University of Southampton

Faculty of Medicine

Cancer Sciences Unit

MHC Class I Protein Dynamics and Antigen Presenting Function

Louise Bolton

Thesis for the degree of Doctor of Philosophy

20th October 2016

Abstract

MHC class I molecules on the cell surface present endogenous peptide to the adaptive immune system. The mature complex is trimeric, consisting of an α-chain encoding membrane-distal peptide binding and membrane-bound α_3 domains, an invariant β_2m protein, and a peptide. The α-chain is highly polymorphic, with the six alleles expressed by an individual capable of presenting all potential antigens due to promiscuous peptide binding. In addition to presenting antigen, MHC class I has a role in selecting high-affinity peptides during complex maturation within the endoplasmic reticulum. There is a large pool of available peptides, but only high-affinity interactions are capable of eliciting an immune response. Whilst the structure of the mature complex has been well documented by X-ray crystallography, the process of peptide optimisation remains unclear. It is apparent that a conformational change within the peptide binding domain is required to exchange low-affinity for high-affinity antigen. Molecular dynamics simulations have been employed to model the dynamics of MHC class I but have not yet been verified by direct observation of such motion.

This thesis investigates the dynamics of MHC class I using nuclear magnetic resonance (NMR) spectroscopy, focusing on the possible existence of alternative conformations within the peptide binding domain that permit peptide exchange.

Protein production methods have been established for reliable generation of mature MHC class I, consisting of the murine H-2D^b α-chain allele, human β_2m, and the high-affinity FAPGNYPAL peptide. Following assignment of backbone resonances, NMR analyses include residual dipolar coupling data revealing the potential for up to a 20° twist between the peptide binding and membrane-proximal domains. Heteronuclear relaxation data shows that MHC class I is inherently dynamic, experiencing multiple conformations in regions throughout the secondary structure. Mobility is largely located within the peptide binding domain and provides confirmation of predictions from molecular dynamics.

This thesis confirms that MHC class I is capable of forming an intermediate conformation during peptide binding, during which it is peptide-receptive.

Contents

List of Tables

List of Figures

List of Abbreviations

α-chain	MHC class I heavy chain
APC	antigen-presenting cell
ARTSY	amide RDCs by TROSY spectroscopy
ATP	adenosine triphosphate
β_2m	beta-2 microglobulin
CDR	complementarity determining region
CNX	calnexin
CRT	calreticulin
CTL	cytotoxic T lymphocyte
ER	endoplasmic reticulum
ERAP	ER aminopeptidase associated with antigen processing
HLA	human leukocyte antigen
HSQC	heteronuclear single quantum correlation
Ig	immunoglobulin
MDS	molecular dynamics simulations
MHC	major histocompatibility complex
NOE	nuclear Overhauser effect
NMR	nuclear magnetic resonance
PBD	peptide binding domain
PBG	peptide binding groove
PC	peptide carboxyl terminus
PLC	peptide loading complex
PN	peptide amino terminus
ppm	parts per million
RDC	residual dipolar coupling
RT	room temperature
TAP	transporter associated with antigen processing
TAPBPR	TAP binding protein related to TPN
TCR	T cell receptor
TPN	tapasin
TROSY	transverse relaxation optimised spectroscopy
τm	rotational correlation time

Declaration

I, Louise Bolton, declare that the thesis entitled 'MHC class I Protein Dynamics and Antigen Presenting Function' and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this university;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signed:

Date:

20th October 2016

Acknowledgements

I thank my supervisors Dr Joern Werner and Professor Tim Elliott for their support and guidance in all aspects of this work. Thanks also go to all members of both research groups for sharing their knowledge and time: particularly Stuart and Alistair for answering my endless questions, and also Andy, Halina, Kelly, Curtis, Denise, and Nasia amongst others.

I would also like to thank the University of Southampton for funding the research presented in this thesis.

I would also like to acknowledge the invaluable assistance offered during NMR data collection: Dr Geoff Kelly and Dr Alain Oregioni at the MRC Biomedical NMR Centre at the National Institute for Medical Research, Dr Sara Whittaker at the Henry Wellcome Building for Biomolecular NMR Spectroscopy at the University of Birmingham, and Professor Christina Redfield at the Department of Biochemistry, University of Oxford.

Finally, I would like to thank friends and family for their support.

Chapter 1

1 Introduction

MHC class I is a protein complex involved in the presentation of antigen to the immune system. As such, the topic of this thesis sits within the scientific discipline of immunology, more specifically within the fields of T-cell immunology and of antigen recognition. The foremost analytical technique applied in this thesis is Nuclear magnetic resonance spectroscopy (NMR), the theory of which will be discussed briefly in this introductory chapter, relevant to the experiments undertaken.

This chapter provides the background information relating to MHC class I, placing into context the significance of MHC class I dynamics, and discusses NMR spectroscopy, highlighting the ability of this technique to reveal dynamic information.

1.1 The immune system of vertebrates

Almost all forms of life possess some form of immune system, most commonly a complement system, to provide defence against pathogens. In vertebrates, immunity consists of physical barriers (skin, mucus membranes etc.) in addition to both innate and adaptive systems. Together, these systems must protect the organism from a wide variety of pathogens, from larger metazoan parasites (e.g. worms) to tiny viruses. Pathogens can multiply rapidly, requiring a fast response by the organism to ensure that they are prevented from spreading to other sites and causing further damage through the secretion of toxins.

In brief, the immune system consists of the lymphoid organs and tissues. The primary lymphoid organs are the bone marrow and thymus, where lymphocytes develop before circulating around the organism via lymphatic vessels, which connect body tissues with lymph nodes. Thus, lymphocytes constantly circulate through an organism in both blood and lymph. This section is designed to give an overview of the various forms of immunity; for more detail see textbooks such as Understanding Immunology[1] or Kuby Immunology[2].

1.1.1 Innate immunity

Upon infection by a pathogen, vertebrates are able to mount an immediate response via the innate immune system. The first requirement of this is that an infection must be recognised; receptors on tissue macrophages (bone-marrow derived white blood

cells) recognise many pathogen-specific molecules. For example, CD14 receptors bind to the lipopolysaccharide component of Gram-negative bacterial cell walls, and Toll-like receptor 3 (TLR3) recognises double-stranded RNA that only occurs during viral infection. One cellular response to recognition of a pathogen is phagocytosis by macrophages or neutrophils. Alternatively, pathogen recognition results in the production and secretion of factors such as cytokines, some of which can inhibit viral replication (interferons) whilst others activate other immune cells (tumour necrosis factor). This recruitment of additional cells is vital if there are insufficient numbers of tissue macrophages to phagocytose all pathogen. Inflammation largely activates macrophages, but will also result in the movement of mast cells and activation of the clotting, complement, and kinin systems.

In the majority of cases, the inflammatory response will succeed in removing the pathogen and thus the response will be confined to the affected tissue. However, if unsuccessful the continued response will increase the levels of cytokines in blood plasma, triggering a more systemic innate response; some cytokines affect the brain and induce fever and somnolence.

The innate immune system is not always successful. Pathogens may not express cell surface molecules that are significantly different to those of host cells and hence are not recognised by innate systems. Alternatively, many pathogens have evolved defence mechanisms; for example whilst non-virulent streptococcal bacteria can be eliminated, virulent strains express a polysaccharide coat layer on their surface that prevents recognition by phagocytes or opsonins[1].

1.1.2 Adaptive immunity

In contrast to the more generic innate immune system, adaptive immunity is highly specific to a particular pathogen. Adaptive immunity arises from B and T lymphocytes originally derived from bone marrow.

B lymphocytes produce antibodies, also referred to as immunoglobulins (Ig). Antibodies are Y-shaped glycoproteins that have four subunits; two heavy chains and two light chains, each further split into domains of approximately 110 amino acids. Each antibody has eight constant domains forming the ‘stem’ of the Y shape (the Fc region) along with four variable domains at the Y branch ends containing the antigen binding sites (the Fab region). Whilst the molecule that the antibody binds to is referred to as the antigen, it is worth noting that only a sub-section of that antigen (the epitope) between 8 and 22 amino acids in size in directly bound. Within the

antibody variable domains there are hyper-variable regions that control the exact conformation of the antigen binding sites and determine complementarity. Non-covalent forces drive protein antigen binding, with higher affinity binding arising from a closer match in the two protein surfaces. The presence of two binding surfaces per antibody means that the avidity of binding can be increased if both Fab regions each bind epitopes on the same molecule. Antibodies can be split into several classes by the sequence of the Fc region. They can be soluble or expressed as an integral membrane protein on a B cell. An individual B lymphocyte expresses IgM and IgD molecules with the same antigen specificity[2] such that it specifically recognises only a single antigen. Antigen recognition gives rise to proliferation, class switching to IgG, IgA, or IgE, and affinity maturation, and can result in a variety of responses including neutralisation of toxins, agglutination, phagocytosis (acting as an opsonin), and activation of complement. A significant difference to the innate immune system is that once activated, B lymphocytes remain in germinal centres as memory cells such that a reinfection with the same pathogen is met with a faster response and higher affinity antibody.

T lymphocytes also originate in bone marrow but mature in the thymus. There are two types of T cell, distinguished by the expression of mutually exclusive cell surface molecules; $CD4^+$ (helper) and $CD8^+$ (cytotoxic) T lymphocytes. Although both express the T-cell receptor (TCR) they are activated in different ways and have different roles in the immune response. The TCR is a membrane-anchored heterodimer complexed with an invariant CD3 chain. Each TCR chain contains both constant and highly variable immunoglobulin-like domains, with the latter containing complementarity-determining regions that mediate the relatively low-affinity interactions with antigen-presenting cells. TCR binding to antigen is degenerate; many TCRs recognise an individual peptide and many antigens are recognised by a single TCR[3]. The enormous TCR repertoire (possible combinations of its gene segments number 10^{15}) means that new T cells are theoretically capable of recognising any soluble antigen as well as self or non-self MHC molecules. T lymphocytes therefore undergo thymic education in which positive selection permits the survival only of those T cells with TCRs capable of self-MHC recognition, and negative selection eliminates T lymphocytes that react too strongly with self-MHC or self-antigen.

Dendritic cells transport pathogenic antigens to lymph nodes, where they activate $CD4^+$ T cells. Activation generally occurs up to five days after antigen presence, and can only occur in the presence of two signals; in addition to antigen binding the TCR,

the antigen-presenting cell (APC) must also bind to co-stimulatory molecules on the T cell such as CD28. The link between the APC and T lymphocyte is termed an immunological synapse, with a high affinity interaction occurring between T cell integrin LFA-1 and ICAM-1 on the APC only after successful antigen recognition. Following activation, $CD4^+$ cells proliferate due to cytokine stimulation; in response to antigen the T lymphocytes both release Il-2 and express Il-2 receptors. The number of antigen-specific $CD4^+$ T cells thus increases 10,000 to 100,000 fold in the days after antigen contact. The lymphocytes then differentiate into Th1 and Th2 cells, and migrate to the specific site of infection. The different effector $CD4^+$ cells release different combinations of cytokines (hence the name 'helper' T cells), with Th1 cells generating cytotoxic T cell and delayed-type hypersensitivity reactions and Th2 cells promoting antibody production from B lymphocytes and regulating responses from mast cells and eosinophils[2].

$CD8^+$ T lymphocytes are activated by cells presenting antigen, and respond by killing the APC (hence the name 'cytotoxic' T cells or CTL). Antigen must bind the TCR with sufficient affinity, and the reaction is stabilised by a second interaction between the CD8 co-receptor and the APC. Similar to $CD4^+$ cells, a second distinct signal is also required. Following a full recognition of a specific antigen, Il-2 receptors are expressed and activated either by cytokines released by Th cells or by a CD4-independent pathway. Activated $CD8^+$ T cells undergo clonal expansion and become effector cells; CTL are able to kill cells by granule exocytosis. They contain cytoplasmic granules that release proteins that lyse the target cells (perforin creates pores in the cell membrane whilst protease granzymes cleave various proteins). Alternatively they may use the Fas apoptosis pathway via expression of Fas-L and binding to Fas on the APC[1]. They also secrete numerous cytokines and chemokines to recruit and activate other effector immune cells such as macrophages and neutrophils. These CTL effector mechanisms have been comprehensively reviewed by Harty, Tvinnereim, and White[4]. Once the pathogen is cleared CTL levels decline to a memory level, able to mount a rapid and heightened response upon re-infection. Antigen may also be presented to natural killer cells, which complements the response by CTL by activating prior to the clonal expansion and differentiation of the $CD8^+$ T cells. They produce lymphokines to recruit additional cells to the site of inflammation, as well as cytokines and chemokines to aid cell destruction[2].

1.1.3 Presentation of antigen

Binding and recognition of antigen by T lymphocytes is dependent upon the presenting molecule; the major histocompatibility complex (MHC). A vital step in adaptive immunity is the pathway by which proteins (foreign or native) are processed and bound by this MHC molecule. Effective sampling of the cell environs is required to ensure that the immune system is alerted to infection rapidly to prevent unnecessary spread.

Antigenic proteins can arise from two sources; exogenous antigen is endocytosed into the cell whereas endogenous antigen is generated within the cell. A number of experiments (for example, those by Morrison and Braciale[5]) have shown that the adaptive immune system utilises different pathways to eliminate extracellular and intracellular antigens; respectively called the endocytic and cytosolic pathways.

In brief, a cell can take up antigen by endocytosis and/or phagocytosis, before it is degraded into peptides in the endocytic processing pathway. This is comprised of three compartments with increasing acidity: early endosomes of pH 6.0, late endosomes of pH 5.5, and lysosomes of pH 5.0. Each compartment contains various hydrolytic enzymes such that the end result is peptides of 13 to 18 amino acid residues in length. These antigenic peptides are then able to bind to MHC molecules for presentation on the surface of the cell[2]. In contrast, the cytosolic pathway for endogenous antigens begins with the ATP-dependent degradation of proteins by a cytoplasmic proteasome complex as part of regulated protein turnover within the cell. Many proteins targeted for this proteolysis are tagged with ubiquitin, and the resultant peptides of 7 to 12 residues in length are then transported to the lumen of the rough endoplasmic reticulum (ER) for loading onto MHC. In both pathways, peptides of inappropriate length for binding by MHC molecules are further degraded by exopeptidases[2].

There are two types of MHC molecule; class I and class II. The exogenous pathway loads peptide onto MHC class II whilst the endogenous pathway loads onto MHC class I. Although both are membrane glycoproteins that function as antigen-presenting molecules able to form stable complexes with peptide ligands, there are structural and functional differences between them. MHC class I is formed of a heavy glycoprotein chain and light protein chain, whereas MHC class II has two non-identical glycoprotein chains. Both possess a peptide binding region with significant polymorphism to permit binding to a broad range of peptides. This is vital because

any one (human) individual expresses only six MHC class I and twelve MHC class II allelic variants; these are able to present all potential antigenic peptides due to promiscuous peptide binding; polymorphism in the peptide binding region ensures variation in peptide binding between alleles[6], but it is not the case that each allele binds to a specific subset of peptides – the repertoires overlap[7]. Whilst all nucleated cells are able to express MHC class I and present antigen to $CD8^+$ T lymphocytes, only specialised antigen-presenting cells express MHC class II and are able to present antigen to $CD4^+$ T cells. Professional antigen-presenting cells include dendritic cells, macrophages, and B cells, and non-professional cells such as fibroblasts and vascular endothelial cells can be induced to express MHC class II[1].

It is important to note that the endocytic and cytosolic pathways are not wholly separate; cross-presentation whereby exogenous antigen is presented to $CD8^+$ T lymphocytes associated with MHC class I (requiring peptides to be passed from the exogenous to the endogenous pathway) can be observed in some antigen-presenting cells[8]. Both types of MHC molecule pass through the rough ER during maturation and although it is here that MHC class I is loaded with peptide, MHC class II instead associates with an invariant chain and proceeds to the late endocytic compartments where this invariant chain is displaced by peptide[2]. Since the pathways diverge post-rough ER, MHC class I should not encounter any exogenous antigen. As such, the pathways by which cross-presentation occurs are not fully understood[8]. An example is that in dendritic cells, exogenous protein in phagosomes can be retro-translocated and thus processed into the rough ER as peptide for loading onto MHC class I. The advantage of this is that dendritic cells are therefore able to capture viral antigens prior to the general spread of the virus, and initiate a primary response by cross-priming $CD8^+$ T cells[9].

1.2 Processing and presentation of endogenous antigens

Endogenous peptide binds to MHC class I following processing in the cytoplasm by a proteasome complex, as outlined above. However, this results in peptides that may be too long for binding to MHC class I and therefore there is additional processing of peptide to ensure high affinity binding.

1.2.1 Production of antigenic peptide

The peptides bound by MHC class I are obtained from endogenously expressed protein, but are unlikely to originate from the natural turnover of full-length protein[10]. Instead, the vast majority of class I ligands are obtained from defective ribosomal products (misfolded, truncated, or non-native protein, or the products of errors in translation or splicing, or protein that is incorrectly post-translationally modified)[11,12]. It is estimated that of the $4x10^6$ proteins synthesised per minute almost a third are proteasomally-degraded within two hours, a quarter within ten minutes[13]. The presentation of epitopes on the cell surface therefore reflects the current set of translational products, rather than already-expressed proteins, thus enabling a rapid targeting of virus-infected cells by the immune system.

The majority of endogenous proteins are degraded by the 20S proteasome which consists of 14 related subunits in a barrel-shaped structure with two outer α-rings and two inner β-rings with protease activity[14]. Proteins targeted for degradation by ubiquitin tagging are recognised by a 19S multi-subunit complex either side of the 20S core which exhibits deubiquitinase and unfoldase activity (thus making the process ATP-dependent)[9]. When cells are exposed to IFN-γ during viral infection the 20S proteolytic β1, β2, and β5 subunits are replaced by homologues to form the immunoproteasome with alternate cleavage motifs[14,15]. Immunoproteasomes are more rapidly assembled than the constitutive version but are less stable and their function is not fully understood[15]. Although variation in the peptide repertoire is advantageous during viral infection, it is likely that immunoproteasomes have additional functions not associated with antigen processing[16]. Proteasomal degradation results in a pool of peptide in the cytosol, only a fraction of which will be suitable for MHC class I binding[17].

1.2.2 Presentation of antigenic peptide

Peptides produced by the proteasome are transported by TAP (transporter associated with antigen processing) from the cytosol to the rough ER, where they are bound by maturing MHC class I molecules. The loading of suitable peptide onto MHC class I is facilitated and regulated by the peptide loading complex (PLC), which consists of calreticulin (CRT), TAP, tapasin (TPN), and ERp57 (Figure 1.1). This process is vital because MHC class I molecules on the cell surface must be sufficiently stable to invoke an immune response.

