the assignments have been made for an antigen multiple antibody binding sites can be swiftly and accurately mapped. The most significant disadvantage often cited against NMR is a size limit of about 30 kDa but continuing advances over the last 10 years have increased this threshold to over 80 kDa (Kanelis *et al.*, 2001; Pervushin, 2000).

A particularly important advance for NMR was the ability to uniformly label proteins with ^{15}N and ^{13}C by over-expression in *E. coli* incubated in minimal media containing sources of ^{15}N and ^{13}C. This enables improved spectral resolution by recording three-dimensional spectra that separate the individual proton resonances according to the chemical shifts of the directly bonded heteronuclei and simultaneously yields information about the system, i.e. ^{15}N and ^{13}C chemical shifts, as shown in figure 1.7. The increased dimensionality in these experiments leads to a reduction in resonance degeneracy. Therefore, a large number of resonance assignments can be made unambiguously, which leads to more structural restraints and hence increased accuracy and precision of NMR derived structures. A set of triple resonance experiments are used to assign the backbone chemical shifts by correlating the ^{1}HN and ^{15}N chemical shifts of a reference amino acid with the $^{13}C\alpha$, $^{13}C\beta$ and/or $^{13}C'$ of the same or preceeding residue. Information about the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts is especially valuable as they are characteristic of the different types of amino acids and can therefore help to position sequentially connected amino acids within the known primary sequence of a protein (Grzesiek & Bax, 1993).

Once the backbone assignments have been established using through bond connectivity, the traditional approach to deriving NMR structures *de novo* is to then assign the sidechain chemical shifts by correlating the ^{1}H and attached ^{13}C chemical shifts. In addition, distance restraints are obtained by the assignment of inter and intra-residue NOEs (nuclear Overhauser effect). These are through space rather than through bond correlations over a limited distance (approximately 5 Å in protonated samples and 8-9 Å in deuterated samples) and so provide information about the spatial orientation of pairs of atoms that are close in space but not necessarily in the primary protein sequence. All

of these restraints together enable the generation of an ensemble of structures that satisfy the data.

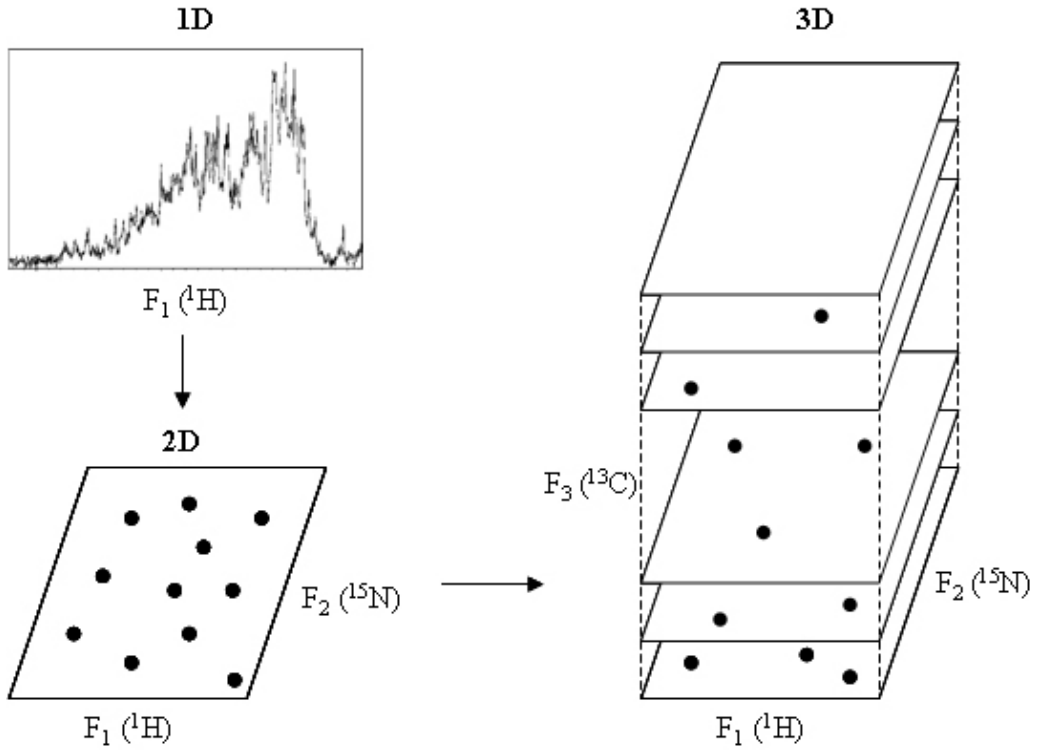

Figure 1.7 A schematic illustration of the progression from 1D NMR to 2D and 3D heteronuclear NMR spectra. The 1D spectrum shows the amide proton section of a representative spectrum for the scFv IC8, showing poor spectral resolution due to the large number of signals. The 2D and 3D spectra show the improved resolution that can be obtained by separating individual proton resonances (represented by black dots) according to the chemical shifts of the directly bonded heteronuclei.

For protonated proteins larger than about 35-40 kDa the excited state relaxes too rapidly for the necessary experimental data to be collected. This is due to larger proteins having longer tumbling times resulting in a faster rate of loss of magnetisation and leading to the generation of broader signals in the spectrum causing a loss of resolution.

Deuteration of proteins, removing all but the backbone amide protons, gives a significant reduction in the relaxation, resulting in a narrowing of linewidths and better resolution. It is now possible to routinely express proteins in *E. coli* in progressively deuterated minimal media (Venters *et al.*, 1995). In addition to deuteration, the introduction of the TROSY sequence (transverse relaxation-optimised spectroscopy) (Pervushin *et al.*, 1997), which selects for the slowly relaxing component, has greatly improved the collection of NMR data for larger protein systems.

Structure determination by NMR has traditionally relied upon the measurement of a large number of local restraints, as outlined previously. A more recent method of obtaining structural information is through the direct magnetic dipole-dipole coupling between spin-½ nuclei (e.g. ^{1}H, ^{13}C, ^{15}N). Depending on the orientation of any two magnetically active nuclei with respect to each other the magnetic field of one nuclei will either reinforce or counteract the effect of the external magnetic field at the other nuclei. The dipolar coupling, D*ij*, between a pair of nuclei is shown in the equation below (Prestegard *et al.*, 2004), where r is the distance between a specific pair of nuclei, γij are the magnetogyric ratios for the nuclei, μ_0 is the permittivity of free space, *h* is Planck's constant and the angle θ relates the inter-nuclear vector between the coupled spins to the external magnetic field

$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h}{(2\pi r)^3}\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle$$

The measured dipolar coupling is an average of all the values of θ, so in isotropic solution rotational Brownian diffusion averages the interaction to zero. In solid state NMR spectroscopy this averaging does not take place and so all dipolar couplings are

seen up to their maximum value and thus the resolution achievable by solution state NMR spectroscopy is lost. It is possible to use solution state NMR with the introduction of a very weak degree of alignment to measure the residuals of dipolar couplings, hence the name residual dipolar couplings (RDCs) (Tolman, 2001). In this case the possible orientations of a protein in the anisotropic solution are not all equally likely, thus the alignment of the protein can be described by an alignment tensor. RDCs can be either positive or negative and are measured as a change in the scalar coupling value. One such method enables calculation of RDC values by comparison of the J couplings measured from HSQC and TROSY experiments fr isotropic and anisotropic samples (Kontaxis *et al.*, 2000).

When the value of the RDC is known the value of θ can be calculated and so the orientation of the bond is known relative to the magnetic field, as shown in figure 1.8. The alignment tensor defines the average orientation of a weakly aligned molecule with respect to the magnetic field, which is defined by eigenvalues (Axx, Ayy, Azz). The alignment tensor can be described by an axial (Da) and rhombic (Dr) component where Da = Azz and Dr = 2/3 (Axx-Ayy). The rhombicity (R) can be derived from Da and Dr. Software packages, such as PALES (Zweckstetter & Bax, 2000) and MODULE (Dosset *et al.*, 2001), can use a known structure or model and set of RDCs to calculate the alignment tensor by singular value decomposition. There is an inherent four-fold degeneracy in the value of the alignment tensor due to the $\cos^2\theta$ dependency of the equation, meaning that an inversion of θ with respect to any of the axes will give the same RDC value. This degeneracy can be reduced by collecting RDC values from an additional alignment medium that orientates the protein in a different way.

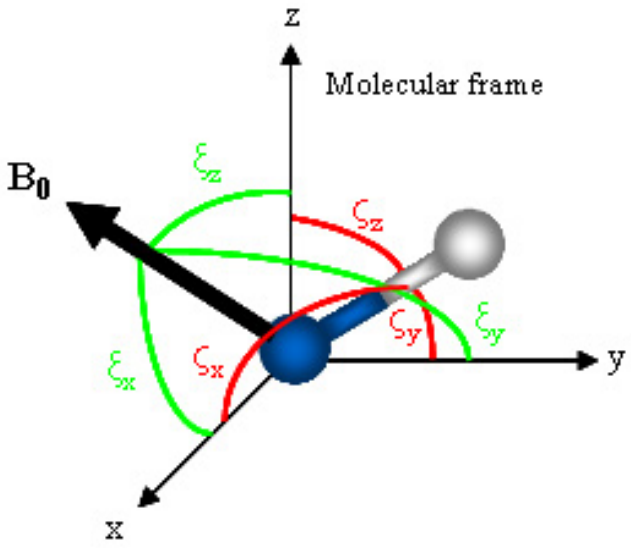

Figure 1.8 Orientation of the magnetic field (B_0) field and the internuclear vector in the molecular frame. The angles show the molecular frame relative to the magnetic field (ξ_x,ξ_y,ξ_z) and the internuclear vector relative to the molecular frame (ς_x, ς_y, ς_z).

A large number of different alignment media have been reported in the literature (Koenig *et al.*, 1999; Prosser *et al.*, 1998; Ruckert & Otting, 2000; Sass *et al.*, 1999; Tycko *et al.*, 2000), of these two of the most popular are filamentous bacteriophage (Hansen *et al.*, 2000) and phospholipid bicelles (Ottiger & Bax, 1998). Bicelles are bilayers consisting of long and short chain phospolipids with the longer hydrocarbon tails making up the edges of the disk and the shorter hydrocarbon tails forming the flat surfaces. At specific temperatures bicelles form lipid crystals and will align with their long edge parallel to a magnetic field, as shown in figure 1.9. Although very useful, the bicelle medium has disadvantages. The phospholipids hydrolyse rapidly if the pH is not carefully monitored, with lipid composition, ionic strength and temperature also being critical to bicelle stability (Bax, 2003). In contrast, the filamentous bacteriophage, Pf1, is remarkably robust, aligning at a wide range of temperatures and concentrations due to the high ratio of length to diameter (Zweckstetter & Bax, 2001), as illustrated in figure

1.9. Pf1 carries a substantial negative charge, causing the interaction between phage and the protein under investigation to be both electrostatic and steric, in contrast to a bicelle system in which all interactions are steric in nature. This makes these two types of alignment media complementary, with the different alignment mechanisms giving different alignment tensors, allowing the removal of RDC degeneracy.

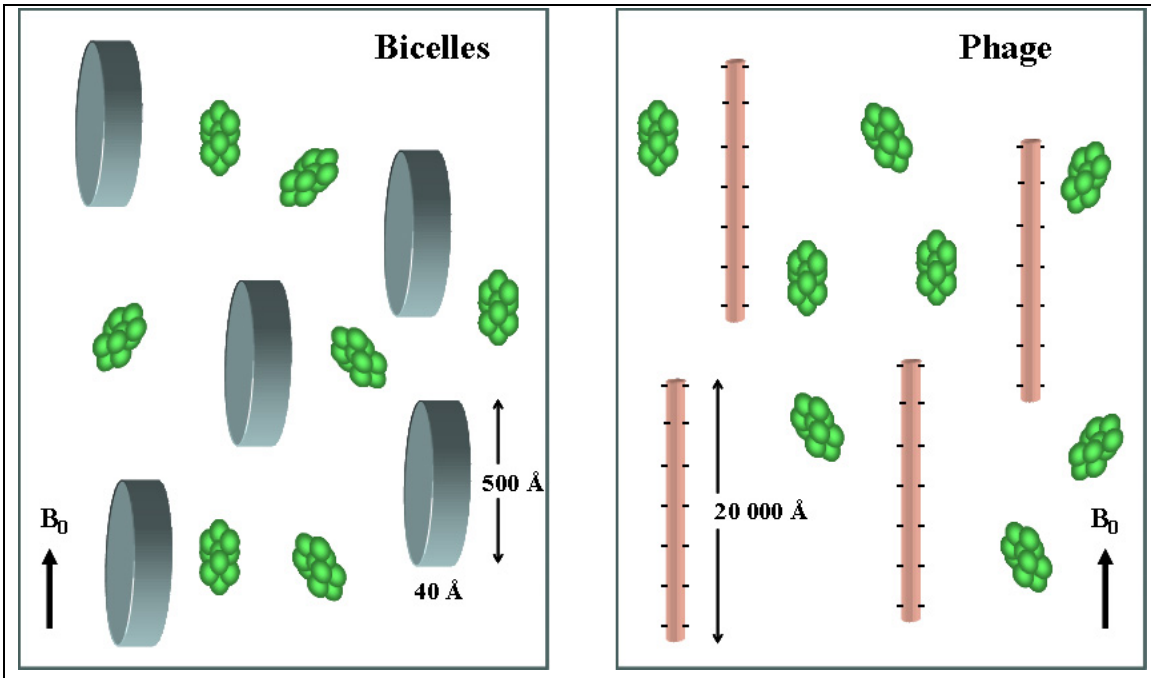

Figure 1.9 A cartoon representation of the partial alignment of a protein sample by phospholipid bicelles and bacteriophage when placed in an external magnetic field (B_0).

Unlike NOEs, RDCs provide orientational information between the coupled nuclei with respect to a global reference frame. This long range orientational information is extremely valuable and compliments the information that can be obtained from NOEs. RDCs are commonly used in sevaral ways. Firstly, they can be used in structure validation, where back-calculated RDCs for the structure can be derived by a software package such as PALES and compared to the measured RDC values. Accurately determined structures will show good correlation, with a quality score, Q (Cornilescu *et*

al., 1998), often used to define the agreement. Secondly, they can be combined with other restraints to refine structures. Finally, and of most relevance to the work reported in this thesis, RDCs can be used to orientate subunits of a multidomain protein or partners in a complex when the structure of the individual components are known (Prestegard *et al.*, 2004).

In addition to the traditional use of RDCs to calculate a global alignment tensor, intervector projection angle restraints can be generated by taking pairs of RDCs and calculating intervector restraints, which are somewhat equivalent to dihedral angle restraints. These restraints have the advantage that they no longer depend on the orientation of the dipole vector with respect to the alignment tensor and instead they restrain the angle between two dipolar vectors. This avoids the burden of working with an alignment tensor in structural calculations and also allows for inter and intra-molecular restraints to be distinguished. This method has been reported to reduce convergence problems that can be associated with the use of RDCs in a simulated annealing process and has also shown promising results when used in combination with an alignment tensor in docking algorithms (Meiler *et al.*, 2000).

In recent years a number of NMR-based structure determination approaches have been proposed that avoid the time consuming assignment of all sidechains and intraresidue NOEs, such as the determination of protein structures from only chemical shifts (Cavalli *et al.*, 2007) or RDCs (Hus *et al.*, 2001) and the use of selective methyl protonation in highly deuterated proteins (Gardner *et al.*, 1997). Another popular approach with complexes for which the individual protein structures are known or can be modelled with reasonable accuracy is NMR restraint-driven docking. This will be discussed more

in chapter 3 but in essence a limited number of NMR restraints (backbone assignments, NH to NH NOEs and RDCs) are used to refine the structures or models and dock the individual proteins. The backbone assignments and NOEs provide dihedral angle and distance restraints respectively, while the RDCs give orientational information.

Until the work reported in this thesis no NMR-derived structures had been determined for free or complexed IgG antibody fragments. The only literature of note with regard to the determination of antibody structures by NMR is for a 15 kDa llama VHH fragment (Renisio *et al.*, 2002) and the Fv and scFv fragments of the anti-phosphorylcholine antibody MCPC603 (Freund *et al.*, 1993; Freund *et al.*, 1994). The structure of the VHH fragment was determined in the classical manner and analysis of its dynamics revealed that the β-sheet core of the antibody was largely rigid with the three CDRs showing a greater mobility. In a fairly limited study, Freund and co-workers initially compared spectra for Fv and scFv MCPC603 fragments and concluded that the only significant differences were due to the presence of the $(Gly_4Ser)_3$ linker residues in the scFv. The similarity suggests that there are no significant structural changes between the scFv and the unlinked Fv fragment. ^{15}N glycine and ^{15}N glycine/serine labelled samples were then used to assign the three serine residues in the linker but assignment of the glycines was not possible due to overlap. Relaxation experiments also showed an enhanced flexibility for the linker, which along with the narrow chemical shift dispersion of the linker residues suggests that the linker region is unstructured and constrained only by the distance between the two domains. A continuation of the specific amino acid labelling strategy enabled sequential assignment of 140 of the 250 residues, however broad linewidths prevented the use of ^{15}N-^{13}C spectra. Despite the lack of assignment for some residues due to overlap, three of the CDRs were found to be well defined. The

authors proposed that the lack of data for the remaining CDRs could potentially be due to an intrinsic dynamic property of these CDRs, although this conclusion is based on assignment for only 56% of the residues.

1.7 Aims

Information about the interaction of an antibody with its target protein is essential for the characterisation, humanisation and development of therapeutic antibodies. High resolution structural data is therefore highly desirable and thus the broad aim of this project was to determine the high resolution structure of a scFv-IL-1β complex by NMR. This can be further broken down into two sections:

1) Characterisation of the behaviour of free scFv fragments and complexes and assessment of their suitability for NMR studies.

ScFvs are the smallest fully functional fragment of an antibody consisting of the V_L and V_H domains connected by a flexible linker. Their small size (28 kDa) allows them to be expressed in *E. coli* and potentially makes them amenable to high resolution structural studies by NMR. Previous studies have reported the multimerisation of scFvs at concentrations above approximately 200 μM but due to the domain swapped nature of this dimerisation the interactions in the monomer and dimer are essentially the same (Lee *et al.*, 2002). The original aim of the work reported in the first experimental chapter of this thesis was to generate a series of scFv expression vectors from pre-existing anti-IL-1β Fab vectors, optimise the expression of reasonable amounts of soluble scFv if possible and characterise the domain-swapped dimer equilibrium. In particular, the concentration dependence of the equilibrium would be investigated as

this has serious implications on the quality of the NMR data that could be obtained at the relatively high concentrations required (>200 μM). Previous NMR studies of free scFvs have shown significantly broader linewidths than would be expected for a protein of the size (Freund *et al.*, 1993; Freund *et al.*, 1994), suggesting the concentration of the domain-swapped dimer is significant, thus increasing the effective molecular weight of the system. Analysis of the quality of the NMR data obtained from all of the free scFv samples and the larger scFv-IL-1β complexes would then enable determination of the viability of a NMR project on free or complexed scFvs.

2) Determination of the high resolution structure of a scFv-target protein complex.

High resolution structural data about the interactions of a potential therapeutic antibody with its target protein, in this case IL-1β, would provide valuable information that could influence the design and humanisation of both the antibody in question and any future therapeutic antibodies. The use of solution state NMR may also provide insights into the mechanisms of antibody binding, in particular the state of the CDRs and effects of antigen binding on the structure of the antibody, that have not been completely answered by crystal structures. Comparison of data collected for the scFv with equivalent Fab fragment data would also enable confirmation that scFvs bind in a similar manner to an equivalent Fab fragment and so can be used as reliable models. Based on the findings of the first experimental chapter, the aim of the work reported in the second experimental chapter of this thesis was to determine the structure of the scFv IC8-IL-1β complex by restraint driven docking. In order to obtain high quality NMR data all protein samples would be deuterated enabling triple resonance experiments to be used to obtain comprehensive backbone assignments (HN, N, Cα, Cβ, C').

Assignment of the backbone amide signals for the scFv IC8-IL-1β complex would allow the residues with significantly perturbed signals upon complex formation to be determined, thus mapping the binding sites on both the scFv IC8 and IL-1β. This information along with backbone amide RDCs and long range NOEs, providing orientational and distance restraints respectively, would enable reliable refinement and docking of a homology model of the scFv IC8 with the crystal structure of free IL-1β.

In parallel to the work detailed in this thesis a second PhD studentship, undertaken by Catherine Hall, aimed to characterise the free and IL-1β bound structures of a Fab fragment containing the same variable domains as the IC8 scFv reported here. This would allow for direct comparison of equivalent scFv and Fab binding to the same antigen.

The structural data will significantly enhance the understanding of the interaction between the scFv IC8 and its target protein IL-1β and will provide the first NMR derived structure for a scFv-target protein complex.

Chapter 2

Construction and Characterisation of Anti-IL-1β ScFvs

2.1 Introduction

2.1.1 The scFv construct

The Fv portion of an antibody, consisting of the V_H and V_L domains, is the smallest fragment that maintains the full binding activity of the intact antibody. However, the Fv fragment tends to have limited stability due to the dissociation of the two domains. This dissociation equilibrium is dependent on the buried surface area and the quality of the interaction at the V_H-V_L interface, which varies from antibody to antibody. V_L dimers are known to occur (Freund *et al.*, 1994; Stevens *et al.*, 1991), thus V_H-V_L interfaces could rearrange to form V_L-V_L homodimers leaving V_H domains as single species which have been shown to have a tendency to aggregate (Ward *et al.*, 1989).

Introduction of a peptide linker between the V_L and V_H domains has been described to create the so-called single chain Fv (scFv) (Bird *et al.*, 1988; Huston *et al.*, 1988). A variety of linkers of different length and sequence have been used (Hennecke *et al.*, 1998; Takkinen *et al.*, 1991; Turner *et al.*, 1997), but the most popular have been linkers of the sequence $(Gly_4Ser)_n$ following the work by Huston and co-workers using a 15 amino acid Gly_4Ser repeat linker (Huston *et al.*, 1988). Linkers of this type offer flexibility and a reduced immunogenicity due to the lack of a well defined tertiary structure, as well as being stable against proteolysis. ScFv fragments can be made in either a V_L-linker-V_H or V_H-linker-V_L orientation, with the distance the linker has to

span being larger in the latter case (Huston *et al.*, 1991), thus a linker of the same length is more strained in this orientation.

Since the first description of Fv and Fab antibody fragment expression in *E. coli* (Better *et al.*, 1988; Skerra & Pluckthun, 1988) this has become the most popular expression system for most antibody fragments and in particular scFvs due to the reduced costs, ease of production and higher expression levels than most other systems. Normally antibodies do not form disulphide bonds in the reducing environment of the cytoplasm (Biocca *et al.*, 1995) and are therefore usually unable to form a stable native fold, resulting in the observation of reduced functional expression levels of cytoplasmically expressed antibody fragments (Cattaneo & Biocca, 1999; Mhashilkar *et al.*, 1995). Cytoplasmic expression and refolding of the protein from the resulting inclusion bodies is possible but low yields are most commonly achieved. Therefore, scFvs are most commonly targeted to the periplasm, where the less reducing environment enables the formation of the intradomain disulphide bond in each of the variable domains, producing correctly folded protein.

2.1.2 Therapeutic uses of scFvs

The use of scFv fragments is advantageous in some circumstances due to the lack of constant domains, which can attribute to increased immunogenicity and, when the Fc domain is present, the activation of the complement pathway. Their small size gives them improved pharmokinetics for tissue penetration, as shown with the CC49 scFv which penetrates tumours in 30 minutes compared to 48 hours and longer measured for the intact IgG (Yokota *et al.*, 1992). However, proteins of less than approximately 50 kDa are quickly removed from the blood stream via the glomerular filtration unit of the

kidney. For this reason scFvs are rarely used as therapeutics unless attached to another protein or compound such as PEG to increase their molecular size as frequent injections of the therapeutic would be required. Nonetheless this makes them ideal candidates for imaging with several scFv fragments currently in clinical trials (Holliger & Hudson, 2005).

2.1.3 ScFv multimerisation

The formation of scFv dimers by domain swapping (Bennett *et al.*, 1994; Bennett *et al.*, 1995) has been noted by many groups and is known to be particularly dependent upon the linker length (Alfthan *et al.*, 1995; Desplancq *et al.*, 1994; Essig *et al.*, 1993; Holliger *et al.*, 1993; Whitlow *et al.*, 1993) and antibody type (Griffiths *et al.*, 1993; Raag & Whitlow, 1995; Wu *et al.*, 1996). Shortening of the linker, originally optimised for monomer formation as a 15 residue $(Gly_4Ser)_3$ linker by Huston and co-workers (Huston *et al.*, 1988), to 5-10 residues has been shown to create a steric hindrance preventing monomeric assembly and giving rise to so-called diabodies (Holliger *et al.*, 1993; Perisic *et al.*, 1994). Reducing the linker further or removing it altogether can result in trimeric scFvs, known as triabodies (Iliades *et al.*, 1997; Kortt *et al.*, 1997; Pei *et al.*, 1997). The crystal structure of the L5MK16 diabody (Perisic *et al.*, 1994) shows that the dimer is formed by the pairing of the V_H domain from one scFv with the V_L domain from the other and vice versa, resulting in a functional dimer. The triabody has three Fv heads with the polypeptides arranged in a cyclic head-to-tail fashion.

Models explaining the domain-swapped dimer formation (figure 2.1) show that the V_H-V_L interface within the free monomeric scFv can dissociate forming an open monomer (Hudson & Kortt, 1999; Worn & Pluckthun, 1999). Domain-swapped association of two

open monomers results in the formation of a domain-swapped dimer. Binding of antigen may prevent the disassociation of the V_H-V_L interface by locking them together and thus stabilising the closed monomer and dimer.

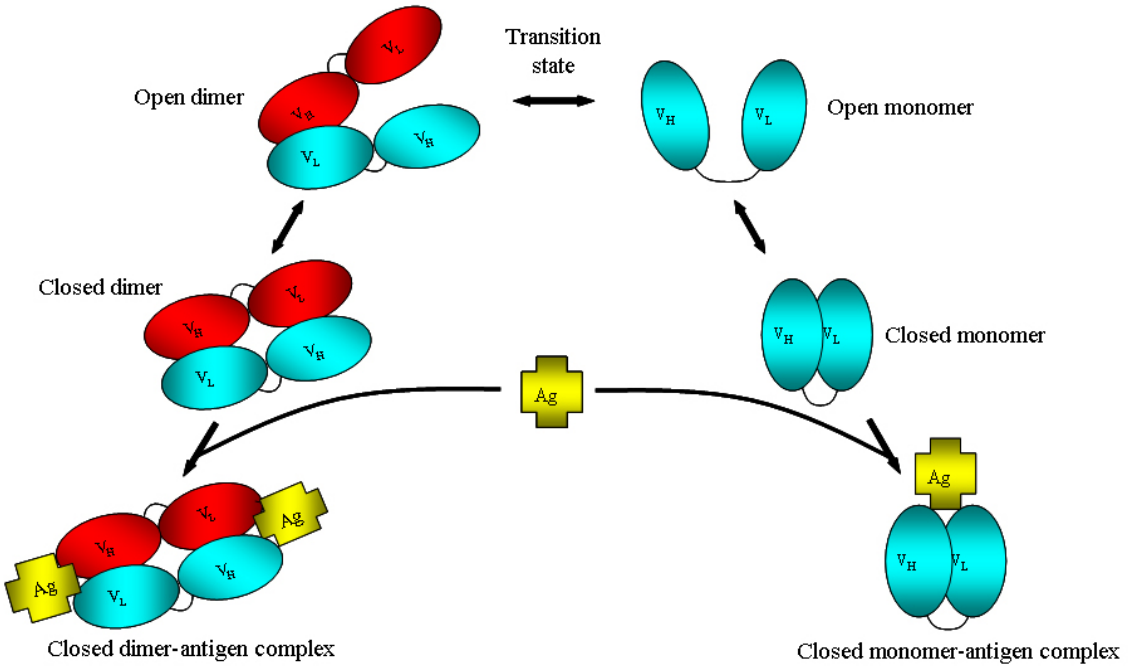

Figure 2.1 A schematic representation of the scFv monomer to domain-swapped dimer equilibrium. Free scFv can exchange between the closed and open forms, enabling a second open monomer (shown in red) to associate with it and form a functional domain-swapped dimer. Binding of antigen (Ag) may prevent the separation of the V_L and V_H domains resulting in the formation of stable monomeric or dimeric scFv-antigen complexes.

A detailed study by Arndt and co-workers investigated a number of factors that influenced the monomer to dimer transition of the scFv McPC603 (Arndt *et al.*, 1998). Comparison of a Fv fragment with the scFv clearly showed that dimerisation requires the presence of the linker, confirming that the dimer is created by domain-swapping and not by non-specific aggregation. The conversion rate of dimer to monomer observed by gel filtration was reported to be heavily dependent upon buffer conditions with higher

pH (measured between pH 7 and 8) and lower ionic strength (measured between 150 and 900 mM sodium chloride) resulting in faster conversion of dimer to monomer. This salt dependency is rationalised by higher salt concentrations stabilising the interface by increasing the strength of hydrophobic interactions, thus disfavouring the open form of the scFv, which is necessary for conversion to monomer. They proposed that the pH dependency of the scFv domain-swapped dimer formation is not necessarily a general feature for all scFv fragments and is more likely a peculiarity of McPC603 due to the nature of the titratable groups at the interface. However, it is likely that the salt dependence is a general feature of scFvs, as all V_H-V_L interfaces have a substantial hydrophobic content (Chothia *et al.*, 1985).

A second comprehensive study by Lee and co-workers (Lee *et al.*, 2002) characterised the stability and concentration dependence of the domain-swapped dimer formation for the anti-tumour scFv MFE-23 by neutron scattering, analytical ultracentrifugation, gel filtration and NMR spectroscopy. MFE-23 was studied in a concentration range of 1.1-8.0 mg/ml (40-300 μM), with backwards extrapolation of the neutron scattering data to zero concentration giving a molecular weight of 27.3 ±1.2 kDa, which corresponds well with the value of 27.2 kDa calculated from the sequence. This shows that MFE-23 is monomeric in dilute solution. Sedimentation equilibrium studies, which determine a molecular mass independent of the macromolecular shape, indicated that the scFv was not a single non-interacting species. At a concentration of 0.4 mg/ml the predicted molecular weight was approximately 25 kDa with this value increasing up to approximately 35 kDa at 3 mg/ml. ^{1}H NMR spectra of the scFv at concentrations between 0.4 and 5.3 mg/ml and temperatures of 25-50 ^{o}C showed thermal stability up to 45 ^{o}C and only very minor changes in two of the six highly shifted methyl groups upon

formation of the dimer, suggesting that there are only subtle structural differences between the monomer and dimer. Analytical gel filtration studies between 0.5 and 8.4 mg/ml showed an increasing proportion of dimer up to approximately 40% at a concentration of 8 mg/ml.

To date the formation of the scFv domain-swapped dimer has limited structural studies of scFvs, particularly by NMR as a mixture of multimeric species is not ideal for the acquisition of high quality data. The work reported in this chapter describes the production of scFv constructs for four antibodies selected for their high affinity to IL-1β. The expression levels of these constructs were characterised and purification of the soluble scFvs performed. Constructs for the scFvs with extended 25 residue linkers or a specific Alanine to Aspartate were also designed to prevent dimer formation. The monomer to domain-swapped dimer equilibrium for the scFvs, in particular its dependence on concentration, was characterised and the potential for high resolution NMR studies on a free scFv or scFv-IL-1β complex assessed.

2.2 Materials and methods

2.2.1 DNA analysis by agarose gel electrophoresis

PCR products and all stages of the scFv construct cloning were analysed by electrophoresis using 1% (w/v) agarose gels run at a constant current of 160 mA for approximately 1 hour. PCR products were prepared for analysis by adding 4 µl loading dye (Promega) and 12 µl deionised water to 4 µl PCR product. For gel purification of double digested PCR product, 6 µl of loading dye was added to the 53 µl double digest solution consisting of 20 µl cleaned PCR product, 5 µl buffer 3, 3 µl BsiWI, 3 µl EcoR1 (New England Biolabs) and 22 µl deionised water. Ethidium bromide was used to visualise the DNA under UV light. All gels were run with a reference lane containing either a 100 bp or 1 kbp DNA ladder (Invitrogen) as appropriate.

2.2.2 SDS-polyacrylamide gel electrophoresis

The expression and purification of protein samples was monitored by SDS-PAGE. 20 µl protein samples were combined with 10 µl 50 mM DTT and 10 µl 4x NuPAGE LDS sample buffer. Samples were heated to 70 ^{o}C for 10 minutes then loaded onto a 4-12% acrylamide gradient pre-cast NuPAGE Bis-Tris gel (Invitrogen). Electrophoresis was carried out at a constant 200 V for 40 minutes in a MES SDS running buffer (Invitrogen). Gels were run with 1 lane containing a 14-66 kDa molecular weight marker (Sigma), which was heated to 70 ^{o}C for 10 minutes prior to loading. After electrophoresis, gels were stained with Coomassie Brilliant Blue (2.5 g/l in 45% (v/v) methanol and 10% (v/v) acetic acid) and destained in a 10% (v/v) methanol and 5% (v/v) acetic acid solution.

2.2.3 Western blotting

The expression of His-tagged scFvs was analysed by western blotting using a mouse anti-His-tag antibody (Novagen) as the primary antibody and a goat anti-mouse IgG horse radish peroxidise (HRP) conjugated antibody (Pierce) as the secondary antibody. A SDS-PAGE gel containing the samples for analysis was run as described in section 2.2.1 but not stained with Coomassie. PVDF transfer membrane was soaked in methanol for between 5 and 15 seconds followed by 20 minutes soaking in transfer buffer (1.45 g Tris, 7.2 g glycine, 200 ml methanol, 800 ml deionised water) along with blotting filter paper (Pierce) and blotting pads from the XCell II blot module (Invitrogen). The wet blot transfer was set up as per the manufacturers' instructions and proteins were transferred from the SDS-PAGE gel to the PVDF membrane for 2 hours at 30 V. After transfer the membrane was blocked for 2 hours at 4 ^{o}C by incubation in blocking solution containing 5% (w/v) milk powder in TBST (135 mM sodium chloride, 15 mM Tris-HCl, 0.1% (v/v) tween-20, pH 7.6). Following blocking the membrane was washed in TBST for 5 minutes followed by overnight incubation at 4 ^{o}C with a 1:4000 dilution of mouse anti-His-tag antibody in a 1% milk (w/v) TBST solution. After overnight incubation the membrane was washed in TBST as before and then incubated for 2 hours at room temperature with a 1:5000 dilution of goat anti-mouse IgG HRP conjugated antibody in a 1% milk (w/v) TBST solution. Finally, the membrane was washed a minimum of 5 times in TBST with 5 minutes for each wash and then electrochemical luminescence (ECL) Plus (GE Healthcare) detection reagents were added as per the manufacturers' instructions. Protein bands were then observed by X-ray film exposure and development.

2.2.4 Construction of scFv expression vectors

2.2.4.1 Wild type scFvs

The coding sequence for the V_H domain of each scFv was amplified by PCR from a pre-existing pTTod-based expression vector for the Fab (figure 2.2), using a 5' primer incorporating a $(Gly_4Ser)_4$ linker and a 3' primer including a C-terminal hexahistidine tag, as shown below.

Forward ggactgcgtacgggtggaggcgggtcaggtggaggcgggtcaggtggaggcgggtcaggt ggaggcgggtcagaggttcagctggtcgagtctggaggc

(BsiWI restriction site underlined)

Reverse cgcgttgaattcttatcagtgatggtgatggtgatgagaggagactgtcacaagggtcccctg

(EcoR1 restriction site underlined)

PCR was carried out using TaqPlus Precision DNA polymerase (Stratagene) in a 50 µl reaction volume. The PCR protocol consisted of a 1 minute denaturation cycle at 94 ^oC followed by 30 cycles of 1 minute at 94 ^oC, 30 seconds at 50 ^oC for annealing and 30 seconds at 72 ^oC for extension. The final extension was 15 minutes at 72 ^oC. A small amount of each PCR product was identified by electrophoresis on a 1% agarose gel followed by PCR cleanup of the remaining PCR product using a QIAquick PCR purification kit (Qiagen). Cleaned eluate was restriction digested in buffer 3 (New England Biolabs) first for 2 hours with BsiWI (New England Biolabs) at 55 ^oC, followed by 2 hours EcoR1 (New England Biolabs) digestion at 37 ^oC. The Fab pTTod vector was also subjected to the same double restriction digest. Products were gel purified using a QIAquick gel extraction kit (Qiagen). The amplified V_H segments were

then ligated in between the restriction sites of the vector using a PCR product to vector molar ratio of approximately 3:1. Ligation was performed overnight with T4 DNA ligase at 16 ^{o}C. This created a V_L-$(Gly_4Ser)_4$-V_H construct for each scFv. ScFv vectors were then transformed into InvαF competent cells (Invitrogen) using a heat shock protocol consisting of 20 minutes on ice, 40 seconds at 42 ^{o}C then 2 minutes on ice.

2.2.4.2 Extended linker scFvs

Extended linker scFvs, containing a 25 residue $(Gly_4Ser)_5$ linker rather than the 20 residue linker in the wild type scFvs, were PCR amplified with the primers below using the wild type scFv vectors as templates.

Forward ggactgcgtacgggtggaggcgggtcaggtggaggcgggtcaggtggaggcgggtcaggt ggaggcgggtcaggtggaggcgggtcagaggttcagctggtcgag

(BSiWI restriction site underlined)

Reverse cgcggaattcttatcagtgatggtgatggtgatggctcgagactgttacaagagtacc

(EcoR1 restriction site underlined)

PCR amplification and subsequent cloning to produce extended linker V_L-$(Gly_4Ser)_5$-V_H constructs was carried out in the same manner as detailed for the wild type scFvs in section 2.2.4.1.

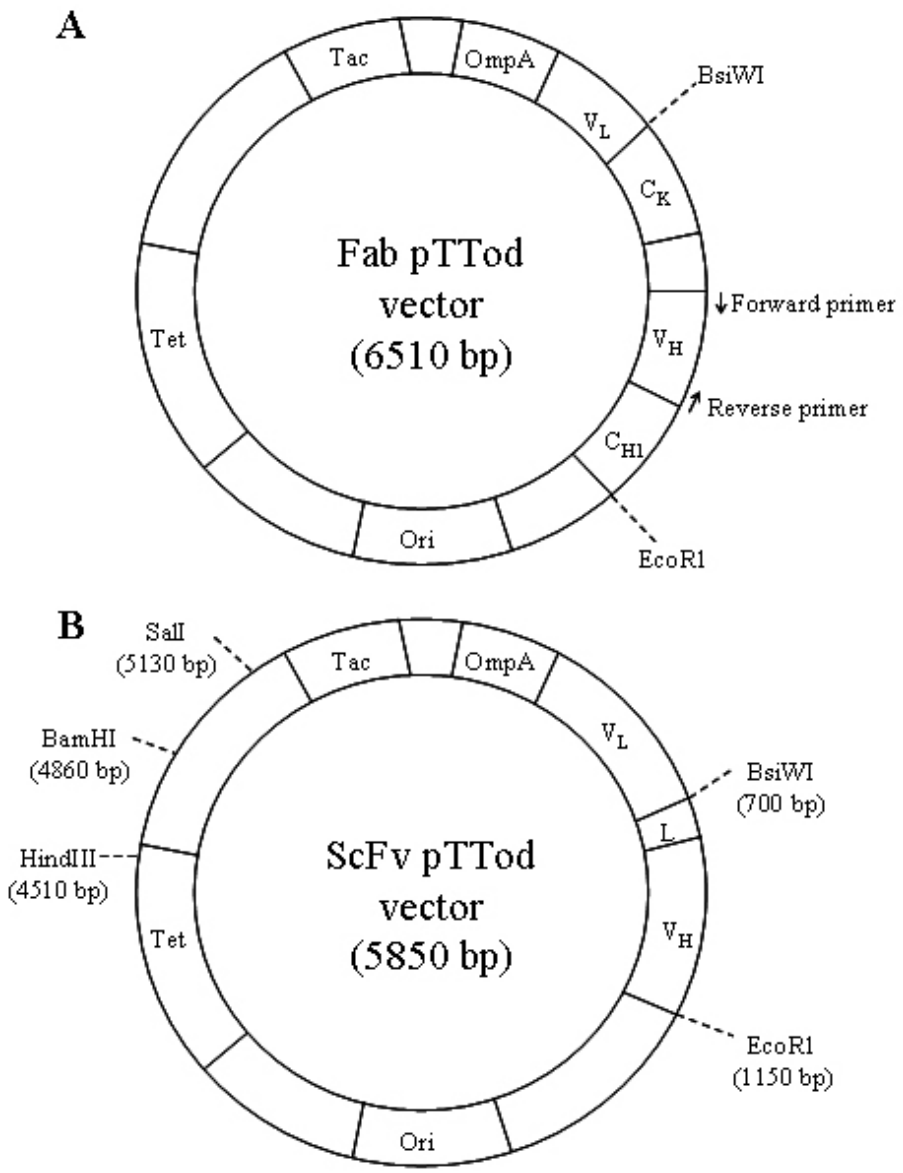

Figure 2.2 Generic maps of the Fab (A) and scFv (B) pTTod vectors. In addition to the constant and variable domains the vectors contain a Tac promoter (Tac), tetracycline resistance gene (Tet) and an OmpA leader to target transport of the protein to the periplasm. The primer and restriction sites for engineering of a scFv construct from the Fab vector are indicated in panel A. Panel B shows the linker region (L) encoded by the forward primer to create a V_L-linker-V_H construct. A selection of restriction sites are also indicated in panel B.

2.2.4.3 Mutant scFvs

To engineer mutant scFv constructs with a specific Alanine to Aspartate mutation within FR 3 of the V_H domain, a two stage PCR amplification protocol was performed using primers A to D listed below. A schematic representation of this is also shown in figure 2.3.

Forward A – cggtcagggcactaaagtag

C – gatgaactctctccgcgacgaggacacagcagtc

Reverse B – gactgctgtgtcctcgtcgcggagagagttcatc

D – cgcggaattcttatcagtgatggtgatggtgatggctcgagactgttacaagagtacc

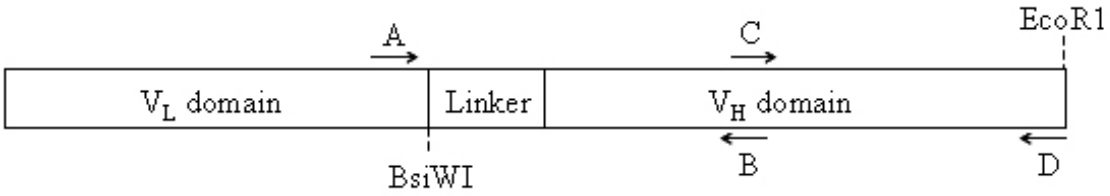

Figure 2.3 A schematic diagram of the scFv construct. A, B, C and D represent the primers used to amplify the variable heavy domain with an Alanine to Aspartate mutation encoded by primers B and C.

Two separate PCR reactions were performed using the same cycle protocol detailed in section 2.2.4.1 with the wild type scFvs as the DNA templates. The first reaction used primers A and B, while the second used C and D. The PCR products were cleaned up using a QIAquick PCR purification kit (Qiagen) and then used as templates in an additional PCR reaction with primers A and D. This generated mutant V_H domains with a single base change, which were then purified and ligated into the vector in a similar manner described for the wild type scFvs in section 2.2.4.1.

Mutant scFvs with the Alanine to Aspartate mutation and an extended 25 residue linker were also engineered using the same procedure but with the extended linker scFvs used as the initial PCR template.

2.2.5 Expression trials

Expression trials for the scFv constructs were performed in 50 ml LB broth cultures with the addition of 10 μg/ml tetracycline. Cultures were seeded from 50 ml overnight cultures to an optical density at 600 nm (OD_{600}) of 0.1, then incubated at 37 ^{o}C until an OD_{600} of between 0.8 and 1.0 was reached. At this point the cultures were either maintained at 37 ^{o}C or cooled to 30, 25, 20 or 15 ^{o}C and protein expression induced overnight by the addition IPTG to a concentration of 0.2 or 0.1 mM. Analysis of the protein expression trials was performed by western blot as described in section 2.2.3.

The expression of deuterated scFvs and IL-1β was initially trialed at a 50 ml culture volume scale. *E. coli* cells transformed with the correct protein expression vector were conditioned to grow in 100% D_2O minimal media by successive overnight incubations in progressively higher percentages of deuterated minimal media (0, 30, 70 and 100%). Glycerol stocks were made from these D_2O conditioned cells. Cultures were then seeded from glycerol stocks and protein expression induced as detailed in section 2.2.6 for the scFvs and IL-1β. *E. coli* doubling times of approximately 4 hours were achieved in 100% D_2O minimal media.

2.2.6 Protein expression and purification

The scFv vectors were transformed into *E. coli* W3110 competent cells (UCB-Celltech) to prepare unlabelled, ^{15}N, $^{15}N/^{2}H$ or $^{15}N/^{13}C/^{2}H$ labelled scFv samples. For labelled samples, cells were grown in minimal media containing 1 g/l >99% ^{15}N ammonium sulphate and 2 g/l >99% ^{13}C glucose if required, as the sole nitrogen and carbon sources, and >99.8% D_2O when appropriate. Transformants were initially grown at 37 ^{o}C in media containing 10 μg/ml tetraycline to an OD_{600} of approximately 1.0, at which point they were cooled to 20 ^{o}C and scFv expression induced by the addition of IPTG to a concentration of 0.2 mM. Cells were harvested approximately 20 hours post-induction in protonated media or 44 hours in deuterated media by centrifugation (6000 g for 20 minutes). The harvested cells were lysed by resuspending in 50 mM Tris-HCl, 2 mM EDTA, and 0.1% triton X100 (v/v) buffer at pH 8, to which was added 100 μg/ml lysozyme, 0.1 mM PMSF, 5 mM magnesium chloride and deoxyribonuclease 1, and incubated for 40 minutes. Insoluble debris was removed by high speed centrifugation (~50,000 g for 30 minutes) and the cleared lysate was then dialysed into the running buffer for the Ni-NTA column (100 mM NaCl, 25 mM Tris, 15 mM imidazole buffer at pH 7.5). The scFv was separated from contaminating proteins on a 6 ml Ni-NTA column (GE Healthcare) and eluted with a linear imidazole gradient (15-500 mM) at a flow rate of 2 ml/min. ScFv containing fractions were then dialysed into the running buffer for gel filtration (100 mM sodium chloride, 25 mM sodium phosphate, 0.02% (w/v) sodium azide buffer at pH 6.5) and subjected to final purification on a Superdex 75 16/60 prep grade gel filtration column (GE Healthcare) at a flow rate of 1 ml/min.

