Development of an NOE driven method for obtaining robust and reliable models of large protein complexes

Shaun Bruton

BSc(Hons) (Leicester)

Thesis submitted for the Degree of Doctor of Philosophy

Department of Biochemistry

University of Leicester

September 2014

Abstract

Development of an NOE driven method for obtaining robust and reliable models of large protein complexes

Shaun Bruton

The majority of proteins interact with other proteins/nucleic acids to form functional complexes that are essential to the proteins cellular role. Solving the structures of these complexes is vital for a full understanding of a proteins function. However, many protein complexes have resisted attempts of structure determination by established methods, making modelling based on experimental data and known structures of individual proteins an attractive alternative.

The work presented here describes the *in silico* testing, experimental validation and application of a technique that uses H_N -CH₃ NOEs to determine sequence-specific ¹³C/¹H NMR assignments for side-chain methyl groups in proteins, which are generally abundant at protein-protein interfaces. The approach developed relies on the preparation of residually protonated protein samples, avoiding limitations imposed by the molecular weight of larger complexes. Using this approach it was possible to obtain comprehensive assignments for the methyl groups of IL-1 β (17 kDa) both in the free form and in complex with a potential therapeutic Fab antibody fragment (a complex of 65 kDa). It was shown that these assignments could be used to identify a number of backbone amide to side chain methyl NOEs across the protein-protein interface. These NOEs provided a significant number of ¹H-¹H distance restraints that made a substantial difference to the accuracy and reliability of docked structures obtained for the IL-1 β /Fab complex by restraint driven docking. This was confirmed by comparison to a crystal structure determined for the complex.

The developed approach is both conceptually and experimentally straight-forward and is expected to be generally applicable to a wide range of protein complexes up to a molecular weight of approximately 100 kDa. The use of a homology model as the starting structure for the Fab fragment also demonstrates that this technique is tolerant of some differences in the starting and final structures.

Acknowledgements

Firstly, I would like to thank Professor Mark Carr for giving me the opportunity to carry out this research, and for his supervision and guidance throughout. I would also like to thank Dr Richard Taylor and Dr Alistair Henry for their continued advice and support during this project. I would also like to thank Dr Cyril Dominguez for helpful input during committee meetings. I acknowledge the Medical Research Council and UCB for providing the funding for this project.

Thank you to all the members of lab 1/12, particularly Dr Lorna Waters and Dr Fred Muskett for sharing their knowledge and expertise in NMR and also to Dr Gareth Hall for providing the crystal structure of the gIC8 Fab/IL-1 β complex.

I would also like to thank Dr Catherine Hall, Dr Ian Wilkinson and everyone at UCB who has contributed to this project before me for providing a strong foundation from the beginning of my Ph.D.

Shaun Bruton, Leicester, September 2014

Table of Contents

Abstract	1
Acknowledgements	2
Table of Contents	3
Abbreviations	6
Chapter 1: General Introduction	8
1.1: Protein complexes	35
1.2: Methods for solving protein structures	35
1.2.1: X-ray crystallography	35
1.2.2: Cryo-electron microscopy	35
1.2.3: NMR	35
1.3: IL-1β	35
1.4: Therapeutic antibodies and antibody fragments	35
1.5: Thesis overview	35
Chapter 2: Development of assignment technique for methyl groups of larg	e proteins 36
2.1: Introduction	36
2.1.1: Methyl Groups as Probes	36
2.1.2: Methyl Labelling Methods	37
2.1.3: Assignment Techniques	37
2.2: Materials and methods	43
2.2.1: Grid searches	43
2.2.2: Expression of IL-1β	43
2.2.3: Purification of IL-1β	44
2.2.4: gIC8 Fab/IL-1β complex preparation and purification	45
2.2.5: NMR Spectroscopy	45
2.3: Results	47

2.3.1: Determination of inter-atomic H _N -CH ₃ distance ranges in proteins47
2.3.2: Assignment of the C^{13}/H^1 resonances for the methyl groups of the cytokine
Interleukin 1β using NOEs55
2.3.3: Optimisation of the NOE-mixing period for observation of intra-residue
(i) H_N -CH ₃ and inter-residue (i+1) H_N -CH ₃ NOEs in residually protonated protein
samples69
2.3.4: Identification and assignment of methyl signals using inter-residue (i) H_N -
CH ₃ and intra-residue (1+1)H _N -CH ₃ NOEs in IL-1 β as part of a 65 kDa complex. 76
2.4: Discussion
Chapter 3: Determination of an NMR based model
3.1: Introduction
3.1.1: Protein Complexes
3.1.2: De Novo Docking Based Modelling of Protein Complexes
3.1.3: Data Driven Docking of Protein Complexes
3.2: Materials and Methods
3.2.1: Expression of isotope labelled gIC8 Fab
3.2.2: Purification of gIC8 Fab
3.2.3: Chemical denaturation and refolding of gIC8 Fab90
3.2.4: NMR Spectroscopy
3.2.5: Refinement of homology model
3.2.6: Defining AIRs
3.2.7: Intermolecular NOEs
3.2.8: Docking
3.2.9: Crystallisation of the IL-1 β /gIC8 Fab complex
3.3: Results
3.3.1: Expression and purification of gIC8 Fab95
3.3.2: Refolding of deuterated gIC8 Fab after chemical denaturation

3.3.3: Assignment of the backbone NMR signals of gIC8 Fab in both the free a	nd
IL-1β bound form	.100
3.3.4: Refinement of the homology model of gIC8 Fab	.107
3.3.5: Input data for docking calculations	.113
3.3.6: Docking the gIC8 Fab/IL-1β complex	.125
3.3.7: Comparison of docked structures	.139
3.3.8: Crystallisation of gIC8 Fab in complex with IL-1β	.144
3.3.9: Comparison of the crystal and docked structures	.146
3.4: Discussion	.164
Chapter 4: Summary	.170
Appendix	.172
A.1: Scripts used for molecular modelling of tri-peptides	.172
A.1.1: Nested loop used to model Gly-Leu-Gly peptide	.172
A.1.2: Awk script used to filter interatomic distances	.173
References:	.174

Abbreviations

AEBSF	4-(2-Aminoethyl)benzenesulfonyl fluoride	
AIR	Ambiguous interaction restraint	
B ₀	Static magnetic field	
CDR	Complementarity determining regions	
СН	Constant heavy domain	
CL	Constant light domain	
Cryo-EM	Cryo-electron microscopy	
CSA	Chemical shift anisotropy	
DD	Dipole-dipole interactions	
EM	Electron microscopy	
Fab	Antigen binding fragment	
Fc	Crystallisable fragment	
FPLC	Fast protein liquid chromatography	
FR	Frame work	
Fv	Variable fragment	
HADDOCK	High ambiguity driven docking approach	
HSQC	Heteronuclear Single Quantum Correlation	
Ig	Immunoglobulin	
IL-1β	Interleukin-1 beta	
INEPT	Insensitive Nuclei Enhanced by Polarization transfer	
IPTG	Isopropyl-1-thio-β-D-galactopyranoside	
mAb	Monoclonal antibody	
NEM	N-Ethylmaleimide	
Ni-NTA	Nickel-nitrilotriacetic acid	
NMR	Nuclear magnetic resonance	
NOE	Nuclear Overhauser effect	
NOESY	Nuclear Overhauser Effect Spectroscopy	
A _{600 nm}	Absorbance at 600 nm	
PALES	Prediction of Alignment from Structure	
PDB	Protein Data Bank	
PEG	Poly(ethylene glycol)	

Q factor	Quality factor	
R	Rhombicity	
RDC	Residual dipolar coupling	
RDCcalc	Theoretical RDC	
RDCobs	Measured RDC	
RMSD	Root mean square deviation	
scFv	Single chain variable fragment	
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis	
T2	Spin-spin relaxation time	
TALOS	Torsion Angle Likelihood Obtained from Shift and sequence	
	similarity	
TEV	Tobacco etch virus	
TROSY	Transverse relaxation-optimised spectroscopy	
VH	Variable heavy domain	
VL	Variable light domain	
WATERGATE	Water suppression through gradient tailored excitation	
γ	Gyromagnetic ratio	
Δδ	Combined chemical shift	
τ_{c}	Rotational correlation time	

Chapter 1: General Introduction

1.1: Protein complexes

Most proteins in organisms act in concert with other proteins and/or nucleic acids, with permanent or transient interactions resulting in the assembly of functional complexes in order to carry out cellular roles. As such, the biological function of a protein is defined by the interactions it makes within the organism (Kastritis and Bonvin, 2013). Protein-protein complexes can be in the form of large multi-subunit modules such as the proteosome and splicing machinery, or dynamic functional units, for example in signalling cascades and cell cycle regulation (Spirin and Mirny, 2003).

Structural biology is an important tool in understanding the biological function of proteins and protein complexes due to the structure-function relationship. Whilst the first structure of a protein complex was determined in 1960 (Perutz et al., 1960), that of haemoglobin, only three years after the determination of the first protein structure (Kendrew and Parrish, 1957), the importance of protein complexes was not recognised until the 1970s, with the co-purification of multi-protein complexes often mistaken for contamination (Kastritis and Bonvin, 2013). Protein complexes now account for roughly 50% of the protein structures deposited in the PDB, although only 12% of structures are of heteromeric protein complexes. It is not clear whether all of the 38% of structures deposited as homomeric complexes form complexes in solution as well as the crystalline state.

Structure determination of protein complexes is important for drug development. Firstly in terms of identification of new opportunities for the development of novel drugs, but also for design and development of potential drugs to achieve the desired physicochemical properties and binding affinities. The development of therapeutic antibodies can also benefit substantially from detailed structural information. For example, it has been shown that near atomic resolution structures of antibody/antigen complexes can be used to model amino acid changes in the antibody CDR loops resulting in a higher affinity interaction with the antigen, even when the starting antibodies have already undergone affinity maturation (Clark et al., 2006; Kiyoshi et al., 2014).

1.2: Methods for solving protein structures

1.2.1: X-ray crystallography

The process of protein structure determination by X-ray crystallography depends on the interpretation of the diffraction pattern obtained from the scattering of X-rays by the electrons in a crystal of the protein being studied. The first protein structure determined by X-ray crystallography was that of sperm whale myoglobin (Kendrew and Parrish, 1957). Since then over 90,000 crystal structures have been deposited in the Protein Data Bank (PDB), accounting for 88% of the protein structures in the PDB.

The first step in structure determination by X-ray crystallography is the production of high quality protein crystals, capable of diffracting X-rays to high resolution. This is performed through slow precipitation of the protein from an aqueous solution, by the gradual increase of protein and precipitant concentrations. This can be achieved in a controlled manner via vapour diffusion in a sealed environment. The optimum conditions for growing crystals of sufficient quality are protein dependent, and determination of these conditions is largely a process of trial and error. However, through the development of sparse matrix screens (for example Newman et al., 2005) and robotics, it is now possible to screen hundreds of different conditions. Once crystals of sufficient quality are obtained, the process of collection of diffraction data, structure calculation and refinement is streamlined and highly automated, provided a similar structure is available for use in molecular replacement.

As with any structure determination by X-ray crystallography, the solution of a proteinprotein complex structure is first dependent on the growth of crystals of the protein complex. For very high affinity protein complexes the process of crystallisation is essentially the same as for monomeric proteins. However, for lower affinity complexes the instability of the complex leads to added difficulty. Analysis of the conditions for the crystallisation of deposited protein complex structures shows that protein complexes crystallise in a narrower range of conditions compared to proteins in general, which indicates the instability of protein complexes (Radaev et al., 2006). This has allowed the development of specialised screens for crystallising protein complexes, for example ProPlexTM. For low affinity protein complexes there is also the issue of free protein present in the crystallisation mix which can affect crystallisation. Despite these challenges the majority of protein complex structures deposited in the PDB are crystal structures and some of these include weak complexes, for example, the crystal structure of Ubiquitin bound to the UEV domain of TSG-101 (an HIV protein) which has a Kd of ~500 μ M (Sundquist et al., 2004). An added complication to the determination of protein complex structures by crystallography is the presence of crystal contacts. These are protein-protein contacts that form as a result of the protein molecules forming an ordered lattice. For some protein complexes it is not possible to determine from the structure which protein complexes are a result of the crystal contacts and which are the protein-protein contacts present in solution, requiring other techniques such as NMR or mutagenesis to fully understand the nature of the protein complex in solution (Kobe et al., 2008).

There is no theoretical upper molecular weight limit to the size of protein assemblies that can be studied by X-ray crystallography. However, in practice, growing crystals of sufficient quality from very large protein assemblies can be challenging. Despite this there are many crystal structures in the PDB of protein assemblies over 1 MDa.

1.2.2: Cryo-electron microscopy

When studying very large protein assemblies it is possible to determine protein structures via cryo electron-microscopy (cryo-EM). This process involves the analysis of thousands of single particles, prepared in a thin layer of vitreous ice, and as such does not require the production of protein crystals (Smith and Rubinstein, 2014). Cryo-EM was initially limited as a technique for solving atomic resolution structures by the low resolution of the images obtained. Resolution better than 10 Å is required to resolve the α -helices of a protein, which appear as rods at this resolution, and resolution better than 4.8 Å is required to resolve β -strands, which has generally been unattainable due to the low signal to noise ratio of the images obtained. For this reason, until recently cryo-EM has served as a method for docking high resolution Crystal or NMR structures into large multi-protein assemblies.

Recent improvements in hardware for single particle cryo-EM have lead to the solution of multiple structures at near atomic resolution (~3 Å), which allows fitting of the structures directly to the density maps obtained. This is possible due to the development of a new detector, capable of single electron counting, allowing for the collection of near noiseless images, at a much higher frame rate than was previously possible (Li et al., 2013). The greater signal to noise of the individual images in conjunction with images captured at 400 frames per second allows the images to be combined in a way that corrects for electron beam induced motion of the sample, a process that would previously have lead to unacceptable degradation of the resolution of the image (Li et al., 2013). This allowed the determination of a 3.3 Å resolution structure of the 700 kDa 20s proteosome (Li et al., 2013).

Whilst cryo-EM is now capable of solving near atomic resolution structures, applications are currently restricted to relatively large protein assemblies of around 500 kDa, preferably with some degree of internal symmetry, although some smaller structures have been reported (Lu et al., 2014). It has also been noted that further development of the tools used to validate cryo-EM structures is also required (Smith and Rubinstein, 2014). The potential of cryo-EM for use in structure based drug discovery has also been demonstrated with the determination of the structure of β -galactosidase (465 kDa) at a resolution of 3.2 Å, for which a bound ligand could be fitted to the density map (Bartesaghi et al., 2014).

1.2.3: NMR

Nuclear magnetic resonance spectroscopy (NMR) is unique among the currently available techniques capable of determining atomic resolution protein structures in that structure calculations are based on observations made on the protein in the solution state. After the first observations of NMR in liquids in 1946 (Bloch et al., 1946), multiple advancements in the field paved the way for the first NMR experiments performed on biological macromolecules. This includes the description of the NOE (Solomon, 1955), the development of Fourier transform NMR (Ernst, 1966), the demonstration of two-dimensional NMR (Aue, 1976), two dimensional NMR of proteins (Nagayama et al., 1977) and two-dimensional NOESY experiments (Jeener et

al., 1979). Pioneering work by Wüthrich and co-workers in the 1980s lead to the first protein structure determined using only NMR data, which was that of bull seminal trypsin inhibitor (Williamson et al., 1985), an accomplishment that was reliant on the development of sequential assignment techniques (Dubs et al., 1979) (Billeter et al., 1982) and distance geometry calculations (Havel and Wüthrich, 1985).

The first protein NMR structures were determined using homonuclear ¹H NMR techniques, which limited solution NMR studies to proteins less than 10 kDa in molecular weight. This was due to the increased signal overlap, broader signals and reduced signal to noise associated with larger proteins. The development of recombinant protein expression and isotope labelling techniques enabled the production of uniformly ¹⁵N labelled protein (LeMaster and Richards, 1985) and ¹³C labelling, which has subsequently allowed the development of multidimensional heteronuclear NMR experiments. Along with improvements in hardware, including higher field strength spectrometers and cryo-probes, these developments have extended the molecular weight range of solution NMR to 25-50 kDa.

A further increase in the molecular weight of proteins that can be studied by NMR was made possible by the development of transverse relaxation optimised spectroscopy (TROSY) (Pervushin et al., 1997). This technique takes advantage of constructive interference between dipole-dipole coupling and chemical shift anisotropy (CSA) for one of the multiplets in a coupled ¹⁵N/¹H HSQC, selection for this multiplet at high field strengths and on large proteins offers superior sensitivity and smaller line widths compared to decoupled HSQC spectra (Pervushin et al., 1997). The TROSY element can be incorporated into most HSQC based pulse sequences.

The introduction of ²H labelling has further extended the applicability of solution NMR to larger proteins. Dipolar interaction of covalently bonded ¹³C-¹H spins dominates ¹³C relaxation in proteins and ~40% of amide proton (H_N) relaxation is attributed to dipolar interactions with side-chain ¹H spins (Cavanagh et al., 2007). Replacement of carbon bound protons with deuterons provides significantly better relaxation properties due to the ~6 fold lower gyromagnetic ratio of ²H compared to ¹H. In combination with TROSY spectroscopy, these techniques now routinely allow the assignment of backbone resonances in proteins of up to 110 kDa, with the largest single chain protein studied to date being the 82 kDa malate synthase G (MSG) (Vitali Tugarinov, 2002).

Deuteration of the protein does limit the structural restraints available from ¹H-¹H interatomic distance restraints to those involving only amide protons. However, it is possible using specific labelling techniques to re-introduce side chain protons at residue specific sites, increasing the restraints available for structure calculations (discussed further in chapter 2).

In recent years developments in data acquisition and processing have allowed a reduction in the often prohibitive length of time required to record 3D and 4D spectra. The need to sample a full Nyquist grid of data points for each indirect dimension often restricts acquisition times in indirect dimensions to less than those optimal for both resolution and sensitivity (Rovnyak et al., 2004). One way to overcome this is to sample only a subset of the data points of the full Nyquist grid. However, direct fourier transform of such non-uniformly sampled (NUS) data results in spectra with a low signal to noise ratio and artefacts. Therefore it is necessary to reconstruct missing data points in non-uniformly sampled dimensions prior to fourier transformation. This can be achieved through the use of maximum entropy reconstruction (Stern et al., 2002) or multi-dimensional decomposition (MDD) (Orekhov et al., 2003). However, these reconstruction algorithms typically require that at least 20% of the data points in the Nyquist grid be sampled (Jaravine et al., 2006). More recent reconstruction algorithms have proven successful with even lower amounts of sampled data, including recursive multidimensional decomposition (R-MDD) (Jaravine et al., 2006) and iterative soft thresholding (IST) (Hyberts et al., 2012). NUS can be used to increase the effective evolution time in each indirect dimension without increasing total acquisition time, or reduce total acquisition time without reducing the effective evolution time in each indirect dimension. It has also been shown that NUS can be used to obtain superior signal to noise per given total acquisition time compared to uniformly sampled data (Hyberts et al., 2013), an effect that becomes even greater in higher dimensional spectra.

Despite these advances in NMR techniques there are still relatively few NMR structures in the PDB with a molecular weight higher than 30 kDa, as illustrated in Figure 1.1, indicating the need for further development of NMR techniques and data analysis for studying large proteins.



Figure 1.1: Number of solution NMR derived protein structures deposited in the PDB

The histogram shows the number of solution NMR derived protein structures deposited in the PDB. The vast majority of the structures have a molecular weight less than 30 kDa.

1.2.3.1: Structure determination by NMR

There is no consensus for the process of structure determination by NMR with software and approaches used varying from lab to lab. Structure determination by NMR spectroscopy is first dependent on the production of isotope labelled samples. This is most easily and cost-effectively achieved in a bacterial expression system, where cultures are grown in minimal media that contains single carbon and nitrogen sources that can be substituted with the ¹³C and ¹⁵N forms depending on the labelling that is

required (Clore and Gronenborn, 1994). ¹³C and ¹⁵N labelling is also readily achieved in yeast and mammalian expression systems albeit at a greater cost. Once a suitable sample or samples have been prepared the NMR spectra can be recorded allowing assignment of the backbone and side chain ¹H, ¹³C and ¹⁵N chemical shifts. The number of interatomic distance restraints has a large effect on the precision and accuracy of the calculated structure (Clore et al., 1993), requiring a high level of assignment for generating a well defined structure. Structure determination usually begins with the automated assignment of NOE cross peaks in 3D ¹⁵N and ¹³C edited NOESY spectra and calculation of a preliminary set of structures which can be achieved using the combined automated NOE assignment and structure determination protocol (CANDID) within the CYANA software (Herrmann et al., 2002) for example. At this stage dihedral angle restraints may be included for example from dihedral angles predicted using the programme TALOS (Shen et al., 2009) which generates restraints based on the assigned backbone ¹H, ¹³C and ¹⁵N chemical shifts via reference to a library of known structures with assigned chemical shifts. From these preliminary structures it may be possible to determine additional restraints, for example, amide groups that form hydrogen bonds tend to exhibit slow exchange with the solvent which can be detected by transferring the sample to a D₂O solvent. If the hydrogen bond acceptor can be identified unambiguously from the preliminary structures, the hydrogen bonds can be included as distance restraints to improve convergence when the final structures are calculated. Final structure calculation using all available restraints can be performed using a torsion angle based simulated annealing protocol combined with redundant dihedral angle constraints (REDAC) (Güntert and Wüthrich, 1991).

Protein complex determination by NMR spectroscopy can be performed essentially as the structure of a monomeric protein would, for example the structure of the secreted *M. tuberculosis* protein complex ESAT-6 CFP-10 (Renshaw et al., 2005), although in this case simplified NOESY spectra where only one component of the complex was isotopically labelled, thus reducing the number of peaks and hence overlap in the spectra, were used. Alternatively, structure determination of a protein complex can be performed with the use of specific intermolecular restraints. These restraints can be obtained through the use of filtered/edited NOESY spectra (Otting et al., 1986) where one component of the complex is ${}^{13}C/{}^{15}N$ labelled and one component is unlabelled. This results in only intermolecular NOEs which can be defined as such when

generating the restraints for structure calculation. An example where this approach has been used is the structure of a complex of an SH3 domain and a LIM domain (Vaynberg et al., 2005). However, filtered/edited NOESY experiments require extra delays that can result in a significant reduction in sensitivity, particularly for large protein complexes (Walters et al., 2001). Another asymmetric labelling approach for recording intermolecular NOEs is where one component of the complex is ¹⁵N/²H labelled (perdeuterated) and the other is unlabelled. Any amide proton to side chain proton NOEs in a ¹⁵N-edited NOESY have to be intermolecular and this doesn't require any modification to the pulse sequence, resulting in higher sensitivity compared to filtered/edited NOESY techniques (Walters et al., 1997).

Structure determination of protein complexes is effective for tight protein complexes such as the obligate secreted protein complex given as an example above (Renshaw et al., 2005) and also for weak protein complexes such as the SH3-LIM complex (Vaynberg et al., 2005) with a Kd of 3 mM, provided that the proteins are soluble enough to achieve a high enough concentration of the complex. However, for some complexes, intermediate exchange can cause signals from residues at the protein interface to broaden, preventing the use of NOEs to generate intermolecular distance restraints.

Despite the ability to solve protein complex structures by NMR, there are few NMR structures of protein complexes deposited in the PDB. The reason for this is the size of protein complexes. Protein structure determination by NMR relies on the assignment of a high level of the backbone and side chain ¹H, ¹³C and ¹⁵N chemical shifts, which as discussed above is not possible in large proteins and protein complexes preventing *de novo* calculation of many protein complex structures. However, NMR derived restraints can be used in protein docking calculations (discussed further in chapter 3).

1.2.3.2: Nuclear Overhauser effect

A phenomenon that is central to protein structure determination by NMR spectroscopy is the Nuclear Overhauser effect (NOE). Broadly speaking the NOE provides information on which ¹H-nuclei in a protein are close together in space (Neuhaus and Williamson, 2000). NOEs are observed as cross peaks in NOESY spectra and occur over relatively short inter-atomic distances of less than 5-6 Å depending on the isotope labelling of the protein. The distance dependence of the NOE allows its use in the definition of inter-atomic distance restraints, thousands of which are required in the generation of a well defined NMR structure.

The theory of the NOE was first proposed by Albert Overhauser (A. Overhauser, 1953; A. W. Overhauser, 1953) and was later described using the Solomon equations (Solomon, 1955). The theory predicts that the Z-component on a spin is altered when the magnetisation on a nearby spin is perturbed from equilibrium, an effect that can be observed in 1D NMR using a difference spectrum or via cross peaks in multidimensional NOESY spectra. The NOE is a relaxation driven process, specifically cross-relaxation which arises from dipolar coupling. Dipolar coupling is a dominant source of longitudinal relaxation (recovery of the Z-component of the magnetisation to its equilibrium value) in proteins and is caused by local magnetic fields in the sample, specifically those generated by nearby spins. When considering dipolar relaxation in a two-spin system, there are four energy levels between which transitions can occur. These include the four single quantum transitions, a zero quantum transition and a double quantum transition. It is the zero and double quantum transitions that give rise to the NOE. The rate constants for these transitions depend on the spectral density at the sum of the larmour frequencies of the two spins for double quantum transitions and the difference in the larmour frequencies of the two spins for zero quantum transitions. In the slow motion limit (which applies to proteins) the spectral density at the sum of the larmour frequencies of two protons is negligible, meaning that in proteins the NOE arises through cross relaxation via zero quantum transitions caused by dipolar coupling (Neuhaus and Williamson, 2000).

1.4: Therapeutic antibodies and antibody fragments

1.4.1: Antibodies

Antibodies have a central role in the adaptive immune system, a defence mechanism present in most vertebrates. With the ability to specifically bind to structural features of invading bacteria, viruses and parasites (antigens), antibodies form the antigen receptors of the adaptive immune response. Binding of an antigen eventually results in the production of soluble antibodies that can bind and neutralise invading organisms by preventing their entry into cells, facilitating phagocytosis and activating complement (a part of the innate immune response) (Schroeder and Cavacini, 2010).

Each individual B-lymphocyte (B-cell) presents, on its surface, identical antibodies that have the same specificity for a known or unknown antigen. Across a population of Bcells, in the order of 10^{11} cells (Trepel, 1974), a vast array of antibodies with different specificities are presented, allowing the recognition of virtually any antigen. This extensive variability in antibodies that are produced by different B-cells stems initially from a repertoire of antibody genes. However, with the number of different antibodies required for an effective immune system in the order of 10^8 (Perelson and Oster, 1979), far exceeding the number of protein coding genes in the human genome, there cannot be separate genes for each antibody. Instead variability is introduced through recombination of various gene fragments during differentiation of B-cells in the primary lymphoid organs (Tonegawa, 1983). In humans all antibodies consist of four polypeptide chains, two identical light chains and two identical heavy chains (discussed further below) (Schroeder and Cavacini, 2010). Genes encoding both the light and heavy chains contain variable and constant regions, with the variable regions coding for the antigen binding domains in the expressed protein. The mature light chain gene is formed from a recombination event that links a variable gene segment (V) with another gene segment (J) that joins the variable segment to the constant gene in a process known as VJ recombination. With multiple different V and J gene segments that can combine in an essentially random manner there is a high degree of diversity in the resulting mature light chain gene. Additional diversity is introduced as the exact site of recombination is not fixed and can vary by several nucleotides. A similar process

results in the formation of the mature heavy chain gene albeit with an extra level of complexity. The heavy chain gene has an additional gene fragment termed the diversity (D) fragment. An initial recombination event links a J fragment to a D fragment, followed by a second recombination event that links a V fragment to the newly formed (D)J fragment in a process known as V(D)J recombination (Oka, 1995).

1.4.2: Antibody Structure

As mentioned previously, all human antibodies are formed of four polypeptide chains; two identical light chains and two identical heavy chains. Initial insight into the structure of antibodies came in 1959 with the discovery that the protease papain cleaves IgG (one of the five human antibody isotypes) into three 50 kDa fragments (Porter, 1959). Two of these fragments bind antigen and are termed the antigen binding fragments or Fabs. Each Fab consists of the entire light chain of the antibody and the Nterminal half of the heavy chain. The third fragment does not bind antigen and is termed the crystallisable fragment or Fc due to it crystallising readily. The Fc contains the two C-terminal halves of the heavy chains. Subsequent structural characterisation of Fabs has shown that they contain four immunoglobulin domains, two from the light chain and two from the heavy chain (Poljak et al., 1973). The Fc also contains four immunoglobulin domains, two from each of the identical heavy chains. The intact antibody forms the characteristic Y-shape shown in Figure 1.2 and has two antigen binding sites. The N-terminal immunoglobulin domains of both the heavy and light chains (the ends of the "arms" of the antibody) are the variable domains forming the antigen binding site and these are encoded by the V(D)J genes described above. The rest of the Ig domains do not change with regard to antigen specificity and so are referred to as constant domains. The C-terminal immunoglobulin domain of the light chain and the second Ig domain of the heavy chain form the remainder of the Fab and these domains are covalently linked by an inter-chain disulphide bond. A linker of varying length known as the hinge region links the first constant domain of the heavy chain to the second constant domain (in the Fc) and at least one inter-chain disulphide bond links the two heavy chains in the hinge region. Each Ig domain is formed of two anti-parallel β sheets stabilised by a disulphide bond. In the variable domains six loops,

three from the light chain and three from the heavy chain, known as complementarity determining regions (CDRs) form the antigen binding site.

Binding of an antigen to cell surface antibodies triggers B-cell signalling, subsequently leading to B-cell proliferation and differentiation. The antigen bound antibody is internalised and the antigen is degraded. Peptide fragments of the antigen are presented on the surface of the B-cell as part of the major histocompatability complx II (MHCII). The peptide associated MHCII is recognised by T-cell receptors on the surface of Tcells. This triggers T-cells to secrete cytokines which in turn cause B-cells to proliferate and differentiate into soluble antibody secreting cells. During proliferation and differentiation a process known as somatic hyper-mutation occurs in the variable regions of the genes encoding the antibody. These mutations can either increase or decrease the affinity of an antibody for the antigen and at this time limited availability of antigen selects for B-cells expressing antibodies with the greatest affinity. This process, known as affinity maturation, can increase the affinity of an antibody more than 1000-fold relative to the initial antibody that triggered the immune response. Secreted antibodies can bind and neutralise invading organisms by preventing their entry into cells, facilitate phagocytosis and activate complement (a part of the innate immune response), the latter two of these processes being driven by the interaction of the constant Ig domains of the Fc with other components of the immune system. Later in the immune response antibody isotype switching occurs. There are 5 classes of antibody heavy chain; μ , δ , γ , ε and α , which determines the antibody isotype, IgM, IgD, IgG, IgE and IgA respectively. Immature B-cells present antibodies of the IgM isotype on their surface and IgM is the first antibody to be secreted after activation of B-cells by antigen binding. The light chain is unaffected by isotype switching, but the V(D)J element of the gene for the heavy chain is recombined with a gene from a different heavy chain. The resulting antibody has identical variable domains to the IgM before isotype switching and so maintains its antigen specificity, but the constant domains of the heavy chain are different (Figure 1.3), allowing the different isotypes to mediate different effector functions through interactions with the Fc.