PLC formation is not fully understood, with two models proposed. The first involves TAP-TPN complexes acting as receptors for peptide-free MHC class I dimers, whilst the second involves almost-complete PLC formation (dimer, TPN, CRT, and ERp57, as observed in TAP-deficient cells) prior to association with TAP[18]. However, each member of the PLC is vital to obtain optimally-loaded MHC class I[19]; for example CRT retains MHC class I dimers[20,21] and CRT-deficient cells have suboptimally loaded complexes[20]. In addition, members of the PLC require the presence of one another for peak activity; CRT[20] and ERp57[23] are required for optimal TPN activity, whilst TPN is required for both efficient formation of TAP[24] and efficient peptide binding to TAP[25].

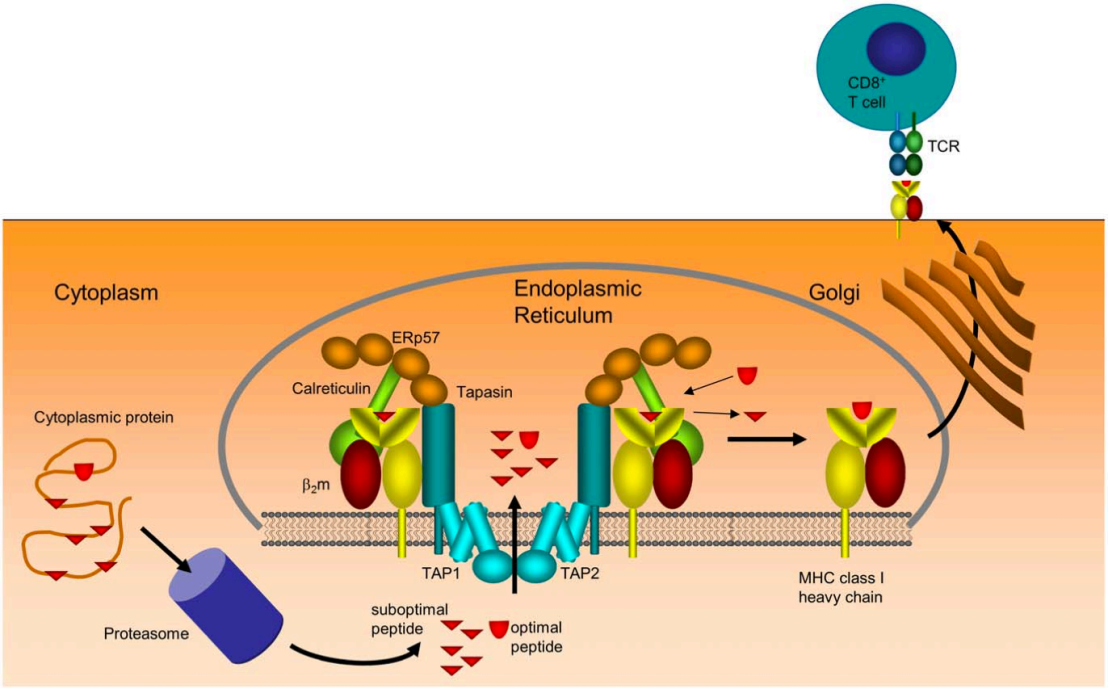


Figure 1.1 – The peptide loading complex
Endogenous proteins are degraded by the proteasome and surviving peptide fragments are translocated into the ER by the TAP heterodimer, co-located with TPN. The PLC also consists of CRT, ERp57, and TPN. The latter binds to MHC class I (yellow and dark red heterodimer) and mediates the exchange of suboptimal peptide for optimal peptide. Mature MHC class I is exported to the cell surface via the Golgi, for interaction with $CD8^+$ T cells. Figure taken from Dalchau et al, 2011[22].

TAP is a heterodimer of TAP1 and TAP2, members of the ABC transporter family containing a C-terminal ATP-binding domain and an N-terminal membrane-anchoring domain. TAP binds a broad range of peptides that are between 7 and 12 residues in length[26], with both subunits required for efficient binding[27,28]. Initial binding of peptide is ATP-independent[27], whilst translocation into the ER is ATP-dependent[29]. The TAP heterodimer is co-located with the PLC via an interaction between the TAP2 N-terminal and TPN[30], and associates with 4 TPN and 4 MHC molecules (each TPN contacts one MHC)[31]. This interaction is not required for MHC class I expression on the surface of cells, but TAP variants lacking this binding region form PLCs that are defective in quality control[30]. There is also evidence that TAP may influence the conformation of MHC, preventing the export of sub-optimally loaded complexes[32].

Peptides that are N-terminal extended beyond the requirements of MHC class I can be shortened by ERAP (ER aminopeptidase associated with antigen processing)[33], which can act on peptides in solution or those bound to MHC. There are opposing theories as to whether peptide trimming occurs via a 'molecular ruler' approach in which contact to the C-terminal of MHC-bound peptide is required[34], or whether it can occur without C-terminal access and is instead dependent on the relative affinity of peptide to MHC class I (a peptide with a higher binding affinity to ERAP is trimmed whilst for a peptide with preference for MHC over ERAP no trimming occurs)[35]. ERAP knock-outs have been shown to decrease surface expression of MHC class I and result in complexes that are sub-optimally loaded; ERAP is vital to shape the peptide repertoire[36].

ERp57 is an oxidoreductase known to be important in MHC assembly and PLC function[23] but which does not require interactions with CRT to promote assembly[37]. Originally thought to act on a disulphide bond within MHC class I unless TPN is present to provide a preferred substrate[38], it is now thought to be a purely structural element that enhances the recruitment and stability of empty MHC associated with CRT[39], although there is evidence of disulphide linkage with TPN[40]. CRT is a multifunctional protein that in the PLC acts as a chaperone of the immature MHC class I molecule by binding to a terminal glucose residue, an interaction that can be disrupted by ATP[41].

Once MHC class I is correctly folded and loaded with high affinity peptide, TPN dissociates[42] and exposes a monoglucosylated glycan on the MHC class I heavy chain to glucosidase II[43]. Once this glycan is removed, MHC complexes cannot associate with the PLC since recruitment is dependent on this $GlcMan_9GlcNAc_2$-CRT

interaction[44], and therefore inhibition of glucosidase II prolongs association with the PLC[45]. MHC class I trimers are transported to the cell surface via COPI vesicles[46] and may cluster together during transport to create an improved target for $CD8^+$ T-cells[47], and potentially initiating the immune response described in Section 1.1.2.

TAPBPR is a polymorphic TPN-related protein that may represent a newly discovered chaperone in the above process. It is not an integral part of the PLC and unlike TPN it does not possess a TAP-binding or ER-retention motif or bind ERp57. Its role in MHC class I maturation remains unclear; it may represent a TPN-independent peptide loading mechanism, or function to prolong PLC association to ensure high affinity peptide binding, although overexpression of TAPBPR results in a decrease in MHC class I expression on the cell surface[48].

1.3 Peptide optimisation and immunodominance

Peptides presented at the cell surface by MHC class I that result in a $CD8^+$ T cell response are referred to as immunodominant; CTL rely on the ability of MHC to selectively present high-affinity, long half-life, optimal peptides from the millions available within the ER. It is thought that MHC binding selectivity gives rise to 90% of immunodominance, with the remaining 10% related to the relationship between MHC and the T cell itself[49]. This process of peptide selection is not fully comprehended despite its importance in the immune response. Why and how does MHC class I choose high-affinity peptide, and what evidence is there for such a process?

Peptide-MHC class I complexes must be of sufficiently high affinity to ensure that the complex is stable for the hours or days required to effectively elicit an immune response[50]. MHC class I on the cell surface is recognised by the TCR, in which the complementarity-determining regions of the variable immunoglobulin-like domains mediate the relatively low-affinity (1-100 μM) interaction with MHC class I (comprehensively reviewed by Rudolph, Stanfield, and Wilson[51] and discussed further in Section 1.4.2.2). TCRs are extremely sensitive to antigen structure[52] (although significant exceptions have been observed[53]) and therefore despite degenerate TCR binding to antigen[3], minute changes to the peptide presented by MHC class I result in considerable, and perhaps vital, changes in the response of the immune system. This is demonstrated by the fact that the immune response is directed toward activated $CD8^+$ T cells with the highest stability interactions with MHC class I[54].

1.3.1 The peptide optimisation concept

The existence of immunodominance implies that MHC class I is capable of editing its peptide repertoire via an active molecular process. A foreign pathogen contains numerous peptides that could be presented by MHC yet only a tiny percentage of these produce a measurable immune response; although 0.5% of potential determinants are capable of immunogenicity (threshold affinity binding (K_d) is at greater than 500 nM), restrictions in the $CD8^+$ T cell repertoire and inefficient antigen processing mean that only 0.05% may achieve immunodominant status[55].

The first optimisation decision point during peptide loading is when the pool of peptide in the ER is acted on by the aminopeptidase ERAP; TAP is able to transport a

range of peptide lengths and ERAP cleaves those that are between 9 and 16 residues in length to make them suitable for binding to MHC class I[34]. This editing is required for optimal presentation and a lack of ERAP decreases surface presentation by 20%[36].

The final decision point is a mechanism in which MHC class I loaded with sub-optimal peptide is retained[42] or retrieved from the cis-Golgi by calreticulin[56,57] and returned to the ER for further processing[57], at which point a conformational change must occur to allow peptide exchange. Misfolded or sub-optimally loaded MHC can also be recovered by UDP-glucose:glycoprotein glucosyltransferase. This enzyme is present in the ER and pre-Golgi compartments[58], and targets MHC back to the PLC for peptide exchange[43,56]. However, peptide optimisation has been observed in post-ER compartments when the PLC is defective[59], and MHC, TPN, and peptide have been shown to continuously cycle between the ER and the Golgi[57,60].

The major quality control checkpoint in the maturation pathway is the TPN-mediated release of MHC class I from the PLC[61].

1.3.2 Tapasin

TPN is a 48 kDa glycoprotein that bridges TAP and the nascent MHC class I molecule, and thus is an integral part of the peptide loading complex. Unlike the other, more generic, components of the PLC TPN is specific to this assembly pathway[62]. Investigation into the role of TPN revealed that it has a large effect on the peptide repertoire[63,64]; it is capable of retaining non-loaded or suboptimally loaded MHC class I in the ER, remaining bound until optimal peptide is present[42]. In the absence of TPN, MHC class I molecules are notably thermolabile and are loaded with suboptimal ligands[65].

The presentation of optimal peptides has been shown to be the remit of TPN and not other members of the PLC through a series of experiments by Howarth *et al*[66]. In the absence of TPN there is decreased interaction between MHC class I and ERp57, TAP, and CRT, and therefore investigation into TPN-deficient cells alone is not sufficient to prove the role of TPN, and the individual effects of these other members must be probed. The peptide binding hierarchy is significantly different in wild-type versus TPN-deficient cells, but when the PLC lacks ERp57 there is no difference, and although a lack of CRT causes a decrease in surface expression of MHC class I this is due to an increase in the export of empty MHC complexes rather than a change in hierarchy[66].

TPN stabilises TAP and bridges to MHC molecules, and the resultant increase in TAP increases the presentation rate of all peptides. However, TPN does not affect the peptide selectivity of TAP[67] and hence the effect of TPN on peptide repertoire is distinct from its interaction with TAP[68]. Another role of TPN is the retention of suboptimally-loaded MHC class I molecules[69,70] but since not all MHC class I alleles show this effect[71] it cannot be the major route of optimisation. Instead, TPN acts as a catalyst for peptide loading such that when it is not present the loading is inefficient even when peptide levels are high[70] and when present high affinity peptides are bound. TPN does not enhance loading of all peptides equally however; immunodominance arises because it is selective for optimal peptides[55]. TPN does this by accelerating the dissociation of low-affinity peptides[72,73], therefore increasing the presentation of peptides with longer half-lives and hence greater stability[66] for improved chance of stimulating T cell activation.

A question that has not yet been answered is how TPN detects an optimal peptide. This is thought to be related to a conformational change within the MHC class I molecule itself, in which TPN may stabilise a conformation from which the dissociation of suboptimal peptides is increased[74-76]. Possible mechanisms for this recognition are discussed in Section 1.4.3.

1.3.3 The kinetics of peptide loading onto MHC class I

A complementary approach examining how peptides are loaded onto MHC class I molecules is to examine the kinetics of the process. Peptide optimisation in the ER is limited to tens of minutes[61], significantly less than the half-lives of many MHC-peptide complexes[77]. The existence of immunodominance means that peptide binding is not a random selection from the available pool of epitopes, but a directed process. Within the ER an unknown agent must be present to permit time to 'run quicker' to enable an individual MHC class I molecule to reject one peptide for another; there is evidence that peptide dissociation rates are higher inside a cell compared to the cell surface[78]. TPN is the best candidate for this role because it has been shown both to increase the dissociation rate of peptide[42,73] and to enhance the ability of MHC complexes to bind new epitopes[78].

It is possible to express the binding of peptide to MHC class I as a reversible binding process (Equation 1.1) with the dissociation constant for peptide calculated by $k_{off} \div k_{on}$. When these variables (dissociation constant, k_{off}, and k_{on}) are independently

measured they do not match[74] and therefore peptide binding must include additional steps such as a two-step mechanism (Equation 1.2)[73].

$$MHC + peptide \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} MHC - peptide$$

Equation 1.1

where k_{on} is the rate of peptide binding, k_{off} the rate of peptide unbinding, *MHC* the HC/β_2m dimer and *MHC-peptide* the trimeric MHC class I complex

$$MHC + peptide \underset{k_{off}^1}{\overset{k_{on}^1}{\rightleftharpoons}} MHC - peptide^* \underset{k_{off}^2}{\overset{k_{on}^2}{\rightleftharpoons}} MHC - peptide$$

Equation 1.2

where *MHC-peptide** denotes an intermediate

Extensive molecular dynamics (MD) simulations focusing on the kinetics have been carried out by Dalchau *et al*, resulting in the formation of a complex peptide loading scheme confirming the presence of an intermediate step in the process[22]. However, there is no direct structural evidence for this intermediate structure; crystal structures for complexes containing immunodominant or non-immunodominant peptides are essentially identical[79], yet we know the latter is exchanged and the former is not. From the research into this intermediate, it is apparent that a conformational change likely occurs within the nascent MHC class I molecule, catalysed by the presence of TPN. This will be further discussed in Section 1.4.3 following details of MHC class I structure, but it is clear that the peptide binding process, and thus immunodominance, is not yet fully understood.

1.4 MHC class I

Major histocompatibility complex class I is expressed on the surface of all nucleated cells and is responsible for the presentation of endogenous antigens to circulating $CD8^+$ T lymphocytes[80], ultimately resulting in the activation of the adaptive immune system when a non-self epitope is recognised. Understanding the structure of MHC class I and how these molecules bind and optimise peptide is vital to understanding how the immune system targets and eliminates disease.

1.4.1 Discovery

Initial evidence for the major histocompatibility complex was found in 1936 by Peter Gorer[81], with the first MHC locus identified in 1951 via tissue transplantation compatibility studies in mice[82] that earned George Snell, Baruj Benacerraf, and Jean Dausset the 1980 Nobel Prize in Physiology or Medicine[83]. A complete genomic sequence for the human MHC was reported in 1999[84].

Following discovery of the major histocompatibility complex region on chromosome 6, work began to characterise the encoded proteins. MHC class I was discovered to consist of two polypeptide chains; the variant heavy chain and invariant β_2m[85,86]. Human chromosome 6p21 encodes the former, with classical class I molecules present on the A, B, and C loci of the 7.6 Mb human leukocyte antigen (HLA) complex[84]. These loci are highly polymorphic, with over 7000 alleles identified[87]. Alleles are co-dominant, such that each nucleated cell expresses both maternal and paternal variants, to produce six class I molecules. There is a similar situation in mice; classical MHC heavy chains are expressed on chromosome 17 in the K, D, and L regions of the H-2 complex, with a heterozygous mouse expressing four variations of class I MHC[82]. To date, there are no major structural or functional differences between murine and HLA molecules, making any observations applicable to both.

After sequence determination[88], the heavy chain was found to consist of three domains, each of approximately 90 amino acids[89]. The α_1 and α_2 domains have weak homology to each other[89], but unlike the α_3 domain (and β_2m) do not contain an immunoglobulin fold[90]. The α_3 domain is followed by a membrane-anchoring transmembrane α-helix, which can be cleaved via papain digestion to leave soluble protein[91]. This method permitted the purification of HLA-A2, the first class I α-chain identified[92], and the acquisition of a 3.5 Å resolution crystallographic structure[93] that was later refined to 2.6 Å[94].

1.4.2 Structure

The 32 kDa MHC class I α-chain associates non-covalently with β_2m, with these four domains arranged as shown in Figure 1.2A. The immunoglobulin folds of the membrane-proximal α_3 and β_2m are not arranged in the fashion of immunoglobulin constant domains, but instead run approximately perpendicular to each other[94]. This interaction is largely stabilised by the A and B strands of the immunoglobulin fold as well as the C-terminus of β_2m, whilst there are several further contacts between β_2m and the α_1 and α_2 domains mainly involving the B-C and D-E loops (Table 1.1). Despite these interactions, it is possible to exchange MHC-bound and serum β_2m[95]. Within the immunoglobulin fold of the α_3 domain there is a disulphide bond between Cys203 and Cys259 on strands B and F respectively, with a second α-chain disulphide link between Cys101 and Cys164 on the β-sheet and α-helix respectively of the α_2 domain[94] (Figure 1.2A and B).