Samples of IL-1β were prepared from a pET-21 based *E. coli* expression vector, which includes a N-terminal hexahistidine tag with a TEV protease cleavage site in the linker.

Tuner (DE3) pLysS cells (Novagen) transformed with the IL-1β expression vector were grown at 30 ^{o}C in media containing 100 μg/ml carbenicillin to an OD_{600} of approximately 0.8, at which point cultures were cooled to 25 ^{o}C and IL-1β expression induced overnight with IPTG to a concentration of 0.01 mM. Cells were harvested, lysed and the His-tagged IL-1β purified in the same manner as described for the scFvs. Removal of the His-tag was achieved by incubation of the purified protein with TEV protease (TEV:IL-1β 1:15 w:w) at room temperature overnight. This mixture was then dialysed into a Ni-NTA running buffer (100 mM NaCl, 25 mM Tris, 15 mM imidazole buffer at pH 7.5) and IL-1β separated from the His-tag and TEV protease by passing through a Ni-NTA column. IL-1β fractions were subsequently dialysed into a running buffer for gel filtration (100 mM sodium chloride, 25 mM sodium phosphate, 0.02% (w/v) sodium azide buffer at pH 6.5), which allowed final purification of IL-1β on a Superdex 75 16/60 prep grade column (GE Healthcare) at a flow rate of 1 ml/min.

ScFv-IL-1β complex was prepared by addition of a 20% molar excess of concentrated IL-1β (typically >2 mg/ml) to a dilute sample of scFv (<0.5 mg/ml). The excess IL-1β, monomeric scFv-IL-1β and dimeric scFv-IL-1β were separated by gel filtration on a Superdex 75 16/60 prep grade column (flow rate of 1 ml/min) pre-equilibrated with a 100 mM sodium chloride, 25 mM sodium phosphate, 0.02% (w/v) sodium azide buffer at pH 6.5.

2.2.7 Circular dichroism spectroscopy

Far UV circular dichroism (CD) spectra were obtained for scFv samples ranging from 10 to 20 μM in a 25 mM sodium phosphate, 100 mM sodium chloride, 0.02% (w/v)

sodium azide buffer at pH 6.5 in a cell with path length 0.1 mm. Spectra were recorded on a Jasco J-715 Spectropolarimeter between 190 and 250 nm at a resolution of 1 nm and scan speed of 20 nm/min, with each spectrum representing the average of 10 accumulations. Spectra were corrected for the buffer and converted to molar CD per residue before secondary structure content was determined using the CDPro package (Sreerama & Woody, 2000). CD thermal denaturation studies to analyse the effect of increasing temperature on the structure were recorded between 5 and 85 ^{o}C with CD spectra collected every 5 ^{o}C. In addition, changes in the intensity of the negative peak at 218 nm were recorded as a function of increasing temperature, with the temperature rising incrementally at 1 ^{o}C/min.

2.2.8 Analytical gel filtration

Analysis of the scFv domain-swapped dimer formation at concentrations between approximately 3 μM and 500 μM was performed in a 25 mM sodium phosphate, 100 mM sodium chloride, 0.02% sodium azide buffer at pHs ranging from 6.0 to 8.0, using an analytical Superdex 75 10/300 GL gel filtration column at a flow rate of 1 ml/min. The column was calibrated using low molecular weight calibration standards (GE Healthcare) and monomer:dimer ratios calculated by peak integration.

2.2.9 NMR spectroscopy

NMR spectra were acquired from 0.35 ml samples of 0.2-0.5 mM free scFv and scFv-IL-1β complex in a 25 mM sodium phosphate, 100 mM sodium chloride, 10 μM EDTA, 100 μM AEBSF, 0.02% sodium azide buffer at pH 6.5, containing 10% D_2O/90% H_2O. All NMR data were acquired between 25 and 50 ^{o}C on a 600 MHz Bruker DRX or 800

MHz Bruker Avance system. $^{15}N/^{1}H$ HSQC spectra (Bodenhausen & Ruben, 1980) were acquired for between 0.5 and 12 hours with acquisition times of 60 ms and 50 ms in F_2 (^{1}H) and F_1 (^{15}N) respectively. The WATERGATE method was used to suppress the water signal when required (Piotto *et al.*, 1992). The NMR data was processed using Topspin (Bruker Biospin Ltd.) with linear prediction used to extend the effective acquisition times by up to two fold in F_2 (^{15}N).

2.3 Results

2.3.1 Bioinformatics

Four antibodies, known as 120B09, 168, 8516 and IC8, had been identified at UCB-Celltech for their high affinity to IL-1β and potential therapeutic importance. The antibodies were selected by the proprietary selected lymphocyte antibody method (SLAM). SLAM is a method for identifying within a complete immune repertoire, rare, single immune cells secreting antibodies with a desired specificity, affinity and functional activity. The genes encoding the antibodies of interest are extracted from the single cells and assembled into an appropriate antibody structure. This permits high throughput screening for antibodies that in addition to binding to antigens possess rare but desirable, functional properties. As the antibodies are selected from a B-cell population it ensures natural V_L-V_H pairings are selected, unlike display technologies, such as phage (Mccafferty *et al.*, 1990) and ribosome (Hanes & Pluckthun, 1997), where pairings that would not naturally occur can be selected.

A multiple sequence alignment of the scFv fragments of the four antibodies is shown in figure 2.4. This shows the leader sequence, which targets the scFv for transport to the periplasm where the leader sequence is removed and the scFv folded, the $(Gly_4Ser)_4$ linker to ensure the association of the two variable domains and the His-tag to allow for simple purification. The six CDRs are clearly the most variable regions of the four antibodies, with the framework regions between them sharing much more sequence homology. Excluding the leader sequence, linker and His-tag, the four sequences share approximately 60% identity. As well as sequence variability, the CDRs also differ in length, with scFv 168 in particular consisting of a very long CDR L1 in comparison to

the other scFvs. The only CDR that is invariant in length between the four scFvs is CDR L2. The highlighted conserved Alanine residue in FR H3 is thought to form part of a hydrophobic patch that is involved in the formation of scFv domain-swapped dimers.

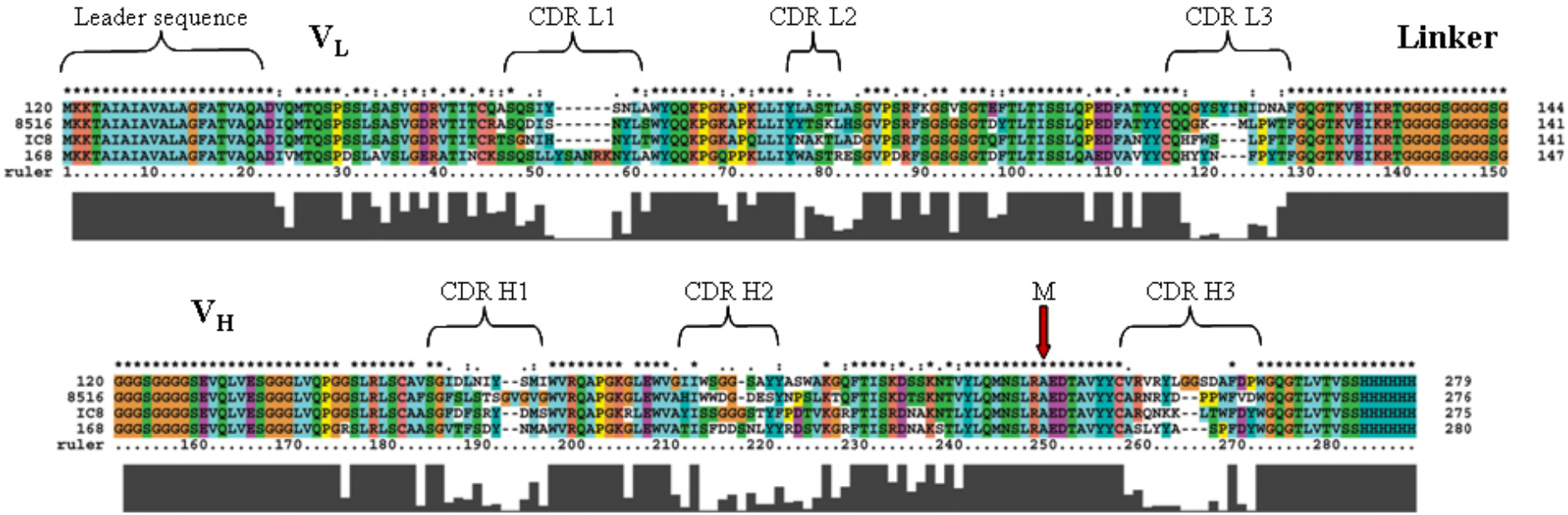

Figure 2.4 Multiple sequence alignment of the four anti-IL-1β scFvs investigated. The variable domains, linker and CDRs are indicated with the consensus between the four sequences shown underneath. As expected for antibodies the sequence homology is very low in the CDRs with the framework regions having a greater similarity. The M indicated in framework region 3 of the V_H domain (FR H3) is a conserved Alanine residue that is part of a hydrophobic patch, which is thought to stabilise the domain-swapped scFv dimer.

2.3.2 Construction of the scFv expression vectors

PCR amplification of the relevant scFv V_H regions was confirmed by agarose gel electrophoresis. A typical gel showing the amplified DNA of correct size is shown in figure 2.5. Double restriction digest (BSiWI and EcoR1) of the PCR products and template Fab vectors, and subsequent ligation of the amplified V_H domains into the pTTod vector was also monitored by agarose gel electrophoresis. Successfully ligated scFv constructs were then transformed into InvαF competent cells (Invitrogen) and positive colonies selected by antibiotic resistance to tetracycline were screened by colony PCR and multiple restriction digests (figure 2.6).

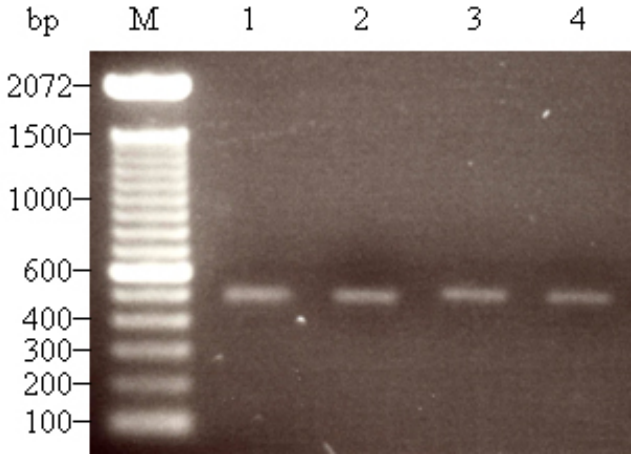

Figure 2.5 A 1% (w/v) agarose gel stained with ethidium bromide and viewed under UV light. Lane M contains a 100 bp marker (Invitrogen) with bands from 100 to 2072 bp as indicated. Lanes 1 to 4 contain purified PCR product for the V_H domains of 8516, IC8, 120B09 and 168 respectively with an approximate size of 450 bp.

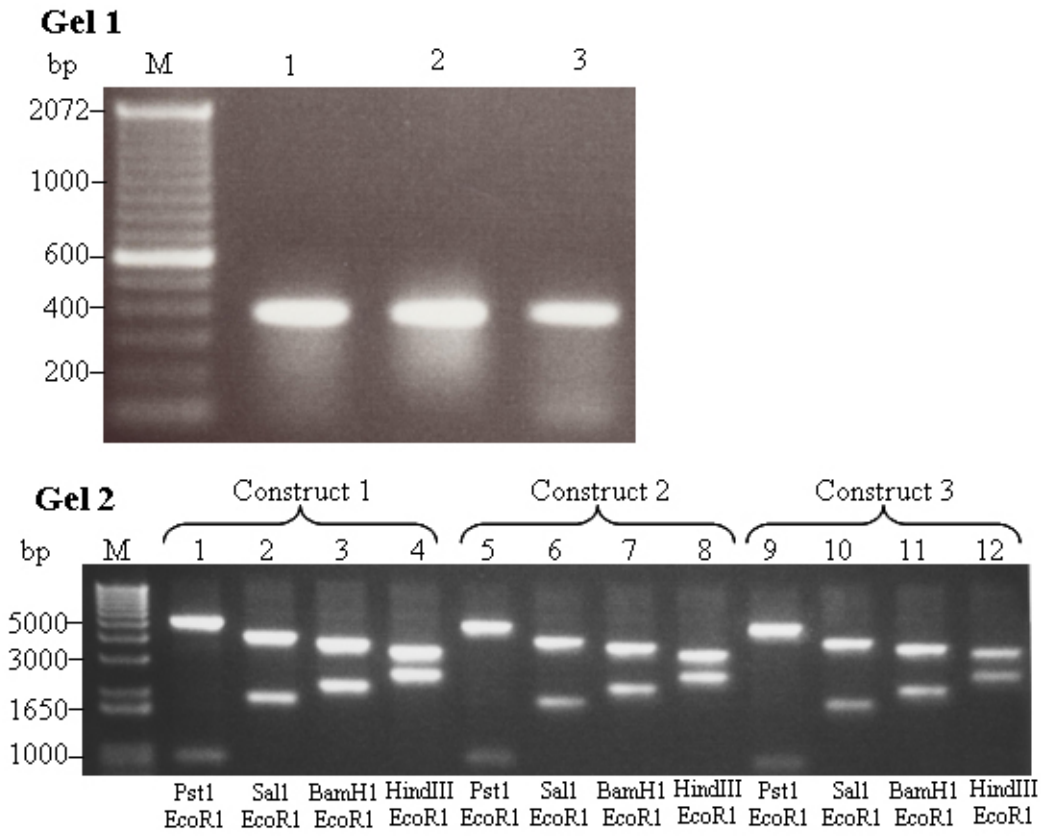

Figure 2.6 1% (w/v) agarose gels analysing 3 positive colonies from the transformation of scFv 8516 constructs. Gel 1 shows the results from a colony PCR screen of the 3 constructs. Lane M contains a 100 bp DNA ladder (Invitrogen) and lanes 1 to 3 contain a positive PCR result from each construct, showing that the V_H domain has been ligated into the vector. Gel 2 shows multiple restriction digests of the same 3 constructs. Lane M contains a 1 kbp DNA ladder (Invitrogen). Lanes 1 to 4 contain fragments from construct 1, lanes 5 to 8 are for construct 2 and lanes 9 to 12 construct 3. Lanes 1, 5 and 9 are the result of double digestion with Pst1 and EcoR1. Lanes 2, 6 and 10 show digestion with Sal1 and EcoR1. Lanes 3, 7, 11 are fragments resulting from digestion with BamH1 and EcoR1. Lanes 4, 8 and 12 are the products of a HindIII and EcoR1 double digest. The pattern of DNA fragments displayed for each construct is as expected for a correctly engineered V_L-linker-V_H construct in a pTTod vector (refer to panel B of figure 2.2 for the location of the restriction sites).

Table 2.1 summarises the scFv constructs produced and their sequences confirmed by DNA sequencing. The table also shows the naming scheme that was adopted for the constructs.

Table 2.1 A table of the scFv constructs successfully engineered from the four Fab vectors 120B09, 168, 8516 and IC8. The wild type scFvs contain the two variable domains from the Fab with a 20 residue Gly_4Ser repeat linker between them. The mutant scFvs contain an Ala to Asp mutation within a hydrophobic patch of framework region 3 of the V_H domain. The extended linker scFvs are identical to the wild type except for containing a 25 residue Gly_4Ser repeat linker. The extended linker and mutant scFvs contain both the 25 residue linker and the Ala to Asp mutation.

Wild type scFv	**Mutant scFv**	**Extended linker scFv**	**Extended linker and mutant scFv**
120B09	120-M	120-X	120-XM
168	168-M	168-X	168-XM
8516	8516-M	8516-X	
IC8	IC8-M		

2.3.3 Expression trials

Previous expression trials on scFvs at UCB-Celltech suggested that maximum protein expression is achieved by induction overnight with IPTG to a concentration of 0.2 mM. This conclusion was confirmed by western blot analysis of expression trials for the four anti-IL-1β scFvs. Expression trials at 37, 30, 25, 20 and 15 ^{o}C (figure 2.7) determined whether the scFvs could be expressed in a soluble form. All forms (wild type, extended linker and mutant) of 8516 and IC8 were soluble when expressed at 20 ^{o}C and below. The expression levels were similar for wild type, extended linker and mutant scFvs

8516 and IC8, as determined by purification of the soluble scFvs. All 168 constructs expressed mainly insoluble scFv even at 15 ^{o}C and in all cases the expression level of this scFv appeared to be lower than observed for 8516 and IC8 by western blot analysis. 120BO9 appeared to show no expression at all conditions. Similar trials to determine expression levels of 8516 and IC8 in minimal media were also performed. Soluble protein was expressed in both cases at 20 ^{o}C but expression levels were approximately 50% of that seen in rich media. On the basis of these expression results scFvs 8516 and IC8 were selected as the proteins to continue to study in this project due to their solubility.

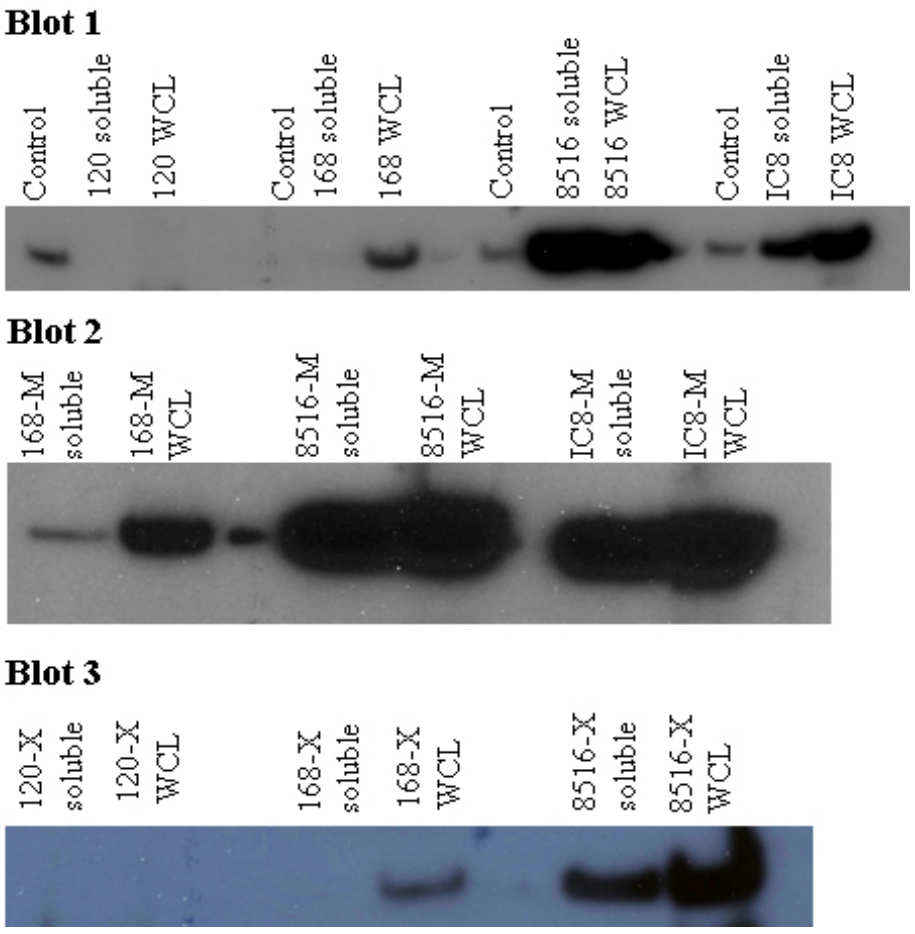

Figure 2.7 Western blots of expression trials for the scFv constructs induced overnight at 15 ^{o}C with IPTG to a concentration of 0.2 mM. The scFvs were detected using a

mouse anti-His tag antibody (Novagen) as the primary antibody to which a HRP conjugated goat anti-mouse IgG secondary antibody (Pierce) was targeted. ECL Plus western blot detection reagents (GE Healthcare) were used to observe the proteins by exposure of X-ray film. Samples were taken from the periplasm and whole cell lysate (WCL) to determine whether the protein was exported to the periplasm in a soluble form. Blot 1 shows samples from all four wild type scFv expression trials. A positive control of purified scFv is run next to each periplasm and WCL sample. The results show that both 8516 and IC8 are expressed in a soluble form in the periplasm, while 168 is mainly found in the WCL and so is mainly insoluble. 120B09 does not appear to be expressed. Blot 2 shows an expression trial for the mutated scFvs. 8516-M and IC8-M are both soluble in the periplasm, whereas only a small amount of 168 is soluble. 120-M is not shown due to its lack of expression. Blot 3 is an expression trial for the extended linker scFvs. 120-X shows no signs of expression, 168-X is expressed but in an insoluble form and 8516-X is expressed in a soluble form in the periplasm.

Worked carried out at UCB-Celltech had determined the optimum conditions for expression of IL-1β in Tuner (DE3) pLysS cells (Novagen). Expression trials were performed to confirm this and evaluate IL-1β expression in minimal media and deuterated media, as illustrated in figure 2.8.

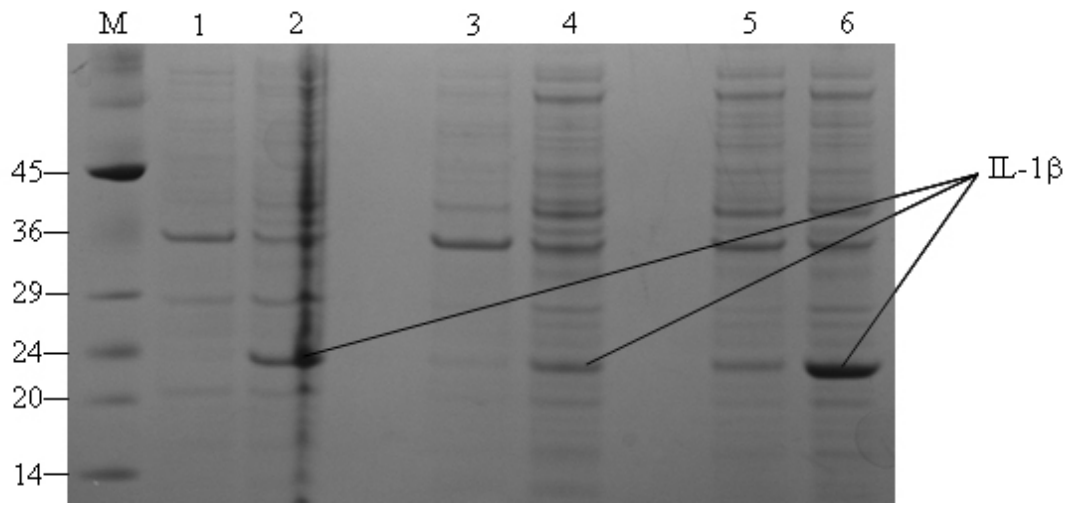

Figure 2.8 Coomassie stained SDS-PAGE gel of IL-1β expression trials. Lane M contains a 14-66 kDa molecular weight marker (Sigma) as indicated. Lanes 1, 3 and 5 are pre-induction whole cell samples from cultures grown in minimal media, deuterated media and LB respectively. Lanes 2, 4 and 6 are overnight post-induction samples from minimal media, deuterated media and LB respectively. All post-induction samples show expression of the 20.5 kDa His-tagged IL-1β.

2.3.4 ScFv expression and purification

The scFv 8516 and IC8 expression vectors were transformed into W3110 *E. coli* cells and protein expression was induced overnight by addition of IPTG to a concentration of 0.2 mM at 20 ^{o}C once an OD_{600} of 0.8-1.0 had been reached, as was determined for optimum protein expression in section 2.3.3. A two-stage purification procedure was used to purify the scFvs. Presence of a polyhistidine tag at the C-terminus of the scFvs allowed initial purification by affinity chromatography (figure 2.9) followed by a final gel filtration purification step, shown in figure 2.10.

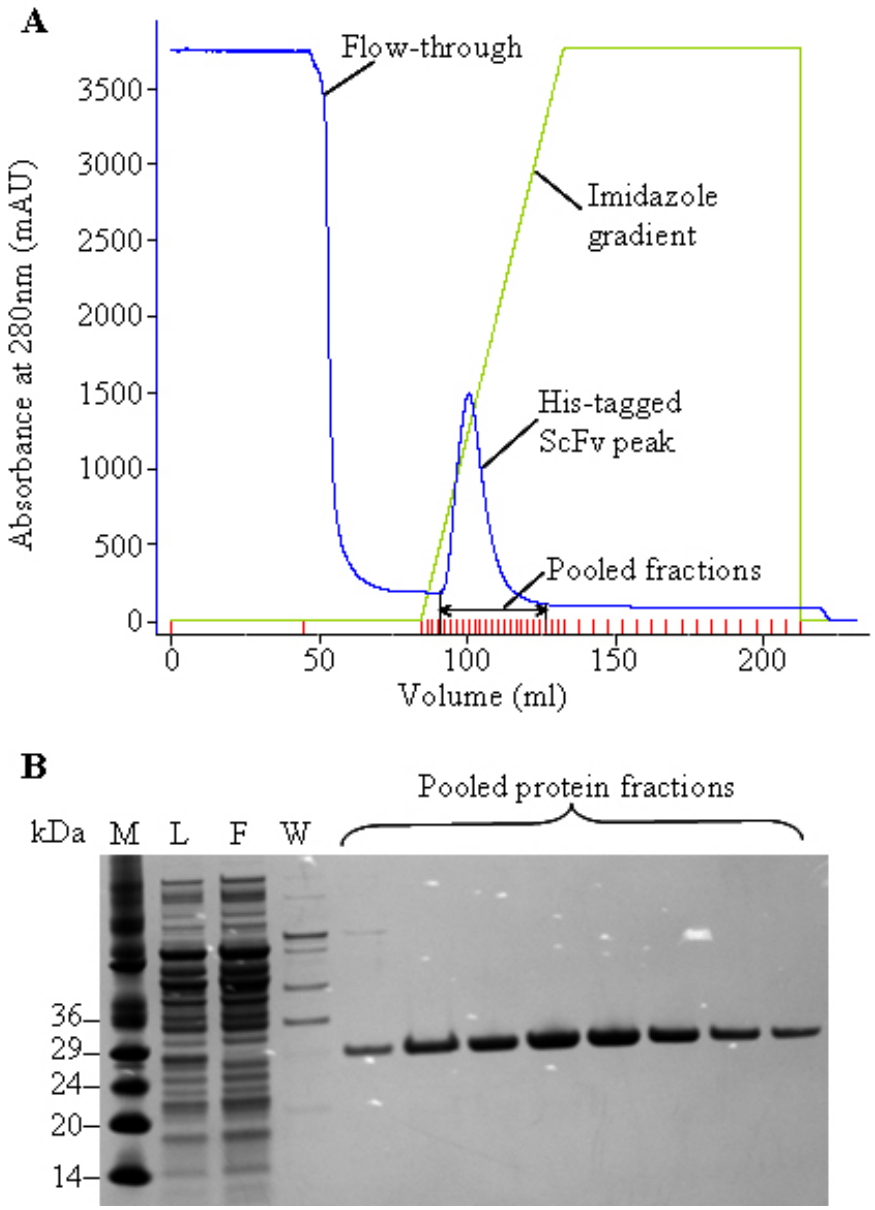

Figure 2.9 Panel A shows a typical FPLC chromatogram obtained for purification of the His-tagged scFv 8516 using a Ni-NTA column (GE Healthcare). The scFv was eluted over a linear gradient from 15 to 500 mM imidazole, as indicated by the green line. Panel B shows a SDS-PAGE gel of fractions from the purification shown in panel A. Lane M contains a 14-66 kDa molecular weight marker (Sigma) as indicated. The next 3 lanes contain samples of the original load onto the column (L), the flow-through (F) and the column wash (W). The remaining lanes represent every second fraction of purified scFv sample from the pooled fractions indicated in panel A.

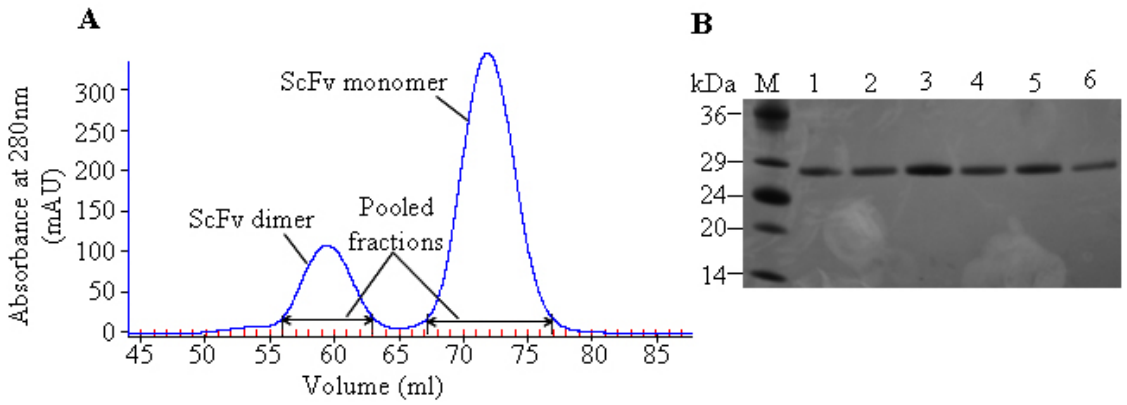

Figure 2.10 Panel A shows a typical chromatogram obtained for gel filtration purification of scFv 8516 on a Superdex 75 16/60 column. The pooled fractions from the monomer and dimer peaks are shown. Panel B shows a non-reducing SDS-PAGE gel of the pooled fractions from the purification shown in panel A. Lane M contains a 14-66 kDa molecular weight marker (Sigma) as indicated. Lanes 1-3 are samples evenly distributed across the dimer peak and lanes 4-6 are samples from the monomer peak.

As shown by the heavily Coomassie stained gel in panel B of figure 2.10, the purified scFv was estimated to be over 98% pure as no other protein bands could be seen. Yields of 4-5 mg of purified protein were obtained per litre of LB culture, with yields reducing by approximately 50% in minimal media. The correct mass of the purified scFvs were confirmed by mass spectrometry.

2.3.5 IL-1β expression and purification

The IL-1β expression vector was transformed into Tuner (DE3) pLysS cells (Novagen) and protein expression was induced overnight as outlined in section 2.2.6. Initial purification of the N-terminal His-tagged IL-1β from the soluble whole cell lysate was by affinity chromatography on a Ni-NTA column, as shown in figure 2.11.

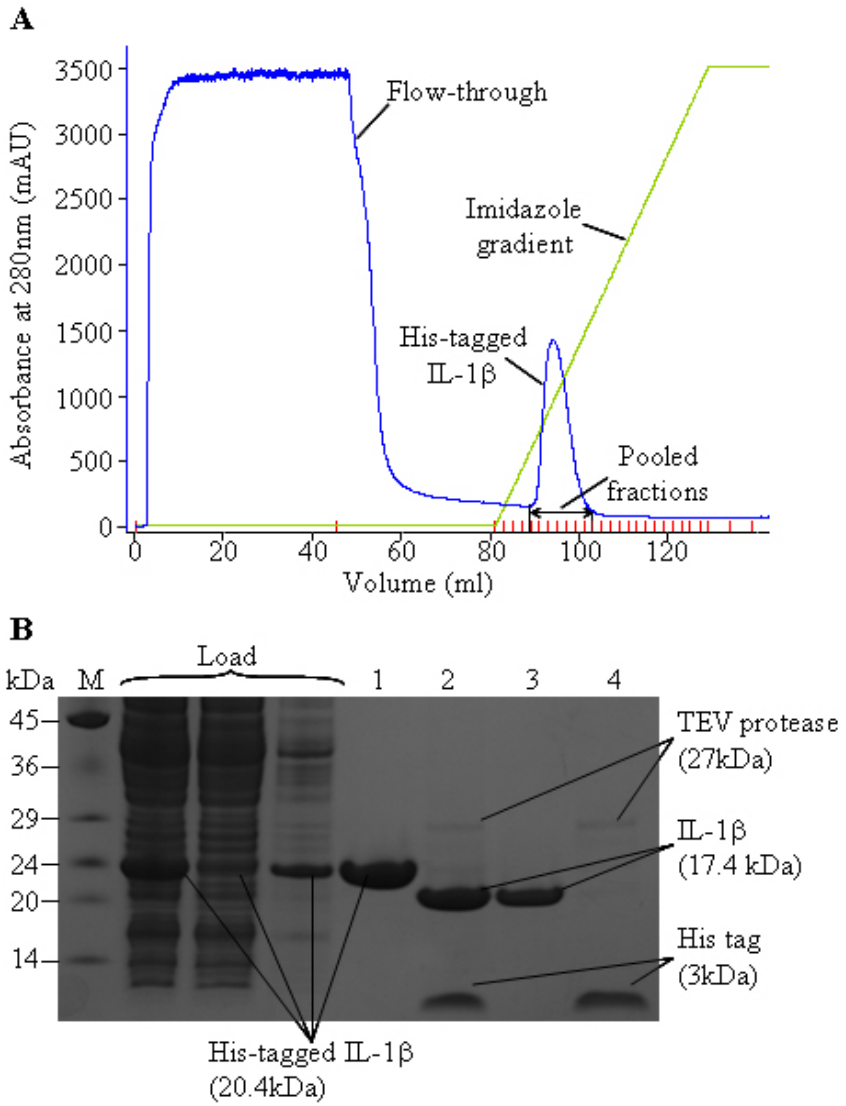

Figure 2.11 Panel A shows a typical chromatogram of affinity purification of His-tagged IL-1β on a Ni-NTA column. The protein was eluted over a linear gradient from 15-500 mM imidazole and the purified IL-1β fractions collected as indicated. Panel B shows a SDS-PAGE gel of fractions from panel A and subsequent cleavage of the His-tag using TEV protease. Lane M contains a 14-66 kDa molecular weight marker as indicated. Lane 1 contains the purified IL-1β from the pooled fractions. Lane 2 shows successful cleavage of the His-tag. Lanes 3 shows the purified untagged IL-1β from the purification step shown in figure 2.12, with lane 4 being the TEV protease and His-tag that have been separated from IL-1β.

Pooled IL-1β fractions were cleaved with TEV protease and dialysed back into Nickel column running buffer. Successful cleavage of the His-tag from IL-1β, shown in panel B of figure 2.11, was followed by purification of the untagged IL-1β on a Ni-NTA column. The cleaved His-tag and His-tagged TEV protease bind to the column when loaded allowing the collection of purified untagged IL-1β in the flow-through (figure 2.12). IL-1β was then subjected to a final gel filtration purification step, shown in figure 2.13.

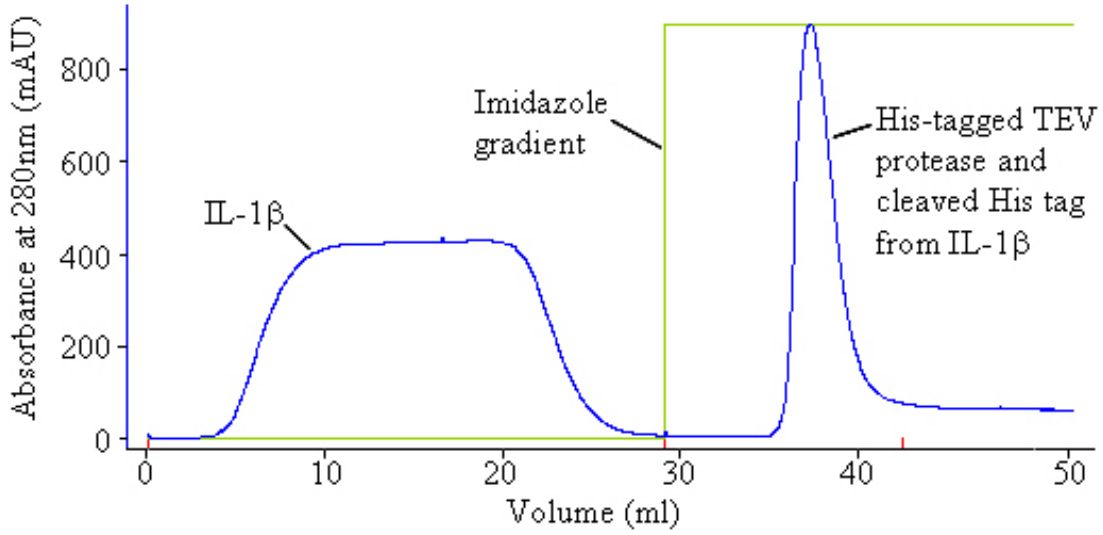

Figure 2.12 Affinity chromatography purification of IL-1β on a Ni-NTA column. After successful TEV cleavage IL-1β flows through the Nickel column without binding and so can be collected while the cleaved IL-1β His-tag and His-tagged TEV protease bind. Once all the IL-1β had been recovered an imidazole step gradient from 15 to 500 mM was applied to elute the bound proteins from the column. A SDS-PAGE gel showing cleavage and purification of the untagged IL-1β is shown in figure 2.11.

The heavily Coomassie stained gel shown in panel B of figure 2.13 shows the purity of IL-1β, which is estimated to be over 98% due to the absence of any visible contaminating bands. Yields obtained for purified IL-1β were over 30 mg of protein per original litre of LB culture, with yields reducing by approximately 50% in minimal media.

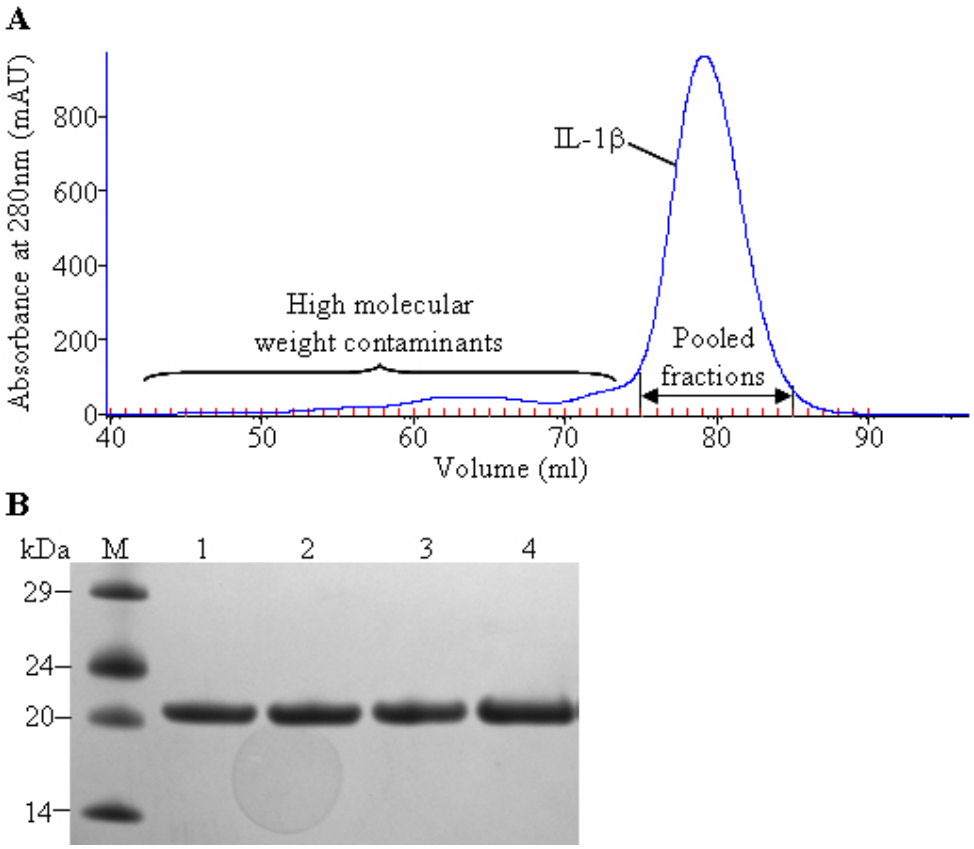

Figure 2.13 Panel A shows a typical gel filtration chromatogram obtained for purification of IL-1β on a Superdex 75 16/60 prep grade column. A small amount of high molecular weight contaminants have been separated and the clean IL-1β fractions were pooled as indicated. Panel B shows a SDS-PAGE gel of some of the pooled fractions from panel A. Lane M contains a 14-66 kDa molecular weight marker as indicated. Lanes 1-4 are samples from the IL-1β peak shown in panel A.

2.3.6 Circular dichroism spectroscopy

CD spectroscopy was used to determine whether the scFvs purified in section 2.3.4 were folded. The far UV CD spectra obtained for scFv IC8 (figure 2.14) is typical of a predominantly β-sheet folded protein. Spectral analysis using the CD Pro software package (Sreerama & Woody, 2000) gave typical secondary structure predictions for the scFvs of 0 ±3% helix, 48 ±8% β-sheet, 22 ±2% turn and 29 ±8% random coil.

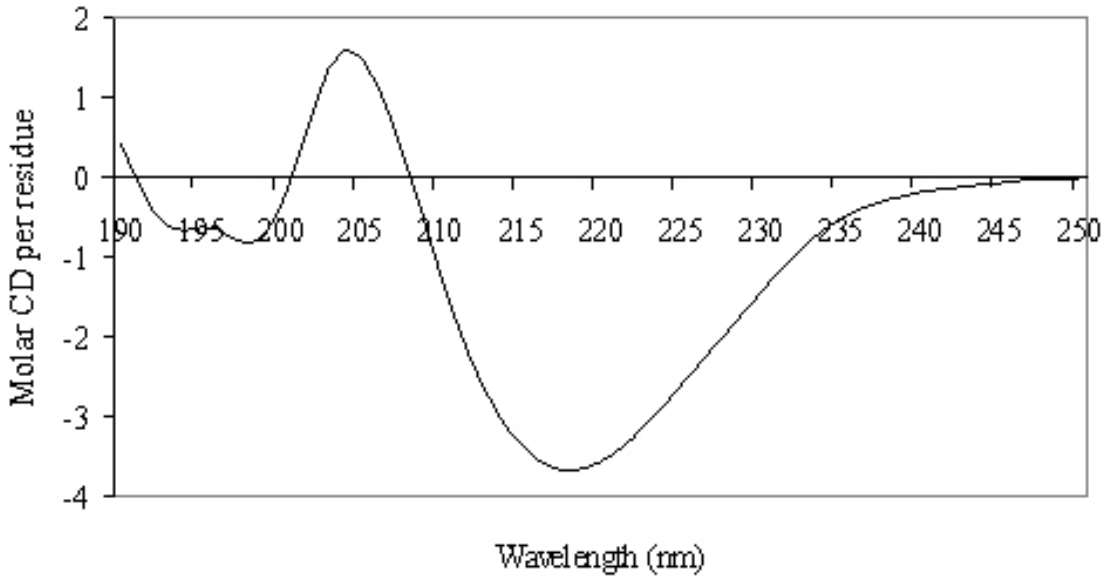

Figure 2.14 A far UV circular dichroism spectrum for a 13.85 μM sample of the scFv IC8 measured between 190 and 250 nm at 20 ^{o}C. The data has been converted to molar CD per residue, buffer corrected and smoothed. The negative peak at 218 nm is indicative of a predominantly β-sheet protein as is confirmed by the CD Pro prediction of 48% β-sheet.

CD thermal denaturation studies monitoring the intensity of the negative peak at 218 nm between 5 and 85 ^{o}C showed unfolding of the scFv once a temperature of about 60 ^{o}C was reached (figure 2.15). The cooperative sigmoidal unfolding curve is indicative of a protein that contains stable tertiary structure and is almost certainly fully folded.

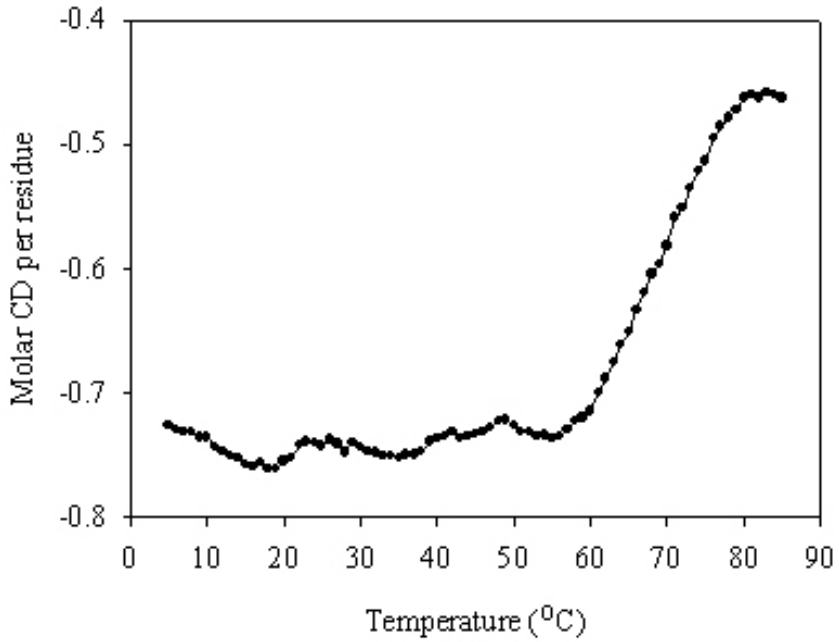

Figure 2.15 A graph showing thermal denaturation of an 11 μM sample of the scFv 8516 as monitored by circular dichroism spectroscopy at a wavelength of 218 nm with measurements taken every 1 °C between 5 and 85 °C. The increase in molar CD per residue after 60 °C shows unfolding of the scFv.