Figure 1.2: General topology of an antibody

The cartoon shows the general structure of an antibody displaying the characteristic Y-shape. Immunoglobulin domains are shown as ovals, with the heavy chains shown in blue and the light chains shown in yellow, inter-chain disulphide bonds are shown as black lines. The structure is a ribbon representation of the backbone of an intact murine IgG, PDB code 1IGT (Harris et al., 1997), with the heavy chains coloured blue and the light chains coloured yellow.



Figure 1.3: General structure of the 5 human antibody isotypes

The cartoons show the general structure of human antibody isotypes. Immunoglobulin domains are shown as ovals, with the heavy chains shown in blue and the light chains shown in yellow, inter-chain disulphide bonds are shown as black lines. IgM and IgA are shown in their multimeric pentamer and dimer forms respectively. Multimerisation is facilitated by the J-chain.

1.4.3: Therapeutic antibodies

The high specificity and affinity of antibodies for their target antigen makes them an ideal candidate for use as therapeutics. The first antibody approved for use was Muromonab in 1986, an anti-CD3 murine antibody used for the treatment of kidney transplant rejection (Smith, 1996). However, the murine antibody was found to elicit an immune response in approximately 50% of patients (Norman et al., 1993). In extreme cases this can result in anaphylaxis, and at best results in fast removal of the therapeutic from the patient's circulation. Immunogenicity can be limited if the murine antibodies are altered to become more human in nature. This was first accomplished by the production of chimeric antibodies (Morrison et al., 1984), where the mouse variable domains were combined with human constant domains and human Fc regions to form an antibody with reduced immunogenicity, for example Abciximab which inhibits platelet aggregation in angioplasty patients (Faulds and Sorkin, 1994). This process was later refined so that the complementarity determining regions of the murine antibodies can be grafted onto human frameworks (Jones et al., 1986), further reducing the murine proportion of the antibody, resulting in lower immunogenicity. Whilst this process has been shown to be very successful, the generation of humanised antibodies can often result in a reduction in binding affinity (Reichert, 2010). This can be understood/improved where structural characterisation of the antibody/antigen interaction is available. With the development of near human antibodies there is the potential for the therapeutic antibody to interact with components of the immune system to induce an immune response. This can be both an advantage and disadvantage depending on the target of the therapeutic. One way to prevent this interaction is removal of the Fc portion of the antibody. The Fab fragment contains the antigen interaction site, so interaction with the antigen is identical to that of the full antibody. This also affects antibody production, as Fab fragments can be expressed in a bacterial expression system whereas full antibodies require expression in a mammalian expression system. This has implications for both manufacturing and NMR based studies of the antigen/antibody interaction, as isotope labelling of the antibody is more favourable in a bacterial expression system. One drawback to using Fab fragments is the reduced size of the antibody, resulting in faster clearance from the patients system, however, it has been shown that the production of a Fab-PEG conjugate to increase the size of the fragment above the glomerular filtration limit can drastically increase the circulating half-life of the Fab (Chapman, 2002). An example of this is the peg-ylated anti-TNF α Fab fragment Cimzia, which is used in the treatment of rheumatoid arthritis and crohns disease (Choy et al., 2002).

1.3: IL-1β

Interleukin-1 β (IL-1 β) is a member of the interleukin-1 (IL-1) family, a group of 11 cytokines that regulate the pro-inflammatory response to infection and injury. IL-1s are produced mainly by macrophages and monocytes in response to activation via pathogen associated molecular patterns (PAMPs) such as lipopolysaccharide (LPS) and danger associated molecular patterns (DAMPs) such as molecules released from dead cells (Eder, 2009). IL-1s are also secreted by fibroblasts, dendritic cells, B-lymphocytes, natural killer cells and epithelial cells. Activation of macrophages and monocytes initially triggers expression of the precursor form of IL-1 β . Proteolytic cleavage of the first 116 N-terminal residues of the precursor protein yields the mature, active IL-1 β , which is secreted from the cells via a non-conventional secretory pathway that is not fully understood (reviewed by Lopez-Castejon et al 2011).

Secreted IL-1 β binds with high affinity to the Interleukin 1 receptor 1 (IL-1R1), which is also able to bind IL-1 α . IL-1 β and IL-1 α share 24% sequence identity and activate the same downstream signalling pathways upon binding to IL-1R1, however, IL-1 α tends to be membrane anchored, having a largely autocrine or juxtocrine effect whereas IL-1 β is secreted, having a paracrine and more systemic effect (Dinarello and Wolff, 1993). Binding of IL-1 α or IL-1 β to IL-1R1 induces a structural rearrangement in the extracellular domains of IL-1R1 which promotes binding of the Interleukin 1 receptor accessory protein (IL-1RAcP). The molecular mechanisms underlying the formation this signalling complex have been revealed in detail by the determination of the crystal structure of the IL-1β/IL-1R1/IL-1RAcP complex (Thomas et al 2012) which is shown in Figure 1.4. Formation of the hetero-trimeric signalling complex brings the Toll and IL-1R like (TIR) intracellular domains of IL-1R1 and IL-1RAcP into close proximity, promoting the formation of an intracellular signalling complex with myeloid differentiation primary response gene 88 (MYD88) and Interleukin 1 receptor activated protein kinase 4 (IRAK4) (Weber et al., 2010). Formation of this stable signalling module facilitates autophosphorylation and activation of IRAK4, which in turn phosphorylates IRAK1 and IRAK2 before recruiting tumour necrosis factor associated factor 6 (TRAF6). TRAF6 and IRAK1 subsequently dissociate from the receptor signalling module and through various signalling intermediates activate p38 MAPK, JNK and NF- κ B, eventually inducing the expression of IL-6, IL-8, MCP-1, COX-2, IL-1 α and IL-1 β (Weber et al., 2010).

Regulation of IL-1 β signalling is complex and occurs at multiple levels. Inhibition of IL-1 β signalling can occur through the decoy receptor IL-1RII, which is homologous to IL-1RI. Upon binding of IL-1 β to IL-1RII, IL-1RAcP is recruited to the receptor complex, however, signal transduction cannot occur as IL-1RII lacks the intracellular TIR domain necessary for signalling (Wang et al., 2010). IL-1 β signalling is also inhibited by the production of IL-1Ra which can bind to IL-1RI but does not form the composite binding site for IL-1RAcP (Greenfeder et al., 1995). In addition to this, all three receptors (IL-1RI, IL-1RII and IL-1RAcP) can be expressed in a soluble form via alternative splicing (Boraschi and Tagliabue, 2006).

Due to its role in inflammation, IL-1 is implicated in a number of inflammatory diseases including rheumatoid arthritis, inflammatory bowel disease and insulin dependent diabetes, making the IL-1 family an important target for therapeutic intervention, as reviewed by (Barksby et al., 2007). For this reason there are several approved therapeutics that target IL-1 β signalling, including recombinant interleukin 1 receptor antagonist (IL-1Ra), which is homologous to IL-1 β and is able to bind IL-1RI, but does not form the composite binding site for IL-1RAcP and so is not able to activate signalling. IL-1 β is also the target of several therapeutic antibodies, one of which (Canakinumab) is approved for the treatment of cryopyrin-associated periodic syndromes (Dhimolea, 2010). The binding modes of this therapeutic antibody along with another (Gevokizumab) have also been determined by X-ray crystallography (Blech et al., 2013) the structures of which are shown in Figure 1.5.

Recombinant mature IL-1 β can be expressed in *E. coli* with high yields, is highly soluble and stable for long periods and for these reasons, in addition to being an important therapeutic target, IL-1 β also serves as a model protein for the investigation of experimental techniques. IL-1 β is particularly amenable to investigations by structural biology, which is evident from the large number of structures available, with 34 deposited in the PDB. The deposited structures of human IL-1 β are summarised in table 1.1. The single 153 residue (17.4 kDa) chain forms a single domain with

predominantly β -sheet secondary structure. 12 β -strands form a β -barrel with an approximate 3-fold symmetry axis, known as a β -trefoil (Finzel et al., 1989).

PDB code	Release Date	Method	Description
1I1B	15/01/90	X-ray	Wild type
2I1B	15/04/90	X-ray	Wild type
411b	15/04/90	X-ray	Wild type
511B	15/01/92	X_ray	Wild type
511B	13/01/92	A-lay	
21BI	15/04/92	X-ray	C71A
	10/01/2	TT Tuy	
31BI	15/04/92	X-ray	C71S
41BI	15/04/92	X-ray	C8A
7I1B	15/10/92	NMR	Wild type
1HIB	31/03/94	X-ray	T9G
1IOB	17/08/96	X-ray/	Wild type – joint refinement using X-ray and NMR data
		NMR	
	0.4/07/17.5		
1ITB	04/02/98	X-ray	IL-1 β in complex with IL-1R1
9ILB	06/01/99	X-ray	Wild type

1L2H	04/02/03	X-ray	F42W/W120F
1S0L	30/03/04	X-ray	F42W
1T4Q	07/12/04	X-ray	F101W
1TOO	07/12/04	X-ray	F146W
1TWE	07/12/04	X-ray	F101Y
1TWM	07/12/04	X-ray	F146Y
1TP0	07/12/04	X-ray	F42W/F101W/F146W
2NVH	12/12/06	X-ray	Wild type
2KH2	08/09/09	NMR/Docking	IL-1 β in complex with a sc-Fv
3040	01/09/10	X-ray	IL-1 β in complex with IL-1R2 and IL-1RaCP
ЗРОК	19/01/11	X-ray	IL-1 β engineered to include a lanthanide binding tag in a loop region
3LTQ	16/02/11	X-ray	IL-1 β engineered to include a lanthanide binding tag in a loop region
4DEP	21/03/12	X-ray	IL-1 β in complex with IL-1R1 and IL-1RaCP

			IL-1β in complex with a Fab
4G6J	19/12/12	X-ray	
4G6M	19/12/12	X-ray	IL-1 β in complex with a sc-Fv
4GAI	20/02/13	X-ray	Chimera of IL-1β and IL-1Ra
4GAF	20/02/13	X-ray	Chimera of IL-1 β and IL-1Ra in complex with IL-1R1

Table 1.1: Summary of IL-1β structures deposited in the PDB

Structures of human IL-1 β structures deposited in the PDB are listed in order of release date.

IL-1 β has also been the subject of pioneering solution NMR based studies in complex with both antigen binding fragment (Fab) and single chain Fv (scFv) antibody fragments (Addis et al., 2014; Hall, 2010; Wilkinson et al., 2009).



Figure 1.4: Structure of the IL-1 receptor signalling complex

Backbone ribbon representations of IL-1 β (green), IL-1RI (red) and IL-1RAcP (Blue) PDB code 4DEP (Thomas et al., 2012). IL-1RI forms a large interface with IL-1 β through all three domains. IL-1 β and IL-1RI form a composite interface for domains two and three of IL-1RAcP.



Figure 1.5: Structure of the IL-1 β in complex with the rapeutic antibodies

Backbone ribbon representations of IL-1 β (blue) in complex with the therapeutic antibody Canakinumab (purple) PDB code 4G6J (Blech et al., 2013). The structure of a second therapeutic antibody, Gevokizumab (Grey) (PDB code 4G6M) is shown in its relative orientation. The antibodies recognise distinct epitopes on IL-1 β .

1.5: Thesis overview

The work in this thesis describes the development of techniques to allow the generation of robust and reliable models of protein complexes via NMR data driven protein-protein docking. The focus of this technique is the identification of intermolecular NOEs across the protein-protein interface to enable the use of unambiguous distance restraints to drive the docking. However, this first requires the unambiguous assignment of side chain resonances.

Chapter two describes the development of a robust and reliable assignment strategy based on the identification of amide proton to methyl proton (H_N -CH₃) NOEs in residually protonated protein samples. The potential of this approach was theoretically proven by *in silico* studies of tri-peptides containing a central methyl-containing residue, prior to being validated on a small protein (IL-1 β), for which side-chain assignments were already available. The extension of this approach to the determination of methyl assignments in large systems was further demonstrated by assignment of the methyl groups of IL-1 β in complex with a Fab antibody fragment, which corresponds to a 65 kDa complex.

Chapter three describes the successful docking of the IL-1 β /gIC8 Fab complex using a range of experimental, NMR based restraints, and evaluation of the relative importance of different types of restraints. The results presented build upon the theses reported previously by Catherine Hall (Hall, 2010) and Ian Wilkinson (Wilkinson, 2009), where chemical shift based AIRs, together with orientational restraints from RDC data were used to produce a docked structure of IL-1 β in complex with gIC8 Fab and gIC8 scFv respectively. Whilst RDC and AIR data appeared to be sufficient to dock the IL-1 β /gIC8 scFv complex, the RDCs and AIRs obtained for the IL-1 β /gIC8 Fab complex did not produce a well converged family of structures, with a variation of approximately 10 Å in the IL-1 β position relative to gIC8 Fab across the ensemble of structures generated (Hall, 2010). Consequently, for the larger IL-1 β /gIC8 Fab complex it was not possible using the data available to determine which of the docked structures most closely resembled the actual structure of the complex.

The work described in this chapter includes the determination of a more complete set of backbone assignments for the free gIC8 Fab than was previously available, as well as

the determination of comprehensive backbone NMR assignments for gIC8 Fab in complex with IL-1 β . Together with the assignments obtained for the methyl groups of IL-1 β when bound to gIC8 Fab (determined in chapter 2) this allowed the identification of H_N-CH₃ intermolecular NOEs, which in addition to more extensive/improved chemical shift based AIR and orientational RDC data was used to dock the IL-1 β /gIC8 Fab complex. The resulting structure of the complex was also validated against a crystal structure obtained for the IL-1 β /gIC8 Fab complex, which had previously proven intractable to crystallography.
Chapter 2: Development of an assignment technique for methyl groups of large proteins

2.1: Introduction

2.1.1: Methyl Groups as Probes

Current techniques for studying large proteins by NMR generally rely on replacement of non-exchangeable protons in the protein with deuterons. Deuterons are approximately 16-fold less effective at causing relaxation due to the six-fold lower gyromagnetic ratio of ²H compared to ¹H (Cavanagh et al., 2007), this significantly attenuates ¹³C relaxation rates through ¹³C-¹H dipolar interactions as well as reducing relaxation rates of amide protons in close proximity to a substituted deuteron. Whilst allowing assignment of the backbone resonances (H_N , N, C α , C β , and CO), complete substitution of all non-exchangeable side chain protons limits the structural information that is available through ¹H-¹H inter-atomic distance restraints to those involving only amide protons. It has been shown that this is sufficient for defining the global fold of a protein, particularly when experimental conditions are optimised for detecting very long range NOEs (Koharudin et al., 2003). However, additional probes are necessary for more accurate definition of protein structure and particularly for defining proteinprotein complexes. For this purpose, methyl groups serve as a valuable tool, due to their abundance in hydrophobic protein cores and at protein-protein interfaces (Janin et al., 1988). The methyl group also consists of three protons, which are degenerate due to fast rotation about the methyl symmetry axis, giving favourable relaxation and increased sensitivity compared to other side chain resonances (Kay et al., 1992). The development of methyl-TROSY NMR has extended the use of solution NMR of large proteins even further (Ollerenshaw et al., 2003) with applications on supramolecular systems up to 1 MDa (Kay, 2011). Despite the advantages of using methyl groups as probes, production of methyl labelled samples is expensive and assignment of these groups in large systems is challenging. Techniques to overcome these challenges are described in the remainder of this section.

2.1.2: Methyl Labelling Methods

The production of isotope labelled protein for NMR studies is most commonly achieved by recombinant protein expression in a bacterial system cultured in a minimal medium containing ¹³C glucose and/or ¹⁵N ammonium sulphate as the sole carbon and nitrogen sources. Early studies of moderately sized proteins relied on random fractional deuteration, where expression is carried out in differing ratios of H_2O/D_2O , retaining some side-chain protonation to allow determination of ¹H-¹H interatomic distance restraints. It was also noted that expression in 100% D₂O minimal media results in a predictable level of deuteration of approximately 80% (Leiting et al., 1998). The deuterium incorporation was not random, the source of protonation is the protonated carbon source, which is metabolised to form the amino acids with protonation occurring mainly at methyl groups. This formed the basis of initial selective methyl labelling protocols with the labelling pattern achieved using ¹H-pyruvate (which is a precursor in the biosynthetic pathway of alanine, valine, leucine and isoleucine) as the carbon source (Rosen et al., 1996), with later characterisation of the labelling pattern achieved using ¹H-glucose as the carbon source (Shekhtman et al., 2002). However, whilst using these labelling techniques provides enrichment of methyl protonation relative to most other side chain groups, the labelling is not entirely specific. Specific labelling protocols were devised enabling highly selective protonation of the methyl groups of isoleucine (δ 1), valine and leucine, through the addition of selectively labelled amino acid precursors, α -ketobutyrate and α -ketoisovalerate (Tugarinov and Kay, 2003). This selective labelling has also been extended to include alanine by the addition of isotope labelled alanine in conjunction with α -ketobutyrate and α -ketoisovalerate to prevent scrambling (Ayala et al., 2009), methionine by the addition of labelled methionine and more recently threonine methyls (Velyvis et al., 2012).

2.1.3: Assignment Techniques

2.1.3.1: Methyl Out and Back Experiments

Assignments for ILV labelled samples can be achieved using a suite of COSY based experiments that correlate the ¹H-methyl (HM) or ¹³C-methyl (CM) chemical shifts with the amide nitrogen and proton chemical shifts in H_N-detected experiments. This requires separate experiments for the HM and CM correlations and separate experiments for the leucine/isoleucine and valine residues (Tugarinov and Kay, 2003), giving total of four experiments; Leu/Ile-HM(CMCGCBCA)NH, Vala HM(CMCBCA)NH, Leu/Ile-(HM)CM(CGCBCA)NH and Val-(HM)CM(CBCA)NH, for which the general coherence transfer pathway is illustrated for a leucine residue in Figure 2.0A. Using this series of experiments on the 82 kDa Malate synthase G (MSG) it was possible to assign 50%, 87% and 97% of the leucine, valine and isoleucine (δ 1) methyl groups respectively, with the main drawbacks being sensitivity and the lack of direct correlation of the ¹³C/¹H chemical shifts of the methyl groups (Tugarinov and Kay, 2003).

A more sensitive set of so called out-and-back experiments that correlate HM/CM chemical shifts with the C α and C β (and C γ in the case of leucine and isoleucine residues) or C` chemical shifts in ¹HM detected experiments (Tugarinov and Kay, 2003) allowed the assignment of 64%, 93% and 95% of the leucine, valine and isoleucine (δ 1) methyl groups respectively. This requires a single HMCM[CG]CBCA experiment for the leucine, isoleucine and valine residues and two experiments, a HMCM(CGCBCA)CO and a HMCM(CBCA)CO for the leucine/isoleucine and valine residues respectively. The general coherence transfer pathway of these experiments is shown in Figure 2.0B for a leucine residue. Whilst these experiments exhibit superior sensitivity to the H_N detected experiments, significant degeneracy of Ca and C β chemical shifts, particularly for leucine residues can prevent assignments (Yang et al., 2004). When both the H_N -detected and out-and-back experiments were combined 86%, 98% and 95% of the leucine, value and isoleucine (δ 1) methyl groups respectively were assigned for MSG (Tugarinov and Kay, 2003). Whilst these experiments have been very successful, full assignment of the methyl groups of MSG required 23 days of acquisition time on a 0.9 mM sample. The success of this technique is also dependent on the linearity of the labelling achieved in the branched chain amino acids (one methyl ${}^{13}\text{C/}{}^{1}\text{H}$ labelled and one methyl ${}^{12}\text{C/}{}^{2}\text{H}$ labelled).

Out-and-back experiments have subsequently been developed for the assignment of isoleucine (γ 2) methyls (Sheppard et al., 2009a), alanine methyls (Sheppard et al.,

2009b) and threonine methyls (Guo and Tugarinov, 2010). It has also been shown that separate correlation of HMCM and CG,CB or CA chemical shifts when obtaining ILV assignments results in increased sensitivity, at the cost of more experiments and hence more acquisition time (Sinha et al., 2013).

2.1.3.2: ¹³C-Detected Experiments

One approach to combat the lack of side chain protons in highly deuterated samples is the use of ¹³C-detected experiments. An experiment that correlates the ${}^{1}H/{}^{13}C$ methyl resonances of selectively methyl labelled side chains with the ¹³C resonances of the entire side-chain can be used in conjunction with experiments that correlate the amide proton and nitrogen resonances with the ¹³C resonances of the side chain. The coherence transfer pathways for these experiments are shown for a leucine residue in figure 2.0C and D. By matching the chemical shifts of the ¹³C side-chain resonances, sequence-specific methyl assignments can be made (Hu et al., 2006) and due to the increased resolution in the directly detected ¹³C dimension, resonances can be matched not only by chemical shift, but also multiplet structure, further reducing ambiguity in assignments. However, this technique has only been demonstrated on a relatively small protein of 18.6 kDa (Hu et al., 2006) and whilst ¹³C-detect experiments are becoming more common in biomolecular NMR, their implementation is not yet routine and is generally reserved for applications such as paramagnetic proteins (Bermel et al., 2003). The applicability of this approach to larger systems is probably limited due to the intrinsically low sensitivity of ¹³C-detected experiments.

2.1.3.3: Structure Based Assignment Methods

In cases where atomic resolution structures are already available from other techniques (such as X-ray crystallography) it is possible to make sequence specific assignments of methyl groups using relatively sparse NMR data, for which multiple automated procedures are available. This has been shown to be successful, resulting in 99% correct assignments for a 300 kDa oligomeric protein (Xu et al., 2009), using only 2D 1 H/ 13 C correlation spectra, 3D H_mC_mC correlation spectra and CCH-NOESY data.

These ideas have been further extended to include data from paramagnetic relaxation enhancement in the FLAMEnGO software (Chao et al., 2014, 2012). These techniques do not rely on prior backbone assignment, and as such are applicable to supramolecular systems, for which backbone assignments are not achievable, with the caveat being the need for an initial high resolution structure.

2.1.3.4: Assignment by mutagenesis

In very large systems many through-bond NMR correlations fail, although it is still possible via methyl-TROSY to acquire high quality ${}^{1}\text{H}/{}^{13}\text{C}$ correlation spectra in systems up to 1 MDa. In these cases, the only way to obtain sequence specific assignments is by mutagenesis and selective methyl labelling. Whilst time consuming, it has been shown that through parallel production of multiple mutant samples and use of fast acquisition methods assignments are feasible for large protein assemblies (Amero et al., 2011).

2.1.3.5: Assignment of Methyl Groups in Residually Protonated Samples

Whilst it is simple to produce samples with semi-selective labelling through residual protonation, it is not possible to assign these samples using the experiments that are commonly used on fully protonated samples due to the lack of protons at sites other than methyl groups. Residually protonated samples are also unsuitable for the methyl out-and-back experiments developed by Kay and co-workers due to the non-linearity of the labelling in the branched chain amino acids. One possible assignment strategy involves a TOCSY based experiment where magnetisation originates on ¹³C, is transferred by homonuclear mixing along the side chain and is detected on methyl protons (Otten et al., 2010), which also relies on selective detection of CHD₂ isotopomers. The coherence transfer pathway for this experiment is shown in Figure 2.0E. This approach was successfully applied to a 34 kDa protein, although the scope for application in systems larger than 40 kDa is not clear (Otten et al., 2010).

2.1.3.6: NOE based assignment of methyl resonances

This chapter describes the development of a robust and reliable assignment strategy based on the identification of amide proton to methyl proton (H_N -CH₃) NOEs in residually protonated protein samples. The potential of this approach was theoretically proven by *in silico* studies of tri-peptides containing a central methyl-containing residue, prior to being validated on a small protein (IL-1 β), for which side-chain assignments were already available. The extension of this approach to the determination of methyl assignments in large systems was further demonstrated by assignment of the methyl groups of IL-1 β in complex with a Fab antibody fragment, which corresponds to a 65 kDa complex.



Figure 2.0: Schematic representations of the coherence transfer pathways employed the methyl assignment techniques described above, demonstrated on a leucine residue

А General coherence transfer pathway for HM(CMCGCBCA)NH/(HM)CM(CGCBCA)NH experiments (Tugarinov and Kay, 2003) (note the linear spin system as one methyl is ${}^{12}C/{}^{2}H$ labelled). B – coherence transfer pathway for HMCMCGCBCA experiment (Tugarinov and Kay, 2003), which also requires a linear spin system. C – coherence transfer pathway for 13 C-detected HMCMC-TOCSY experiment (Hu et al., 2006), which does not require a linear spin system. D - coherence transfer pathway for ¹³C-detected HN(CA)C-TOCSY experiment (Hu et al., 2006), which also does not require a linear spin system. E coherence transfer pathway for C-TOCSY-CHD₂ experiment (Otten et al., 2010), which is effective in residually protonated samples. Green arrows indicate the coherence transfer pathways for ¹H-excite/¹H-detect experiments, red arrows indicate ¹Hexcite/¹³C-detect experiments and blue arrows indicate ¹³C-excite/¹H-detect experiments.

2.2: Materials and methods

2.2.1: Grid searches

Modelling of Gly-X-Gly tri-peptides (X being a methyl containing residue) was performed using CYANA 3.0 (Güntert et al., 1991). All dihedral angles that affect the inter-atomic distance between the methyl group(s) of the central residue and the amide proton of each of the three residues were varied systematically in 30° increments using a nested loop whilst the phi and psi angles of the C-terminal glycine were fixed at -60°, consistent with α -helical secondary structure. The set dihedral angles were reported along with the inter-atomic distances and the structural statistics for each conformer, including van der Waals violations. Inter-atomic distances and dihedral angles were reported for sterically allowed conformers by filtering for individual van der Waals violations of less than 0.25 Å and violations across the whole peptide of less than 1 Å, which was determined to limit the allowed conformers to the generally allowed regions of the Ramachandran plot. The scripts used to perform these manipulations are shown in appendix A.1.

2.2.2: Expression of IL-1β

IL-1 β with a TEV cleavable N-terminal His₆ tag, cloned into a pET-21 based vector was provided by UCB. IL-1 β plasmid was transformed into chemically competent *E. coli* BL21 (DE3) cells (Novagen) with selection for transformants achieved by plating on LB agar containing 100 µg/ml Carbenicillin. Single colonies were used to inoculate 50 ml LB starter cultures containing 100 µg/ml Carbenicillin. For the preparation of unlabelled samples, the starter cultures were used to inoculate 500 ml LB media, containing 100 µg/ml Carbenicillin in 2.5 l baffled flasks, to a starting A_{600nm} of 0.1. Cultures were incubated at 37 °C, 200rpm until the A_{600nm} reached 0.8-1.0. IPTG was added to a concentration of 0.5 mM to induce expression. Cultures were incubated at 18 °C, 200rpm for 18 hours before harvesting at 4000 g for 10 minutes. Cell pellets were stored at -20 °C prior to protein extraction and purification. For the preparation of ²H labelled samples cells were conditioned to grow in D_2O minimal media. LB starter cultures, prepared using H₂O, were used to inoculate 5 ml minimal media in a 50 ml Falcon tube, also prepared using H_2O , containing 100 µg/ml Carbenicillin to a starting A_{600nm} of 0.1 and incubated at 37 °C, 200 rpm. When the A_{600nm} reached 0.8-1.0 cells were washed by centrifugation at 10,000 g for 1 minute followed by re-suspension in minimal media prepared using 30% D₂O, and used to inoculate 5 ml of 30% D₂O minimal media containing 100 µg/ml Carbenicillin to a starting A_{600nm} of 0.1. Cultures were incubated at 37 °C, 200rpm and the process repeated in 70% and 100% D₂O minimal media, with the final 100% D₂O cultures prepared using 50 ml D₂O minimal media in 250 ml baffled flasks, to form starter cultures for the main expression media. Conditioned cells were washed twice (by centrifugation and re-suspension in 100% D₂O minimal media) to remove traces of unlabelled media and used to inoculate 500 ml D₂O minimal media cultures containing 100 μ g/ml Carbenicillin in 2.51 baffled flasks to a starting A_{600nm} of 0.1. Cultures were incubated at 37 °C, 200 rpm until the A_{600nm} reached 0.8-1.0 and induced, incubated and harvested as described above for the preparation of unlabelled IL-1 β samples.

2.2.3: Purification of IL-1β

Cell pellets were re-suspended in 30 ml of 100 mM NaCl, 25 mM Tris pH 7.5, 15 mM Imidazole, 1 mM DTT, 100 μ M AEBSF and lysed by French press. Insoluble cell debris was removed by centrifugation at 40000 g for 30 minutes. The soluble lysate was loaded onto a 1 ml Ni-NTA column, pre-equilibrated in 100 mM NaCl, 25 mM Tris pH 7.5, 15 mM Imidazole, 1 mM DTT, at 1 ml per minute. After washing with 20 column volumes of the running buffer, bound protein was eluted using a linear imidazole gradient from 15 to 500 mM over a volume of 10 ml. Fractions were analysed by SDS PAGE and those containing IL-1 β were pooled. TEV protease was added to the pooled IL-1 β which was subsequently dialysed against 41 of 100 mM NaCl, 25 mM Tris pH 7.5, 15 mM Imidazole, 1 mM DTT overnight at room temperature. The IL-1 β /TEV protease was loaded onto a 1 ml Ni-NTA column, pre-equilibrated in 100 mM NaCl, 25 mM Tris pH 7.5, 15 mM Imidazole, 1 mM DTT, at 1 ml per minute. Cleaved IL-1 β was collected in the flowthrough and bound TEV/cleaved His-tag eluted using a step

gradient from 15-500 mM imidazole. The purified IL-1 β was loaded onto a gel filtration column pre-equilibrated in 100 mM NaCl, 25 mM sodium phosphate pH 6.5, 1 mM DTT, 10 μ M EDTA, 10 μ M AEBSF. Fractions containing monomeric IL-1 β were pooled and concentrated by ultra-filtration to 400-700 μ M prior to running NMR experiments.