The membrane-distal α_1 and α_2 domains together form the peptide binding domain (PBD). This region contains a mixture of secondary structure; both α_1 and α_2 donate four β-strands to an eight-stranded antiparallel β-sheet to form the base of the peptide binding platform and each containing a C-terminal α-helix to form the sides of the peptide binding groove (PBG) (Figure 1.2B), a highly conserved structure (found in e.g. human; HLA-A2[94], HLA-Aw68[96], HLA-B27[97], murine; H-2D^b [98], H-2K^b [99]). The ends of the PBG are closed off by a number of highly conserved predominantly aromatic residues (Figure 1.2C), in addition to a number of less-conserved residues. This closure of the ends restricts the distance between peptide antigen termini to around 25 Å (between 8 and 10 residues[100]) and tethers the peptide termini via extensive hydrogen bonding[101]. In comparison to the conserved residues at the ends of the PBG, residues lining the remainder of the cleft are highly polymorphic; indeed the most polymorphic of the α-chain[102]. Arrangement of these side chains gives rise to a number of binding pockets[96]. Studies of HLA-A2[94], HLA-B27[100], and H-2K^b [103] have identified six such pockets (commonly labelled A through F). The presence of binding pockets and a closed cleft results in allele-specific peptide anchor residues, which are required to possess side chains complementary to the binding surface. Commonly P2 (peptide position 2) or P3 is hydrophobic or polar uncharged, and P9 is hydrophobic and commonly Ile or Leu[97,99]. Young, Nathenson, and Sacchettini comprehensively review both binding pockets and anchor residues[105]. Research in this thesis focuses on the murine H-2D^b α-chain, whose PBG and binding pockets are shown in Figure 1.2D. The cleft contains a hydrophobic ridge that disrupts the common P2/3 anchor

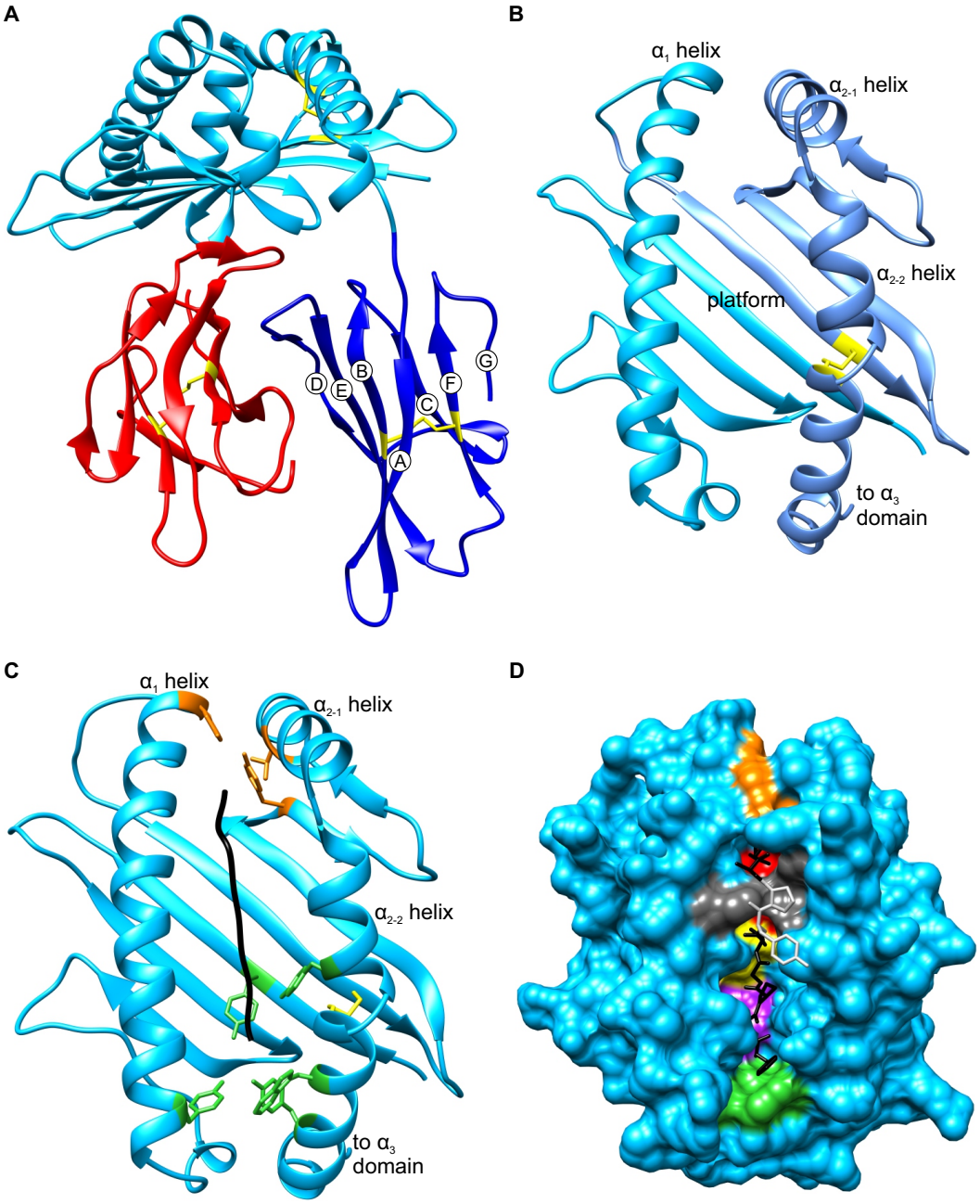


Figure 1.2 – Structure of MHC class I.
(A) Heavy chain domain arrangement (PDB: 3HLA[94]). The α-chain can be split into three domains; α_1 (residues 1-90), α_2 (residues 91-181), together the PBD (light blue), and the α_3 domain (dark blue), with the fourth MHC class I domain of β_2m (red). Disulphide bonds (yellow) are present in the α_2 (tethering the α_2 helix to the floor of the peptide binding groove) and α_3 domains (connecting strands B and F of the immunoglobulin fold, strands of which are labelled by letter, A-G). **(B)** The peptide binding domain (PDB: 3HLA[94]). The α_1 (light blue) and α_2 (mid-blue) domains contribute equally to form a platform of antiparallel β-sheet, and each also contains an α-helix. The α_2 helix has a kinked structure at residue 151 and can be split into the $\alpha_{2\text{-}1}$ and $\alpha_{2\text{-}2}$ helices. **(C)** The ends of the peptide binding cleft are closed (PDB: 1CE6[98]). In H-2D^b Tyr7, Tyr59, Tyr159, Trp167, and Tyr171 (green) enclose the peptide N-terminus (PN; all peptide residues denoted P*n*). Tyr84, Tyr123, and Thr143 (orange) enclose PC. **(D)** The peptide binding groove of H-2D^b (PDB: 1CE6[98]). On the surface of the PBD (light blue) residues are highlighted for enclosing PN (green), B pocket for P2 binding (pink; 9, 24, 45, 67), C pocket for P5 binding (yellow; 9, 97), hydrophobic ridge (dark grey; 73, 147, 156), F pocket for PC binding (red; 95, 116, 123), and for enclosing PC (orange). The peptide backbone is shown (black), with the arching of P6 and P7 highlighted (light grey).

Table 1.1 – Contacts between β_2m and MHC class I heavy chain.
Adapted from Tysoe-Calnon[104]; HLA-A2.1 and HLA-Aw68.1.

β_2m Contact	Side/Main Chain	MHC class I Contact	Side/Main Chain	Type of Interaction
Ile1 N-terminus	MC	Asp119 C=O	MC	H-bond
Lys6 NH_3^+	SC	Glu232 CO_2^-	SC	Ionic
Gln8 NH_2	SC	Glu232 C=O	MC	H-bond
Gln8 C=O	SC	Arg234 NH_2	SC	H-bond
Tyr10 OH	SC	Pro235 C=O	MC	H-bond
Ser11 CO	MC	Gln242 NH_2	SC	H-bond
Arg12 NH_2	SC	Gly237 C=O	MC	H-bond
Arg12 CO	MC	Gln242 NH_2	SC	H-bond
Asn24 NH_2	SC	Ala236 C=O	MC	H-bond
Tyr26 CH	SC	Pro235 CH	SC	Hydrophobic
His31 NH	SC	Gln96 C=O	SC	H-bond
Asp53 CO_2^-	SC	Gln32 NH_2	SC	H-bond
Asp53 CO_2^-	SC	Arg35 NH_2^+	SC	H-bond
Phe56 CH	SC	Phe8 CH	SC	Hydrophobic
Trp60 C=O	MC	Gln96 NH_2	SC	H-bond
Trp60 NH	SC	Asp122 CO_2	SC	H-bond
Phe62 (ring)	SC	Gln96 NH_2	SC	H-bond
Tyr63 OH	SC	Tyr27 OH	SC	H-bond
Asp98 C=O	MC	Arg202 NH	SC	H-bond
Asp98 C=O	MC	Arg202 NH_2	SC	H-bond
Met99 C-terminus	MC	Arg234 NH_2^+	SC	Ionic

and creates a P5 requirement for Asn by causing the peptide to arch, as well as creating a C-terminal bulge and a P8 anchor constraint of Met, Ile, or Leu. The ridge consists of the conserved Trp147 coupled with Trp73 and Tyr156[106], a combination found only in H-2L^d, L^q, D^q, and D^b alleles[107].

Despite a limit of six distinct MHC class I alleles per individual, all prospective antigenic peptides can be presented to the TCR due to promiscuous peptide binding; the polymorphic binding pockets in the PBG give rise to a huge range of potential peptide binding sites to accommodate the various residue side chains[6], with different MHC α-chain alleles having overlapping binding repertoires to ensure coverage of potential epitopes[7]. The shape of peptide is not required to exactly mirror that of the binding pockets; bound water molecules stabilise the interaction through formation of hydrogen bonding networks and by filling empty spaces in the PBG[108,109] thus ensuring high affinity peptides binding[77].

A clear example of PBG polymorphism is found in murine H-2D^b, where there is a PBD ridge that forces the peptide to arch, introduces conformational variability around P6, and improves TCR discrimination between antigens and between self and non-self antigens[107]. In addition, the arch-induced carboxyl-terminal bulge increases the number of contacts between peptide and the TCR[106]. Interestingly, HLA alleles maintain a flat binding groove[107] suggesting they evolved an alternative strategy to improve TCR discrimination.

1.4.2.1 Maturation and degradation of MHC class I

MHC maturation occurs in the endoplasmic reticulum (ER), with the assistance of numerous chaperones. β_2m and α-chain are both co-translationally inserted into the ER. During translocation the α-chain is glycosylated on Asn86 with the glycan $Glc_3Man_9GlcNAc_2$, before the first two glucose residues are removed by glucosidase I and II respectively to permit binding of calnexin (CNX)[110]. The function of glycosylation is purely related to chaperoning and has no structural effects. CNX is a 67 kDa membrane-bound lectin-like protein that assists N-linked glycoprotein folding, as well as retaining unfolded protein in the ER. For the α-chain CNX chaperones the folding of the α_3 domain[111], before β_2m binding promotes the folding of the α_1 and α_2 domains[112] (also promoted by the presence of peptide[40]). CNX chaperones the formation of the two α-chain disulphide bonds[113], which along with the presence of β_2m alter the preference of the heterodimer for CNX, inducing a swap for calreticulin (CRT)[113,114]. The monoglucosylated glycan assists in the recruitment of the β_2m/α-chain dimer into the peptide loading complex[43,115] (Section 1.2.2, Figure 1.1).

MHC class I complexes on the cell surface are eventually endocytosed for proteasome-mediated degradation, with β_2m released into the circulation[95]. Additional degradation pathways are available during complex maturation; for example MHC class I that is misfolded, mistranslated, or that fails to acquire appropriate peptide within the ER, can be flipped into the cytosol and tagged with ubiquitin for targeting by the proteasome[116].

1.4.2.2 Binding to T cell receptor

The side chains of the MHC class I peptide anchor residues face into the PBG whilst the remaining residues face outward towards the T cell for TCR recognition of antigen. The TCR is a membrane-anchored αβ heterodimer (Figure 1.3A) complexed with an invariant CD3 chain. Each TCR chain consists of a short cytoplasmic tail,

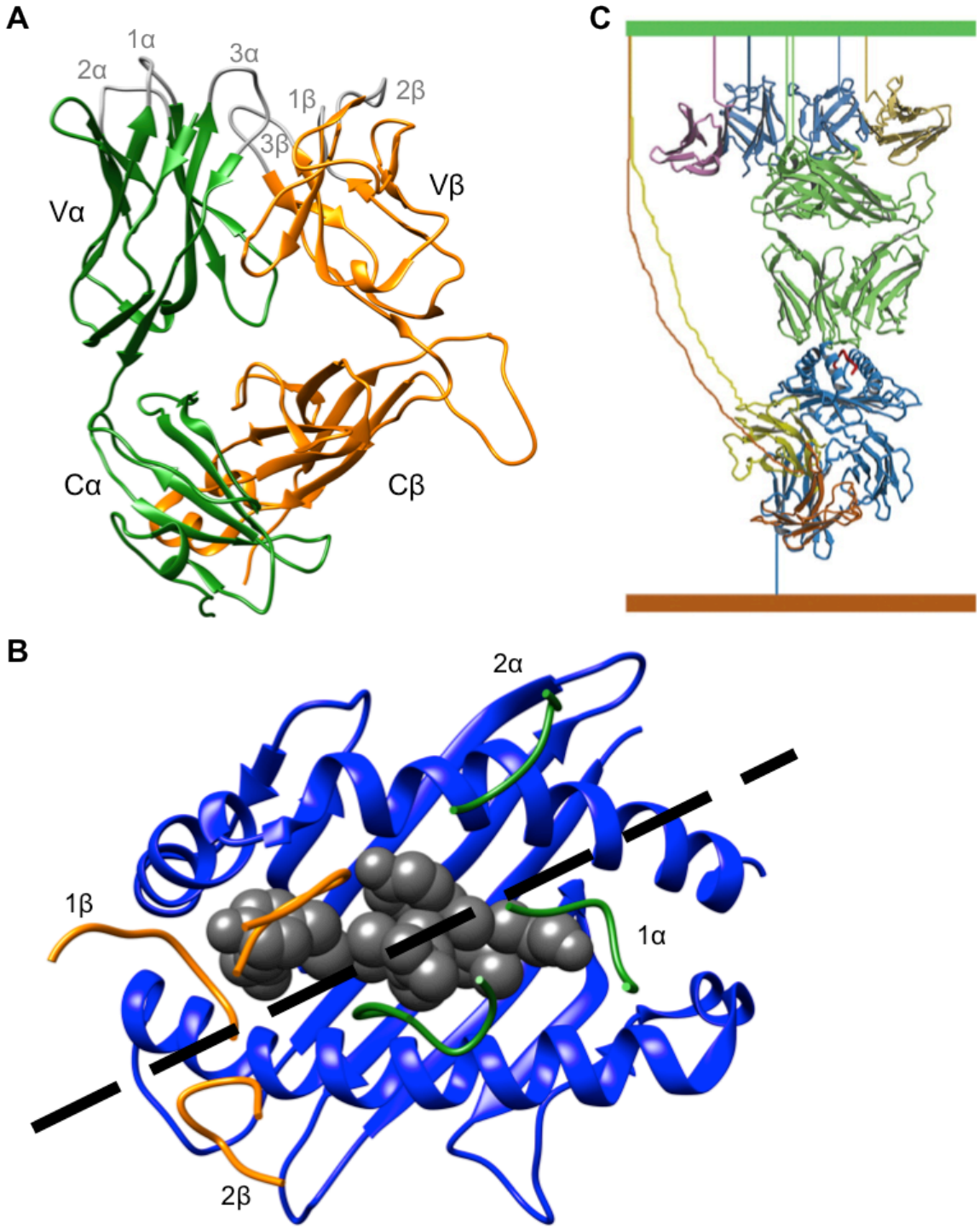


Figure 1.3 – Interaction between MHC class I and TCR complex.
(A) Structure of αβ TCR. The heterodimer chains mirror each other in secondary structure, with membrane-proximal constant (Cα green/Cβ orange) and membrane distal variable (Vα/Vβ) domains. Each variable domain contains three CDR loops, labelled 1-3 in grey. **(B)** Interaction surface between MHC class I and TCR. The PBD of MHC class I is shown in blue, with peptide in spherical representation in dark grey. Location of CDR loops from αβ chains are shown. Approximate axis of interaction is shown as superimposed black dashed line. PDB: 1G6R[120]. **(C)** Hypothetical TCR/MHC/CD3εδ/CD3εγ/CD8 complex. MHC/CD8 (1AKJ) and TCR/MHC (1QRN) are superimposed to show TCR (green), MHC (dark blue), peptide (red) and CD8αα (yellow and orange). The CD3εδ (1XIW, pink and blue) and CD3εγ (1SY6, gold and blue) are shown placed in approximate binding positions. Lines are drawn to depict tethers connecting the different subunits to the TCR cell membrane (top, green) or the APC membrane (brown, bottom). Figure taken from Rudolph, Stanfield, and Wilson[51].

transmembrane domain, connecting region with an inter-chain disulphide bond, membrane-proximal constant domain, and membrane-distal variable domain. Both constant and variable domains are Ig-like, and the latter contains three hypervariable complementarity-determining regions (CDR) to form the antigen binding site[2] and mediate the interaction with MHC class I[51]. The variability of the CDRs means that TCR binding to antigen is degenerate; many TCR recognise an individual peptide and many antigens are recognised by a single TCR[3,117].

Initial contact between TCR and APC occurs by contact between CDRs and MHC class I, with secondary 'scanning' of the presented peptide determining the duration of binding[118]. The major interactions between TCR and peptide occur via CDR3[51,119,120] although contacts with peptide termini from CDR1 are also significant[121], particularly CDR1β[119] (Figure 1.3B). When bound to MHC class I, the TCR is orientated diagonally relative to the long axis of the peptide binding groove, with the Vα domain above the peptide amino-terminal and the Vβ domain above the carboxyl terminal. When MHC class I complexes are compared it is clear that the TCR engages with MHC class I at a number of possible angles, and it is thought that the initial TCR approach is diagonal and then modified to maximise peptide contact[51]. This agrees with findings that the peptide-induced conformational change within the TCR does not correlate to the strength of induced T cell signals[52], suggesting that binding geometry instead is significant.

The stability of TCR-peptide binding may be partly governed by a dynamic rather than static interaction; although the majority of TCR-MHC interactions have been studied using X-ray crystallography, NMR studies suggest that there is motion in the CDR3α loop that mirrors motion in the nearby peptide. Matching structural fluctuations is linked to improved stability, and may form part of the TCR 'scanning' that occurs secondary to MHC contact[122], thus providing a reason for TCR binding orientation to alter. This is further evidenced by a study in which a modified peptide showed increased affinity for MHC class I but also an increased flexibility that reduced the subsequent immune response[123], perhaps due to a mis-match in structural fluctuations.

As a co-receptor to the TCR, $CD8^+$ T lymphocytes express the CD8 glycoprotein which binds to the α_3 domain of HC[124] in addition to potential contacts in the α_2 and β_2m domains[51,124]. CD8 is a disulphide-linked αβ heterodimer or αα homodimer with chains of 30-38 kDa consisting of an Ig-like domain, a stalk region, a hydrophobic transmembrane domain, and a cytoplasmic tail[2] (Figure 1.3C). The interaction

between the TCR and MHC class I complex is enhanced by various co-receptors. The affinity for TCR-MHC binding is relatively weak at 1-100 μM[51] but is strengthened by various cell-adhesion molecules (e.g. CD2, CD28, CD45R) on the T lymphocyte independently binding to other ligands on the APC. If a T cell is activated through contact with an antigenic peptide there is a transient increase in the membrane expression of these cell-adhesion molecules which results in closer contact for more efficient transfer of cytotoxic molecules[2].

Yet another co-receptor is required for actual recognition of the antigen-MHC complex. This is CD3, consisting of ζζ homodimers, εδ or εγ heterodimers, each comprising an Ig-like domain, transmembrane region, and long cytoplasmic tail[51] (Figure 1.3C). CD3 is required for the TCR to be expressed on the cell membrane and its role is in signal transduction following recognition of an antigenic peptide; following antigen engagement by the TCR there is a conformational change in CD3ε resulting in the immunoreceptor tyrosine-based activation motifs of the cytoplasmic tails of the CD3 chains interacting with tyrosine kinases within the cell as part of the signalling pathway for clonal expansion into CTLs[51].

1.4.2.3 Pathophysiology

The densely concentrated gene region of the MHC means linkage disequilibrium makes it difficult to identify a single allele or set of alleles that independently contribute to disease, but MHC class I is associated with transplant rejection[125], HIV-AIDS disease progression[126], as well as several auto-immune diseases[127]. Understanding the mechanism of epitope presentation to the immune system may provide new avenues to explore treatment.