2.3.7 Analysis of the scFv domain-swapped dimer equilibrium

Initial gel filtration purification of the scFvs 8516 and IC8 on a Superdex 75 16/60 prep grade column showed the presence of multimeric species. Calibration of the column using a set of low molecular weight calibration standards (GE Healthcare) enabled identification of the three main peaks as monomer, dimer and trimer (figure 2.16). The void volume, Vo, for the column was determined as 44.13 ml by measuring the elution volume, Ve, of the 2 MDa oligosaccharide Blue Dextran. The column volume, Vc, is 120 ml, which means the Kav for each protein can be calculated by the following formula (Laurent & Killander, 1964): $Kav = (Ve - Vo)/(Vc - Vo)$

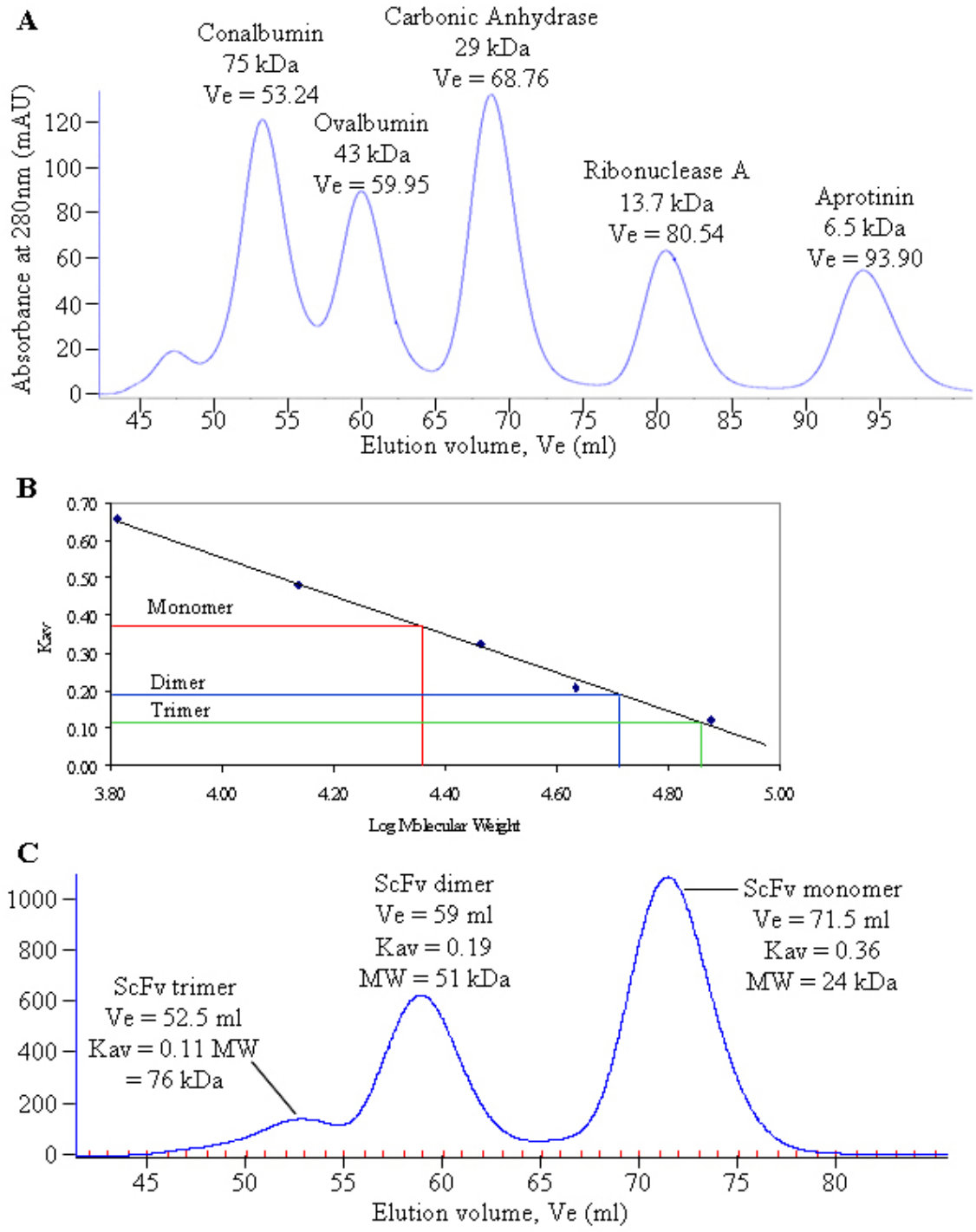

Figure 2.16 Calibration of the Superdex 75 16/60 prep grade gel filtration column (GE Healthcare) using low molecular weight calibration standards. Panel A shows a chromatogram of the protein standards with their elution volumes (Ve) and molecular weights indicated. Panel B is a calibration curve of Kav against log of protein molecular weight, such that $kav = (Ve - Vo)/(Vc - Vo)$, where Vo is the column void volume and

Vc is the column volume. The Kav for each peak shown in the gel filtration chromatogram of scFv 8516 (panel C) is indicated on the graph together with the predicted molecular weights, which identifies peaks arising from monomer, dimer and trimer. The molecular weights determined by gel filtration correspond well with the predicted values based on protein sequence of 27.3, 54.6 and 81.9 kDa for monomer, dimer and trimer respectively.

The dependence of the monomer to dimer equilibrium on concentration was analysed in detail by analytical gel filtration on a Superdex 75 10/300 GL column. Calibration of the column was performed in a similar manner to the Superdex 75 16/60 prep grade column using low molecular weight calibration standards (GE Healthcare). Figure 2.17 shows how the percentage of dimer content within the purified scFv sample increases with concentration until an approximate 1:1 ratio of monomer:dimer is achieved at 500 μM. An increase in the amount of precipitation was also noted upon increasing scFv concentration. Samples of scFv at low protein concentration (<20 μM) remained stable at 4^{o}C for long periods of time, while more concentrated samples (>100 μM) under the same conditions showed signs of protein precipitation over a period of days and weeks. This is possibly due to the formation of higher multimeric species in the more concentrated samples.

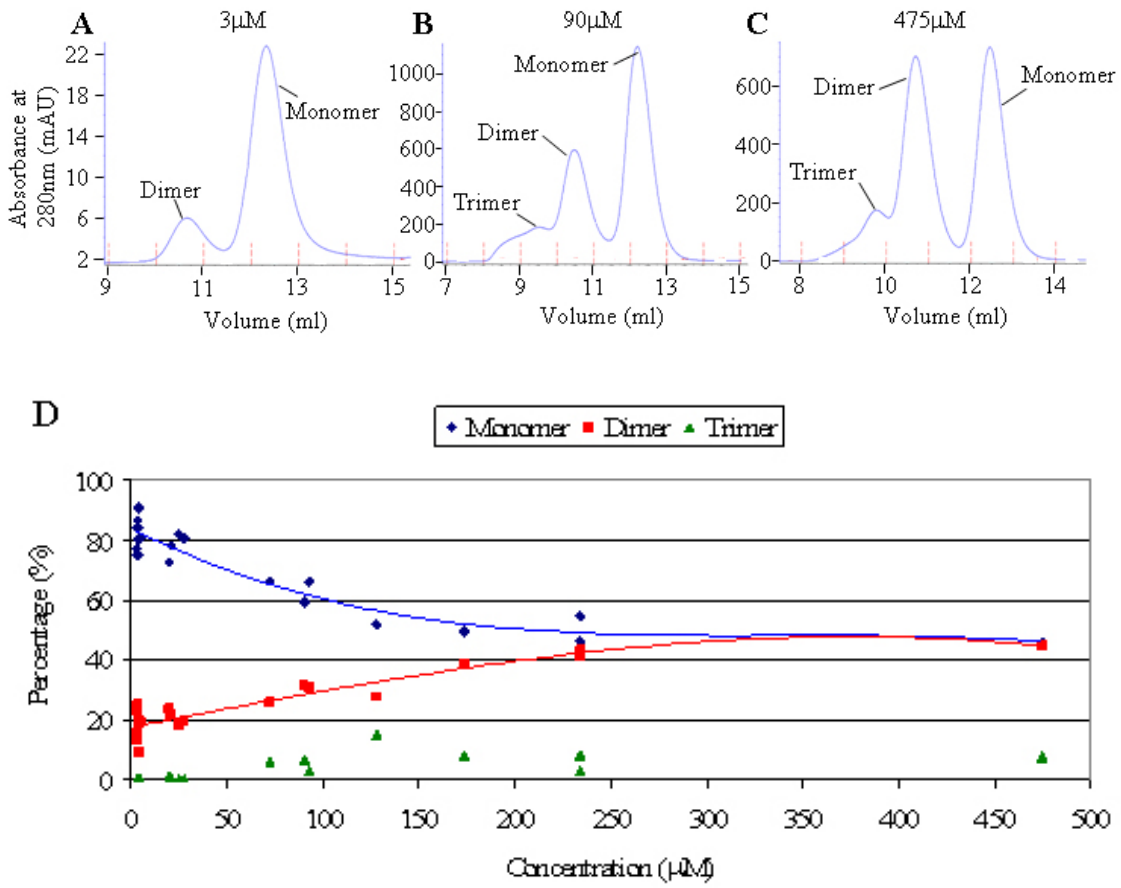

Figure 2.17 Panels A, B and C show analytical gel filtration chromatograms for the same scFv 8516 sample at 3, 90 and 475 µM respectively, with different volumes of sample loaded onto the column in each case. Higher concentrations result in an increase in the percentage of dimer present. Panel D summarises all the analytical gel filtration data observed for scFv 8516 at pH 6.5, showing a decrease in the percentage of monomer and concurrent increase in the percentage of dimer as concentration of the protein sample increases up to 500 µM.

It has been suggested in a previous study (Arndt *et al.*, 1998) that pH can effect the balance of the monomer:dimer equilibrium. However, no change in the ratio of monomer to dimer was seen for the scFvs 8516 and IC8 at pHs between 6.0 and 8.0. Many other factors such as scFv orientation, linker length and antibody type have also

been reported to have an impact on the domain-swapped dimer equilibrium, as described previously in section 2.1.3. For this reason scFvs with extended 25 residue linkers and with an Ala to Asp mutation were engineered as described in section 2.2.4, in the hope of disrupting the formation of the dimer to enable free monomeric scFv to be analysed at high concentration. The extended linker scFvs showed no difference in their behaviour to the wild type (data not shown). The mutant scFvs displayed a significantly lower percentage of dimer at high concentration compared to the wild type on initial analysis. However, over a period of approximately 72 hours the monomer:dimer ratio returned to the levels seen for the wild type (figure 2.18).

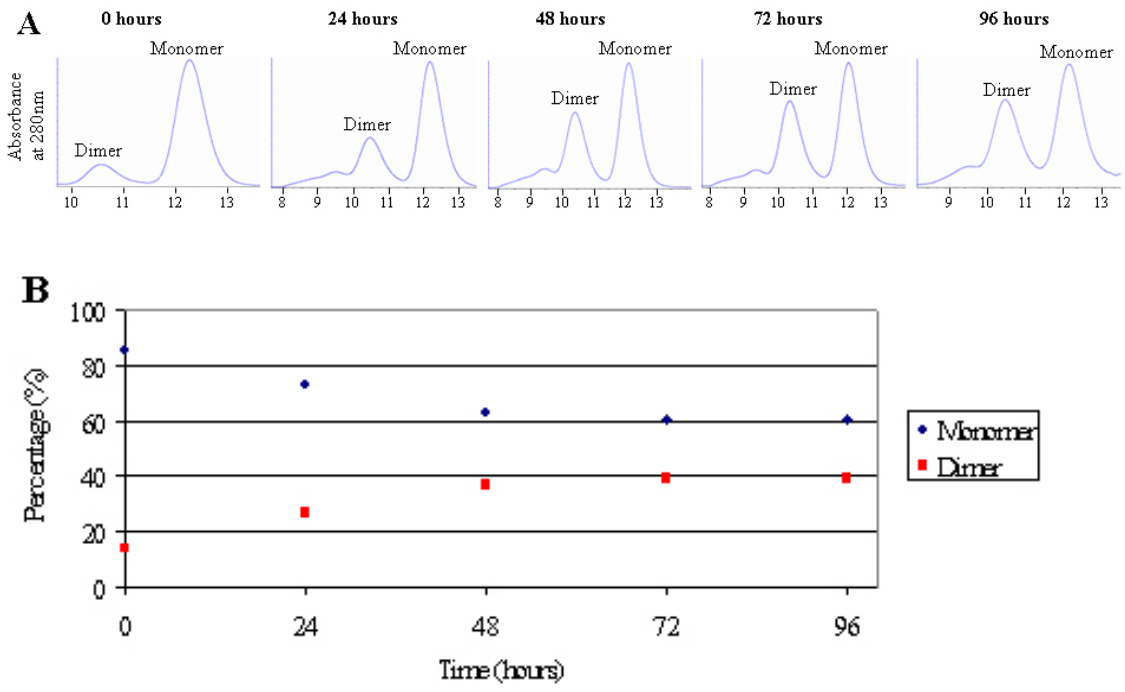

Figure 2.18 Analysis of the monomer-dimer equilibrium for the scFv 8516-M. Panel A shows analytical gel filtration chromatograms of a 189 µM sample of scFv 8516-M over a period of 96 hours after initial concentration. The peaks were integrated and relative percentages of the monomer and dimer are shown in the graph in panel B. The monomer-dimer equilibrium stabilises after approximately 72 hours.

2.3.8 ScFv-IL-1β complex purification

Formation of scFv-IL-1β complex was achieved by addition of a 20% molar excess of purified IL-1β to a dilute sample of purified scFv (typically less than 20 μM). The resulting complex was then purified on a Superdex 75 16/60 gel filtration column (figure 2.19), producing resolved protein peaks for both monomeric and dimeric scFv-IL-1β complex, and the excess of free IL-1β. The monomeric complex was pooled and used in subsequent NMR experiments.

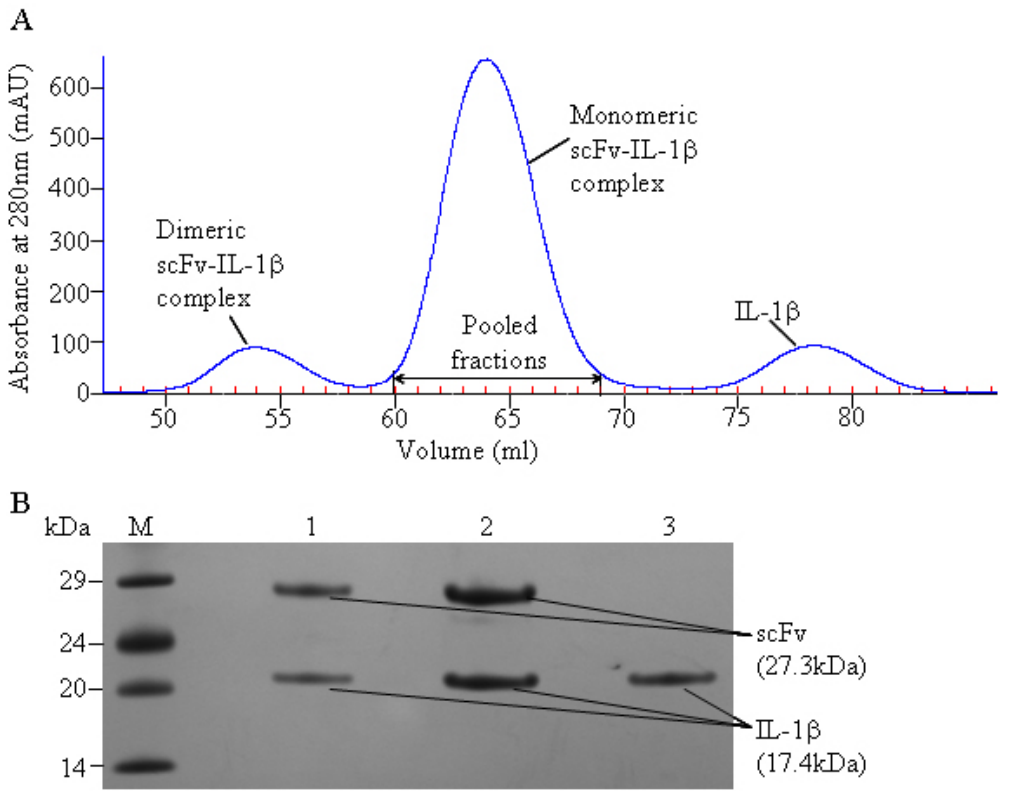

Figure 2.19 Panel A shows a typical gel filtration purification chromatogram for the scFv IC8-IL-1β complex, with resolved peaks for the monomeric and dimeric IC8-IL-1β complex as well as the 20% excess of free IL-1β added to form the complex. Panel B is a SDS-PAGE gel of samples from the three peaks in panel A. Lane M contains a 14-66 kDa molecular weight marker (Sigma). Lane 1 contains dimeric IC8-IL-1β complex, lane 2 contains monomeric IC8-IL-1β complex and lane 3 contains free IL-1β.

2.3.9 NMR spectroscopy

$^{15}N/^{1}H$ HSQC spectra were collected from ^{15}N labelled samples of both free scFv 8516 and IC8 (figure 2.20), with the knowledge that these would consist of an approximate 1:1 ratio of monomer:dimer. Spectra were collected at a range of temperatures between 35 and 50 ^{o}C and the backbone amide proton linewidths of a consistent selection of peaks were measured. An optimum temperature of 40 ^{o}C was selected as a compromise between proton linewidth and protein precipitation. $^{15}N/^{1}H$ HSQC spectra obtained from samples of the scFvs containing nearly 50% domain-swapped dimer were surprisingly good (figure 2.20A), which in part reflects the conservation of the V_H-V_L domain interface in the dimer and results in only a few signals being shifted on dimer formation. In contrast, dimer formation has a marked influence on the signal linewidth resulting in significantly reduced sensitivity and resolution, with average backbone amide proton linewidths of 31.5 ± 5.5 Hz compared to <25 Hz expected for a monomeric 27 kDa protein. For the spectrum shown in panel A of figure 2.20 approximately 75% of the possible backbone amide signals are visible, with the remaining signals possibly missing due to overlap.

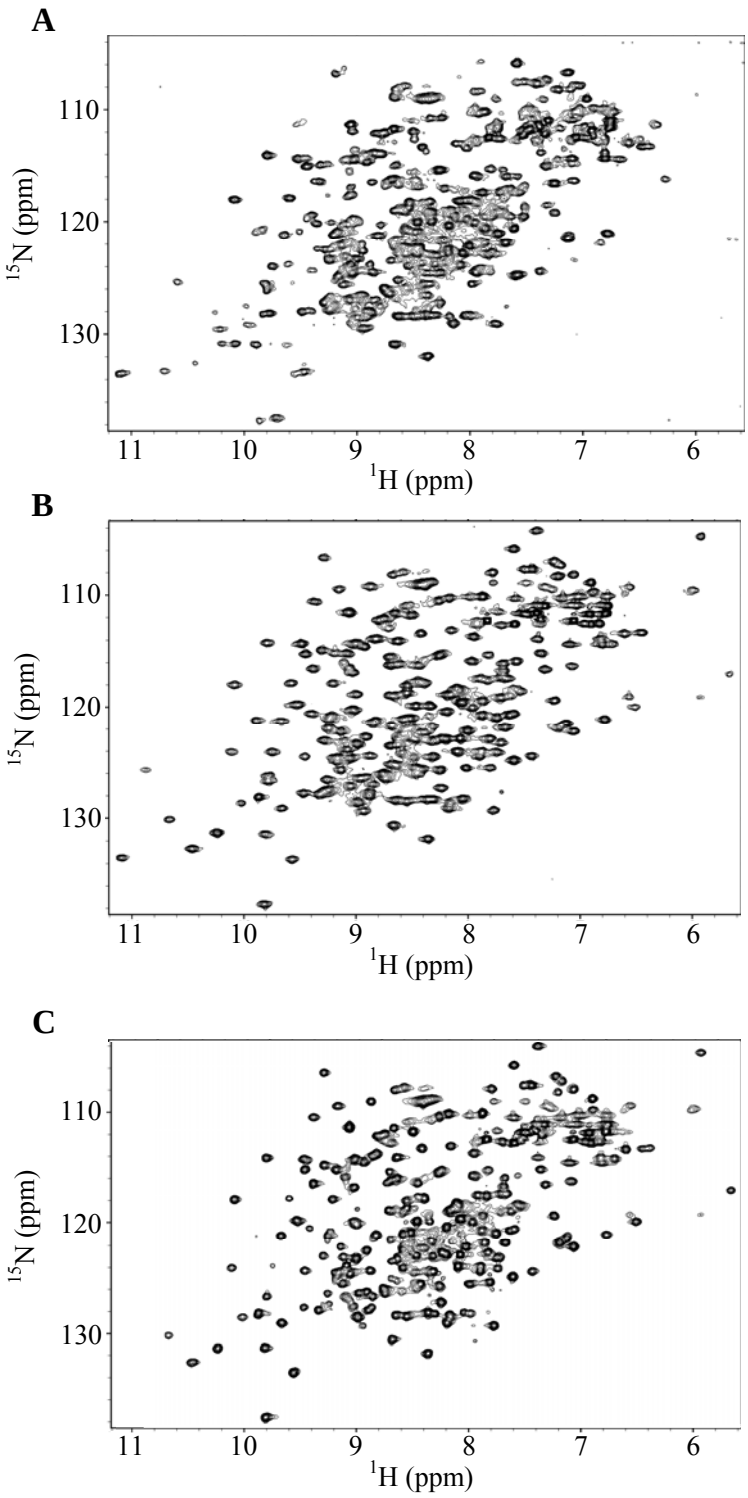

Figure 2.20 ^{15}N/^{1}H HSQC spectra of a scFv in the free and bound state. Panel A is a spectrum of free ^{15}N labelled scFv IC8 (400 μM). Panels B and C are spectra of ^{15}N IC8-unlabelled IL-1β (650 μM) and ^{15}N/^{2}H IC8-unlabelled IL-1β complex (300 μM) respectively. The improved backbone amide proton linewidths and resolution in the

complex due to removal of the domain-swapped dimer equilibrium present in the free scFv samples can clearly be seen by comparison of panels A and B, with deuteration reducing the backbone amide proton linewidths still further as evident in panel C.

Preventing the formation of the dimer by engineering new scFv constructs had proved impossible, as shown previously. However, it was reasoned that formation of monomeric scFv IC8-IL-1β complex at low concentration may stabilise the bound scFv and allow subsequent concentration of the monomeric complex for detailed NMR analysis without formation of the domain-swapped scFv dimers. This approach proved very successful and enabled preparation of 200-700 μM samples of scFv-IL-1β complexes that contained no detectable dimer. $^{15}N/^{1}H$ HSQC spectra of these samples showed significantly reduced backbone amide proton linewidths (27.5 ± 3.0 Hz) compared to the scFv alone, resulting in increased resolution and sensitivity as illustrated in figure 2.20, with approximately 85% of the backbone amide peaks visible.

From the experience of previous structural biology projects undertaken within the laboratory it is known that the collection of good quality triple resonance data and the subsequent determination of the protein or protein-protein complex structure becomes difficult once the backbone amide proton linewidths exceed approximately 25 Hz. Therefore, to obtain backbone amide proton linewidths below this limit it was decided that deuteration of the protein samples would be necessary. Table 2.2 shows a comparison of the backbone amide proton linewidths for the free scFv and scFv-IL-1β samples, together with measurements made for other protein and protein-protein complexes studied within the group. The backbone amide proton linewidths for the scFv 8516 are broader than would be expected for a 27 kDa monomeric protein, as shown by

comparison with the backbone amide proton linewidth values for the 24 kDa RV0288-RV0287 complex and 33 kDa Pdcd4 2MA-3 protein. Despite the increase in molecular weight both the protonated scFv-IL-1β complexes show a reduction in the backbone amide proton linewidths compared to the free scFv to values closer to those expected for a complex of that size. Deuteration of the scFv IC8-IL-1β complex reduces the average backbone amide proton linewidth to a similar value to that observed for the 20 kDa ESAT-6/CFP-10 complex, thus reaching a level that was deemed amenable to the collection of high quality NMR data.

Table 2.2 A table of backbone amide proton linewidth measurements from $^{15}N/^{1}H$ HSQC spectra for eight different free protein or protein-protein complexes studied within the laboratory. The structures of these proteins and complexes have either already been solved or are in the process of being determined. The free scFv 8516 sample has the broadest of all linewidths despite being the fourth smallest protein by molecular weight in the table. This can be attributed to the formation of domain-swapped dimer. Large reductions in the linewidths of the complex can be obtained by perdeuteration, as shown in the final column for the triple labelled scFv IC8-IL-1β complex.

Protein	Pdcd4 MA-3C	ESAT-6/CFP-10	Rv0288/Rv0287	Pdcd4 2MA-3	8516	8516/IL-1β	IC8/IL-1β	IC8/IL-1β
Mass (Da)	15487	20564	24495	33002	27291	44680	44753	48137
No. of residues	136	195	233	298	255	408	407	407
Labelling	^{15}N	^{15}N	$^{15}N/^{13}C$	^{15}N	^{15}N	^{15}N	^{15}N	$^{15}N/^{13}C/^{2}H$
Temp	25°C	35°C	35°C	25°C	40°C	40°C	40°C	40°C
	17.2	18.9	22.1	28.4	21.6	28.6	22.6	19.0
	15	19.65	22.5	28.7	25.3	29.2	24.4	16.6
	21.9	15.39	21.4	26.4	37.3	23.8	28.0	16.0
	15.6	19.57	29.8	25.1	41.6	31.6	26.8	16.0
	14.1	17.44	36.6	25.4	33.0	24.4	31.0	16.0
	15	16.15	29.2	27.1	31.5	29.8	30.4	17.2
	15.7	15.73	28.8	29.5	28.7	29.2	28.6	22.0
	17.5	18.36	16.2	22.7	28.4			
	18.2	15.02	22.4	21.1	31.2			
	18.1	15.39	23.7	24	27.3			
Mean (Hz)	**16.8**	**17.2**	**25.3**	**25.8**	**31.3**	**28.1**	**27.4**	**17.5**
S.D.	2.3	1.8	5.8	2.7	5.8	2.9	3.1	2.2

2.4 Discussion

The results presented here clearly show that two of the four selected scFvs can be efficiently expressed and purified following the protocol described in section 2.2.6, leading to expression levels in LB broth of approximately 5 mg of purified protein per original litre of culture. ScFv 168 showed lower expression levels on western blot analysis and was mainly expressed as insoluble protein even at low temperatures and IPTG concentrations, while scFv 120B09 was not expressed at all (figure 2.7). The apparent lack of expression of scFv 120B09 could be due to toxicity of the gene, mRNA instability, mRNA secondary structure or protein instability but each of these possibilities seems unlikely given the sequence similarity of the four scFvs. Insoluble expression is most likely due to a slow folding rate for the protein, with the *E. coli* diverting the material into inclusion bodies.

The expression yields of antibody fragments reported in the literature varies markedly (Bedzyk *et al.*, 1990; Carter *et al.*, 1992; Denzin *et al.*, 1991; Worn & Pluckthun, 1998; Worn & Pluckthun, 1999) but a direct comparison of functional yields is difficult due to the use of many different vector systems, strains, experimental conditions and types of antibody fragment. To the best of my knowledge there has been only one detailed investigation into the differential periplasmic expression yields of antibody fragments (Knappik & Pluckthun, 1995). Knappik and Pluckthun chose a model system consisting of the murine anti-phosphorylcholine antibody McPC603 (Satow *et al.*, 1986) and the anti-HER2 antibody hu4D5 (Carter *et al.*, 1992). Despite sharing a high degree of sequence homology, approximately 75% for each variable domain, the functional yields of the antibody fragments differed greatly. In an identical *E. coli* host and vector system there was an enormous increase in the amount of folded periplasmic hu4D5 compared

to McPC603. This observation was independent of whether Fv, scFv or Fab fragments were studied. It was also noted that cells expressing McPC603 became leaky and eventually lysed with large parts of the expressed protein being insoluble aggregate. These effects were absent for cells expressing the hu4D5 fragments, leading to the conclusion that the primary sequence must be responsible for the differences in both expression yield and cell stability during expression. Domain switching experiments showed that the leakiness phenotype and the folding problems were associated with the V_H domain. Eight mutations in the McPC603 V_H domain were analysed, with the mutations selected based on a comparison of the framework sequences of the McPC603 and hu4D5 V_H sequences. The leakiness was attributed to one residue in FR H2 and the production of insoluble protein to two consecutive residues in FR H3, as shown in figure 2.21. Substitution of the three McPC603 residues for the equivalent hu4D5 residues resulted in the production of large amounts of soluble antibody fragment in the periplasm with no leakiness or cell death. These three residues are all located in turns between β-sheets. The simplified conclusions from this study based on only two antibody sequences is that Pro at site 1 and Ser and Ala at sites 2 and 3 are unfavourable for cell integrity and folding respectively, while Ala at site 1 and Ala and Asp at sites 2 and 3 are favourable (figure 2.21).

Further investigation by Knappik and Pluckthun into the basis of the observed differences in expression yields showed that the mutations discussed above caused no significant increase in the thermodynamic stability of the intact antibody fragment. However, the kinetics of a slow folding step, measured by urea-induced fluorescence changes, differed dramatically between the wild type and triple mutant McPC603 fragments. The wild type fragments did not fold during the time of measurement but

rather aggregated very rapidly, whereas the mutant fragment interconverted slowly into a native-like structure. It was therefore concluded that a folding intermediate differs between the wild type and mutant protein in its tendency to aggregate. Based on the observation that the favourable Asp residue in hu4D5 is exposed (Eigenbrot *et al.*, 1993) and the main chain of McPC603 and hu4D5 superimpose almost perfectly in this region, the authors suggest that the different aggregation behaviour may be due to a greater ordering of water around the Ala in McPC603 compared to the equivalent Asp in hu4D5. Therefore, less entropy of activation is required to overcome the barrier for forming the crucial transition state with increased hydrophobic surface, which leads to aggregation. This single amino acid dependency of folding efficiency has also been shown for other proteins (Deng *et al.*, 1994; Jappelli *et al.*, 1992; Mitraki & King, 1992), with these residues also lying in turns between β-sheets. This theory may, therefore, reflect the periplasmic folding of recombinant proteins in general.

The multiple sequence alignment of the four anti-IL-1β scFvs studied in this chapter with the hu4D5 and McPC603 V_H sequences, shown in figure 2.21, shows that all of the anti-IL-1β sequences contain the favourable Ala residue at site 1, thought to prevent a leaky phenotype. This is reflected in the fact that no scFv leakage from the periplasm or cell death during expression has been noted during the expression trials for the four anti-IL-1β scFvs. There is much greater sequence variability at sites 2 and 3 making comparison of the expression yields and association with residue type at these positions difficult. From the antibody sequences analysed it is difficult to decipher a pattern for expression based on the two residues at these locations. This may reflect the lack of sequences and expression yields analysed or may indicate that the soluble periplasmic

expression of antibody fragments is also determined by other residues in the V_H sequence or other domains.

Soluble expression has been achieved for all four of the equivalent anti-IL-1β Fab fragments at UCB-Celltech. This suggests that the sequence of the V_H domain is not the only contributing factor to whether an antibody fragment folds correctly in the periplasm, although Fab 120B09 was very poorly expressed and Fab 168 only moderately expressed, which is a similar pattern to that seen for the scFvs. The Fab fragment is a complete unit within the antibody, separated from the rest of the molecule by a flexible hinge region with no further protein-protein contacts. However, by generating a scFv fragment the protein domains are taken out of their natural biological context by removing the variable domains from the constant domains. The residues at the former variable-constant domain interface, which are usually buried, remain exposed to the solvent and form an extended hydrophobic patch. These hydrophobic patches may influence the *in vivo* folding pathway, perhaps promoting the aggregation of folding intermediates. Mutation of some of these residues to hydrophilic residues has led to some success (Nieba *et al.*, 1997) but it appears that different antibodies may have different problems and so solutions need to be tailor-made for a given antibody.

Based on previous studies at UCB-Celltech and reports in the literature it was known that purification of inclusion bodies and refolding of scFvs can lead to low protein yields (Worn & Pluckthun, 1999), which is not ideal for the amounts of protein required for NMR studies. Therefore, attempts to work with scFv 168, as well as scFv 120B09 which was not expressed in any of the trials, were not pursued.

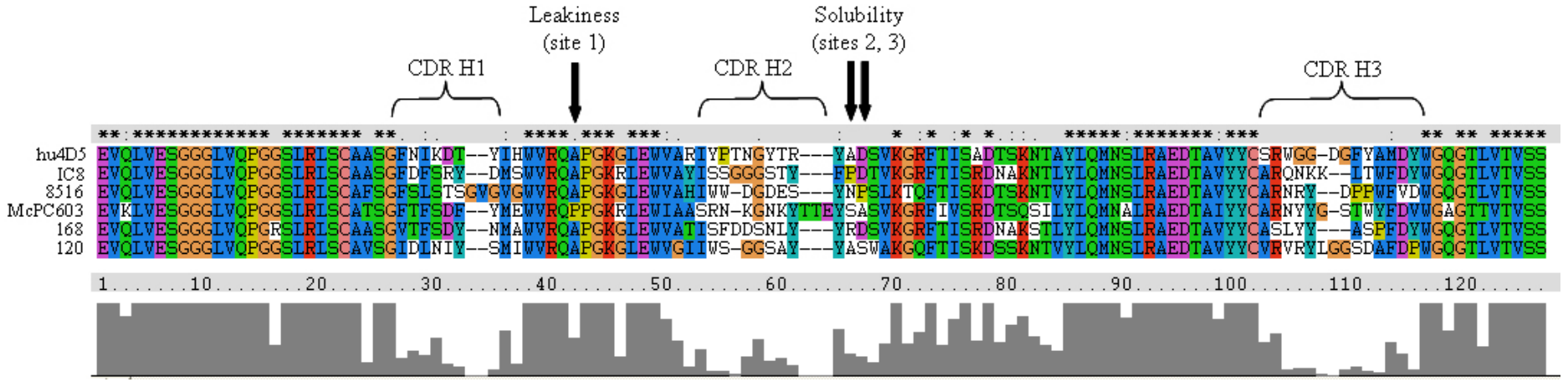

Figure 2.21 Multiple sequence alignment of the V_H domains of the four anti-IL-1β scFvs, IC8, 8516, 168 and 120B09, and two other antibodies hu4D5, which has a favourable sequence for soluble periplasmic expression, and McPC603, which has an unfavourable sequence for soluble periplasmic expression (Knappik & Pluckthun, 1995). The sequences are ordered such that the top three express soluble protein and the bottom three express insoluble protein or no protein at all in the case of 120B09. Knappik and Pluckthun used back engineering from sequence comparisons to identify the three highlighted residues that are responsible for the leakiness and insoluble protein aggregation shown by McPC603 antibody fragments but not hu4D5 fragments.

Initial gel filtration purification of the two soluble scFvs, 8516 and IC8, revealed the presence of multimers. Confirmation that these were due to hydrophobic interactions rather than disulphide cross-linkage is shown by the non-reducing SDS-PAGE gel in panel B of figure 2.10, where the fractions from the two gel filtration chromatogram peaks both run at the same molecular weight of approximately 28 kDa. Calibration of the Superdex 16/60 gel filtration column (figure 2.16) enabled estimation of the molecular weights for each peak and thus determination of the composition of the sample. Table 2.3 shows the estimated molecular weights from gel filtration and the actual molecular weights based on protein sequence, thus confirming that the scFv multimers peaks corresponded to monomer, dimer and trimer.

Table 2.3 A table comparing the molecular weights (MW) of the scFv as determined experimentally by gel filtration and by computational analysis of the known protein sequence.

	Experimental MW (kDa)	**Actual MW (kDa)**
Monomer	24	27.4
Dimer	51	54.8
Trimer	76	82.2

A model explaining the occurrence of scFv multimers, initially proposed in 1998 (Arndt *et al.*, 1998), showing the dimeric form of the scFv as a product of domain-swapping (Bennett *et al.*, 1994; Bennett *et al.*, 1995) is shown in figure 2.22. As discussed in detail in section 2.1.3, the position and rate of the monomer-dimer equilibrium has been reported to be affected by pH, scFv orientation, protein concentration, linker length and antibody type. All four anti-IL-1β scFv constructs had been engineered in the orientation V_L-linker-V_H, which gives the shortest distance for the linker to span (Huston *et al.*, 1991) reducing the strain and causing preferential formation of the

monomer. Analysis of scFvs 8516 and IC8 samples at pHs between 6.0 and 8.0 showed no significant influence of pH on the position of the equilibrium. This suggests that pH dependence of the equilibrium is antibody specific as proposed previously (Arndt *et al.*, 1998).

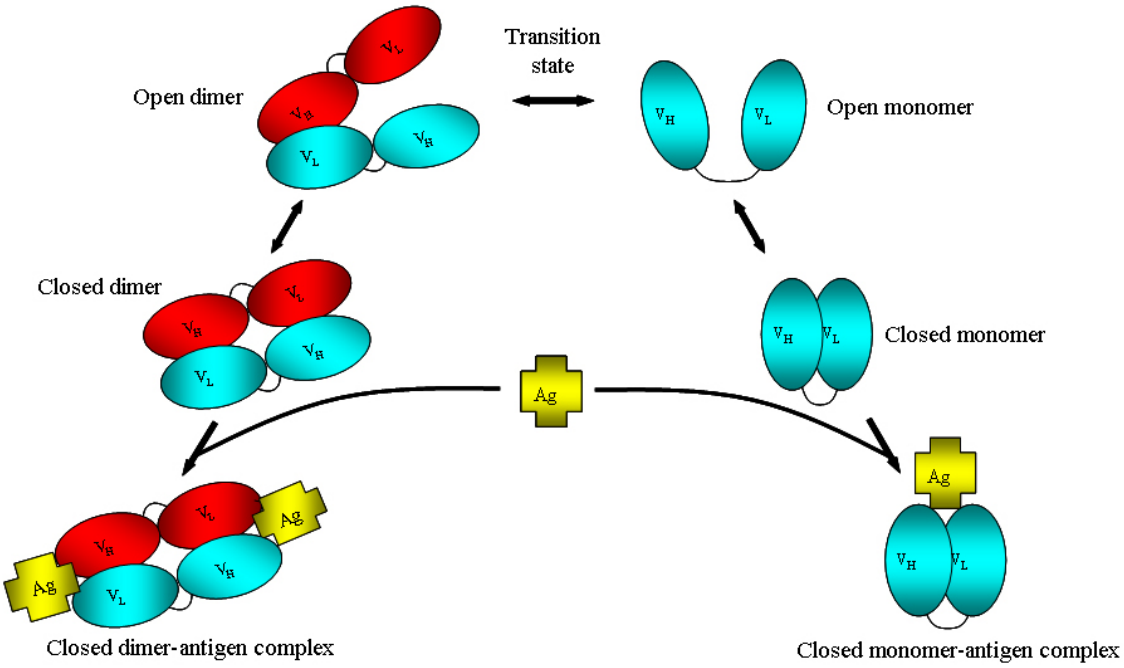

Figure 2.22 A schematic representation of the scFv monomer to domain-swapped dimer equilibrium. Free scFv can exchange between the closed and open forms, enabling a second open monomer (shown in red) to associate with it and form a functional domain-swapped dimer. Tight binding of antigen (Ag) prevents the separation of the V_L and V_H domains resulting in the formation of stable monomeric or dimeric scFv-antigen complexes.

Of all the studies on the effects of dimerisation of scFvs only one (Lee *et al.*, 2002) has involved analyses of scFv samples at concentrations of up to 300 μM. These results are very similar to those observed for scFv 8516, shown in figure 2.17, with a dimer concentration of approximately 40% for the anti-tumour scFv MFE-23 at a protein

concentration of 8 mg/ml (~300 μM). The interface between the V_L and V_H domains is not very stable in the free scFv, with a dissociation constant of 10 μM estimated for one particular anti-phosphocholine scFv McPC603 (Glockshuber *et al.*, 1990). Thus, the open form of the monomer should be present to a significant amount and will undergo an intramolecular association to form a monomer or an intermolecular association to produce dimer. As the protein concentration increases the likelihood of the open form of the monomeric scFv associating with another open monomer to form the domain-swapped dimer increases, thus resulting in the observed increase in dimer with concentration.

It has been estimated that the concentration of a scFv in the periplasm is approximately 500 μM (13 mg/ml). This is based on the assumption that 5 mg of functional protein are expressed in a 2 litre culture of *E. coli* at OD_{600} of 2.0, with the cell treated as a capped cylinder 1 μM in diameter and 2 μM in length and the periplasmic space occupying 10% of the total cell volume (Arndt *et al.*, 1998). If this is indeed the case then approximately 50% of the scFv in the periplasm will be dimer. During the purification process the concentration of the sample is reduced to <20 μM ensuring that the sample consists of mainly monomer upon initial gel filtration purification. This high concentration within the cell could also account for the insoluble expression of some scFvs if they have a tendency to form multimers at slightly lower concentrations. It has been observed that concentrated samples (>100 μM) of the scFvs investigated undergo precipitation over long periods of time, presumably due to the formation of higher multimeric species that do not occur at the lower concentrations (<20 μM) at which the scFv samples are relatively stable. A similar occurrence could be happening in the cell causing the insoluble expression of some scFvs.

The wild type scFvs were each engineered with a 20 residue Gly_4Ser repeat linker under the assumption that this would be long enough to allow monomer formation, as has been reported on many occasions (Holliger *et al.*, 1993; Huston *et al.*, 1988; Todorovska *et al.*, 2001). From the work reported here it was clear that the constructs with a 20 residue linker did not form 100% monomer, so to ascertain whether this was due to a limitation in the linker length longer linker constructs containing a 25 residue Gly_4Ser repeat linker were cloned. Analysis of the monomer-dimer equilibrium by analytical gel filtration for the purified scFv 8516-X showed an identical ratio of monomer to dimer as seen for the wild type scFv 8516. This suggests that increasing the linker length to 25 residues does not further stabilize the scFv and prevent the formation of the dimer. It can, therefore, be assumed that the original 20 residue linker is not responsible for the occurrence of the dimer in agreement with the previous studies stating that a linker length of 15 residues and above is adequate for correct formation of the monomeric form of the scFv.

It is hypothesised that the domain-swapped dimer forms due to the exposure of a hydrophobic patch when the variable domains of the scFv dissociate producing the open monomer, enabling the domain-swapped dimer to form by hydrophobic interaction between two open monomers. In an attempt to destabilise the interaction that forms the dimer a specific Ala to Asp mutation was engineered into framework region 3 of the V_H domain, as described in section 2.2.4.3. Mutant versions of each scFv were successfully cloned and confirmed by DNA sequencing and expression of the protein was shown to be similar to the wild type (figure 2.7). Analysis of the equilibrium at high concentration for scFv 8516-M (figure 2.18) and IC8-M (data not shown), showed that the rate of formation of the dimer was slowed down with the percentage of dimer

stabilising over approximately 72 hours to similar levels recorded for the wild type scFvs.

Detailed structural studies of a protein that consists of a monomer-dimer equilibrium at the concentrations required for NMR is not ideal. However, the suitability of free scFvs for study by NMR was assessed by preparing ^{15}N labelled samples for the collection of $^{15}N/^{1}H$ HSQC spectra. The signal dispersion of the spectra obtained was excellent, as shown in figure 2.20. Backbone amide proton linewidths were also measured to assess the quality of the NMR data by measuring the full width at half the height of the peak. The resonance linewidth is inversely proportional to the relaxation rate constant. Therefore, the faster the signal decays the broader the linewidth. The relaxation rate increases with protein size due to the longer tumbling time of larger systems, and so in general larger proteins give broader linewidths. Observed linewidths significantly broader than expected for a monomeric 27 kDa protein imply that aggregation due to the scFv monomer-dimer equilibrium is increasing the apparent rotational correlation time or that chemical exchange effects contribute to the inhomogeneous linewidth. As is shown visually in figure 2.20 and numerically in table 2.2 the resolution of the $^{15}N/^{1}H$ HSQC spectra of the scFvs 8516 and IC8 is worse than would be expected for a monomeric 27 kDa protein. This observation must be due to presence of between 40 and 50% dimer in the free scFv samples, along with small amounts of higher multimeric species, which will increase the overall correlation time and produce broader signals. From previous structural studies in the lab it was known that it was difficult to collect sufficient reliable NMR data from protein samples that gave average backbone amide proton linewidths of over approximately 25 Hz. Therefore, based on the quality of the

scFv data it was concluded that determination of a high resolution structure for a free scFv by NMR would not be viable.

The model of the scFv domain-swapped dimer equilibrium (figure 2.22) shows that the monomer or dimer may be trapped in either state by addition of antigen. The presence of the antigen locks the scFv into either the monomeric or dimeric form by preventing the dissociation of the variable domains. The dissociation constant of affinity matured antibodies with their target proteins is so low (< nM) that the dissociation of the antigen from the scFv and formation of dimer is highly unfavourable. Therefore, purification of monomeric scFv-target protein complex enabled monomeric scFv to be analysed without the presence of multimeric species but as part of a larger complex.

Samples of both 8516-IL-1β and IC8-IL-1β were successfully purified and separated from the dimeric complex as shown by the typical chromatogram in figure 2.19. These samples were stable for prolonged periods of time and could be concentrated to high concentrations without any of the precipitation problems that had occurred with the free scFv samples. Analysis of table 2.2 shows that the backbone amide proton linewidths for ^{15}N labelled scFv-unlabelled IL-1β complexes were narrower than those for the free scFv 8516. Despite the increase in molecular weight from the 27 kDa free scFv to a 44 kDa scFv-IL-1β complex, which would normally be associated with a broadening of linewidths, the $^{15}N/^{1}H$ HSQC data for the scFv-IL-1β complex showed greater resolution, as shown by the comparison in figure 2.20, with narrower proton linewidths. This is due to the prevention of the formation of a domain-swapped dimer, which as described earlier causes a broadening of the proton linewidths and some heterogeneity.

Based on the backbone amide proton linewidth information in table 2.2 for those proteins and protein-protein complexes whose structures have been solved recently in our laboratory, it was decided that high quality NMR data leading to the determination of a reliable high resolution structure could be obtained for protein samples that gave backbone amide proton linewidths of less than approximately 25 Hz. The measurements for the scFv-IL-1β complexes were above this limit and so perdeuteration of the protein samples was attempted. A deuteron has a 6.5 fold lower gyromagnetic ratio than a proton, so replacing all the non-exchangable protons with deuterons causes a significant increase the transverse relaxation time thus reducing the signal linewidths (Clore & Gronenborn, 1998).

A perdeuteration strategy for the scFvs was successfully optimized producing approximately 2.5 mg of purified scFv per original litre of 100% deuterated minimal media. The resolution of the $^{15}N/^{1}H$ HSQC spectrum of $^{15}N/^{2}H$ labeled IC8-unlabelled IL-1β was much improved with the average proton linewidth of 17.5 ±2.2 Hz being similar to those obtained for the 20 kDa *mycobacterium tuberculosis* complex ESAT-6/CFP-10. Therefore, it was concluded from this data that determination of a high resolution structure of a scFv using NMR would be possible if the structure of the scFv-target protein complex was determined rather than the free scFv and that perdeuteration of the proteins would be necessary.

2.5 Conclusion

The work reported in this chapter has described the construction of scFv expression vectors for four separate anti-IL-1β antibodies. Expression trials confirmed the soluble expression of two of the scFvs, 8516 and IC8, under the optimized conditions, with scFv 168 being insoluble and scFv 120 was not expressed at all.

Purification of the soluble scFvs, giving purified yields of 4-5 mg of protein per original litre of LB culture, revealed the presence of a monomer to domain-swapped dimer equilibrium. Further analysis of this equilibrium showed a protein concentration dependency, with the amount of dimer increasing from approximately 20% at low concentration (<10 μM) to almost 50% at concentrations required for NMR (>250 μM). Disruption of the domain-swapped dimer formation by engineering of scFv constructs with either an extended linker or a specific Ala to Asp mutation proved unsuccessful.

Assessment of the suitability of scFv samples, consisting of an approximate 1:1 ratio of monomer to dimer at high concentrations, for study by NMR showed that signal dispersion in $^{15}N/^{1}H$ HSQC spectra was very good but the sensitivity and resolution as measured by backbone amide proton linewidths was poor, probably due to the presence of multimeric species.