2.2.4: gIC8 Fab/IL-1β complex preparation and purification

Purified unlabelled samples of gIC8 Fab (supplied by UCB) were dialysed into 100 mM NaCl, 25 mM sodium phosphate pH 6.5, 10 μ M EDTA, 10 μ M AEBSF. Purified samples of IL-1 β were added to the gIC8 Fab at a 10% molar excess of IL-1 β and incubated at room temperature for one hour. The gIC8 Fab/IL-1 β mixture was then loaded onto a gel filtration column pre-equilibrated with 100 mM NaCl, 25 mM sodium phosphate pH 6.5, 10 μ M EDTA, 10 μ M AEBSF. Fractions containing gIC8 Fab/IL-1 β complex were pooled and concentrated by ultra-filtration to a concentration of 250-400 μ M.

2.2.5: NMR Spectroscopy

NMR samples were typically 380-400 µl in volume with a protein concentration of 200-700 µM in 5 mm Shigemi tubes. Buffer conditions were 100 mM NaCl, 25 mM Sodium Phosphate pH 6.5, 10 µM EDTA, 10 µM AEBSF, 0.02% NaN₃ for experiments recorded on the gIC8 Fab/IL-1 β complex and the same buffer with the addition of 1 mM TCEP HCl for experiments recorded on free IL-1 β (IL-1 β has two solvent exposed cysteine residues which enable disulphide mediated oligomerisation in the absence of gIC8 Fab). All spectra were recorded at 35 °C for free IL-1 β and 45 °C for the gIC8 Fab/IL-1 β complex on either 600 MHz Bruker DRX or 800 MHz Bruker Avance spectrometers fitted with room temperature or cryogenically cooled probes. The following two, three and four dimensional experiments were recorded to complete methyl ¹³C/¹H assignments for IL-1 β : ¹⁵N/¹H HSQC (Bodenhausen and Ruben, 1980), ¹⁵N/¹H TROSY (Pervushin et al., 1997), ¹³C/¹H HSQC, constant-time ¹³C/¹H HSQC

(Vuister and Bax, 1992), ¹⁵N/¹H NOESY-HSQC (Marion et al., 1989), ¹⁵N/¹H NOESY-TROSY (Zhu et al., 1999), ¹³C/¹H NOESY-HSQC (Muhandiram et al., 1993a), ¹³C/¹H HSQC-NOESY-¹⁵N/¹H HSQC (Muhandiram et al., 1993b). Typical acquisition times for 2D experiments were 45 ms in F1 (15N) and 80 ms in F2 (1H). NOESY-TROSY/NOESY-HSQC experiments were acquired with typical acquisition times of 16 ms in F1 (¹H indirect), 10 ms in F2 (¹⁵N) and 80 ms in F3 (¹H direct). Total acquisition times ranged from 30-50 minutes for 2D experiments and 20-90 hours for 3D experiments. Solvent suppression was achieved using the WATERGATE (Piotto et al., 1992) method where appropriate. Acquisition times were reduced where appropriate through the use of NUS (non-uniform sampling), typically recording between 40 and 50% of the Nyquist grid. Data was processed using either Topsin3.1 (Bruker Biospin ltd) or NMRPipe, with reconstruction of NUS data performed using the Harvard iterative soft thresholding method (IST) (Hyberts et al., 2012). The 4D NOESY experiment was acquired with acquisition times of 3 ms in F1 (13 C), 10 ms in F2 (1 H indirect), 6 ms in F3 (¹⁵N) and 80 ms in F4 (¹H) direct. 25% of the Nyquist grid was sampled for the 4D experiment with reconstruction of the indirect dimensions performed using 4D MDD (Orekhov and Jaravine, 2011). Data analysis was performed using SPARKY (Goddard and Kneller).

2.3: Results

2.3.1: Determination of inter-atomic H_N-CH₃ distance ranges in proteins

To investigate the feasibility of using NOEs to link methyl signals to the assigned backbone amide signals of a protein, in order to obtain sequence specific assignment of the methyl signals, it was first necessary to determine the range of inter-atomic distances that are possible in protein structures. This was achieved using molecular modelling of tri-peptides as described in section 2.2.1. By systematically varying the backbone and side-chain dihedral angles that can affect these distances and subsequent filtering based on steric clashes it was possible to determine the distance ranges that are physically possible in protein structures.

Initial testing of the modelling protocol was performed on a Gly-Ala-Gly peptide with trans-peptide bonds. The only dihedral angle that affects the intra-residue $(i)H_N$ -CH₃ inter-atomic distance is phi (φ), which is a rotation about the N-C α bond. When this dihedral angle is varied in 30° increments the distance between the amide proton (H_N) and the methyl group of the central alanine residue varies in a cyclical manner, as illustrated in figure 2.1A. The only dihedral angle that affects the inter-residue $(i+1)H_{N}$ -CH₃ inter-atomic distance is psi (ψ), which is a rotation about the C α -C bond. When this dihedral angle is varied in 30° increments the distance between the amide proton of the C-terminal glycine and the methyl group of the central alanine also varies in a cyclical manner, as shown in Figure 2.1B. These simulations give the geometrical maximum distance ranges for this peptide, however, not all combinations of these dihedral angles are possible due to steric clashes of the atoms. When conformers of the peptide are filtered based on van der Waals violations, the distance ranges are representative of what is physically possible in a protein structure. A cut off of 0.25 Å for any single violation and a sum of 1.0 Å for violations in the whole peptide was shown to limit the resulting conformers to acceptable regions of the Ramachandran plot as shown in Figure 2.2.



Figure 2.1: Inter-atomic H_N -CH₃ distances at varying backbone dihedral angles in a Gly-Ala-Gly peptide

The plots show how the inter-residue (i) H_N -CH₃ (A) and intra-residue (i+1) H_N -CH₃ (B) inter-atomic distances change with varying φ and ψ angles respectively, of the central Ala in a modelled Gly-Ala-Gly peptide. Distances were measured between H_N and the methyl pseudoatom MB.



Figure 2.2: Combinations of φ and ψ dihedral angles for the central alanine residue in sterically allowed conformers of a modelled Gly-Ala-Gly peptide

The Ramachandran plot shows the combinations of backbone dihedral angles present in the conformers of the Gly-Ala-Gly peptide after filtering based on vdw violations. Blue areas indicate that there are no allowed conformers whilst green-orange indicates an increasing number of allowed conformers. The conformers predominantly occupy the generally allowed regions of the Ramachandran plot.

The grid search was extended to include all possible combinations of the φ and ψ dihedral angles of the N-terminal glycine residue of the Gly-Ala-Gly peptide, allowing determination of the inter-atomic (i-1)H_N-CH₃ distance ranges.

This approach was successfully extended to include all of the methyl containing amino acids using Gly-X-Gly peptides, where X is Val, Thr, Leu, Ile or Met. These residues have longer side-chains requiring the sampling of all combinations of side chain dihedral angles (χ 1,2,3 e.t.c) in addition to the backbone dihedral angles. For the Gly-Val-Gly peptide the dihedral angles φ and chi1 (χ 1) are the only angles that affect the intra-residue (i)H_N-CH₃ inter-atomic distances for both methyl groups (MG1 and MG2). As with the Gly-Ala-Gly peptide, the inter-atomic distances are found to vary cyclically with the changing φ angle and the same is true for the changing $\chi 1$ angle, which is illustrated in the 3D plots shown in Figure 2.3A and B. The pattern of the inter-atomic distances is found to be the same for both of the methyl groups (H $\gamma 1$ and H $\gamma 2$) with the exception of a phase shift of approximately 120° with respect to $\chi 1$, as is expected given their geometric relationship.





The plots show how the inter-residue (i) H_N - $H\gamma 1$ (A) and (i) H_N - $H\gamma 2$ (B) and intraresidue (i+1) H_N - $H\gamma 1$ (C) and (i+1) H_N - $H\gamma 2$ (D) inter-atomic distances change with varying φ and $\chi 1$ angles for plots A and B, and ψ and $\chi 1$ angles for plots B and C, in a modelled Gly-Val-Gly peptide. Distances were measured between the amide proton and the methyl pseudoatoms MG1 and MG2. As with the Gly-Ala-Gly peptide, the conformers were filtered using the same cut offs for vdw violations. When the φ and ψ dihedral angles from the central value residue of the filtered conformers are plotted on a Ramachandran plot, as shown in Figure 2.4, the conformers are found to largely populate only the allowed regions of the phi/psi space, demonstrating that the inter-atomic distances reported from these conformers are likely to be representative of those found in proteins.



Figure 2.4: Physically possible combinations of φ and ψ dihedral angles for the central value residue in a modelled Gly-Val-Gly peptide

The Ramachandran plot shows the combinations of backbone dihedral angles present in the conformers of the Gly-Val-Gly peptide after filtering based on vdw violations. Blue areas indicate that there are no allowed conformers whilst green-orange indicates an increasing number of allowed conformers. The conformers predominantly occupy the generally allowed regions of the Ramachandran plot.

This process was repeated for Gly-Leu-Gly peptides, with the longer side-chain of the leucine resulting in an additional dihedral angle to systematically vary which is chi2 (χ 2). As before a group of conformers of the peptide was generated by systematically

varying all of the dihedral angles (backbone and side-chain) that affect the inter-atomic distance between the methyl groups of the leucine and the amide protons of the N-terminal glycine, leucine and C-terminal glycine. After filtering based on vdw violations, as described above, the resulting side chain ($\chi 1/\chi 2$) dihedral angles for the leucine residue in each conformer were plotted in a Janin plot (Janin et al., 1978), as illustrated in Figure 2.5.



Figure 2.5: Combinations of $\chi 1$ and $\chi 2$ dihedral angles for the central leucine residue in a modelled Gly-Leu-Gly peptide

The Janin plot shows the combinations of side-chain dihedral angles present in the conformers of the Gly-Leu-Gly peptide after filtering based of vdw violations. Blue areas indicate that there are no allowed conformers whilst green-red indicates an increasing number of allowed conformers. The conformers predominantly occupy the generally allowed regions of the Janin plot.

The dihedral angle plots and distance plots above show that the modelling protocol for the tri-peptides, described in section 2.2.1, is able to correctly and systematically sample all possible conformations of the torsion angle space available for these peptides. The cut-offs for filtering the resulting conformers based on vdw violations is also able to limit most of the conformers to generally allowed regions of the Ramachandran/Janin plots. This shows that the inter-atomic distances obtained from this investigation are likely to be representative of those found in protein structures. The inter-atomic H_N -CH₃ distances were determined from filtered conformers of Gly-X-Gly peptides containing all of the methyl containing amino acids (alanine, valine, leucine, isoleucine, methionine and threonine). The maximum and minimum distances from the allowed conformers were determined and are summarised in table 2.1.

Residue	Methyl	NH	Max Å	Min Å
Ala	Beta CH₃	Gly-1	7.23	4.14
		Ala-2	3.49	2.74
		Gly-3	4.37	2.86
Val	Gamma CH ₃	Gly-1	8.61	2.79
		Val-2	4.82	2.69
		Gly-3	5.27	2.88
Leu	Delta CH ₃	Gly-1	9.75	2.09
		Leu-2	5.99	2.46
		Gly-3	6.83	2.33
lle	Delta CH ₃	Gly-1	9.82	2.39
		lle-2	6.06	2.45
		Gly-3	6.81	2.46
	Gamma CH₃	Gly-1	8.6	3.67
		lle-2	4.81	2.7
		Gly-3	5.36	2.36
Thr	Gamma CH₃	Gly-1	8.61	3.75
		Thr-2	4.81	2.73
		Gly-3	5.78	2.36

Table 2.1: Inter-atomic distance ranges of atoms in sterically allowed conformers of Gly-X-Gly peptides

The maximum and minimum inter-atomic distances between the methyl groups and each of the three amide protons in the modelled tri-peptides are shown. Distances over which NOEs are expected are highlighted in green (≤ 6 Å), whilst distances over which NOEs are generally not expected, but are potentially possible in deuterated samples, are highlighted in yellow (6-8 Å). Distances were measured to pseudoatoms of the methyl groups.

The possible inter-atomic distances determined (table 2.1) show that intra-residue (i)H_N-CH₃ NOEs are always expected to be observed. The maximum distances range from 3.49-6.06 Å, with the δ -methyls of the longer side chains of leucine and isoleucine displaying the largest inter-atomic distances of 5.99 Å and 6.06 Å respectively, as would be expected. The residues with γ -methyl groups (valine and threonine) display similar maximum interatomic distances of 4.82 Å and 4.81 Å respectively, and the β methyl of alanine shows the shortest maximum inter-atomic distance of 3.49 Å. Based on these inter-atomic distances it is expected that intra-molecular (i)H_N-CH₃ NOEs should always be observed for alanine, valine and threonine residues. The maximum distance for observing NOEs in a fully protonated sample is approximately 5-5.5 Å and depends on the mixing time used for NOE build up. This increases to 6 Å when considering NOEs involving methyl groups, due to the presence of the 3 magnetically equivalent methyl protons. The distance for observing NOEs can be further extended up to 8 Å with the use of deuterated protein samples (Koharudin et al., 2003). It is expected in a deuterated sample that intra-residue (i)H_N-CH₃ NOEs should also be observed for leucine and isoleucine.

The inter-residue (i+1)H_N-CH₃ inter-atomic distances show maximum distances that range from 4.37 Å-6.83 Å. As with the intra-residue (i)H_N-CH₃ inter-atomic distances the methyl groups at equivalent positions in the side chain (β , γ or δ) display similar maximum distances. It is also expected that inter-residue (i+1)H_N-CH₃ NOEs should always be observed in deuterated protein samples. However, the grid searches also showed that inter-residue (i-1)H_N-CH₃ NOEs are not always expected, due to maximum inter-atomic distances that range from 7.23 Å for alanine residues to 9.82 Å for the δ -methyl of isoleucine. Whilst it has been shown that H_N-H_N NOEs can be observed up to a distance of 8 Å in a highly deuterated sample (Koharudin et al., 2003), in a deuterated sample with substantial residual methyl protonation it is unlikely that H_N-CH₃ NOEs will be possible over these large distance ranges for the (i-1)H_N-CH₃ NOEs (7.2-9.8 Å).

2.3.2: Assignment of the ${}^{13}C/{}^{1}H$ resonances for the methyl groups of the cytokine Interleukin 1 β using NOEs

Having shown in section 2.3.1 that it is theoretically possible to correlate methyl signals with those of backbone amide signals using NOEs (based on inter-atomic distances), it is still possible that this is not sufficient to make reliable assignments. NOEs correlate atoms through space making it possible for NOEs to occur between a methyl residues amide proton and methyl protons from other residues that are close in space but distant in amino acid sequence. One approach to addressing this ambiguity could be by only assigning methyl groups if matching i and i+1 NOEs are observed, however, depending on the tertiary structure of the protein using through space correlations rather than through bond correlations could still lead to uncertainty in the assignments obtained. IL-1 β was used as a test case to determine whether it is possible to detect and use these NOEs to make reliable assignments for the ${}^{13}C/{}^{1}H$ resonances of the methyl groups. Due to its biological interest and relatively small size, full backbone and side chain assignments for IL-1 β obtained using conventional NMR techniques have been deposited in the BMRB (Clore et al., 1990; Driscoll et al., 1990) and more recently backbone assignments have been obtained under the conditions used for the work described here (Addis et al., 2014; Wilkinson et al., 2009).

2.3.2.1: Expression and purification of II-1β:

When the expression vector for IL-1 β was transformed into BL21 DE3 cells and expression induced with an IPTG concentration of 0.5 mM at 18 °C there was a good post induction expression level with no pre-induction expression detectable by SDS PAGE. Post induction cultures in LB media grew to a final A_{600nm} of 5.5-6.0. Growth in H₂O minimal media and D₂O minimal media was reduced with a final A_{600nm} of 4.0-4.5, yet still yielded a high level of protein expression. The expressed IL-1 β was purified as described previously (section 2.2.3) using a two step Ni affinity purification followed by gel filtration (Wilkinson et al., 2009). SDS PAGE analysis of the single eluted peak from the first Ni-NTA column confirmed the presence of IL-1 β as well as some contaminants that had also bound the column (Figure 2.7). Cleavage of the His tag by Tev protease and subsequent Ni affinity purification resulted in pure IL-1 β with no contaminants detectable by SDS PAGE (Figure 2.8). Final purification was carried out by size exclusion chromatography which resulted in a single eluted peak (Figure 2.9) at the correct approximate elution volume for monomeric IL-1 β (as determined from known molecular weight standards). The yield of purified deuterated protein was approximately 18 mg/l culture.



Figure 2.7: Ni-Affinity purification of His₆-tagged IL-1β

The chromatogram (A) shows a single eluted peak (2) during a linear Imidazole gradient from 15-500 mM over 10 column volumes. SDS PAGE analysis (B) of pre and post induction samples of the whole cell lysate (pre/post) shows a band at approximately 22 kDa (His₆-IL-1 β) which is present in the soluble lysate (L) and absent in the flow-through (1). Bands in the fractions from the eluted peak (2) are consistent with His₆-IL-1 β , and higher molecular weight contaminants are also present.



Figure 2.8: Ni-Affinity purification of TEV-cleaved His₆-tagged IL-1β

The chromatogram (A) shows a single eluted peak (2) after a step Imidazole gradient from 15-500 mM. SDS PAGE analysis (B) of the pooled fractions from the first Niaffinity purification before and after adding TEV show near complete removal of the His-tag. The flow-through contained only cleaved IL-1 β (1) whilst the eluted peak (2) contains uncleaved IL-1 β , the cleaved His-tag, His₆-tagged TEV and higher molecular weight contaminants.



Figure 2.9: Size exclusion chromatography of purified human IL-1β

The chromatogram (A) shows a single eluted peak after loading purified IL-1 β on an S75 size exclusion column. SDS PAGE analysis (B) confirmed that only IL-1 β (17.4 kDa) was present.

2.3.2.2: Comparison of deuteration levels obtained from different labelling schemes

The high yield of purified IL-1 β enabled the investigation of different labelling schemes. A sample of fully protonated ¹³C labelled IL-1 β was produced to compare the level of deuteration achieved from different labelling schemes. The production of residually protonated IL-1 β was achieved by expression in 100% D₂O minimal media using ¹³C/¹H glucose as the sole carbon source. Specifically labelled samples of IL-1 β with ¹³C/¹H Leu/Val methyl groups were produced by expression in 100% D₂O minimal media using ¹²C/²H glucose as the carbon source, with the addition of 2-keto-3-(methyl-D₃)-butyric acid-4-¹³C one hour prior to induction.

Comparison of the labelling achieved by observing the peaks present in ${}^{13}C/{}^{1}H$ HSQC spectra shows that, as expected, both of the deuterated samples contain only a subset of the signals present in the fully protonated sample (Figure 2.10). The labelling patterns achieved when employing residual protonation have been extensively characterised (Otten et al., 2010; Shekhtman et al., 2002) by comparison of peak intensities. The labelling pattern obtained for IL-1 β shows a high level of protonation for all of the methyl groups, with the Ser- β and Pro- δ groups also showing a high level of protonation. The absence of signals for the α groups shows that there is a high level of deuteration at these positions, with the remaining CH₂/CH groups showing some protonation, but to a lower extent than the methyl groups. The specific Leu/Val methyl labelling was successful, showing protonation of only the Leu/Val methyl groups, with the only other peaks in the spectrum resulting from a small molecule contaminant.



2.3.2.3: Identification and assignment of methyl signals using inter-residue (i) H_N -CH₃ and intra-residue (i+1) H_N -CH₃ NOEs observed in spectra of IL-1 β

The production of purified ${}^{15}N/{}^{2}H$ labelled IL-1 β expressed with ${}^{12}C/{}^{1}H$ glucose as the sole carbon source allowed the collection of ¹⁵N/¹H NOESY-HSOC data displaying a large number of H_N -CH₃ NOEs, as shown by the ${}^1H/{}^1H$ projection in Figure 2.11A. A relatively long NOE mixing time of 600 ms was used during the experiment in order to observe NOEs up to the maximum expected inter-atomic distance of 6.8 Å. This mixing time was chosen based on previous NOESY experiments on similarly labelled samples (Veverka et al., 2008). H_N-CH₃ NOEs were identified by analysing strips of the 3D spectrum taken at the amide ${}^{15}N/{}^{1}H$ chemical shifts of the peaks located in the assigned 2D HSQC spectra. The majority of ¹⁵N strips show multiple H_N-CH₃ NOEs. For strips relating to a methyl residue, any H_N-CH₃ NOEs are matched to NOEs in the strip of the following residue (i+1). If a matching NOE exists it was assigned to the methyl (i) residue. In cases where more than the expected number of matching NOEs were present, for example more than one for alanine or more than two for valine, a confident assignment could not be made, however an ambiguous assignment could be made. In some cases, unique ¹H chemical shifts allowed identification of peaks in a 2D ${}^{13}C/{}^{1}H$ HSQC, which instantly provided the ¹³C chemical shift of the methyl group allowing full assignment of the methyl group. In other cases, the ¹H chemical shift was too similar to that of multiple peaks in the ${}^{13}C/{}^{1}H$ HSQC to allow identification. A 3D $^{13}C/^{1}H$ NOESY-HSQC such as the one shown in Figure 2.11B could potentially resolve any ambiguity, by correlating the methyl ¹H/¹³C resonances with the ¹H resonance of the amide group. This allows assignment of the methyl signals based on equivalent NOEs in both of the 3D NOESY spectra, as illustrated for the residue V40 in Figure 2.12. Once all of the confident assignments are made, ambiguous assignments could be re-evaluated.

Remaining ambiguity was further resolved by using a 4D ${}^{13}C/{}^{1}H$ HSQC-NOESY- ${}^{15}N/{}^{1}H$ HSQC experiment, in which H_{N} -CH₃ NOEs are characterised by both the ${}^{13}C$ and ${}^{1}H$ chemical shift of the methyl group and the ${}^{15}N$ and ${}^{1}H$ chemical shift of the amide group. This is illustrated for the residue L31 in Figure 2.13.



Figure 2.11: 3D ${}^{15}N/{}^{1}H$ and ${}^{13}C/{}^{1}H$ NOESY spectra recorded from residually protonated IL-1 β

2D F_1/F_3 ¹H/¹H projections of a 3D ¹⁵N/¹H NOESY-HSQC recorded on ¹⁵N/²H labelled IL-1 β expressed using ¹H glucose as the carbon source (A) and a 3D ¹³C/¹H NOESY-HSQC recorded on ¹³C/²H labelled IL-1 β expressed using ¹H glucose as the carbon source. The spectra were recorded with an NOE mixing period of 600 ms in A and 450 ms in B. The spectra display a good number H_N-CH₃ NOEs (boxed). The negative peaks (red) arise from resonances that fall outside of the spectral width in the indirect ¹⁵N/¹³C dimensions.



Figure 2.12: Example F_1/F_3 strips from 3D $^{15}N/^{1}H$ and $^{13}C/^{1}H$ NOESY spectra recorded from residually protonated IL-1 β samples

Strips corresponding to residues V40 and V41 from the ${}^{15}N/{}^{1}H$ NOESY-HSQC spectra (shown in figure 2.10A) are shown in red and similarly, strips corresponding to the two methyl groups of V40 from the ${}^{13}C/{}^{1}H$ NOESY-HSQC (figure 2.10B) are shown in green. The diagonal peaks are labelled with their corresponding assignments and the boxed peaks highlight the equivalent NOEs identified in all four strips.



Figure 2.13: Example strips from 3D and 4D ¹⁵N and ¹³C-edited NOESY spectra recorded from residually protonated IL-1β

Panel A shows F_1/F_3 strips corresponding to residue L31 and the i+1 residue (Q32) from the ${}^{15}N/{}^{1}H$ NOESY-HSQC spectra (shown in figure 2.11A) in red and similarly, a ${}^{13}C/{}^{1}H$ plane from the 4D ${}^{13}C/{}^{1}H$ HSQC-NOESY- ${}^{15}N/{}^{1}H$ HSQC taken at the L31 amide ${}^{15}N$ and ${}^{1}H$ chemical shifts is shown in blue, illustrating how the 4D NOESY data can provide the ${}^{13}C$ chemical shifts of the methyl groups assigned by matching i and i+1 NOEs. A 2D F_2/F_4 ${}^{1}H/{}^{1}H$ projection of the 4D NOESY experiment is shown in panel B, indicating the quality of the 4D NOESY data.

Using the approach developed and described above it was possible to assign the majority of the methyl groups of IL-1 β . Some of the methyl groups could not be confidently assigned due to the lack of a backbone assignment for the i+1 residue, in some cases this was due to the i+1 residue being a proline. In cases where there are multiple sequential methyl containing residues in the IL-1 β sequence it was possible to make assignments by starting with the residue closest to the C-terminal and working backwards, an example being the unambiguous assignment of the methyl groups of V41 which subsequently allowed the methyl groups of V40 to be unambiguously assigned (as shown in Figure 2.14). Using all of the available NOESY data (3D ¹³C/¹⁵N and 4D) assignments were made for 90% of the methyl groups of IL-1 β which are summarised in table 2.1. The assignments are also shown in the labelled ¹³C/¹H HSQC in figure 2.15. Many of the peaks in the ¹³C/¹H HSQC show multiple maxima. This is expected due to the isotopomeric methyl groups obtained using this labelling scheme, where CH₃, CH₂D and CHD₂ methyl groups are present (Otten et al., 2010; Shekhtman et al., 2002).

Desidue tures	Total no. of	Confidently	Tentatively	Unable to
Kesidue type.	residues	assigned	assigned	assign
Alanine	4	4	0	0
Valine	10	9	1	0
Threonine	5	4	0	1
Leucine	15	11	3	1
Isoleucine-gamma	5	4	0	1
Isoleucine-delta	5	3	1	1

Table 2.1: Summary of sequence-specific assignments obtained for the methyl residues of IL-1β

For the approach developed, side chains are confidently assigned when NOEs to the i and i+1 residue are matched (as illustrated in figure 2.12). Tentative assignments are made when there is no backbone assignment for the i+1 residue (e.g. it is a Proline) or no signals are present for the i+1 residue.



Figure 2.14: Example F₁/F₃ strips $^{15}N/^{1}H$ from a 3D NOESY recorded on a protonated IL-1β

Strips corresponding to residues V40, V41 and F42 from the ${}^{15}N/{}^{1}H$ NOESY-HSQC spectra (shown in figure 2.10A) are shown. Matching i and i+1 NOEs are boxed, with assignments indicated for the methyl ¹H NOEs. The initial unambiguous assignment of the V41 methyls reduced the number of possible for assignments for V40, allowing the subsequent unambiguous assignment of the



Figure 2.15: ¹³C/¹H CT-HSQC spectrum of ¹³C/²H labelled IL-1β

A typical 13 C/ 1 H HSQC spectrum of 13 C/ 2 H labelled IL-1 β (expressed using 1 H glucose) in 100 mM NaCl, 25 mM Sodium Phosphate pH 6.5, 1 mM TCEP-HCl, 10 μ M EDTA, 10 μ M AEBSF, 0.02% NaN₃ at 35 °C. Assigned peaks are labelled by residue and atom type and number. Assignments were made by matching intra-residue (i)H_N-CH₃ and inter-residue (i+1)H_N-CH₃ NOEs as illustrated in figure 2.11. Most peaks display multiple maxima due to the isotopomeric methyl groups obtained with this kind of methyl labelling.

2.3.3: Optimisation of the NOE-mixing period for observation of intra-residue (i)H_N-CH₃ and inter-residue (i+1)H_N-CH₃ NOEs in residually protonated protein samples

The H_N -CH₃ inter-atomic distances determined in section 2.3.1 show that intra-residue (i)H_N-CH₃ and inter-residue (i+1)H_N-CH₃ NOEs should be observed for all methyl containing residues excluding methionine. ¹H/¹H NOEs develop during a mixing period in the pulse sequence of the NOESY experiment. The NOE is a relaxation driven process, and as such it takes time for NOEs to build up. In order to use NOEs to make sequence-specific methyl assignments in proteins it is vital that these NOEs are observed in the spectra with the maximum possible signal intensity. If the mixing time is too short the NOEs will not reach maximum intensity, whereas if it is too long NOE intensity will reduce due to relaxation. The build up of the NOE is also dependent on the inter-atomic ¹H/¹H distance, which in the case of intra- and inter-(i+1) residue H_N-CH₃ can vary between 2.3 Å and 6.8 Å (as shown in section 2.3.1) depending on residue type. It is therefore necessary to optimise the length of the mixing period in the pulse sequence for observation of these NOEs.

To investigate the optimum NOE mixing period for methyl containing residues four 3D ¹⁵N/¹H NOESY-TROSY spectra were recorded with NOE mixing periods ranging from 150 to 600 ms on a sample of ${}^{15}N/{}^{2}H$ labelled IL-1 β expressed using ${}^{1}H$ glucose as the carbon source. Non-uniform sampling of the indirectly detected dimensions was employed in order to reduce the total acquisition time of these four experiments. By sampling only 40% of the normally required points the acquisition time for each experiment was reduced to approximately 28 hours. Reconstruction of the data was carried out as described in section 2.2.5 using the Harvard IST protocol. Despite the relatively short acquisition time for these experiments the reconstructed data displayed excellent signal to noise and resolution equivalent to that of uniformly sampled spectra, as illustrated in the ${}^{1}\text{H}/{}^{1}\text{H}$ 2D projections shown in Figure 2.16. Peaks corresponding to intra-residue (i)H_N-CH₃ and inter-residue (i+1)H_N-CH₃ NOEs were selected using the methyl ¹H assignments for IL-1β obtained in section 2.3.2. Only NOEs that were not significantly overlapped were used to observe the NOE build up, unless the overlapped NOEs correspond to methyl groups from the same residue (e.g MD1 and MD2 of the same leucine residue). The NOE build-ups were observed by plotting the peak height of the individual NOEs against the length of the NOE mixing period and are illustrated in Figures 2.17 to 2.20.



Figure 2.16: Reconstructed ¹⁵N/¹H NOESY-TROSY data recorded on IL-1β

2D F_1/F_3 ¹H/¹H projections of 3D ¹⁵N/¹H NOESY-TROSY spectra recorded on ¹⁵N/²H labelled IL-1 β expressed using ¹H glucose as the carbon source. The spectra were recorded with an NOE mixing period of; 150 ms (A), 300 ms (B), 450 ms (C) and 600 ms (D). The spectra are of excellent quality considering the relatively short acquisition time of 28 hours per spectrum and display good signal to noise and resolution in the indirectly detected ¹H dimension (y axis).