The progression from HIV infection to AIDS varies between 1 and 35 years, with several HLA-B alleles or groups of alleles linked to a slow development. This slow development can be due to peptide binding motifs; B*27 requires a PN anchor of Arg and a P2 of Arg/Lys, with epitopes restricted in this way thought to arise from areas of protein structural significance which are resistant to mutation. Individual residues can also confer HIV-resistance; position 9 (forming the B pocket) controls the peptide repertoire of B*4201 (Tyr) and B*4202 (His) – the former is able to present multiple immunodominant GAG epitopes whilst the latter presents none. Position 97 (forming the C pocket) has also been implicated in several alleles, including the protective B*5801 and the susceptible B*5802 that differ only by 94, 95, and 97[126].

In terms of autoimmune disease, MHC class I has been linked to multiple sclerosis[128], and type 1 diabetes[129] amongst others. The clearest link between MHC class I genotype and autoimmune disease is for ankylosing spondylitis (AS); of the 27 HLA-B alleles known, three have been shown to predispose to AS (HLA-B*2701, 2704, and 2705) whilst two are protective (*2706 and *2709)[127]. Interestingly, *2705 and *2709 differ in residue 116 only (Asp and His respectively), which is involved in the F pocket. For *2705, the negative charge imparted by Asp116 is likely neutralised via salt bridge or hydrogen bond by the positive charge of His114. However, for *2709 the positive charge of His116 is thought to strongly affect the F pocket, altering PC specificity. Indeed, these alleles have been shown to have varying binding affinities for peptide, which has been postulated to result in differential development of HLA-B27-restricted CTL repertoires in *2705- and *2709- positive individuals[130].

1.4.2.4 Beta-2 microglobulin

Structure of β_2m

β_2m is a 99-residue, 11.8 kDa protein. Initially observed in the urine of patients with tubular proteinuria[131], the protein was sequenced in 1973[132] and found to have homology with portions of the immunoglobulin G constant region. β_2m contains an immunoglobulin fold, first observed in class I MHC-bound β_2m by X-ray crystallography in 1987[93,94] and confirmed to exist in monomeric β_2m through 2D-NMR secondary structure studies in 1992[133], with detailed crystal structures only available in 2002[134,135]. The fold is prototypic, with the β-strands A, B, E, and D forming one antiparallel β-pleated sheet and strands C, F, and G comprising the other (Figure 1.4). A single disulphide bond is formed between Cys25 and Cys80 in strands B and F respectively. Whether in isolation or bound by MHC, there are few unstructured regions.

The structures of monomeric β_2m suggest the presence of multiple species; NMR studies reveal a D-strand bulge of residues Ser52, Asp53, and Leu54[134] (Figure 1.4A) whilst in the crystal structure residues 51-56 form a continuous β-strand with six hydrogen bonds to strand E[135] (Figure 1.4B). In comparison to the X-ray structure, the NMR structure lacking the β-bulge contains a rotation of 180° in His51 and a re-arrangement of the N-terminal hydrogen bonding, whilst the C strand is split into two sections with an additional area of secondary structure in the C-D loop[135]. In addition to this structural disparity, the NMR analysis revealed that the A-B (residues 13-18) and D-E (residues 53-62) loop regions are dynamic and not uniquely structured in

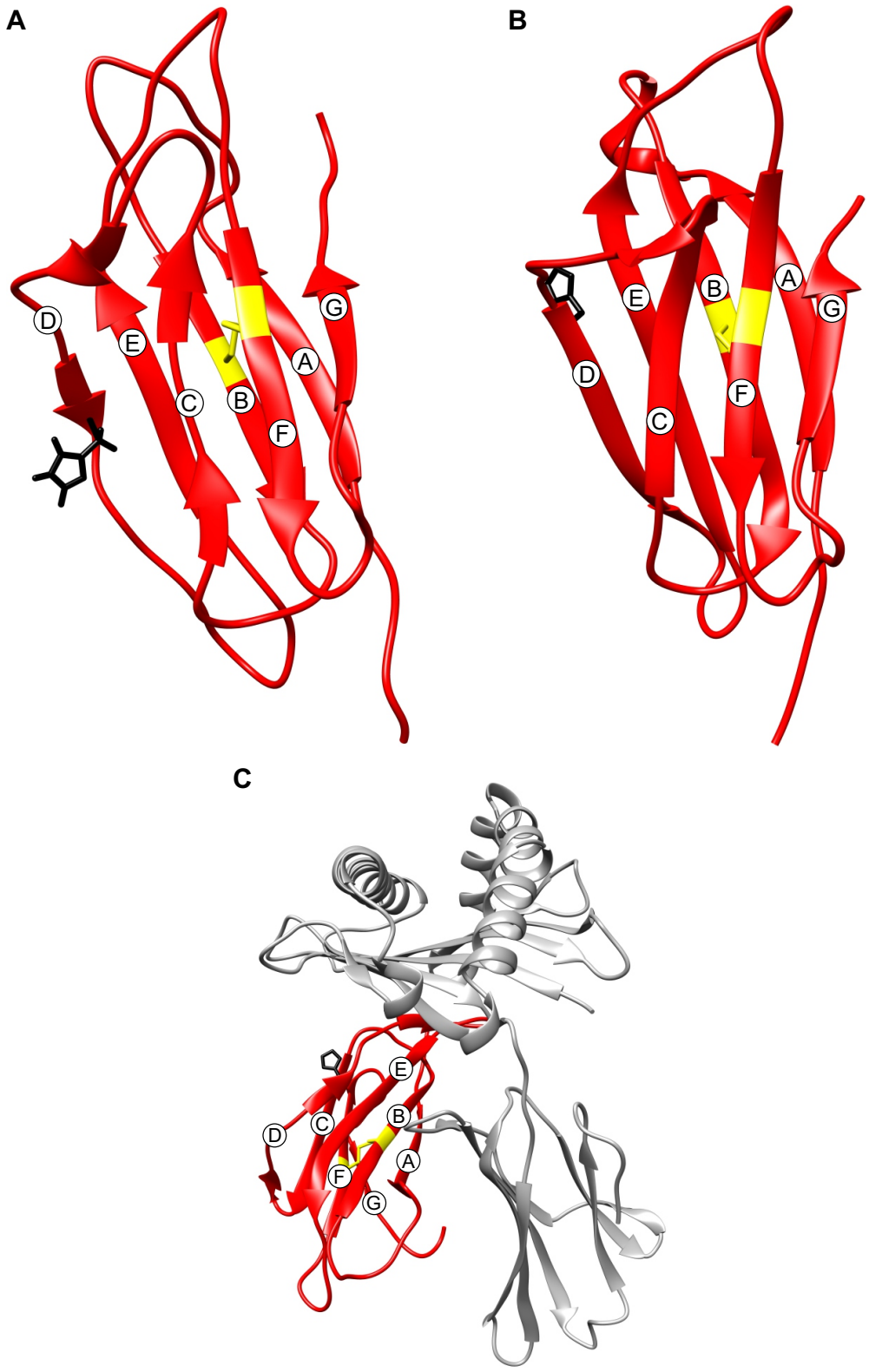


Figure 1.4 – Structure of hβ$_2$m.
(A) Structure of monomeric β2m obtained by NMR (PDB: 1JNJ[134]). For clarity, only a single representative chain of the structural ensemble is depicted. β-strands forming the antiparallel β-sheet are labelled by letter, A-G. The disulphide bond between Cys25 and Cys80 is shown in yellow. The side chain of His31 is shown in black. **(B)** Structure of monomeric β$_2$m obtained by X-ray crystallography (PDB: 1LDS[135]). In the absence of the β-bulge the D strand is continuous, and the side chain of His31 has rotated 180°. **(C)** Structure of MHC class I-bound β$_2$m obtained by X-ray crystallography (PDB: 3HLA[94]). A-chain is shown in grey. The D strand structure of β$_2$m (red) (including the His31 side chain) is consistent with the structure of monomeric β$_2$m obtained by NMR.

solution; it appears that the crystal structure is of a rare conformer. β_2m maintains the immunoglobulin fold when present as the light chain of the MHC class I complex (Figure 1.4C), with a D strand consistent with the solution-state structure of monomeric β_2m. The D-strand bulge twists the edge strand to facilitate binding to the HC, which contacts a number of residues as listed in Table 1.1. Comparison of isolated and MHC-bound β_2m shows that the dynamics of the structure are reduced when it is present in the MHC class I complex[136].

β_2m Physiology

Synthesised by all nucleated cells, β_2m is commonly found as the invariant chain in membrane-anchored peptide- or lipid- presenting protein complexes as well as in Fc receptors. In addition to MHC class I, β_2m is also present in a number of other MHC class I-like complexes, including CD1 which presents antigenic lipids to natural killer T cells[137]. On average 10^5-10^6 β_2m molecules are present on the surface of lymphocytes[138].

Monomeric β_2m is present at low levels (128-256 nM) in serum[139] from MHC class I degradation; the α-chain is internalised whilst β_2m is released into circulation where it has unknown function[95]. The kidneys filter over 99.9% of serum β_2m, re-absorb it in the proximal tubule and degrade it[139]. Serum levels can rise (e.g. due to chronic immunoinflammatory conditions) without resulting in disease, but when the glomerular filtration rate is reduced (e.g. due to diabetic nephropathy) β_2m can no longer be efficiently removed from the bloodstream. This results in prolonged elevated concentrations and amyloidosis[140].

β_2m and Amyloidosis

Amyloidosis is the deposition, and aggregation into highly ordered fibrils, of insoluble plaques comprising misfolded proteins. In addition to protein, plaques commonly contain minor constituents of proteoglycans, sugars, and serum amyloid P component, amongst others[140]. Deposits are classified as amyloid if they exhibit apple-green birefringence when stained with Congo red dye and viewed under cross-polarised light[141]. Amyloid deposits are associated with numerous diseases, including neurodegenerative conditions such as Alzheimer's disease, non-neuropathic localised diseases such as cataracts, and non-neuropathic systemic diseases such as the haemodialysis-related amyloidosis involving β_2m[140].

The formation of amyloid has not been fully elucidated, but it is thought to be a two-step nucleation-dependent process that requires the partial unfolding of a natively

globular protein[140]. Electron microscopy has shown that amyloid fibrils are highly ordered, indefinite length, unbranched chains[142], whilst X-ray diffraction reveals a cross-β structure[143]. Solid-state NMR atomic-level structures contain variations of closely-packed parallel β-sheets[144].

β_2m conversion into fibrils *in vivo* remains unknown, although *in vitro* formation is possible through lowering the pH, the use of amyloid seeds, and the introduction of copper (associated with haemodialysis), with a number of post-translational modifications also implicated[145]. The most significant structural finding relating to β_2m fibrillation is the removal of the D-strand β-bulge (Figure 1.4A vs. Figure 1.4B). Edge-to-edge association of adjacent monomers is enhanced by long, regular β-strands at the edge of β-sheets. In both monomeric and MHC-bound β_2m the A and G edge strand pair are protected by an inward-pointing charge contribution from Lys91, whilst the C and D edge pair is protected by the disruption of the D strand as well as the inward-pointing His31[135]. Removal of the β-bulge, postulated to be via cis-trans isomerisation of the His31-Pro32 amide bond[145], eliminates both protective features.

1.4.3 Dynamics of MHC class I

Static structures are not enough to explain all MHC features and abilities, particularly immunodominance and peptide optimisation. There is evidence from a number of structural biology techniques, and from extensive MD simulations, that MHC class I undergoes conformational fluctuations during both peptide loading and whilst present on the cell surface.

1.4.3.1 MHC class I acts as a peptide selector in tandem with TPN

It is clear that, of the available pool of peptides in the ER, MHC class I does not bind to the first encountered and immediately progress to the cell surface. The presence of immunodominance means that there is a selection of which peptide is presented by MHC class I. As per Section 1.3.1, there are a number of stages during which peptide may be selected. The most significant of these is the TPN-mediated release of MHC from the PLC. TPN links TAP and the MHC dimer by binding to residues 128-136[146], immediately prior to the $\alpha_{2\text{-}1}$ helix, and to residues 222-229[147] of the α_3 D-strand. But how does TPN detect sub-optimal peptide and mediate the release of such for higher-affinity peptide? Not all MHC class I alleles are affected by TPN to the same extent; whilst some exhibit markedly altered peptide binding hierarchies when TPN is present, for others the range of peptides found on the cell surface does not alter. This

has made it possible to investigate the selection of peptide further, using a combination of molecular biology and MDS approaches.

A great deal of literature deals with the TPN-dependency conferred by position 116, a residue on the platform of the PBG that contributes to the PC pocket. Originally identified in 2002[148], it is a residue easily investigated due to the existence of two alleles that are identical apart from this position; HLA-B*4402 contains Asp116 and is TPN-dependent whilst B*4405 is Tyr116 and able to load peptide in the absence of the entire PLC[149]. Extensive molecular dynamics (MD) simulations have shown that, despite a similar structure when peptide is bound, the empty binding grooves have distinctly different conformations in which the $\alpha_{2\text{-}1}$ helix in B*4402 moves outwards[150,151]. This means that B*4405 remains peptide-receptive, whereas B*4402 does not maintain the conformation[150] required to bind the C-terminus of peptide[152]. TPN induces a conformational change of the $\alpha_{2\text{-}1}$ helix to stabilise the peptide-receptive state[153]. Thus, whilst B*4405 is able to maintain a stable wide-PBG conformation, B*4402 requires chaperoning by TPN[154,155]. The mechanism by which TPN promotes peptide optimisation may therefore be that TPN acts to widen the PBG and only high-affinity peptides are able to overcome this force to close the binding groove and enter the mature conformation that can then travel to the cell surface[156].

The effect of altered peptide optimisation can have pathophysiological ramifications. Residue 116 is also the sole difference between HLA-B*2705 (Asp) and B*2709 (His). IR spectroscopy has shown that B*2705 is more flexible, with MDS suggesting this flexibility is located within the PBG and permits the groove to widen[157]. Further MD simulations examining the effect of different peptides bound to these two alleles implicated the α_1 helix as a source of motion within MHC class I[158]. It is clear that the conformation of the PBG has an effect on TCR recognition, with the same peptide presented by these two MHC alleles having different immunogenicity[158], possibly due to differences in the flexibility of peptide itself; B*2709 exhibits pronounced flexibility[130] and a greater ability to stimulate a T cell response[158]. This may explain why B*2705 is associated with ankylosing spondylitis whilst B*2709 is not.

In other areas of MHC class I α-chain, residue 156 forms parts of the D and E pockets in HLA, and part of the ridge in H-2D^b. It is known to have an effect on T lymphocyte alloreactivity, and in TPN-deficient cells only a B*44/156Arg mutant can be expressed on the cell surface, likely due to the extended side chain of Arg being able to stabilise non-optimal peptides, despite 156Arg and 156Asp requiring identical peptide anchor motifs[159]. Residues 97, 114, and 156 tether the $\alpha_{2\text{-}2}$ helix to the floor of

the PBG via hydrogen bonding, and 114 is also implicated in TPN-dependency; an acidic residue confers high TPN reliance whilst basic residues confer low reliance[160].

However, none of these residues or regions contact TPN directly; the interaction with the PBG appears indirectly coupled to the TPN binding region of 222-229 in the α_3 domain, and possibly also to the pre-α_2 128-136 region. A key MDS investigation has shown that (in the B*4402 allele) it is possible to affect the PBG conformation by altering the plasticity of the 220-227 α_3 region[155]. This matches studies into chicken MHC alleles, in which position 220 was shown *in vitro* to influence the effect of TPN on peptide dissociation[161]. It can therefore be postulated that flexibility is somehow communicated through the secondary structure of MHC class I from the TPN binding site to the PBG, with some MD simulations hinting at such an allosteric coupling[76,155]. Currently, there is limited direct evidence for systemic MHC class I dynamics and this requires further investigation.

1.4.3.2 An intermediate during peptide loading?

Even for TPN-independent MHC class I, there is still evidence of peptide selection, and for TPN-dependent molecules the dependence correlates with conformational flexibility such that TPN is acting as a chaperone to stabilise MHC class I conformation[153]. Significant efforts have been made to investigate conformational change within MHC class I using molecular dynamics simulations, which have shown that peptide binding is generally under kinetic control. TPN introduces a partial thermodynamic control by inducing different MHC conformations around the $\alpha_{2\text{-}1}$ helix to increase the off-rate of all peptides[162] such that in the presence of TPN the binding groove adopts a wider, more peptide-receptive conformation that is capable of universally binding peptide more readily[152], thus evading the restricted time spent within the ER[163]. These MDS studies implicate flexibility in the PC/F pocket adjacent to the $\alpha_{2\text{-}1}$ helix as being required for peptide binding and optimisation[150] (explaining why residues in this area of the PBD confer TPN dependency[152]). It is clear that peptide binding at the amino-terminus follows a different, less significant conformational change[151], and this is reinforced by molecular biology research into partial peptide epitopes that found C-terminal shortened peptides were unable to stabilise MHC class I whilst those missing N-terminal anchor residues resulted in a stable, mature trimeric complex; PC anchors are more critical than PN[79]. The availability of such extensive MD simulation data enables a mechanistic scheme of the peptide binding and optimisation process to be described in such detail that experimental data can be combined with it to enable quantification of the range of

peptides presented at the cell surface[22,164]. This 'peptide filtering' scheme is a significant step forward in understanding the implications and relative effects of each stage of peptide processing. The best example of this relates to the TPN dependency of the B*4402 and B*4405 alleles; in combination with pulse-chase and thermostability *in vivo* assays, the MD-predicted intermediate steps are shown to be significant in peptide selection[155].

Distinct from TPN-dependence, other areas of the α-chain have shown flexibility during peptide loading. This includes residues 47-51, immediately before the α_1 helix and supporting the PN pocket, which are solvent-exposed in the peptide-free state but not when peptide is bound. This suggests that peptide binding stabilises the 47-51 region, allowing the binding pockets to fully form and increase the affinity of the interaction[165].

Moving away from MD simulations, more established molecular biology techniques have also provided evidence for an intermediate state. Some antibodies are sensitive to the folding of the PBG, and thus act as markers for MHC class I associated with optimal peptide. Investigation has suggested the presence of transitional states between the empty-PBG and optimal peptide-bound forms[166]. Antibodies can also be used to investigate a variety of peptides for their ability to stabilise MHC conformation, showing that although many peptides can be bound not all are capable of stabilising the mature conformation[167]. Further investigation using FRET (fluorescence resonance energy transfer) suggests the peptide-bound structure is more compact than peptide-empty[168] but although FRET provides a good measure of the kinetics of complex formation, and provides evidence of at least two conformations of α-chain, it is unable to distinguish between the effect of β_2m or peptide binding[169].