Formation of scFv-IL-1β complex effectively locked together the variable domains of the scFv, preventing the formation of the open form of the scFv thus disrupting the monomer-dimer equilibrium. This allowed purification and concentration of 200-700 μM samples of scFv-IL1β complex that contained no detectable dimer. $^{15}N/^{1}H$ HSQC spectra of these samples showed significantly reduced backbone amide proton

linewidths compared to the free scFv despite the increase in molecular weight. Deuteration of the scFv-IL-1β complex provided a further reduction in backbone amide linewidths to a point at which a study of the high resolution structure of the scFv-IL-1β complex became feasible.

Chapter 3

Solution structure of the scFv IC8-IL-1β complex

3.1 Introduction

3.1.1 Protein-protein docking

Protein-protein interactions play a central role in many aspects of biochemistry, such as protein localisation, competitive inhibition, gene regulation and signal transduction, while the disruption of protein-protein interactions can cause biochemical diseases related to their function. Knowledge of the structure of a protein-protein complex can give an insight to its function as well as providing information that can be used in the engineering of the interface or drugs targeted to it in order to manipulate properties such as binding specificity, affinity or kinetics.

The number of expected biomolecular complexes is predicted to exceed the number of proteins in the proteome by at least an order of magnitude (Bonvin, 2006), with a significant number of these likely to be extremely difficult to study by the traditional structural methods of NMR spectroscopy and X-ray crystallography. There is also a shortfall in the PDB with the majority of structures being for single proteins, most of which are involved in one or more protein-protein interactions but the structures of these complexes have not been determined. This deficiency can be reduced by the use of computational methods such as docking to predict the structure of a complex based on the structures of its known constituents.

The underlying theory behind all docking predictions is that the structure of the complex is the lowest free energy state. Therefore, docking programs need to give a score or energy function that can discriminate correctly or nearly-correctly docked orientations from incorrect ones and have a search or sampling method that can find a correctly docked orientation. Since the 1970s there has been a long progression of computerised approaches using increasingly efficient sampling methods and increasingly accurate energy functions.

The earliest docking studies used low resolution models and sampling over the orientation of one docking partner surface (Wodak & Janin, 1978). These approaches treated the proteins as rigid bodies in order to reduce the search space for optimal structures of the complexes. Ignoring flexibility in this way could prevent the algorithms from determining native associations, and so incorporating flexibility into the docking algorithms is essential for the accuracy of the docked complexes. This is particularly important when docked structures are being determined from homology models (Marti-Renom *et al.*, 2000). Allowing flexibility introduces a high number of degrees of freedom, which not only increases the running time but also results in a higher rate of false positives. It is thus essential to score these correctly in order to identify near-native results (Andrusier *et al.*, 2008).

The general scheme of flexible docking can be divided into multiple stages as depicted in figure 3.1. The structures of the two individual proteins, determined by structural techniques such as NMR or X-ray crystallography or produced by homology modeling, are treated as rigid bodies and the six-dimensional space of relative protein orientations (translational and rotational) is searched and a set of candidate structures identified

using a simple scoring function. These structures are then re-scored using a more expensive energy function that is better at discriminating near-native orientations. In the final stages movement of the sidechains and possibly the backbone is allowed followed by minimisation by a further energy function.

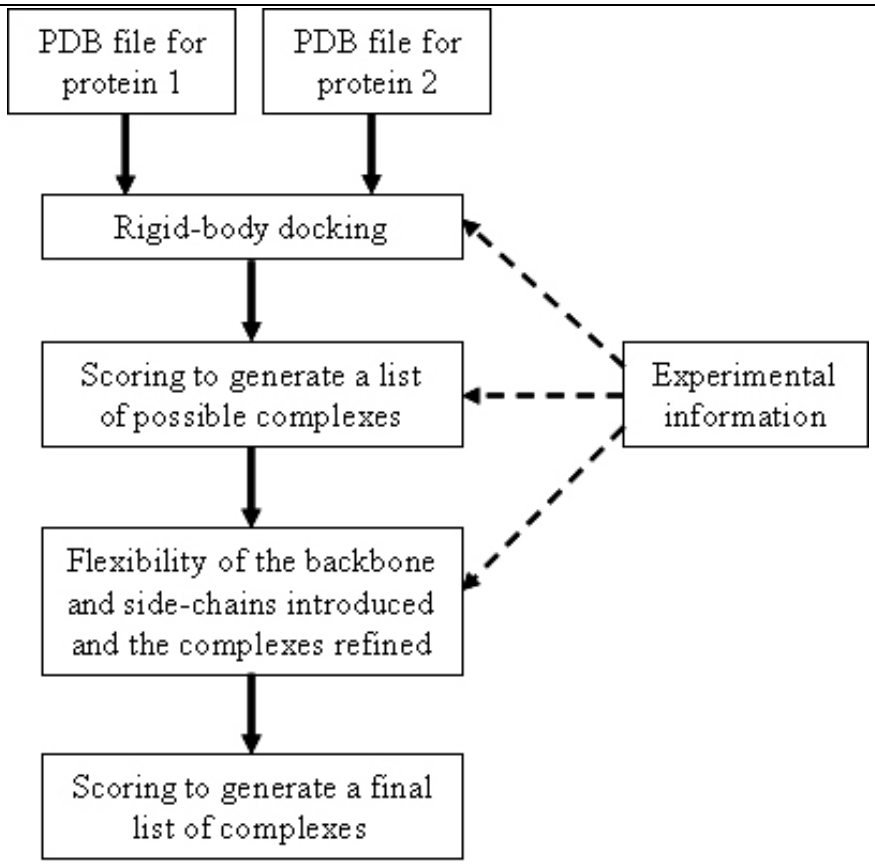

Figure 3.1 A general scheme for flexible protein-protein docking. The two input proteins are freely rotated in all orientations and docked by a rigid-body method. The energies of the resulting complexes are calculated and scored. A set number of near-native complexes are then refined at which point flexibility is introduced. The final set of docked complexes is then scored. Experimental data that provides information about the protein-protein interface, orientation of the two proteins with respect to each other or flexibility of regions within either protein can be input at several stages of the docking procedure.

More recently restraint-driven docking programs have been designed or adapted to incorporate biological information to constrain the structure of possible docked complexes. It is common to have experimental information about some of the residues that most probably lie at or near the protein-protein interface, determined for example by mutagenesis. It is also possible to extract information from the sequence alignment of homologous proteins as a cluster of well conserved residues, in particular polar residues, can often be used to identify a binding site (Aloy *et al.*, 2001; Hu *et al.*, 2000). This information can be used to define ambiguous interaction restraints (AIRs), which are treated by the docking algorithms as part of the protein-protein interface, thus narrowing down the search and increasing the likelihood of obtaining a near-native complex structure.

NMR spectroscopy is an especially useful tool to determine restraints for docking with many structures having been determined in this way (Clore, 2000; Morelli *et al.*, 2001). Mapping of the chemical shifts observed between free and bound spectra can be used to determine those residues that show a change in chemical environment and so are more likely to be at the protein-protein interface. Orientational information for the two proteins with respect to each other can be determined by measuring RDC values and intra and inter-molecular NOEs provide distance restraints within and between the two proteins.

The Rosetta suite of software (Gray *et al.*, 2003), including RosettaDock and RosettaNMR, is able to dock two known or predicted protein structures based on NMR derived restraints, with the fundamental assumption that the backbone conformation of the proteins does not change much upon complex formation. RosettaNMR utilises Cα

and Cβ chemical shifts to generate dihedral angles and utilises NOEs and RDCs as distance and orientation restraints respectively. These restraints are then utilised in the scoring of the docked complexes with violations increasing the scoring accordingly. Prior to docking the sidechains of the native proteins are removed and low resolution rigid body docking is performed with any available biochemical data used to localise the binding site to a small region on one or both proteins, such that only a small region of space is explored around the defined binding site. This generates 1000 docked structures to which sidechains are added using the Rosetta sidechain packing algorithm to prevent errors in docking due to irregularities (e.g. crystal contacts) in the native protein structures. A high resolution refinement is then performed by Monte Carlo simulated annealing and a list of the 10 lowest scoring docked complexes is created.

Xplor-NIH (Clore *et al.*, 1998; Schwieters *et al.*, 2003) is a package for biomolecular structure determination from experimental NMR data with known or predicted geometric data. This is achieved by searching for the minimum target function comprising terms for the experimental NMR restraints, covalent geometry and non-bonded contacts using a variety of minimisation procedures including molecular dynamics in Cartesian space and Monte Carlo minimisation methods. Using the Xplor software Marius Clore demonstrated the power of NMR restraint-driven docking to accurately determine the structure of complexes by using the EIN-HPr complex as a test case (Clore, 2000). The structure of the 40 kDa complex was determined by NMR over the course of approximately 2 years and 3500 hours of experimental time (Garrett *et al.*, 1999). The most time consuming part of the analysis was the assignment of the NOEs, the majority of which are intramolecular NOEs. Clore proposed that an approach to assign only the intermolecular NOEs would lead to a dramatic saving in time and effort

and that as high resolution structures of the free proteins are available and no significant changes in the backbone occur upon complex formation, it should be possible to dock the two proteins using intermolecular NOE and backbone amide RDC restraints to provide the necessary translational and orientational information. He showed that accurate docking can be achieved by rigid body minimisation using a target function that consists of only intermolecular NOE and backbone amide RDC restraints and an intermolecular van der Waals repulsive potential, obtaining a docked EIN-HPr complex with a RMSD of only 1.17 Å from the original NMR structure for the complex.

One of the most proven programs for restraint-driven protein-protein docking and the one which will be used in this chapter is HADDOCK (high ambiguity driven protein-protein docking) (Dominguez *et al.*, 2003). It was one of the first programs to be specifically designed to incorporate experimental data into the docking of proteins and is one of the few algorithms to allow flexibility of the protein backbone as well as the sidechains (Bonvin, 2006). The docking protocol consists of: randomisation of orientations and rigid body energy minimisation; semi-rigid simulated annealing in torsion angle space; final refinement in Cartesian space in an explicit water shell. Typically, 1000 complex conformations are calculated in the first stage and the best 200 in terms of intermolecular energies are then refined. During the simulated annealing steps backbone and sidechains at the interface are allowed to move to allow for some conformational changes. The final structures are clustered using the pairwise backbone RMSD at the interface and the resulting clusters are ranked based on their average interaction energies and buried surface area. Any form of biochemical or biophysical information that provides information about residues at the interface or distance and orientation restraints can be used to restrain the docking protocol at any stage. The

HADDOCK authors also tested the accuracy of their docking approach on the EIN-HPr complex but with the use of only NMR chemical shift perturbation data to define AIRs at the interface. From the structures of the free proteins a docked complex with a RMSD of 1.70 Å from the NMR structure for the complex was produced. To date 54 novel protein-protein complex structures have been determined using HADDOCK along with over 30 protein-peptide, ligand, DNA or RNA complexes. Work by Jain and co-workers using HADDOCK to dock the 9 kDa ACP protein to the 86 kDa homotrimeric protein LpxA using chemical shift perturbations, backbone amide RDCs and mutagenesis data highlights how such techniques can extend the size range of NMR spectroscopy (Jain *et al.*, 2004). The crystal structure of *E. coli* LpxA and the crystal and NMR structures of *E. coli* ACP had been solved previously but attempts to determine the structure of the complex had failed. Perdeuteration of ACP allowed for collection of high quality backbone amide RDC data by measuring the difference in peak splitting in a pair of spectra for the isotropic and phage aligned phases, and determination of the binding site on ACP by chemical shift perturbations upon complex formation. The binding epitope on LpxA was defined from mutagenesis data and RDCs were back-calculated from the LpxA crystal structure using the program MODULE (Dosset *et al.*, 2001). Docking with HADDOCK resulted in a cluster of structures with an overall backbone RMSD of 1.8 Å and no AIR or RDC restraint violations. The structure is consistent with previous mutagenesis and chemical modification experiments and explains the significance of conserved residues that are involved in salt-bridges at the interface.

The performance of current docking methods is monitored by a blind docking competition known as CAPRI (Critical Assessment of Predicted Interactions) (http://capri.ebi.ac.uk). The CAPRI results indicate that excellent predictions can be

obtained on the whole for targets that only show small backbone conformational changes. For targets that undergo a conformation change upon binding (even as small as 2 Å) accurate docking becomes much more challenging (Mendez *et al.*, 2005). It can be concluded from the CAPRI results observed to date that the best targets for docking are single-domain small proteins with known monomer structures, micromolar or better binding affinity, and minimal backbone conformational change upon binding, with experimental data greatly improving the likelihood of success. The docking programs detailed above have performed consistently well in all of the CAPRI rounds that they have entered.

As discussed previously in chapters 1 and 2, the determination of high resolution structures for scFvs has been limited due to the formation of domain-swapped dimers. The formation of dimer can be prevented by binding to antigen, allowing purification of monomeric scFv-target protein complex. The work reported in this chapter describes the collection, assignment and analysis of NMR data for the scFv IC8-IL-1β complex. Essentially complete backbone resonance assignments (NH, N, Cα, Cβ and C') allowed analysis of the chemical shifts perturbed upon formation of the scFv IC8-IL-1β complex, thus mapping the binding sites on each protein. This information as well as backbone amide RDCs and long-range NH to NH NOEs has enabled a homology model for the scFv IC8 to be refined and docked to the structure of IL-1β using HADDOCK. $^{15}N/^{1}H$ HSQC data for scFv and Fab IC8-IL-1β complexes has shown that scFvs bind in an almost identical manner to the equivalent Fab, making them ideal models. This, therefore, clearly demonstrates that NMR restraint-driven docking can be successfully used to determine the structure of a scFv-target protein complex, which provides a means for structure based rational design of therapeutic antibodies.

3.2 Materials and methods

3.2.1 NMR spectroscopy

NMR spectra were acquired from 0.35 ml samples of 0.2-0.7 mM scFv IC8-IL-1β complex in a 25 mM sodium phosphate, 100 mM sodium chloride, 10 μM EDTA, 100 μM AEBSF, 0.02% (w/v) sodium azide buffer at pH 6.5, containing 10% D_2O/90% H_2O. All NMR data were acquired at 35 or 40 ^oC on a 600 MHz Bruker DRX or 800 MHz Bruker Avance system. The 2D and TROSY-based 3D spectra (Pervushin *et al.*, 1997) recorded to obtain sequence specific backbone assignments for the scFv IC8-IL-1β complex were: $^{15}N/^{1}H$ HSQC (Bodenhausen & Ruben, 1980), $^{15}N/^{13}C/^{1}H$ HNCACB (Wittekind & Mueller, 1993), HN(CO)CACB (Grzesiek & Bax, 1992), HNCA, HN(CO)CA, and HNCO (Kay *et al.*, 1990). Typical acquisition times for 3D experiments were 8-9 ms in F_1 (^{13}C) (except for the HNCO, which was 24 ms), 21-24 ms in F_2 (^{15}N) and 70 ms in F_3 (^{1}H). The 3D spectra were collected for between 64 and 80 hours and the $^{15}N/^{1}H$ HSQC spectra for between 0.5 and 2 hours, with acquisition times of 60 ms in F_2 (^{1}H) and 50 ms in F_1 (^{15}N). Long range NOE data were obtained using $^{15}N/^{1}H$ NOESY-HSQC (Marion *et al.*, 1989) experiments with a mixing time of 450 ms and collected over approximately 90 hours.

RDC data were collected on an 800 MHz Bruker Avance system by the addition of 4 mg/ml Pf1 phage (Asla Biotech) to 0.35 ml samples of 0.2-0.3 mM scFv IC8-IL-1β complex in a 25 mM sodium phosphate, 100 mM sodium chloride, 10 μM EDTA, 100 μM AEBSF, 0.02% sodium azide buffer at pH 6.5, containing 10% D_2O/90% H_2O. A stock of 50 mg/ml Pf1 phage was prepared by dialysis into 25 mM sodium phosphate, 100 mM sodium chloride, 2 mM magnesium chloride, 0.02% sodium azide buffer at pH

6.5. Backbone amide RDC values were derived from the differences between the ^{15}N-^{1}H scalar couplings for isotropic and partially aligned samples using $^{15}N/^{1}H$ HSQC and TROSY spectra (Kontaxis *et al.*, 2000), with acquisition times of 60 ms in F_2 (^{1}H) and 50 ms in F_1 (^{15}N) and collected for approximately 10 hours.

The WATERGATE method was used to suppress the water signal when required (Piotto *et al.*, 1992). All NMR data was processed using Topspin (Bruker Biospin Ltd.) with linear prediction used to extend the effective acquisition times by up to 2 fold in ^{15}N. The spectra were analysed using the Sparky package (Goddard & Kneller).

3.2.2 Sequence specific assignments

Sequence specific backbone resonance assignments (N, NH, Cα, Cβ, C') were obtained for the IC8-IL-1β complex from the identification of intra and inter-residue connectivities in $^{15}N/^{13}C/^{1}H$ HNCACB, HN(CO)CACB, HNCA, HN(CO)CA and HNCO spectra. Intra and inter-residue amide NH to Cα and Cβ peaks were identified in the triple resonance spectra and used to search for signals from neighbouring residues in the sequence. Where possible, signals were confirmed by the observation of NH to NH NOEs in the $^{15}N/^{1}H$ NOESY-HSQC spectra. The CSI (Wishart & Sykes, 1994) and TALOS (Cornilescu *et al.*, 1999) programs were used to determine the positions of elements of regular secondary structure from the chemical shift data.

3.2.3 Chemical shift mapping of the scFv IC8-IL-1β protein-protein interface

The minimal shift approach (Veverka *et al.*, 2008; Waters *et al.*, 2007; Williamson *et al.*, 1997) was used to estimate the changes in signal positions of the scFv IC8 upon

binding to IL-1β. The minimum change for all available backbone amide peaks between the free and bound IC8 was calculated using Microsoft Excel to calculate the combined chemical shift difference in ^{15}N and ^{1}H for each assigned peak in the $^{15}N/^{1}H$ HSQC spectrum of the bound IC8 to all observed peaks in the $^{15}N/^{1}H$ HSQC spectrum of the free scFv IC8. The combined amide proton and nitrogen chemical shift difference (Δδ) was defined by $\Delta\delta = \sqrt{(\Delta\delta_{HN})^2 + (\Delta\delta_N \cdot \alpha_N)^2}$ where $\Delta\delta_{HN}$ and $\Delta\delta_N$ correspond to the ^{1}H and ^{15}N chemical shifts between pairs of compared $^{15}N/^{1}H$ HSQC peaks and α_N is a scaling factor of 0.2 to account for the differences in range of amide proton and nitrogen chemical shifts. For each $^{15}N/^{1}H$ HSQC peak, the minimal shift upon binding was taken as the lowest possible combined shift (Δδ). Backbone amide resonance assignments were determined for free and complexed IL-1β, so the actual chemical shift could be determined for each residue. The combined chemical shift difference (Δδ) was determined in the same was as above for each residue that had been assigned in both the free and bound IL-1β $^{15}N/^{1}H$ HSQC spectra.

3.2.4 Homology modelling

A homology model for the scFv IC8 was produced by Jiye Shi from UCB-Celltech. Antibody Modelling Tool (Riley & Shi, 2005) was used to build a homology model of the Fv portion of the scFv. This tool performed two BLAST (Altschul *et al.*, 1997) searches against the Protein Data Bank (Berman *et al.*, 2000), one with the V_H domain of the scFv and the other with the V_L domain. Antibody structures that have E-values lower than 1^{-10} in both searches were retained as potential structural templates and aligned to the Fv portion of the scFv using FUGUE (Shi *et al.*, 2001). Each of the pairwise alignments were scored based on sequence similarity within the framework

region, the conservation of solvent inaccessible residues, the similarity in CDR length, and the resolution of the template structure. The alignment with the best score was used as the input to MODELLER (Sali & Blundell, 1993) to build a homology model of the Fv region.

The model was then subjected to manual refinement. HARMONY3 (Shi, 2001) was used to highlight potentially problematic regions within the model. All the potential structural templates were examined for local similarity to the scFv within those regions. A refined model was produced by forcing MODELLER to use the most similar fragments for modelling the problematic regions. A flexible linker was then added to convert the Fv model to a scFv model.

The PDB structure 1L7I (Vajdos *et al.*, 2002) was selected by the Antibody Model Tool as the best structural template for the modelling of the scFv. The initial model was analyzed by HARMONY3, and CDR H2 was identified as a potentially problematic region. Following manual examination of other structural templates, the CDR H2 of 1L7I was replaced by that of 1KNO (Charbonnier *et al.*, 1995) during another round of the modelling process. The resulting model of the Fv region was converted to a scFv by adding the flexible linker.

In addition to the generation of a 'good' scFv IC8 model described above, a 'poor' model was also produced for the docking analysis. This was essential produced by purposely misaligning the β-sheets of the framework and inserting structures for the CDR loops that did not match the predicted canonical structures for the scFv IC8 primary sequence.

3.2.5 Docking of the scFv IC8-IL-1β complex

The structure of the scFv IC8-IL-1β complex was determined by NMR restraint driven docking (chemical shift changes, NOEs and RDCs) of IL-1β and the scFv using HADDOCK (Dominguez *et al.*, 2003). A homology model of the scFv IC8 produced as described in section 3.2.4 and a crystal structure of free IL-1β (PDB code 2I1B (Priestle *et al.*, 1989)) were used as starting points for the docking calculations. To dock the scFv IC8 and IL-1β AIRs were selected to define the protein-protein interaction surface. AIRs are defined as being either active or passive. Active residues are ones that have been experimentally identified as being involved in the interaction and are also solvent exposed, with the passive residues being all the solvent accessible surface neighbours of active residues. Analysis of the chemical shift perturbation data and solvent accessibilities using NACCESS (Hubbard & Thornton, 1993) resulted in the identification of active residues as scFv residues with a minimal shift greater than 0.075 ppm and 20% solvent accessibility, and IL-1β residues with a combined shift of greater than 0.1 ppm and 20% solvent accessibility. For the scFv 17 active and 25 passive residues were selected and for IL-1β 26 active and 19 passive residues were identified, as shown below.

IC8 active: G27, N31, Y32, W92, S93, R160, D162, G183, G184, G185, S186, T187, Y188, D191, G195, K230, K231

IC8 passive: I2, R24, T25, S26, N28, I29, H30, Y49, N50, S67, L94, D157, F158, S159, Y161, S182, F189, P190, T192, V193, K194, R201, N213, R227, L232

IL-1β active: R4, S5, L6, R11, S13, Q15, H30, Q39, V41, F46, G49, E50, E51, S52, D54, K55, I56, E105, N107, N108, K109, T147, Q149, V151, S152, S153

IL-1β passive: A1, P2, V3, D12, Q14, M36, E37, Q38, V40, V47, Q48, N53, K63, K93, I106, D145, F146, M148, F150

Residues at the protein-protein interface or in loops that are thought to be highly flexible can be defined as semi-flexible or fully flexible for the docking. The side-chain and backbone atoms of semi-flexible residues are allowed to move during the semi-flexible refinement stage while the fully flexible residues are allowed to move at all stages of iteration 1 (see definition below). The semi-flexible interface on each protein was either left as the default setting (±2 residues from each AIR) or manually defined as listed below to include the scFv IC8-IL-1β protein-protein interface and the V_H-V_L interface within the scFv. In addition the scFv linker (residues 110-129) was defined as being fully flexible.

IL-1β semi-flexible regions: 1-17, 30-58, 103-111, 143-153

IC8 semi-flexible regions : 23-37, 43-53, 65-69, 86-101, 155-167, 172-197, 223-241

The axial tensor (Da) and rhombicity (R) components of the alignment tensor for partially aligned samples of IL-1β were calculated using PALES (Zweckstetter & Bax, 2000) and together with the backbone amide RDC data were used to incorporate restraints defining the orientation of the scFv and IL-1β. Intervector projection angle

restraints (Meiler *et al.*, 2000) were also derived from the RDC data. Long-range interresidue NOEs were also included as additional restraints during the docking calculations. The NOEs were calibrated on the basis of peak intensity and determined to correspond to a ^{1}H-^{1}H distance restraint of <5 Å, 5-6.5 Å or 6.5-8 Å. The calibration was performed by determining average i to i+1 and i to i+2 peak intensities and using these values as thresholds for the <5 Å and 5-6.5 Å groups respectively.

The docking protocol consists of three stages: randomisation of the orientations of the two binding proteins and energy minimisation of each (iteration 0); semi-rigid-body docking of the proteins using a simulated annealing protocol (iteration 1); refinement of the complexes obtained by molecular dynamics in an explicit water shell (iteration 2). In the first stage, the proteins were placed about 25 Å apart, randomly rotated with respect to each other, energy minimised and from 1000 starting combinations the lowest 200 energy solutions were selected for further refinement. The next stage involved three rounds of simulated annealing-based docking: rigid body simulated annealing; semi-flexible simulated annealing with flexible side-chains at the interface; semi-flexible simulated annealing with fully flexible interface (both backbone and side-chains). In the final step the structures obtained were refined by molecular dynamics in an 8 Å water shell. The AIR and NOE restraints were active during all iterations. The direct use of RDC values as a tensor was only active during iteration 2, while the indirect use of RDC values as intervector projection angles was active during iterations 0 and 1. In most cases the recommended clustering cut-off of 7.5 Å grouped all 200 structures into one cluster with an overall backbone RMSD of less then 4 Å. This value was used to recluster the structures into a greater number (typically about eight) of more well defined clusters.

To check the scFv IC8-IL-1β complex obtained, backbone amide RDCs calculated from the complex structures were compared with the experimentally determined RDCs using PALES. This allows the calculation of a Cornilescu quality factor (Q), which evaluates the agreement between the structure and the observed RDC values, whereby $Q = rms(RDC_{meas} - RDC_{pred}) / rms(RDC_{meas})$ (Cornilescu *et al.*, 1998).

Analysis of the clusters of structures obtained was carried out using HADDOCK (Dominguez *et al.*, 2003) and MolMol (Koradi *et al.*, 1996). All figures in this thesis showing protein structures were created in either MolMol or PyMol (DeLano, 2002).

3.2.6 Robustness of the docking procedure

The robustness of the HADDOCK docking procedure and in particular its reliance on the completeness of the backbone amide RDC data was evaluated by repeating the docking calculations with subsets of the RDC data randomly removed. Residues for which RDC data had been obtained were randomly selected using a random number generator plugin (AbleBits) for Excel and 10, 30, 50 or 70% of the RDC values removed. The RDC measurements from the scFv IC8 and IL-1β were treated separately and the random selection was repeated five times at each percentage. The docking was then performed as outlined in section 3.2.5.

The impact of the accuracy of the initial input structure on the docking protocol for the scFv IC8-IL-1β complex was also investigated by comparing the output structures obtained using either a good or poor scFv IC8 model as the input structure. The docking protocol was then run as outlined in section 3.2.5.

3.3 Results

3.3.1 Sequence specific assignments

As discussed in chapter 2, samples of free scFv approached 50% domain-swapped dimer at concentrations above 200 μM, producing broad backbone amide proton linewidth values and reduced sensitivity and resolution in $^{15}N/^{1}H$ HSQC spectra. Therefore, it was not possible to obtain sequence specific assignments for the free scFv. It was reasoned that the formation of monomeric scFv-IL-1β complex at low concentrations would stabilise the bound scFv and allow subsequent concentration of the complex for detailed NMR analysis without formation of domain-swapped scFv dimers. This approach proved very successful and enabled the preparation of 200-700 μM samples of scFv-IL-1β complexes that contained no detectable dimer, with reduced backbone amide proton linewidths in $^{15}N/^{1}H$ HSQC spectra especially upon perdeuteration of the proteins.

This strategy gave rise to well resolved spectra, illustrated by the $^{15}N/^{1}H$ HSQC spectra shown in figures 3.2 and 3.3, which allowed comprehensive sequence specific assignments to be made for the scFv IC8-IL-1β complex. For the scFv IC8 essentially complete backbone (NH, N, Cα, Cβ and C') assignments were obtained for all 228 residues excluding the linker and C-terminal His-tag apart from D1, P8, S9, N28, P40, G41, K42, A43, Y87, C88, Q89, R108, T109, D157, V177, Q211 and C225 (92%), and for IL-1β all 154 residues except A1, Q34, N53, K63, E64, K65, N66, D75, G139 and Q141 (93%). A representative set of ^{15}N strips from the HNCACB spectra of the scFv IC8-IL-1β complex used to assign the backbone atoms (NH, N, Cα and Cβ) are shown in figures 3.4 and 3.5. $^{15}N/^{1}H$ NOESY-HSQC spectra of the scFv IC8-IL-1β complex

were used to assign long-range NH to NH NOEs, as shown by the representative set of ^{15}N strips in figure 3.6. For the scFv IC8 387 NOEs were assigned (119 sequential (i,i + 1), 102 medium (i,i ≤ 4) and 166 long range (i,i ≥ 5)), and for IL-1β assignments were made for 240 NOEs (78 sequential (i,i + 1), 58 medium (i,i ≤ 4) and 104 long range (i,i ≥ 5)).

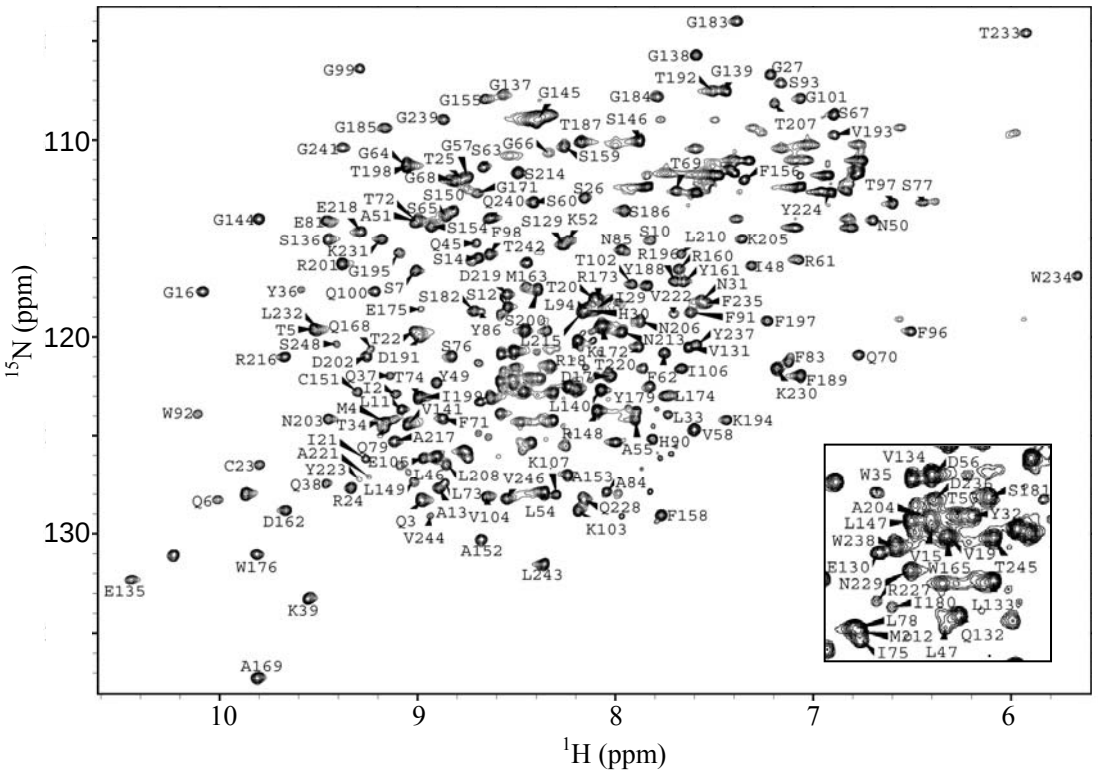

Figure 3.2 A $^{15}N/^{1}H$ HSQC spectrum of $^{15}N/^{13}C/^{2}H$ labelled scFv IC8-unlabelled IL-1β complex. The assignment of backbone amide signals for the scFv are indicated by residue type and number. The insert shows an enlarged image of the crowded central region of the spectrum.

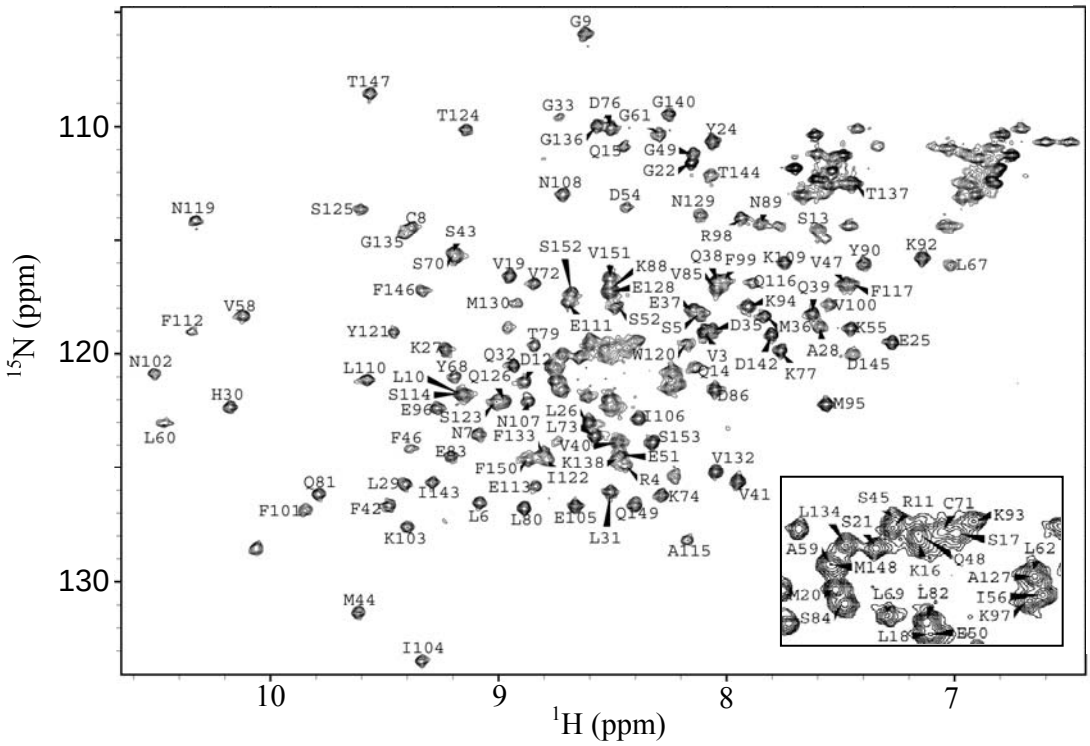

Figure 3.3 A ^{15}N/^{1}H HSQC spectrum of ^{15}N/^{13}C/^{2}H labelled IL-1β-unlabelled scFv IC8 complex. The assignment of backbone amide signals for IL-1β are indicated by residue number and type. The insert shows an enlarged image of the crowded central region of the spectrum.

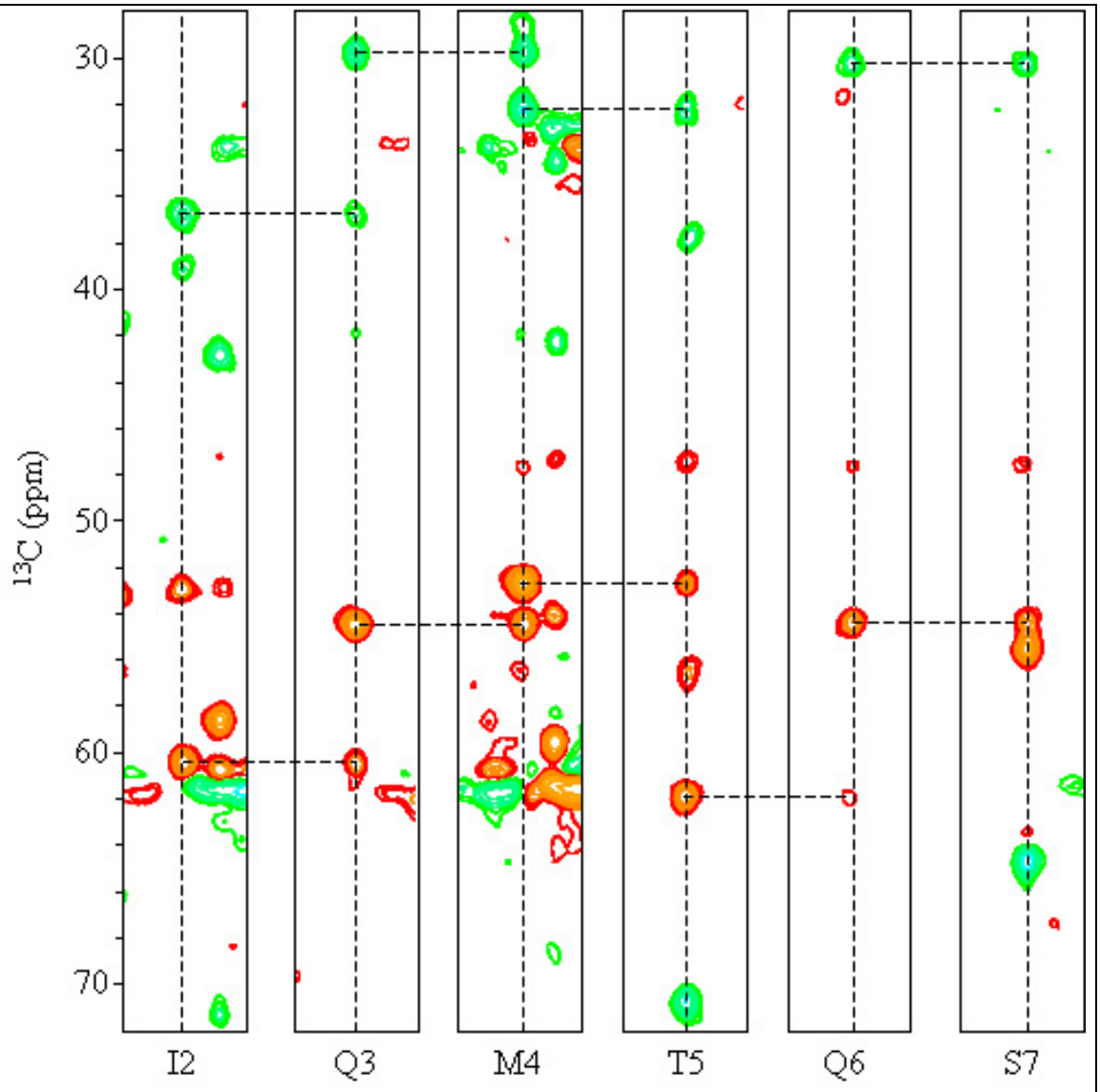

Figure 3.4 ^{15}N strips from the HNCACB spectrum of ^{15}N/^{13}C/^{2}H labelled scFv IC8-unlabelled IL-1β complex, showing the intra and inter-residue crosspeaks linking residues I2 to S7. Cα signals are show in red-yellow and Cβ signals in green. The horizontal dashed black lines indicate sequential Cα and Cβ connections.

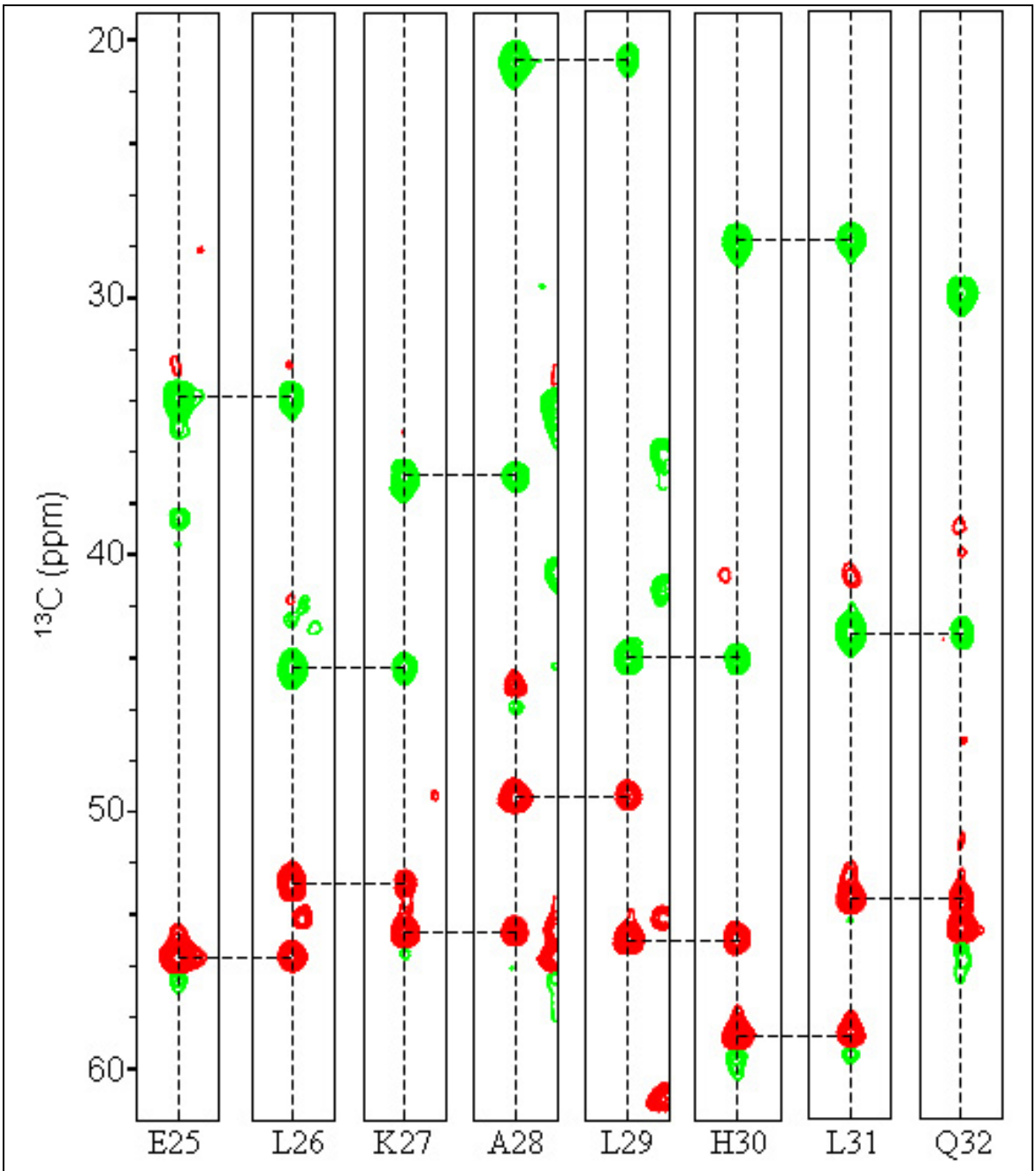

Figure 3.5 ^{15}N strips from the HNCACB spectrum of ^{15}N/^{13}C/^{2}H labelled IL-1β-unlabelled scFv IC8 complex, showing the intra and inter-residue crosspeaks linking residues E25 to Q32. Cα signals are shown in red and Cβ signals in green. The horizontal dashed black lines indicate sequential Cα and Cβ connections.

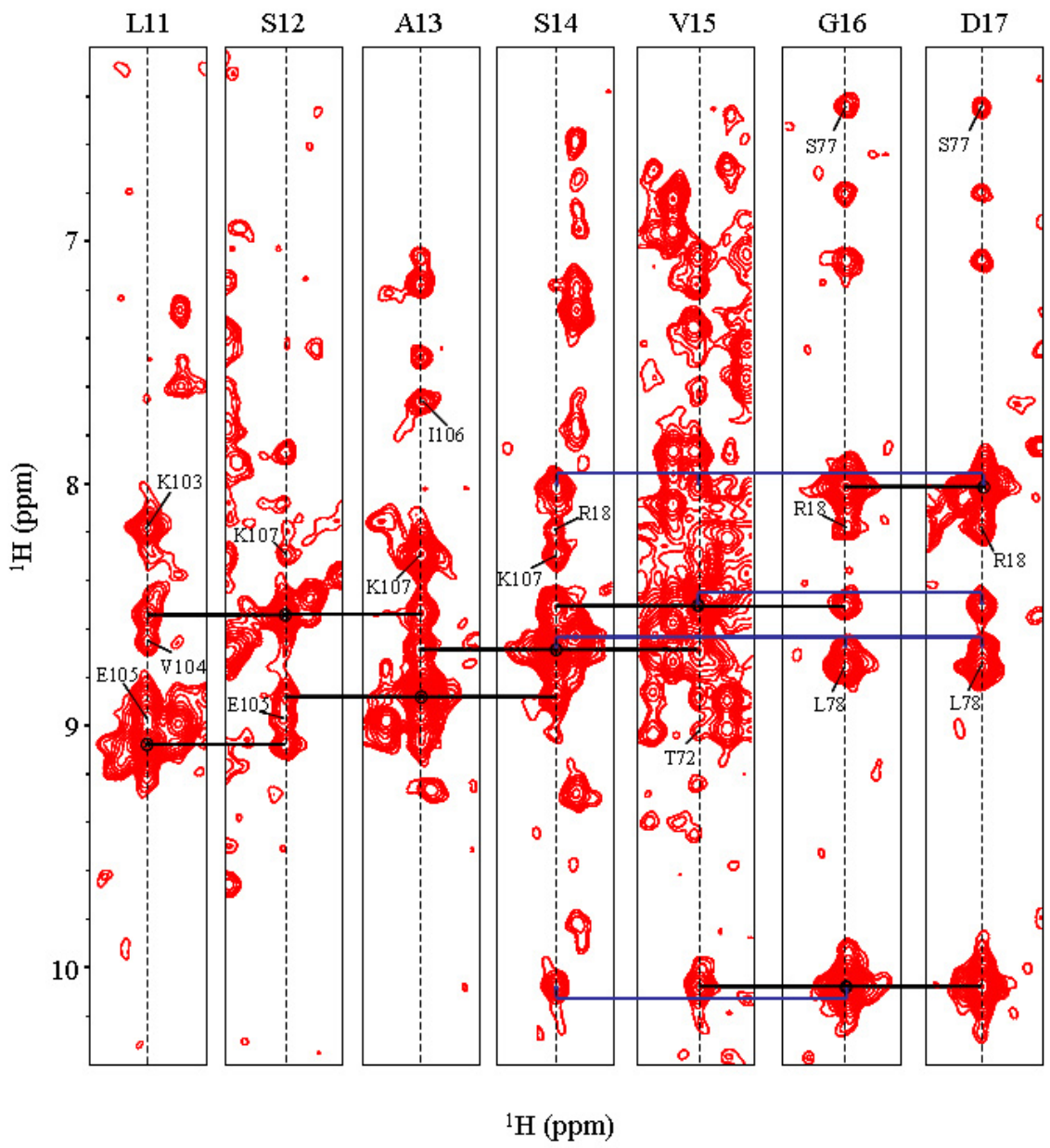

Figure 3.6 ^{15}N strips from the NOESY-HSQC spectrum of ^{15}N/^{2}H labelled scFv IC8-unlabelled IL-1β complex showing the NOE connectivities for residues L11 to D17. The thick black lines indicate sequential NH to NH NOEs and the blue lines indicate i to i+2 and i+3 NOEs. Additional long-range NOEs from residues further away in the primary sequence are labelled.