Figure 2.17: NOE build ups for intra-residue (i) H_N -CH₃ and inter-residue (i+1) H_N -CH₃ NOEs for the alanine residues of IL-1 β

The graph shows the change in intensity of individual NOEs with an increasing NOE mixing period in the NOESY experiment. NOE intensities were determined from the peak heights of the NOE cross peaks in ${}^{15}N/{}^{1}H$ NOESY-TROSY spectra recorded on ${}^{15}N/{}^{2}H$ labelled IL-1 β expressed using ${}^{1}H$ glucose as the carbon source. After the initial increase in intensity, the NOE intensity plateaus and in some cases starts to decrease due to T₁ relaxation during the NOE mixing period.


Figure 2.18: NOE build ups for intra-residue (i) H_N -CH₃ and inter-residue (i+1) H_N -CH₃ NOEs for the value residues of IL-1 β

The graph shows the change in intensity of a representative set of individual NOEs with an increasing NOE mixing period in the NOESY experiment. NOE intensities were determined from the peak heights of the NOE cross peaks in ¹⁵N/¹H NOESY-TROSY spectra recorded on ¹⁵N/²H labelled IL-1 β expressed using ¹H glucose as the carbon source. After the initial increase in intensity, the NOE intensity plateaus and in some cases starts to decrease due to T1 relaxation during the NOE mixing period.



Figure 2.19: NOE build ups for intra-residue (i) H_N -CH₃ and inter-residue (i+1) H_N -CH₃ NOEs for the leucine residues of IL-1 β

The graph shows the change in intensity of individual NOEs with an increasing NOE mixing period in the NOESY experiment. NOE intensities were determined from the peak heights of the NOE cross peaks in ${}^{15}N/{}^{1}H$ NOESY-TROSY spectra recorded on ${}^{15}N/{}^{2}H$ labelled IL-1 β expressed using ${}^{1}H$ glucose as the carbon source. After the initial increase in intensity, the NOE intensity plateaus and in some cases starts to decrease due to T1 relaxation during the NOE mixing period.



Figure 2.20: NOE build ups for intra-residue (i) H_N -CH₃ and inter-residue (i+1) H_N -CH₃ NOEs for the isoleucine residues of IL-1 β

The graph shows the change in intensity of individual NOEs with an increasing NOE mixing period in the NOESY experiment. The NOE build-ups for the γ -methyls are represented with the dashed lines and the δ -methyls with the solid lines. NOE intensities were determined from the peak heights of the NOE cross peaks in ¹⁵N/¹H NOESY-TROSY spectra recorded on ¹⁵N/²H labelled IL-1 β expressed using ¹H glucose as the carbon source. After the initial increase in intensity, the NOE intensity plateaus and in some cases starts to decrease due to T1 relaxation during the NOE mixing period.

When the intra-residue (i) H_N -CH₃ and inter-residue (i+1) H_N -CH₃ NOE intensities for the alanine methyl groups of IL-1 β are plotted against the length of the NOE mixing period in the NOESY experiment as shown in Figure 2.17 a typical NOE build up curve is observed. The NOE intensities increase with increasing mixing time before reaching a plateau at around 300 ms for the majority of the NOEs. A few of the NOEs show a decline in intensity between 300 ms and 600 ms mixing time. The same is true of the NOE intensities for the valine methyl groups (Figure 2.18), the majority of which reach a plateau or show only a small increase in intensity between 300 and 450 ms. Some of the NOEs show a decrease in intensity between 450 and 600 ms. Some of the lower intensity NOEs show an erratic build up but have a maximum intensity between 450 and 600 ms mixing time. Many of the leucine NOEs (Figure 2.19) show increased signal intensity up to 450 ms, with only a small increase or no change in signal intensity between 450 and 600 ms. The same is true for the isoleucine NOEs (Figure 2.20).

Therefore, a mixing time of 450 ms appears to be the optimal compromise to observe NOEs over the distance range required. At this mixing time there is not too much loss of NOE intensity from the shorter side chains of alanine and valine, whilst allowing enough time for the NOEs of the longer side chains of leucine and isoleucine to build up.

2.3.4: Identification and assignment of methyl signals using inter-residue (i) H_N -CH₃ and intra-residue (i+1) H_N -CH₃ NOEs in IL-1 β as part of a 65 kDa complex

Samples of purified IL-1 β /gIC8 Fab complex were prepared as described in section 2.2.4. Purified unlabelled gIC8 Fab, supplied by UCB, formed a tight complex with IL-1 β (K_D=80 pM (Addis et al., 2014)) that could be purified by size exclusion chromatography on an S75 column (Figure 2.21). An approximate 10% excess of IL-1 β was added to ensure that all of the Fab was bound to IL-1 β , as the complex could be easily resolved from the free IL-1 β using the S75 column.

Backbone assignments for IL-1 β when in complex with gIC8 Fab had previously been determined (Hall, 2010; Wilkinson et al., 2009) and were used to interpret ¹⁵N/¹H NOESY-TROSY spectra recorded on residually protonated ¹⁵N/²H IL-1 β in complex with unlabelled gIC8 Fab. The NOESY mixing period was adjusted to 150 ms to account for the size of the complex and the temperature at which the spectrum was recorded (45 °C) (equivalent to the 450 ms used when recording data on free IL-1 β). The ¹⁵N-edited NOESY spectrum was used with a ¹³C/¹H NOESY-HSQC recorded on residually protonated ¹³C/²H labelled IL-1 β in complex with unlabelled gIC8 Fab to assign the methyl groups by matching intra-residue (i)H_N-CH₃ and inter-residue (i+1)H_N-CH₃ NOEs as described for the assignments obtained for the free IL-1 β . Example i and i+1 strips for V40 are shown in Figure 2.22. It was possible to assign almost all of the methyl groups that were assigned for the free IL-1 β for IL-1 β in complex with gIC8 Fab using this approach.



Figure 2.21: Purification of the gIC8 Fab/IL-1β complex by size exclusion chromotography

The chromatogram (A) shows two eluted peaks. SDS PAGE analysis (B) confirms the presence of gIC8 Fab and IL-1 β in peak 1. No band is observed by SDS PAGE for peak 2, however, the elution volume and peak intensity are consistent with the 10% excess of IL-1 β that was added to gIC8 Fab to make the complex.



Figure 2.22: Example F_1/F_3 strips from 3D $^{15}N/^{1}H$ and $^{13}C/^{1}H$ NOESY spectra recorded from residually protonated IL-1 β in complex with gIC8 Fab

Strips corresponding to residue V40 and the i+1 residue (V41) from the ${}^{15}N/{}^{1}H$ NOESY-TROSY spectra are shown in red and similarly, strips corresponding to the two methyl groups of V40 from the ${}^{13}C/{}^{1}H$ NOESY-HSQC are shown in green. The diagonal peaks are labelled with their corresponding assignments and the boxed peaks highlight the equivalent NOEs identified in all four strips.



Figure 2.23: Assigned ¹³C/¹H CT-HSQC spectrum of ¹³C/²H labelled IL-1β in complex with unlabelled gIC8 Fab

 13 C/¹H HSQC spectrum of 13 C/²H labelled IL-1 β (expressed using ¹H glucose) bound to unlabelled gIC8 Fab, in 100 mM NaCl, 25 mM Sodium Phosphate pH 6.5, 1 mM TCEP-HCl, 10 μ M EDTA, 10 μ M AEBSF, 0.02% NaN₃ at 35 °C. Assigned peaks are labelled by residue and atom type and number. Assignments were made by matching intra-residue (i)H_N-CH₃ and inter-residue (i+1)H_N-CH₃ NOEs as illustrated in Figure 2.22. Most peaks display multiple maxima due to the isotopomeric methyl groups obtained with this kind of methyl labelling.

2.3.4: Identification and assignment of Serine H β signals using inter-residue (i)H_N-CH₃ and intra-residue (i+1)H_N-CH₃ NOEs in IL-1 β as part of a 65 kDa complex

During the assignment of the methyl signals of IL-1 β using matching inter-residue (i)H_N-CH₃ and intra-residue (i+1) H_N-CH₃ NOEs it was noted that a significant number of matching inter-residue (i)H_N-CH₂ and intra-residue (i+1) H_N-CH₂ NOEs were observed for Serine residues. This is due to the significant protonation of Serine H β groups that occurs when a protein is expressed in 100% D₂O minimal media using ¹H-glucose as the carbon source. This is evident in the ¹³C/¹H HSQC of IL-1 β shown in Figure 2.10B and is also entirely consistent with previously reported deuteration levels achieved using this labelling scheme. The proximity of the Serine H β groups to the backbone amide protons is also consistent with the observation of (i/i+1)H_N-CH₂ NOEs and has been previously reported with maximum distances of 4.01 and 4.64 Å for (i)H_N-CH₂ and (i+1)H_N-CH₂ respectively (Wüthrich et al., 1983). This has allowed the assignment of the Serine H β resonances for 9 of the 13 Serine residues of IL-1 β for which backbone assignments are available. Example strips from an ¹⁵N-edited NOESY-TROSY spectrum are shown in Figure 2.24.



Figure 2.24: Example F_1/F_3 strips from 3D $^{15}N/^{1}H$ edited NOESY spectrum recorded from residually protonated IL-1 β in complex with gIC8 Fab

Strips corresponding to residues S17, L18, S21, G22 S84 and V85 from a $^{15}N/^{1}H$ NOESY-TROSY spectrum are shown. Boxed peaks highlight the equivalent NOEs identified in the i and i+1 strips corresponding to the serine H β groups between 3.4 and 4.2 ppm in the F1 dimension (i+1/i-1 H_N-H_N NOEs are boxed for residues S21, G22, S84 and V85).

2.4: Discussion

The results presented in this chapter demonstrate the theoretical validation, testing, optimisation and application of an NOE based assignment approach on both a 17 kDa protein (IL-1 β) and 65 kDa protein-protein complex (IL-1 β in complex with gIC8 Fab), allowing the successful and reliable assignment of the methyl signals of IL-1 β in both the free and gIC8 Fab bound form.

Molecular modelling was used to determine the maximum possible inter-atomic distance between the protons of the methyl group(s) in a residue and the amide proton of that residue, as well as the amide protons of both the following residue (i+1) and the preceding residue (i-1). The Gly-X-Gly peptides used for the modelling protocol (X being a methyl containing residue) should be the least restrictive in terms of sterically allowed dihedral angles, allowing all possible conformations present in protein structures to be reported. The conformers produced by a grid search of all the dihedral angles of the peptide were filtered based on vdws violations, so that only inter-atomic distances from the sterically allowed conformers would be reported. The cutoffs for vdws violations were determined empirically to restrict the conformers to the generally allowed regions of the backbone (Hovmöller et al., 2002) and side-chain (Janin et al., 1978) torsion angle spaces, but were left deliberately loose to avoid filtering out any conformers that could occur in a protein structure. Observation of the dihedral angles from the filtered conformers showed that they largely conformed to the allowed regions of the Ramachandran/Janin plots, confirming the validity of this approach. The resulting inter-atomic distances showed, as expected, that the largest distances were associated with the longest side chains corresponding to leucine and isoleucine. The smallest maximum inter-atomic distances were the intra-residue distances, with the largest being the inter-residue (i-1) distances.

The inter-atomic distances determined largely agree with those reported by Wüthrich and co workers (Wüthrich et al., 1983), however, no consideration was given to steric hindrance in that study, which explains why some of the inter-atomic distances presented here are slightly shorter. A similar study investigating the use of i and i+1 NOEs to assign the entire side chains in protein used inter-atomic distances determined from a library of over 500 protein structures (Xu et al., 2009). It was found that 100% of the β and γ groups (i and i+1) showed inter-atomic distances of less than 5.5 Å. For the δ groups 99% of the intra-residue (i) distances were less than 5.5 Å and 77% of the inter-residue (i+1) distances were less than 5.5 Å. These findings are consistent with the maximum distance ranges presented in this chapter. In this previous study it was shown that in the two proteins studied, matching i and i+1 NOEs were obtained for only 50% and 75% of the methyl groups present (Xu et al., 2005). This low number is likely to be due to the use of fully protonated protein samples as well as a relatively short NOE mixing time of 50 ms.

Whilst it has been demonstrated in two separate studies that H_N-H_N NOEs can be observed up to a distance of 7 and 8 Å in perdeuterated samples with an optimised NOE mixing period (Koharudin et al., 2003; Mal et al., 1998) and CH₃-CH₃ NOEs can be observed up to a distance of 12 Å in perdeuterated, specifically Ile MD labelled samples (Sounier et al., 2007), no studies have characterised the use of long mixing times to observe long range NOEs in residually protonated samples in detail. It is unlikely that NOEs are visible up to these very long distances in residually protonated samples. However, we have shown that the substantial deuteration level that decreases the longitudinal relaxation rates, coupled with the increased NOE mixing time leads to the observation of nearly all of the expected i and i+1 H_N-CH₃ NOEs in the ¹⁵N-edited NOESY spectra recorded on IL-1 β . When compared to the structure of IL-1 β this includes NOEs over distances up to 6.55 Å. However, when additional observed NOEs are compared to the structure of IL-1 β it is clear that some NOEs are observable over inter-atomic distances up to at least 7.5 Å whilst some expected NOEs over shorter distances (6-7 Å) are not observed. The inter-atomic distance over which NOEs are detected in residually protonated samples is likely to be affected by the local proton density. As the protonation level is different depending on residue type it is possible that some areas of the protein will have higher protonation levels compared to others, an obvious example being the core of the protein where methyl residues will be abundant. It is also possible that the apparent increase in the distance over which the NOEs occur is due partly to spin diffusion. However, these NOEs are not being used for structure calculation and so the potential presence of spin diffusion is not detrimental to their use in making assignments.

The method outlined in this chapter enabled the assignment of 90% of the methyl groups of IL-1 β (excluding methyl groups of methionine residues and of the four N-

terminal residues for which no backbone assignments are available) with the majority of the assignments confirmed by matching i and i+1 NOEs. These assignments were obtained from samples which were residually protonated, which occurs when the expression is performed in D₂O minimal media with ¹H glucose as the sole carbon source. The deuteration levels achieved with this labelling scheme were consistent with published data (Rosen et al., 1996; Shekhtman et al., 2002). Virtually all expected NOEs were observed in the ¹⁵N-edited NOESY data. Some NOEs present in the 3D spectra were not observed in the 4D spectra, probably due to the decrease in sensitivity in the 4D experiment. Fewer NOEs were also observed in the ¹³C-edited NOESY experiment. The reason for the difference in the number of NOEs observed in the 3D ¹⁵N-edited and ¹³C-edited spectra is probably the introduction of the ¹³C label in the samples used to collect the ¹³C-edited data, which through dipolar interactions will increase the relaxation rate of the covalently attached ¹H in the methyl groups.

The residual protonation results in isotopomers of the methyl groups, where a mixture of CH₃, CH₂D, CHD₂ and CD₃ methyls are present. The proportion of each isotopomer is residue specific and varies significantly, although CH₂D and CHD₂ are generally the most abundant (12-49% and 30-56% respectively), with CH₃ being the least abundant (1-8%) (Otten et al., 2010). This results in the presence of three peaks for each methyl group due to the deuterium isotope effects on the ${}^{13}C/{}^{1}H$ chemical shifts of the methyl groups. The isotopomers did not affect the appearance of the ¹⁵N-edited NOESY spectra as the resolution in the indirect ¹H dimension was not sufficient to resolve them. The isotopomers were visible in the 13 C-edited NOESY data and 13 C/ 1 H HSQC spectra. It is possible to simplify the spectra, by filtering for only the CHD₂ component of the isotopomers, however, this would result in a 20-40% reduction in sensitivity (Liao and Tugarinov, 2011; Otten et al., 2010) and was considered to be unnecessary in this case. The presence of the isotopomers does prevent the utilisation of the methyl-trosy effect as it has been shown that there is no gain for the CH₂D and CHD₂ isotopomers when using HMQC compared to HSQC pulse sequences (Ollerenshaw et al., 2005). Highly selective isotope labelling solves the problem of isotopomers, but can require multiple samples and is more expensive (Ollerenshaw et al., 2005). The identification of H_N-CH₃ NOEs also requires that the protein be dissolved in H₂O rather than D₂O. Whilst the use of methyl-TROSY has been reported in H₂O, the effect is maximised in D₂O (Guo and Tugarinov, 2009). Specific labelling of IL-1 β , with ¹³C/¹H Leu/Val methyls on a ${}^{12}\text{C}/{}^{2}\text{H}$ background did result in some extra H_{N} -CH₃ NOEs, but still not as many as in the ${}^{15}\text{N}$ -edited NOESY, indicating that the use of specifically labelled samples may not offer significant improvement over residual protonation.

Residual protonation also offered the opportunity for assignment of additional side chain protons using only NOEs, namely Serine H β s. Serine (i)H_N-H β and (i+1)H_N-H β were observed for most of the serine residues of IL-1 β due to the high protonation level of Ser H β and their proximity to the backbone. C β assignments for the serine residues were already available from the backbone assignments (Wilkinson, 2009). Whilst not considered typical probes for studying large protein complexes by NMR, assignment of serine H β s could prove useful for characterisation of protein complexes. Serine residues are not found to be abundant at protein-protein interfaces in general (Ofran and Rost, 2003), however, serine is found to be enriched in paratopes (the antigen contacting surface of an antibody) (Soga et al., 2010).

The successful application of the assignment technique to the 65 kDa complex of IL-1 β and gIC8 Fab proved the applicability of the method to large protein complexes. As for the free IL-1 β , almost all of the expected NOEs were observed in the ¹⁵N-edited NOESY spectra recorded on the complex, with fewer NOEs observed in the ¹³C-edited NOESY spectra. No 4D NOESY data was available for the complex, although knowledge of the methyl assignments of free IL-1 β in conjunction with the high level of NOE completeness in the ¹⁵N-edited NOESY spectra allowed the assignment of the majority of the IL-1 β methyl groups in the IL-1 β /gIC8 Fab complex.

In summary, the general approach proposed and developed appears to represent a useful and reliable addition to the NMR-based structural biology toolkit, with particular relevance to studies of larger proteins and complexes.

Chapter 3: Determination of an NMR based model of the IL-1β/gIC8 Fab complex

3.1: Introduction

3.1.1: Protein Complexes

The majority of proteins form functional complexes through interactions with other proteins and/or nucleic acids, and the number of protein complexes is expected to exceed the number of proteins in the proteome by an order of magnitude (Bonvin, 2006). Despite this, heteromeric protein complexes represent less than 12% of the structures deposited in the PDB. This indicates that solving protein-protein complexes is more challenging than solving structures of individual proteins.

3.1.2: De Novo Docking Based Modelling of Protein Complexes

When structures of individual components of a protein complex are available, it may be possible to predict the structure of the complex using protein-protein docking approaches. This was originally tried through rigid body docking based approaches (Wodak and Janin, 1978) where comprehensive sampling over the surface of the component proteins was used to determine the lowest energy states available to the system. This *de novo* approach is generally highly reliant on shape complementarity of the proteins to be docked, although solvation and electrostatic energy terms can be included when scoring the resulting docked structures. As expected, rigid body docking is generally unable to deal with complexes where structural changes occur between the free and bound proteins, and is even highly affected by changes in side-chain conformations. So for cases where large structural changes of the backbone do not occur, incorrect conformations of side-chains in the starting model, due for example to flexibility in solvent exposed side-chains, or in crystal structures where surface side-chain conformations may be dictated by crystal contacts, can prevent the docking resulting in complexes that are close to the actual structure of the complex. For similar

reasons of uncertainty in surface loop and side chain conformations, rigid body docking is not generally suitable when using homology models as the starting structures.

The limitations of rigid body docking have lead to the introduction of degrees of flexibility in docking protocols (Bonvin, 2006), for example through soft docking, where initially some overlap of the protein surfaces is allowed, meaning that the protein surfaces to not have to be entirely complementary to achieve a good result. Another approach has been termed ensemble docking, where variations across the ensemble allow for near-native solutions (Bonvin, 2006). Introduction of some flexibility has allowed correct protein complex structures to be obtained through protein-protein docking as long as the RMSD between the starting free structures and those in the correct complex structures is relatively low (<2 Å). However, difficulty remains in selecting the correct complex structures over low scoring incorrect complex structures without the use of experimental data. This ambiguity or lack of confidence in which *de novo* docked structures most closely represent the correct complex structure is a significant barrier to using *de novo* docked complexes in structure based drug discovery.

3.1.3: Data Driven Docking of Protein Complexes

The inclusion of experimental constraints in the docking process is an obvious route to improving the reliability and accuracy of the process. Determination of correctly docked complexes from false positives can be achieved through filtering of the docked structures against experimental data, or alternatively experimental data can be incorporated into the docking protocol as structural restraints, for which a number of software packages have been developed. Perhaps however, the most widely used software for data driven docking is HADDOCK (Dominguez et al., 2003), with over 50 reports of its use to dock protein-protein complexes in the literature. HADDOCK uses ambiguous interaction restraints (AIRs) to restrict sampling during rigid body docking to regions which have been experimentally determined to be at the interaction site. The regions involved in the protein-protein interfaces can be determined by various methods including mutagenesis, H/D exchange and chemical shift perturbation from NMR data. The rigid body docking phase generates a large number of complex structures, which

are ranked by energy scores, with a subset of the structures (those with the lowest HADDOCK score) used as inputs in molecular dynamics (MD) simulated annealing with flexible side chains at the protein-protein interface. The resulting structures are further refined in a water shell, with a fully flexible protein-protein interface. HADDOCK also allows the use of non-ambiguous structural restraints in docking calculations including distance restraints from intermolecular ¹H-¹H NOEs and orientational restraints from residual dipolar couplings (RDCs).

This chapter describes the successful docking of the IL-1 β /gIC8 Fab complex using a range of experimental, NMR based restraints, and evaluation of the relative importance of different types of restraints. The results presented build upon the work reported previously by Catherine Hall (Hall, 2010) and Ian Wilkinson (Wilkinson, 2009), where chemical shift based AIRs, together with orientational restraints from RDC data were used to produce a docked structure of IL-1 β in complex with gIC8 Fab and gIC8 scFv respectively. Whilst RDC and AIR data appeared to be sufficient to dock the IL-1 β /gIC8 scFv complex, the RDCs and AIRs obtained for the IL-1 β /gIC8 Fab complex did not produce a well converged family of structures, with a variation of approximately 10 Å in the IL-1 β position relative to gIC8 Fab across the ensemble of structures generated (Hall, 2010). Consequently, for the larger IL-1 β /gIC8 Fab complex it was not possible using the data available to determine which of the docked structures most closely resembled the actual structure of the complex.

The work described in this chapter includes the determination of a more complete set of backbone assignments for the free gIC8 Fab than was previously available, as well as the determination of comprehensive backbone NMR assignments for gIC8 Fab in complex with IL-1 β . Together with the assignments obtained for the methyl groups of IL-1 β when bound to gIC8 Fab (determined in chapter 2) this allowed the identification of H_N-CH₃ intermolecular NOEs, which in addition to more extensive/improved chemical shift based AIR and orientational RDC data was used to dock the IL-1 β /gIC8 Fab complex. The resulting structure of the complex was also validated against a crystal structure obtained for the IL-1 β /gIC8 Fab complex, which had previously proven intractable to crystallography.

3.2: Materials and Methods

3.2.1: Expression of isotope labelled gIC8 Fab

gIC8 Fab, with an N-terminal leader sequence for targeting expressed protein to the periplasm, cloned into a pTTOD vector was provided by UCB. gIC8 plasmid was transformed into chemically competent *E. coli* W3110 cells (also provided by UCB) with selection for transformants achieved by plating on LB agar containing 15μ g/ml tetracycline. Single colonies were used to inoculate 10 ml LB starter cultures containing 15μ g/ml tetracycline. Cells were conditioned to grow in D₂O minimal media as described in section 2.2.2. Conditioned cells were used to inoculate 500 ml D₂O minimal media cultures containing 15μ g/ml tetracycline in 2.51 baffled flasks to a starting A_{600nm} of 0.1. Cultures were incubated at 37 °C, 200rpm for 15 hours until the A_{600nm} reached 0.8-1.0. IPTG was added to a final concentration of 200 μ M to induce expression. Cultures were incubated at 37 °C, 200rpm for 28 hours before harvesting at 6000 g for 30 minutes. Cell pellets were stored at -20 °C prior to protein extraction and purification. The supernatant containing any secreted gIC8 Fab was passed through a 0.22 μ M filter prior to ultrafiltration.

3.2.2: Purification of gIC8 Fab

To recover secreted gIC8 Fab from the expression media, the filtered media was concentrated by ultrafiltration using a tangential flow system (Vivaflow 200 1000 MWCO PES Sartorius). Cell pellets were re-suspended in 30 ml of 75 mM sodium citrate/sodium phosphate buffer pH 6.0 and lysed by French press. Insoluble cell debris was removed by centrifugation at 40000g for 30 minutes. The soluble lysate was pooled with the concentrated media and buffer exchanged into 75 mM sodium citrate/sodium phosphate buffer pH 6.0 by several rounds of dilution and concentration by ultrafiltration. The pooled lysate/media was then loaded onto a 10 ml protein G column at 0.5 ml per minute. Bound protein was eluted using a step gradient to 0.1M glycine HCl pH 2.7. Fractions containing gIC8 Fab were pooled and dialysed into 100 mM NaCl, 25 mM sodium phosphate pH 6.5. Disulphide linked di-Fab was reduced by

the addition of TCEP-HCl to a concentration of 2 mM. After incubation for 1 hour at room temperature free thiols were capped by addition of NEM to a concentration of 50 mM. After incubation at room temperature for 1 hour reduced and capped gIC8 Fab was dialysed into 100 mM NaCl, 25 mM sodium phosphate pH 6.5, before loading onto a gel filtration column pre-equilibrated in the same buffer. Fractions containing monomeric gIC8 Fab were pooled, dialysed into distilled H₂O and freeze dried prior to chemical denaturation and refolding.

3.2.3: Chemical denaturation and refolding of gIC8 Fab

To characterise the chemical denaturation of gIC8 Fab, unlabelled Fab (provided by UCB) was transferred by dilution (2 mM to 1 μ M) into buffers containing increasing concentrations of guanidine (0-6 M guanidine HCl, 100 mM NaCl, 25 mM sodium phosphate pH 6.5). The extent of protein denaturation was measured by recording the fluorescence emission spectrum of the tryptophan residues in the protein. Samples with guanidine concentrations high enough to partially or fully denature gIC8 Fab, as determined from the wavelength of the peak of the fluorescence emission, were refolded by dialysis into 100 mM NaCl, 25 mM sodium phosphate pH 6.5. Fluorescence spectra of the dialysed samples were used to determine if the protein had refolded, and A_{280nm} measurements were used to calculate the yields of refolded protein. The refolding experiment was repeated at a protein concentration of 50 µM to determine suitability for preparative scale refolding of deuterated gIC8 Fab samples. Freeze dried deuterated gIC8 Fab (section 3.2.2) was re-suspended in 6M guanidine HCl, 100 mM NaCl, 25 mM sodium phosphate pH 6.5 in a sufficient volume to give a protein concentration of 50 µM. After 4 hours at room temperature, to allow complete exchange of all backbone amide groups, denatured gIC8 Fab was refolded by dialysis twice against 2l of 100 mM NaCl, 25 mM Sodium Phosphate pH 6.5, 10 µM EDTA, 10 µM AEBSF, 0.02% NaN₃. Refolded gIC8 Fab was concentrated by ultrafiltration to 300-400 µM.

3.2.4: NMR Spectroscopy

NMR samples were typically 380-400 µl in volume with a protein concentration of 200-400 µM in 5 mm Shigemi tubes. Buffer conditions were 100 mM NaCl, 25 mM Sodium Phosphate pH 6.5, 10 µM EDTA, 10 µM AEBSF, 0.02% NaN₃. All data was recorded at a temperature of 45 °C on either 600 MHz Bruker DRX or 800 MHz Bruker Avance spectrometers fitted with room temperature or cryogenically cooled probes. The following two and three dimensional experiments were recorded to obtain complete backbone assignments and obtain structural restraints: TROSY (Pervushin et al., 1997), TROSY-HNCO (Kay et al., 1990), TROSY-HNCACB (Wittekind and Mueller, 1993) and NOESY-TROSY (G Zhu et al., 1999). Typical acquisition times for 2D experiments were 35 ms in F1 (¹⁵N) and 60 ms in F2 (¹H) and 20 ms in F1 (¹⁵N), 8 ms in F2 (¹³C) and 60 ms in F3 (¹H) for 3D triple resonance experiments, with the exception of the TROSY-HNCO where 20 ms in F2 (¹³C) was used. ¹⁵N/¹H NOESY-TROSY experiments were acquired with typical acquisition times of 16 ms in F1 (¹H indirect), 10 ms in F2 (¹⁵N) and 60 ms in F3 (¹H) direct. Total acquisition times ranged from 30-50 minutes for 2D experiments and 20-90 hours for 3D experiments. Solvent suppression was achieved using the WATERGATE method (Piotto et al., 1992) for all experiments described here. Total acquisition times were reduced where appropriate through the use of NUS (non-uniform sampling), typically recording between 20 and 50% of the Nyquist grid. Data was processed using either Topsin3.1 (Bruker Biospin ltd) or NMRPipe, with reconstruction of NUS data performed using the Harvard iterative soft thresholding method (IST) (Hyberts et al., 2012). Data analysis was performed using SPARKY (Goddard and Kneller).

3.2.5: Determination of RDCs for the gIC8 Fab/IL-1β complex

RDCs were calculated from spectra previously acquired by Catherine Hall (Hall, 2010). Isotropic samples consisted of either 300 μ M ¹⁵N/²H gIC8 Fab in complex with unlabelled IL-1 β or 200 μ M unlabelled gIC8 Fab in complex with ¹⁵N/²H IL-1 β in 100 mM NaCl, 25 mM Sodium Phosphate pH 6.5, 0.02% NaN₃. Partially aligned samples were the same but with the addition of 3 mg/ml Pf1 phage (Asla biotech). RDCs were calculated as the difference in ¹⁵N/¹H scalar couplings between isotropic and partially aligned samples, which were measured as double the difference in peak position in ¹⁵N/¹H HSQC and TROSY spectra (Kontaxis et al., 2000). RDCs for IL-1 β were used

as calculated by Catherine Hall (Hall, 2010) and RDCs for gIC8 Fab were recalculated (using the spectra recorded by Catherine Hall) to include the extended assignments for gIC8 Fab when in complex with IL-1 β .