Immunodominance is vital in eliciting an effective immune response, and therefore the initial peptide bound by MHC class I within the ER may require optimisation. This occurs via motion within the MHC molecule, particularly within the PBD, creating at least one intermediate conformation. Evidence from molecular biology and MD simulations suggest that the motion is localised to the PBD with the intermediate state possessing a wider PBG to facilitate peptide exchange. However, it is the kinetic stability of the MHC class I-peptide complex that is vital for eliciting an immune response[54]. For example, if an intermediate conformation has a wide PBG the interaction with peptide will be lower affinity due to disrupted hydrogen bonding. If this lower-stability complex is exported to the cell surface it will not display the

ability to sustain T cell expansion[170] and thus this intermediate must undergo conversion to a final closed-PBG conformation. This MHC class I plasticity appears to be an evolutionarily conserved feature[161], with a strong correlation between the degree of plasticity demonstrated by an allele and the ability to select high affinity peptide[171]. A more dynamic structure is able to experience more conformations and is therefore quicker at binding to a high-affinity peptide that is able to close the PBG and ensure a stable mature conformation[75,76,161,171].

It is worth noting that such conformational fluctuations have not been noted in X-ray crystallography structures; immunodominant and non-immunodominant peptides produce essentially identical structures although studies have shown that the PC portion of peptide must be present to stabilise the PBG (i.e. suggesting there is motion in the $\alpha_{2\text{-}1}$ helix that must be abrogated to narrow the PBG)[79]. To confirm the MDS findings alternative techniques need to be utilised. NMR is well suited to this task.

1.5 Nuclear magnetic resonance spectroscopy

The material found within this section, and within subsequent introductory sections of later chapters, is well explained in a number of textbooks[172-174] and selects only theory particularly relevant to this project. For a more detailed explanation, please see Spin Dynamics (2nd edn.) by Malcolm Levitt (2008)[174].

1.5.1 Discovery

Spectroscopy is the study of the interaction between matter and electromagnetic radiation, and how matter is influenced is dependent on the frequency of the radiation; thus there are many forms of spectroscopy. NMR spectroscopy makes use of the magnetic properties of some atomic nuclei.

NMR was discovered in 1945 by Bloch and Purcell, and a working group including Proctor and Yu in the late 1940s measured the first chemical shift. In 1951 it was found that the nuclear resonance frequency of a particular nucleus is dependent on its chemical environment. This made the technique particularly useful in the determination of the chemical composition of small molecules and compounds. In 1952, Bloch and Purcell shared the Nobel Prize in Physics '*for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith*'[175]. A fascinating personal account by Proctor of the early discoveries in NMR has been made available by Sykora[176]. Initially the low field strengths of NMR magnets prevented the analysis of large compounds, and the first protein spectrum was reported in 1957. The subsequent introduction of superconducting magnets permitted the use of higher strength magnetic fields and thus it was only after 1967 that NMR began to regularly be used to complement crystal structures. The first complete protein structure determined solely by NMR was published in 1985[177]. Today, NMR spectroscopy is an efficient method for the determination of atomic resolution structure.

1.5.2 Fundamentals

Briefly, a NMR spectrometer consists of a magnet, radiofrequency transmitter, probe, radiofrequency receiver, frequency synthesisers, amplifiers, and a control computer. The superconducting magnet must generate a field that is strong and stable, with higher field strengths producing greater sensitivity. The field must also be highly homogeneous, requiring the additional use of shim coils. The radiofrequency produced by the transmitter is converted to radiofrequency pulses using a gate

controlled by a pulse generator, in turn controlled by the computer. The amplified pulses are sent to the NMR probe, which contains a wire coil to induce a magnetic field in response to these pulses. The NMR signal elicited from the sample induces an oscillating voltage, which is amplified and sent to the radiofrequency receiver. The computer can then digitise the signal.

The frequencies detected by this method consist of $\omega_{rf} + \Omega$, where the transmitter frequency of $\omega_{rf}/2\pi$ is commonly between 600 and 950 MHz whilst the resonance offsets Ω (the chemical shift of the nucleus being studied) are typically kHz. The signal detected in the coil is mixed with the spectrometer reference frequency ω_{rf} so that only the desired Ω signal is passed to the computer. The result of this mixing is equivalent to observing the generated signal in a frame rotating at ω_{rf}; hence the term *rotating frame* in the discussion of signal collection.

Quadrature detection is used to detect the free induction decay, using two detection channels that are orthogonal in the rotating frame. For each resonance two signals are acquired; the cosine and sine functions of the offset frequency Ω, which can be considered the real and imaginary parts of a complex time-domain signal. Fourier transformation is then used to convert this into a frequency-domain spectrum:

$$S(\Omega) = \int_0^\infty s(t) exp\{-i\Omega t\} dt$$

Equation 1.3

where 'input' s is a function of time t, 'output' S is a function of a frequency variable Ω and is called the spectrum

However, this assumes that the absorption lineshape is in the real part of the spectrum and the dispersion lineshape is in the imaginary part. In order to obtain a sharp and uniform peak, the spectrum must be phased by taking linear combinations of the real and imaginary parts.

1.5.2.1 Nuclear spin

NMR spectroscopy relies on the nuclear spin characteristic, a form of angular momentum that is an intrinsic property of nuclei. This means that different energy levels are present, the number of which depend on the spin value (0, 1/2, 1 etc.). Normally degenerate, when a magnetic field is applied these energy levels are separated due to Zeeman splitting, with a stronger magnetic field causing a larger split. These nuclei are now able to absorb a radiofrequency wave with energy equivalent to the difference in energy levels, and this absorption produces a resonance signal detectable by the NMR spectrometer.

1.5.2.2 Chemical shift

The NMR frequency of a nucleus is dependent on the strength of the magnetic field it is experiencing, and its gyromagnetic ratio:

$$\nu = \frac{\gamma B}{2\pi}$$

Equation 1.4

where ν is the NMR frequency of the nucleus, γ gyromagnetic ratio, and B magnetic field strength

Therefore all nuclei of a particular species (protons, carbons etc.) will resonate at a similar frequency to one another. However, the frequency is also dependent on the local electron distribution around the nucleus i.e. the immediate chemical environment. This effect is the chemical shift and means that, for example, the Cα of alanine can be distinguished from the Cα of arginine.

The field B experienced by nuclei is not identical to the external magnetic field B_0; the presence of B_0 induces the circulation of electrons within their orbitals. This induces an additional small magnetic field B', which can augment or oppose B_0. The nucleus is therefore shielded or deshielded from the external field by its surrounding electrons:

$$B = B_0 - B' = B_0(1 - \sigma)$$

Equation 1.5

where σ is the constant of proportionality between B' and B_0; the shielding constant

Alternatively, if a nucleus is bonded to a more electronegative atom, it will be exposed to the external field when its electrons are pulled away. The chemical shift is defined in terms of the difference in resonance frequency between the nucleus of interest and a reference nucleus, using a dimensionless parameter δ, with δ values given in parts per million (ppm).

1.5.2.3 Dipole-dipole coupling

Each nuclear spin is magnetic and therefore generates a magnetic field in the direction of the spin magnetic moment. When a second nuclear spin is present it will interact with the magnetic field of the first, whilst also generating its own field that the first will interact with. This is through-space dipole-dipole coupling. Fluctuating around zero, its primary effect in solution NMR is relaxation. The most prominent exploitation of dipolar relaxation is the NOE effect, which can be used to estimate the distance between spins.

1.5.2.4 J-coupling

J-coupling, or indirect coupling, arises from the influence of electrons involved in chemical bonds on the magnetic fields between two nuclear spins; two nuclear spins will have a J-coupling only if they are linked together through a small number of chemical bonds (including hydrogen bonds). There are both isotropic and anisotropic components; the former averages out and the latter gives rise to spin-spin splittings (a single resonance may be split into a multiplet).

1.5.2.5 The Hamiltonian

In quantum mechanics, the Hamiltonian operator represents the total energy of the system in question. The nuclear spin Hamiltonian encompasses all interactions in the system, and is the sum of the Hamiltonians for chemical shift, dipole-dipole coupling, and J-coupling[174]. Solution-state NMR spectroscopy detects only the isotropic components of the chemical shift and J-coupling due to the rapid molecular tumbling averaging out the anisotropic interactions. In special applications a small degree of spatial anisotropy can be induced in a solution sample leading to measurable residual dipolar couplings[178]. The anisotropic component of chemical shift and dipole-dipole coupling influences line-widths, and along with chemical shift anisotropy is important for relaxation properties.

1.5.3 Solution-state NMR study of proteins

The property of chemical shift produces a spectrum of absorption against frequency in which each NMR-active nucleus in a protein is distinct. However, the peaks cannot always be fully resolved in a one-dimensional spectrum. To yield an interpretable spectrum two or more dimensions can be introduced, where each dimension defines a particular type of nucleus. The correlations between dimensions can either be homonuclear (^{1}H-^{1}H) or heteronuclear (typically ^{1}H-^{15}N and/or ^{1}H-^{13}C). It is also possible to be selective about the nuclei observed; some pulse sequences are designed so that only backbone nuclei (HN, Cα, and CO) are visible whilst others focus on the side chain nuclei.

The typical starting point for NMR study of a protein is a ^{1}H-^{15}N heteronuclear single-quantum correlation (HSQC) experiment that produces a 'fingerprint' spectrum containing all the ^{1}H-^{15}N correlations, such that each residue produces a single signal on a two-dimensional spectrum (with ^{1}H in the direct and ^{15}N in the indirect dimension). Proline residues are a special case due to their cyclised structure rendering them invisible to the HSQC experiment, whilst side chain NH groups from

arginine, asparagine, glutamine, histidine, and tryptophan are visible in addition to the backbone NH group. However, these side chain NHs appear in predictable areas of the spectrum and in general have little overlap with backbone resonances.

Using the HSQC as a basis, the aim is to identify to which residue each peak belongs in a process called assignment. This involves the collection of three-dimensional spectra, usually CBCANH and CBCA(CO)NH (Figure 1.5A and B) that together can be used to identify the HN, N, Cα, and Cβ chemical shifts for each residue (Figure 1.5C). This forms the backbone resonance assignment, with further data available from side chain assignment using spectra such as HBHA(CO)NH, H(CCO)NH, and CC(CO)NH[179]. These spectra are of limited use when applied to MHC class I (Section 1.5.3.2).

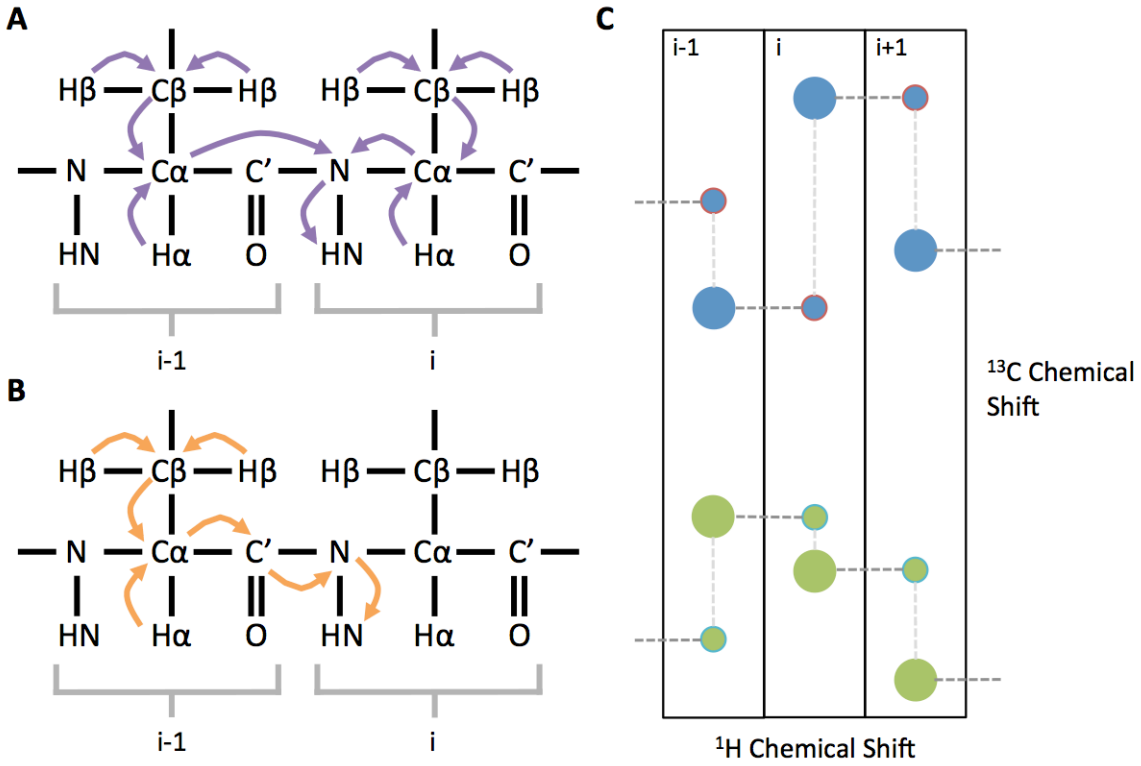


Figure 1.5 – Triple resonance backbone resonance assignment.
(A) Magnetisation transfer during a CBCANH experiment. Starting on the Hα and Hβ groups, magnetisation passes to the Cα and Cβ nuclei respectively, and then from Cβ to Cα. It then moves to the nitrogen nucleus of residue i from both $C\alpha_i$ and $C\alpha_{i-1}$, before passing to the amide HN for detection. For residue i Cα and Cβ peaks from both i and i-1 are visible. **(B)** Magnetisation transfer during a CBCA(CO)NH experiment. Again, magnetisation starts on the Hα and Hβ groups before passing to the Cα and Cβ nuclei respectively, and then from Cβ to Cα. It then moves to the carbonyl group before the N and HN group. For residue i, the observable Cα and Cβ peaks belong to residue i-1. **(C)** Sequential linking of residues using CBCANH and CBCA(CO)NH spectra. Each residue produces a ^{1}H-^{13}C strip of resonances containing four peaks; Cα peaks are positive (blue, red), Cβ peaks negative (green, cyan). In the CBCANH spectrum (blue and green) the $C\alpha_i$ and $C\beta_i$ typically produce stronger signals than $C\alpha_{i-1}$ and $C\beta_{i-1}$. As a further way to separate the i and i-1 peaks, the CBCA(CO)NH experiment (red and cyan) shows only i-1. During the assignment process sequential links are made between residues as shown for i-1, i, and i+1. Some residues produce distinctive Cα or Cβ chemical shifts and with a long sequence it is possible to identify the residues in question and assign the protein. Figure adapted from *protein-nmr.org.uk*[179].

Once a protein spectrum is assigned, there are a number of analyses that can be carried out. These include the study of dynamic processes (see Section 1.5.4) and the determination of protein structure, utilising proton-proton distances derived from ^{1}H-^{1}H NOE measurements.

1.5.3.1 Isotope labelling

As discussed previously, NMR spectroscopy requires nuclear spin. The spin of a nucleus is dependent upon its protons and neutrons, with different isotopes of an element possessing different levels of spin (Table 1.2). A minority of isotopes possess zero spin and hence do not display Zeeman splitting. An alternative isotope must be therefore be used to investigate these NMR-silent nuclei, for example substituting ^{13}C for ^{12}C.

Nuclei that possess spin-1/2 are desirable for NMR investigation because they behave as perfect spheres. They have no electric energy terms that depend on the orientation or internal structure of the nucleus, and therefore behave as magnetic gyroscopes mounted on frictionless bearings. In contrast, nuclei with spin greater than 1/2 (quadrupolar nuclei) do not have spherically symmetrical electric charge distribution. This means that the electric energy of the nucleus is dependent on its orientation with respect to the remainder of the molecule; the electric quadrupole moment is dependent on the electric field gradient of the surrounding space. For this reason nuclei with spin-1 are generally avoided, and to investigate nitrogen nuclei protein must be labelled with the ^{15}N isotope.

1.5.3.2 Complications of investigating large proteins

The use of NMR spectroscopy to investigate proteins over 40 kDa in size has several complications. Larger proteins tumble slowly, which means their NMR signal decays more quickly and spectral peaks are broader. There will also be more observable

Table 1.2 – Nuclear isotope properties relevant to isotopic protein labelling
Adapted from Spin Dynamics Table 1.2[174].

Isotope	Nuclear Spin	Natural Abundance (%)
^{1}H	½	99.9885
^{2}H	1	0.0115
^{12}C	0	98.93
^{13}C	½	1.07
^{14}N	1	99.632
^{15}N	½	0.368

peaks, meaning a common problem is overlapping resonances. Although the use of quadrupolar nuclei is generally avoided in NMR spectroscopy, one solution is to isotopically label with ^{2}H (deuterium). This has only a small electric quadrupolar moment and therefore the orientation-dependent electric term is small. The advantage of using ^{2}H is twofold: the ^{2}H nucleus is not visible in ^{1}H-detected spectra leading to a significant reduction in

spectral complexity. In addition, replacing protons by ^{2}H reduces the cross relaxation of protons with their neighbours, leading to significantly enhanced lifetimes for the remaining proton resonances. Since dipolar coupling of ^{1}H protons to ^{13}C nuclei provides the dominant relaxation mechanism for ^{13}C nuclei, replacement of these protons with ^{2}H leads to significantly enhanced ^{13}C lifetimes. As a result, ^{13}C signals in complex triple resonance experiments of slowly tumbling large proteins are detectable only when the adjacent protons are substituted by ^{2}H.

As discussed above, HSQC experiments are the standard for NMR but are not well suited to the investigation of large proteins. HSQCs consist of an insensitive nuclei enhanced by polarisation transfer (INEPT) sequence, followed by a t_1 evolution during which a 180° pulse is applied to refocus the magnetisation resulting from J-coupling, and then a reverse INEPT (Figure 1.6A). The NMR signal from the first spin (^{1}H) is labelled with the modulation arising from the second spin (^{15}N) and then returned to the first for detection. When large proteins are under investigation it is common to use the transverse relaxation optimised spectroscopy (TROSY) experiment (Figure 1.6B), in which the NMR signal is acquired without heteronuclear decoupling. This pulse sequence has been designed so that it excites only the slowly-decaying component of the four-line ^{1}H-^{15}N multiplet. It exploits the fact that two relaxation mechanisms (in this case CSA and dipolar relaxation) on the same chemical bond can destructively interfere and cancel one another so that the resulting signal has a significantly increased lifetime.