3.3.2 Chemical shift mapping

Residues involved in the interaction between the scFv IC8 and IL-1β were identified by the comparison of $^{15}N/^{1}H$ HSQC spectra acquired for the free and bound proteins, as described in section 3.2.3. It was not possible to obtain assignments for the free scFv and so significantly perturbed backbone amide signals were identified by minimal shift analysis. Assignments have been previously reported for free IL-1β (Driscoll *et al.*, 1990) and essentially complete backbone resonance assignments for free IL-1β under the experimental conditions used in my laboratory had been made by Vaclav Veverka prior to my work. It proved relatively straightforward to obtain nearly complete backbone resonance assignments for IL-1β bound to the scFv IC8, which allowed the actual combined shifts in amide ^{15}N and ^{1}H to be determined. The observed changes in backbone amide signals for both proteins on complex formation are summarised in the histograms shown in figures 3.7 and 3.8, which clearly indicate an extensive and specific interaction surface. The interaction surfaces are approximately equal size on both proteins (~3500 $Å^2$) with the signals from 29 scFv residues (R24, G27, N31, Y32, F91, W92, S93, F96, R160, Y161, D162, W176, G183, G184, G185, S186, T187, Y188, F189, D191, G195, T198, S200, N203, K230, K231, L232, T233 and F234) showing a significant shift at the interaction surface (figure 3.7C) and 27 peaks from IL-1β residues (R4, S5, L6, N7, T9, D12, Q15, S17, L18, Q39, V47, E50, E51, S52, D54, I56, V58, E105, I106, N107, K109, L110, T147, Q149, V151, S152 and S153) showing a significant shift at the interaction surface (figure 3.8C). For the scFv, residues with significantly perturbed backbone amides mainly lie in the CDRs, as expected for antibody binding, with shifts in signals from other nearby residues perhaps reflecting small conformational changes.

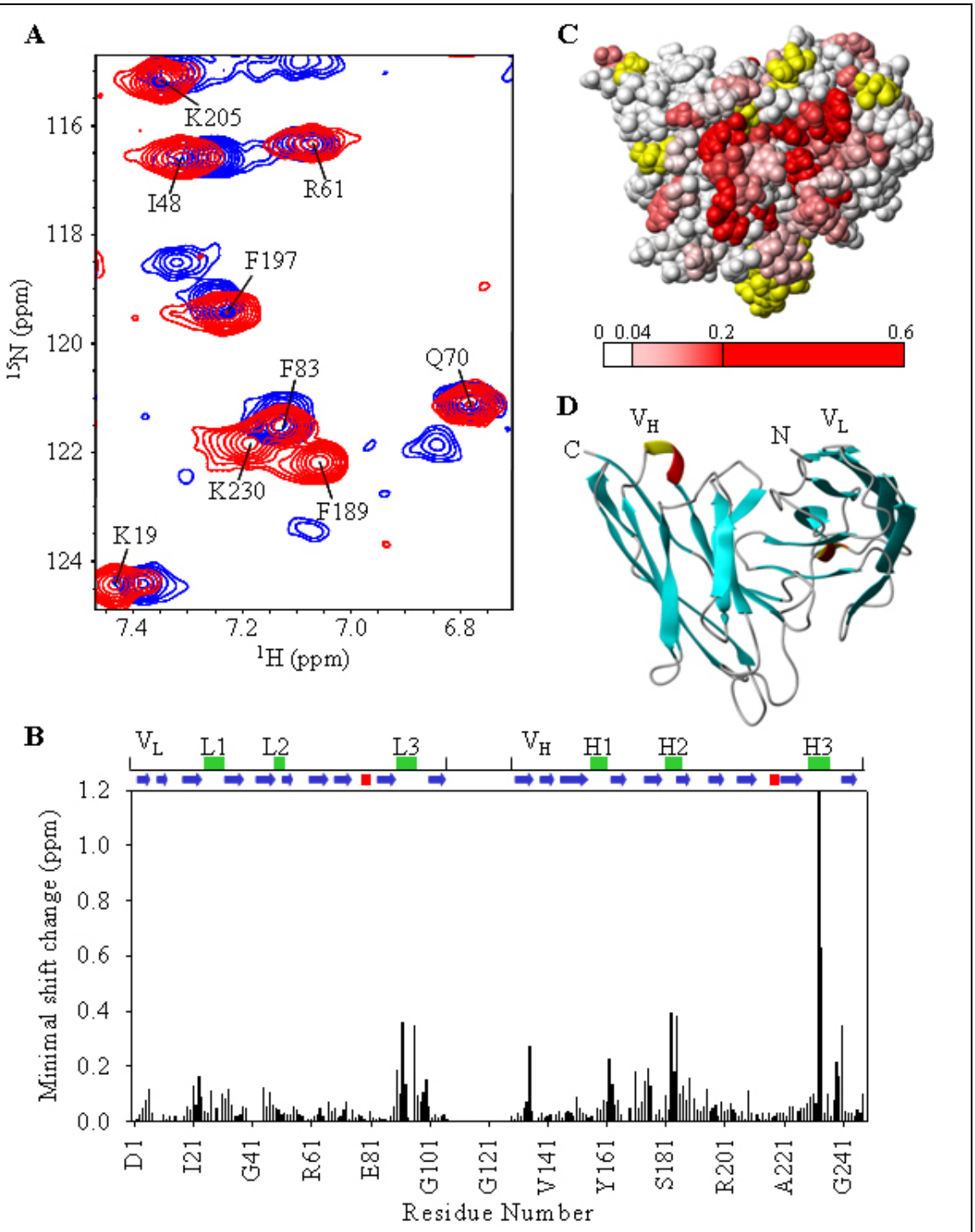

Figure 3.7 Mapping of the IL-1β binding site on the scFv by minimal shift analysis. Panel A shows a small region from the $^{15}N/^{1}H$ HSQC spectra for $^{15}N/^{2}H$ labelled scFv IC8-unlabelled IL-1β complex in red and free $^{15}N/^{2}H$ labelled scFv in blue, with the assignments for the complex indicated. It is clear that signals for R61 and Q70 show no significant shift, while the backbone amide peaks for F189 and K230 show substantial

changes. The histogram shown in panel B summarises the backbone amide minimal shifts seen for the scFv upon complex formation with IL-1β. The positions of the CDRs are indicated above the histogram by green bars, whilst regions of regular secondary structure are indicated by red bars for helices and blue arrows for β-sheets. The minimal shift data is mapped onto the space-filled view of the scFv shown in panel C with significantly perturbed residues (shift >0.04 ppm) coloured with a gradient from white to red. Residues shown in yellow correspond to ones for which no minimal shift data is available. Panel D shows a ribbon representation of the scFv structure in the same orientation as panel C.

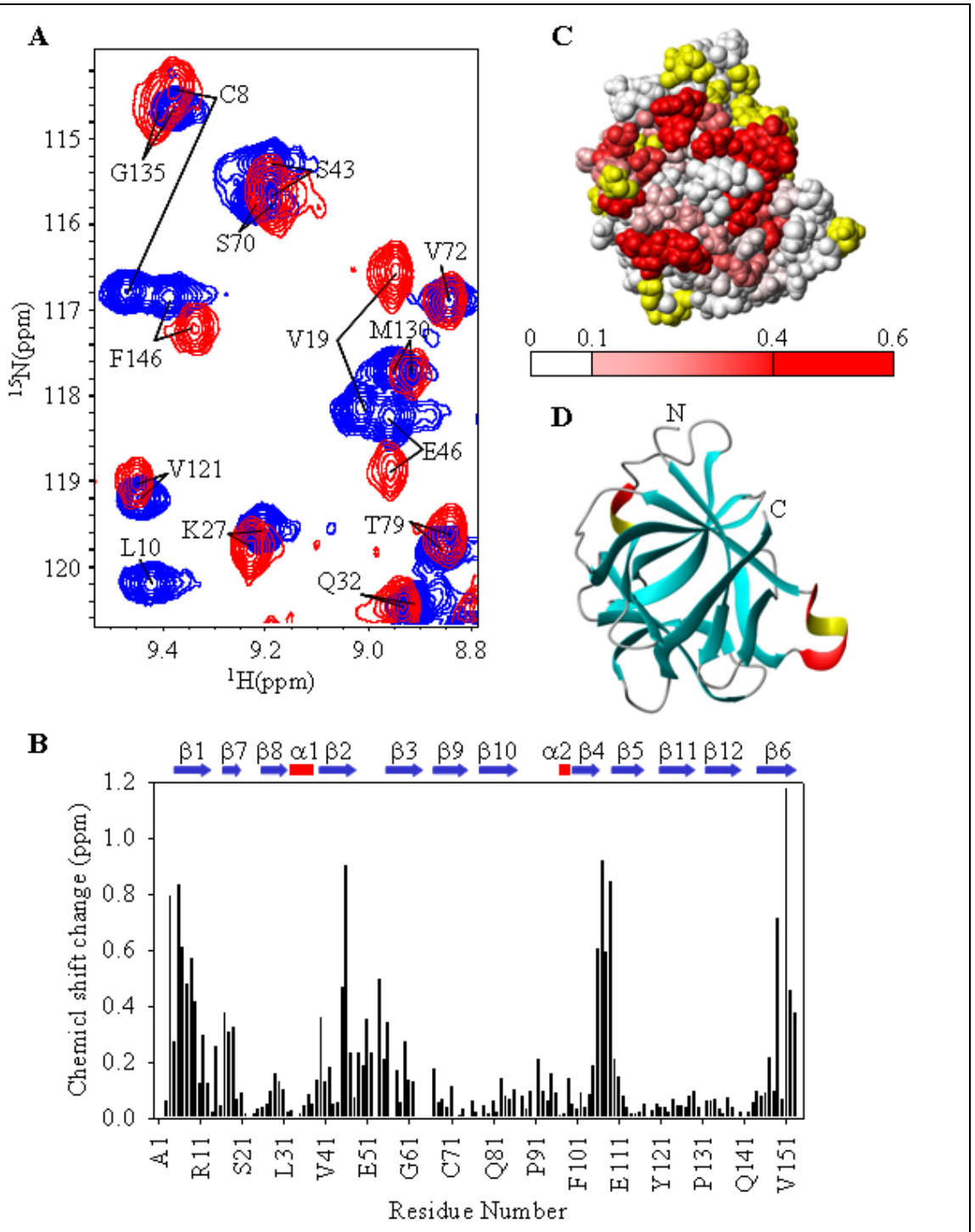

Figure 3.8 Mapping of the scFv binding site on IL-1β by combined amide backbone analysis. Panel A shows a small region from the $^{15}N/^{1}H$ HSQC spectra for $^{15}N/^{2}H$ labelled IL-1β-unlabelled scFv IC8 complex in red and free ^{15}N labelled IL-1β in blue, with assignments indicated for both. Signals from a number of residues, such as C8 and V19, clearly undergo significant shifts on complex formation, while others, such as

V72, remain unperturbed. The histogram shown in panel B summarises the combined backbone amide shifts seen for IL-1β on binding to the scFv IC8. Regions of regular secondary structure are indicated by red bars for helices and blue arrows for β-sheets. In panel C the shift data is mapped onto the space-filled view of IL-1β, with significantly perturbed residues (shift >0.1 ppm) coloured on a gradient from white to red. Residues highlighted in yellow are ones for which no chemical shift perturbation data was obtained. Panel D shows a ribbon representation of the IL-1β structure in the same orientation as panel C.

The chemical shift perturbation data for the scFv also revealed that the backbone amide signals of residues away from the IL-1β interaction site and located at the V_H-V_L interface were significantly perturbed upon formation of the scFv IC8-IL-1β complex. Ignoring the residues present at or very close to the IL-1β binding site, 10 residues from the V_L domain (T34, Q38, Q45, L46, A55, Q89, T97, F98, G99, Q100) and 14 residues from the V_H domain (V134, R160, Y161, G171, R173, L174, E175, W176, T233, W234, D236, W238, Q239, Q240) have peaks that clearly show a significant chemical shift at the V_H-V_L interface, as displayed in figure 3.9. The shifts seen in the V_L domain are on average much smaller, potentially due to these residues being located in crowded regions of the $^{15}N/^{1}H$ HSQC spectra for the free and bound scFv IC8, and also more data is missing for the V_L residues located at the V_H-V_L interface.

Thus this shows that movement of the V_H and V_L domains with respect to each other occurs upon binding of IL-1β, providing evidence of a potential B-cell membrane bound antibody signalling mechanism via a subtle structural rearrangement upon binding of the target protein.

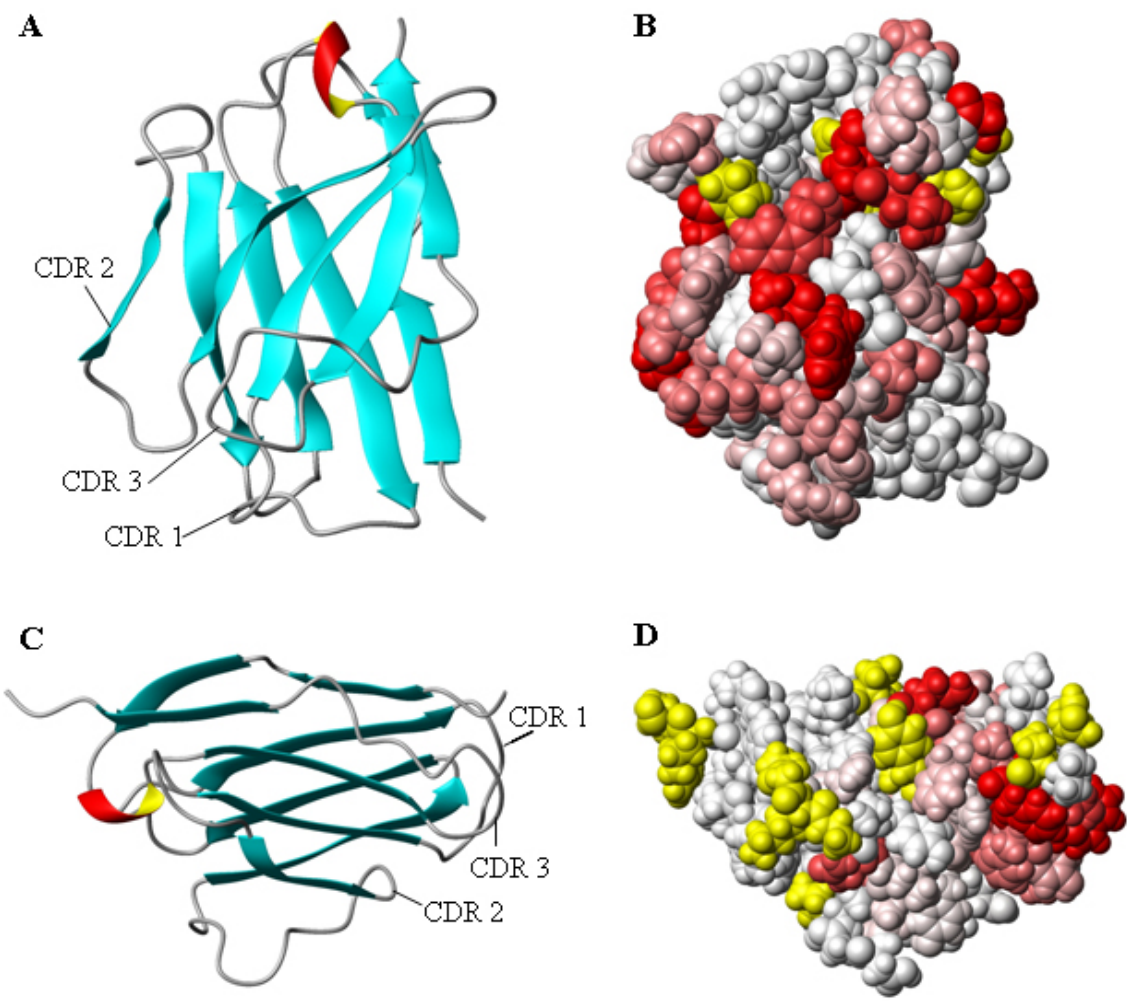

Figure 3.9 Chemical shift perturbation data showing the movement of residues located at the V_H-V_L interface. Panels A and C show ribbon representations of the V_H and V_L domains respectively and are orientated such that the residues located at the V_H-V_L interface are shown. Panel B shows a space filled view of the V_H domain in the same orientation as panel A and panel D shows a space filled view of the V_L domain in the same orientation as panel C. Residues with significantly perturbed backbone amide signals (shift >0.04 ppm) are coloured with a gradient from white to red with yellow residues corresponding to those for which no minimal shift data is available.

3.3.3 Secondary structure prediction

Analysis of the Cα, Cβ and C' assignments obtained for the scFv IC8-IL-1β complex, using TALOS (Cornilescu *et al.*, 1999) and CSI (Wishart & Sykes, 1994), revealed that both proteins have an essentially unchanged secondary structure in the complex (figure 3.10), which strongly suggests that neither of the variable domains or IL-1β undergo significant rearrangements of their secondary structure on complex formation. This clearly indicates that the backbone amide signal perturbations seen on complex formation reflect changes mainly at the interaction sites and supports a restraint-driven docking approach to determine the structure of the scFv-IL-1β complex.

3.3.4 Residual dipolar couplings

Information concerning the relative orientations of the scFv IC8 and IL-1β was obtained from backbone amide RDC measurements for the complex. Partial alignment was achieved using Pf1 phage, with initial optimisation of the concentration producing average RDC values of 3.3 and 6.0 Hz at 600 MHz with phage concentrations of 2.5 and 4.0 mg/ml. The data collected for use in the docking procedure, shown in figure 3.11, was collected at 800 MHz with a phage concentration of 4.0 mg/ml, which gave an average splitting of 10.8 Hz over a much broader range. Attempts to collect an orthogonal set of RDC values for the scFv IC8-IL-1β complex using a bicelle alignment medium proved unsuccessful.

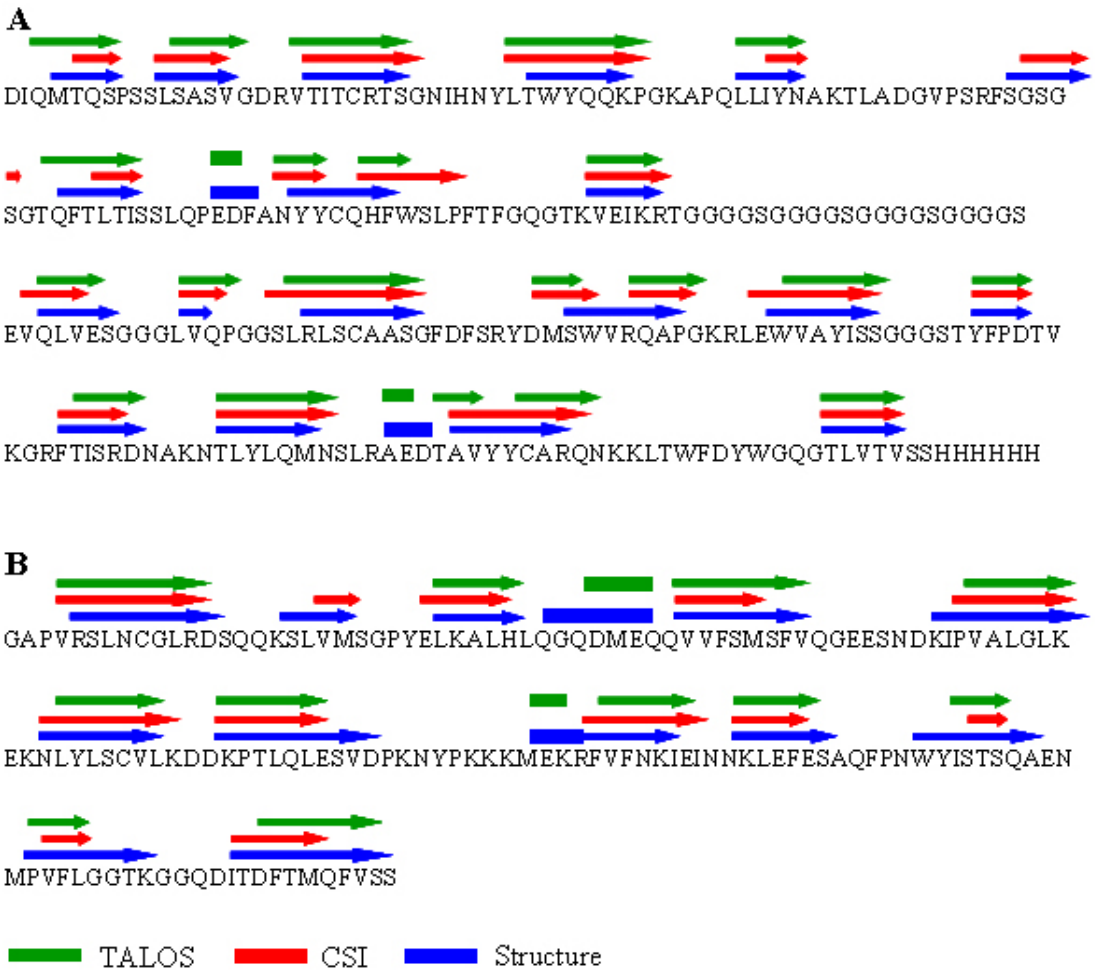

Figure 3.10 A comparison of the secondary structure predictions for the scFv IC8 (panel A) and IL-1β (panel B) from analysis of the Cα, Cβ and C' chemical shifts with TALOS (green) and CSI (red) with the secondary structure from the free scFv model and free IL-1β crystal structure (blue). β-sheets are represented by arrows and α-helices by rectangular bars. Both proteins show an essentially unchanged secondary structure upon complex formation.

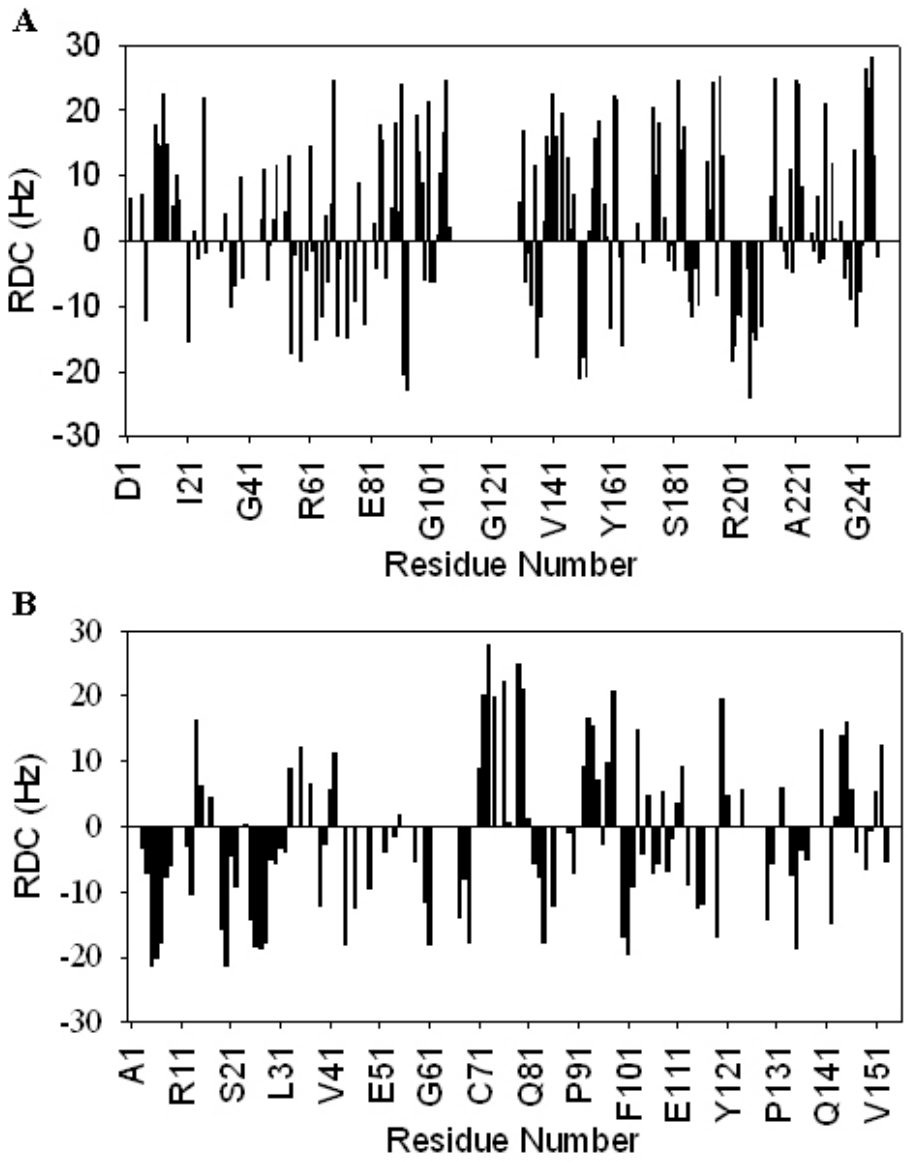

Figure 3.11 Histograms summarising the backbone amide RDC values determined for the scFv IC8 (A) and IL-1β (B) as part of the scFv IC8-IL-1β complex.

3.3.5 Optimisation of restraint driven docking

Optimisation of the input format for the restraints and the docking protocol was performed by detailed analysis of the two driving factors for docking, the RDCs and AIRs. As explained in section 3.2.5, RDC values can be used as direct restraints and/or intervector projection angles. Use of RDCs only as direct restraints results in the

grouping of the 200 output docked complexes into two equally populated but distinct clusters. Combination of the direct RDC use with intervector projection angles, which are utilised at an earlier stage in the docking protocol as recommended by the HADDOCK authors, grouped all 200 docked structures into one reasonably well defined cluster with an overall backbone RMSD of approximately 2 Å. Refinement of the clustering threshold by lowering the default value of 7.5 Å typically grouped the 200 structures into between 6 and 10 clusters with only one or two of these containing more than 15 structures. The most populated cluster (typically 70-90 structures) gave the best agreement with the NMR restraints, had the lowest overall energy and was very well defined with an average overall backbone RMSD of the order of 1 Å.

Correct weighting of the backbone amide chemical shift data to define the AIRs proved to be crucial. Initial thresholds of 0.1 ppm and 20% solvent accessibility for the scFv IC8, giving 11 active and 11 passive residues, and 0.25 ppm and 20% solvent accessibility for IL-1β, giving 13 active and 10 passive residues, gave relatively small AIR interfaces on the proteins. Lowering the combined shift thresholds to 0.075 ppm for the scFv, giving 17 active and 25 passive residues, and 0.1 ppm for IL-1β, giving 26 active and 19 passive residues, produced AIR interfaces of similar size to the combined shift interaction surfaces, as illustrated in figure 3.12. This improved the initial clustering of the 200 docked structures with respect to both overall backbone RMSD and agreement with the combined shift data.

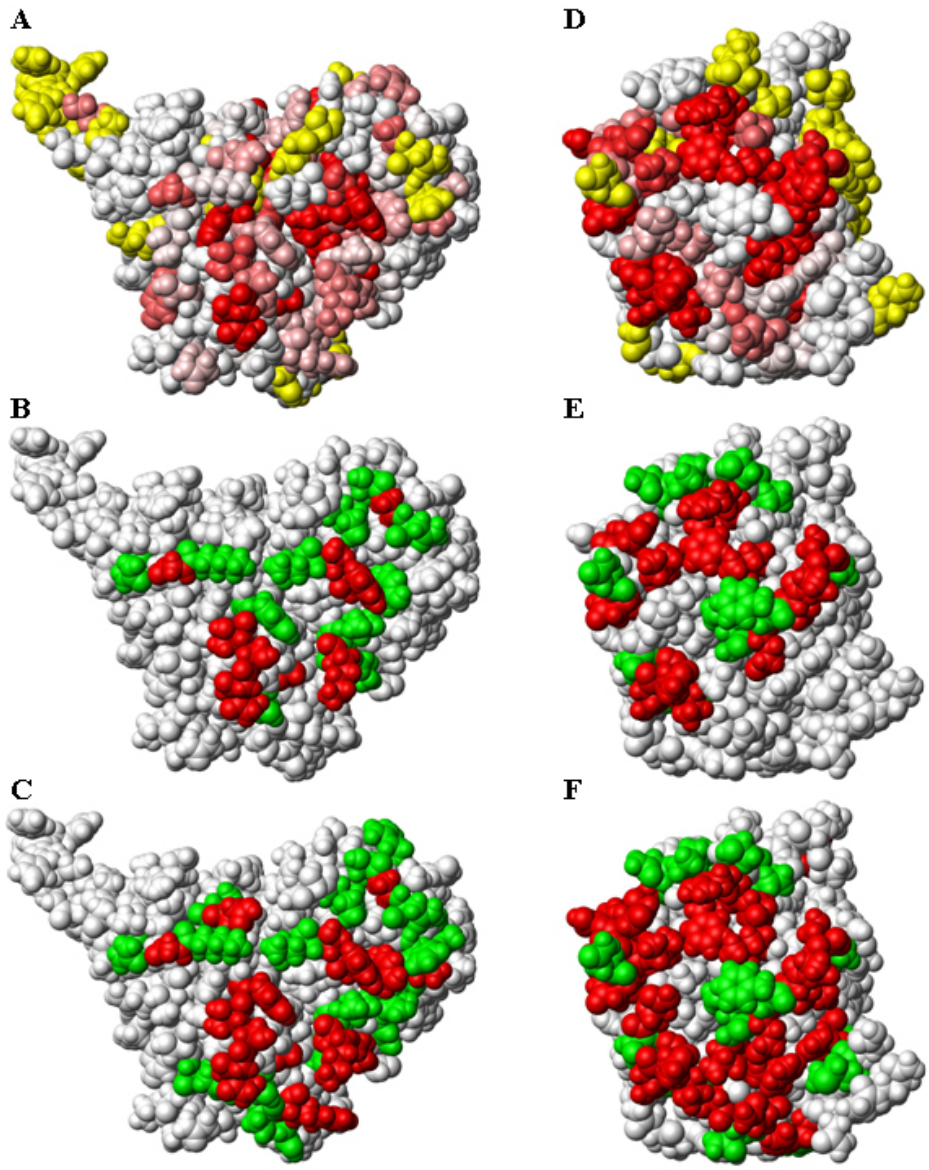

Figure 3.12 Space filled views of the scFv IC8 model (A-C) and free IL-1β crystal structure (D-F). Panels A and D show the binding surfaces on each protein upon formation of the scFv IC8-IL-1β complex as determined by chemical shift perturbation from $^{15}N/^{1}H$ HSQC data for the free and bound proteins. Residues with signals that are significantly perturbed are coloured in a gradient from white to red with yellow residues corresponding to those for which no chemical shift data is available. Panels B, C, E and F show the AIRs with red representing the active residues and green the passive residues. Panel B shows the AIRs for the scFv with a threshold for the chemical shift data of 0.1 ppm and 20% solvent accessibility and panel C shows the AIRs when the

threshold is lowered to 0.075 ppm. Panel E indicates the AIRs for IL-1β with a threshold for the chemical shift data of 0.25 ppm and 20% solvent accessibility and panel F shows the AIRs when the threshold is lowered to 0.1 ppm.

To take account of the potential structural changes at a protein-protein interface HADDOCK enables movement of the backbone and side-chain atoms of a defined set of residues at the interface. By default this incorporates all AIR residues ±2 residues. As shown in section 3.2.3 (figure 3.9) significant chemical shift perturbations were observed at the V_H-V_L interface of the scFv as well as at the IL-1β binding site, suggesting that a reorientation of the two variable domains with respect to each other may occur upon formation of the scFv IC8-IL-1β complex. Defining the semi-flexible interface on each protein as AIR residues ±2 made approximately 44% of IL-1β and 24% of the scFv IC8 semi-flexible. Manual addition of residues at or close to the V_H-V_L interface to the list of semi-flexible residues increased this value to 41% of the scFv, as represented in figure 3.13. A subtle difference was observed between these two approaches with respect to the positioning of the two domains upon docking to IL-1β, suggesting that defining the V_H-V_L interface as semi-flexible allows for subtle movement between the domains. Incorporating all residues as semi-flexible had no further affect on the positioning of the variable domains but reduced the backbone RMSD and population of the main cluster. To reflect the flexible nature of the linker this was always assigned as fully flexible. Defining all residues as fully flexible proved impossible without turning on solvated docking so that a water shell was always present. However, this introduced too much flexibility into the system with it proving impossible to obtain a small number of well defined and significantly populated clusters.

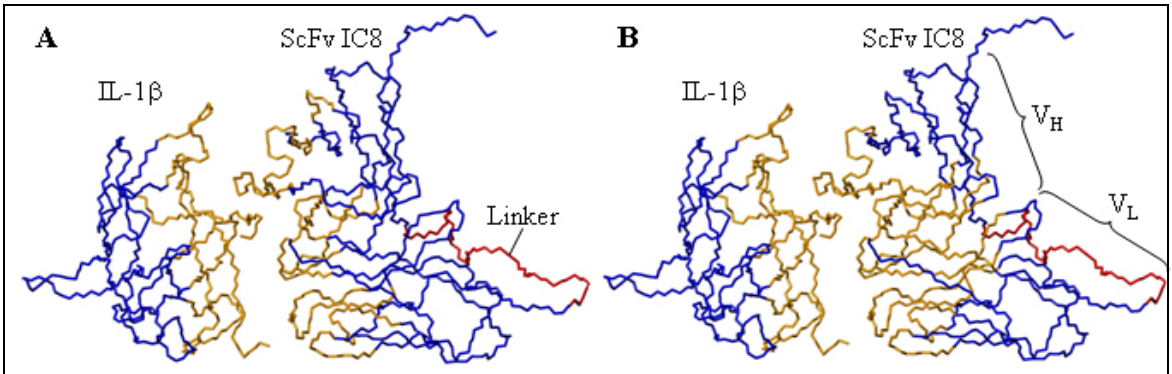

Figure 3.13 Line representations of the scFv IC8-IL-1β complex. The residues coloured in orange are semi-flexible and those coloured in red are fully flexible. Panel A shows the residues defined as flexible using the HADDOCK default setting of ±2 residues from every AIR. Panel B shows the optimised manual definition of flexible residues that includes all continuous loops containing AIRs and all residues at the V_H-V_L interface within the scFv.

Incorporation of the intramolecular NH to NH NOEs as distance restraints proved to be unnecessary as their removal had no effect on the docked scFv IC8-IL-1β complex. This is due to the initial quality of the input scFv model and free IL-1β crystal structure as the NOEs were already satisfied by the input structures. However, the NOE restraints were crucial when the poor scFv model was used as the input structure as some of the β-sheets were misaligned.

3.3.6 Determination of the scFv IC8-IL-1β complex structure

The docking calculations were primarily driven by a knowledge of the interaction surface from backbone amide chemical shift perturbations, which were incorporated as AIRs, and information on the relative orientation of the two proteins from backbone amide RDCs. Backbone amide chemical shift perturbation data was used to identify the

scFv IC8-IL-1β interface and to define the AIRs for docking, which resulted in 17 active and 25 passive residues being selected for the scFv and 26 active and 19 passive residues for IL-1β. Information concerning the relative orientation of the two proteins was obtained using backbone amide RDC measurements for the complex (figure 3.11), with partial alignment achieved using 4 mg/ml Pf1 phage. RDC values were obtained for 107 residues of IL-1β (70%) and 166 of the scFv residues (65%), with data for the remaining residues missing due to overlap in the $^{15}N/^{1}H$ HSQC and TROSY spectra. For the scFv, 387 NOE-derived distance restraints were also included in the docking calculations (119 sequential (i,i + 1), 102 medium (i,i ≤ 4) and 166 long range (i,i ≥ 5)), and for IL-1β 240 NOE-derived distance restraints (78 sequential (i,i + 1), 58 medium (i,i ≤ 4) and 104 long range (i,i ≥ 5)). No intermolecular NOEs were identified.

The optimised docking calculation produced one main cluster for the scFv IC8-IL-1β complex, which contained 77 of the 200 calculated structures and is shown in figure 3.14. A best fit superposition of the structures for the backbone atoms of residues in elements of regular secondary structure in both IL-1β and the scFv gives backbone atom (N, Cα, C') RMSDs to the mean of 0.7 ± 0.1 Å for IL-1β and 0.87 ± 0.2 Å for the scFv. The remaining structures were grouped into eight sparsely populated clusters, which all showed poorer agreement with the NMR data and higher energies. The agreement between the measured RDC data and the two proteins in the scFv IC8-IL-1β complex was greatly improved after docking and refinement using HADDOCK, with Q values reducing from 0.311 to 0.106 for IL-1β and 0.477 to 0.109 for the scFv IC8 (figures 3.15 and 3.16). The Q value is calculated by PALES and can range from zero to one, with values closer to zero showing a better agreement. Figure 3.17 shows the agreement of the back-calculated RDC values for all structures in the final cluster with the

measured RDC values. The final set of converged scFv-IL-1β complexes are entirely consistent with the NMR-derived constraints used for docking with no significant or consistent violations, which clearly indicates that the approach reported here can produce reliable structures for scFv-target protein complexes. Validation of the structure was also performed by comparison of the backbone dihedral angle ranges suggested by TALOS from analysis of the chemical shift assignments (166 ϕ and ψ angle constraints for IL-1β and 276 for the scFv IC8) to those observed in the cluster of 77 scFv IC8-IL-1β complex structures. No consistent dihedral angle violations were reported apart from two for IL-1β (ψ of E96 and I106) and six for the scFv (ϕ of A51 and ψ of T69, F91, G184, N229 and K230), which probably reflects the 2% error rate for TALOS predictions (Cornilescu *et al.*, 1999).

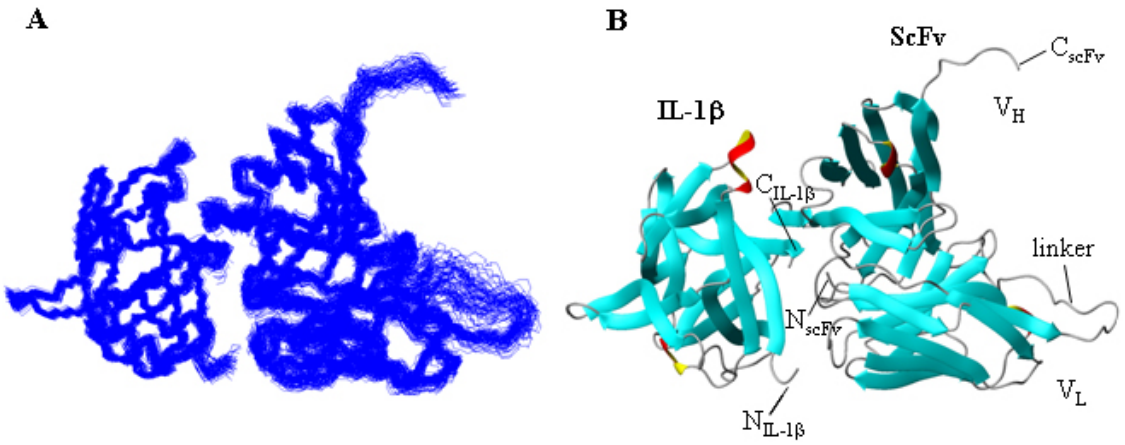

Figure 3.14 Solution structure of the scFv IC8-IL-1β complex (PDB code 2KH2). Panel A shows a best fit superposition of the 77 converged structures obtained for the scFv IC8-IL-1β complex. The structures were superimposed using the backbone atoms in elements of regular secondary structure in both proteins, which gives a backbone atom RMSD to the mean of 0.741 ±0.2 Å for the complex. Panel B is a ribbon representation of the structure closest to the mean in the same orientation as panel A.

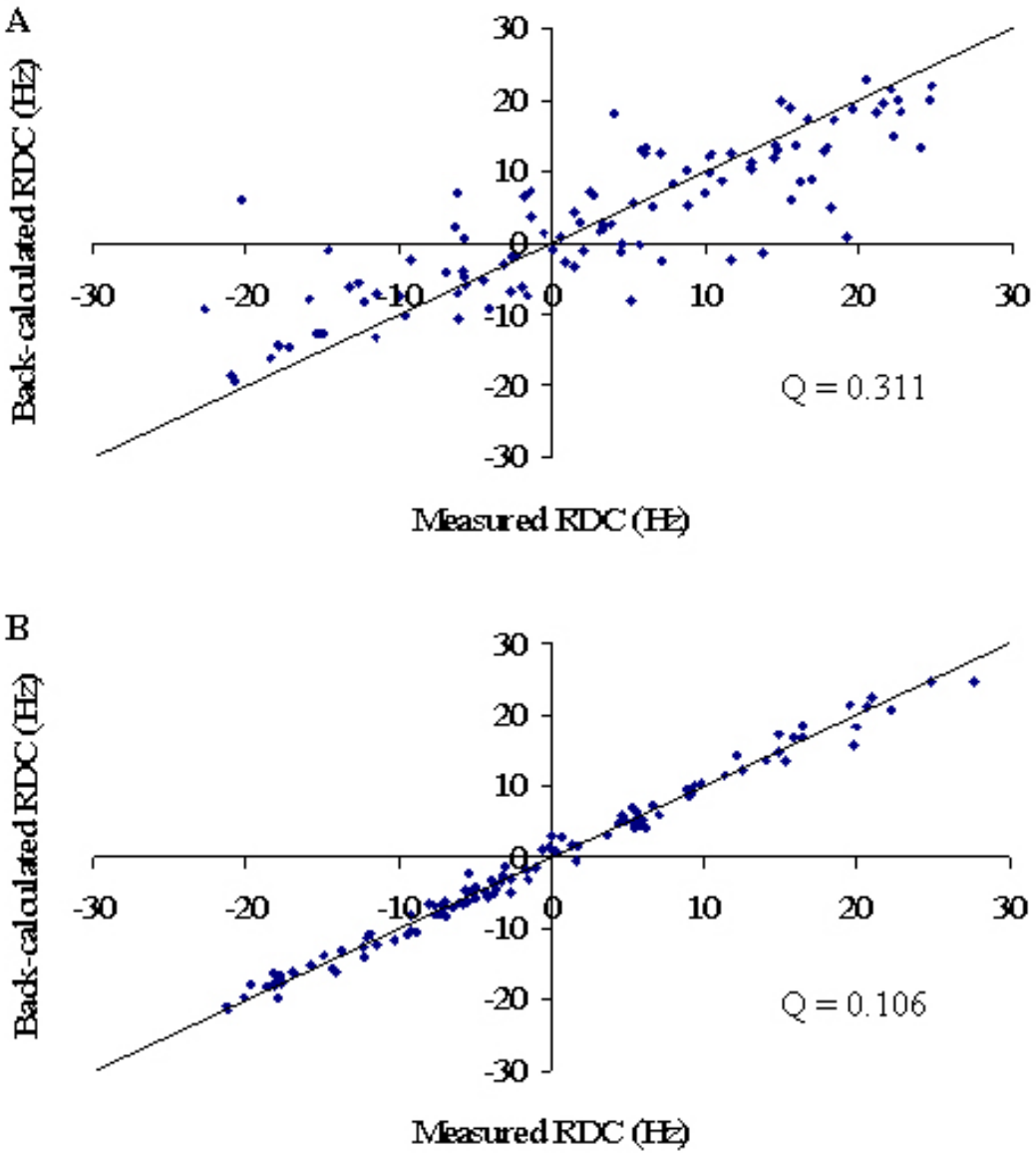

Figure 3.15 Correlation graphs of the measured backbone amide RDC values against back-calculated RDC values for the original IL-1β crystal structure used as the starting point for docking (A) and the IL-1β structure obtained as part of the scFv IC8-IL-1β complex from HADDOCK (B). The values shown in panel B are for the closest structure to the mean from the cluster obtained from the docking protocol. The line on each graph is the line that would be achieved for a perfect agreement of the data. The reduction in the Q value from panel A to B represents the much greater agreement of the data after docking and refinement.

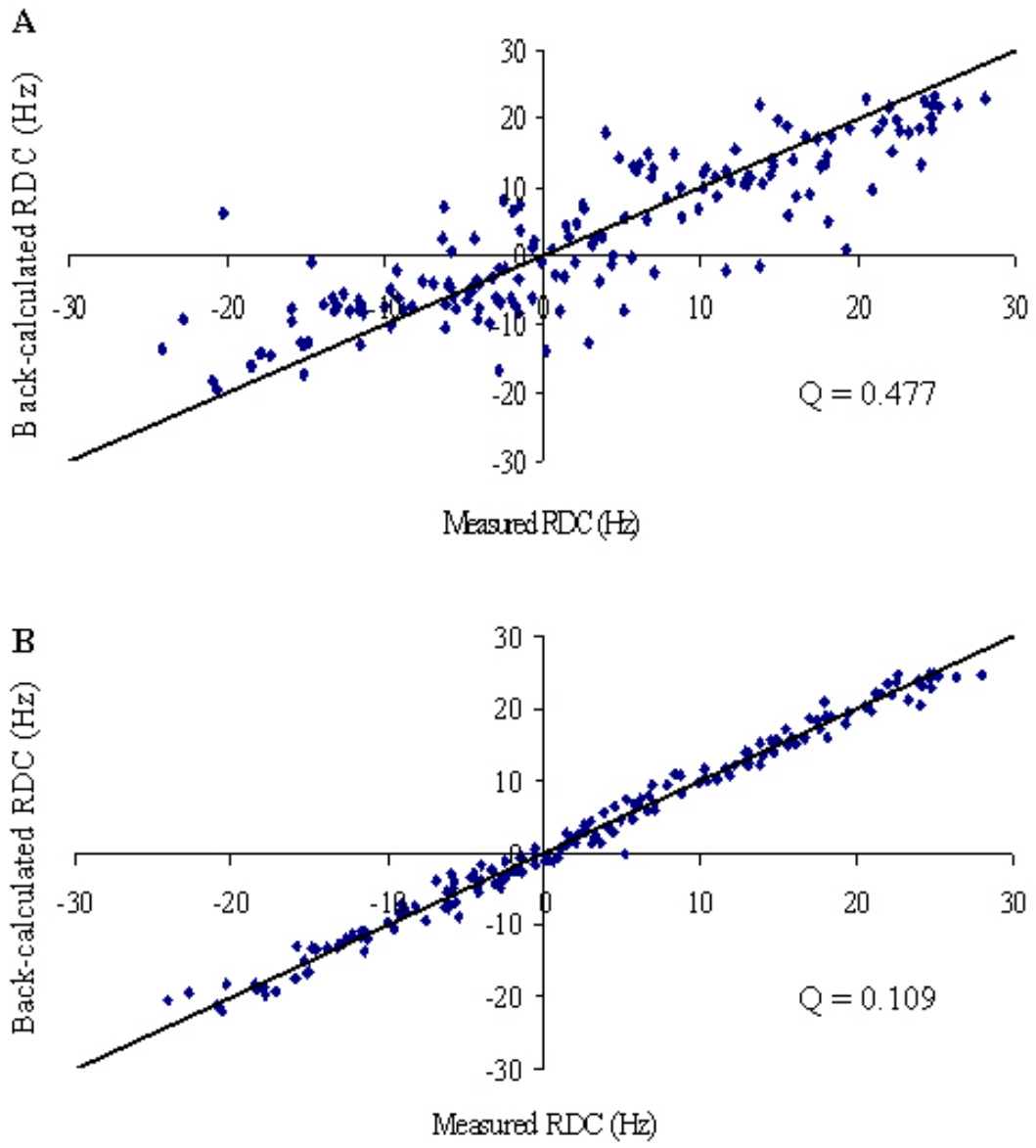

Figure 3.16 Correlation graphs of the measured backbone amide RDC values against back-calculated RDC values for the original scFv IC8 model used as the starting point for docking (A) and the scFv IC8 structure obtained as part of the scFv IC8-IL-1β complex from HADDOCK (B). The values shown in panel B are for the closest structure to the mean from the cluster obtained from the docking protocol. The line on each graph is the line that would be achieved for a perfect agreement of the data. The reduction in the Q value from panel A to B represents the much greater agreement of the data after docking and refinement.