3.2.6: Refinement of a homology model of gIC8 Fab

A homology model of gIC8 Fab had previously been generated by Jiye Shi (UCB). Refinement of the homology model for gIC8 Fab was carried out using XPLOR-NIH (Schwieters et al., 2006, 2003), using RDC and dihedral angle restraints. Dihedral angle restraints were generated using TALOS+ (Shen et al., 2009) with automatic correction of chemical shifts applied to account for deuterium shift. Only unambiguous predictions were used to generate restraints and these were limited to a minimum range of $\pm 15^{\circ}$. RDC data were included as direct restraints with the axial and rhombic components estimated using PALES (Zweckstetter and Bax, 2000). Generation of the input homology model has already been described in detail.

3.2.7: Defining ambiguous interaction restraints (AIRs)

Potential residues at the interaction site with IL-1 β on gIC8 Fab were identified using chemical shift perturbation upon binding of IL-1 β to gIC8 Fab. The combined backbone amide shift was calculated using a scaling factor of 0.2 for ¹⁵N shifts to account for the difference in spectral widths. Residues with a combined amide chemical shift perturbation greater than 0.1 ppm, and solvent accessibility greater than 10%, were defined as active. Residues with solvent accessibility greater than 10% that were adjacent to active residues were defined as passive. Active and passive residues for IL-1 β had previously been determined (Hall, 2010).

3.2.8: Intermolecular NOEs

Intermolecular H_N -CH₃ NOEs between IL-1 β and gIC8 Fab were identified using an asymmetric isotope labelling approach, where one component of the complex is ¹⁵N/²H labelled (expressed using ²H glucose as the carbon source) and the other is unlabelled. ¹⁵N/¹H NOESY-TROSY spectra recorded on such samples will show intra-molecular and inter-molecular H_N - H_N NOEs, but any H_N -CH_x NOEs will have to be intermolecular, as any CH_x groups in the ¹⁵N labelled protein will be deuterated. This approach offers greater sensitivity compared to those that rely on isotope filtered experiments. Assignment of the intermolecular NOEs was done in the first instance by matching ¹H chemical shifts to assigned methyl signals (±0.03ppm). Where this leaves multiple possible assignments, potential assignments were filtered based on proximity to the presumed interaction site (assumed to be residues used as AIRs, defined in section 3.2.5). Any remaining ambiguity was resolved based on proximity to groups already assigned to intermolecular NOEs.

3.2.9: Docking

Docking the gIC8 Fab/IL-1 β complex was achieved using the docking protocol in HADDOCK (Dominguez et al., 2003). Input structures for the docking calculations were a crystal structure of IL-1 β (2I1B) (Priestle et al., 1989) and a refined homology model of gIC8 Fab (produced as described in section 3.2.4). Intermolecular restraints for docking included ambiguous interaction restraints (AIRs) (section 3.2.5), unambiguous distance restraints (NOEs) (section 3.2.6) and RDCs (section 3.2.4) which were included as both direct restraints and intervector projection angle restraints. TALOS restraints (section 3.2.4) and H_N-H_N NOEs were included as intramolecular restraints to help define the structure of any residues that become flexible as part of the docking protocol. Semi-flexible residues were defined using default HADDOCK parameters. Three docking runs were carried out using different combinations of the intermolecular restraints to assess the impact of the different restraints on the docking outcome. This included one run using only RDCs and AIRs as intermolecular restraints, one run using only NOEs and AIRs and a final run using all of the available data (RDCs, NOEs and AIRs). 5% of the AIRs were randomly excluded from each structure calculation. For each run 1000 structures were generated in the initial rigid body docking stage, with the best 200 structures used during the simulated annealing and water refinement stages.

3.2.10: Crystallisation of the IL-1β/gIC8 Fab complex

Crystallisation trials were set up using the MorpheusTM, PACT, JCSG+, MIDAS, and Clear Strategy 1 screens (Molecular Dimensions). Purified IL-1 β /gIC8 Fab complex (prepared as described in section 2.2.4) at a protein concentration of 10 mg/ml was mixed in a 1:1 ratio with precipitant giving a final volume of 0.2 µl. Crystals grew in several conditions within one day at 17 °C. Optimisation of the conditions was achieved by varying pH and precipitant concentration in 2 µl hanging drop (1:1 protein and precipitant) format.

3.3: Results

3.3.1: Expression and purification of gIC8 Fab

Conditions for expression of isotopically labelled gIC8 Fab using an E.coli expression system had been previously determined within the lab (Hall, 2010). Expression in W3110 cells (UCB) using 200 µM IPTG induction at 37 °C for 28 hours gave cultures that grew to a post induction A_{600nm} of 1.5-2.0 and showed a good post induction expression level when analysed by SDS PAGE (Figure 3.1B). It had also been previously observed that, during post induction incubation, periplasmic gIC8 Fab leaked into the expression media. To maximise the yield of purified protein, gIC8 Fab was extracted from the periplasm via French press, and recovered from the media using ultra-filtration. Purification of the resulting lysate/concentrated media by protein G affinity chromatography resulted in a single eluted peak that was confirmed to contain gIC8 Fab when analysed by SDS PAGE (Figure 3.1A). When analysed by SDS PAGE under non-reducing conditions it was clear that a significant proportion of the purified gIC8 Fab had formed disulphide linked dimers via a solvent exposed thiol in the C terminal region of the heavy chain (Figure 3.2B). After reduction of the disulphide linked dimeric Fab using TCEP and capping of free thiols using NEM, size exclusion chromatography gave a single eluted peak of pure monomeric gIC8 Fab. The yield of purified deuterated gIC8 Fab was approximately 6 mg/l culture.



Figure 3.1: Protein G affinity purification of gIC8 Fab

The chromatogram (A) shows a single eluted peak after a step gradient to 0.1M Glycine HCl pH 2.7. SDS PAGE analysis under reducing conditions is shown in panel B with lanes containing marker (M), the load (L) and fractions collected from peaks 1 and 2 (A). Fractions from the eluted peak (2) show a band between the 20 and 30 kDa marker, consistent with both the heavy chain (25 kDa) and the light chain (23 kDa), which have not been resolved on the gel. The band at approximately 50 kDa is consistent with disulphide linked heavy and light chain.



Figure 3.2: Size exclusion chromatography of purified gIC8 Fab

The chromatogram (A) shows one eluted peak, consistent with monomeric gIC8 Fab. SDS PAGE analysis under non-reducing conditions (B) confirmed the presence of gIC8 Fab. The band at 25 kDa is consistent with both the heavy and light chain and the band at 50 kDa is consistent with disulphide linked heavy and light chain. The two higher molecular weight bands are consistent with 3 and 4 disulphide linked chains (dimeric Fab).

3.3.2: Refolding of deuterated gIC8 Fab after chemical denaturation

Proteins expressed in D_2O minimal media will initially have deuterated amide groups, which does not allow the use of NMR experiments where the signal is excited or detected on the backbone amide proton. In order to record such experiments the deuterated amide groups must be exchanged with protons to give protonated amide groups. This occurs at solvent exposed amide groups during protein purification in H₂O buffer, however, amide groups in the core of the protein, or those involved in hydrogen bonds can remain deuterated. It is therefore necessary to fully denature the protein, allowing complete backbone amide exchange prior to refolding.

Purified gIC8 Fab was denatured as described in section 3.2.3 using Guanidine HCl in H₂O buffer. Unfolding of unlabelled samples of gIC8 Fab was monitored using the intrinsic tryptophan fluorescence emission spectrum. Unfolding of the protein began at a guanidine concentration of 2.5-3M and was complete at 4.5M (Figure 3.3) and showed typical cooperative unfolding. Unfolded samples of gIC8 Fab were refolded by dialysis into buffer that did not contain guanidine HCl as described in section 3.2.3. No protein precipitation was observed on refolding and the fluorescence emission spectrum post dialysis was indicative of a folded protein (data not shown). The yield of refolded protein was determined to be greater than 95% after calculation of protein concentrations from the measured A_{280nm}. The chemical denaturation and refolding protocol was successfully scaled up as described in section 3.2.3 to allow refolding of milligram quantities of deuterated gIC8 Fab. TROSY spectra from the original, nonrefolded and refolded gIC8 Fab overlay very well (Figure 3.3D), indicating that the refolded gIC8 Fab has adopted an identical structure to that of the non-refolded material. Comparison of the TROSY spectra shows a small number of extra peaks in the refolded spectrum and a significant number of peaks showing increased signal intensity, indicating that the amide groups in the core of the protein had exchanged with the solvent and have been fully protonated.



Figure 3.3: Chemical denaturation of gIC8 Fab to exchange deuterated backbone amide groups

Cooperative unfolding of gIC8 Fab is shown by the increase of the emission wavelength of tryptophan fluorescence from 330nm (buried) to 350nm (solvent exposed) (B). $^{15}N/^{1}H$ TROSY spectra of non-refolded (red A) and refolded (black C) gIC8 Fab shows a good overlay of peak position when the spectra are superimposed (D), with multiple extra peaks in the black (refolded) spectrum such as those highlighted with the arrows.

3.3.3: Assignment of the backbone NMR signals of gIC8 Fab in both the free and IL-1 β bound form.

The production of milligram quantities of purified, ¹⁵N/¹³C/²H labelled gIC8 Fab allowed the acquisition of a series of high quality triple resonance and 3D NOESY NMR data sets. Despite the size of the protein (48 kDa/444 residues) the majority of peaks in both 2D and 3D data sets were well resolved (Figure 3.4/7). A significant proportion of the backbone assignments for gIC8 Fab had already been completed using non-refolded protein (Hall, 2010). These backbone assignments were confirmed and extended using 3D trHNCACB data (Figure 3.6) recorded on the refolded gIC8 Fab with fully back-exchanged amide groups. The trHNCACB data is viewed as 2D strips taken at the amide proton (H_N) and amide Nitrogen $({}^{15}N)$ chemical shifts of the peaks (as shown in Figure 3.6) which are located using the 2D TROSY data (Figure 3.4). Each strip contains up to 4 peaks (assuming no overlap of the spin systems), 2 intraresidue peaks which correlate the C α and C β signals with the amide proton and nitrogen signals and two inter-residue peaks which correlate the amide proton and amide nitrogen signals with the C α and C β of the preceding (i-1) residue (figure 3.6). C α peaks have a positive phase and $C\beta$ peaks have a negative phase. Signals from sequential residues are identified by matching the intra-residue peaks from one strip to the inter-residue peaks of another strip as illustrated by the dashed lines in Figure 3.6. Sequence-specific assignments were completed by matching signals from stretches of sequential residues to the amino acid sequence using residues with distinctive signals, for example glycine which has no C β signals or alanine, serine and threonine which have distinctive C β signals. Backbone assignments were further confirmed by the presence of i±1 H_N-H_N NOEs in ¹⁵N/¹H NOESY-TROSY data recorded on refolded gIC8 Fab. This allowed assignment of a high proportion of the backbone atoms (HN, N, $C\alpha$, $C\beta$ and CO) of gIC8 Fab, with 333 (79%) of the 419 assignable backbone amide groups identified (excluding prolines and the two N-terminal residues), as summarised in Figure 3.8. Backbone amide signals could not be assigned for Ile², Gln⁶-Ser⁷, Ser⁹, Asn²⁸, Trp³⁵, Gly⁴¹, Gln⁴⁵, Gly⁶⁶, Trp⁹²-Leu⁹⁴, Phe⁹⁶, Asp¹²², Ser¹⁵⁶, Gly¹⁵⁷, Ser¹⁶⁸, Ser²⁰², Ser²⁰³, Gly²¹²-Cys²¹⁴, Glu²¹⁵, Val²¹⁶, Gly²²², Asp²⁴², Asp²⁴⁷, Leu²⁵⁹, Val²⁶²-Tyr²⁶⁴, Gly²⁶⁸, Gly²⁷⁰, Ser²⁷¹, Tyr²⁷³, Phe²⁷⁴, Asp²⁷⁶, Thr²⁷⁷, Gly²⁸⁰, Tyr³⁰⁸-Asp³²¹, Ser³³³, Ser³⁴⁷-Thr³⁵⁵, Val³⁶²-Phe³⁶⁵, Thr³⁸⁰-Gly³⁸², Ser³⁹², Ser³⁹³, Ser⁴⁰⁶, Ser⁴⁰⁷ and Lys⁴³⁴-Ala⁴⁴³ as reported recently (Addis et al., 2014).

Purified ¹⁵N/¹³C/²H labelled gIC8 Fab in complex with unlabelled IL-1β (produced as described in section 2.2.4) also allowed the acquisition of high quality 2D and 3D data sets (Figure 3.6/7) that were used to obtain essentially complete backbone assignments for gIC8 Fab in the bound form, as described above. Assignments for the bound gIC8 Fab were more complete than those of the free protein with 358 (85%) of the 419 assignable backbone amide groups identified. Backbone amide signals could not be assigned for Ile², Gln⁶-Ser⁷, Ser⁹, Asn²⁸, Ile²⁹, Gly⁴¹, Gln⁴⁵, Ser¹²¹, Asp¹²², Ser¹⁵⁶, Gly¹⁵⁷, Gln¹⁶⁶, Ser¹⁶⁸, Ser²⁰², Ser²⁰³, Gly²¹²-Cys²¹⁴, Glu²¹⁵, Val²¹⁶, Gly²²², Asp²⁴², Trp²⁵⁰, Val²⁶², Tyr³⁰⁹, Cys³¹⁰, Ser³³³, Ser³⁴⁷-Thr³⁵⁵, Val³⁶²-Phe³⁶⁵, Thr³⁸⁰-Gly³⁸², Thr³⁸⁵, Ser³⁹², Ser³⁹³, Ser⁴⁰⁶, Ser⁴⁰⁷ and Lys⁴³⁴-Ala⁴⁴³.



Figure 3.4: Assigned ¹⁵N/¹H TROSY spectrum of ¹⁵N/²H labelled refolded gIC8 Fab

A typical ¹⁵N/¹H TROSY spectrum of gIC8 Fab in 100 mM NaCl, 25 mM Sodium Phosphate pH 6.5, 10 μ M EDTA, 10 μ M AEBSF, 0.02% NaN₃ at 45 °C. Assigned peaks are labelled by residue type and number. The spectrum displays good signal dispersion despite the protein having 444 residues, with 383 peaks visible compared to the 419 expected for gIC8 Fab. The lower spectrum is a larger view of the crowded central region of the spectrum above.



Figure 3.5: Assigned $^{15}\text{N/}^{1}\text{H}$ TROSY spectrum of $^{15}\text{N/}^{2}\text{H}$ labelled gIC8 Fab in complex with unlabelled IL-1β

A typical ¹⁵N/¹H TROSY spectrum of gIC8 Fab in complex with unlabelled IL-1 β in 100 mM NaCl, 25 mM Sodium Phosphate pH 6.5, 10 μ M EDTA, 10 μ M AEBSF, 0.02% NaN₃ at 45 °C. Assigned peaks are labelled by residue type and number. The spectrum displays good signal dispersion despite the protein having 444 residues and being part of a 65 kDa complex, with 396 peaks visible compared to the 419 expected for gIC8 Fab.



recorded on ¹⁵N/¹³C/²H labelled gIC8 Fab

Strips from a 3D trHNCACB spectrum taken at the backbone amide ${}^{15}N/{}^{1}H$ chemical shift for residues Tyr 36-Lys 39 showing intra and inter-residue cross peaks with C α peaks shown in red and C β peaks shown in green. Dashed lines indicate sequential connectivities. These strips correspond to peaks that are only present in the spectra of refolded gIC8 Fab.



Figure 3.7: Strip plots representative of the 3 dimensional trHNCACB data recorded on ${}^{15}N/{}^{13}C/{}^{2}H$ labelled gIC8 Fab in complex with IL-1 β

 15 N strips for residues Tyr 36-Lys 39 showing intra and inter-residue cross peaks with C α peaks shown in red and C β peaks shown in green. Dashed lines indicate sequential connectivities.

Light	chain:					
	1 <u>0</u>	2 <u>0</u>	3 <u>0</u>	4 <u>0</u>	5 <u>0</u>	6 <u>0</u>
Free	DIQMTQSPSS	LSASVGDRVT	ITCRTSGNIH	NYLI <mark>WYQQ</mark> KP	GKAPQLLIYN	AKTLAD <mark>GVPS</mark>
Bound	DIQMTQSPSS	LSASVGDRVT	ITCRTSGNIH	NYLIWYQQKP	GKAPQLLIYN	AKTLADGVPS
	7 <u>0</u>	8 <u>0</u>	90	10 <u>0</u>	11 <u>0</u>	12 <u>0</u>
Free	RFSGSGSGTQ	FTLTISSLQP	EDFANYYCQH	FWSLPFTFGQ	GTKVEIKRTV	AAPSVFIFPP
Bound	RFSGSGSGTQ	FTLTISSLQP	EDFANYYC <mark>Q</mark> H	FWSLPFT <mark>F</mark> GQ	GTKVEIKRTV	AAPSVFIFPP
	130	140	150	160	170	180
Free	SDEOLKSGTA	SVVCLUNNFY	PREAKVOWKV	DNALOSGNSO	ESVTEODSKD	STYSLSSTLT
Bound	SDEQLKSGTA	SVVCLLNNFY	PREAKVOWKV	DNALOSGNSO	ESVTEODSKD	STYSLSSTLT
	~		~	~ ~ ~ ~	~~~	
	19 <u>0</u>	20 <u>0</u>	21 <u>0</u>			
Free	LSKADYEKHK	VYACEVTHQG	LSSPVTKSFN	RGEC		
Bound	LSKADYEKHK	VYACEVTHQG	LSSPVTKSFN	RGEC		
	, ·					
неаvу	chain:			220	230	240
Free				EVOLVE	SGGGLVOPGG	ST.RI.SCAASG
Bound				EVOLVE	SGGGLVOPGG	SLRLSCAASG
				~	~ ~ ~	
	250	26 <u>0</u>	270	280	29 <u>0</u>	30 <u>0</u>
Free	FDFSRYDMSW	VRQAPGKRLE	WVAYISSGGG	STYFPDTVKG	RFTISRDNAK	NTLYLQMNSL
Bound	FDFSRYDMSW	VRQAPGKRLE	WVAYISSGGG	STYFPDTVKG	RFTISRDNAK	NTLYLQMNSL
	21.0	220	220	240	250	2.00
Free	DAEDTAUVVC	V DONKKI TWE			VEDT V DOCKO	
Bound	RAEDIAVIIC	ARQNKKL.TWF		VSSASINGES	VFPLAPSSKS	TSGGTAALGC
Bound	INEDIAVIIC	AIQUILIDIWI	DIWGQIIVI	V D D D D D D D D D D D D D D D D D D D	VEL DAL SSI(S	ISGGIAADGC
	370	380	390	400	410	420
Free	LVKDYFPEPV	TVSWNSGALT	SGVHTFPAVL	QSSGLYSLSS	VVTVPSSSLG	TQTYICNVNH
Bound	LVKDYFPEPV	TVSWNSGALT	SGVHTFPAVL	QSSGLYSLSS	VVTVPSSSLG	TQTYICNVNH
	4.2.0					
Erco	43 <u>0</u>	44 <u>0</u>				
rree Bound	KESNIKVDKK	VEPRSCUATH				
	ILL DIVITION DIVIN	A THUNCOUTU	TOUU			

Figure 3.8: Summary of backbone assignments for free and bound gIC8 Fab

Amino acid sequence of gIC8 Fab, residues from the light chain are numbered 1-214 and residues from the heavy chain are numbered 215-444. Residues from the CDR loops are boxed in green. Residues for which backbone assignments were determined are highlighted in grey, the upper sequence representing assignments for free gIC8 Fab and the lower representing gIC8 Fab in complex with IL-1 β . Significant stretches of missing assignments are present only in the constant domain of the heavy chain (346-355, 362-366 and 433-443), CDR3 of both chains in the free gIC8 Fab, and CDR2 of the heavy chain in the free gIC8 Fab.

3.3.4: Refinement of the homology model of gIC8 Fab

As no structure of gIC8 Fab was available at the onset of this project, structural characterisation of the gIC8 Fab/IL-1 β complex required the generation of a homology model. A homology model of the variable domains of gIC8 Fab, in the form of a single chain Fv (scFv), had already been produced by Jiye Shi (UCB), as reported previously (Wilkinson et al., 2009). This model was subsequently refined against backbone amide RDC data recorded on gIC8 scFv in complex with IL-1 β (Wilkinson et al 2009). This refined model was used as a template for the variable domains when producing a homology model of gIC8 Fab (also performed by Jiye Shi). This second generation homology model was subsequently refined against backbone amide RDCs, TALOS restraints and backbone amide (H_N-H_N) NOEs using XPLOR as described in section 3.2.5, before being used as the input structure for docking calculations of the gIC8 Fab/IL-1 β complex.

3.3.4.1: Determination of backbone amide RDC values for gIC8 Fab when in complex with IL-1 β

Backbone amide N-H RDCs had previously been determined for the gIC8-IL-1 β complex by measuring the difference between ¹⁵N/H_N scalar couplings in isotropic and partially aligned samples (Hall, 2010). However, this RDC data was incomplete due to lack of assignments for gIC8 Fab in the bound form, allowing calculation of backbone amide RDCs only for residues for which assignments could be confidently transferred from free gIC8 Fab data to bound. This resulted in 201 RDCs covering 46% of the protein. The spectra of isotropic and partially aligned gIC8 Fab in complex with IL-1 β recorded by Catherine Hall (Hall, 2010) were used with the extended assignments for gIC8 Fab in the bound form, obtained in this study, to determine RDCs for 242 backbone amide groups, covering 56% of the protein (summarised in Figure 3.9). The ¹⁵N/H_N scalar couplings were determined from the difference in peak position between TROSY and HSQC spectra. Due to signal overlap in the 2D spectra used it was not possible to determine a more complete set of backbone amide RDCs.


Figure 3.9: Summary of backbone amide RDC values obtained for gIC8 Fab bound to IL-1β

Backbone amide RDC values obtained for ${}^{15}N/{}^{2}H$ labelled gIC8 Fab in complex with unlabelled IL-1 β are shown for residues in the light chain (upper plot) and the heavy chain (lower plot) (spectra recorded by Catherine Hall). Partial alignment was achieved using 3 mg/ml Pf1 phage. The measured RDCs range from +18.9 Hz to -17.3 Hz. Errors for the RDCs were set to 1.4 Hz based on a conservative estimate for the error in peak positions in the processed spectra. The green bars indicate the location of the CDRs in the variable domains of each of the chains.

3.3.4.1: Estimation of the backbone dihedral angles of gIC8 Fab when in complex with IL-1β from backbone chemical shift data

The backbone assignments for gIC8 Fab in complex with IL-1 β allowed prediction of phi and psi dihedral angles using TALOS+ (Shen et al., 2009), which uses chemical shifts to predict backbone dihedral angles by comparison to a reference database of structures. The samples used to obtain backbone assignments for gIC8 Fab were deuterated, which has an effect on the chemical shift of the C α and C β chemical shifts (Venters et al., 1996). Automatic correction of C α and C β chemical shifts was applied within TALOS+ to account for deuterium shifts present. Ambiguous predictions and predictions of non-regular secondary structure were discarded and upper bounds for angle restraints were limited to greater than $\pm 15^{\circ}$, resulting in 634 dihedral angle restraints covering 71% of the protein as illustrated in Figure 3.10.





Dihedral angle restraints are shown for the light chain (upper plot) and heavy chain (lower plot) of gIC8 Fab. Constraints for φ are shown in blue and those for ψ in red, with error bars indicating the upper and lower bounds of each constraint.

3.3.4.2: Refinement of the homology model of gIC8 Fab against backbone amide RDC and TALOS restraints

The homology model for gIC8 Fab was refined using XPLOR-NIH (Schwieters et al 2006) by energy minimisation against 242 experimentally determined RDCs and 634 backbone dihedral angle restraints. The initial homology model showed reasonable agreement with the experimentally determined RDCs with a Q value of 0.35 when analysed using PALES (Figure 3.11). Refinement did not result in any large structural changes with an average C α RMSD of 1.1 Å (Figure 3.12). The refined model showed very good agreement between experimentally observed and calculated RDCs with a Q value of 0.09, however during refinement, Ramachandran outliers increased from 2 (0.5%) in the starting homology model to 22 (5.1%) in the refined homology model. Overall the refined homology model was deemed to be of sufficient quality to use as an input for docking calculations of the gIC8 Fab I-1 β complex.



Figure 3.11: Comparison of observed and calculated backbone amide RDCs for the gIC8 homology model before and after refinement against the experimental data

The two graphs show the agreement between observed backbone amide RDCs and those calculated from the homology model using PALES. Q factors for each of the fits are indicated on each plot. The error bars on the red data point indicate the 1.4 Hz estimated error present for each RDC.



Figure 3.12: Comparison of the backbone topology of the gIC8 Fab homology model before and after refinement against backbone amide RDCs and TALOS restraints.

The backbone ribbon representation of the homology model for gIC8 Fab is shown (A) next to the model that has been energy minimised against the experimental data (B). An overlay of the structures (C) shows that they are very similar with a C α RMSD of 1.1 Å for all residues. The variable and constant domains for the heavy and light chains are labelled on the homology model (A).



3.3.5: Input data for docking calculations

3.3.5.1: Identification of gIC8 Fab residues involved in binding to IL-1β

As comprehensive backbone NMR assignments have been obtained for gIC8 Fab in both the free and IL-1ß bound form it was possible to use direct comparison of backbone amide ¹⁵N/H_N chemical shifts to identify residues potentially involved in the interaction with IL-1B. Combined backbone amide chemical shift changes are shown in Figure 3.13. Significant chemical shift changes were observed only in the variable domains of both the light and heavy chains. Specifically, CDR1 and residues flanking CDR3, as well as some framework residues of the light chain show large chemical shift changes. CDR1, CDR2, residues flanking CDR3 and some framework residues in the heavy chain also show large chemical shift changes. A number of residues that were assigned in the IL-1 β bound form of gIC8 Fab could not be identified in spectra recorded on free gIC8 Fab. As noted previously (Hall 2009 and Wilkinson et al 2009), it is likely that some residues of the gIC8 Fab are present in multiple conformations, exchanging on an intermediate NMR timescale. This would result in broadening of the NMR signals, precluding assignment of these residues. These residues could be stabilised into one conformation upon binding to IL-1β, allowing their assignment in the IL-1β bound gIC8 Fab. The majority of these residues are located in CDR3 of both the light and heavy chains and CDR2 of the heavy chain as illustrated in Figure 3.13.

When mapped onto surface views of the refined model of gIC8 Fab the largest shifts are seen at or near the CDR loops, at the top of the structure as represented in Figure 3.14. However, there are also considerable shifts observed in residues of the framework of the variable domains that are distant from the expected antigen binding site. These shifts are unlikely to be caused by direct interaction with IL-1 β and could possibly indicate structural changes in gIC8 Fab that occur upon antigen binding. The lack of significant chemical shift changes in the constant domains of gIC8 Fab confirms that any potential structural changes that occur upon IL-1 β binding are not transmitted through to the constant domains. The majority of the shifts in the variable domains that are not related to the antigen binding site are found at the interface of the variable domains, as shown in the ribbon representation of the structure in Figure 3.14. This could be due to a change in the relative orientation of the variable domains that occurs upon antigen binding, as has been proposed previously (Hall, 2010).



Figure 3.13: Mapping of the IL-1β interaction site on the gIC8 Fab sequence

Combined backbone amide (N and H_N) chemical shift changes between free gIC8 Fab and gIC8 Fab in complex with IL-1 β . Residues marked in red indicate that no chemical shift data is available. The blue bars indicate that assignments are available for these residues in the IL-1 β bound form of gIC8 Fab but not in the free form. The CDR positions are marked in green.



surface views of the refined gIC8 Fab homology model. Residues with combined shifts <0.02 ppm are shown in white, between 0.02 and 0.1 ppm on a linear gradient from white to red and >0.1 ppm in red. Residues for which no chemical shift data is available are shown in yellow. Residues for which assignments are available for gIC8 Fab in complex with IL-1 β but not in the free form are shown in blue.

3.3.5.2: Defining active and passive residues for docking using HADDOCK

The chemical shift changes between free and IL-1 β bound gIC8 Fab, illustrated in Figures 3.13 and 3.14, allowed the definition of active and passive residues for docking using HADDOCK, as described in section 3.2.5.

The following residues were defined as active and passive:

Active

T5, R24, H30, N31, K39, N50, W92, S93, L94, Q100, V219, R245, Y246, R258, E260, G268, G269, G270, S271, T272, Y273, D276, T277, G280, I284, Y294, N314, K315, K316, L317, Y322, Q325.

Passive

Q3, S26, N28, Y32, P40, K42, Y49, K52, T53, S67, T69, Q70, E81, K103, V216, Q217, R233, A237, D242, S244, K257, K279, T283, S285, R286, D287, K290, N298.

The location of the active and passive residues on the gIC8 Fab structure is illustrated in Figure 3.15. The majority of the active residues are at the top of the structure as represented in figure 3.15 covering the CDR loops which form the expected antigen binding site. However, there are also active residues defined close to the VL-VH interface, distinct from the CDR loops, due to large chemical shift changes in these residues. It is highly unlikely that these residues in a Fab would be involved in direct interactions with the antigen, however, in protein complexes where an expected interaction site is not known it would not be possible to exclude these residues. To assess the ability of the docking protocol to cope with chemical shift changes at sites distinct from the interaction site, these residues were kept as active residues for the docking runs.



Figure 3.15: Residues defined as active and passive for gIC8 Fab during docking with IL-1 β

Surface views of the refined gIC8 Fab homology model. Residues defined as active during docking with IL-1 β are shown in red and passive residues are shown in green. The majority of the active residues are found near the CDR loops (top of the Fab as represented here) with some active residues defined in the framework of the variable domains.

Assignments for IL-1 β in both the free and bound form were already available (Wilkinson, 2009) and chemical shift changes between free IL-1 β and IL-1 β in complex with gIC8 Fab had already been determined (Hall, 2010). Active residues for IL-1 β were defined as those with a chemical shift difference >0.15 ppm and solvent accessibility >10% and passive residues as those adjacent to active residues with solvent accessibility >10%, as described previously (Hall, 2010).

The following residues were defined as active and passive:

Active

S5, L6, Q15,Q39, F46, Q48, G49, E50, S52, D54, K55, I56, K92, K93, E105, I106, N107, N108, K109, T147, M148, Q149, F150, V151, S152, S153

Passive

A1, P2, R4, R11, S13, Q14, M36, E37, V41, E51, N53, P91, K94, E96, D145

The locations of the active and passive residues are highlighted on the structure of IL- 1β shown in Figure 3.16. The active and passive residues form a continuous patch at the open end of the β -barrel/trefoil, close to the N and C-termini.