There are three parameters that control spectra quality during the collection of any experiment:

$$res = \frac{1}{at} = \frac{2sw}{np}$$

Equation 1.6

where res is the digital resolution, at is the acquisition time, sw is the sweep width, and np is the number of points

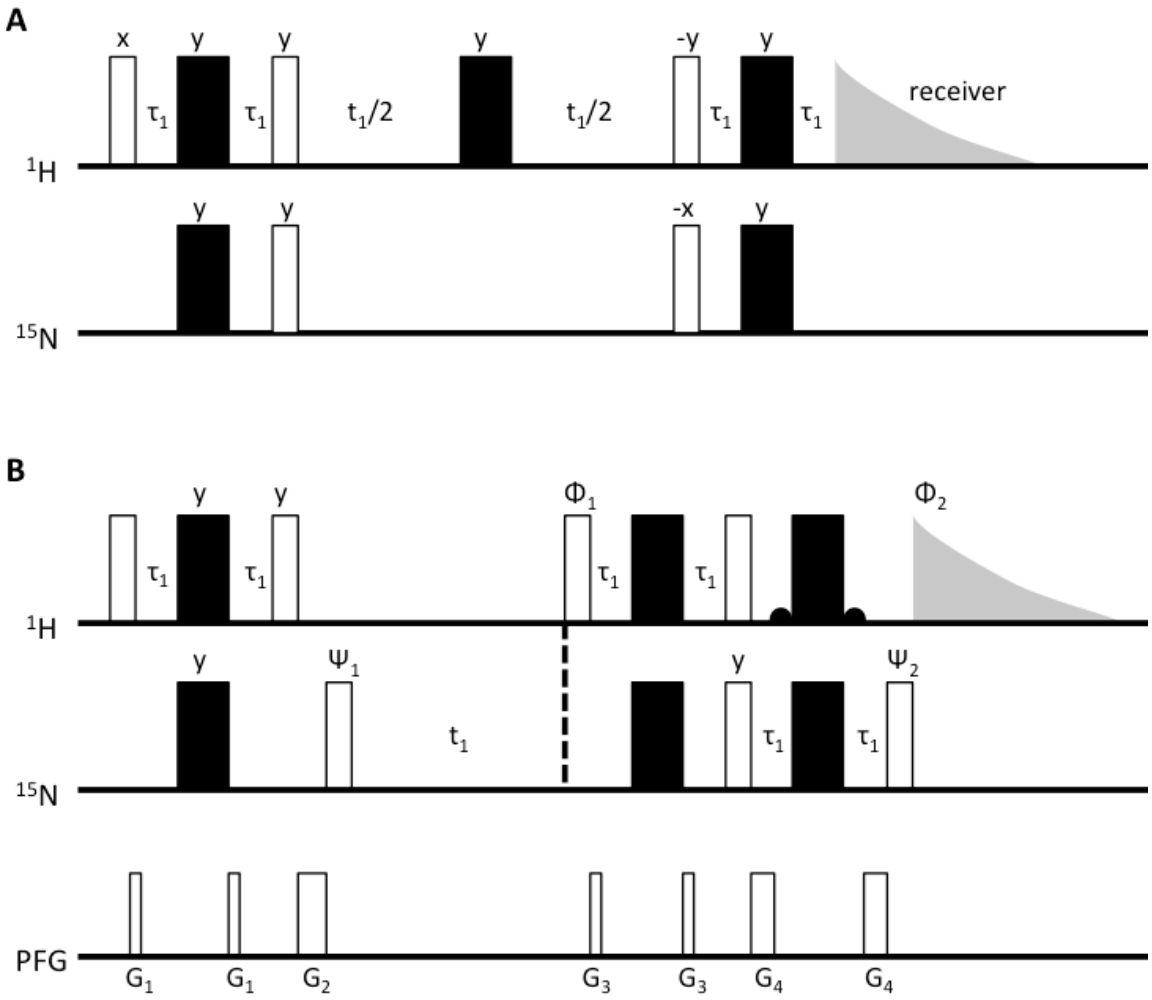


Figure 1.6 – HSQC and TROSY pulse sequences.
(A) HSQC pulse sequence. This can be enhanced using pulsed field gradients[180]. Open rectangles refer to 90° pulses, filled rectangles to 180° pulses. $\tau_1 = 1 \div (4\ J_{NH})$. **(B)** TROSY pulse sequence[181]. Water suppression via WATERGATE; two curved shapes indicate off-resonance radiofrequency pulses on -x. Pulses are on x unless stated otherwise; Ψ_1 = [y, -y, -x,x,y, -y, -x,x], Ψ_2 = [4(x),4(-x)], Φ_1 = [4(y),4(-y)], Φ_2 (receiver) = [x, -x, -y,y,x, -x,y, -y]. PFG (pulsed field gradients) are applied along the z-axis; G_1 = -30 G/cm, 0.4 ms, G_2 = -60 G/cm, 1 ms, G_3 = 50 G/cm, 0.4 ms, G_4 = 48 G/cm, 0.6 ms. A full account of the TROSY pulse sequence has been published by Peruvshin, Riek, Wider, and Wüthrich[181].

The first of these is the acquisition time of the spectrum; increasing the number of data points will initially increase resolution but the signal (the free induction decay) has a finite lifetime and an acquisition time significantly longer than theinevitable decrease in signal will result in a worse signal-to-noise ratio. On the other hand if the data receiver is switched off before the signal has mostly decayed, spectral artefacts will arise. A long acquisition time in the indirect dimension(s) will significantly increase the time required for data collection, and therefore a compromise is usually made between acquisition time and spectral resolution. The second parameter to alter is the number of points of data that are recorded during the analogue to digital conversion of data; this produces a series of points along the free induction decay

curve. The more points used to define the signal, the higher the digital resolution will be. The third important option to consider is the sweep width, which is the chemical shift range over which data is recorded.

During data processing, it is possible to increase spectral quality using zero filling or linear prediction. Zero filling is the addition of data points of zero intensity to the end of the free induction decay; thus adding data without adding noise. However, it should be noted that this only improves the apparent resolution. Another option during data processing is to use apodisation, a data weighting function that can emphasise the tail end of the signal to increase resolution (although also causing a decrease in signal:noise).

Magnetic field strength also has an effect on spectral quality. An improvement in quality comes about because NMR signals are proportional to the difference between the nuclear energy levels caused by Zeeman splitting. The energies of the levels are proportional to field strength and they separate in higher fields. Secondly, it is easier to resolve individual peaks (or multiplets) at higher fields because the difference between the chemical shifts at which nuclei resonate is proportional to field strength. The spin-spin coupling that gives rise to peak multiplets is not sensitive, and therefore at a higher field the spectral width (in Hz) will increase.

The inevitable degradation of spectral quality for larger proteins complicates the process of resonance assignment. Although the standard CBCANH and CBCA(CO)NH experiments can be carried out, where deuteration is necessary the order of nuclei must be reversed such that the experiments run as HNCACB and HN(CO)CACB. Typically these spectra are not sufficient for the backbone resonance assignment of proteins above 40 kDa, and other sets of three-dimensional spectra are obtained. These can include the HNCA and HN(CO)CA, and the HNCO and HN(CA)CO experiments (Figure 1.7), which have shorter magnetisation transfer pathways and better spectral quality.

1.5.4 Solution-state NMR study of dynamic processes within proteins

While X-ray crystallography has proved invaluable in the determination of protein structures, it is generally less able to describe protein dynamics. NMR spectroscopy is a technique suited to this area of study, with an ever-growing number of experiments designed solely to observe dynamic processes. This section deals briefly with the fundamentals, with more detailed explanation in Chapters 2 and 3, pertinent to the experiments conducted.

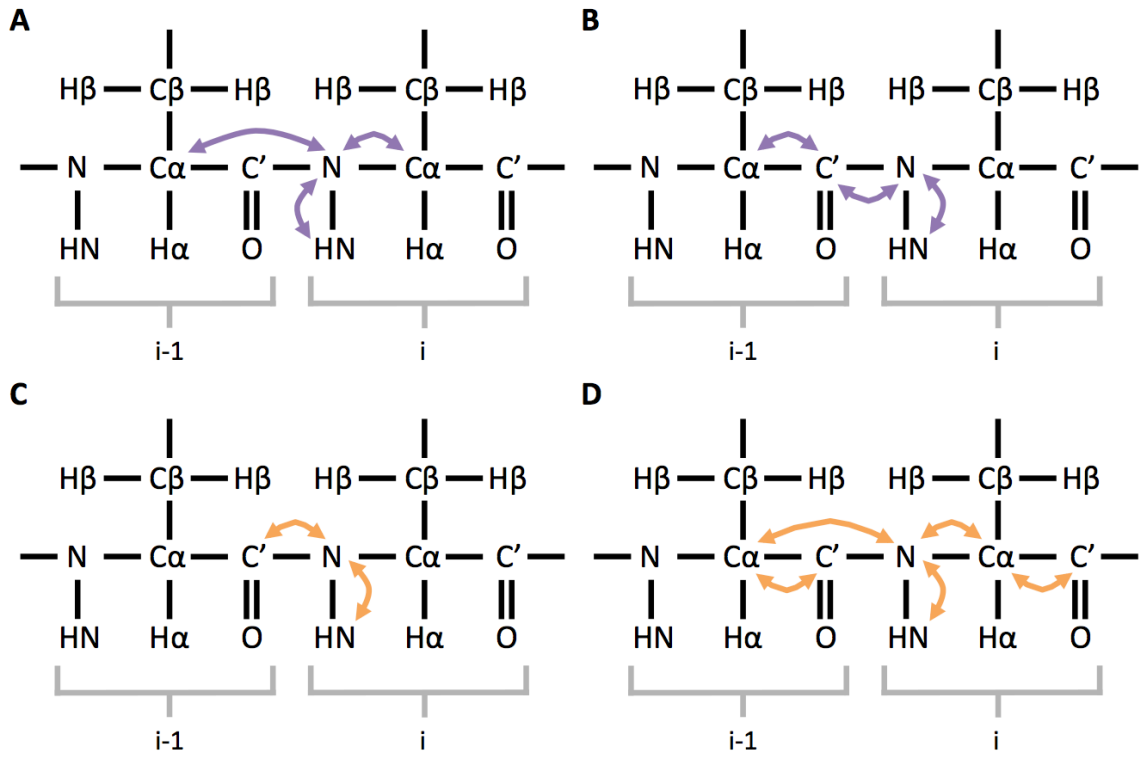


Figure 1.7 – Triple resonance backbone assignment for large proteins.
(A) Magnetisation transfer during a HNCA experiment. Starting on the amide HN, magnetisation passes to the nitrogen nuclei, then to either i or i-1 alpha carbons, before returning along the same path. For residue i peaks will be observed for $C\alpha_i$ and $C\alpha_{i-1}$, allowing the sequential linking of peaks. **(B)** Magnetisation transfer during a HN(CO)CA experiment. Starting on the amide HN, it passes to N and then to C' before returning along the same path. For residue i only the Cα of residue i-1 will be observed. **(C)** Magnetisation transfer during a HNCO experiment. Magnetisation travels from the HN to the nitrogen and then to the carbonyl of the previous residue before returning along the same path. For residue i a single peak for C'_{i-1} will be observed. **(D)** Magnetisation transfer during a HN(CA)CO experiment. Beginning on the amide HN, it passes to N and then to the i and i-1 carbonyls via the N-Cα J-coupling, before returning along the same path. For residue i peaks will be observed for C'_i and C'_{i-1}, allowing sequential linking of peaks. Figure adapted from *protein-nmr.org.uk*[179].

1.5.4.1 Chemical exchange

The chemical shift of a nucleus is extremely sensitive to the immediate electronic environment and any slight change is observable on a spectrum. This means that minute changes in conformation can be detected. This phenomenon of chemical exchange is seen where two or more resonance peaks are related by a form of structural exchange and therefore can be used to detect internal motions as well as interactions with other molecules. The simplest example is that of a nucleus that exists in two conformations i.e. is in two-site conformational exchange, and that has equal rates of interconversion between two equally populated conformers. The two conformations will produce different chemical shifts, but two peaks are not always observable. If the exchange rate is slow compared to the duration of the NMR signal encoding time (and hence slow compared to the frequency difference, in Hz, between the resonances), two equally intense, narrow peaks at the resonance frequencies of

the two sites will be observed. As the rate of exchange increases, the peak linewidths will increase and the peaks will appear to move towards one another, before broadening further until they coalesce into a single, very broad resonance. At this point, the exchange rate is equal in magnitude to the difference in resonance frequencies of the two original sites. If the exchange rate increases further the intensity of the observed resonance will increase whilst linewidth remains the same. In the fast exchange regime the observed chemical shift represents the population-weighted average of the two contributing resonances. This process of peak change is known as slow, intermediate, and fast exchange (Figure 1.8A). For the more common case of unsymmetrical two-site exchange, the two resonance peaks will not be of equal intensity in slow exchange and will broaden unequally as the exchange rate increases, with the single resonance observed for fast exchange weighted in the favour of the more common state (Figure 1.8B).

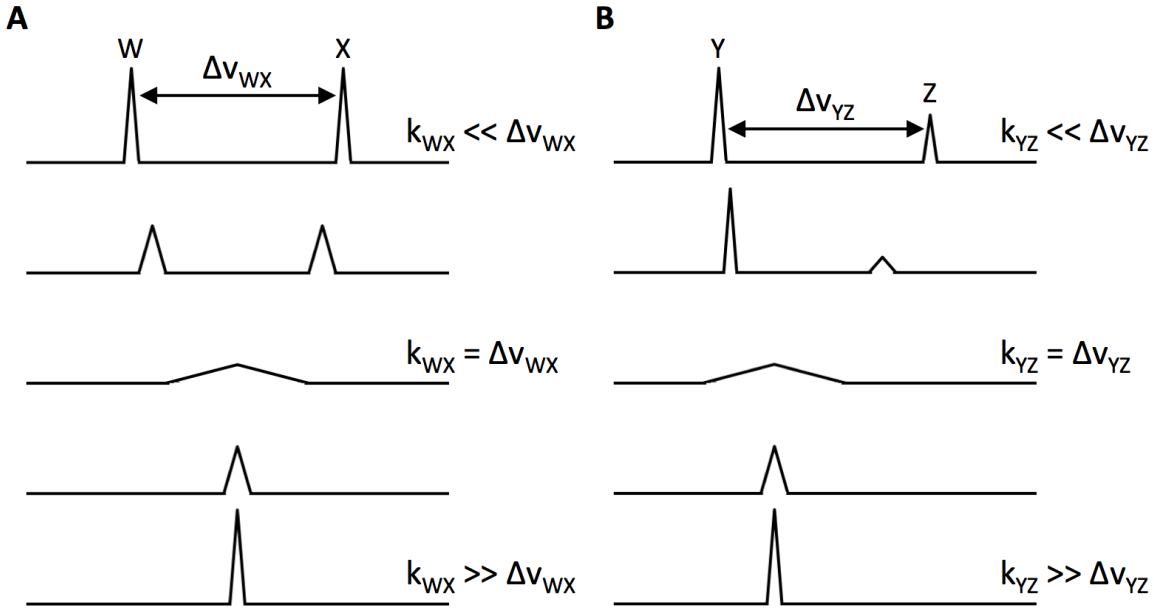


Figure 1.8 – Two-site chemical exchange.
(A) Two-site symmetrical exchange. A nucleus is present in two equally populated conformations, W and X, giving rise to peaks of equal intensity. In slow exchange (top) both peaks are observed because the exchange constant k_{WX} is smaller than the difference in resonance between the two peaks, $\Delta\nu_{WX}$. In intermediate exchange (middle) Y and Z have merged and k_{WX} is equal to $\Delta\nu_{WX}$. In fast exchange (bottom), k_{WX} is greater than $\Delta\nu_{WX}$ and only a single intense peak is observed at the midpoint between the fast exchange peaks. **(B)** Two-site unsymmetrical exchange. A nucleus is present in two conformations, Y and Z, where Y has a higher population and hence greater peak intensity when in slow exchange (top). As the exchange rate increases, the peak corresponding to Z broadens first, and the position of the merged peak is weighted towards population Y in both intermediate (middle) and fast exchange (bottom).

1.5.4.2 Heteronuclear relaxation

The way in which the signals generated during NMR spectroscopy decay to their equilibrium depends on the motions of the nucleus under observation. This enables dynamic processes on a picosecond to millisecond timescale to be observed.

Following the perturbation of nuclei by a radiofrequency pulse, the macroscopic magnetisation returns to equilibrium via two relaxation times; longitudinal spin-lattice (T_1) relaxation and transverse spin-spin (T_2) relaxation. The former is the decay constant for the recovery of the Z component of nuclear spin towards thermal equilibrium and involves energy transfer to the surroundings, whilst the latter is the decay constant for the X-Y component as a result of loss of coherence in this component. In general, T_1 and T_2 can be quantitatively described as a sum of contributions from the active relaxation mechanism (dipolar and CSA, and chemical exchange), and the number of accessible states in the surroundings (lattice) encoded by the spectral density.

The energy levels of a spin split apart in a magnetic field due to Zeeman splitting. The populations of the states are unequal; in T_1 relaxation the spins flip between levels to establish Boltzmann population differences, releasing energy into the surrounding lattice. T_1 relaxation therefore reduces the lifetime of spin states and causes line broadening, similar to slow chemical exchange. For a model of the surroundings as a continuum of random rotational fluctuation the spectral density takes a simple analytical form:

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2\tau_c^2}$$

Equation 1.7

where *J(ω)* is the spectral density, ω the angular frequency in radians s^{-1}, and τ_c the rotational correlation time

The T_1 relaxation rate is large when the tumbling time is significantly smaller, or significantly larger, than the angular frequency. It reaches a minimum when tumbling time and angular frequency match and $\omega_0\tau_c=1$ (Figure 1.9). At this minimum the relaxation rate is at maximum, for example if the rotational motion of a NH bond vector matches the transition frequencies of the NH spin system. In this example the T_1 becomes longer if the motion is either faster or slower. A fast T_1 relaxation means that the surrounding lattice has a large number of available fluctuations that can absorb the energy that a nuclear transition releases.

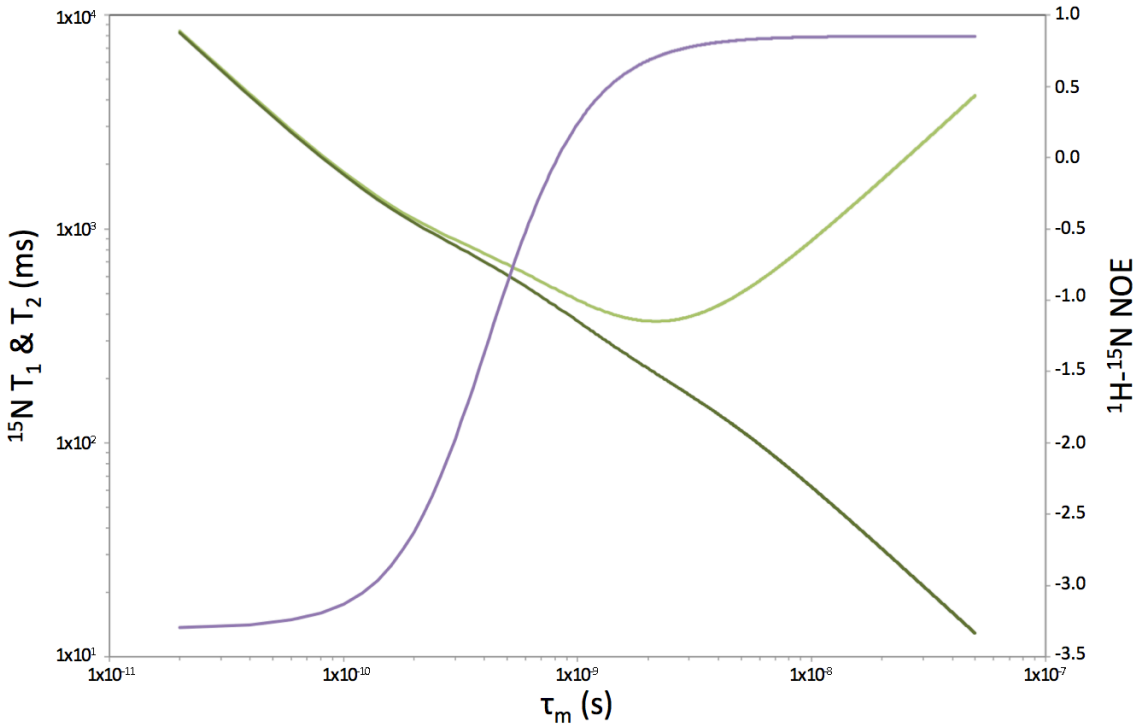


Figure 1.9 – The relationship of ^{15}N T_1, ^{15}N T_2, and ^{1}H-^{15}N NOE with rotational correlation.
The relationship between T_1 (light green), T_2 (dark green), and NOE (purple) and the rotational correlation time (τ_m) is dependent on field strength; data shown here is for 700 MHz. Data generated from a custom Linux script created by Dr Joern Werner.