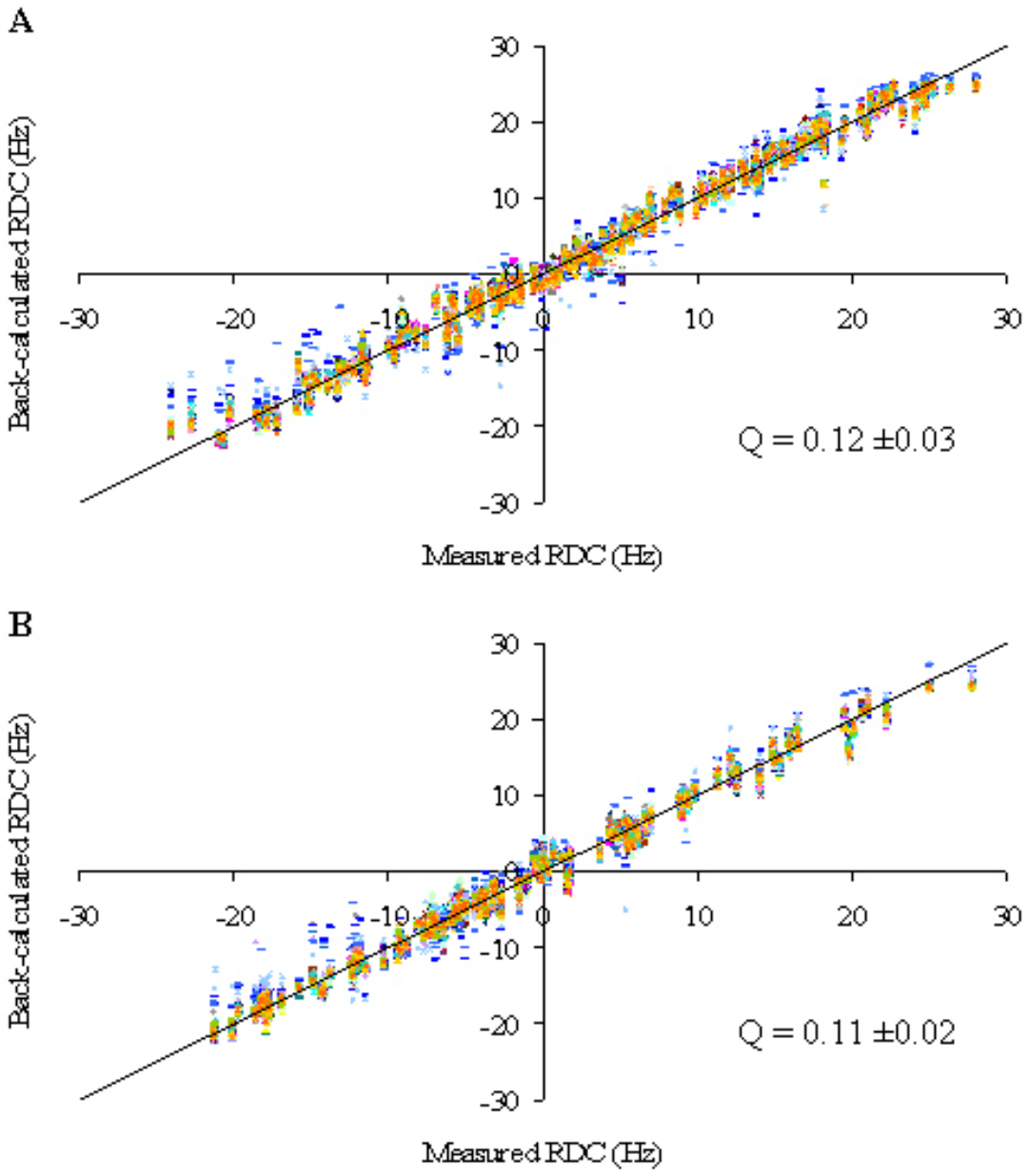

Figure 3.17 Correlation graphs of the measured backbone amide RDC values against the back-calculated RDC values for all 77 structures in the main cluster obtained from HADDOCK for the scFv IC8-IL-1β complex. Panel A shows the agreement of the RDC values for the scFv IC8 and panel B shows the agreement of the RDC values for IL-1β. The line on each graph represents the line that would be obtained for a perfect agreement of the data. The Q values are a mean for all 77 structures with the standard deviation shown.

3.3.7 Robustness of the HADDOCK docking protocol

The robustness of the docking procedure and in particular its dependence on the completeness of the backbone amide RDC data was evaluated by repeating the docking calculations with subsets of the RDC data randomly removed. The full set of backbone amide RDC data obtained covered 65% of the scFv residues and 70% of the IL-1β residues, with reliable data for the remaining residues mainly missing due to overlap in the $^{15}N/^{1}H$ HSQC and TROSY spectra. Removal of up to 30% of the RDC data obtained, resulting in coverage of 46% of the scFv residues and 48% of the IL-1β residues, resulted in the production of several equally populated clusters from the docking calculations, however, the cluster of complex structures with the lowest energies and best agreement with the NMR data were closest to the scFv IC8-IL-1β complex obtained with the full data set (backbone RMSDs of 1.5 ± 0.15 Å for all residues). As expected, removal of over 50% of the RDC data collected resulted in significant variability between the clusters of scFv IC8-IL-1β complexes produced by docking, with no reliable criteria on which to select the correct complex structure. This analysis clearly illustrates the importance of obtaining a good coverage of RDC measurements to obtain a reliable complex structure by NMR restraint-driven docking. For the scFv IC8-IL-1β complex reported here this corresponds to backbone amide RDC data for over 60% of the residues in each protein, which is likely to reflect the requirement for tight protein complexes in general.

The necessity for a good starting model was analysed by comparison of the docked scFv IC8-IL-1β structures produced when using either a good or poor scFv model as an input structure. As shown in figure 3.18, the resulting scFv structures are clearly from the

same family of structures but subtle differences are seen in particular in some of the CDRs.

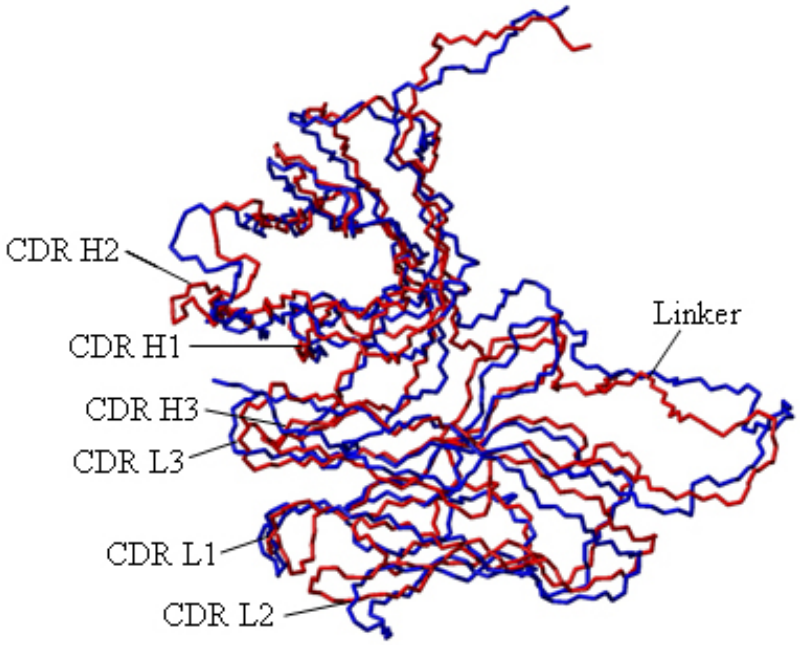

Figure 3.18 An overlay of the docked scFv structures produced when the good model (red) and poor model (blue) were used as the input structures for the docking calculations. The backbone RMSD between the two structures is 3.2 Å. The CDRs are indicated with the largest differences in positioning shown by CDRs L2 and H2.

The docked complex from the poor model satisfies all NOE distance restraints, which it did not do prior to the docking protocol, indicating that the intramolecular NH to NH NOEs are necessary in this instance. The agreement of the back-calculated RDC data for the docked complex with the measured backbone amide RDC data is greatly improved with the Q value reducing from 0.743 to 0.182 but this is not quite as good as the fit obtained for the docked scFv from the good model (figure 3.19). An additional docking procedure using a representative output structure for the docked scFv from the poor model as a starting point does not improve the agreement further, suggesting that a local minimum has been reached that cannot be overcome with the available data.

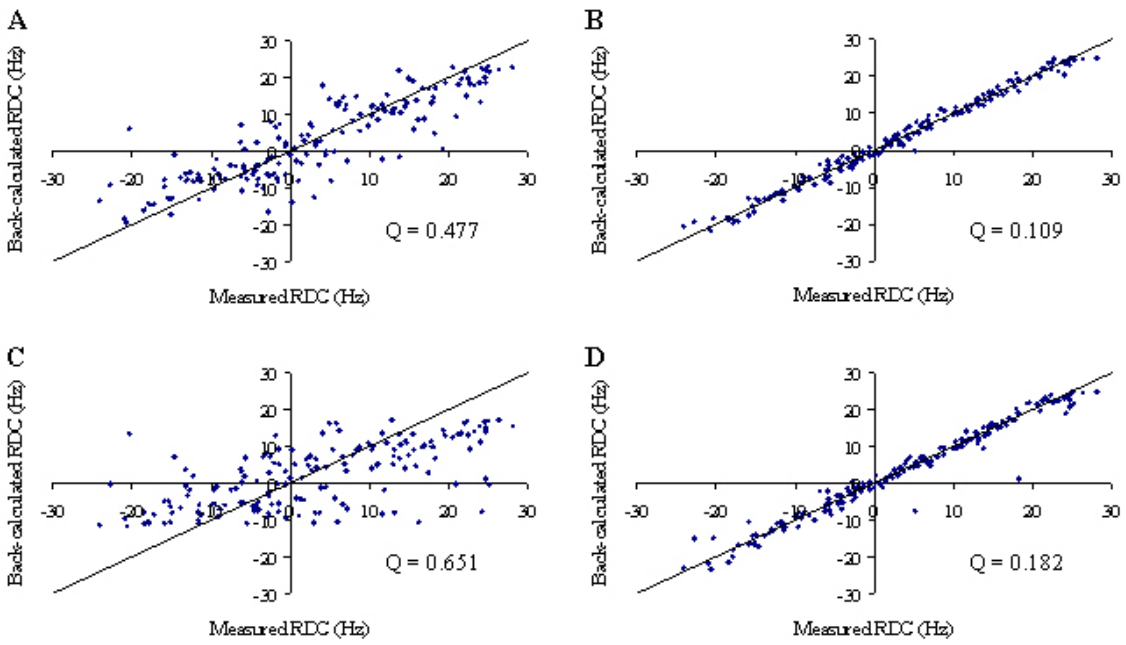

Figure 3.19 Correlation graphs of the measured backbone amide RDC values obtained using 4 mg/ml phage alignment media against back-calculated RDC values from PALES for the good scFv IC8 model (A), the docked scFv IC8 structure obtained using the good model as an input structure (B), the poor scFv IC8 model (C) and the docked scFv IC8 structure obtained using the poor model as an input structure (D). The line on each graph is the line that would be achieved for a perfect agreement of the data. The values shown for the docked scFvs in panel B and D are for the closest structure to the mean from the major cluster obtained in the docking procedure. The data shows a significant improvement in the agreement of the backbone amide RDC values from the models to the docked scFv structures. The docked scFv structure obtained from the poor model, however, does not reach the same level of agreement as the structure from the good model.

3.3.8 ScFvs as models for Fab fragments

No direct evidence has been previously presented to show that a scFv and Fab with identical V_L and V_H domains bind in an essentially identical manner to a target protein, although this has been shown for Fab and Fv D1.3-HEL complexes (Bhat *et al.*, 1990). A comparison of the $^{15}N/^{1}H$ HSQC spectra for ^{15}N labelled IL-1β bound to an equivalent scFv and Fab is shown in figure 3.20. The spectra of IL-1β bound to the scFv and Fab look nearly identical in terms of the positions of signals, which provides evidence that scFvs bind in a very similar manner to an equivalent Fab. Very subtle chemical shift differences can be seen at the N and C terminal regions of IL-1β which are involved in the interaction with the antibody fragments. However, these chemical shift differences are approximately one tenth of those induced by scFv or Fab binding (figure 3.8) and must reflect only very subtle differences in the interactions with the Fab and scFv, which may not be detectable by other NMR measurements such as intermolecular NOEs nor by X-ray crystallography.

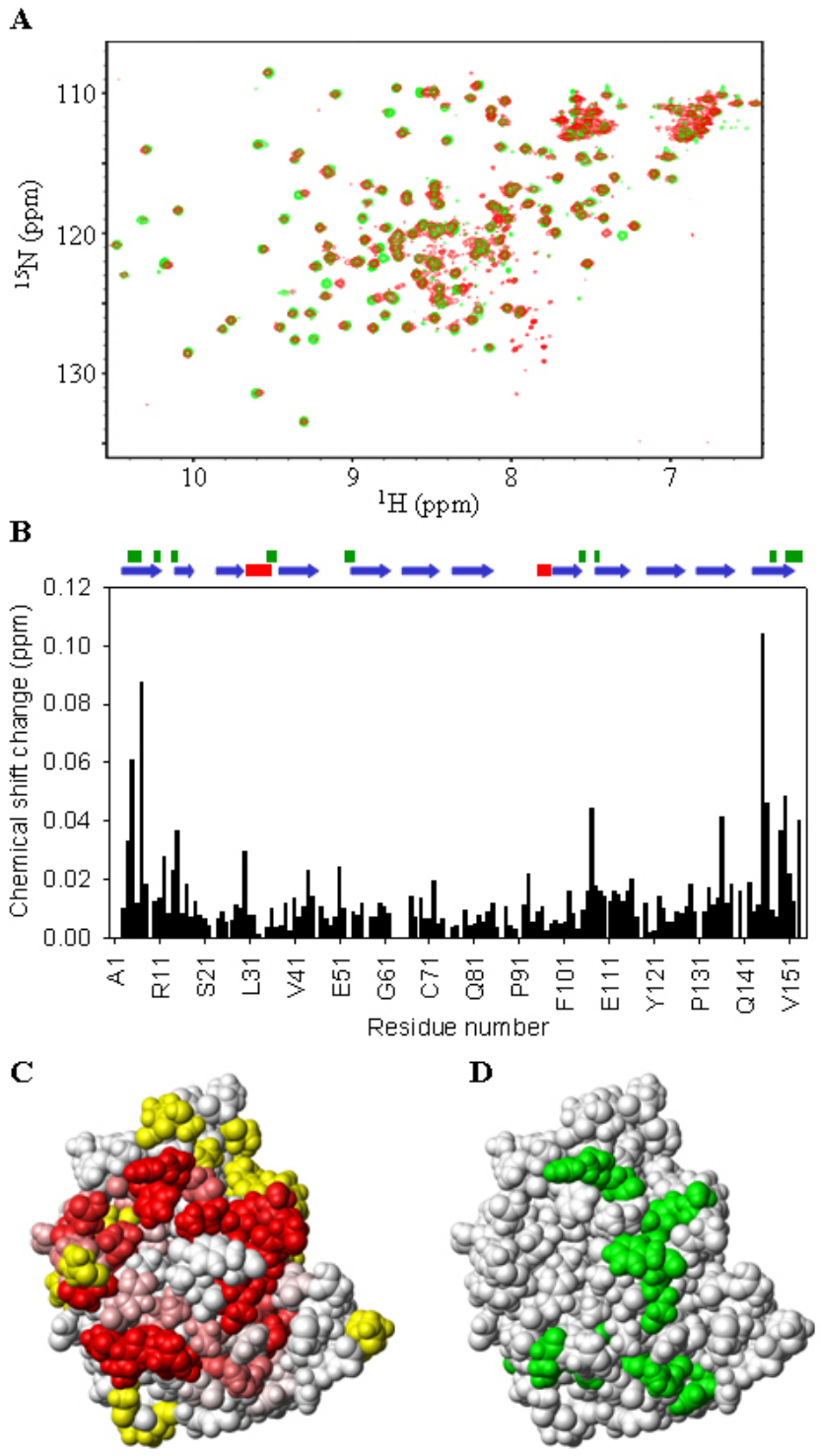

Figure 3.20 Comparison of IL-1β binding to equivalent scFv and Fab IC8 antibody fragments. Panel A shows overlaid $^{15}N/^{1}H$ HSQC spectra for ^{15}N labelled IL-1β bound to unlabelled scFv IC8 in red and unlabelled Fab IC8 in green. The spectra overlay extremely closely, showing that the interactions with IL-1β are essentially identical

when bound to either the scFv or Fab. Combined amide backbone chemical shift differences are summarised in the histogram (B), with the position of elements of regular secondary structure represented by red bars for helices and blue arrows for β-sheets. IL-1β regions that interact with the scFv are indicated by the green bars. Panel C shows the scFv binding site on IL-1β determined by backbone amide chemical shift perturbation analysis with residues with significantly shifted signals coloured in a gradient from white to red and residues for which no data is available coloured in yellow. Panel D is a space filled view of IL-1β in the same orientation as panel C with the green residues representing those which have signals showing a small but significant chemical shift (shift >0.03 ppm). These residues are primarily located at the antibody binding site showing that there are very subtle differences between the binding of the scFv and Fab IC8 antibody fragments to IL-1β.

3.4 Discussion

The work reported in this chapter clearly demonstrates that the problems discussed in chapter 2 associated with scFv multimerisation can be avoided by working with the scFv-target protein complex. Perdeuteration of the proteins enables high quality NMR data to be collected and comprehensive sequence specific assignments to be made as shown by the assigned $^{15}N/^{1}H$ HSQC spectra and ^{15}N strips from HNCACB and NOESY-HSQC spectra in figures 3.2 to 3.6. A NMR restraint-driven docking approach using chemical shift changes, NOEs and RDCs can then be successfully used to determine the structure of a scFv-target protein complex.

The structure obtained for the scFv IC8-IL-1β complex (PDB code 2KH2), shown in more detail in figures 3.21 and 3.22, is typical of that expected for an antibody-target protein complex with residues in the CDR loops responsible for most of the contacts with IL-1β. The complex features a fairly large interface between the two proteins with the buried surface area upon complex formation corresponding to 1930 ± 130 $Å^2$, as calculated with MolMol (Koradi *et al.*, 1996) by loss of solvent accessibility. Residues were considered to be involved in intermolecular contacts at the interface if the distance between neighbouring atoms corresponded to less than the sum of their van der Waals radii plus 0.5 Å (Tsai *et al.*, 1996). On this basis, 18 residues from both the scFv IC8 and IL-1β contribute to the protein-protein interface, as summarised in table 3.1. For the scFv, four of the six available CDRs (L1, L3, H2 and H3) make significant interactions with IL-1β and interestingly three framework (FR) residues (D1, D191 and K194) are also found at the interface. The contact residues from the scFv are equally spread between the V_H and V_L domains, with CDRs L1 and H2 making the most contacts with IL-1β. One notable difference to the majority of antibody structures reported previously

is the peripheral involvement of CDR H3, with only one residue from this CDR being involved in intermolecular interactions. Antibody-antigen intermolecular interactions are typically dominated by residues from the CDR 3 loops of both variable domains, with CDR H3 often accounting for the greater proportion of contacts (MacCallum *et al.*, 1996). Figure 3.22 clearly shows that CDR H3 lies opposite a cavity in IL-1β, meaning that only the long side chain of K231 makes any interactions across the interface.

Table 3.1 A summary of the contact residues and interactions made in the **s**cFv IC8-IL-1β complex. In addition to the expected contacts from CDR loops in the scFv, three framework (FR) residues also interact with IL-1β. Residues highlighted with an asterisk (*) are involved in hydrogen bonds across the interface.

		ScFv	**IL-1β**
V_L	**FR L1**	Asp 1*	Ser 152, Ser 153*
	CDR L1	Asn 28	Ala 1
		Ile 29*	Arg 4*
		His 30	Arg 4
		Asn 31*	Arg 4*
		Tyr 32	Arg 4
	CDR L3	Trp 92	Arg 4, Ser 5, Phe 46
		Ser 93	Leu 6
		Leu 94	Phe 150, Val 151, Ser 152
V_H	**CDR H2**	Ser 181	Asn 108
		Gly 183	Asn 108
		Gly 185	Gln 15
		Ser 186*	Asn 108*, Thr 147
		Thr 187*	Arg 11*
		Tyr 188	Phe 150
	FR H3	Asp 191*	Val 151, Ser 152*
		Lys 194*	Glu 37*, Gln 39*
	CDR H3	Lys 231*	Asn 53*, Asp 54, Glu 105*

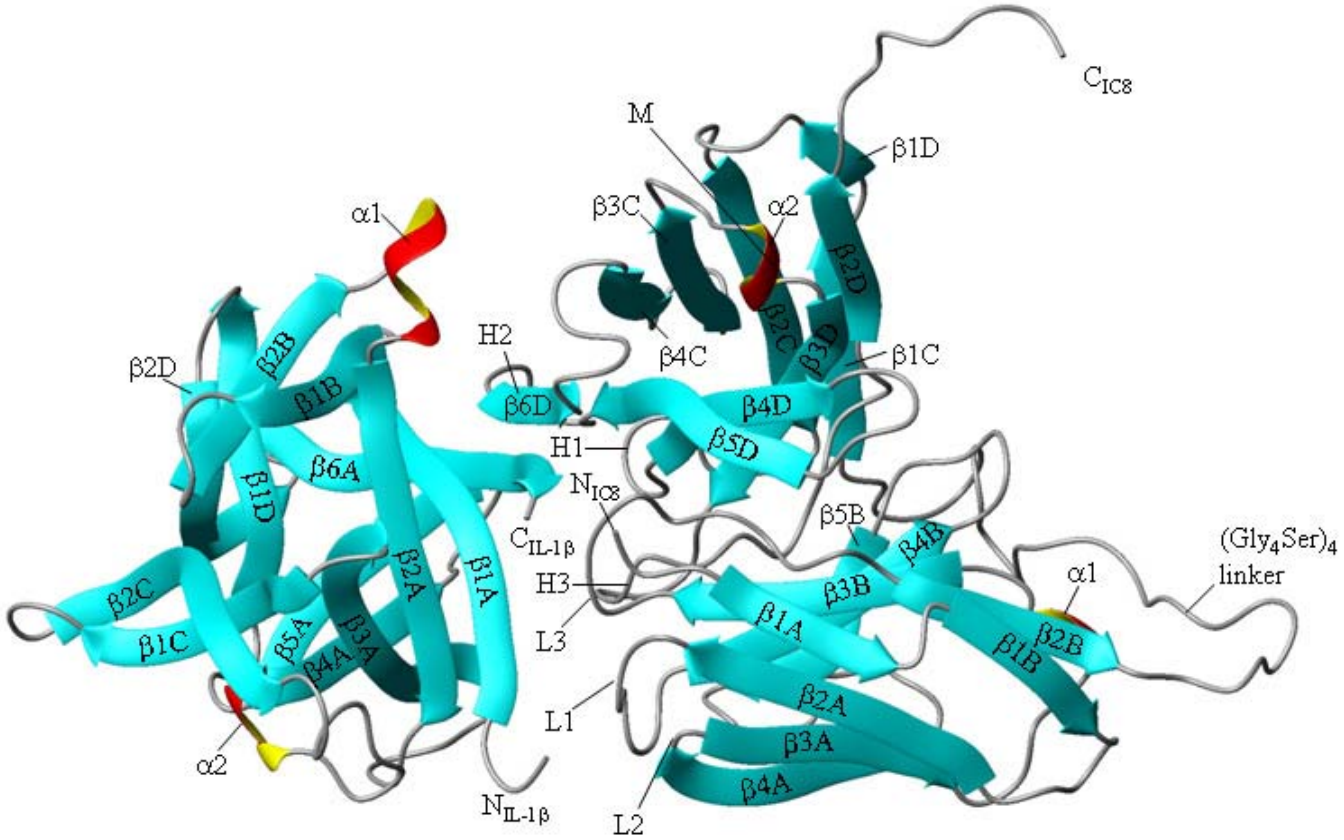

Figure 3.21 Ribbon diagram of the solution structure of the scFv IC8-IL-1β complex (PDB code 2KH2). The CDR loops are labelled as L1, L2 and L3 for the variable light domain and H1, H2 and H3 for the variable heavy domain. The site of the specific Ala to Asp mutation, which was engineered to try and disrupt the formation of domain-swapped dimer, is located in the α2 helix of the scFv as indicated by M.

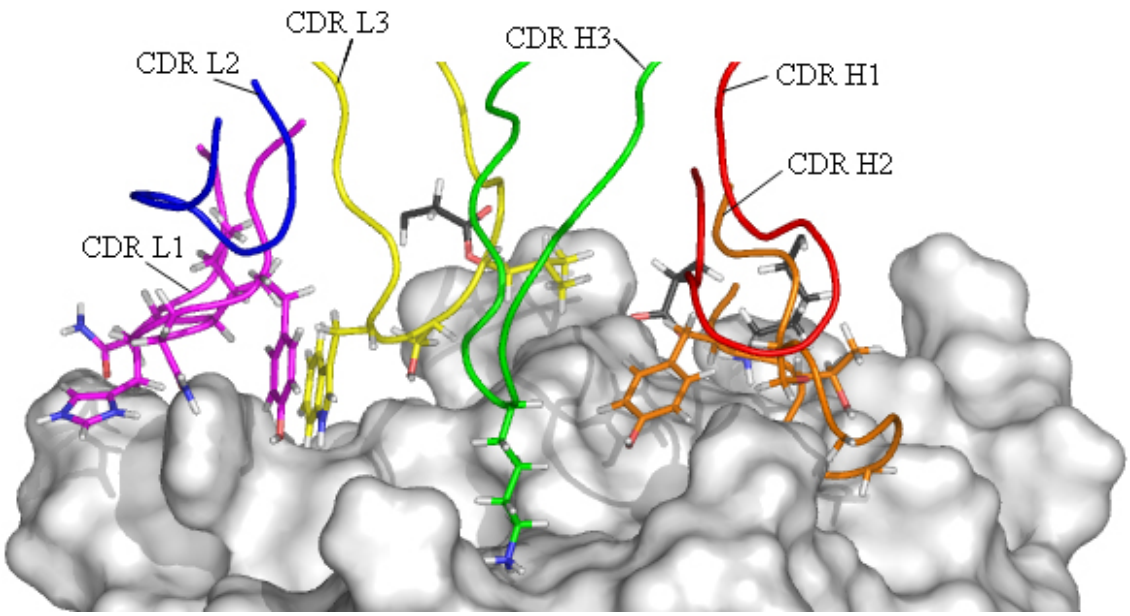

Figure 3.22 A view of the antibody binding site in the scFv IC8-IL-1β complex. A surface representation of IL-1β is shown with the backbone atoms of the scFv CDRs shown as coloured loops. The side chains of scFv IC8 residues that are involved in intermolecular interactions are shown as line representations with nitrogen atoms coloured in blue, oxygens in red and hydrogens in white. CDR L1 is coloured in magenta, L2 in blue, L3 in yellow, H1 in red, H2 in orange, H3 in green and interacting framework residues in dark grey.

Detailed analysis of the scFv IC8-IL-1β residues at the protein-protein interface and in direct contact is shown in table 3.2. The interface and contact surface on IL-1β is slightly less charged and much more polar than the rest of the solvent exposed surface of the protein. The interaction and contact surface on the scFv IC8 has a much higher aromatic content than the rest of the antibody, which is a general characteristic of antibody binding sites.

Table 3.2 Analysis of the residue characteristics within the scFv IC8-IL-1β complex. Interface residues are those that lose solvent accessibility upon formation of the complex (calculated using MolMol) with residues considered to be in contact if the distance between neighbouring atoms corresponded to less than the sum of their van der Waals radii plus 0.5 **Å** (Tsai *et al.*, 1996). Surface exposed residues are all residues with a greater than 5% solvent accessibility with residues below this threshold defined as buried (Jones & Thornton, 1996; Miller *et al.*, 1987).

Amino acid type	IL-1β					IC8			
	Interface	**Contact**	**Surface**	**Buried**		**Interface**	**Contact**	**Surface**	**Buried**
Charged	25.0%	27.8%	32.1%	9.1%		20.7%	27.8%	21.4%	8.5%
Polar	50.0%	44.4%	28.6%	29.5%		31.0%	33.3%	41.3%	22.5%
Non-polar	25.0%	27.8%	39.3%	61.4%		48.3%	38.9%	37.3%	69.0%
Aromatic	8.3%	11.1%	7.1%	15.9%		27.6%	22.2%	5.6%	21.1%

Comparison of the characteristics of the scFv IC8-IL-1β complex with a selection of other scFv and Fab-target protein complexes is shown in table 3.3. The nature of the scFv interface in the scFv IC8-IL-1β complex, in particular the high aromatic and non-polar content is in very good agreement with the typical values observed for the selected antibody-target protein complexes. All the complexes have a large buried surface area of approximately 1900 ±330 $Å^2$ consisting of about 27 residues with between 5 and 14 hydrogen bonds between the antibody and target protein. These values are also comparable to those obtained by Lo Conte and co-workers, who analysed 19 antibody fragment complexes giving an average buried surface area of 1680 ±260 $Å^2$ and 9 hydrogen bonds across the interface (Lo Conte *et al.*, 1999).

Table 3.3 A comparison of the interface features of a representative selection of antibody-target protein complexes (Ay *et al.*, 2000; Faelber *et al.*, 2001; Hwang *et al.*, 2006; Josephson *et al.*, 2002; Li *et al.*, 2008; Malby *et al.*, 1998; Mirza *et al.*, 2000; Nybakken *et al.*, 2005), as well as the scFv IC8-IL-1β complex. Buried surface area (BSA), interface residues and hydrogen bonds were determined using MolMol. The interface is defined by residues that lose solvent accessibility on complex formation. Residues were considered to be involved in intermolecular contacts at the interface if the distance between neighbouring atoms corresponded to less than the sum of their van der Waals radii plus 0.5 Å (Tsai *et al.*, 1996).

PDB code	**Antibody type**	**Target Protein**	**BSA (Å^2)**	**Hydrogen bonds**	**Interface (%)**				**Interface residues**	**Contact residues**
					Aromatic	**Non-polar**	**Polar**	**Charged**		
1NMC	scFv	Neuraminidase	1416	6	31.6	45	35	20	19	11
1DZB	scFv	Lysozyme	1856	6	32.1	46.4	14.3	39.3	29	19
2GHW	scFv	SARS spike	2528	5	13.5	27	43.2	29.7	37	27
1LK3	Fab	IL-10	1732	14	26.9	50	34.6	15.4	26	18
1FSK	Fab	Pollen antigen	1761	5	41.7	54.2	29.2	16.7	24	14
1ZTX	Fab	West Nile protein	1691	10	26.9	42.3	34.6	23.1	26	18
1JPS	Fab	Tissue factor	2103	18	25.8	45.2	22.6	26.1	31	20
3B2V	Fab	EGF receptor	1869	8	30.8	61.5	30.8	7.7	26	15
MEAN			**1870**	**9**	**28.7**	**46.5**	**30.5**	**22.3**	**27**	**18**
S.D.			**329**	**5**	**8**	**10**	**9**	**10**	**5**	**5**
2KH2	scFv	IL-1β	1930	11	27.6	48.3	31	20.7	29	18

Early analysis of protein-protein recognition sites based on a small number of structures of mainly protease-inhibitor and antibody-lysozyme complexes revealed a homogeneity in the size of interfaces and suggested that protein surfaces that form recognition sites do not differ greatly in their chemical character from the rest of the solvent-accessible surface (Chothia & Janin, 1975; Janin & Chothia, 1990). The interfaces have a large non-polar component which favours complex association to become buried, and a polar component providing hydrogen bonds. More recent studies used a much larger sample of a greater variety of complexes (Jones & Thornton, 1996; Lo Conte *et al.*, 1999). The study by Jones and Thornton analysed 32 nonhomologous homodimers and 26 heterodimers, of which 10 were enzyme-inhibitor and 7 antibody-antigen complexes. The work by Lo Conte and co-workers inspected 75 heterodimers, comprising of 24 protease-inhibitor complexes, 19 antibody-antigen complexes and 32 complexes of other kinds. Both studies gave similar conclusions finding no significant correlation between the size of the protein-protein interface and the size of the proteins in the complex, with an average buried surface area for heterodimer complexes of approximately 1950 $Å^2$. In contrast to homodimer complexes, which have hydrophobic interfaces with few hydrogen bonds, the heterodimer complex interfaces are on average as non-polar as the rest of the solvent accessible surface of the proteins, have a greater proportion of neutral polar groups giving rise to an average of 10 hydrogen bonds and are less abundant in charged residues. Homodimers rarely occur as functional monomers, and so their hydrophobic interfaces are permanently buried within the complex. The majority of the heterodimer complexes analysed do occur as monomers in solution, hence the interfaces cannot be as hydrophobic as this would be energetically unfavourable. The interfaces are rich in aromatic residues (21% at the interfaces versus 8% on the rest of the surface) but are depleted in all charged residues except Arg, which

is the residue that makes the largest overall contribution to interfaces (10%). It is noted that the spread around these average heterodimer complex values is large with significantly polar and non-polar interfaces observed ranging in size from 1140 to 4660 $Å^2$.

The interfaces of antibody-target protein complexes are significantly more polar than the average protein-protein interface with Tyr contributing the most to the interface. The values obtained on analysis of the scFv IC8-IL-1β complex (table 3.2) and other antibody-target protein complexes (table 3.3) are broadly in agreement with the above conclusions. The interface size (1930 $Å^2$) and number of hydrogen bonds (11) are in very close agreement with the average protein-protein heterodimer complex values. As would be predicted the interface on the scFv IC8 is highly aromatic with Tyr representing 2 of the 18 contact residues.

Of particular importance for the design and humanisation of therapeutic antibodies is a knowledge of which residues from the antibody make interactions with the target protein and are thus necessary for therapeutic activity. Residues not necessary for binding can be considered for mutation to reduce immunogenicity as required, which is particularly useful if humanisation of the CDRs is desired. CDRs not involved in binding, which on average is one of the six CDRs, could potentially be mutated to more human like CDR sequences without adversely affecting the binding activity of the therapeutic antibody. In the scFv IC8-IL-1β complex structure CDRs L1 and H2 make the most contacts with IL-1β while CDRs L2 and H1 play no role in binding. Of the eight antibody-target protein complexes selected for comparison from the PDB, three (2GHW, 1LK3, 1DZB) had interactions with their target protein involving framework

residues as well as CDR loops (defined using the IMGT numbering system (Lefranc, 1997)). This feature is also seen in the scFv IC8-IL-1β complex structure with three scFv framework residues contacting IL-1β. Mutation of these framework residues involved in binding would almost certainly have a dramatic affect on binding activity of the antibody, which is why a knowledge based approach to antibody design and humanisation using a high resolution antibody-target protein structure is so desirable.

A comparison of the interface characteristics of antibodies selected as therapeutics with ones not selected specifically for this purpose (table 3.4) reveals no obvious differences in the size of the contact surface, types of interactions or the locations of residues involved in interactions.

Table 3.4 A comparison of the interface features in complexes formed with therapeutic and non-therapeutic antibodies (Ay *et al.*, 2000; Faelber *et al.*, 2001; Hwang *et al.*, 2006; Josephson *et al.*, 2002; Li *et al.*, 2008; Malby *et al.*, 1998; Mirza *et al.*, 2000; Nybakken *et al.*, 2005).

Mean	**Therapeutic**	**Normal**
Aromatics (%)	29.6 ± 6.5	36 ± 14.2
Charged (%)	23.1± 6.5	21.2 ± 9.5
Polar (%)	30.4 ± 7.3	30.3 ± 17.4
Non-polar (%)	46.4 ± 8.9	48.4 ± 13.2
BSA (Å^2)	1869 ± 161	1859 ± 409
No. of H bonds	11 ± 4.0	7 ± 3.8
No. of interface residues	28 ± 2.3	27 ± 6.6
No. of contact residues	18 ± 1.8	18 ± 6.1
No. of contact atoms	160 ± 13	164 ± 48
Non-CDR residue contacts	0.8 ± 1.3	2.6 ± 4.0
%VL residues	40 ± 8.9	43 ± 14.8%
%VH residues	60 ± 8.9	57 ± 14.8%

The restraint-driven docking approach discussed in this chapter has been shown to be heavily reliant on the backbone amide RDC data. In particular removal of RDC values reduced the accuracy of the docking protocol significantly once below approximately 60% of total residue coverage, resulting in the output of several equally populated clusters. The closest cluster to the cluster obtained using all the available data could be selected based upon analysis of the overall energy levels and agreement with the input restraints. Reduction of the RDC coverage to below approximately one third of the residues produced a large diversity of sparsely populated clusters. Blind selection of the closest cluster to the cluster obtained using all the data proved impossible at this point. Therefore, when using this approach it is vital to ensure that backbone amide RDC values are measured for as many residues as possible. To reduce the loss of data due to overlap from the two dimensional spectra utilised in this thesis, a three dimensional approach to the collection of RDC could be utilised as has been reported previously (Kontaxis *et al.*, 2000). Collection of a second RDC type, such as nitrogen-α-carbon or nitrogen-carbonyl, from the same alignment medium may also be useful as an additional restraint. It is advantageous to collect RDC data from an orthologous alignment medium, which can be used to validate the structure. Unfortunately it has proved impossible to collect accurate backbone amide RDC values for the scFv IC8-IL-1β complex from a bicelle alignment medium.

The very similar $^{15}N/^{1}H$ HSQC spectra seen for IL-1β bound to equivalent scFv and Fab IC8 antibody fragments (figure 3.20) provide direct evidence that scFvs bind target proteins in an essentially identical manner to Fabs, which makes scFv-target protein complexes an attractive target for structural studies directed at a knowledge based approach to rational design and humanisation of antibodies. Interestingly, very subtle

differences can be seen in the shifts of some interface residues but must reflect very small changes in interactions. The work reported here strongly suggests that the use of NMR spectroscopy to determine high resolution structures for scFv-target protein complexes has the potential to have a major impact on the rational design, optimisation and humanisation of therapeutic antibodies. Therefore, this confirms that the manner of binding shown by equivalent scFvs and Fabs, at least in this case, are so similar that the scFv-target protein system could be used as a model to infer the structure of the variable domains and hence the target protein binding site for the larger therapeutic Fab or full antibody. Thus the approach shown here to determine a high resolution structure of a scFv-target protein complex by NMR restraint-driven docking can be used for future therapeutic antibody-target protein complexes to accurately determine the residues involved in binding and to influence the design and humanisation of the therapeutic antibody.

The chemical shift perturbation data for the scFv IC8 revealed a very interesting feature, with the two variable domains appearing to move relative to each other upon binding to IL-1β (figure 3.9). The selection of antibodies displayed on the surface of B-cells is dependent on antigen binding to the antibody inducing a signal for the selection of that particular antibody. It has been hypothesised previously that binding of the antigen could induce a conformational change within the antibody binding site which could pass down through the entire antibody causing the signal to be passed to the cell membrane, promoting selection of the antibody. Initially, the chemical shift perturbation data showing a ripple of movement through the variable domains supported this idea. However, further work by my colleague, Catherine Hall, on the Fab IC8-IL-1β complex revealed that comparison of $^{15}N/^{13}C/^{1}H$ HNCO spectra for free and bound Fab IC8

showed no chemical shift perturbations being transmitted through the constant domains (figure 3.23). Instead, in addition to the shifted signals seen for residues at the antibody-target protein and V_H-V_L interfaces that had been observed in the scFv IC8 system, significantly perturbed signals were also seen for residues at the interface between the V_L and C_L domains. This suggests a rotation of the V_L domain upon binding of IL-1β, causing movement at the V_H-V_L and V_L-C_L interfaces but no further transduction of a signal through the antibody.

IgG antibodies are extremely dynamic due to the flexible hinge region between the two Fab arms and the Fc domain (Oi *et al.*, 1984). The end-to-end distance for IgG1 in solution has been determined to be on average 75 Å, with a broad range over approximately 57-143 Å (Zheng *et al.*, 1992) (for a planar extended IgG1 the estimated end-to-end length is in the range of 120-170 Å (Dangl *et al.*, 1988)). Zheng and co-workers demonstrated that upon binding of the IgG1 Fc domain to the FcγR1 receptor the flexibility of the antibody reduced with an apparent end-to-end distance of 85 Å. From the experimental evidence reported previously in this chapter it is tantalising to propose that this inherent flexibility of IgG antibodies could play a role in signalling. Rearrangement of the variable domains upon antigen binding has been shown to induce a conformational change at the V_L-C_L interface (figure 3.23) and this may either alter the flexibility of the antibody directly, or reveal a binding site possibly close to the V_L-C_L interface allowing interaction with either the Fc domain or a membrane associated signalling protein. This hypothesis would go someway to explaining how a membrane bound antibody can activate a B-cell upon monomeric antigen binding without a conformational change being passed through the entire antibody.

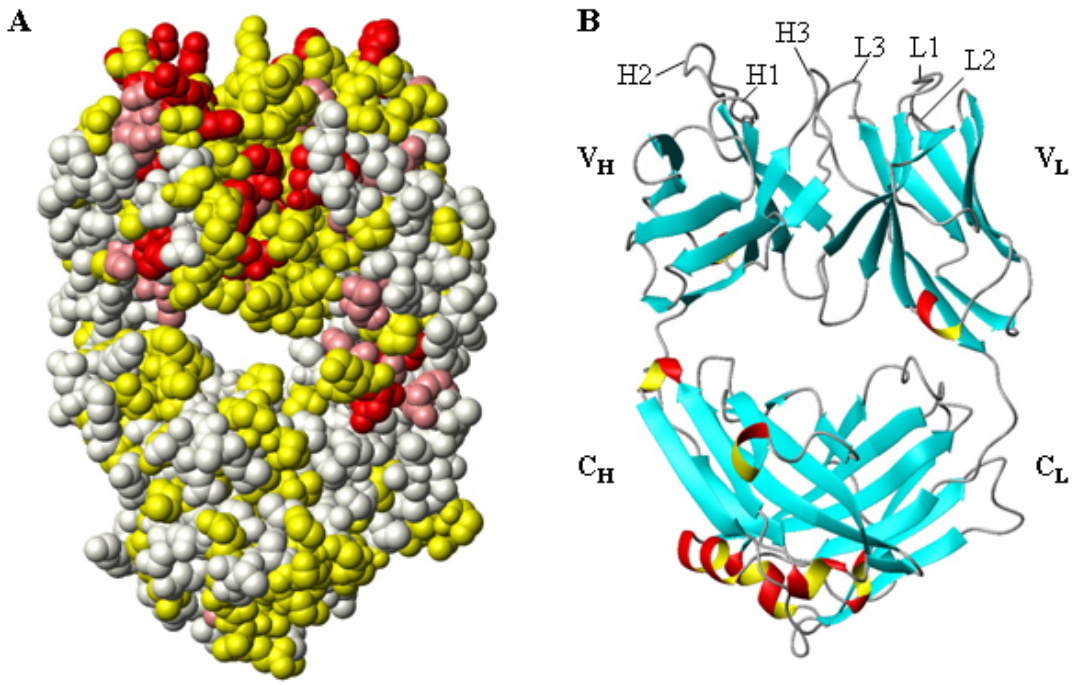

Figure 3.23 Mapping of the IL-1β binding site on the Fab IC8 determined by combined minimal shift from $^{15}N/^{13}C/^{1}H$ HNCO spectra for the free and IL-1β bound Fab IC8. This data was collected and analysed by Catherine Hall. Panel A shows a space filled model of the Fab with residues that have significantly perturbed signals coloured in a gradient from white to red with yellow residues corresponding to those for which no minimal shift data is available. Panel B displays a ribbon representation of the four domains of the Fab in the same orientation as panel A, with the CDRs numbered appropriately. Residues are significantly perturbed at the Fab-IL-1β, V_H-V_L and V_L-C_L interfaces.

Structural studies on the molecular evolution of antibody binding sites, in particular for the anti-fluorescein antibody 4-4-20, have demonstrated using molecular dynamics that antibodies become less flexible as they incorporate somatic mutations to evolve from germline to affinity matured antibodies (Thorpe & Brooks, 2007; Zimmermann *et al.*, 2006). Some of the most significant changes in binding kinetics and thermodynamics

resulted from mutations at residues distant from the antibody binding site, suggesting that they altered the protein structure or dynamics and increased affinity for the antigen. To the best of my knowledge the only none *in silico* structural evidence of this are the free and hapten bound crystal structures of the affinity matured antibody 48G7 and its germline precursor (Wedemayer *et al.*, 1997). Comparison of the free Fab and Fab-hapten complex showed no significant rearrangement of the variable domains upon complex formation. However, for the free and hapten bound structures of the germline antibody the relative domain association changed by 4.6 º, suggesting that affinity matured antibodies are less inducible than the germline precursors. It is unclear whether these differences arise from direct interaction of the ligand (induced fit) or by preferential binding and selection of a preexisting subpopulation of antibody conformers. This distinction is of significant immunological consequence as two or more conformers in spontaneous equilibrium would form part of the humoral repertoire, whereas an induced conformation existing only once part of a complex would not contribute to diversity in the same way (Foote & Milstein, 1994). The lack of variable domain reorientation for the affinity matured 48G7 antibody upon binding to its antigen may be due to the small size of the hapten. IC8 is an affinity matured antibody that binds to a target protein and upon doing so clearly shows signs of domain reorientation, suggesting that this feature is not just a means of increasing antibody variability but potentially serves as an additional signalling mechanism in all antibodies.