Figure 3.16: Residues defined as active and passive for IL-1 β during docking with gIC8 Fab

Surface and ribbon views of the IL-1 β crystal structure 2I1B (Priestle et al., 1989). Residues defined as active during the docking are coloured red and those defined as passive are coloured green.

3.3.5.3: Identification and assignment of intermolecular NOEs between gIC8 Fab and IL-1 β

Intermolecular backbone amide to side chain methyl (H_N -CH₃) NOEs were identified as described in section 3.2.6 using gIC8 Fab/IL-1 β complexes with asymmetric labelling where one component was uniformly ¹⁵N/²H labelled and the other unlabelled. In this way any backbone amide to side chain NOEs in an ¹⁵N-edited NOESY-TROSY experiment can only arise from intermolecular ¹H-¹H NOEs. Projections from 3D spectra of NOESY data recorded on such samples show a substantial number of intermolecular NOEs (Figure 3.17).

Multiple backbone amide groups of IL-1 β show NOEs to the same two methyl groups of gIC8 Fab (highlighted in Figure 3.17). The distinctive chemical shifts of these methyl groups allowed their identification in a CT-HSQC of residually protonated ¹³C/²H labelled gIC8 Fab in complex with unlabelled IL-1 β . From the ¹³C/¹H shifts these methyl groups were identified as either leucine or valine. Analysis of the refined homology model of gIC8 Fab showed only two leucine residues near the potential antigen binding site (L94 and L317). ¹⁵N/¹H NOESY-TROSY data recorded on residually protonated ¹⁵N/²H labelled gIC8 Fab in complex with unlabelled IL-1 β showed that the amide group of L94 showed an NOE to a methyl group that matched the chemical shift of one of the methyls in question. Both methyl groups were assigned to L94.

Four amide groups of gIC8 Fab showed NOEs to three methyl groups of IL-1 β (highlighted in Figure 3.17). One of the intermolecular NOEs could be assigned by matching to the unique ¹H chemical shift of the I56 δ -methyl. The other NOEs were too similar in ¹H methyl chemical shift to multiple IL-1 β methyl groups to assign based on chemical shift alone. The potential assignments were filtered based on proximity to the interaction site (using active residues defined in section 3.3.5), leaving L6 and L110 as the potential methyl assignments. The position of these residues relative to the residues already assigned to intermolecular NOEs left only one possible combination of assignments.

In total, 10 intermolecular H_N -CH₃ NOEs were identified and assigned between residues L94 (CDR L3), G270 (CDR H2), S271 (CDR H2) and K316 (CDR H3) of gIC8 Fab and residues L6, N7, I56, L110, V151 and S153 of IL-1 β , as observed in the

¹⁵N strips of the NOESY-TROSY spectra in Figure 3.17. When the positions of these residues are displayed on the surface of the gIC8 Fab and IL-1 β structures they show good spatial dispersion across the interface as illustrated in Figure 3.18.

In addition to the H_N-CH₃ intermolecular NOEs observed between gIC8 Fab and IL-1β there are multiple intermolecular H_{N} to side-chain proton NOEs visible in the 2D $^{1}\text{H-}^{1}\text{H}$ projections of the NOESY-TROSY data, between approximately 6 and 1 ppm in the ¹H (F1) dimension (Figure 3.17). As described in chapter 2 full side chain assignments are not available for the gIC8 Fab/IL-1β complex due to its large size (65 kDa), making it difficult to assign most of the non-methyl side-chain to backbone amide intermolecular NOEs. However, it was possible to assign two intermolecular H_N-CH₂ (serine) NOEs. Assignment of the serine HB resonances was achieved by matching intra-residue H_N-CH₂ and inter-residue (i+1) H_N -CH₂ NOEs in ${}^{15}N/{}^{1}H$ NOESY-TROSY spectra, which are should always be observed based on interatomic distances (Wüthrich et al., 1983), recorded on ${}^{15}N/{}^{2}H$ labelled samples expressed using protonated glucose as the carbon source. The approach is entirely analogous to the methyl assignment protocol described in chapter 2 and is possible because of the semi-selective protonation of the serine $H\beta$ groups when using the residual protonation labelling scheme. Assignment of the intermolecular H_N-CH₂ NOEs was achieved by matching the ¹H chemical shifts to the assigned serine H β resonances.



Figure 3.17: Identification of intermolecular H_N -CH₃ NOEs between gIC8 Fab and IL-1 β

2D 1 H/ 1 H projections of 3D NOESY-TROSY spectra recorded on uniformly labelled 15 N perdeuterated IL-1 β complexed with unlabelled gIC8 Fab (A) and unlabelled IL-1 β complexed with uniformly labelled 15 N perdeuterated gIC8 Fab (B). Backbone amide to CH_x 1 H- 1 H NOEs will have arisen from intermolecular interactions given the labelling scheme used. Several NOEs in the methyl region are visible in each spectrum (highlighted in the red boxes).



Figure 3.18: Assignment of intermolecular $H_{\text{N}}\text{-}\text{CH}_3$ NOEs between gIC8 Fab and IL-1 β

¹⁵N strips from ¹⁵N/¹H NOESY-TROSY spectra recorded on uniformly ¹⁵N/²H labelled IL-1 β in complex with unlabelled gIC8 Fab (red) and unlabelled IL-1 β in complex with uniformly ¹⁵N/²H labelled gIC8 Fab (blue). Assignments of intermolecular NOEs (top) are indicated.



Figure 3.19: Locations of residues found to show intermolecular NOEs in the gIC8 Fab/IL-1β complex.

Surface views are shown for gIC8 Fab (A) and IL-1 β (B). Residues with amide groups involved in intermolecular H_N-CH₃ NOEs are shown in blue and similarly residues with methyl groups involved in intermolecular H_N-CH₃ NOEs are shown in green.

3.3.6: Docking the gIC8 Fab/IL-1β complex

A docked structure of the gIC8 Fab/IL-1 β complex had previously been determined using only chemical shift perturbation from minimal shift to define ambiguous interaction restraints (AIRs) and backbone amide RDCs to provide orientation restraints (Hall, 2010), however, this proved only partially successful. The 132 converged structures obtained formed 5 equally populated clusters and had an average backbone RMSD to the mean structure of 2.39 ±0.70 Å. The orientation of gIC8 Fab and IL-1 β in the complex structures was uniform across the ensemble, however, there was a difference in the IL-1 β position relative to gIC8 Fab of approximately 10 Å across the 5 clusters (Hall, 2010). The docking did not result in a well defined interaction surface on gIC8 Fab, probably due to the high degree of chemical shift perturbation observed in the gIC8 Fab variable domains upon binding to IL-1 β . The large degree of variability between the resulting structures made it impossible to confidently identify which residues were essential for complex formation, limiting the applicability of the structure to knowledge based development of gIC8 Fab as a therapeutic.

The work presented in this chapter has resulted in the determination of comprehensive backbone assignments for gIC8 Fab in both the free and IL-1 β bound form. This has allowed the use of chemical shift perturbation from actual shifts, rather than minimal shift, to define the potential IL-1 β interaction site on gIC8 Fab. This has also allowed determination of a more complete set of backbone amide RDCs covering a larger proportion of the residues of gIC8 Fab. The backbone assignments of gIC8 Fab in the IL-1 β bound form were also used to identify and assign intermolecular H_N-CH₃ intermolecular NOEs using the techniques described in section 3.3.5 in conjunction with the IL-1 β methyl assignments determined in chapter 2. This data was used to generate the input restraints used to dock the gIC8 Fab/IL-1 β complex. However, it is possible that for some systems not all of this data would be available to use for docking calculations for the gIC8 Fab/IL-1 β complex were carried out. In the first, the backbone amide RDC data was used in conjunction with the AIRs as intermolecular restraints, a second run utilising only intermolecular NOEs and AIRs as intermolecular restraints

was also carried out along with a third run using all of the available data (intermolecular NOEs, RDCs and AIRs).

3.3.6.1: Docking of the gIC8 Fab/IL-1ß complex using RDCs and AIRs

A docked structure of the gIC8 Fab/IL-1 β complex was produced using HADDOCK as described in section 3.2.8, using AIRs as intermolecular restraints and backbone amide RDCs to define the relative orientations of each of the proteins. The docking calculations resulted in one major cluster of 152 structures and a minor cluster of 43 structures. A best fit superposition of the structures of the most populated, lowest scoring cluster to the backbone atoms of the closest to the mean structure showed an average RMSD of 1.17 ±0.60 Å (Figure 3.20). The individual components of the complex were well defined, with alignment of the structures on the backbone atoms of IL-1 β in the closest to the mean structure showing an average RMSD of 0.64 ±0.22 Å for the backbone atoms of IL-1 β residues. Alignment of the structures on the backbone atoms of 0.48 ±0.09 Å for the backbone atoms of gIC8 Fab in the closest to the mean structure showed an average RMSD for the alignment of the structures on the entire complex compared to the individual components indicates that there is some variability in the position of IL-1 β relative to gIC8 Fab as illustrated in Figure 3.20B and 3.20C.

Analysis of the closest to the mean structure using PALES showed a near perfect agreement with the backbone amide RDC data with a Q value of 0.10 and a correlation between observed and calculated RDCs of 0.995 for the whole complex. When analysed separately gIC8 Fab and IL-1 β showed equally good agreement with Q values of 0.11 and 0.06 and correlations of 0.993 and 0.998 respectively (Figure 3.21).



С

Figure 3.20: Ensemble of gIC8 Fab/IL-1β complex structures obtained for the docking calculations run using RDCs and AIRs as intermolecular restraints

Backbone representations of all 152 gIC8 Fab (black)/IL-1 β (blue) complex structures from the lowest scoring cluster are shown aligned on the backbone atoms of the entire complex of the closest to the mean structure (A), with an average backbone RMSD of 1.17 ±0.60

Å, aligned on the backbone atoms of the residues of IL-1 β (B), with an average backbone RMSD for IL-1 β residues of 0.64 ±0.22 Å, and aligned on the backbone atoms of the residues of gIC8 Fab (C), with an average backbone RMSD for gIC8 Fab residues of 0.48 ±0.09 Å.



Figure 3.21: Comparison of observed and calculated backbone amide RDCs for the closest to the mean gIC8 Fab/IL-1 β complex structure of the family of structures obtained from the docking calculations using RDCs and AIRs as intermolecular docking restraints.

The three graphs show the agreement between observed backbone amide RDCs and those calculated (from the closest to the mean structure of the ensemble shown in figure 3.20) using PALES when fitted to the entire complex structure (A), and when fitted seperately to the gIC8 Fab structure (B) and the IL-1 β structure (C). Q factors for each of the fits are indicated on each plot. The error bars on the red data point indicate the 1.4 Hz estimated error present for each RDC.

The assigned intermolecular H_N -CH₃ NOEs were not used as restraints for this docking run and so could be used to assess the accuracy of the docked structures. Cross validation against the assigned intermolecular NOEs showed poor agreement with the calculated structures, with just 3 of the 10 intermolecular distance restraints satisfied in greater than 15% of the 152 structures of the most populated cluster, and 3 distance restraints that are not satisfied by any of the structures.

gIC8 Fab – IL-1β	Distance range (Å)	No. of structures that would satisfy distance restraints (5.5 Å)
L94 H _N - L6 QD	6.3-13.6	2
S271 H _N - L110 MD1	3.1-11.9	120
G270 H _N - L110 MD1	3.9-10.7	118
K316 H _N - I56 MD	5.4-11.6	11
L94 MD1 - N7 H _N	4.2-10.4	22
L94 MD2 - N7 H _N	3.9-9.5	97
L94 MD1 - V151 H _N	9.0-14.9	0
L94 MD2 - V151 H _N	6.7-14.0	0
L94 MD1 - S153 H _N	7.4-14.6	0
L94 MD2 - S153 H _N	6.3-14.7	3

Table 3.1: Inter-atomic distance ranges of atoms in the docked structures calculated using RDCs

Inter-atomic distance ranges from the 152 docked structures calculated using RDCs and AIRs are shown for atoms that have been assigned to intermolecular NOEs. The number of structures that would satisfy each of the NOE restraints is indicated.

Stats	Cluster 1	Cluster 2
Number of structures	152	43
Distance constraints (Å) ambiguous	1.75 (±2.95E-2)	1.84 (3.93E-2)
Distance constraints (Å) unambiguous	N/A	NA
R.M.S Sani (Hz)	1.11 (±0.14)	1.12(±0.15)
Bond Lengths (Å)	5.02E-3(±6.44E-4)	5.02E-3(±6.56E-4)
Bond Angles (°)	0.63(±7.02E-2)	0.64(±7.84E-2)
Impropers (°)	0.70(±7.85E-2)	0.70(±9.04E-2)
RMSD to closest to mean Interface	1.13 ±0.50	0.82 ±0.19
Buried surface area $(Å^2)$	1761.77(±144.587)	1617.36 (±109.30)
Favored	82.8	82.4
Allowed	12.2	12.4
Outliers	5	5.5

 Table 3.2: Statistics for the docked structures calculated using RDCs and AIRs

3.3.6.2: Docking of the gIC8 Fab/IL-1β complex using NOEs and AIRs

A docked structure of the gIC8 Fab/IL-1 β complex was produced using HADDOCK as described in section 3.2.8, with just a combination of AIRs, intermolecular H_N-CH₃ NOEs and H_N-CH₂ (Ser) NOEs used as intermolecular restraints. The use of intermolecular NOEs in the docking protocol had a dramatic effect on the diversity of the complex structures obtained, resulting in all 200 calculated structures forming a single, tightly defined cluster (Figure 3.22). When aligned on the backbone atoms of the entire complex the average backbone RMSD to the closest to the mean structure was 0.47 ±0.07 Å. The structures of the individual components of the complex were well defined with little variation in the gIC8 Fab and IL-1 β structures across the ensemble, with average backbone RMSDs of 0.37 ±0.03 Å and 0.36 ±0.02 Å for IL-1 β and gIC8 Fab respectively, when aligned separately. This similarity in the RMSDs of the ensemble of complex structures when analysed as a whole and as individual complex components shows that there is only a small variation in the IL-1 β position relative to gIC8 Fab across the ensemble, which is visualised in figure 3.22C.

None of the intermolecular NOEs used in the docking were violated by greater than 0.5 Å in any of the calculated structures. The backbone amide RDCs were not used as restraints to generate this docked structure and so could be used to assess the accuracy of the docked structures. Cross validation of the calculated structures against the RDC data showed poor agreement. When analysed in PALES the closest to the mean structure showed a correlation between observed and calculated RDCs of 0.76 with a Q value of 0.65 (Figure 3.23). Separate analysis of the gIC8 Fab and IL-1 β showed a correlation of 0.87 and 0.85 with Q values of 0.50 and 0.52 for IL-1 β and gIC8 Fab respectively, slightly worse than the input structures for the docking.



Figure 3.22: Ensemble of gIC8 Fab/IL-1β complex structures obtained for the docking calculations run using NOEs and AIRs as intermolecular restraints

Backbone representations of all 200 gIC8 Fab (black)/IL-1 β (blue) complex structures are shown aligned on the backbone atoms of the entire complex of the closest to the mean structure (A), with an average backbone RMSD of 0.47 ±0.07 С

Å, aligned on the backbone atoms of the residues of IL-1 β (B), with an average backbone RMSD for IL-1 β residues of 0.37 ±0.03 Å, and aligned on the backbone atoms of the residues of gIC8 Fab (C), with an average backbone RMSD for gIC8 Fab residues of 0.36 ±0.02 Å.

В



Figure 3.23: Comparison of observed and calculated backbone amide RDCs for the closest to the mean gIC8 Fab/IL-1 β complex structure of the family of structures obtained from the docking calculations using NOEs and AIRs as intermolecular docking restraints.

The three graphs show the agreement between observed backbone amide RDCs and those calculated (from the closest to the mean structure of the ensemble shown in figure 3.22) using PALES when fitted to the entire complex structure (A), and when fitted seperately to the gIC8 Fab structure (B) and the IL-1 β structure (C). Q factors for each of the fits are indicated on each plot. The error bars on the red data point indicate the 1.4 Hz estimated error present for each RDC.

Stats	Cluster 1
Number of structures	200
Distance constraints (Å) ambiguous	0.25 (±0.77)
Distance constraints (Å) unambiguous	1.14E-2(±1.96E-3)
R.M.S Sani (Hz)	N/A
Bond Lengths (Å)	3.09E-3(±2.88E-5)
Bond Angles (°)	0.44(±4.28E-3)
Impropers (°)	0.41(±6.05E-3)
RMSD Interface	0.54 ±0.07
Buried surface area $(Å^2)$	1954.34±82.84
Favored	89.3
Allowed	9.6
Outliers	1.1

 Table 3.3: Statistics for the docked structures calculated using NOEs and AIRs

3.3.6.3: Docking of the gIC8 Fab/IL-1β complex using NOEs, RDCs and AIRs

A final docked structure of the gIC8 Fab/IL-1 β complex was produced using HADDOCK, as described in section 3.2.8, but making use of all the available NMR data (AIRs, NOEs and RDCs). This gave a result that, upon initial inspection, was similar to the docked structure generated using just using RDCs and AIRs. Alignment of all 200 calculated structures gave a backbone RMSD to the closest to the mean structure of 1.40 ±0.68 Å. However, a large number of the structures (87 out of the 200 calculated) were found to have violations of greater than 0.5 Å for the intermolecular NOE restraints. Removal of the structures with NOE restraint violations greater than 0.5 Å and re-clustering of the remaining 113 structures resulted in a major cluster of 103 structures and a minor cluster of 10 structures. Analysis of the largest cluster showed a backbone RMSD to the closest to the mean structure of 0.68 ±0.21 Å (Figure 3.24). The structures of the individual components of the complex were well defined with little variation in the gIC8 Fab and IL-1 β structures across the ensemble, with average backbone RMSDs of 0.55 ±0.21 Å and 0.45 ±0.21 Å for IL-1 β and gIC8 Fab respectively, when aligned separately.

When fitted to the RDC data using PALES the closest to the mean structure showed a Q value of 0.12 and a correlation of 0.993 (Figure 3.25). Separate analysis of IL-1 β and gIC8 Fab showed Q values of 0.064 and 0.158 and correlations of 0.998 and 0.987 respectively. The final cluster of 103 structures therefore fit all available experimental data within experimental error.



В

С



Figure 3.24: Ensemble of gIC8 Fab/IL-1β complex structures obtained for the docking calculations run using NOEs, RDCs and AIRs as intermolecular restraints

Backbone representations of the 103 gIC8 Fab (black)/IL-1 β (blue) converged complex structures from the lowest scoring cluster are shown aligned on the backbone atoms of the entire complex of the closest to the mean structure (A), with an average backbone



RMSD of 0.68 \pm 0.21 Å, aligned on the backbone atoms of the residues of IL-1 β (B), with an average backbone RMSD for IL-1 β residues of 0.55 \pm 0.21 Å, and aligned on the backbone atoms of the residues of gIC8 Fab (C), with an average backbone RMSD for gIC8 Fab residues of 0.45 \pm 0.10 Å. The variable heavy/light and constant heavy/light domains of the Fab are labelled.



Figure 3.25: Comparison of observed and calculated backbone amide RDCs for the closest to the mean gIC8 Fab/IL-1β complex structure of the family of structures obtained from the docking calculations using NOEs, RDCs and AIRs as intermolecular docking restraints.

The three graphs show the agreement between observed backbone amide RDCs and those calculated (from the closest to the mean structure of the ensemble shown in figure 3.24) using PALES when fitted to the entire complex structure (A), and when fitted seperately to the gIC8 Fab structure (B) and the IL-1 β structure (C). Q factors for each of the fits are indicated on each plot. The error bars on the red data point indicate the 1.4 Hz estimated error present for each RDC.

Stats	Cluster 1	Cluster 1 (no viol)	Cluster 2 (no viol)	
Number of structures	200	103	10	
Distance constraints (Å) ambig	0.34 (±0.17)	0.21 (±6.24E-2)	0.27 (±0.13)	
Distance constraints (Å) unambig	0.10(±0.11)	1.93E-2(±5.68E- 3)	2.40E-2(9.27E- 3)	
R.M.S Sani (Hz)	1.5 (±1.99)	1.19 (±0.16)	1.35 (±0.14)	
Bond Lengths (Å)	6.19E-3(±1.33E- 3)	5.92E-3 (±5.09E- 4)	6.68E- 3(±7.79E-4)	
Bond Angles (°)	0.87(±1.4)	0.83 (±5.23E-2)	0.89 (±7.47E-2)	
Impropers (°)	1.50(±0.43)	1.43 (±0.13)	1.40 (±0.53)	
RMSD Interface	1.72±0.89	0.71 ±0.21	1.19 ±0.76	
Buried surface area ($Å^2$)	1619.36±146.13	1701.21±117.85	1616.16 (±139/08)	
Favored	80.2	81.5	79	
Allowed	12.9	12.2	13.5	
Outliers	6.9	6.3	7.6	

Table 3.4: Statistics for the docked structures calculated using NOEs, RDCs andAIRs

3.3.7: Comparison of docked structures

Comparison of the resulting ensembles of structures from the various HADDOCK runs in section 3.3.6 shows some difference in the diversity of the structures among each of the ensembles. This is reflected in the difference of the average backbone RMSDs to the closest to the mean structures from each ensemble (0.47-1.17 Å).

Comparison of the structures with the lowest HADDOCK scores from each of the HADDOCK runs presented reveals that the resulting structures are fairly similar (Figure 3.26). The largest difference is seen between the structure calculated using RDCs and AIRs and the structure calculated using NOEs and AIRs, with an RMSD of 3.37 Å, when aligned on the backbone atoms of the entire complex. The smallest difference is seen between the structure calculated using NOEs and AIRs and the structure calculated using NOEs and AIRs and the structure calculated using NOEs and AIRs and the structure calculated using NOEs and AIRs with an RMSD of 1.65 Å. Comparison of the structure calculated using RDCs and AIRs and the structure calculated using NOEs, RDCs and AIRs with an RMSD of 2.01 Å.

Separate comparisons of the gIC8 Fab and IL-1 β structures showed relatively low RMSDs ranging from 0.50 Å to 0.56 Å and 0.60 to 0.65 Å respectively, indicating that the larger RMSDs for the complex structures were arising from a difference in the position of the IL-1 β relative to gIC8 Fab. This is best observed when the lowest scoring structures from each of the HADDOCK runs is aligned on the backbone atoms of the gIC8 Fab structure as illustrated in Figure 3.26B. When the structures are aligned in this way the RMSDs between the IL-1 β backbone atoms show large differences between the structures with RMSDs ranging from 3.84 Å to 8.43 Å which are summarised in Table 3.5.



Figure 3.26: Comparison of the gIC8 Fab/IL-1β structures obtained from docking calculations using different intermolecular restraints

Backbone bond representations of the lowest scoring structures from the clusters obtained from docking calculations run using; RDCs and AIRs (orange), NOEs and AIRs (blue) and RDCs, NOEs and AIRs (green) as intermolecular restraints. Structures on the left (A) are aligned on the backbone atoms of the entire complex and structures on the right (B) are aligned on the backbone atom of gIC8 Fab. RMSDs between each of the structures are summarised in Table 3.5.

		Whole Complex	gIC8 Fab	IL-1β	IL-1β when aligned on gIC8 Fab	Interface
AIRs & RDCs	AIRs & NOEs	3.37 Å	0.56 Å	0.64 Å	8.43 Å	3.59 Å
AIRs & RDCs	AIRs, RDCs & NOEs	2.01 Å	0.53 Å	0.65 Å	5.22 Å	2.32 Å
AIRs & NOEs	AIRs, RDCs & NOEs	1.65 Å	0.50 Å	0.60 Å	3.84 Å	1.67 Å

 Table 3.5: Comparison of the lowest scoring structures from the three docking runs

RMSDs are shown between the docked structures calculated using the restraints indicated in the left hand column. The RMSDs are shown for; the structures when aligned on the backbone atoms of the entire complex, the individual components of the complex when aligned separately, for the backbone atoms of IL-1 β when the complexes are aligned on the backbone atoms of gIC8 Fab and for the interface residues when aligned on the backbone atoms of the residues at the interface.

Analysis of intermolecular van der Waals interactions show a similar pattern of contact residues across the lowest scoring structures from each of the HADDOCK runs which are summarised if Figure 3.27. The majority of contacts are between CDR 2 and 3 of the heavy chain of gIC8 Fab and the β 4- β 5 and β 8- β 9 loops and the C-terminal region of IL-1 β . The N-terminal region of IL-1 β makes significant contacts with CDR 3 of the light chain of gIC8 Fab in the docked structure from the run using RDCs and AIRs, but these contacts are not identified in the structure from the run using NOEs and AIRs. In total, 29 contact residues were identified in the complex structure obtained using RDCs and AIRs, which is greater than the 21 identified in the structure obtained using RDCs, NOEs and AIRs. However, the contacts identified in the structure obtained using all of the available data (RDCs, NOEs and AIRs) reflect the networks present in both of the other structures.



Figure 3.27: Comparison of the gIC8 Fab/IL-1β contact residues from the lowest scoring structures obtained from docking calculations using different intermolecular restraints

Intermolecular contact maps are shown for the lowest scoring structures from the ensembles calculated using AIRs and NOEs (A), AIRs and RDCs (B), and AIRs, RDCs and NOEs (C). The locations of the CDR residues are indicated with the blue bars under each of the sequences.

The structures generated from the various docking runs presented show that the use of RDCs and AIRs gives the most diversly populated cluster of structures, which could give a lower confidence in the accuracy of the derived structures, whereas the use of NOEs and AIRs gives a very tight family of highly similar structures. In the absence of a full solution structure of the complex it is difficult to assess which of the intermolecular restraints has the greatest impact on the accuracy of the derived structures. For reasons described in detail in chapter 2 it is not possible to solve the solution structure of a complex of this size. However, crystallographic studies of the complex would reveal details of the intermolecular interactions for comparison to the docked structures.
3.3.8: Crystallisation of gIC8 Fab in complex with IL-1β

Initial crystallisation screens of approximately 500 conditions gave several hits, the most promising being 20% PEG 2000 MME, 0.2 M Trimethylamine N-oxide, 0.1 M Tris-HCl pH 8.5. Optimisation of the crystallisation conditions was achieved by varying the PEG concentration between 16% and 24%, and pH between 8.0 and 9.0. Large crystals grew as clusters of thin plates (Figure 3.28) in multiple conditions ranging between 16 and 24% PEG 2000 MME, 0.2 M Trimethylamine N-oxide, 0.1 M Tris-HCl pH 8.5-9.0. Crystals diffracted to 2.45 Å and were of sufficient quality to solve the structure of the gIC8 Fab/IL-1 β complex. Acquisition of diffraction data, processing and structure refinement were carried out idependently by Gareth Hall (University of Leicester). The X-ray diffraction and structure refinement statistics are summarised in Table 3.6. The resulting structure was used to assess the accuracy of the docked HADDOCK structures presented in this chapter.



Figure 3.28: Crystals of the gIC8 Fab/IL-1β complex

Example of the crystals that grew in optimised crystallisation conditions.

Resolution range (Ã…)	44.69 - 2.45 (2.538 - 2.45)
Space group	P 1 21 1
Unit cell	51.896 69.01 171.312 90 90.91 90
Total reflections	141931 (12462)
Unique reflections	44097 (4376)
Multiplicity	3.2 (2.8)
Completeness (%)	98.37 (98.36)
Mean I/sigma(I)	7.78 (1.53)
Wilson B-factor	38.51
R-merge	0.1031 (0.5068)
R-meas	0.1238
CC1/2	0.993 (0.82)
CC*	0.998 (0.949)
Reflections used for R-free	
R-work	0.2051 (0.3193)
R-free	0.2547 (0.3518)
CC(work)	
CC(free)	
Number of non-hydrogen atoms	8954
macromolecules	8903
ligands	
water	51
Protein residues	1151
RMS(bonds)	0.003
RMS(angles)	0.84
Ramachandran favored (%)	97
Ramachandran allowed (%)	
Ramachandran outliers (%)	0.27
Clashscore	5.79
Average B-factor	43.50
macromolecules	43.60
ligands	
solvent	33.40

Table 3.6: X-ray diffraction analysis statistics for the IL-1 β /gIC8 Fab complex

3.3.9: Comparison of the crystal and docked structures

The general appearance of the crystal structure of the gIC8 Fab/IL-1 β complex is similar to the docked structures obtained in section 3.3.6. IL-1 β sits off-centre relative to the variable domains forming a slightly larger interaction surface with the variable domain of the heavy chain. Residues in the light chain (CDR 1 and2) of gIC8 Fab contact IL-1 β at the N and C termini whilst the heavy chain (CDR 1, 2 and 3) makes contacts with the β 4- β 5 and β 8- β 9 loops. Residues in CDR2 of the heavy chain form a parallel β -sheet with residues in β 12 at the C-terminus of IL-1 β .

Alignment of the crystal structure and the lowest scoring structure from docking using RDCs and AIRs gave a backbone RMSD of 2.60 Å (Figure 3.29A). When the structures are aligned on the backbone atoms of gIC8 Fab a large displacement of the IL-1 β structures is observed (Figure 3.29C) with an RMSD of 6.43 Å for the backbone atoms of IL-1 β . This displacement of the backbone atoms reduces to 4.86 Å when only considering the residues of IL-1 β that form the interface with gIC8 Fab (summarised in Table 3.7).

Alignment of the crystal structure and the lowest scoring structure from docking using NOEs and AIRs gave a backbone RMSD of 2.06 Å (Figure 3.30A). When the structures are aligned on the backbone atoms of gIC8 Fab a smaller displacement of the IL-1 β structures is observed (Figure 3.30C) with an RMSD of 4.71 Å for the backbone atoms of IL-1 β . This displacement of the backbone atoms reduces to 3.84 Å when only considering the residues of IL-1 β that form the interface with gIC8 Fab.





Figure 3.29: Comparison of the crystal structure of the gIC8 Fab/IL-1 β complex with the structure docked using RDCs and AIRs

Backbone representations of the gIC8 Fab/IL-1 β complex crystal structure (black) and the lowest scoring docked structure are shown aligned on the backbone atoms of entire complex (A), with a backbone RMSD of 2.60 Å, aligned on the backbone atoms of the residues of IL-1 β (B), with a backbone RMSD for IL-1 β residues of 1.10 Å, and aligned on the



backbone atoms of the residues of gIC8 Fab (C), with a backbone RMSD for gIC8 Fab residues of 1.33 Å. When aligned on the backbone atoms of gIC8 Fab (as in C) the RMSD for the backbone atoms of the IL-1 β residues is 6.43 Å.