Linewidths are also affected by T_2 relaxation; whilst T_1 is spin-lattice relaxation and requires exchange of energy with the lattice to return spin populations to the Boltzmann equilibrium, T_2 does not exchange energy with the surrounding lattice but instead between coupled spins (spin-spin relaxation) and describes the loss of phase coherence between spins. Neighbouring nuclei produce NMR signals that are split many times into multiplets, and when the motion in the molecule observed is small (i.e. the molecule is large) the spectral peak observed is broad:

$$1/\pi T_2 = \Delta v$$

Equation 1.8

where Δv is the linewidth associated with relaxation processes

A small molecule has faster tumbling and the resultant motion of neighbouring spins modulates the dipolar coupling experienced. Similar to fast chemical exchange, the splittings of the NMR signal become averaged, as long as the motion is fast compared to the frequency differences being modified. Increased motion therefore increases T_2 relaxation (Figure 1.9). In a coupled 2-spin system such as a ^{1}H-^{15}N spin pair the processes that lead to de-coherence of spins with each other are particularly sensitive to low frequency molecular motions such as the molecular rotation or the exchange between two or more conformations. When Δv is measured as an average over the

entire detection volume of the coil, small fluctuations of the magnetic inhomogeneity over the sample significantly contribute to the observed linewidth but these effects can be efficiently supressed when the T_2 is measured in a spin echo train typically implemented as a Carr-Purcell-Meiboom-Gill (CPMG) pulse train[182]. In this implementation the variation of the spacing between the spin echo pulses can be used to systematically probe fluctuations on a timescale from microseconds to milliseconds[174].

When nuclei are dipolar-coupled, they experience correlated time-dependent local magnetic fields and therefore their relaxation is coupled. This produces the NOE, in which this cross-relaxation causes transfer of spin polarisation from one nuclear spin to another. NOE is therefore the change in resonance line intensity caused by dipolar cross-relaxation between neighbouring spins with perturbed energy level populations. The steady state NOE is the population distribution of neighbouring spins that derive from two counter-acting effects. The first of these is the NOE build-up from cross-relaxation with the targeted spin, and the second is the effect of T_1 relaxation of the spins that are altered from the Boltzmann distribution due to the NOE and are returning to equilibrium.

T_1, T_2, and NOE relaxation parameters are therefore dependent on the rotational diffusive motions of the nucleus and the orientation of its chemical bonds relative to the external field. Molecular motion results in the nucleus experiencing energy fluctuations to cause transitions between the Zeeman energy levels, resulting in relaxation. This is dependent on both overall motion (the rotational correlation time, τ_m) and internal dynamic processes.

1.6 Aim of thesis

The dynamics of MHC class I will be investigated using nuclear magnetic resonance (NMR) spectroscopy. The first step is to establish reliable protein production protocols for the generation of MHC class I that is suitable for analysis by NMR spectroscopy, before backbone resonance assignment permits the investigation of plasticity via measurement of residual dipolar coupling and heteronuclear relaxation data. This data will be compared to existing analysis of MHC class I using X-ray crystallography and molecular dynamics simulations.

The aim of this research is to discover the dynamic processes within the solution state of MHC class I, and explore how they are relevant to the MHC class I function of antigen processing, particularly peptide exchange.

Chapter 2

2 Expression, purification, and characterisation of MHC class I

The first step in the analysis of the dynamics of MHC class I is the production of a protein complex that will meet the requirements of a typical NMR experiment; protein should be stable (i.e. will not degrade or undergo any form of aggregation) for several days at room temperature.

This chapter describes the production of soluble, stable MHC class I.

2.1 Introduction

There are a large variety of MHC α-chain alleles to choose between; the majority of MHC class I-related publications focus on either human or murine moieties and there are established protocols for the production of a wide range of alleles for both. These protein production methods employ several techniques, and the selection of α-chain allele will depend partly on the choice of production method.

2.1.1 Methods of protein expression

Protein expression involves the processes of transcription, translation, and post-translational modification, each of which has a number of key components. For example, transcription requires a RNA polymerase whilst translation needs a ribosome, tRNA, mRNA, and amino acids, along with other cofactors such as ATP. For some proteins post-translational modification is vital for the protein to mature; this might include the formation of disulphide bridges, glycosylation, and phosphorylation. As a result, it is common to express proteins within cells so that the required machinery for these processes is present. Cell-free expression systems, provide the machinery essential to each step but are not a cost- or time-effective way of producing the quantity of protein needed for NMR spectroscopy. A final option is the chemical synthesis of proteins, which is useful if unnatural amino acids are required but which has a low yield and high cost that makes it suitable only for relatively small polypeptides.

There are a variety of *in vivo* protein expression systems to choose between, including both prokaryotic (*E. coli*) and eukaryotic (yeast, baculovirus-infected insect cells, and mammalian cells) methods. Generally these systems begin with the transfection of cells with a DNA source containing the template, followed by culturing

the cells and allowing them to transcribe and translate their proteome that now includes the protein of interest, and then lysing the cells to obtain the expressed protein. There are a variety of methods for delivery of DNA to the cell, with the use of plasmids common for *E. coli* and yeast, baculovirus for insect cells, and lentivirus or adenovirus for mammalian cells.

One of the most common systems in use is the bacterium *Escherichia coli*, to which DNA is normally introduced using a plasmid expression vector. It is relatively simple to quickly culture *E. coli* at low cost to obtain significant overexpression of the protein of interest and yields can be very high. However, proteins can become insoluble and get segregated into inclusion bodies that can make protein recovery difficult, and bacteria do not possess the correct machinery for much post-translational modification so multi-domain proteins are often non-functional. Moving on to eukaryotic protein expression systems, yeast (often *S. cerevisiae*) permits many post-translational modifications and has the additional advantage that it can be cultured to very high cell densities. However, culturing is not as simple as for *E. coli* and the cost is higher. Insect cells are able to carry out extensive post-translational modifications and are often used to obtain correctly folded protein that is insoluble in systems such as *E. coli* and yeast, although the cost is significantly higher because of the increased demands, difficulty, and length of cell culture. Mammalian cell lines are the most expensive and complex expression system, and are only used for proteins that cannot be produced using an alternative method.

2.1.2 Methods of protein purification

Following extraction of expressed protein from the chosen expression system, the protein of interest must be purified. Cell harvesting is the separation of cells from the culture medium and usually involves centrifugation or filtration, with the exact protocol dependent on the protein expression system. During this step it is important to use the gentlest procedure possible to avoid denaturing the protein of interest. Protease inhibitors are often added to prevent the digestion of protein. The cell extract must then be purified via a second centrifugation or filtration step, usually immediately prior to chromatographic purification. It is common to use precipitation agents, de-salting columns, or fractional precipitation during this initial purification.

Chromatographic purification of a protein relies on the protein of interest possessing different properties to the rest of the sample. If the property to be used as the basis for purification is size, then size-exclusion chromatography is applied; the sample is

run over a column containing beads with pores of a set size. Large molecules cannot enter the pores and quickly elute from the column, whilst the target protein and low-molecular weight substances enter the pores. Molecules are washed from the pores such that the smallest molecules elute last. Alternatively, if the purification property is hydrophobicity then hydrophobic interaction chromatography is used. The separation involves protein binding to the hydrophobic surface of a column and this binding being reversed by alteration of salt concentration. A protein can also be purified by affinity chromatography; if the protein reversibly binds to a specific ligand (e.g. an antibody) the sample can be run over a column containing that ligand and eluted using a competitive ligand or by changing pH or ionic strength. However, affinity chromatography, size, and hydrophobicity are not always appropriate and in these situations tags are added to the target protein. A common tag is the addition of between six and ten histidine residues; histidine has a strong affinity for Ni^{2+} and hence the protein will bind strongly to a Ni-Sepharose column whilst much of the remaining sample will pass through. The protein of interest can then be eluted from the column using an increasing concentration of imidazole.

An exception to these common methods of protein purification is when protein is insoluble during protein expression in *E. coli* and is aggregated into inclusion bodies, which increases the steps required during protein purification and therefore expression conditions that produce inclusion bodies are commonly avoided. However, high yields can be obtained because over 50% of inclusion body content may be the protein of interest, and protein production can be carried out quickly and can produce proteins that are otherwise toxic to the cell. For the purification of protein from inclusion bodies, cell harvest is by centrifugation and cells are typically disrupted using mechanical lysis before being washed in detergent. Protein contained within inclusion bodies is likely to be mis-folded and hence the washed inclusion bodies are solubilised using denaturants and reducing agents before the refolding process is initiated. The conditions required for protein refolding will vary depending on the protein of interest and may require extensive screening.

2.1.3 Selection of MHC class I complex

The choice of protein expression system for MHC class I targets is relatively straightforward; *in vitro* methods are not suitable because they are not capable of producing sufficient protein for NMR spectroscopy in a cost-effective manner. Chemical synthesis is a good source for the antigenic peptide fragment as, although

relatively high-cost, alternative methods would be time-consuming. Of the *in vivo* protein expression systems, mammalian or insect cell cultures are not required because neither β_2m nor α-chain have extensive post-translational modifications and thus these methods are overly complex. This narrows the choice to yeast or *E. coli*; both permit the isotopic labelling that is required for NMR and are likely to produce good yields of protein. However, the yeast expression system requires slightly more time to culture and is slightly higher cost, and therefore *E. coli* has been chosen. *E. coli* is not capable of the N-glycosylation on α-chain Asn86, but this post-translational modification is required only during the chaperoned folding of α-chain in the ER, is not thought to have any structural effects and therefore is not required for folding in the absence of chaperones or for mature MHC class I.

A number of research groups have chosen the *E. coli* expression system for MHC class I and hence there are developed protocols for the expression and purification of β_2m and α-chain. The proteins have been expressed separately in various *E. coli* strains using various plasmid vectors, such that insoluble protein is expressed into inclusion bodies. These are purified using the above method of centrifugation, lysis, and solubilisation, before the protein is denatured and refolded to produce mature MHC class I prior to final purification via size-exclusion chromatography.

There are a number of *E. coli* strains commercially available and an appropriate one must be selected. An important consideration is codon bias; prokaryotes have a different codon usage to eukaryotes and therefore a strain that corrects for the codon bias should be used. This strain should also permit overexpression of protein; the BL21-CodonPlus(DE3)-RIPL strain fulfils these criteria and is cost-effective. An additional consideration is the maintenance of a stock of plasmid containing the β_2m or α-chain, and for production of a high quantity of DNA the DH5α strain was chosen.

DNA for β_2m and α-chain was initially obtained in separate pET-11A plasmid vectors, in which protein expression is under the control of the lac promoter. This is a promoter induced by IPTG (isopropyl β-D-1-thiogalactopyranoside), and because this induction is strong and reliable it was not necessary to alter the pET-11A plasmid to improve yield.

Having selected the protein expression method the most important decision to be made is that of the exact MHC class I moiety. There are a number of well-characterised α-chain alleles, including the human HLA-B*27[97] and the murine H-2K^b [99] and H-2D^b [98] –based complexes. In terms of protein expression, in *E. coli* human

alleles have a codon bias that typically reduces expression more than for the murine alleles. A number of experiments have shown that human and murine MHC class I are very similar in both structure and function (Section 1.4). Of the murine alleles there are more structural studies focusing on H-2D^b and these cover MHC class I formed using a variety of antigenic peptides; of great interest when considering the question of how the dynamics of MHC class I may affect the antigen presenting function. H-2K^b is better described in terms of kinetic and *in vitro* data, however the well-characterised biology and structure for H-2D^b coupled with the availability of a number of MD studies makes the murine H-2D^b allele a more sensible choice.

The next choice is therefore which of these peptides to select for characterisation; peptide must be of high affinity because the trimeric complex must be stable for several days at room temperature for NMR spectroscopy, but as discussed in Chapter 1 the presence of a lower affinity peptide may allow dynamics within MHC class I. Therefore three peptides of varying affinities have been chosen; FAPGNYPAL (FL) is a immunodominant nonameric peptide, FAPGNYPAA (FA) lacks the optimal position 9 anchor residue (L to A) and hence is expected to be lower affinity, whilst NYPAL (NL) has a truncation that removes the amino-terminal section including the position 2 anchor residue and is also expected to be of lower affinity for the α-chain than FL.

The final choice within the MHC class I complex is between murine or human β_2m. When hβ_2m is used to produce a complex with murine α-chain it increases the stability of MHC class I compared to mβ_2m. Structural studies have shown that this is because hβ_2m is capable of forming more hydrogen bonds with the underside of the peptide binding groove of the α-chain, thus stabilising the peptide-receptive, mature conformation. This does not affect the conformation of the α-chain or peptide[98,106], but it may have a small effect on the binding of CD8[183]. Since hβ_2m produces significantly higher yields of refolded protein during the purification of MHC class I and CD8 is not involved in this study it is the logical choice.

For clarity, in the remainder of the thesis 'MHC class I' refers to a generic complex whilst 'H-2D^b/FL' or 'H-2D^b/FA' or 'H-2D^b/NL' will refer to the specific individual MHC class I-peptide complex, and 'H-2D^b α-chain' refers only to the heavy-chain portion of the complex.

2.2 Materials and methods

The sequences of each MHC class I constituent can be found in Appendix B; peptides were obtained from Peptide Protein Research Ltd. Recipes for the buffers and media used can be found in Appendix C.

2.2.1 Bacterial protein expression

To establish a DNA stock to be stored at -80°C, hβ_2m and H-2D^b α-chain from separate pET-11A vectors (Novagen) were transformed into DH5α cells (Thermo Fisher Scientific) and plated onto LB (lysogeny broth) agar before overnight incubation at 37°C. DNA was then purified using a QIAprep spin Miniprep kit (Qiagen).

For protein expression, hβ_2m and H-2D^b α-chain from the separate pET-11A vectors were transformed into BL21-CodonPlus(DE3)-RIPL cells (Agilent Technologies) and plated onto LB agar before overnight incubation at 37°C.

To assess the level of protein expression, 10 ml LB cultures were each inoculated with a single colony and grown at 37°C, shaking at 170 rpm, until OD_{600} was between 0.7 and 0.8 (measured using a ThermoScientific Nanodrop 2000c Spectrophotometer). 1 ml was pelleted using a Heraeus Biofuge Pico centrifuge (13000 rpm for 4 minutes) and the remainder treated with IPTG to 1 mM and cultured for a further 3 hours before a second 1 ml was pelleted. Pre- and post-induced samples were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gel.

2.2.1.1 Culture in lysogeny broth

10 ml LB cultures were grown overnight at 37°C and 170 rpm, with colonies selected based on the expression-check SDS-PAGE result. 2 l baffled flasks were filled with 500 ml of LB and each inoculated with 10 ml of overnight culture. Following growth at 37°C and 170 rpm, flasks were treated with IPTG to 1 mM at OD_{600} 0.7-0.8. Cells were grown for a further 3 hours and pelleted using a Beckman Avanti J-20i centrifuge with a JLA-8,100 rotor at 5000 rpm for 20 minutes at 4°C, before being frozen at -20°C prior to purification.

2.2.1.2 Culture in minimal media

M9 minimal media can be used to produce a variety of desired isotopic labelled protein. For non-deuterated protein (^{13}C and/or ^{15}N), cells from overnight 10 ml LB cultures were pelleted using a Sorvall Legend RT centrifuge at 2500 rpm for 15 minutes before being resuspended in 20 ml minimal M9 media for 20 minutes. This

20 ml was then used to inoculate each 500 ml of culture media, and the cells grown as above until induction with IPTG to 1 mM at OD_{600} 0.7-0.8. Cells were cultured overnight (approximately 16 hours) before being pelleted and frozen.

In addition to isotopic labelling, minimal media protein expression can be used for selective amino acid (un)labelling. For this, *E. coli* was cultured as standard with 120 mg/l of unlabelled desired amino acid added alongside the IPTG at induction.

For perdeuteration (^2H isotopic labelling), the culture protocol required optimisation for maximal protein expression. D_2O used in media was recycled using a rotary evaporator.

2.2.2 Protein purification and refolding

Cell pellets were stored at -20°C for variable amounts of time with no impact on the protein yield.

2.2.2.1 Purification from inclusion bodies

For hβ_2m, Triton X-100 detergent was used. Per 2 g of frozen cell pellet 15 ml of Triton buffer, 150 µl 2 mg/ml DNase, 10 µl 1 ku/µl lysozyme, and $MgCl_2$ to 5 mM were added at 4°C. After homogenisation the solution was gently rocked at room temperature (RT) for 20 minutes, before centrifuging at 16000 rpm for 20 minutes at 4°C using a Beckman J2-21 centrifuge with a F21-8x50y rotor. To the resultant pellet was added a further 15 ml of Triton buffer, 150 µl DNase, 10 µl lysozyme, and $MgCl_2$ to 1 mM. After homogenisation the solution was rocked at RT for 10 minutes, then centrifuged at 8000 rpm for 20 minutes at 4°C. 15 ml of 10% Triton buffer was added to the pellet, which was homogenised and centrifuged at 10000 rpm for 20 minutes at 4°C; this washing step was repeated two further times. To the resultant pellet was added 15 ml of 10% Triton buffer, the solution homogenised, and centrifuged at 16000 rpm for 20 minutes at 4°C. 10 ml resolubilisation buffer was added to the pellet, and the solution rocked overnight at RT before a final centrifugation step of 16000 rpm for 25 minutes at 10°C to obtain a supernatant containing purified but misfolded hβ_2m.