Comparison of the assignments made for the scFv IC8-IL-1β complex with those of the free Fab IC8 also revealed an interesting feature of the CDRs. Essentially complete assignment of the CDRs residues was obtained in the scFv complex, whereas in the free assigned Fab and unassigned scFv spectra a subset of signals were missing. Overlay of

the $^{15}N/^{13}C/^{1}H$ HNCO spectra of $^{15}N/^{13}C/^{2}H$ labelled IC8 scFv-unlabelled IL-1β complex, $^{15}N/^{13}C/^{2}H$ labelled Fab IC8-unlabelled IL-1β complex and $^{15}N/^{13}C/^{2}H$ labelled Fab IC8 clearly reavealed that approximately 50% of the assigned CDR signals were missing in the spectrum of the free antibody. This feature can almost certainly be attributed to chemical exchange broadening, where the CDR loops that are essentially invisible are in a dynamic exchange between two or more states. At an intermediate exchange rate this leads to a severe broadening of the NMR signals to the extent that they are no longer visible. Therefore, this data indicates that some of the CDR loops, in this case the longer CDR 3 loops of both variable domains in particular, are in exchange between two or more states, while the remaining CDRs have an apparently more rigid structure.

3.5 Conclusion

The work reported in this chapter has shown that high quality NMR data can be collected for a perdeuterated scFv IC8-IL-1β complex, enabling comprehensive assignment of the backbone residues and long-range NOEs. Analysis of the backbone amide chemical shifts upon complex formation was used to map the binding site on each protein in the complex, which in combination with backbone amide RDCs and NOEs was used to drive a restraint-driven docking procedure to successfully produce a docked scFv IC8-IL-1β complex that satisfied all the NMR data.

The backbone amide chemical shifts for the IC8 scFv and Fab upon complex formation with IL-1β show that the signals of residues located at the V_H-V_L and V_L-C_L interfaces are significantly perturbed in addition to those at the IL-1β interaction site. This suggests that a rotation of the V_L domain occurs upon the antibody binding to IL-1β. This could potentially give an insight into antibody selection from the surface of B-cells, with rearrangement of the V_L domain perhaps signalling for selection of the antibody.

Analysis of the actual combined backbone shifts for IL-1β upon binding to either the IC8 scFv or Fab revealed that the two equivalent antibody fragments show very similar binding patterns as expected and thus the scFv can be used as a reliable model for larger therapeutic antibodies. However, very subtle differences at the binding site were seen that would most likely only be detected by this technique.

Chapter 4

Conclusions

Therapeutic antibodies are currently the fastest growing area of biopharmaceuticals (Pavlou & Belsey, 2005) and humanisation of the variable domains to reduce their immunogenicity is clearly vital to their success in the clinic. Humanisation of antibodies is currently performed either based solely on the assumption that antibodies interact with antigens via residues in the CDRs or in combination with data from indirect techniques such as site-directed mutagenesis or analysis of backbone amide exchange by mass spectrometry (Ehring, 1999; Prasad *et al.*, 1993). While supplying some information about the binding epitope on the antibody and target protein, these techniques do no offer the detailed information provided by a high resolution structure. X-ray crystallography has provided some success in this area but the inherent flexibility and solubility of antibodies probably contributes to the apparent limited success in crystallisation trials. Being a solution method NMR spectroscopy does not suffer from this problem but instead size becomes the limitation. For this reason the work I have presented in this thesis set out to investigate the feasibility of utilising the smallest fragment of an antibody, the scFv, to study the structure of therapeutic antibodies and the complexes formed with their target proteins by NMR.

Studies of scFvs have previously shown that they are prone to the formation of multimers via a monomer to domain-swapped dimer equilibrium (Desplancq *et al.*, 1994; Essig *et al.*, 1993; Holliger *et al.*, 1993). The work reported in chapter 2 shows the formation of domain-swapped dimers for two anti-IL-1β scFvs, with the relative

amounts of dimer increasing with protein concentration. At the concentrations required for NMR studies (>200 μM) the amount of domain-swapped dimer approached nearly 50%, which is similar to the results obtained for the only other study of scFvs at relatively high concentrations (Lee *et al.*, 2002). This reflects the weak association of the V_L and V_H domains, resulting in the formation of open monomers and upon increasing protein concentration the likelihood of two open forms of the monomer associating to produce a domain-swapped scFv dimer increases.

The length of the linker between the variable domains is a crucial factor in scFv dimer formation, with linkers shorter than 15 residues causing a steric hindrance preventing the formation of monomeric scFv (Holliger *et al.*, 1993; Perisic *et al.*, 1994). It was expected that the $(Gly_4Ser)_4$ linkers engineered in the anti-IL-1β scFvs would be sufficient to allow stable folding of the monomer, which was true but domain-swapped dimers still formed at higher concentrations as described above. Construction of scFvs with extended $(Gly_4Ser)_5$ linkers had no effect on the monomer to dimer equilibrium, suggesting that the linker was not contributing to the instability of the monomer. Perturbation of the equilibrium by generating mutant constructs with a specific Ala to Asp residue change within the hydrophobic patch thought to enable formation of the dimer also proved unsuccessful. This specific residue had been identified by modelling studies at Celltech in the early 1990s. The high resolution structure, shown in figure 3.21, highlights the location of the mutation within the α-helix towards the rear of the V_H domain. The mutation was expected to prevent the back-to-back interactions between the variable domains that lead to the formation of domain-swapped dimer. However, it appears that the mutation is too far from this site to impact upon the potential interactions that occur within the dimer.

$^{15}N/^{1}H$ HSQC spectra from samples of the free scFv containing almost 50% domain-swapped dimer showed significantly reduced sensitivity and resolution compared to comparable size proteins. The broad linewidths from scFv samples has been cited as a problem previously (Freund *et al.*, 1993; Freund *et al.*, 1994) and it would appear that the collection of NMR data of the quality required for determination of high resolution structures for free scFvs would be very challenging, if not impossible, which clearly explains why no NMR structures of scFv fragments have previously been reported.

Formation of the scFv IC8-IL-1β complex stabilised the bound scFv and allowed subsequent concentration of the complex for detailed NMR analysis without the formation of the domain-swapped scFv dimers. The work reported in chapter 3 described the essentially complete assignment of the backbone chemical shifts (NH, N, Cα, Cβ, C') for both the scFv IC8 and IL-1β enabling analysis of the chemical shifts perturbed upon formation of the complex. In combination with backbone amide RDCs and NH to NH NOEs this enabled the docking and refinement of a reliable model for the scFv-IL-1β complex. The final set of converged scFv IC8-IL-1β complexes (PDB code 2KH2) are entirely consistent with the NMR-derived constraints used for docking, with no significant or consistent violations and the structure is typical of that expected for an antibody-target protein complex, which clearly indicates that the approach described in this thesis can produce reliable structures for scFv-target protein complexes. A comparison of the $^{15}N/^{1}H$ HSQC spectra for ^{15}N labelled IL-1β bound to equivalent scFv and Fab IC8 fragments reveals only very subtle changes in the positions of some peaks. This indicates that scFvs bind in an essentially identical manner to equivalent Fabs thus making scFvs excellent models for larger fragments or full antibodies. I would, therefore, expect this NMR restraint-driven docking procedure to

be widely utilised in the future to provide a knowledge based approach for the design and humanisation of candidate therapeutic antibodies.

Displacement of the antibody variable domains relative to each other upon binding to an antigen has been reported previously (Bhat *et al.*, 1990; Rini *et al.*, 1992; Stanfield *et al.*, 1990; Stanfield *et al.*, 1993). This movement at the V_H-V_L interface was also evident on analysis of the chemical shift perturbation data for the free and IL-1β bound scFv IC8. No previous studies of antibody-target protein crystal structures have reported evidence of structural rearrangements at any other locations in the antibody upon binding to the antigen. However, analysis of the backbone amide shifts seen upon formation of the Fab IC8-IL-1β complex clearly showed that the signals from residues located at the V_L-C_L interface are significantly perturbed in addition to those signals at the IL-1β and V_H-V_L interfaces that were also seen for the scFv IC8-IL-1β complex. This data suggests that a rotation of the V_L domain occurs upon binding leading to potentially very subtle structural rearrangements at the V_H-V_L and V_L-C_L interfaces.

It is beyond the scope of this thesis to determine the implications of this domain rearrangement upon antigen binding but it is perhaps a mechanism for signaling that the antibody has bound to its antigen. This would allow B-cells to determine when a monomeric antigen has bound to a membrane bound antibody and initiate the activation of the B-cell and internalisation of the antigen for degradation and presentation. This activating signal could be caused by binding of a site at the V_L-C_L interface revealed by the rearrangement of the domains to either a signaling protein associated with the membrane bound antibody or to the Fc domain of the antibody. This proposal is highly speculative at the moment but further investigation may eventually lead to a better

understanding of the signaling mechanisms associated with membrane-bound antibody receptors.

Appendix

A.1 Media and reagents

A.1.1 LB Broth

10 g NaCl, 10 g tryptone and 5 g yeast extract per litre. Make up to one litre with water and autoclave. Once cool add desired antibiotic. For LB agar add 15 g bacteriological agar prior to autoclaving.

A.1.2 Minimal media

Minimal media for the isotopic labelling of expressed proteins was prepared as follows per litre of media:

- 1 g $(NH_4)_2SO_4$ (1)
- 100 ml of 10X PO_4/NaCl stock solution made as follows:
  - 17 g Na_2HPO_4
  - 7.5 g KH_2PO_4
  - 1.25 g NaCl
  - Make up to 250 ml with water (2)
- 100 ml of 100X EDTA trace elements stock solution made up as follows:
  - 1 g EDTA
  - Dissolve in 160 ml water (2) and adjust to pH 7 with NaOH
  - 0.32 g $MnCl_2$
  - 0.1 g $FeCl_3$
  - 10 mg $ZnCl_2$
  - 2 mg $CuCl_2$
  - 2 mg $CoCl_2$
  - 2 mg H_3BO_3
  - Make up to 200 ml with water (2) and adjust to pH 7
- Add 864 ml of water (2), autoclave and leave to cool
- 1 ml of 1000X $MgSO_4$ stock solution made as follows:
  - 24.6 $MgSO_4.7H_20$ in 100 ml of water (2)
- 1 ml of 1000X $CaCl_2$ stock solution made as follows:
  - 4.4 g $CaCl_2.2H_20$ in 100 ml of water (2)

- 1 ml of 1000X d-biotin stock solution made as follows:
  - 10 mg d-biotin in 5 ml water (2) and 5 ml ethanol

- 1 ml of 1000X thiamine stock solution made as follows:
  - 10 mg in 10 ml of water (2)

- 20 ml of 50X autoclaved glucose (3) giving 4 g/l made as follows:
  - 20 g glucose in 100 ml of water (2)

- 1 ml of 1000X antibiotic

(1) For ^{15}N labelled samples the ammonium sulphate was replaced with >98% $(^{15}NH_4)_2SO_4$ (Sigma).

(2) For ^{13}C labelled samples the glucose was replaced with 2 g/l >99% ^{13}C D-glucose (Sigma).

(3) For ^{2}H labelled samples the water was replaced with >99.8% D_2O (Sigma).

A.1.3 Tetracycline

Stock solution was prepared at 10 mg/ml in 50% ethanol, filter sterilised through a 0.2 μM syringe filter and stored at 4 ^{o}C. Add to media as required to a final concentration of 10 μg/ml.

A.1.4 Carbenicillin

Stock solution was prepared at 100 mg/ml in dH_2O, filter sterilised through a 0.2 μM syringe filter and stored at -20 ^{o}C. Add to media as required to a final concentration of 100 μg/ml.

A.2 Chemical shift data

A.2.1 Sequence specific assignments for scFv IC8 bound to IL-1β

1	Asp	C	172.71
		CA	53.01
		CB	39.11
2	Ile	C	175.58
		CA	60.52
		CB	36.74
		H	9.11
		N	123.18
3	Gln	C	175.59
		CA	54.45
		CB	29.73
		H	8.98
		N	128.61
4	Met	C	176.55
		CA	52.72
		CB	32.26
		H	9.16
		N	124.59
5	Thr	C	174.23
		CA	61.90
		CB	70.76
		H	9.52
		N	119.82
6	Gln	C	175.61
		CA	54.35
		CB	30.32
		H	10.01
		N	128.55
7	Ser	CA	55.51
		CB	64.73
		H	9.01
		N	116.89
9	Ser	C	175.56
		CA	60.92
10	Ser	C	171.96
		CA	56.97
		CB	64.56
		H	7.82
		N	115.34
11	Leu	C	174.22
		CA	54.62
		CB	44.57
		H	9.08
		N	123.93
12	Ser	C	173.65
		CA	56.06
		CB	63.78
		H	8.55
		N	118.71
13	Ala	C	174.96
		CA	50.66
		CB	22.37
		H	8.88
		N	127.95
14	Ser	C	173.85
		CA	57.07
		CB	64.26
		H	8.69
		N	116.23
15	Val	C	177.99
		CA	64.66
		CB	30.34
		H	8.51
		N	122.68
16	Gly	C	174.86
		CA	44.03
		H	10.08
		N	117.95
17	Asp	C	174.67
		CA	54.58
		CB	40.90
		H	8.02
		N	122.23
18	Arg	C	176.77
		CA	54.35
		CB	30.37
		H	8.19
		N	120.41
19	Val	C	173.78
		CA	59.66
		CB	34.59
		H	8.47
		N	122.96
20	Thr	C	173.21
		CA	61.09
		CB	71.19
		H	8.11
		N	118.35

21	Ile	C	174.71
		CA	60.54
		CB	39.97
		H	9.26
		N	126.44
22	Thr	C	173.92
		CA	62.05
		CB	70.55
		H	9.02
		N	120.04
23	Cys	C	172.00
		CA	55.87
		CB	47.22
		H	9.79
		N	126.71
24	Arg	C	175.85
		CA	53.91
		CB	32.48
		H	9.33
		N	127.94
25	Thr	C	175.51
		CA	59.53
		CB	70.03
		H	8.77
		N	112.29
26	Ser	C	173.98
		CA	59.13
		CB	63.02
		H	8.16
		N	113.17
27	Gly	CA	43.41
		H	7.22
		N	106.95
28	Asn	C	175.08
		CA	53.60
		CB	36.97
29	Ile	CA	59.75
		CB	38.30
		H	8.09
		N	118.87
30	His	C	173.59
		CA	58.11
		CB	26.94
		H	8.16
		N	119.05
31	Asn	C	175.90
		CA	51.55
		CB	37.05
		H	7.57
		N	118.43
32	Tyr	C	173.66
		CA	57.05
		CB	33.23
		H	8.37
		N	122.31
33	Leu	C	176.29
		CA	51.87
		CB	44.78
		H	7.73
		N	124.17
34	Thr	C	171.64
		CA	59.52
		CB	71.73
		H	9.19
		N	124.58
35	Trp	C	175.58
		CA	55.41
		H	8.69
		N	121.58
36	Tyr	C	174.47
		CA	56.04
		CB	63.88
		H	9.59
		N	117.83
37	Gln	C	173.92
		CA	53.05
		CB	33.42
		H	9.14
		N	122.20
38	Gln	C	174.91
		CA	54.29
		CB	31.05
		H	9.46
		N	127.71
39	Lys	CA	54.96
		CB	30.31
		H	9.56
		N	133.53
44	Pro	C	172.98
		CA	61.91
		CB	29.23
45	Gln	C	175.73
		CA	53.16
		CB	32.03
		H	8.70
		N	115.49
46	Leu	C	174.73
		CA	56.49
		CB	41.67
		H	8.90

46	Leu	N	126.29
47	Leu	C	175.07
		CA	54.89
		CB	45.57
		H	8.47
		N	125.98
48	Ile	C	173.90
		CA	58.05
		CB	43.45
		H	7.31
		N	116.63
49	Tyr	C	174.00
		CA	54.84
		CB	38.98
		H	8.90
		N	122.56
50	Asn	C	178.27
		CA	55.14
		CB	36.19
		H	6.69
		N	114.30
51	Ala	C	178.15
		CA	57.39
		CB	18.55
		H	9.01
		N	114.42
52	Lys	C	176.19
		CA	55.37
		CB	34.79
		H	8.24
		N	115.30
53	Thr	C	172.25
		CA	62.99
		CB	68.24
		H	8.43
		N	122.35
54	Leu	C	177.30
		CA	55.41
		CB	41.71
		H	8.35
		N	128.17
55	Ala	C	176.70
		CA	51.12
		CB	17.92
		H	7.89
		N	124.40
56	Asp	C	177.84
		CA	56.01
		CB	40.15
		H	8.51
		N	121.00
57	Gly	C	174.36
		CA	45.13
		H	8.74
		N	112.20
58	Val	CA	60.41
		CB	31.81
		H	7.60
		N	124.94
59	Pro	C	177.05
		CA	62.81
		CB	32.85
60	Ser	C	174.35
		CA	60.07
		CB	62.50
		H	8.41
		N	113.38
61	Arg	C	175.71
		CA	56.49
		CB	28.21
		H	7.07
		N	116.35
62	Phe	C	174.30
		CA	58.10
		CB	39.82
		H	7.75
		N	121.02
63	Ser	CA	57.00
		CB	65.14
		H	8.66
		N	111.64
64	Gly	C	172.48
		CA	43.52
		H	9.06
		N	111.52
65	Ser	C	174.25
		CA	56.74
		CB	65.58
		H	8.85
		N	114.08
66	Gly	C	171.97
		CA	44.71
		H	8.33
		N	110.92
67	Ser	C	174.52
		CA	58.04
		CB	64.21
		H	6.89
		N	108.96

68	Gly	C	172.40
		CA	47.03
		H	8.80
		N	112.32
69	Thr	C	174.92
		CA	59.91
		CB	70.73
		H	7.69
		N	112.79
70	Gln	C	172.75
		CA	55.93
		CB	30.59
		H	6.77
		N	121.16
71	Phe	C	175.48
		CA	56.44
		CB	41.92
		H	8.87
		N	124.36
72	Thr	C	172.63
		CA	60.13
		CB	72.14
		H	9.01
		N	114.32
73	Leu	C	171.97
		CA	52.99
		CB	41.15
		H	8.86
		N	127.71
74	Thr	C	173.22
		CA	60.75
		CB	70.32
		H	9.00
		N	123.32
75	Ile	C	176.30
		CA	59.74
		CB	37.38
		H	8.74
		N	126.46
76	Ser	C	173.66
		CA	60.12
		CB	62.44
		H	8.82
		N	121.20
77	Ser	C	173.86
		CA	56.48
		CB	62.02
		H	6.44
		N	113.40
78	Leu	C	176.90
		CA	56.40
		CB	42.06
		H	8.76
		N	126.00
79	Gln	CA	54.48
		CB	28.47
		H	9.16
		N	124.91
80	Pro	C	178.35
		CA	66.11
		CB	30.97
81	Glu	C	176.36
		CA	57.41
		CB	27.75
		H	9.46
		N	114.32
82	Asp	C	177.49
		CA	54.34
		CB	40.89
		H	8.34
		N	119.93
83	Phe	C	174.16
		CA	59.28
		CB	36.86
		H	7.12
		N	121.53
84	Ala	C	174.35
		CA	51.16
		CB	20.08
		H	8.05
		N	128.15
85	Asn	C	174.87
		CA	51.44
		CB	40.02
		H	7.97
		N	115.82
86	Tyr	CA	56.88
		H	8.67
		N	118.97
89	Gln	C	174.37
		CA	55.05
		CB	32.03
90	His	C	173.98
		CA	53.97
		CB	34.55
		H	7.80
		N	125.44
91	Phe	C	173.83
		CA	54.23

91	Phe	CB	35.96
		H	7.62
		N	119.00
92	Trp	C	173.59
		CA	58.23
		CB	31.41
		H	10.11
		N	124.11
93	Ser	C	172.51
		CA	54.97
		CB	64.45
		H	7.16
		N	107.36
94	Leu	CA	52.84
		CB	40.00
		H	8.14
		N	118.83
95	Pro	C	175.82
		CA	62.83
		CB	33.21
96	Phe	C	175.43
		CA	55.01
		CB	35.46
		H	6.50
		N	119.99
97	Thr	C	171.94
		CA	59.76
		CB	72.18
		H	6.59
		N	113.46
98	Phe	C	178.81
		CA	56.31
		CB	43.09
		H	8.63
		N	116.04
99	Gly	C	172.77
		CA	44.47
		H	9.29
		N	106.63
100	Gln	C	175.86
		CA	57.11
		CB	26.72
		H	9.21
		N	117.95
101	Gly	C	172.00
		CA	45.06
		H	7.06
		N	108.15
102	Thr	C	173.79
		CA	61.72
		CB	72.65
		H	7.91
		N	117.57
103	Lys	C	174.66
		CA	55.77
		CB	32.78
		H	8.18
		N	129.08
104	Val	C	174.81
		CA	61.86
		CB	31.71
		H	8.65
		N	128.40
105	Glu	C	174.38
		CA	54.13
		CB	32.31
		H	8.97
		N	126.39
106	Ile	C	176.49
		CA	59.92
		CB	37.76
		H	7.66
		N	121.86
107	Lys	CA	56.34
		CB	31.94
		H	8.30
		N	128.26
110	Gly	CA	44.25
		H	8.39
		N	109.42
111	Gly	CA	44.25
		H	8.39
		N	109.42
112	Gly	CA	44.25
		H	8.39
		N	109.42
113	Gly	CA	44.60
		H	8.39
		N	109.42
114	Ser	CA	57.90
		CB	63.78
		H	8.26
		N	115.62
115	Gly	CA	44.25
116	Gly	CA	44.25
		H	8.39
		N	109.42

117	Gly	CA	44.25
		H	8.39
		N	109.42
118	Gly	CA	44.60
		H	8.39
		N	109.42
119	Ser	CA	57.90
		CB	63.78
		H	8.26
		N	115.62
120	Gly	CA	44.25
121	Gly	CA	44.25
		H	8.39
		N	109.42
122	Gly	C	44.25
		CA	44.25
		H	8.39
		N	109.42
123	Gly	CA	44.60
		H	8.39
		N	109.42
124	Ser	CA	57.90
		CB	63.78
		H	8.26
		N	115.62
125	Gly	CA	44.25
126	Gly	CA	44.25
		H	8.39
		N	109.42
127	Gly	CA	44.25
		H	8.39
		N	109.42
128	Gly	C	174.26
		CA	44.77
		H	8.39
		N	109.42
129	Ser	C	174.68
		CA	57.74
		CB	63.78
		H	8.26
		N	115.57
130	Glu	C	176.25
		CA	56.81
		CB	29.39
		H	8.68
		N	123.57
131	Val	C	175.50
		CA	62.81
		CB	30.95
		H	7.63
		N	120.78
132	Gln	C	173.99
		CA	54.86
		CB	31.56
		H	8.42
		N	125.61
133	Leu	C	175.54
		CA	53.45
		CB	43.05
		H	8.31
		N	124.49
134	Val	C	176.45
		CA	61.99
		CB	33.85
		H	8.58
		N	121.13
135	Glu	C	176.90
		CA	56.74
		CB	29.79
		H	10.45
		N	132.54
136	Ser	C	173.47
		CA	57.72
		CB	65.42
		H	9.45
		N	115.28
137	Gly	C	174.44
		CA	44.47
		H	8.57
		N	108.01
138	Gly	C	173.01
		CA	45.16
		H	7.59
		N	105.99
139	Gly	C	170.94
		CA	44.46
		H	7.43
		N	107.78
140	Leu	C	176.84
		CA	54.11
		CB	42.23
		H	8.07
		N	122.95
141	Val	C	173.17
		CA	59.44
		CB	34.46

141	Val	H	9.05
		N	124.70
142	Gln	CA	53.49
		CB	27.25
		H	8.19
		N	122.90
143	Pro	C	177.76
		CA	63.35
		CB	30.15
144	Gly	C	175.27
		CA	44.64
		H	9.80
		N	114.24
145	Gly	C	181.66
		CA	44.19
		H	8.39
		N	109.36
146	Ser	C	173.04
		CA	56.17
		CB	66.49
		H	7.87
		N	110.25
147	Leu	C	173.35
		CA	54.49
		CB	46.58
		H	8.58
		N	122.50
148	Arg	C	175.70
		CA	54.12
		CB	31.53
		H	8.09
		N	124.02
149	Leu	C	176.86
		CA	52.73
		CB	42.94
		H	9.05
		N	127.12
150	Ser	C	173.30
		CA	56.39
		CB	66.23
		H	8.81
		N	113.83
151	Cys	C	171.16
		CA	53.32
		CB	41.64
		H	9.30
		N	123.02
152	Ala	C	176.39
		CA	50.46
		CB	19.45
		H	8.67
		N	130.53
153	Ala	C	176.55
		CA	50.12
		CB	23.21
		H	8.24
		N	127.26
154	Ser	C	174.14
		CA	57.54
		CB	65.22
		H	8.93
		N	114.61
155	Gly	C	173.37
		CA	45.30
		H	8.66
		N	108.21
156	Phe	CA	54.30
		CB	39.17
		H	7.35
		N	112.24
157	Asp	C	175.33
		CA	52.73
		CB	38.19
158	Phe	C	175.80
		CA	60.57
		CB	39.48
		H	7.77
		N	129.32
159	Ser	C	173.28
		CA	60.46
		CB	62.94
		H	8.26
		N	110.57
160	Arg	C	176.72
		CA	54.81
		CB	29.92
		H	7.68
		N	116.84
161	Tyr	C	174.67
		CA	57.97
		CB	40.96
		H	7.65
		N	117.45
162	Asp	C	173.21
		CA	53.70
		CB	39.52
		H	9.66
		N	129.06

163	Met	C	174.47
		CA	51.94
		CB	35.71
		H	8.45
		N	117.73
164	Ser	C	174.33
		CA	57.85
		CB	63.68
		H	9.08
		N	115.12
165	Trp	CA	57.73
		C	175.84
		CA	55.86
		CB	29.74
		H	8.46
		N	123.04
166	Val	C	173.63
		CA	60.30
		CB	34.71
		H	7.83
		N	122.85
167	Arg	C	173.98
		CA	52.20
		CB	32.63
		H	9.74
		N	123.99
168	Gln	C	174.92
		CA	54.65
		CB	31.39
		H	9.23
		N	120.86
169	Ala	CA	49.86
		CB	17.44
		H	9.81
		N	137.52
170	Pro	C	178.39
		CA	64.03
		CB	30.49
171	Gly	C	174.39
		CA	45.47
		H	8.70
		N	112.95
172	Lys	C	176.28
		CA	54.17
		CB	34.08
		H	8.07
		N	119.70
173	Arg	C	176.96
		CA	55.45
		CB	30.65
		H	8.08
		N	118.52
174	Leu	C	175.20
		CA	55.94
		CB	40.51
		H	7.74
		N	123.22
175	Glu	C	175.50
		CA	54.50
		CB	32.90
		H	8.99
		N	118.83
176	Trp	CA	55.76
		CB	29.00
		H	9.81
		N	131.30
177	Val	C	172.78
		CA	61.75
		CB	33.21
178	Ala	C	174.69
		CA	52.03
		CB	24.23
		H	7.56
		N	115.85
179	Tyr	C	173.15
		CA	56.69
		CB	42.92
		H	7.89
		N	124.03
180	Ile	C	172.05
		CA	58.36
		CB	40.06
		H	8.64
		N	125.33
181	Ser	C	178.17
		CA	58.04
		CB	63.76
		H	8.32
		N	121.72
182	Ser	C	173.58
		CA	60.80
		CB	63.80
		H	8.71
		N	118.90
183	Gly	C	176.00
		CA	45.44
		H	7.38
		N	104.26

184	Gly	C	175.03
		CA	43.58
		H	7.78
		N	108.08
185	Gly	C	172.28
		CA	45.86
		H	9.16
		N	109.61
186	Ser	C	173.38
		CA	56.09
		CB	64.90
		H	7.95
		N	113.78
187	Thr	C	173.82
		CA	59.52
		CB	74.84
		H	8.18
		N	110.35
188	Tyr	C	171.33
		CA	57.87
		CB	41.06
		H	7.84
		N	117.67
189	Phe	CA	53.74
		CB	41.33
		H	7.06
		N	122.22
190	Pro	C	177.20
		CA	61.70
		CB	31.95
191	Asp	C	177.21
		CA	57.44
		CB	39.80
		H	8.99
		N	120.30
192	Thr	C	176.48
		CA	63.29
		CB	68.03
		H	7.50
		N	107.79
193	Val	C	175.15
		CA	60.42
		CB	31.26
		H	6.89
		N	109.99
194	Lys	C	176.96
		CA	57.90
		CB	31.36
		H	7.43
		N	124.46
195	Gly	C	174.35
		CA	44.94
		H	9.09
		N	116.05
196	Arg	C	175.37
		CA	56.84
		CB	29.67
		H	7.70
		N	117.42
197	Phe	CA	52.45
		CB	39.01
		H	7.23
		N	119.43
198	Thr	C	172.64
		CA	60.79
		CB	71.38
		H	9.06
		N	111.65
199	Ile	C	172.04
		CA	58.66
		CB	42.78
		H	9.00
		N	123.33
200	Ser	C	171.85
		CA	57.41
		CB	65.12
		H	8.39
		N	117.85
201	Arg	C	172.78
		CA	54.33
		CB	34.74
		H	9.38
		N	116.56
202	Asp	C	177.18
		CA	51.98
		CB	41.52
		H	9.25
		N	121.23
203	Asn	C	177.16
		CA	54.98
		CB	36.56
		H	9.45
		N	124.43
204	Ala	C	179.06
		CA	53.64
		CB	17.83
		H	8.52
		N	122.39
205	Lys	C	176.43
		CA	54.58

205	Lys	CB	32.28
		H	7.36
		N	115.27
206	Asn	C	172.26
		CA	53.76
		CB	36.02
		H	7.87
		N	119.43
207	Thr	C	171.18
		CA	60.67
		CB	72.49
		H	7.19
		N	108.41
208	Leu	C	173.74
		CA	53.20
		CB	45.87
		H	8.85
		N	126.76
209	Tyr	C	175.05
		CA	56.67
		CB	42.84
		H	9.25
		N	121.81
210	Leu	CA	52.61
		CB	44.41
		H	7.67
		N	116.03
211	Gln	CA	54.31
		CB	27.97
212	Met	C	173.67
		CA	54.32
		CB	34.44
		H	8.74
		N	126.12
213	Asn	C	173.31
		CA	51.17
		CB	40.89
		H	7.96
		N	120.00
214	Ser	C	175.80
		CA	57.24
		CB	61.48
		H	8.49
		N	111.91
215	Leu	C	177.20
		CA	56.21
		CB	42.00
		H	8.16
		N	118.92
216	Arg	C	176.95
		CA	53.34
		CB	33.04
		H	9.67
		N	121.27
217	Ala	C	180.75
		CA	55.69
		CB	17.25
		H	9.12
		N	125.56
218	Glu	C	176.05
		CA	58.25
		CB	27.38
		H	9.28
		N	114.89
219	Asp	C	176.92
		CA	54.72
		CB	41.43
		H	8.55
		N	118.11
220	Thr	C	174.43
		CA	64.23
		CB	69.62
		H	7.88
		N	120.76
221	Ala	C	175.69
		CA	52.38
		CB	20.35
		H	9.24
		N	127.39
222	Val	C	175.25
		CA	62.43
		H	7.71
		N	119.12
223	Tyr	C	180.38
		CA	58.23
		H	9.29
		N	127.55
224	Tyr	CA	57.21
		H	6.91
		N	112.48
225	Cys	C	175.76
		CA	53.16
		CB	34.21
226	Ala	C	174.82
		CA	50.18
		CB	18.76
		H	8.23
		N	122.79

227	Arg	C	173.86
		CA	55.12
		CB	33.47
		H	8.69
		N	125.13
228	Gln	C	174.65
		CA	54.31
		CB	30.36
		H	8.16
		N	128.59
229	Asn	C	174.47
		CA	53.66
		CB	37.11
		H	8.58
		N	124.12
230	Lys	C	174.77
		CA	56.45
		CB	32.81
		H	7.18
		N	121.85
231	Lys	C	173.50
		CA	56.69
		CB	26.45
		H	9.18
		N	115.28
232	Leu	C	175.98
		CA	55.90
		CB	37.80
		H	9.51
		N	119.94
233	Thr	C	173.96
		CA	58.60
		H	5.92
		N	104.83
234	Trp	C	172.79
		CA	56.55
		CB	28.33
		H	5.66
		N	117.11
235	Phe	C	176.10
		CA	56.43
		CB	37.55
		H	7.54
		N	118.54
236	Asp	C	176.18
		CA	53.95
		CB	40.62
		H	8.51
		N	121.88
237	Tyr	C	172.91
		CA	59.82
		CB	42.20
		H	7.59
		N	120.64
238	Trp	C	177.22
		CA	55.62
		CB	31.43
		H	8.63
		N	123.35
		C	172.75
239	Gly	CA	45.23
		H	8.86
		N	109.22
240	Gln	C	177.87
		CA	56.69
		CB	29.33
		H	8.63
		N	114.16
241	Gly	C	173.54
		CA	43.64
		H	9.37
		N	110.63
242	Thr	C	171.99
		CA	59.59
		CB	70.40
		H	8.44
		N	116.46
243	Leu	C	176.01
		CA	55.62
		CB	41.76
		H	8.35
		N	131.78
244	Val	C	173.96
		CA	61.65
		CB	32.74
		H	8.93
		N	129.38
245	Thr	C	172.24
		CA	61.38
		CB	69.80
		H	8.31
		N	123.05
246	Val	C	176.05
		CA	59.75
		CB	30.13
		H	8.55
		N	128.49
247	Ser	C	176.59
		CA	56.97
		CB	64.28

247	Ser	H	8.45	248	Ser	CA	56.76
		N	119.94			H	9.41
						N	120.63

A.2.2 Sequence specific assignments for IL-1β bound to scFv IC8

2	Pro	C	176.20	11	Arg	C	176.20
		CA	62.44			CA	54.04
		CB	31.14			CB	34.08
						H	8.60
3	Val	C	176.20			N	119.58
		CA	60.39				
		CB	33.20	12	Asp	C	176.20
		H	8.10			CA	52.93
		N	119.03			CB	40.57
						H	8.88
4	Arg	C	176.20			N	121.21
		CA	56.03				
		CB	30.50	13	Ser	C	176.20
		H	8.44			CA	59.99
		N	124.91			CB	61.58
						H	7.60
5	Ser	C	176.20			N	114.54
		CA	57.25				
		CB	63.90	14	Gln	C	176.20
		H	8.12			CA	54.76
		N	118.24			CB	27.66
						H	8.14
6	Leu	C	176.20			N	120.58
		CA	52.77				
		CB	42.47	15	Gln	C	176.20
		H	9.09			CA	57.12
		N	126.57			CB	24.51
						H	8.46
7	Asn	C	176.20			N	110.93
		CA	52.73				
		CB	38.80	16	Lys	C	176.20
		H	9.09			CA	58.46
		N	123.53			CB	29.50
						H	8.52
8	Cys	C	176.20			N	119.97
		CA	55.39				
		CB	31.83	17	Ser	C	176.20
		H	9.38			CA	59.00
		N	114.44			CB	64.71
						H	8.43
9	Gly	C	176.20			N	119.74
		CA	43.46				
		H	8.62	18	Leu	C	176.20
		N	105.94			CA	54.67
						CB	40.99
10	Leu	C	176.20			H	8.50
		CA	53.64			N	122.33
		CB	46.33				
		H	9.16	19	Val	C	176.20
		N	121.79			CA	59.15
						CB	35.55

19	Val	H	8.95
		N	116.58
20	Met	C	176.20
		CA	55.26
		CB	31.86
		H	8.75
		N	121.22
21	Ser	C	176.20
		CA	56.80
		CB	62.62
		H	8.64
		N	120.11
22	Gly	CA	44.15
		H	8.15
		N	111.59
23	Pro	C	176.20
		CA	63.88
		CB	30.79
24	Tyr	C	176.20
		CA	55.81
		CB	38.63
		H	8.06
		N	110.64
25	Glu	C	176.20
		CA	55.67
		CB	33.90
		H	7.27
		N	119.52
26	Leu	C	176.20
		CA	52.83
		CB	44.41
		H	8.61
		N	123.04
27	Lys	C	176.20
		CA	54.78
		CB	36.98
		H	9.24
		N	119.82
28	Ala	C	176.20
		CA	49.44
		CB	20.78
		H	7.60
		N	118.82
29	Leu	C	176.20
		CA	55.03
		CB	44.06
		H	9.41
		N	125.77
30	His	C	176.20
		CA	58.86
		CB	27.80
		H	10.18
		N	122.29
31	Leu	C	176.20
		CA	53.53
		CB	43.16
		H	8.51
		N	126.10
32	Gln	C	176.20
		CA	54.68
		CB	29.95
		H	8.93
		N	120.48
33	Gly	CA	46.23
		H	8.73
		N	109.60
34	Gln	C	176.20
		CA	56.65
		CB	27.72
35	Asp	C	176.20
		CA	54.96
		CB	40.49
		H	8.06
		N	118.98
36	Met	C	176.20
		CA	57.16
		CB	31.25
		H	7.83
		N	118.35
37	Glu	C	176.20
		CA	57.36
		CB	28.62
		H	8.15
		N	118.06
38	Gln	C	176.20
		CA	55.84
		CB	28.78
		H	8.04
		N	116.79
39	Gln	C	176.20
		CA	56.01
		CB	28.60
		H	7.62
		N	118.24
40	Val	C	176.20
		CA	61.96
		CB	31.99
		H	8.48
		N	123.94

41	Val	C	176.20
		CA	61.51
		CB	31.70
		H	7.95
		N	125.63
42	Phe	C	176.20
		CA	57.06
		CB	40.75
		H	9.48
		N	126.70
43	Ser	C	176.20
		CA	56.70
		CB	63.32
		H	9.20
		N	115.58
44	Met	C	176.20
		CA	54.09
		CB	35.55
		H	9.62
		N	131.38
45	Ser	C	176.20
		CA	56.61
		CB	65.68
		H	8.60
		N	119.39
46	Phe	C	176.20
		CA	58.49
		CB	37.81
		H	9.39
		N	124.15
47	Val	C	176.20
		CA	59.47
		CB	33.32
		H	7.48
		N	116.94
48	Gln	C	176.20
		CA	55.91
		CB	28.88
		H	8.54
		N	119.78
49	Gly	C	176.20
		CA	44.71
		H	8.15
		N	111.16
50	Glu	C	176.20
		CA	56.61
		CB	28.95
		H	8.51
		N	122.35
51	Glu	C	176.20
		CA	55.24
		CB	30.99
		H	8.46
		N	124.46
52	Ser	CA	56.88
		CB	64.41
		H	8.48
		N	117.91
53	Asn	C	176.20
		CA	55.21
54	Asp	C	176.20
		CA	54.71
		CB	40.73
		H	8.44
		N	113.50
55	Lys	C	176.20
		CA	55.69
		CB	34.22
		H	7.46
		N	118.89
56	Ile	CA	57.58
		CB	41.26
		H	8.21
		N	121.33
57	Pro	C	176.20
		CA	62.49
		CB	30.86
58	Val	C	176.20
		CA	58.70
		CB	36.70
		H	10.12
		N	118.34
59	Ala	C	176.20
		CA	50.40
		CB	22.83
		H	8.75
		N	120.59
60	Leu	C	176.20
		CA	53.68
		CB	45.01
		H	10.47
		N	123.00
61	Gly	C	176.20
		CA	42.44
		H	8.29
		N	110.34
62	Lys	CA	54.01
		CB	40.66
		H	8.24

62	Lys	N	120.73
66	Asn	C	176.20
		CA	53.93
67	Leu	C	176.20
		CA	52.68
		CB	46.69
		H	7.02
		N	116.10
68	Tyr	C	176.20
		CA	57.09
		CB	42.48
		H	9.20
		N	121.02
69	Leu	C	176.20
		CA	56.48
		CB	41.41
		H	8.62
		N	121.86
70	Ser	C	176.20
		CA	55.80
		CB	66.57
		H	9.19
		N	115.75
71	Cys	C	176.20
		CA	56.48
		CB	29.81
		H	8.48
		N	119.72
72	Val	C	176.20
		CA	58.42
		CB	35.55
		H	8.84
		N	116.91
73	Leu	C	176.20
		CA	54.23
		CB	41.96
		H	8.58
		N	123.60
74	Lys	CA	55.44
		CB	33.42
		H	8.28
		N	126.24
75	Asp	C	176.20
		CA	55.87
		CB	38.92
76	Asp	C	176.20
		CA	55.61
		CB	39.83
		H	8.51
		N	110.06
77	Lys	CA	52.83
		CB	33.99
		H	7.76
		N	119.87
78	Pro	C	176.20
		CA	62.97
		CB	30.59
79	Thr	C	176.20
		CA	61.87
		CB	72.29
		H	8.84
		N	119.65
80	Leu	C	176.20
		CA	53.72
		CB	43.64
		H	8.89
		N	126.82
81	Gln	C	176.20
		CA	53.32
		CB	31.32
		H	9.79
		N	126.22
82	Leu	C	176.20
		CA	53.99
		CB	41.20
		H	8.52
		N	122.03
83	Glu	C	176.20
		CA	54.52
		CB	32.69
		H	9.21
		N	124.53
84	Ser	C	176.20
		CA	58.38
		CB	63.01
		H	8.73
		N	121.55
85	Val	C	176.20
		CA	58.08
		CB	34.47
		H	8.05
		N	117.12
86	Asp	CA	50.74
		CB	41.79
		H	8.05
		N	121.57
87	Pro	C	176.20
		CA	63.65

88	Lys	C	176.20
		CA	57.54
		CB	31.42
		H	8.52
		N	117.21
89	Asn	C	176.20
		CA	52.65
		CB	41.52
		H	7.85
		N	114.24
90	Tyr	CA	56.08
		CB	41.04
		H	7.40
		N	116.04
91	Pro	C	176.20
		CA	62.42
92	Lys	C	176.20
		CA	54.12
		CB	34.17
		H	7.14
		N	115.78
93	Lys	C	176.20
		CA	58.96
		CB	31.80
		H	8.39
		N	119.44
94	Lys	C	176.20
		CA	53.86
		CB	30.11
		H	7.91
		N	117.93
95	Met	C	176.20
		CA	54.07
		CB	33.91
		H	7.57
		N	122.22
96	Glu	C	176.20
		CA	59.31
		CB	27.62
		H	9.27
		N	122.39
97	Lys	C	176.20
		CA	59.53
		CB	31.51
		H	8.25
		N	121.49
98	Arg	C	176.20
		CA	57.98
		CB	27.26
		H	7.94
		N	114.02
99	Phe	C	176.20
		CA	57.60
		CB	40.74
		H	8.01
		N	116.76
100	Val	C	176.20
		CA	63.12
		CB	31.26
		H	7.55
		N	117.83
101	Phe	C	176.20
		CA	55.79
		CB	41.65
		H	9.85
		N	126.90
102	Asn	C	176.20
		CA	51.74
		CB	38.98
		H	10.51
		N	120.83
103	Lys	C	176.20
		CA	56.13
		CB	30.84
		H	9.40
		N	127.63
104	Ile	C	176.20
		CA	60.49
		CB	40.45
		H	9.34
		N	133.46
105	Glu	C	176.20
		CA	55.52
		CB	28.61
		H	8.67
		N	126.72
106	Ile	C	176.20
		CA	59.50
		CB	39.20
		H	8.39
		N	122.87
107	Asn	C	176.20
		CA	54.45
		CB	37.04
		H	8.87
		N	122.09
108	Asn	C	176.20
		CA	54.60
		CB	37.45
		H	8.72

108	Asn	N	112.97
109	Lys	C	176.20
		CA	54.26
		CB	36.50
		H	7.75
		N	116.02
110	Leu	C	176.20
		CA	53.25
		CB	44.84
		H	9.58
		N	121.13
111	Glu	C	176.20
		CA	53.56
		CB	33.11
		H	8.70
		N	117.73
112	Phe	C	176.20
		CA	56.15
		CB	41.34
		H	10.35
		N	119.02
113	Glu	C	176.20
		CA	53.05
		CB	32.67
		H	8.83
		N	125.87
114	Ser	C	176.20
		CA	58.74
		CB	62.65
		H	9.17
		N	121.78
115	Ala	C	176.20
		CA	53.30
		CB	17.58
		H	8.18
		N	128.20
116	Gln	C	176.20
		CA	56.37
		CB	28.62
		H	7.89
		N	116.86
117	Phe	CA	52.97
		CB	38.34
		H	7.45
		N	116.94
118	Pro	C	176.20
		CA	63.07
		CB	31.08
119	Asn	C	176.20
		CA	55.40
		CB	37.69
		H	10.33
		N	114.12
120	Trp	C	176.20
		CA	57.01
		CB	28.55
		H	8.17
		N	119.58
121	Tyr	C	176.20
		CA	56.66
		CB	41.33
		H	9.46
		N	119.07
122	Ile	C	176.20
		CA	65.61
		CB	36.09
		H	8.79
		N	124.66
123	Ser	C	176.20
		CA	57.23
		CB	66.72
		H	9.01
		N	122.12
124	Thr	C	176.20
		CA	58.49
		CB	70.37
		H	9.14
		N	110.14
125	Ser	C	176.20
		CA	56.65
		CB	63.83
		H	9.61
		N	113.61
		C	176.20
126	Gln	CA	58.09
		CB	28.21
		H	8.97
		N	122.08
127	Ala	C	176.20
		CA	51.06
		CB	18.97
		H	8.23
		N	120.88
128	Glu	C	176.20
		CA	56.09
		CB	30.49
		H	8.51
		N	117.30

129	Asn	C	176.20
		CA	54.42
		CB	37.54
		H	8.11
		N	113.91
130	Met	CA	51.64
		CB	31.10
		H	8.92
		N	117.76
131	Pro	C	176.20
		CA	62.58
		CB	32.90
132	Val	C	176.20
		CA	62.47
		CB	31.52
		H	8.05
		N	125.23
133	Phe	C	176.20
		CA	55.38
		CB	41.36
		H	8.80
		N	124.42
134	Leu	C	176.20
		CA	53.02
		CB	43.03
		H	8.72
		N	120.06
135	Gly	C	176.20
		CA	44.73
		H	9.41
		N	114.74
136	Gly	C	176.20
		CA	43.85
		H	8.58
		N	109.96
137	Thr	C	176.20
		CA	61.01
		CB	69.99
		H	7.46
		N	112.48
138	Lys	CB	32.39
		H	8.48
		N	124.64
139	Gly	C	176.20
		CA	45.35
140	Gly	CA	44.40
		H	8.25
		N	109.46
141	Gln	C	176.20
		CA	56.40
		CB	28.90
142	Asp	C	176.20
		CA	54.18
		CB	41.39
		H	7.81
		N	119.17
143	Ile	C	176.20
		CA	61.27
		CB	36.23
		H	9.29
		N	125.67
144	Thr	C	176.20
		CA	59.99
		CB	69.56
		H	8.07
		N	112.10
145	Asp	C	176.20
		CA	52.16
		CB	42.79
		H	7.43
		N	120.06
146	Phe	C	176.20
		CA	56.65
		CB	44.32
		H	9.35
		N	117.25
147	Thr	C	176.20
		CA	60.84
		CB	71.32
		H	9.57
		N	108.57
148	Met	C	176.20
		CA	53.72
		CB	37.98
		H	8.76
		N	120.54
149	Gln	C	176.20
		CA	54.05
		CB	30.18
		H	8.40
		N	126.68
150	Phe	C	176.20
		CA	59.47
		CB	39.34
		H	8.87
		N	124.64
151	Val	C	176.20
		CA	59.50

151	Val	CB	34.93
		H	8.51
		N	116.68
152	Ser	C	176.20
		CA	59.47
		CB	63.43
		H	8.68
		N	117.35
153	Ser	CA	61.04
		CB	63.71
		H	8.33
		N	123.87

References

Adams GP, Schier R, McCall AM, Simmons HH, Horak EM, Alpaugh RK, Marks JD and Weiner LM. (2001). High affinity restricts the localization and tumor penetration of single-chain fv antibody molecules. *Cancer Research*, **61,** 4750-5.