Figure 3.30: Comparison of the crystal structure of the gIC8 Fab/IL-1β complex with the structure docked using NOEs and AIRs

Backbone representations of the gIC8 Fab/IL-1 β complex crystal structure (black) and the lowest scoring docked structure are shown aligned on the backbone atoms of entire complex (A), with backbone RMSD of 2.06 Å, aligned on the backbone atoms of the residues of IL-1 β (B), with a backbone RMSD for IL-1 β residues of 1.12 Å, and aligned on the backbone atoms of the backbone atoms of the residues of gIC8 Fab (C), with a backbone RMSD for gIC8 Fab residues of 1.24 Å. When



aligned on the backbone atoms of gIC8 Fab (as in C) the RMSD for the backbone atoms of the IL-1 β residues is 4.71 Å.

Superposition of the crystal structure and the lowest scoring structure from docking using RDCs, NOEs and AIRs gave a backbone RMSD of 1.81 Å (Figure 3.31A). When the structures are aligned on the backbone atoms of gIC8 Fab the smallest displacement of the IL-1 β structures is observed (Figure 3.31C) with an RMSD of 4.66 Å for the backbone atoms of IL-1 β . This displacement of the backbone atoms reduces to 3.50 Å when only considering the residues of IL-1 β that form the interface with gIC8 Fab.

		Whole Complex	gIC8 Fab	IL-1β	IL-1β when aligned on gIC8 Fab	Interface
X-ray	RDCs and AIRs	2.60 Å	1.33 Å	1.10 Å	6.43 Å	2.62 Å
X-ray	NOEs and AIRs	2.06 Å	1.24 Å	1.12 Å	4.71 Å	1.84Å
X-ray	RDCs, NOEs and AIRs	1.81 Å	1.28 Å	1.02 Å	4.66 Å	1.96 Å

Table 3.7: Comparison of the lowest scoring structures from the three docking runs to the crystal structure of the gIC8 Fab/IL-1β complex

RMSDs are shown between the docked structures calculated using the restraints indicated in the left hand column and the crystal structure. The RMSDs are shown for; the structures when aligned on the backbone atoms of the entire complex, the individual components of the complex when aligned separately and for the backbone atoms of IL- 1β when the complexes are aligned on the backbone atoms of gIC8 Fab.





Figure 3.31: Comparison of the crystal structure of the gIC8 Fab/IL-1β complex with the structure docked using NOEs, RDCs and AIRs

Backbone representations of the gIC8 Fab/IL-1 β complex crystal structure (black) and the lowest scoring docked structure are shown aligned on the backbone atoms of entire complex (A), with

backbone RMSD of 1.81 Å, aligned on the backbone atoms of the residues of IL-1 β (B), with a backbone RMSD for IL-1 β residues of 1.02 Å, and aligned on the backbone atoms of the residues of gIC8 Fab (C), with a backbone RMSD for gIC8 Fab residues of 1.28 Å. When aligned on the backbone atoms of gIC8 Fab (as in C) the RMSD for the backbone atoms of the IL-1 β residues is 4.66 Å.

С

a

The docked structure calculated using AIRs, NOEs and RDCs showed the most structural similarity to the crystal structure. A detailed analysis of the intermolecular interactions is illustrated in Figure 3.32. Of the 17 intermolecular van der Waals contacts identified in the crystal structure, 8 were correctly identified in the docked structure. A further 4 contacts were identified in the docked structure that did not match the crystal structure, however these typically followed a pattern similar to the interactions in the crystal structure, but were displaced by 1 or 2 residues. Interactions between residues at the N-terminus of IL-1 β and CDR 1 of the light chain of gIC8 Fab that were found in the crystal structure were not present in the docked structure as well as an interaction between the C-terminus of IL-1 β and CDR3 of the heavy chain of gIC8 Fab.

The positions of the contact residues of gIC8 Fab (identified in Figure 3.32) relative to IL-1 β are illustrated in Figure 3.33. With the exception of some of the side chain orientations, and the position of K316 and L317 from CDR3 of the heavy chain, the position of these residues is remarkably similar between the docked structure and the crystal structure.



Figure 3.32: Comparison of the gIC8 Fab/IL-1β contact residues from the lowest scoring structure obtained from docking calculations using AIRs, RDCs and NOEs and the crystal structure

Intermolecular contact maps are shown for the lowest scoring structure from the ensemble calculated using AIRs, RDCs and NOEs and the crystal structure. Black lines show interactions present in the docked structure, blue lines show interactions in the crystal structure and green lines show interactions that are present in both structures. The locations of the CDR residues are indicated by the blue bars under each of the sequences of the Fab variable domains.



Figure 3.33: Comparison of the crystal structure of the gIC8 Fab/IL-1β complex with the structure docked using NOEs, RDCs and AIRs

Surface views of IL-1 β are shown with the contact residues of gIC8 Fab represented as sticks, with the lowest scoring structure from the ensemble of structures generated using AIRs, RDCs and NOEs shown in panel A, the crystal structure in B and a larger view of the crystal structure superimposed with the contact residues of gIC8 Fab (from A) shown in panel C. The relative positions of the residues are fairly similar, aside from some side-chain positions (Y273 for example), with the exception of K316 and L317 from CDRH3 which are highlighted with the arrows.

3.3.10: Comparison of the crystal structure and input structures for docking

Comparison of the IL-1 β structure from the complex crystal structure and the input crystal structure for docking showed a backbone RMSD of 0.87 Å for all residues and a backbone RMSD of 0.9 Å when considering only interface residues. This indicates that there is very little change in the IL-1 β structure between free IL-1 β and IL-1 β in complex with gIC8 Fab. Comparison of the gIC8 Fab structure from the complex crystal structure with the input structure for docking showed a backbone RMSD of 1.36 Å for all residues and 1.36 Å for interface residues. The larger RMSD for gIC8 Fab compared to IL-1 β is expected due to the input structure for gIC8 Fab being a homology model.





Figure 3.34: Comparison of the backbone topology of the gIC8 Fab homology model before and after refinement against backbone amide RDCs and TALOS restraints.

The backbone ribbon representation of the homology model for gIC8 Fab is shown (A) next to the model that has been energy minimised against the experimental data (B). An overlay of the structures (C) shows that they are very similar with a C α RMSD of 1.1 Å for all residues. The variable and constant domains for the heavy and light chains are labelled on the homology model (A).



3.3.11: Comparison of the crystal structure with input restraints for docking

The crystal structure of gIC8 Fab in complex with IL-1 β was compared to the RDC data for the complex using PALES. When fitted to the entire complex a Q value of 0.59 was calculated with a correlation of 0.81 between observed and calculated RDCs. Fitting the individual components of the complex showed Q values of 0.44 and 0.36 and correlations of 0.90 and 0.93 for gIC8 Fab and IL-1 β respectively. Plots of observed and calculated RDCs are shown in Figure 3.35. Given the resolution of the crystal structure (2.45 Å) the agreement between the observed and calculated RDCs is in line with what it expected (Zweckstetter and Bax, 2000).

The crystal structure of the gIC8 Fab/IL-1 β complex was also compared to the 10 intermolecular H_N-CH₃ distance restraints used to dock the complex. The distances (shown in Table 3.8) range from 3.92-7.13 Å and are therefore consistent with the 6.5 Å restraint, after application of a pseudoatom correction factor (~1.5 Å).



Figure 3.35: Comparison of observed and calculated backbone amide RDCs for the crystal structure of the gIC8 Fab/IL-1β complex.

The three graphs show the agreement between observed backbone amide RDCs and those calculated (from the crystal structure) using PALES when fitted to the entire complex structure (A), and when fitted seperately to the gIC8 Fab structure (B) and the IL-1 β structure (C). Q factors for each of the fits are indicated on each plot. The error bars on the red data point indicate the 1.4 Hz estimated error present for each RDC.

	Distance (Å)	Distance (Å) (Pseudoatom for	
gIC8 Fab – IL-Iβ	(Pseudoatom for two methyls)	each methyl)	
L94 H _N - L6 QD	4.71	6.05/3.52	
S271 H _N - L110 MD1	5.1	4.34/6.14	
G270 H _N - L110 MD1	5.61	4.64/6.79	
K316 H _N - I56 MD	-	3.92	
L94 MD1 - N7 H _N	-	4.61	
L94 MD2 - N7 H _N	-	6.97	
L94 MD1 - V151 H _N	-	4.15	
L94 MD2 - V151 H _N	-	7.13	
L94 MD1 - S153 H _N	-	5.57	
L94 MD2 - S153 H _N	_	6.42	

Table 3.8: Inter-atomic distances of atoms in the crystal structure of the IL-1β gIC8 Fab complex

Inter-atomic distances are shown for atoms that have been assigned to intermolecular NOEs. Distances were measured to pseudoatoms for the methyl groups. For residues where methyl resonances are degenerate, or stereospecific assignment of the methyl groups are not available, distances are given for each methyl group and the pseudoatom representing both methyl groups.

3.3.12: Testing the robustness of docking using intermolecular NOEs

Previous docking runs of the gIC8 Fab IL-1 β complex using unambiguous interatomic distance restraints have relied on 12 intermolecular NOEs, 4 involving methyl groups of IL-1 β , 6 involving methyl groups of gIC8 Fab and 2 involving serine H β groups (one from IL-1 β and one from gIC8 Fab). Full assignments of the IL-1 β methyls when in complex with gIC8 Fab were completed using the method developed in chapter 2, allowing unambiguous identification of the intermolecular NOEs involving these methyl groups. However, unambiguous assignments for the methyl groups of gIC8 Fab were not available and intermolecular NOEs were assigned based on the presence of a single leucine residue at the interface identified using chemical shift mapping. Unambiguous assignment of the serine H β groups was possible using the method described in section 2.3.4.

To test the robustness of the intermolecular H_N -CH₃ NOEs for use in docking, a further docking run was carried out with the same parameters as the run in section 3.3.6.2, however, intermolecular NOEs were restricted to those involving only the unambiguously assigned methyl groups of IL-1 β . The resulting 200 structures formed a single cluster. Two of the calculated structures showed NOE restraint violations of greater than 0.5 Å and were removed. Alignment of the remaining 198 structures gave an average backbone RMSD of 0.63 ±0.15 Å to the closest to the mean structure. When aligned separately IL-1 β and gIC8 Fab showed average backbone RMSDs to the closest to the mean structure of 0.39 ±0.04 Å and 0.38 ±0.02 Å respectively.

The resulting structures (calculated using the restricted set of distance restraints) were compared to the original set of 10 intermolecular H_N -CH₃ distance restraints used in previous docking runs and the distance ranges are shown in Table 3.9. Despite not being used as restraints during the docking, the 6 excluded distance restraints are satisfied in more than 50% of the calculated structures, showing that the 4 unambiguously assigned NOEs are sufficient to dock the complex accurately.





Figure 3.36: Ensemble of gIC8 Fab/IL-1β complex structures obtained for the docking calculations run using a restricted set of NOEs as intermolecular restraints

Backbone representations of the 198 gIC8 Fab (black)/IL-1 β (blue) complex structures are shown aligned on the backbone atoms of the entire complex of the closest to the mean structure (A), with an average backbone RMSD of 0.63 ±0.15 Å, aligned on the backbone atoms of the residues of IL-1 β (B), with an average backbone RMSD for IL-1 β residues of 0.39 ±0.04 Å, and aligned on the backbone atoms of



the residues of gIC8 Fab (C), with an average backbone RMSD for gIC8 Fab residues of 0.38 ± 0.02 Å. The variable heavy/light and constant heavy/light domains of the Fab are labelled.

gIC8 Fab – IL-1β	Distance range (Å)	No. of structures that would satisfy distance restraints (5.5 Å)
L94 H _N - L6 QD	3.32-6.11	198
S271 H_N - L110 MD1	3.89-6.45	198
G270 H _N - L110 MD1	5.10-6.57	198
K316 H _N - I56 MD	5.08-6.75	198
L94 MD1 - N7 H _N	3.77-7.88	193
L94 MD2 - N7 H _N	3.51-9.00	134
L94 MD1 - V151 H _N	4.36-9.98	117
L94 MD2 - V151 H _N	3.85-8.70	128
L94 MD1 - S153 H _N	2.59-9.05	171
L94 MD2 - S153 H _N	2.44-9.65	170

Table 3.9: Inter-atomic distance ranges of atoms in the docked structures calculated using a restricted set of NOEs

Inter-atomic distance ranges from the 198 docked structures calculated using only NOEs involving unambiguously assigned IL-1 β methyl groups are shown for atoms that have been assigned to intermolecular NOEs. The number of structures that would satisfy each of the NOE restraints is indicated. Interatomic distance ranges corresponding to the intermolecular distance restraints used are shown in red.

However, when compared to the crystal structure the lowest scoring docked structure from this run shows RMSDs of 2.96 Å, 1.33 Å, 0.92 Å and 3.25 Å when considering the backbone atoms of the whole complex, gIC8 Fab residues, IL-1 β residues and interface residues respectively. This compares to RMSDs of 2.06 Å, 1.24 Å, 1.12 Å and 1.84 Å (Table 3.7) for the same comparisons of the equivalent structure docked using the full set of intermolecular distance restraints.

3.3.13: Investigating the source of differences in the crystal and docked structures of the gIC8 Fab/IL-1β complex

Previous sections showed the effect of different restraints used in docking the gIC8 Fab IL-1 β complex and how the resulting docked structures differed from a crystal structure of the complex. These differences could be a result of there being insufficient data to dock the complex or due to structural differences between the proteins in the complex and the input structures used for the docking. To establish the source of the differences in the crystal and docked structures a final docking run was performed with the same parameters as in section 3.3.6.2, however, the input structures were replaced with the IL-1 β and gIC8 Fab coordinates taken from the crystal structure of the complex.

The resulting 200 calculated structures formed a well defined single cluster with an average backbone RMSD of 0.42 ± 0.03 Å when fitted to the closest to the mean structure (Figure 3.37). When aligned separately IL-1 β and gIC8 Fab showed average backbone RMSDs to the closest to the mean structure of 0.39 ± 0.03 Å and 0.40 ± 0.03 Å respectively.

Comparison of the lowest scoring docked structure from this ensemble with the crystal structure shows a backbone RMSD of 1.73 Å. When aligning the individual components of the complex separately the RMSD is 1.96 Å and 0.48 Å for gIC8 Fab and IL-1 β respectively. When considering only the interface residues the RMSD between the docked structure and the crystal structure is 0.62 Å. The most significant difference between this docking run and the equivalent docking run using the gIC8 Fab homology model and the free IL-1 β crystal structure as the input structures is the reduction in the interface RMSD between the docked structure in section 3.3.6.2 to 0.62 Å in this docked structure.





Figure 3.37: Ensemble of gIC8 Fab/IL-1β complex structures obtained for the docking calculations run using crystal structure coordinates as input structures

Backbone representations of the 200 gIC8 Fab (black)/IL-1 β (blue) complex structures are shown aligned on the backbone atoms of the entire complex of the closest to the mean structure (A), with an average backbone RMSD of 0.42 ±0.03 Å, aligned on the backbone atoms of the residues of IL-1 β (B), with an average backbone RMSD for IL-1 β

residues of 0.39 \pm 0.03 Å, and aligned on the backbone atoms of the residues of gIC8 Fab (C), with an average backbone RMSD for gIC8 Fab residues of 0.40 \pm 0.03 Å. The variable heavy/light and constant heavy/light domains of the Fab are labelled.

3.4: Discussion

Expression and purification of ${}^{13}C/{}^{15}N/{}^{2}H$ labelled gIC8 Fab was successfully accomplished using previously optimised protocols, however, completion of the backbone assignments to a high level required denaturation of gIC8 Fab to exchange deuterated backbone amide groups from the core of the protein with the solvent. Refolding proved simple and successful with a high yield of correctly folded protein. The correct refolding was evident from the near perfect overlay of the signals visible in TROSY spectra of the non-refolded and refolded protein. Whilst refolding did result in a number of extra peaks appearing in the TROSY spectrum showing that the deuterated amide groups in the core of the protein had exchanged with the solvent, this did not result in a dramatic improvement in the overall assignment level of gIC8 Fab, although a number of dubious assignments from very weak peaks in the non-refolded data could be confidently confirmed. The majority of the remaining unassigned residues were found in three stretches of 5-10 residues in the constant domain of the heavy chain, which could be attributed to conformational exchange in these regions. This is supported by the crystal structure of the gIC8 Fab/IL-1 β complex where no electron density is observed for residues 346-355 and 433-443. Assignments were also missing for CDR3 of both the light and the heavy chain and some residues in CDR2 of the heavy chain, suggesting conformational heterogeneity at the antigen binding site as reported previously (Addis et al., 2014). This has also been observed in an antibody/IL-6 interaction (Addis et al., 2014). Completing backbone assignments for gIC8 Fab in complex with IL-1 β was assisted by comparison with the assignments of the corresponding scFv in complex with IL-1 β . The resulting backbone assignments for gIC8 Fab when in complex with IL-1 β covering 85% of the protein gave excellent coverage of the antigen binding site and so were sufficient to use when generating restraints for the docking.

It was possible using the approach described to assign 12 intermolecular H_N -CH₃/H_N-CH₂ NOEs for use in the docking calculations. This is a very small number compared to the several thousand NOEs used in traditional NMR structure determination, however, it has been shown that a small number of unambiguous intermolecular distance restraints can be sufficient to dock a protein-protein complex. The solution structure of the EIN-HPr complex (40 kDa) was solved using a full complement of

NMR data including 110 intermolecular NOEs, however, this required approximately 3500 hours of spectrometer time and over 2 years to analyse the data (Garrett et al., 1999). It was subsequently shown that docking of the crystal structures of the free proteins using as few as 10 intermolecular H_N-CH₃ NOEs, and backbone amide RDCs, could produce a structure that was very similar to the full NMR structure with an RMSD of 1.3 Å (Clore, 2000). More recently it has been shown that the use of very sparse NOE data (3 intermolecular CH₃-CH_x NOEs) could be used to dock a structure that was only 2 Å away from the original NMR structure (Tang and Clore, 2006). The use of asymmetrically labelled samples and conventional ¹⁵N/¹H NOESY-TROSY experiments to identify intermolecular H_N-CH₃ NOEs offers superior sensitivity to any combination of ${}^{15}N/{}^{13}C$ labelling with filtered/edited experiments (Otting et al., 1986). Sensitivity could have been further improved by replacing the unlabelled half of the complex with selectively labelled protein with ${}^{12}C/{}^{1}H$ methyl groups on an otherwise $^{12}C/^{2}H$ background. This would result in slower relaxation of the methyl signals, and with the use of long NOE mixing times could result in the identification of more NOEs over larger inter-atomic distances. No intermolecular CH₃-CH₃ NOEs were observed for the complex, despite a leucine and a value residue from IL-1 β being within range of two Leucine residues of gIC8 Fab in the crystal structure to show NOEs. This is potentially due to the labelling scheme used. ¹³C/¹H NOESY HSQC data was recorded on complex samples where IL-1 β was ¹³C/²H methyl protonated and gIC8 Fab was unlabelled. This is similar to other asymmetric labelling schemes (Gross et al., 2003), however, the methyl labelling of IL-1 β was only semi-selective. It is possible that the addition of the ¹³C label coupled with the increased relaxation of the gIC8 Fab methyls in the unlabelled sample have made it difficult to detect these NOEs in a complex of this size. It may have been possible to detect these NOEs using a sample where IL-1 β methyls were selectively ${}^{13}C/{}^{1}H$ labelled on an otherwise ${}^{12}C/{}^{2}H$ background and gIC8 Fab methyls were selectively ¹²C/¹H labelled on a ¹²C/²H background. However, this would still result in a small number of NOEs that would not offer very good coverage of the interaction site, and so would probably not offer any improvement to the docking results already obtained using the H_N-CH₃ NOEs.

Docking was initially run using the previously determined RDC data accompanied with newly determined AIRs, TALOS and intra-molecular NOE restraints so as to give an accurate comparison to what happens when intermolecular NOEs are included. The lowest scoring and most populated cluster of structures showed an average RMSD of 1.2 Å to the closest to the mean structure. This is significantly better than the 2.4 Å RMSD cluster that was observed during previous attempts to dock the complex. It is unlikely that the additional input data for this run of HADDOCK would have such a drastic effect, the difference is more likely related to the definition of semi-flexible residues in gIC8 Fab in the original HADDOCK runs, where a substantial proportion of the residues were defined as semi-flexible to allow for potential domain/domain reorientation during the docking. The input model for gIC8 Fab in the docking runs described here has already been minimised against the experimental data and so semiflexible residues were defined automatically by HADDOCK. This resulted in a tighter cluster of structures but was still not sufficient to define a discrete protein-protein interface. This was not improved when considering only the best structures from the lowest scoring cluster as even the 20 structures with the lowest HADDOCK score displayed the full spread of the entire cluster. The intermolecular NOEs were not used as restraints in this docking run and so could be used to validate the resulting structures. Only 3 of the 10 assigned NOEs were satisfied in a significant number of the calculated structures, showing that the RDCs are not sufficient to dock the complex with complete accuracy.

Docking using intermolecular NOEs resulted in all 200 calculated structures forming a single tightly defined cluster, leaving little uncertainty in the resulting structure. Whilst 10 NOEs is a relatively low number for NMR structure determination, the lack of ambiguity compared with other potential restraints that are frequently used for docking, combined with excellent coverage of the interaction surface, leaves a single relative orientation of the two proteins that is able to satisfy the restraints. The RDC data had not been used as a restraint in this docking run so could be used to validate the resulting structures. The decrease in the agreement between the calculated structures and the RDC data for the individual components of the complex compared to the input structures would be expected as the structures have changed to fit the restraints used in the docking without the constraints of the RDCs for IL-1 β are underestimated, which shows that the two halves of the complex could not be fitted to the same alignment tensor. When NOEs and RDCs were used in the docking protocol the resulting

structures fitted all of the available experimental data, but the resulting structures were not significantly different to the structures calculated using only the NOEs.

In this case RDC and AIR data alone was not sufficient to confidently dock the protein complex. This is possibly due to the low coverage with RDC data available for only 56% of the backbone amide groups of gIC8 Fab. Increasing the coverage of RDCs could be possible using 3D experiments to determine the RDCs and RDC data could be improved by the use of different alignment media giving a different alignment tensor, or measuring different couplings. However, despite exhaustive efforts it was not possible to find a different alignment media for gIC8 Fab that gave consistent RDCs and other backbone RDCs that can be measured tend to be smaller than J_{N-H} which makes accurate measurement challenging.

Comparisons can be drawn between the use of RDCs to dock the gIC8 Fab/IL-1 β complex and RDCs to dock the gIC8 scFv/IL-1 β complex. The scFv is approximately half the size of the Fab resulting in fewer peaks in TROSY/HSQC spectra allowing determination of a more complete set of RDC data (65%). This docking resulted in one highly populated cluster and 5 sparsely populated clusters of structures. The lowest scoring, most populated cluster showed a low RMSD with the final result being a highly convincing docked structure. However, the scFv/IL-1 β complex shows fairly large RMSDs when compared to any of the docked gIC8 Fab/IL-1 β complex structures presented in this chapter. It is possible that the scFv is not interacting with IL-1 β in an identical manner to the Fab, but this is unlikely given the similarity of the amide ${}^{15}N/{}^{1}H$ chemical shifts for IL-1 β bound to the scFv or the Fab. When compared with the crystal structure for gIC8 Fab the backbone atoms of IL-1 β show an RMSD of 8.34 Å (when aligned on the variable domains of the Fab and scFv). The RMSD increases to 8.88 Å when considering only interface residues. This shows that even with a relatively full set of RDC data the docked structure can be some way away from reality. The likely cause of the incorrectly docked gIC8 scFv/IL-1 β complex is the large chemical shift changes that occur throughout the variable domains upon binding to IL-1 β , resulting in failure to define a discrete interaction site.

Docking using all the available data resulted in the structure that is closest to the crystal structure with a backbone RMSD of 1.81 Å and an IL-1 β displacement of 4.66 Å. Analysis of the structures showed that a similar number of residues were making

intermolecular contacts (17 for the crystal structure and 16 for the docked structure). 5 of these intermolecular contacts were shared between the two structures, with the majority of the other intermolecular contacts differing by only one or two residues. Alignment of the two structures on the backbone atoms of gIC8 Fab shows that the refined homology model of the Fab is fairly accurate, with a backbone RMSD of 1.3 Å. The CDR loops have also been modelled correctly with the exception of CDR 2 of the heavy chain, which is projecting further away from the framework of the variable domain when compared to the crystal structure, which can be visualised in Figure 3.38. CDR2 of the heavy chain forms the major part of the interaction interface with IL-1 β . When the structures are aligned on the backbone atoms of IL-1 β the position of CDR H2 relative to IL-1 β is the same in both the crystal structure and the docked structure. It is possible that the incorrect position of CDR H2 in the gIC8 Fab model is responsible for the differences seen between the crystal and docked structures of the complex. CDR H2 was identified as problematic during the homology modelling of the initial scFv model due to low sequence identity with the template structure and so was modelled using the structure of a CDR from a different structure (Wilkinson, 2009).



Figure 3.38: Comparison of gIC8 Fab/IL-1β crystal structure and the lowest scoring docked structure calculated using all of the available data

Cartoon representations of the gIC8 Fab/IL-1 β complex crystal structure (green) and the lowest scoring docked structure (blue) aligned on the superposed on the backbone atoms of IL-1 β (B and C only). B shows a close up of the boxed area in A, highlighting the similarity in the position of CDR H2 relative to IL-1 β , forming a parallel β -sheer with the resides of β -12 in IL-1 β . When the structures are superposed on the backbone atoms of gIC8 Fab (C) the displacement of CDR H2 relative to the framework is evident, whilst CDRs H3 and L3 appear to have been modelled correctly (IL-1 β atoms are not shown for clarity).

Chapter 4: Summary

The majority of proteins form complexes with other proteins and/or nucleic acids and these interactions are essential to the proteins cellular role. Solving the structures of these complexes is vital for a full understanding of the proteins function. The structure of proteins also provides a substantial benefit to knowledge based drug development. To date, many protein complexes have resisted attempts of structure determination by established methods. This makes an approach based on modelling complexes using experimental data an attractive alternative when structures of the individual proteins are available.

The work presented in this thesis describes a strategy for producing reliable models of protein complexes using docking calculations driven by experimental restraints. A key aspect of the envisaged approach was the identification of intermolecular NOEs across protein-protein interfaces, which are generally rich in methyl containing amino acids. Given the importance of methyl groups at protein-protein interfaces the initial focus was on a strategy for assigning methyl groups using an NOE based approach in order to avoid the limitations imposed by the molecular weight of larger complexes.

Solution NMR studies of large proteins generally rely on ²H labelling of the protein where all carbon bound protons are replaced by deuterons. Whilst allowing assignment of the backbone resonances in large proteins, deuteration precludes the assignment of side chain ¹H signals, limiting the structural information available from ¹H-¹H interatomic distance restraints. The work presented in chapter two describes the *in silico* testing, validation and application of a technique that uses H_N-CH₃ NOEs to determine ¹³C/¹H assignments for the methyl groups in residually protonated protein samples. The result is a potentially robust and reliable approach which has applications in a range of studies from small proteins to large protein complexes. The benefits of this approach include the use of standard NOESY based experiments and samples that do not require highly specific isotope labelling. The samples required for this technique are also significantly less expensive to produce compared to labelled samples produced using specifically labelled biosynthetic precursors.

The applicability of this approach was demonstrated by obtaining comprehensive assignments for the methyl groups of IL-1 β (17 kDa) both in the free form and in complex with a potential therapeutic Fab antibody fragment (a complex of 65 kDa). It was shown that these assignments could be used to identify a number of backbone amide to side chain methyl NOEs across the protein-protein interface. These NOEs provided a significant number of structural restraints that made a substantial difference to the accuracy and reliability of docked structures of the IL-1 β /gIC8 Fab complex, which has subsequently been demonstrated by comparison to a crystal structure of the complex.

The developed approach is expected to be generally applicable to a wide range of protein complexes up to a molecular weight of approximately 100 kDa as this is currently the limit for which backbone assignments are achievable. The use of a homology model as the starting structure for the Fab fragment also demonstrates that this technique is tolerant of small differences in the starting and final structures.

Appendix

A.1: Scripts used for molecular modelling of tri-peptides

A.1.1: Nested loop used to model Gly-Leu-Gly peptide

```
echo:=off
do psil -180 180 30
   do phi1 -180 180 30
      do psi2 -180 180 30
         do phi2 -180 180 30
            do chi1 0 360 30
               do chi2 0 360 30
      print "@" >>1.vdw
      angles set "PSI 1" value=psi1 info=none
      angles set "PHI 1" value=phil info=none
      angles set "PSI 2" value=psi2 info=none
      angles set "PHI 2" value=phi2 info=none
      angles set "CHI1 2" value=chi1 info=none
      angles set "CHI2 2" value=chi2 info=none
            angles set "PSI 3" value=-60 info=none
            angles set "PHI 3" value=-60 info=none
            angles set "CHI31 2" value=0 info=none
            angles set "CHI32 2" value=0 info=none
            write pdb 1
            read pdb 1
            structure list >>1.vdw
            d1=dist('QD1 2','H 1')
            d2=dist('QD2 2','H 1')
            d3=dist('QD1 2','H 2')
            d4=dist('QD2 2','H 2')
            d5=dist('QD1 2','H 3')
            d6=dist('QD2 2','H 3')
            print "psil $psil phil $phil psi2 $psi2 phi2 $phi2 chil
$chi1 chi2 $chi2 QD1H1 $d1 QD2H1 $d2 QD1H2 $d3 QD2H2 $d4 QD1H3 $d5
QD2H3 $d6" >>1.vdw
               end do
            end do
         end do
      end do
   end do
end do
```

The dihedral angles that affect the interatomic (i-1,i and i+1) H_N -CH₃ distances are varied in 30° increments, with the van der Waals violations, set dihedral angles and interatomic distances written to a text file.

A.1.2: Awk script used to filter interatomic distances

```
#!/usr/bin/awk -f
BEGIN {
   FS ="[\n]"
   RS ="@"
        }
   { split($7,a," ")
   }
   {
    if ($7=a[5] < 0.25 && $7=a[4] < 1)
        print $9, $7=a[2], $7=a[4], $7=a[5]
   }
}</pre>
```

The interatomic distances generated in CYANA using the script in A.1.1 is filtered with only the distances from conformers with maximum individual van der Waals violations <0.25 Å and a sum of violations for the whole structure <1 Å.

References:

- Addis, P.W., Hall, C.J., Bruton, S., Veverka, V., Wilkinson, I.C., Muskett, F.W., Renshaw, P.S., Prosser, C.E., Carrington, B., Lawson, A.D.G., Griffin, R., Taylor, R.J., Waters, L.C., Henry, A.J., Carr, M.D., 2014. Conformational heterogeneity in antibody-protein antigen recognition: implications for high affinity protein complex formation. J. Biol. Chem. 289, 7200–10. doi:10.1074/jbc.M113.492215
- Amero, C., Asunción Durá, M., Noirclerc-Savoye, M., Perollier, A., Gallet, B., Plevin, M.J., Vernet, T., Franzetti, B., Boisbouvier, J., 2011. A systematic mutagenesisdriven strategy for site-resolved NMR studies of supramolecular assemblies. J. Biomol. NMR 50, 229–36. doi:10.1007/s10858-011-9513-5
- Aue, W.P., 1976. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. J. Chem. Phys. 64, 2229. doi:10.1063/1.432450
- Ayala, I., Sounier, R., Usé, N., Gans, P., Boisbouvier, J., 2009. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. J. Biomol. NMR 43, 111–9. doi:10.1007/s10858-008-9294-7
- Barksby, H.E., Lea, S.R., Preshaw, P.M., Taylor, J.J., 2007. The expanding family of interleukin-1 cytokines and their role in destructive inflammatory disorders. Clin. Exp. Immunol. 149, 217–25. doi:10.1111/j.1365-2249.2007.03441.x
- Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A., Subramaniam, S., 2014. Structure of β -galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. Proc. Natl. Acad. Sci. U. S. A. 111, 11709–14. doi:10.1073/pnas.1402809111
- Bermel, W., Bertini, I., Felli, I.C., Kümmerle, R., Pierattelli, R., 2003. 13C direct detection experiments on the paramagnetic oxidized monomeric copper, zinc superoxide dismutase. J. Am. Chem. Soc. 125, 16423–9. doi:10.1021/ja037676p
- Billeter, M., Braun, W., Wüthrich, K., 1982. Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. J. Mol. Biol. 155, 321–346. doi:10.1016/0022-2836(82)90008-0
- Blech, M., Peter, D., Fischer, P., Bauer, M.M.T., Hafner, M., Zeeb, M., Nar, H., 2013. One target-two different binding modes: structural insights into gevokizumab and canakinumab interactions to interleukin-1β. J. Mol. Biol. 425, 94–111. doi:10.1016/j.jmb.2012.09.021
- Bloch, F., Hansen, W., Packard, M., 1946. The Nuclear Induction Experiment. Phys. Rev. 70, 474–485. doi:10.1103/PhysRev.70.474
- Bodenhausen, G., Ruben, D.J., 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem. Phys. Lett. 69, 185–189. doi:10.1016/0009-2614(80)80041-8

- Bonvin, A.M.J.J., 2006. Flexible protein-protein docking. Curr. Opin. Struct. Biol. 16, 194–200. doi:10.1016/j.sbi.2006.02.002
- Boraschi, D., Tagliabue, A., 2006. The interleukin-1 receptor family. Vitam. Horm. 74, 229–54. doi:10.1016/S0083-6729(06)74009-2
- Cavanagh, J., Fairbrother, W., Palmer III, A., Rance, M., Skelton, N., 2007. Protein NMR spectroscopy. Principles and Practice., Elselvier Academic Press. doi:10.1093/cid/cis040
- Chao, F.-A., Kim, J., Xia, Y., Milligan, M., Rowe, N., Veglia, G., 2014. FLAMEnGO 2.0: An enhanced fuzzy logic algorithm for structure-based assignment of methyl group resonances. J. Magn. Reson. 245, 17–23. doi:10.1016/j.jmr.2014.04.012
- Chao, F.-A., Shi, L., Masterson, L.R., Veglia, G., 2012. FLAMEnGO: a fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. J. Magn. Reson. 214, 103–10. doi:10.1016/j.jmr.2011.10.008
- Chapman, A.P., 2002. PEGylated antibodies and antibody fragments for improved therapy: a review. Adv. Drug Deliv. Rev. 54, 531–545. doi:10.1016/S0169-409X(02)00026-1
- Choy, E.H.S., Hazleman, B., Smith, M., Moss, K., Lisi, L., Scott, D.G.I., Patel, J., Sopwith, M., Isenberg, D.A., 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 (Oxford). 41, 1133–7.
- Clark, L. a, Boriack-Sjodin, P.A., Eldredge, J., Fitch, C., Friedman, B., Hanf, K.J.M., Jarpe, M., Liparoto, S.F., Li, Y., Lugovskoy, A., Miller, S., Rushe, M., Sherman, W., Simon, K., Van Vlijmen, H., 2006. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. Protein Sci. 15, 949–60. doi:10.1110/ps.052030506
- Clore, G.M., 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. Proc. Natl. Acad. Sci. U. S. A. 97, 9021–5.
- Clore, G.M., Bax, A., Driscoll, P.C., Wingfield, P.T., Gronenborn, A.M., 1990. Assignment of the side-chain 1H and 13C resonances of interleukin-1 beta using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. Biochemistry 29, 8172–84.
- Clore, G.M., Gronenborn, A.M., 1994. Multidimensional heteronuclear nuclear magnetic resonance of proteins. Methods Enzymol. 239, 349–63.
- Clore, G.M., Robien, M.A., Gronenborn, A.M., 1993. Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. J. Mol. Biol. 231, 82–102. doi:10.1006/jmbi.1993.1259

Dhimolea, E., 2010. Canakinumab. MAbs 2, 3–13.

- Dinarello, C.A., Wolff, S.M., 1993. The role of interleukin-1 in disease. N. Engl. J. Med. 328, 106–13. doi:10.1056/NEJM199301143280207
- Dominguez, C., Boelens, R., Bonvin, A.M.J.J., 2003. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. 125, 1731–7. doi:10.1021/ja026939x
- Driscoll, P.C., Clore, G.M., Marion, D., Wingfield, P.T., Gronenborn, A.M., 1990.
 Complete resonance assignment for the polypeptide backbone of interleukin 1.beta. using three-dimensional heteronuclear NMR spectroscopy. Biochemistry 29, 3542–3556. doi:10.1021/bi00466a018
- Dubs, A., Wagner, G., Wüthrich, K., 1979. Individual assignments of amide proton resonances in the proton NMR spectrum of the basic pancreatic trypsin inhibitor. Biochim. Biophys. Acta - Protein Struct. 577, 177–194. doi:10.1016/0005-2795(79)90020-5
- Eder, C., 2009. Mechanisms of interleukin-1beta release. Immunobiology 214, 543–53. doi:10.1016/j.imbio.2008.11.007
- Ernst, R.R., 1966. Application of Fourier Transform Spectroscopy to Magnetic Resonance. Rev. Sci. Instrum. 37, 93. doi:10.1063/1.1719961
- Faulds, D., Sorkin, E.M., 1994. Abciximab (c7E3 Fab). A review of its pharmacology and therapeutic potential in ischaemic heart disease. Drugs 48, 583–98.
- Finzel, B.C., Clancy, L.L., Holland, D.R., Muchmore, S.W., Watenpaugh, K.D., Einspahr, H.M., 1989. Crystal structure of recombinant human interleukin-1β at 2·0 Å resolution. J. Mol. Biol. 209, 779–791. doi:10.1016/0022-2836(89)90606-2
- Garrett, D.S., Seok, Y.J., Peterkofsky, a, Gronenborn, a M., Clore, G.M., 1999. Solution structure of the 40,000 Mr phosphoryl transfer complex between the Nterminal domain of enzyme I and HPr. Nat. Struct. Biol. 6, 166–73. doi:10.1038/5854

Goddard, T., Kneller, D., n.d. SPARKY 3, University of California, San Francisco.

- Greenfeder, S.A., Nunes, P., Kwee, L., Labow, M., Chizzonite, R.A., Ju, G., 1995. Molecular cloning and characterization of a second subunit of the interleukin 1 receptor complex. J. Biol. Chem. 270, 13757–65.
- Gross, J.D., Gelev, V.M., Wagner, G., 2003. A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. J. Biomol. NMR 25, 235–42.
- Güntert, P., Braun, W., Wüthrich, K., 1991. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the

program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. J. Mol. Biol. 217, 517–530. doi:10.1016/0022-2836(91)90754-T

- Güntert, P., Wüthrich, K., 1991. Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. J. Biomol. NMR 1, 447–56.
- Guo, C., Tugarinov, V., 2009. Identification of HN-methyl NOEs in large proteins using simultaneous amide-methyl TROSY-based detection. J. Biomol. NMR 43, 21–30. doi:10.1007/s10858-008-9285-8
- Guo, C., Tugarinov, V., 2010. Selective 1H- 13C NMR spectroscopy of methyl groups in residually protonated samples of large proteins. J. Biomol. NMR 46, 127–33. doi:10.1007/s10858-009-9393-0
- Hall, C.J., 2010. Comparisons of Solution Structures of Free and IL-1β bound Fab Suggests a Model for B Cell Receptor Signalling. Ph. D. University of Leicester
- Harris, L.J., Larson, S.B., Hasel, K.W., McPherson, A., 1997. Refined structure of an intact IgG2a monoclonal antibody. Biochemistry 36, 1581–97. doi:10.1021/bi962514+
- Havel, T.F., Wüthrich, K., 1985. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. J. Mol. Biol. 182, 281–294. doi:10.1016/0022-2836(85)90346-8
- Herrmann, T., Güntert, P., Wüthrich, K., 2002. Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. J. Mol. Biol. 319, 209–227. doi:10.1016/S0022-2836(02)00241-3
- Hovmöller, S., Zhou, T., Ohlson, T., 2002. Conformations of amino acids in proteins. Acta Crystallogr. D. Biol. Crystallogr. 58, 768–76.
- Hu, K., Vögeli, B., Clore, G.M., 2006. 13C-detected HN(CA)C and HMCMC experiments using a single methyl-reprotonated sample for unambiguous methyl resonance assignment. J. Biomol. NMR 36, 259–66. doi:10.1007/s10858-006-9090-1
- Hyberts, S.G., Milbradt, A.G., Wagner, A.B., Arthanari, H., Wagner, G., 2012. Application of iterative soft thresholding for fast reconstruction of NMR data nonuniformly sampled with multidimensional Poisson Gap scheduling. J. Biomol. NMR 52, 315–27. doi:10.1007/s10858-012-9611-z
- Hyberts, S.G., Robson, S. a, Wagner, G., 2013. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. J. Biomol. NMR 55, 167–78. doi:10.1007/s10858-012-9698-2

- Janin, J., Miller, S., Chothia, C., 1988. Surface, subunit interfaces and interior of oligomeric proteins. J. Mol. Biol. 204, 155–164. doi:10.1016/0022-2836(88)90606-7
- Janin, J., Wodak, S., Levitt, M., Maigret, B., 1978. Conformation of amino acid sidechains in proteins. J. Mol. Biol. 125, 357–386. doi:10.1016/0022-2836(78)90408-4
- Jaravine, V., Ibraghimov, I., Orekhov, V.Y., 2006. Removal of a time barrier for highresolution multidimensional NMR spectroscopy. Nat. Methods 3, 605–7. doi:10.1038/nmeth900
- Jeener, J., Meier, B.H., Bachmann, P., Ernst, R.R., 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. J. Chem. Phys. 71, 4546. doi:10.1063/1.438208
- Jones, P.T., Dear, P.H., Foote, J., Neuberger, M.S., Winter, G., 1986. Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature 321, 522–5. doi:10.1038/321522a0
- Kastritis, P.L., Bonvin, A.M.J.J., 2013. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. J. R. Soc. Interface 10, 20120835. doi:10.1098/rsif.2012.0835
- Kay, L.E., 2011. Solution NMR spectroscopy of supra-molecular systems, why bother? A methyl-TROSY view. J. Magn. Reson. 210, 159–70. doi:10.1016/j.jmr.2011.03.008
- Kay, L.E., Bull, T., Nicholson, L., Griesinger, C., Schwalbe, H., Bax, A., Torchia, D., 1992. The measurement of heteronuclear transverse relaxation times in ax3 spin systems via polarization-transfer techniques. J. Magn. Reson. 100, 538–558. doi:10.1016/0022-2364(92)90058-F
- Kay, L.E., Ikura, M., Tschudin, R., Bax, A., 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J. Magn. Reson. 89, 496– 514. doi:10.1016/0022-2364(90)90333-5
- Kendrew, J.C., Parrish, R.G., 1957. The Crystal Structure of Myoglobin. III. Sperm-Whale Myoglobin. Proc. R. Soc. A Math. Phys. Eng. Sci. 238, 305–324. doi:10.1098/rspa.1957.0002
- Kiyoshi, M., Caaveiro, J.M.M., Miura, E., Nagatoishi, S., Nakakido, M., Soga, S., Shirai, H., Kawabata, S., Tsumoto, K., 2014. Affinity improvement of a therapeutic antibody by structure-based computational design: generation of electrostatic interactions in the transition state stabilizes the antibody-antigen complex. PLoS One 9, e87099. doi:10.1371/journal.pone.0087099
- Kobe, B., Guncar, G., Buchholz, R., Huber, T., Maco, B., Cowieson, N., Martin, J.L., Marfori, M., Forwood, J.K., 2008. Crystallography and protein-protein

interactions: biological interfaces and crystal contacts. Biochem. Soc. Trans. 36, 1438–41. doi:10.1042/BST0361438

- Koharudin, L.M., Bonvin, A.M.J., Kaptein, R., Boelens, R., 2003. Use of very longdistance NOEs in a fully deuterated protein: an approach for rapid protein fold determination. J. Magn. Reson. 163, 228–235. doi:10.1016/S1090-7807(03)00149-6
- Kontaxis, G., Clore, G.M., Bax, A., 2000. Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. J. Magn. Reson. 143, 184–96. doi:10.1006/jmre.1999.1979
- Leiting, B., Marsilio, F., O'Connell, J.F., 1998. Predictable deuteration of recombinant proteins expressed in Escherichia coli. Anal. Biochem. 265, 351–5. doi:10.1006/abio.1998.2904
- LeMaster, D.M., Richards, F.M., 1985. 1H-15N heteronuclear NMR studies of Escherichia coli thioredoxin in samples isotopically labeled by residue type. Biochemistry 24, 7263–8.
- Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D. a, Cheng, Y., 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods 10, 584–90. doi:10.1038/nmeth.2472
- Liao, X., Tugarinov, V., 2011. Selective detection of ¹³CHD₂ signals from a mixture of ¹³CH₃/¹³CH₂D/¹³CHD₂ methyl isotopomers in proteins. J. Magn. Reson. 209, 101–7. doi:10.1016/j.jmr.2010.12.014
- Lu, P., Bai, X., Ma, D., Xie, T., Yan, C., Sun, L., Yang, G., Zhao, Y., Zhou, R., Scheres, S.H.W., Shi, Y., 2014. Three-dimensional structure of human γ-secretase. Nature 512, 166–170. doi:10.1038/nature13567
- Mal, T.K., Matthews, S.J., Kovacs, H., Campbell, I.D., Boyd, J., 1998. Some NMR experiments and a structure determination employing a [15N,2H] enriched protein. J. Biomol. NMR 12, 259–76.
- Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., Bax, A., 1989. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. J. Am. Chem. Soc. 111, 1515–1517. doi:10.1021/ja00186a066
- Morrison, S.L., Johnson, M.J., Herzenberg, L.A., Oi, V.T., 1984. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. Proc. Natl. Acad. Sci. U. S. A. 81, 6851–5.
- Muhandiram, D.R., Farrow, N.A., Xu, G.Y., Smallcombe, S.H., Kay, L.E., 1993a. A Gradient 13C NOESY-HSQC Experiment for Recording NOESY Spectra of 13C-Labeled Proteins Dissolved in H2O. J. Magn. Reson. Ser. B 102, 317–321. doi:10.1006/jmrb.1993.1102
- Muhandiram, D.R., Xu, G., Kay, L., 1993b. An enhanced-sensitivity pure absorption gradient 4D 15N, 13C-edited NOESY experiment. J. Biomol. NMR 3. doi:10.1007/BF00176011
- Nagayama, K., Wüthrich, K., Bachmann, P., Ernst, R.R., 1977. Two-dimensional Jresolved 1H n.m.r. spectroscopy for studies of biological macromolecules. Biochem. Biophys. Res. Commun. 78, 99–105. doi:10.1016/0006-291X(77)91226-8
- Neuhaus, D., Williamson, M.P., 2000. The nuclear Overhauser effect in structural and conformational analysis. Wiley-VCH.
- Newman, J., Egan, D., Walter, T.S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J.L., Stuart, D.I., Perrakis, A., 2005. Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. Acta Crystallogr. D. Biol. Crystallogr. 61, 1426–31. doi:10.1107/S0907444905024984
- Norman, D.J., Chatenoud, L., Cohen, D., Goldman, M., Shield, C.F., 1993. Consensus statement regarding OKT3-induced cytokine-release syndrome and human antimouse antibodies. Transplant. Proc. 25, 89–92.
- Ofran, Y., Rost, B., 2003. Analysing six types of protein-protein interfaces. J. Mol. Biol. 325, 377–87.
- OKA, C., 1995. V(D)J recombination of immunoglobulin genes. Adv. Biophys. 31, 163–180. doi:10.1016/0065-227X(95)99390-B
- Ollerenshaw, J.E., Tugarinov, V., Kay, L.E., 2003. Methyl TROSY: explanation and experimental verification. Magn. Reson. Chem. 41, 843–852. doi:10.1002/mrc.1256
- Ollerenshaw, J.E., Tugarinov, V., Skrynnikov, N.R., Kay, L.E., 2005. Comparison of 13CH3, 13CH2D, and 13CHD2 methyl labeling strategies in proteins. J. Biomol. NMR 33, 25–41. doi:10.1007/s10858-005-2614-2
- Orekhov, V.Y., Ibraghimov, I., Billeter, M., 2003. Optimizing resolution in multidimensional NMR by three-way decomposition. J. Biomol. NMR 27, 165– 173. doi:10.1023/A:1024944720653
- Orekhov, V.Y., Jaravine, V.A., 2011. Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. Prog. Nucl. Magn. Reson. Spectrosc. 59, 271– 92. doi:10.1016/j.pnmrs.2011.02.002
- Otten, R., Chu, B., Krewulak, K.D., Vogel, H.J., Mulder, F. a a, 2010. Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. J. Am. Chem. Soc. 132, 2952–60. doi:10.1021/ja907706a

- Otting, G., Senn, H., Wagner, G., Wüthrich, K., 1986. Editing of 2D 1H NMR spectra using X half-filters. combined use with residue-selective 15N labeling of proteins. J. Magn. Reson. 70, 500–505. doi:10.1016/0022-2364(86)90144-7
- Overhauser, A., 1953. Polarization of Nuclei in Metals. Phys. Rev. 92, 411–415. doi:10.1103/PhysRev.92.411
- Overhauser, A.W., 1953. Paramagnetic Relaxation in Metals. Phys. Rev. 89, 689–700. doi:10.1103/PhysRev.89.689
- Perelson, A.S., Oster, G.F., 1979. Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination. J. Theor. Biol. 81, 645–670. doi:10.1016/0022-5193(79)90275-3
- PERUTZ, M.F., ROSSMANN, M.G., CULLIS, A.F., MUIRHEAD, H., WILL, G., NORTH, A.C.T., 1960. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. Nature 185, 416– 422. doi:10.1038/185416a0
- Pervushin, K., Riek, R., Wider, G., 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. Proc. Natl. Acad. Sci. 94, 12366–12371. doi:10.1073/pnas.94.23.12366
- Piotto, M., Saudek, V., Sklenář, V., 1992. Gradient-tailored excitation for singlequantum NMR spectroscopy of aqueous solutions. J. Biomol. NMR 2, 661–665. doi:10.1007/BF02192855
- Poljak, R.J., Amzel, L.M., Avey, H.P., Chen, B.L., Phizackerley, R.P., Saul, F., 1973. Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2,8-A resolution. Proc. Natl. Acad. Sci. U. S. A. 70, 3305–10.
- Porter, R.R., 1959. The hydrolysis of rabbit y-globulin and antibodies with crystalline papain. Biochem. J. 73, 119–26.
- Priestle, J.P., Schär, H.P., Grütter, M.G., 1989. Crystallographic refinement of interleukin 1 beta at 2.0 A resolution. Proc. Natl. Acad. Sci. U. S. A. 86, 9667–71.
- Radaev, S., Li, S., Sun, P.D., 2006. A survey of protein-protein complex crystallizations. Acta Crystallogr. D. Biol. Crystallogr. 62, 605–12. doi:10.1107/S0907444906011735

Reichert, J.M., 2010. Antibody-based therapeutics to watch in 2011. MAbs 3, 76–99.

Renshaw, P.S., Lightbody, K.L., Veverka, V., Muskett, F.W., Kelly, G., Frenkiel, T.A., Gordon, S. V, Hewinson, R.G., Burke, B., Norman, J., Williamson, R.A., Carr, M.D., 2005. Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. EMBO J. 24, 2491–8. doi:10.1038/sj.emboj.7600732

- Rosen, M.K., Gardner, K.H., Willis, R.C., Parris, W.E., Pawson, T., Kay, L.E., 1996. Selective methyl group protonation of perdeuterated proteins. J. Mol. Biol. 263, 627–36. doi:10.1006/jmbi.1996.0603
- Rovnyak, D., Hoch, J.C., Stern, A.S., Wagner, G., 2004. Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy. J. Biomol. NMR 30, 1–10. doi:10.1023/B:JNMR.0000042946.04002.19
- Schroeder, H.W., Cavacini, L., 2010. Structure and function of immunoglobulins. J. Allergy Clin. Immunol. 125, S41–52. doi:10.1016/j.jaci.2009.09.046
- Schwieters, C.D., Kuszewski, J.J., Clore, G.M., 2006. Using Xplor-NIH for NMR Molecular Structure Determination. ChemInform 37. doi:10.1002/chin.200644278
- Schwieters, C.D., Kuszewski, J.J., Tjandra, N., Marius Clore, G., 2003. The Xplor-NIH NMR molecular structure determination package. J. Magn. Reson. 160, 65–73. doi:10.1016/S1090-7807(02)00014-9
- Shekhtman, A., Ghose, R., Goger, M., Cowburn, D., 2002. NMR structure determination and investigation using a reduced proton (REDPRO) labeling strategy for proteins. FEBS Lett. 524, 177–82.
- Shen, Y., Delaglio, F., Cornilescu, G., Bax, A., 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR 44, 213–23. doi:10.1007/s10858-009-9333-z
- Sheppard, D., Guo, C., Tugarinov, V., 2009a. Methyl-detected "out-and-back" NMR experiments for simultaneous assignments of Alabeta and Ilegamma2 methyl groups in large proteins. J. Biomol. NMR 43, 229–38. doi:10.1007/s10858-009-9305-3
- Sheppard, D., Guo, C., Tugarinov, V., 2009b. 4D 1H-13C NMR spectroscopy for assignments of alanine methyls in large and complex protein structures. J. Am. Chem. Soc. 131, 1364–5. doi:10.1021/ja808202q
- Sinha, K., Jen-Jacobson, L., Rule, G.S., 2013. Divide and conquer is always best: sensitivity of methyl correlation experiments. J. Biomol. NMR 56, 331–5. doi:10.1007/s10858-013-9751-9
- Smith, M.T.J., Rubinstein, J.L., 2014. Beyond blob-ology. Science (80-.). 345, 617–619. doi:10.1126/science.1256358
- Smith, S.L., 1996. Ten years of Orthoclone OKT3 (muromonab-CD3): a review. J. Transpl. Coord. 6, 109–19; quiz 120–1.
- Soga, S., Kuroda, D., Shirai, H., Kobori, M., Hirayama, N., 2010. Use of amino acid composition to predict epitope residues of individual antibodies. Protein Eng. Des. Sel. 23, 441–8. doi:10.1093/protein/gzq014

- Solomon, I., 1955. Relaxation Processes in a System of Two Spins. Phys. Rev. 99, 559–565. doi:10.1103/PhysRev.99.559
- Sounier, R., Blanchard, L., Wu, Z., Boisbouvier, J., 2007. High-accuracy distance measurement between remote methyls in specifically protonated proteins. J. Am. Chem. Soc. 129, 472–3. doi:10.1021/ja067260m
- Spirin, V., Mirny, L.A., 2003. Protein complexes and functional modules in molecular networks. Proc. Natl. Acad. Sci. U. S. A. 100, 12123–8. doi:10.1073/pnas.2032324100
- Stern, A.S., Li, K.-B., Hoch, J.C., 2002. Modern Spectrum Analysis in Multidimensional NMR Spectroscopy: Comparison of Linear-Prediction Extrapolation and Maximum-Entropy Reconstruction. J. Am. Chem. Soc. 124, 1982–1993. doi:10.1021/ja0116690
- Sundquist, W.I., Schubert, H.L., Kelly, B.N., Hill, G.C., Holton, J.M., Hill, C.P., 2004. Ubiquitin Recognition by the Human TSG101 Protein. Mol. Cell 13, 783–789. doi:10.1016/S1097-2765(04)00129-7
- Tang, C., Clore, G.M., 2006. A simple and reliable approach to docking protein-protein complexes from very sparse NOE-derived intermolecular distance restraints. J. Biomol. NMR 36, 37–44. doi:10.1007/s10858-006-9065-2
- Thomas, C., Bazan, J.F., Garcia, K.C., 2012. Structure of the activating IL-1 receptor signaling complex. Nat. Struct. Mol. Biol. 19, 455–7. doi:10.1038/nsmb.2260
- Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302, 575–581. doi:10.1038/302575a0
- Trepel, F., 1974. Number and distribution of lymphocytes in man. A critical analysis. Klin. Wochenschr. 52, 511–515. doi:10.1007/BF01468720
- Tugarinov, V., Kay, L.E., 2003. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. J. Am. Chem. Soc. 125, 13868–13878.
- Vaynberg, J., Fukuda, T., Chen, K., Vinogradova, O., Velyvis, A., Tu, Y., Ng, L., Wu, C., Qin, J., 2005. Structure of an ultraweak protein-protein complex and its crucial role in regulation of cell morphology and motility. Mol. Cell 17, 513–23. doi:10.1016/j.molcel.2004.12.031
- Velyvis, A., Ruschak, A.M., Kay, L.E., 2012. An economical method for production of (2)H, (13)CH3-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. PLoS One 7, e43725. doi:10.1371/journal.pone.0043725
- Venters, R.A., Farmer, B.T., Fierke, C.A., Spicer, L.D., 1996. Characterizing the use of perdeuteration in NMR studies of large proteins: 13C, 15N and 1H assignments of human carbonic anhydrase II. J. Mol. Biol. 264, 1101–16.

- Veverka, V., Crabbe, T., Bird, I., Lennie, G., Muskett, F.W., Taylor, R.J., Carr, M.D., 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– 95. doi:10.1038/sj.onc.1210693
- Vitali Tugarinov, R.M., 2002. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. J. Am. Chem. Soc. 124, 10025 35.
- Vuister, G.W., Bax, A., 1992. Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling. J. Magn. Reson. 98, 428–435. doi:10.1016/0022-2364(92)90144-V
- Walters, K., Matsuo, H., Wagner, G., 1997. A simple method to distinguish intermonomer nuclear overhauser effects in homodimeric proteins with C 2 symmetry. J. Am. ... 7863, 9015–9016.
- Walters, K.J., Ferentz, A.E., Hare, B.J., Hidalgo, P., Jasanoff, A., Matsuo, H., Wagner, G., 2001. Nuclear Magnetic Resonance of Biological Macromolecules - Part B, Methods in Enzymology, Methods in Enzymology. Elsevier. doi:10.1016/S0076-6879(01)39316-3
- Wang, D., Zhang, S., Li, L., Liu, X., Mei, K., Wang, X., 2010. Structural insights into the assembly and activation of IL-1β with its receptors. Nat. Immunol. 11, 905–11. doi:10.1038/ni.1925
- Weber, A., Wasiliew, P., Kracht, M., 2010. Interleukin-1 (IL-1) pathway. Sci. Signal. 3, cm1. doi:10.1126/scisignal.3105cm1
- Wilkinson, I.C., 2009. Solution Structure of Interleukin-1β Bound To a Potential Therapeutic Antibody. Ph. D. University of Leicester
- Wilkinson, I.C., Hall, C.J., Veverka, V., Shi, J.Y., Muskett, F.W., Stephens, P.E., Taylor, R.J., Henry, A.J., Carr, M.D., 2009. High resolution NMR-based model for the structure of a scFv-IL-1beta complex: potential for NMR as a key tool in therapeutic antibody design and development. J. Biol. Chem. 284, 31928–35. doi:10.1074/jbc.M109.025304
- Williamson, M.P., Havel, T.F., Wüthrich, K., 1985. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. J. Mol. Biol. 182, 295–315. doi:10.1016/0022-2836(85)90347-X
- Wittekind, M., Mueller, L., 1993. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. J. Magn. Reson. Ser. B 101, 201–205. doi:10.1006/jmrb.1993.1033

- Wodak, S.J., Janin, J., 1978. Computer analysis of protein-protein interaction. J. Mol. Biol. 124, 323–42.
- Wüthrich, K., Billeter, M., Braun, W., 1983. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J. Mol. Biol. 169, 949–61.
- Xu, Y., Lin, Z., Ho, C., Yang, D., 2005. A General Strategy for the Assignment of Aliphatic Side-Chain Resonances of Uniformly 13C, 15N-labelled large proteins. J. Am. Chem. Soc. 127, 11920–11921.
- Xu, Y., Liu, M., Simpson, P.J., Isaacson, R., Cota, E., Marchant, J., Yang, D., Zhang, X., Freemont, P., Matthews, S., 2009. Automated assignment in selectively methyl-labeled proteins. J. Am. Chem. Soc. 131, 9480–1. doi:10.1021/ja9020233
- Yang, D., Zheng, Y., Liu, D., Wyss, D.F., 2004. Sequence-specific assignments of methyl groups in high-molecular weight proteins. J. Am. Chem. Soc. 126, 3710–1. doi:10.1021/ja039102q
- Zhu, G., Kong, X.M., Sze, K.H., 1999. Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. J. Biomol. NMR 13, 77–81. doi:10.1023/A:1008398227519
- Zhu, G., Xia, Y., Sze, K.H., Yan, X., 1999. 2D and 3D TROSY-enhanced NOESY of 15N labeled proteins. J. Biomol. NMR 14, 377–381. doi:10.1023/A:1008386525465
- Zweckstetter, M., Bax, A., 2000. Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. J. Am. Chem. Soc. 122, 3791–3792. doi:10.1021/ja0000908