Alfthan K, Takkinen K, Sizmann D, Soderlund H and Terri TT. (1995). Properties of a Single-Chain Antibody Containing Different Linker Peptides. *Protein Engineering*, **8,** 725-731.

Aloy P, Querol E, Aviles FX and Sternberg MJ. (2001). Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. *Journal of molecular biology*, **311,** 395-408.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W and Lipman DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25,** 3389-3402.

Amit AG, Mariuzza RA, Phillips SE and Poljak RJ. (1986). Three-dimensional structure of an antigen-antibody complex at 2.8 A resolution. *Science*, **233,** 747-53.

Andrusier N, Mashiach E, Nussinov R and Wolfson HJ. (2008). Principles of flexible protein-protein docking. *Proteins-Structure Function and Bioinformatics*, **73,** 271-289.

Anthony RM, Nimmerjahn F, Ashline DJ, Reinhold VN, Paulson JC and Ravetch JV. (2008). Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG fc. *Science*, **320,** 373-376.

Arndt KM, Muller KM and Pluckthun A. (1998). Factors influencing the dimer to monomer transition of an antibody single-chain Fv fragment. *Biochemistry*, **37,** 12918-12926.

Ay J, Keitel T, Kuttner G, Wessner H, Scholz C, Hahn M and Hohne W. (2000). Crystal structure of a phage library-derived single-chain Fv fragment complexed with turkey egg-white lysozyme at 2.0 angstrom resolution. *Journal of Molecular Biology*, **301,** 239-246.

Barlow DJ, Edwards MS and Thornton JM. (1986). Continuous and discontinuous protein antigenic determinants. *Nature*, **322,** 747-8.

Bax A. (2003). Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Science*, **12,** 1-16.

Bedzyk WD, Weidner KM, Denzin LK, Johnson LS, Hardman KD, Pantoliano MW, Asel ED and Voss EW, Jr. (1990). Immunological and structural characterization of a high affinity anti-fluorescein single-chain antibody. *Journal of biological chemistry*, **265,** 18615-20.

Bennett MJ, Choe S and Eisenberg D. (1994). Domain Swapping - Entangling Alliances between Proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **91,** 3127-3131.

Bennett MJ, Schlunegger MP and Eisenberg D. (1995). 3d Domain Swapping - a Mechanism for Oligomer Assembly. *Protein Science*, **4,** 2455-2468.

Bentley GA, Boulot G, Riottot MM and Poljak RJ. (1990). Three-dimensional structure of an idiotope-anti-idiotope complex. *Nature*, **348,** 254-7.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE. (2000). The Protein Data Bank. *Nucleic Acids Research*, **28,** 235-242.

Better M, Chang CP, Robinson RR and Horwitz AH. (1988). Escherichia-Coli Secretion of an Active Chimeric Antibody Fragment. *Science*, **240,** 1041-1043.

Bhat TN, Bentley GA, Fischmann TO, Boulot G and Poljak RJ. (1990). Small Rearrangements in Structures of Fv and Fab Fragments of Antibody D1.3 on Antigen-Binding. *Nature*, **347,** 483-485.

Biocca S, Ruberti F, Tafani R, Pierandreiamaldi P and Cattaneo A. (1995). Redox State of Single-Chain Fv Fragments Targeted to the Endoplasmic-Reticulum, Cytosol and Mitochondria. *Bio-Technology*, **13,** 1110-1115.

Bird RE, Hardman KD, Jacobson JW, Johnson S, Kaufman BM, Lee SM, Lee T, Pope SH, Riordan GS and Whitlow M. (1988). Single-Chain Antigen-Binding Proteins. *Science*, **242,** 423-426.

Bodenhausen G and Ruben DJ. (1980). Natural Abundance N-15 Nmr by Enhanced Heteronuclear Spectroscopy. *Chemical Physics Letters*, **69,** 185-189.

Bonvin AM. (2006). Flexible protein-protein docking. *Current Opinion in Structural Biology*, **16,** 194-200.

Boulianne GL, Hozumi N and Shulman MJ. (1984). Production of functional chimaeric mouse/human antibody. *Nature*, **312,** 643-6.

Braden BC and Poljak RJ. (1995). Structural Features of the Reactions Between Antibodies and Protein Antigens. *Faseb Journal*, **9,** 9-16.

Brekke OH and Loset GA. (2003). New technologies in therapeutic antibody development. *Current Opinion in Pharmacology*, **3,** 544-550.

Burnens A, Demotz S, Corradin G, Binz H and Bosshard HR. (1987). Epitope mapping by chemical modification of free and antibody-bound protein antigen. *Science*, **235,** 780-3.

Campbell ID and Downing AK. (1998). NMR of modular proteins. *Nature Structural Biology*, **5 Suppl,** 496-9.

Carter P, Kelley RF, Rodrigues ML, Snedecor B, Covarrubias M, Velligan MD, Wong WL, Rowland AM, Kotts CE, Carver ME and et al. (1992). High level Escherichia coli expression and production of a bivalent humanized antibody fragment. *Biotechnology (N Y)*, **10,** 163-7.

Cattaneo A and Biocca S. (1999). The selection of intracellular antibodies. *Trends in Biotechnology*, **17,** 115-121.

Cavalli A, Salvatella X, Dobson CM and Vendruscolo M. (2007). Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America*, **104,** 9615-9620.

Chapman AP. (2002). PEGylated antibodies and antibody fragments for improved therapy: a review. *Advanced Drug Delivery Reviews*, **54,** 531-545.

Charbonnier JB, Carpenter E, Gigant B, Golinellipimpaneau B, Eshhar Z, Green BS and Knossow M. (1995). Crystal-Structure of the Complex of a Catalytic Antibody Fab Fragment with a Transition-State Analog - Structural Similarities in Esterase-Like Catalytic Antibodies. *Proceedings of the National Academy of Sciences of the United States of America*, **92,** 11721-11725.

Cheetham GM, Hale G, Waldmann H and Bloomer AC. (1998). Crystal structures of a rat anti-CD52 (CAMPATH-1) therapeutic antibody Fab fragment and its humanized counterpart. *Journal of molecular biology*, **284,** 85-99.

Chothia C and Janin J. (1975). Principles of protein-protein recognition. *Nature*, **256,** 705-8.

Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, Colman PM, Spinelli S, Alzari PM and Poljak RJ.

(1989). Conformations of immunoglobulin hypervariable regions. *Nature*, **342,** 877-83.

Chothia C, Novotny J, Bruccoleri R and Karplus M. (1985). Domain Association in Immunoglobulin Molecules - the Packing of Variable Domains. *Journal of Molecular Biology*, **186,** 651-663.

Choy EHS, Hazleman B, Smith M, Moss K, Lisi L, Scott DGI, Patel J, Sopwith M and Isenberg DA. (2002). Efficacy of a novel PEGylated humanized anti-TNF fragment (CDP870) in patients with rheumatoid arthritis: a phase II double-blinded, randomized, dose-escalating trial. *Rheumatology*, **41,** 1133-1137.

Christiansen J and Rajasekaran AK. (2004). Biological impediments to monoclonal antibody-based cancer immunotherapy. *Molecular cancer therapeutics*, **3,** 1493-501.

Clark M. (2000). Antibody humanization: a case of the 'Emperor's new clothes'? *Immunology Today*, **21,** 397-402.

Clark MR. (1997). IgG effector mechanisms. *Chemical immunology*, **65,** 88-110.

Clore GM. (2000). Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. *Proceedings of the National Academy of Sciences of the United States of America*, **97,** 9021-5.

Clore GM and Gronenborn AM. (1998). NMR structure determination of proteins and protein complexes larger than 20 kDa. *Current opinion in chemical biology*, **2,** 564-70.

Clore GM, Gronenborn AM and Tjandra N. (1998). Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *Journal of Magnetic Resonance*, **131,** 159-62.

Cornilescu G, Delaglio F and Bax A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular Nmr*, **13,** 289-302.

Cornilescu G, Marquardt JL, Ottiger M and Bax A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *Journal of the American Chemical Society*, **120,** 6836-6837.

Cortez-Retamozo V, Backmann N, Senter PD, Wernery U, De Baetselier P, Muyldermans S and Revets H. (2004). Efficient cancer therapy with a nanobody-based conjugate. *Cancer Research*, **64,** 2853-7.

Cwirla SE, Peters EA, Barrett RW and Dower WJ. (1990). Peptides on phage: a vast library of peptides for identifying ligands. *Proceedings of the National Academy of Sciences of the United States of America*, **87,** 6378-82.

Dal Porto JM, Gauld SB, Merrell KT, Mills D, Pugh-Bernard AE and Cambier J. (2004). B cell antigen receptor signaling 101. *Molecular Immunology*, **41,** 599-613.

Dangl JL, Wensel TG, Morrison SL, Stryer L, Herzenberg LA and Oi VT. (1988). Segmental flexibility and complement fixation of genetically engineered chimeric human, rabbit and mouse antibodies. *Embo Journal*, **7,** 1989-94.

Davies DR and Chacko S. (1993). Antibody Structure. *Accounts of Chemical Research*, **26,** 421-427.

Davies DR and Padlan EA. (1992). Twisting into shape. *Current Biology*, **2,** 254-6.

DeLano WL. (2002). *The PyMOL User's Manual*: DeLano Scientific, Palo Alto, Ca, USA.

Deng SJ, MacKenzie CR, Sadowska J, Michniewicz J, Young NM, Bundle DR and Narang SA. (1994). Selection of antibody single-chain variable fragments with

improved carbohydrate binding by phage display. *Journal of biological chemistry*, **269,** 9533-8.

Denzin LK, Whitlow M and Voss EW, Jr. (1991). Single-chain site-specific mutations of fluorescein-amino acid contact residues in high affinity monoclonal antibody 4-4-20. *Journal of biological chemistry*

266, 14095-103.

Desplancq D, King DJ, Lawson ADG and Mountain A. (1994). Multimerization Behavior of Single-Chain Fv Variants for the Tumor-Binding Antibody B72.3. *Protein Engineering*, **7,** 1027-1033.

Devlin JJ, Panganiban LC and Devlin PE. (1990). Random peptide libraries: a source of specific protein binding molecules. *Science*, **249,** 404-6.

Diaz M and Casali P. (2002). Somatic immunoglobulin hypermutation. *Current opinion in immunology*, **14,** 235-40.

Dinarello CA. (1994). The interleukin-1 family: 10 years of discovery. *Faseb Journal*, **8,** 1314-25.

Dinarello CA. (1998). Interleukin-1 beta, interleukin-18, and the interleukin-1 beta converting enzyme. *Molecular Mechanisms of Fever*, **856,** 1-11.

Dinarello CA. (2000). Proinflammatory cytokines. *Chest*, **118,** 503-8.

Dinarello CA and Wolff SM. (1993). The role of interleukin-1 in disease. *New England Journal of Medicine*, **328,** 106-13.

Dominguez C, Boelens R and Bonvin AMJJ. (2003). HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, **125,** 1731-1737.

Dosset P, Hus JC, Marion D and Blackledge M. (2001). A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *Journal of biomolecular NMR*, **20,** 223-31.

Driscoll PC, Clore GM, Marion D, Wingfield PT and Gronenborn AM. (1990). Complete Resonance Assignment for the Polypeptide Backbone of Interleukin-1-Beta Using 3-Dimensional Heteronuclear Nmr-Spectroscopy. *Biochemistry*, **29,** 3542-3556.

Edmundson AB, Ely KR, Girling RL, Abola EE, Schiffer M, Westholm FA, Fausch MD and Deutsch HF. (1974). Binding of 2,4-dinitrophenyl compounds and other small molecules to a crystalline lambda-type Bence-Jones dimer. *Biochemistry*, **13,** 3816-27.

Ehring H. (1999). Hydrogen exchange/electrospray ionization mass spectrometry studies of structural features of proteins and protein/protein interactions. *Analytical biochemistry*, **267,** 252-9.

Eigenbrot C, Randal M, Presta L, Carter P and Kossiakoff AA. (1993). X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling. *Journal of molecular biology*, **229,** 969-95.

Eisenberg RJ, Long D, Pereira L, Hampar B, Zweig M and Cohen GH. (1982). Effect of monoclonal antibodies on limited proteolysis of native glycoprotein gD of herpes simplex virus type 1. *Journal of virology*, **41,** 478-88.

Essig NZ, Wood JF, Howard AJ, Raag R and Whitlow M. (1993). Crystallization of single-chain Fv proteins. *Journal of Molecular Biology*, **234,** 897-901.

Faelber K, Kirchhofer D, Presta L, Kelley RF and Muller YA. (2001). The 1.85 A resolution crystal structures of tissue factor in complex with humanized Fab D3h44 and of free humanized Fab D3h44: revisiting the solvation of antigen combining sites. *Journal of molecular biology*, **313,** 83-97.

Fischer E. (1894). Einfluss der configuration auf die wirkung derenzyme. *Ber Dt Chem Ges*, **27,** 2985-2993.

Foote J and Milstein C. (1994). Conformational isomerism and the diversity of antibodies. *Proceedings of the National Academy of Sciences of the United States of America*, **91,** 10370-4.

Freund C, Ross A, Guth B, Pluckthun A and Holak TA. (1993). Characterization of the Linker Peptide of the Single-Chain F(V)-Fragment of an Antibody by Nmr-Spectroscopy. *Febs Letters*, **320,** 97-100.

Freund C, Ross A, Pluckthun A and Holak TA. (1994). Structural and Dynamic Properties of the Fv Fragment and the Single-Chain Fv Fragment of an Antibody in Solution Investigated by Heteronuclear 3-Dimensional Nmr-Spectroscopy. *Biochemistry*, **33,** 3296-3303.

Gardner KH, Rosen MK and Kay LE. (1997). Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry*, **36,** 1389-1401.

Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM and Clore GM. (1999). Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nature Structural Biology*, **6,** 166-73.

Glockshuber R, Malia M, Pfitzinger I and Pluckthun A. (1990). A Comparison of Strategies to Stabilize Immunoglobulin Fv-Fragments. *Biochemistry*, **29,** 1362-1367.

Goddard TD and Kneller DG. University of California: San Francisco.

Goldsby RA, Kindt TJ, Osbourne BA and Kuby JI. (2000). *Kuby immunology*. W.H. Freeman: New York.

Graves BJ, Hatada MH, Hendrickson WA, Miller JK, Madison VS and Satow Y. (1990). Structure of interleukin 1 alpha at 2.7-A resolution. *Biochemistry*, **29,** 2679-84.

Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA and Baker D. (2003). Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of molecular biology*, **331,** 281-99.

Green LL, Hardy MC, Maynardcurrie CE, Tsuda H, Louie DM, Mendez MJ, Abderrahim H, Noguchi M, Smith DH, Zeng Y, David NE, Sasai H, Garza D, Brenner DG, Hales JF, Mcguinness RP, Capon DJ, Klapholz S and Jakobovits A. (1994). Antigen-Specific Human Monoclonal-Antibodies from Mice Engineered with Human Ig Heavy and Light-Chain Yacs. *Nature Genetics*, **7,** 13-21.

Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC and Flajnik MF. (1995). A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature*, **374,** 168-73.

Greenfield N and Fasman GD. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, **8,** 4108-16.

Griffiths AD, Malmqvist M, Marks JD, Bye JM, Embleton MJ, Mccafferty J, Baier M, Holliger KP, Gorick BD, Hughesjones NC, Hoogenboom HR and Winter G. (1993). Human Anti-Self Antibodies with High Specificity from Phage Display Libraries. *Embo Journal*, **12,** 725-734.

Grzesiek S and Bax A. (1992). Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr. *Journal of the American Chemical Society*, **114,** 6291-6293.

Grzesiek S and Bax A. (1993). Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly C-13/N-15-Enriched Proteins. *Journal of Biomolecular Nmr*, **3,** 185-204.

Hanes J and Pluckthun A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America*, **94,** 4937-4942.

Hansen MR, Hanson P and Pardi A. (2000). *Rna-Ligand Interactions Pt A, Vol. 317: Methods in Enzymology*. Academic Press Inc: San Diego, pp 220-240.

Hennecke F, Krebber C and Pluckthun A. (1998). Non-repetitive single-chain Fv linkers selected by selectively infective phage (SIP) technology. *Protein Engineering*, **11,** 405-410.

Herron JN, He XM, Ballard DW, Blier PR, Pace PE, Bothwell AL, Voss EW, Jr. and Edmundson AB. (1991). An autoantibody to single-stranded DNA: comparison of the three-dimensional structures of the unliganded Fab and a deoxynucleotide-Fab complex. *Proteins*, **11,** 159-75.

Holliger P and Hudson PJ. (2005). Engineered antibody fragments and the rise of single domains. *Nature Biotechnology*, **23,** 1126-1136.

Holliger P, Prospero T and Winter G. (1993). Diabodies - Small Bivalent and Bispecific Antibody Fragments. *Proceedings of the National Academy of Sciences of the United States of America*, **90,** 6444-6448.

Hu Z, Ma B, Wolfson H and Nussinov R. (2000). Conservation of polar residues as hot spots at protein interfaces. *Proteins*, **39,** 331-42.

Hubbard SJ and Thornton JM. (1993). Department of biochemistry and molecular biology, University College London.

Hudson PJ and Kortt AA. (1999). High avidity scFv multimers; diabodies and triabodies. *Journal of Immunological Methods*, **231,** 177-189.

Hus JC, Marion D and Blackledge M. (2001). Determination of protein backbone structure using only residual dipolar couplings. *J Am Chem Soc*, **123,** 1541-2.

Huston JS, Levinson D, Mudgetthunter M, Tai MS, Novotny J, Margolies MN, Ridge RJ, Bruccoleri RE, Haber E, Crea R and Oppermann H. (1988). Protein Engineering of Antibody-Binding Sites - Recovery of Specific Activity in an Anti-Digoxin Single-Chain Fv Analog Produced in Escherichia-Coli. *Proceedings of the National Academy of Sciences of the United States of America*, **85,** 5879-5883.

Huston JS, Mudgetthunter M, Tai MS, Mccartney J, Warren F, Haber E and Oppermann H. (1991). Protein Engineering of Single-Chain Fv Analogs and Fusion Proteins. *Methods in Enzymology*, **203,** 46-&.

Hwang WC, Lin YQ, Santelli E, Sui JH, Jaroszewski L, Stec B, Farzan M, Marasco WA and Liddington RC. (2006). Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. *Journal of Biological Chemistry*, **281,** 34610-34616.

Hwang WYK and Foote J. (2005). Immunogenicity of engineered antibodies. *Methods*, **36,** 3-10.

Iliades P, Kortt AA and Hudson PJ. (1997). Triabodies: Single chain Fv fragments without a linker form trivalent trimers. *Febs Letters*, **409,** 437-441.

Inbar D, Givol D and Hochman J. (1972). Localization of Antibody-Combining Sites within Variable Portions of Heavy and Light Chains. *Proceedings of the National Academy of Sciences of the United States of America*, **69,** 2659-&.

Jain NU, Wyckoff TJ, Raetz CR and Prestegard JH. (2004). Rapid analysis of large protein-protein complexes using NMR-derived orientational constraints: the 95

kDa complex of LpxA with acyl carrier protein. *Journal of molecular biology*, **343,** 1379-89.

Janin J and Chothia C. (1990). The structure of protein-protein recognition sites. *Journal of biological chemistry*, **265,** 16027-30.

Jappelli R, Luzzago A, Tataseo P, Pernice I and Cesareni G. (1992). Loop mutations can cause a substantial conformational change in the carboxy terminus of the ferritin protein. *Journal of molecular biology*, **227,** 532-43.

Jemmerson R and Paterson Y. (1986). Mapping epitopes on a protein antigen by the proteolysis of antigen-antibody complexes. *Science*, **232,** 1001-4.

Jones PT, Dear PH, Foote J, Neuberger MS and Winter G. (1986). Replacing the Complementarity-Determining Regions in a Human-Antibody with Those from a Mouse. *Nature*, **321,** 522-525.

Jones S and Thornton JM. (1996). Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **93,** 13-20.

Josephson K, Jones BC, Walter LJ, DiGiacomo R, Indelicato SR and Walter MR. (2002). Noncompetitive antibody neutralization of IL-10 revealed by protein engineering and x-ray crystallography. *Structure*, **10,** 981-7.

Junghans RP and Anderson CL. (1996). The protection receptor for IgG catabolism is the beta2-microglobulin-containing neonatal intestinal transport receptor. *Proc Natl Acad Sci U S A*, **93,** 5512-6.

Kaneko Y, Nimmerjahn F and Ravetch EV. (2006). Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science*, **313,** 670-673.

Kanelis V, Forman-Kay JD and Kay LE. (2001). Multidimensional NMR methods for protein structure determination. *IUBMB Life*, **52,** 291-302.

Kay LE, Ikura M, Tschudin R and Bax A. (1990). 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins. *Journal of Magnetic Resonance*, **89,** 496-514.

Kipriyanov SM and Le Gall F. (2004). Recent advances in the generation of bispecific antibodies for tumor immunotherapy. *Current opinion in drug discovery & development*, **7,** 233-42.

Knappik A and Pluckthun A. (1995). Engineered turns of a recombinant antibody improve its in vivo folding. *Protein Engineering*, **8,** 81-9.

Koenig BW, Hu JS, Ottiger M, Bose S, Hendler RW and Bax A. (1999). NMR measurement of dipolar couplings in proteins aligned by transient binding to purple membrane fragments. *Journal of the American Chemical Society*, **121,** 1385-1386.

Kohler G and Milstein C. (1975). Continuous Cultures of Fused Cells Secreting Antibody of Predefined Specificity. *Nature*, **256,** 495-497.

Kontaxis G, Clore GM and Bax A. (2000). Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. *Journal of Magnetic Resonance*, **143,** 184-196.

Koradi R, Billeter M and Wuthrich K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics*, **14,** 51-55.

Kortt AA, Lah M, Oddie GW, Gruen CL, Burns JE, Pearce LA, Atwell JL, McCoy AJ, Howlett GJ, Metzger DW, Webster RG and Hudson PJ. (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 Engineering*, **10,** 423-433.

Lane DP and Stephen CW. (1993). Epitope Mapping Using Bacteriophage Peptide Libraries. *Current Opinion in Immunology*, **5,** 268-271.

Laurent TC and Killander J. (1964). A theory of gel filtration and its exeperimental verification. *Journal of chromatography*, **14,** 317-330.

Ledford H. (2007). Fever pitch. *Nature*, **450,** 600-1.

Lee YC, Boehm MK, Chester KA, Begent RHJ and Perkins SJ. (2002). Reversible dimer formation and stability of the anti-tumour single-chain Fv antibody MFE-23 by neutron scattering, analytical ultracentrifugation, and NMR and FT-IR spectroscopy. *Journal of Molecular Biology*, **320,** 107-127.

Lefkovits J, Ivanhoe RJ, Califf RM, Bergelson BA, Anderson KM, Stoner GL, Weisman HF and Topol EJ. (1996). Effects of platelet glycoprotein IIb/IIIa receptor blockade by a chimeric monoclonal antibody (abciximab) on acute and six-month outcomes after percutaneous transluminal coronary angioplasty for acute myocardial infarction. EPIC investigators. *The American journal of cardiology*, **77,** 1045-51.

Lefranc MP. (1997). Unique database numbering system for immunogenetic analysis; Current literature. *Immunology Today*, **18,** 509-509.

Lemieux UR and Spohr U. (1994). How Emil Fischer was led to the lock and key concept for enzyme specificity. *Advances of Carbohydrate Chemistry and Biochemistry*, **50,** 1-20.

Li S, Kussie P and Ferguson KM. (2008). Structural basis for EGF receptor inhibition by the therapeutic antibody IMC-11F8. *Structure*, **16,** 216-27.

Lo Conte L, Chothia C and Janin J. (1999). The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology*, **285,** 2177-2198.

Lu JR, Witcher DR, White MA, Wang XL, Huang LH, Rathnachalam R, Beals JM and Kuhstoss S. (2005). IL-1 beta epitope mapping using site-directed mutagenesis and hydrogen - Deuterium exchange mass spectrometry analysis. *Biochemistry*, **44,** 11106-11114.

Luheshi GN. (1998). Cytokines and fever. Mechanisms and sites of action. *Annals of the New York Academy of Sciences*, **856,** 83-9.

MacCallum RM, Martin ACR and Thornton JM. (1996). Antibody-antigen interactions: Contact analysis and binding site topography. *Journal of Molecular Biology*, **262,** 732-745.

Malby RL, McCoy AJ, Kortt AA, Hudson PJ and Colman PM. (1998). Three-dimensional structures of single-chain Fv-neuraminidase complexes. *Journal of Molecular Biology*, **279,** 901-910.

Marion D, Kay LE, Sparks SW, Torchia DA and Bax A. (1989). 3-Dimensional Heteronuclear Nmr of N-15-Labeled Proteins. *Journal of the American Chemical Society*, **111,** 1515-1517.

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F and Sali A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*, **29,** 291-325.

Mccafferty J, Griffiths AD, Winter G and Chiswell DJ. (1990). Phage Antibodies - Filamentous Phage Displaying Antibody Variable Domains. *Nature*, **348,** 552-554.

Meiler J, Blomberg N, Nilges M and Griesinger C. (2000). A new approach for applying residual dipolar couplings as restraints in structure elucidation. *Journal of Biomolecular Nmr*, **16,** 245-252.

Mendez R, Leplae R, Lensink MF and Wodak SJ. (2005). Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins*, **60,** 150-69.

Mhashilkar AM, Bagley J, Chen SY, Szilvay AM, Helland DG and Marasco WA. (1995). Inhibition of Hiv-1 Tat-Mediated Ltr Transactivation and Hiv-1 Infection by Anti-Tat Single-Chain Intrabodies. *Embo Journal*, **14,** 1542-1551.

Miller S, Lesk AM, Janin J and Chothia C. (1987). The Accessible Surface-Area and Stability of Oligomeric Proteins. *Nature*, **328,** 834-836.

Mirza O, Henriksen A, Ipsen H, Larsen JN, Wissenbach M, Spangfort MD and Gajhede M. (2000). Dominant epitopes and allergic cross-reactivity: complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. *Journal of Immunology*, **165,** 331-8.

Mitraki A and King J. (1992). Amino acid substitutions influencing intracellular protein folding pathways. *Febs Letters*, **307,** 20-5.

Moelling K, Scott A, Dittmar KE and Owada M. (1980). Effect of p15-associated protease from an avian RNA tumor virus on avian virus-specific polyprotein precursors. *Journal of virology*, **33,** 680-88.

Morea V, Lesk AM and Tramontano A. (2000). Antibody modeling: Implications for engineering and design. *Methods*, **20,** 267-279.

Morelli XJ, Palma PN, Guerlesquin F and Rigby AC. (2001). A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data. *Protein Science*, **10,** 2131-7.

Morris GE. (1996). *Epitope mapping protocols*. Humana: Totowa, N.J.

Morrison SL, Johnson MJ, Herzenberg LA and Oi VT. (1984). Chimeric Human-Antibody Molecules - Mouse Antigen-Binding Domains with Human Constant Region Domains. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **81,** 6851-6855.

Murphy KM, Travers P, Walport M and Janeway C. (2007). *Janeway's immunobiology*. Garland Science ; London : Taylor & Francis: New York.

Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA and Hamers R. (1994). Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Engineering*, **7,** 1129-35.

Nieba L, Honegger A, Krebber C and Pluckthun A. (1997). Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved in vivo folding and physical characterization of an engineered scFv fragment. *Protein Engineering*, **10,** 435-44.

Nuttall SD, Humberstone KS, Krishnan UV, Carmichael JA, Doughty L, Hattarki M, Coley AM, Casey JL, Anders RF, Foley M, Irving RA and Hudson PJ. (2004). Selection and affinity maturation of IgNAR variable domains targeting Plasmodium falciparum AMA1. *Proteins*, **55,** 187-97.

Nybakken GE, Oliphant T, Johnson S, Burke S, Diamond MS and Fremont DH. (2005). Structural basis of West Nile virus neutralization by a therapeutic antibody. *Nature*, **437,** 764-9.

Oi VT, Vuong TM, Hardy R, Reidler J, Dangle J, Herzenberg LA and Stryer L. (1984). Correlation between segmental flexibility and effector function of antibodies. *Nature*, **307,** 136-40.

Oka C and Kawaichi M. (1995). V(D)J Recombination of Immunoglobulin Genes. *Advances in Biophysics*, **31,** 163-180.

Ottiger M and Bax A. (1998). Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *Journal of Biomolecular Nmr*, **12,** 361-372.

Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, Rademacher TW, Mizuochi T, Taniguchi T, Matsuta K, Takeuchi F, Nagano Y, Miyamoto T and Kobata A. (1985). Association of Rheumatoid-Arthritis and Primary Osteo-Arthritis with Changes in the Glycosylation Pattern of Total Serum Igg. *Nature*, **316,** 452-457.

Pavlou AK and Belsey MJ. (2005). The therapeutic antibodies market to 2008. *European Journal of Pharmaceutics and Biopharmaceutics*, **59,** 389-396.

Pei XY, Holliger P, Murzin AG and Williams RL. (1997). The 2.0-angstrom resolution crystal structure of a trimeric antibody fragment with noncognate V-H-V-L domain pairs shows are arrangement of V-H CDR3. *Proceedings of the National Academy of Sciences of the United States of America*, **94,** 9637-9642.

Perisic O, Webb PA, Holliger P, Winter G and Williams RL. (1994). Crystal structure of a diabody, a bivalent antibody fragment. *Structure*, **2,** 1217-26.

Pervushin K. (2000). Impact of Transverse Relaxation Optimized Spectroscopy (TROSY) on NMR as a technique in structural biology. *Quarterly Reviews of Biophysics*, **33,** 161-197.

Pervushin K, Riek R, Wider G and Wuthrich K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America*, **94,** 12366-71.

Piotto M, Saudek V and Sklenar V. (1992). Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions. *Journal of Biomolecular Nmr*, **2,** 661-665.

Poljak RJ, Amzel LM, Avey HP, Chen BL, Phizacke.Rp and Saul F. (1973). 3-Dimensional Structure of Fab' Fragment of a Human Immunoglobulin at 2.8-a Resolution. *Proceedings of the National Academy of Sciences of the United States of America*, **70,** 3305-3310.

Prasad L, Sharma S, Vandonselaar M, Quail JW, Lee JS, Waygood EB, Wilson KS, Dauter Z and Delbaere LTJ. (1993). Evaluation of Mutagenesis for Epitope Mapping - Structure of an Antibody-Protein Antigen Complex. *Journal of Biological Chemistry*, **268,** 10705-10708.

Prestegard JH, Bougault CM and Kishore AI. (2004). Residual dipolar couplings in structure determination of biomolecules. *Chemical reviews*, **104,** 3519-40.

Priestle JP, Schar HP and Grutter MG. (1989). Crystallographic refinement of interleukin 1 beta at 2.0 A resolution. *Proceedings of the National Academy of Sciences of the United States of America*, **86,** 9667-71.

Prosser RS, Losonczi JA and Shiyanovskaya IV. (1998). Use of a novel aqueous liquid crystalline medium for high-resolution NMR of macromolecules in solution. *Journal of the American Chemical Society*, **120,** 11010-11011.

Przepiorka D, Kernan NA, Ippoliti C, Papadopoulos EB, Giralt S, Khouri I, Lu JG, Gajewski J, Durett A, Cleary K, Champlin R, Andersson BS and Light S. (2000). Daclizumab, a humanized anti-interleukin-2 receptor alpha chain antibody, for treatment of acute graft-versus-host disease. *Blood*, **95,** 83-9.

Qiu XQ, Wang H, Cai B, Wang LL and Yue ST. (2007). Small antibody mimetics comprising two complementarity-determining regions and a framework region for tumor targeting. *Nature Biotechnology*, **25,** 921-9.

Raag R and Whitlow M. (1995). Single-Chain Fvs. *Faseb Journal*, **9,** 73-80.
Raju TS. (2008). Terminal sugars of Fc glycans influence antibody effector functions of IgGs. *Current Opinion in Immunology*, **20,** 471-478.
Reichert JM and Valge-Archer VE. (2007). Outlook - Development trends for monoclonal antibody cancer therapeutics. *Nature Reviews Drug Discovery*, **6,** 349-356.
Renisio JG, Perez J, Czisch M, Guenneugues M, Bornet O, Frenken L, Cambillau C and Darbon H. (2002). Solution structure and backbone dynamics of an antigen-free heavy chain variable domain (VHH) from llama. *Proteins-Structure Function and Genetics*, **47,** 546-555.
Riedel S. (2005). Edward Jenner and the history of smallpox and vaccination. *Proceedings (Baylor University Medical Centre)*, **18,** 21-5.
Riley D and Shi J. (2005). Unpublished.
Rini JM, Schulze-Gahmen U and Wilson IA. (1992). Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science*, **255,** 959-65.
Royer CA. (2006). Probing protein folding and conformational transitions with fluorescence. *Chem Rev*, **106,** 1769-84.
Ruckert M and Otting G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society*, **122,** 7793-7797.
Sali A and Blundell TL. (1993). Comparative Protein Modeling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, **234,** 779-815.
Sandborn WJ, Feagan BG, Stoinov S, Honiball PJ, Rutgeerts P, Mason D, Bloomfield R and Schreiber S. (2007). Certolizumab pegol for the treatment of Crohn's disease. *New England Journal of Medicine*, **357,** 228-38.
Sass J, Cordier F, Hoffmann A, Cousin A, Omichinski JG, Lowen H and Grzesiek S. (1999). Purple membrane induced alignment of biological macromolecules in the magnetic field. *Journal of the American Chemical Society*, **121,** 2047-2055.
Satow Y, Cohen GH, Padlan EA and Davies DR. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 A. *Journal of molecular biology*, **190,** 593-604.
Scallon BJ, Tam SH, McCarthy SG, Cal AN and Raju TS. (2007). Higher levels of sialylated Fc glycans in immunoglobulin G molecules can adversely impact functionality. *Molecular Immunology*, **44,** 1524-1534.
Schiffer M, Girling RL, Ely KR and Edmundson AB. (1973). Structure of a lambda-type Bence-Jones protein at 3.5-A resolution. *Biochemistry*, **12,** 4620-31.
Schwieters CD, Kuszewski JJ, Tjandra N and Clore GM. (2003). The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, **160,** 65-73.
Schwyzer M, Weil R, Frank G and Zuber H. (1980). Amino acid sequence analysis of fragments generated by partial proteolysis from large simian virus 40 tumor antigen. *Journal of biological chemistry*, **255,** 5627-34.
Scott JK and Smith GP. (1990). Searching for peptide ligands with an epitope library. *Science*, **249,** 386-90.
Segal DM, Padlan EA, Cohen GH, Rudikoff S, Potter M and Davies DR. (1974). The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proceedings of the National Academy of Sciences of the United States of America*, **71,** 4298-302.
Sela M. (1969). Antigenicity: some molecular aspects. *Science*, **166,** 1365-74.

Sharma S, Georges F, Delbaere LT, Lee JS, Klevit RE and Waygood EB. (1991). Epitope mapping by mutagenesis distinguishes between the two tertiary structures of the histidine-containing protein HPr. *Proceedings of the National Academy of Sciences of the United States of America*, **88,** 4877-81.

Sheshberadaran H and Payne LG. (1988). Protein antigen-monoclonal antibody contact sites investigated by limited proteolysis of monoclonal antibody-bound antigen: protein "footprinting". *Proceedings of the National Academy of Sciences of the United States of America*, **85,** 1-5.

Shi J. (2001). Cambridge University, UK.

Shi JY, Blundell TL and Mizuguchi K. (2001). FUGUE: Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of Molecular Biology*, **310,** 243-257.

Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, Weikert SHA and Presta LG. (2002). Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fc gamma RIII and antibody-dependent cellular toxicity. *Journal of Biological Chemistry*, **277,** 26733-26740.

Shinkawa T, Nakamura K, Yamane N, Shoji-Hosaka E, Kanda Y, Sakurada M, Uchida K, Anazawa H, Satoh M, Yamasaki M, Hanai N and Shitara K. (2003). The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *Journal of Biological Chemistry*, **278,** 3466-3473.

Sims JE. (2002). IL-1 and IL-18 receptors, and their extended family. *Current opinion in Immunology*, **14,** 117-22.

Skerra A and Pluckthun A. (1988). Assembly of a Functional Immunoglobulin-Fv Fragment in Escherichia-Coli. *Science*, **240,** 1038-1041.

Sreerama N and Woody RW. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical biochemistry*, **287,** 252-60.

Stanfield RL, Dooley H, Flajnik MF and Wilson IA. (2004). Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science*, **305,** 1770-1773.

Stanfield RL, Fieser TM, Lerner RA and Wilson IA. (1990). Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 A. *Science*, **248,** 712-9.

Stanfield RL, Takimoto-Kamimura M, Rini JM, Profy AT and Wilson IA. (1993). Major antigen-induced domain rearrangements in an antibody. *Structure*, **1,** 83-93.

Stephen CW and Lane DP. (1992). Mutant conformation of p53. Precise epitope mapping using a filamentous phage epitope library. *Journal of molecular biology*, **225,** 577-83.

Stevens FJ, Solomon A and Schiffer M. (1991). Jones,Bence Proteins - a Powerful Tool for the Fundamental-Study of Protein Chemistry and Pathophysiology. *Biochemistry*, **30,** 6803-6805.

Streltsov VA, Varghese JN, Carmichael JA, Irving RA, Hudson PJ and Nuttall SD. (2004). Structural evidence for evolution of shark Ig new antigen receptor variable domain antibodies from a cell-surface receptor. *Proceedings of the National Academy of Sciences of the United States of America*, **101,** 12444-9.

Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD and Panoskaltsis N. (2006). Cytokine storm in a phase 1 trial of the anti-CD28

monoclonal antibody TGN1412. *New England Journal of Medicine*, **355,** 1018-1028.

Takkinen K, Laukkanen ML, Sizmann D, Alfthan K, Immonen T, Vanne L, Kaartinen M, Knowles JKC and Teeri TT. (1991). An Active Single-Chain Antibody Containing a Cellulase Linker Domain Is Secreted by Escherichia-Coli. *Protein Engineering*, **4,** 837-841.

Teeling JL, French RR, Cragg MS, van den Brakel J, Pluyter M, Huang H, Chan C, Parren PW, Hack CE, Dechant M, Valerius T, van de Winkel JG and Glennie MJ. (2004). Characterization of new human CD20 monoclonal antibodies with potent cytolytic activity against non-Hodgkin lymphomas. *Blood*, **104,** 1793-800.

Thorpe IF and Brooks CL. (2007). Molecular evolution of affinity and flexibility in the immune system. *Proceedings of the National Academy of Sciences of the United States of America*, **104,** 8821-8826.

Todorovska A, Roovers RC, Dolezal O, Kortt AA, Hoogenboom HR and Hudson PJ. (2001). Design and application of diabodies, triabodies and tetrabodies for cancer targeting. *Journal of Immunological Methods*, **248,** 47-66.

Tolar P, Sohn HW and Pierce SK. (2008). Viewing the antigen-induced initiation of B-cell activation in living cells. *Immunological reviews*, **221,** 64-76.

Tolman JR. (2001). Dipolar couplings as a probe of molecular dynamics and structure in solution. *Current Opinion in Structural Biology*, **11,** 532-539.

Tonegawa S. (1983). Somatic Generation of Antibody Diversity. *Nature*, **302,** 575-581.

Tsai CJ, Lin SL, Wolfson HJ and Nussinov R. (1996). A dataset of protein-protein interfaces generated with a sequence-order-independent comparison technique. *Journal of Molecular Biology*, **260,** 604-620.

Turner DJ, Ritter MA and George AJT. (1997). Importance of the linker in expression of single-chain Fv antibody fragments: Optimisation of peptide sequence using phage display technology. *Journal of Immunological Methods*, **205,** 43-54.

Tycko R, Blanco FJ and Ishii Y. (2000). Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *Journal of the American Chemical Society*, **122,** 9340-9341.

Vajdos FF, Adams CW, Breece TN, Presta LG, de Vos AM and Sidhu SS. (2002). Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis. *Journal of Molecular Biology*, **320,** 415-428.

Van Berkel PHC, Gerritsen J, Perdok G, Valbjorn J, Vink T, van de Winkel JGJ and Parren P. (2009). N-Linked Glycosylation is an Important Parameter for Optimal Selection of Cell Lines Producing Biopharmaceutical Human IgG. *Biotechnology Progress*, **25,** 244-251.

Van de Water J, Deininger SO, Macht M, Przybylski M and Gershwin ME. (1997). Detection of molecular determinants and epitope mapping using MALDI-TOF mass spectrometry. *Clinical Immunology and Immunopathology*, **85,** 229-235.

Van Regenmortel MH. (1989). Structural and functional approaches to the study of protein antigenicity. *Immunology Today*, **10,** 266-72.

Venters RA, Huang CC, Farmer BT, 2nd, Trolard R, Spicer LD and Fierke CA. (1995). High-level 2H/13C/15N labeling of proteins for NMR studies. *Journal of biomolecular NMR*, **5,** 339-44.

Veverka V, Crabbe T, Bird I, Lennie G, Muskett FW, Taylor RJ and Carr MD. (2008). Structural characterization of the interaction of mTOR with phosphatidic acid and a novel class of inhibitor: compelling evidence for a central role of the FRB

domain in small molecule-mediated regulation of mTOR. *Oncogene*, **27,** 585-595.

Vigers GP, Caffes P, Evans RJ, Thompson RC, Eisenberg SP and Brandhuber BJ. (1994). X-ray structure of interleukin-1 receptor antagonist at 2.0-A resolution. *Journal of biological chemistry*, **269,** 12874-9.

Wabl M and Steinberg C. (1996). Affinity maturation and class switching. *Current Opinion in Immunology*, **8,** 89-92.

Ward ES, Gussow D, Griffiths AD, Jones PT and Winter G. (1989). Binding Activities of a Repertoire of Single Immunoglobulin Variable Domains Secreted from Escherichia-Coli. *Nature*, **341,** 544-546.

Waters LC, Veverka V, Bohm M, Schmedt T, Choong PT, Muskett FW, Klempnauer KH and Carr MD. (2007). Structure of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4 and characterization of its interaction with eIF4A. *Oncogene*, **26,** 4941-4950.

Wedemayer GJ, Patten PA, Wang LH, Schultz PG and Stevens RC. (1997). Structural insights into the evolution of an antibody combining site. *Science*, **276,** 1665-1669.

Whitlow M, Bell BA, Feng SL, Filpula D, Hardman KD, Hubert SL, Rollence ML, Wood JF, Schott ME, Milenic DE, Yokota T and Schlom J. (1993). An Improved Linker for Single-Chain Fv with Reduced Aggregation and Enhanced Proteolytic Stability. *Protein Engineering*, **6,** 989-995.

Williamson RA, Carr MD, Frenkiel TA, Feeney J and Freedman RB. (1997). Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry*, **36,** 13882-9.

Wilson IA and Stanfield RL. (1993). Antibody-Antigen Interactions. *Current Opinion in Structural Biology*, **3,** 113-118.

Wishart DS and Sykes BD. (1994). The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. *Journal of Biomolecular Nmr*, **4,** 171-180.

Wittekind M and Mueller L. (1993). Hncacb, a High-Sensitivity 3d Nmr Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance*, **101,** 201-205.

Wodak SJ and Janin J. (1978). Computer-Analysis of Protein-Protein Interaction. *Journal of Molecular Biology*, **124,** 323-342.

Worn A and Pluckthun A. (1998). Mutual stabilization of V-L and V-H in single-chain antibody fragments, investigated with mutants engineered for stability. *Biochemistry*, **37,** 13120-13127.

Worn A and Pluckthun A. (1999). Different equilibrium stability behavior of ScFv fragments: Identification, classification, and improvement by protein engineering. *Biochemistry*, **38,** 8739-8750.

Wright A and Morrison SL. (1997). Effect of glycosylation on antibody function: Implications for genetic engineering. *Trends in Biotechnology*, **15,** 26-32.

Wright A and Morrison SL. (1998). Effect of C2-associated carbohydrate structure on Ig effector function: Studies with chimeric mouse-human IgG1 antibodies in glycosylation mutants of Chinese hamster ovary cells. *Journal of Immunology*, **160,** 3393-3402.

Wu AM, Chen WG, Raubitschek A, Williams LE, Neumaier M, Fischer R, Hu SZ, OdomMaryon T, Wong JYC and Shively JE. (1996). Tumor localization of anti-

CEA single-chain Fvs: Improved targeting by non-covalent dimers. *Immunotechnology*, **2,** 21-36.

Wuhrer M, Stam JC, van de Geijn FE, Koeleman CAM, Verrips CT, Dolhain R, Hokke CH and Deelder AM. (2007). Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics*, **7,** 4070-4081.

Yokota T, Milenic DE, Whitlow M and Schlom J. (1992). Rapid Tumor Penetration of a Single-Chain Fv and Comparison with Other Immunoglobulin Forms. *Cancer Research*, **52,** 3402-3408.

Zhang HX, Tanji K, Mori F and Wakabayashi K. (2008). Epitope mapping of 2E2-D3, a monoclonal antibody directed against human TDP-43. *Neuroscience Letters*, **434,** 170-174.

Zheng Y, Shopes B, Holowka D and Baird B. (1992). Dynamic conformations compared for IgE and IgG1 in solution and bound to receptors. *Biochemistry*, **31,** 7446-56.

Zimmermann J, Oakman EL, Thorpe IF, Shi X, Abbyad P, Brooks CL, 3rd, Boxer SG and Romesberg FE. (2006). Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **103,** 13722-7.

Zweckstetter M and Bax A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *Journal of the American Chemical Society*, **122,** 3791-3792.

Zweckstetter M and Bax A. (2001). Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *Journal of Biomolecular Nmr*, **20,** 365-377.