For the washing of H-2D^b α-chain inclusion bodies, BugBuster detergent was used. Per 2 g of frozen cell pellet, 10 ml of BugBuster, 200 µl 2 mg/ml DNase, 10 µl 1 ku/µl lysozyme, and $MgCl_2$ to 1 mM were added at 4°C. After homogenisation the solution was gently rocked at RT for 30 minutes, before centrifuging at 8000 rpm for 20 minutes at 4°C. To the resultant pellet was added a further 10 ml of BugBuster and 10

µl lysozyme. After homogenisation the solution was rocked at RT for 10 minutes, then centrifuged at 8000 rpm for 20 minutes at 4°C. 30 ml of 10% v/v BugBuster was added to the pellet, which was homogenised and centrifuged at 10000 rpm for 20 minutes at 4°C. The inclusion bodies were then washed three times by addition of 30 ml 10% BugBuster, homogenisation, and centrifugation at 16000 rpm for 20 minutes at 4°C. To the resultant pellet was added 12 ml of 10% v/v BugBuster, the solution homogenised, and centrifuged at 16000 rpm for 20 minutes at 4°C. 10 ml resolubilisation buffer was added to the pellet, and the solution rocked overnight at RT before a final centrifugation step of 16000 rpm for 30 minutes at 10°C to obtain a supernatant containing purified but misfolded α-chain.

2.2.2.2 Protein refolding

The concentration of protein was determined spectrophotometrically at 280 nm via Nanodrop, and any pellet from the final centrifugation step of the purification discarded. At this stage protein is denatured by the urea solution and must be refolded into native conformation. Refolding buffer was cooled to 4°C and stirred to form a slight vortex without drawing air into solution. A fine-bore needle attached to a syringe was suspended above the buffer, and the protein solution allowed to drip through under gravity.

For refolding of hβ_2m, a single addition of protein was made to the refolding buffer such that the final concentration was less than or equal to 10 µM. The solution was stirred until the β_2m was fully dissipated, before storage at 4°C for 24 hours. For H-2D^b α-chain refolding, 4 µM β_2m (fully purified by size exclusion chromatography) and 10 µM peptide were added to the refolding buffer. Three equal additions of protein solution were made, separated by 6 to 16 hours, to a final concentration of 2 µM H-2D^b α-chain. For each addition the solution was stirred until protein was fully dispersed before storage at 4°C until the next step. After the third addition, the solution containing MHC class I was stored for 48 hours at 4°C.

2.2.2.3 Concentration and gel filtration

After refolding, the protein solution was concentrated using 10000-MWCO Vivaspin units (polyethersulphone membrane, Generon) by centrifugation at 3900 rpm at 4°C for varying lengths of time to obtain hβ_2m and MHC class I to concentrations not exceeding 4 mg/ml and 2 mg/ml respectively, as determined by Nanodrop.

A Sephadex 75 HiLoad 16/600 column (GE Healthcare) attached to an ÄKTApurifier was equilibrated with de-gassed phosphate buffered saline (PBS) buffer and 0.1%

sodium azide. Per run, up to 5 ml of 0.45 µm-filtered protein solution was loaded, and eluted at a rate no faster than 1 ml/minute. The fraction(s) containing isolated hβ_2m or MHC class I were identified by SDS-PAGE and comparison of elution profile with known standards.

hβ_2M or MHC class I was subsequently concentrated to between 4 and 8 mg/ml using clean 10000-MWCO Vivaspin units, and stored at 4°C for up to 48 hours or flash-frozen and stored at -20°C for longer periods of time.

2.2.5 Fluorescence polarisation

Fluorescence polarisation assays are generally used to provide information on the formation or degradation of protein complexes. The assay relies on the presence of fluorescence anisotropy, which is where the light emitted by a fluorophore has different intensities along planes parallel and perpendicular to the plane of excitation, such that the degree of polarisation is a ratio of the two measurements. The level of fluorescence polarisation observed is dependent on the mobility (or the rotation correlation time) of a molecule, such that binding of a small fluorophore to a larger, slowly rotating, molecule results in a high polarisation value.

The immunodominant peptide ASNKNMETM was labelled with TAMRA (tetramethylrhodamine) on the lysine side chain, with the expectation that it would displace FL, FA, or NL and become bound to the hβ_2m/ H-2D^b α-chain heterodimer if functional and stable MHC class I had been produced during the refold step of protein purification. H-2D^b/FL, H-2D^b/FA, or H-2D^b/NL at a final concentration of 0.4 µM was reacted with varying concentrations of ASNK*NMETM such that the total reaction volume was 60 µl. Reactions were carried out in a microplate at 4°C, with measurements taken at RT at various time points using an Analyst AD (Molecular Devices) with 530 nm excitation and 580 nm emission filters, and a 561 nm dichroic mirror. Dr Andy van Hateren at the Faculty of Medicine, University of Southampton kindly carried out the fluorescence polarisation assay.

2.2.4 Mass spectrometry

Mass spectrometry required at least 35 µl of 1.5 mg/ml for either isolated hβ_2M or MHC class I. A P6 acrylamide column (600 µl of acrylamide resin dissolved in water) was used to remove the 150 mM NaCl of the PBS buffer, before methanol to 50% and formic acid to 1% (v/v) were added. Readings were taken on a Waters Micromass LCT mass spectrometer equipped with an electrospray ionisation source, and the

maximum entropy technique used to determine protein mass. Mass spectrometry was kindly carried out by Dr David Staunton at the Department of Biochemistry, University of Oxford.

2.2.5 NMR spectroscopy

For preparation of NMR samples, PBS buffer was exchanged 100-fold for HEPES buffer, decreasing the level of NaCl from 150 mM to 1.5 mM to improve spectral quality; the required volume of protein stock was concentrated in 10000 MWCO Vivaspin units by centrifuging at 3900 rpm at 4°C to 1 ml, before dilution to 10 ml with HEPES buffer. This was concentrated back to 1 ml, diluted to 10 ml again, and then concentrated to 300 or 600 µl as required to obtain a 600 µM solution. All NMR spectroscopy sample buffers contained 50 mM HEPES pH 7 and 0.05% w/v sodium azide. All experiments were carried out at room temperature (RT)

The one-dimensional NMR described in this chapter was carried out by Dr Joern Werner at the University of Southampton at 600 MHz on a Varian Inova spectrometer using a triple-resonance PFG cryoprobe.

2.2.6 Data processing

NMR data was processed using nmrPipe[184], with mild resolution enhancement and zero filling in the proton dimension, and linear prediction in the nitrogen dimension. Multidimensional NMR spectra were analysed in CCPN Analysis[185]. Raw data output from CCPN (collaborative computational project for NMR) was further analysed and manipulated in Microsoft Excel; this software was also used to produce graphs. RDC data analysis was carried out using Module 1.0[186]. If required, images were manipulated using CorelDRAW X6 or Microsoft Powerpoint before importing into Microsoft Word. All molecular visualisation has utilised the program UCSF Chimera 1.8[187].

2.3 Results

The presence of hβ$_2$m and H-2D^b α-chain was verified at each stage of bacterial expression and purification using SDS-PAGE, before final characterisation by mass spectrometry and NMR.

2.3.1 Protein production

The original DNA stocks of hβ$_2$m and H-2D^b α-chain were provided by the Elliot lab (University of Southampton). These master stocks were stored at -80°C, with the Qiagen Miniprep producing working stocks of 97.5 ng/µl for hβ$_2$m and of 105.3 ng/µl for H-2D^b α-chain.

2.3.1.1 Expression of MHC class I in *E. coli*

hβ$_2$m and H-2D^b α-chain were successfully expressed in *E. coli*, as shown by SDS-PAGE of pre- and post-induction *E. coli* samples (Figure 2.1). Solubilisation of hβ$_2$m inclusion bodies resulted in yields of up to 120 mg/l for unlabelled LB cultures and of up to 90 mg/l for minimal media growth labelling with ^{13}C and/or ^{15}N. For H-2D^b α-chain the yields were lower; up to 90 mg/l for unlabelled cultures and 70 mg/l for ^{13}C and/or ^{15}N labelling.

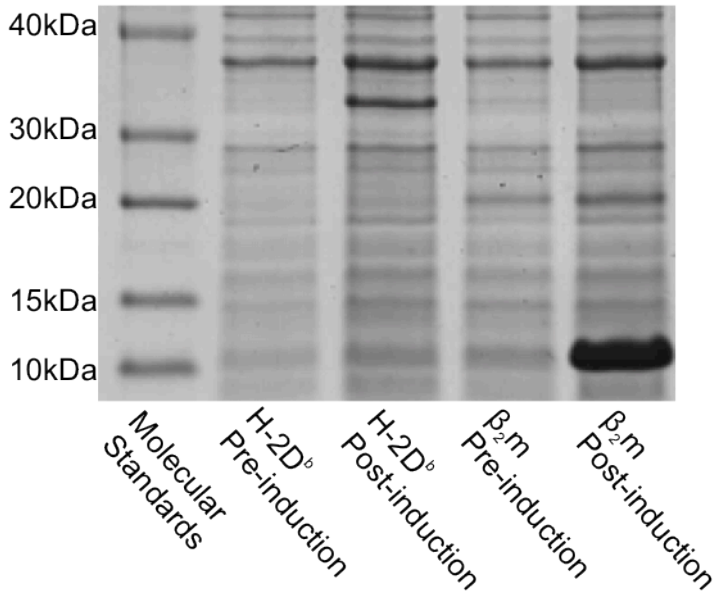


Figure 2.1 – Confirmation of hβ$_2$m and H-2D^b α-chain expression in *E. coli*.
1 ml samples from culture were pelleted at 13000 rpm and run on a 12% polyacrylamide gel alongside Novex Sharp Protein Standard, and visualised with Coomassie stain (see Appendix B). A protein band corresponding to approximately 32 kDa is present in the H-2D^b sample taken three hours after induction but not in the pre-induction sample. A large protein band corresponding to approximately 12 kDa is present in the β2m post-induction sample but not in the pre-induction sample.

2.3.1.2 Deuteration of MHC class I in *E. coli*

Bacteria require acclimatisation to deuterated media due to the isotope effect; deuterium is more electronegative than protium, and therefore forms stronger bonds that require more energy to break. Bacterial cultures demonstrated significantly slower growth rates, and on many occasions cell death was apparent. Using a trial-and-error approach a new protein expression protocol was developed. This protocol was partly developed during laboratory work by myself at the University of Southampton whilst working toward a MBioSci Master of Biomedical Sciences.

For the acclimatisation of *E. coli*, based on the expression-check SDS-PAGE colonies positive for protein expression were sequentially transferred to LB agar plates containing 50% D_2O, 75% D_2O, and 100% D_2O. They were then transferred to an agar plate containing 50% LB, 50% minimal media, 100% D_2O before a final agar plate of 100% minimal media. 100% D_2O fully acclimatises the cells. An additional expression check was carried out using this final plate.

For the culturing of *E. coli* the existing minimal media protocol was followed, but cells were allowed to grow for additional time following induction with IPTG. The rate of cell growth after induction was followed using OD_{600} measurements as a guide. Once the rate of growth had slowed, an additional 0.25 g/l of glucose (isotopically labelled as appropriate to the growth) was added. This normally occurred 16-20 hours after induction and cells were then allowed to grow for a further 8 hours, giving a total culture of time after induction of 24-28 hours. The additional glucose was found to give a significant boost to cell growth; a plateau of OD_{600}=1.550 was increased to 1.729 for $^{2}H^{15}N$-labelled hβ_2m.

Use of this protocol ensured reliable bacterial growth, but despite these improvements protein expression was still decreased compared to non-deuterated minimal media growths. Yields of up to 65 mg/ml for hβ_2m and of up to 55 mg/ml for H-2D^b α-chain were obtained.

2.3.1.3 Purification of MHC class I

Refolding of hβ_2m converted around 60% of solubilised protein to a monomeric state, with size exclusion chromatography resulting in 4 distinct peaks (Figure 2.2A). Three peaks are β_2m species with the fourth generated by residual buffer components and DNA (determined by OD_{260} and comparison with size exclusion chromatography of protein-free refold buffer). Protein molecular weight standards indicated that the third peak was monomeric hβ_2m, and this was confirmed by SDS-PAGE.

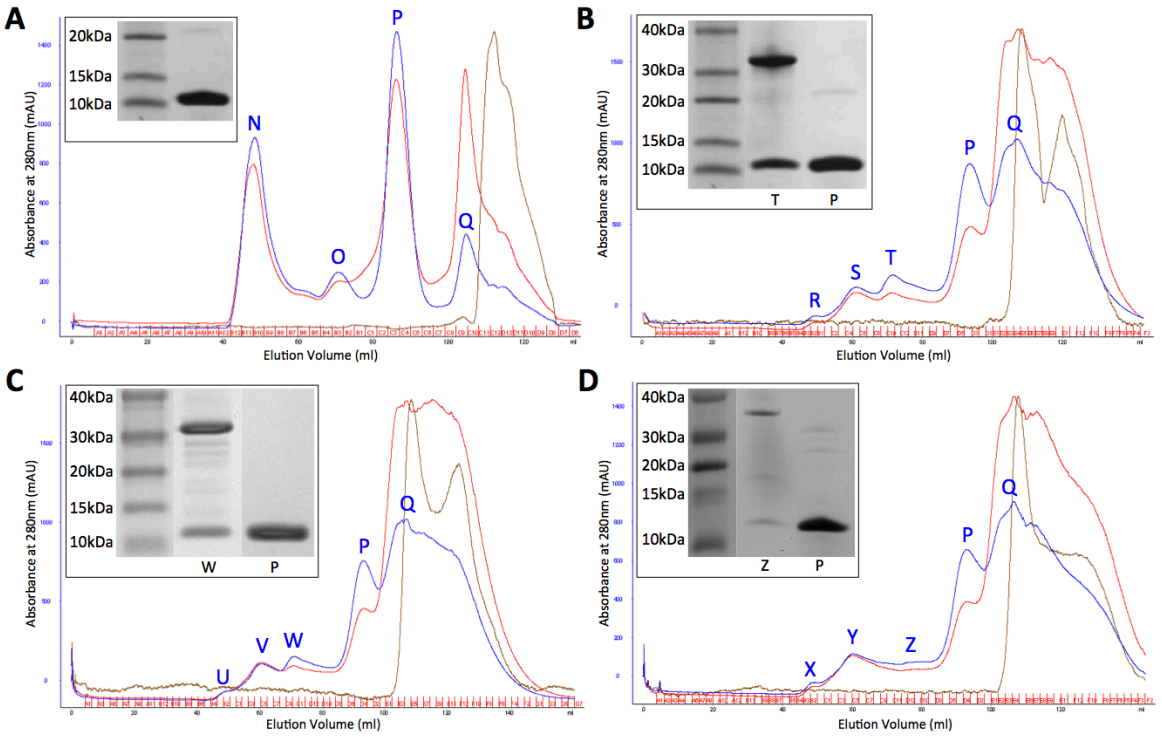


Figure 2.2 – Purification of hβ$_2$m and MHC class I.
(A) Size exclusion chromatography of hβ$_2$m following refolding. In blue is the OD_{280} (protein), in red the OD_{260} (DNA), and in brown the elution of salt. 4 peaks are present; N and O correspond to β$_2$m multimers, P to the monomeric species, and Q to elution of DNA. Inset: SDS-PAGE of P; a band is observed at approximately 12 kDa. **(B)** Size exclusion chromatography of H-2D^b/FL following refolding. R and S correspond to multimers, whilst T is the trimeric complex. Inset: SDS-PAGE of T and P; for T, bands are observed at approximately 12 kDa and 32 kDa. **(C)** Size exclusion chromatography of H-2D^b/FA following refolding. U and V correspond to multimers, whilst W is the trimeric complex. Inset: SDS-PAGE of W and P. **(D)** Size exclusion chromatography of H-2D^b/NL following refolding. X and Y correspond to multimers. The expected peak at Z, for the trimeric complex, is not immediately obvious. Inset: SDS-PAGE of Z and P.

Refolding of H-2D^b α-chain into MHC class I converted around 50% of solubilised protein to the trimeric form. Size exclusion chromatography resulted in five distinct peaks for H-2D^b/FL (Figure 2.2B) and H-2D^b/FA (Figure 2.2C), whilst only four peaks were observed for H-2D^b/NL (Figure 2.2D), with a peak observed for the excess hβ$_2$m at the same elution volume as in Figure 2.2A. For H-2D^b/FL and H-2D^b/FA, it is the third peak that corresponds to the expected elution volume for a 45 kDa molecule, with SDS-PAGE showing two bands at the correct molecular weights (peptide is too small to be observed on these gels). This peak is not present for H-2D^b/NL, suggesting this refold has failed, but the same eluted fractions were collected as for the other two complexes. SDS-PAGE showed a band for hβ$_2$m as expected but a higher than expected band (approximately 35 kDa rather than 32 kDa) for the H-2D^b α-chain.

Production of unlabelled hβ_2m had a yield of 75-100mg of refolded, monomeric protein per litre of bacterial culture, H-2D^b/FL a yield of up to 25 mg/l. Labelling with ^{15}N resulted in a drop to approximately 60 mg/l and 20 mg/l respectively, whilst a ^{13}C^{15}N scheme resulted in around 40 mg/l and 15 mg/l respectively. Deuteration resulted in yields of up to 30 mg/l for β_2m and 8 mg/l for H-2D^b/FL.

2.3.2 Protein characterisation

The first step in selecting a stable MHC class I complex was to characterise the complexes refolded using the FL, FA, and NL peptides. This was done using fluorescence polarisation, an assay measuring decrease in concentration over time, mass spectrometry, and finally one-dimensional NMR spectroscopy.

The fluorescence polarisation assay suggested that the MHC class I complexes were functional peptide selectors but that H-2D^b/NL may be unstable. At low concentrations of tracer peptide (Figure 2.3A) all three complexes were able to bind; the slowest exchange was observed for H-2D^b/FL whilst H-2D^b/FA and H-2D^b/NL exchanged at approximately the same rate, suggesting that H-2D^b α-chain has a much higher affinity for FL. As the concentration of tracer peptide increases, H-2D^b/FL remains the slowest exchanger but H-2D^b/NL becomes almost as slow (Figure 2.3B). This is not an indication of stability but rather that the number of available MHC complexes for H-2D^b/NL is lower than for H-2D^b/FL or H-2D^b/FA; it is likely that H-2D^b/NL is experiencing aggregation. As the concentration of tracer peptide increases further it begins to reach the point at which ASNK*NMETM has displaced all original peptide (Figure 2.3C). The maximum value for H-2D^b/FA is now higher than that of either H-2D^b/FL or H-2D^b/NL, suggesting that it has more peptide-receptive complexes. However, H-2D^b/FL and H-2D^b/FA had similar plateau levels in lower tracer concentrations. As the concentration of ASNK*NMETM becomes far in excess (Figure 2.3D) the reaction can no longer be followed due to poor signal-to-noise.

The stability of the MHC class I complexes over time was assessed using concentration as a guide. By centrifuging the protein solution prior to measuring the concentration, insoluble aggregates were removed. This method will not remove soluble aggregates that may form as an intermediate between soluble native protein and insoluble aggregates, and therefore concentration is an indirect assessment of protein stability. This assay suggested that H-2D^b/FL is the most stable MHC class I complex at both 4°C and at RT (Figure 2.4). There is increased discrimination between H-2D^b/FA and H-2D^b/NL, with the former more stable than the latter.

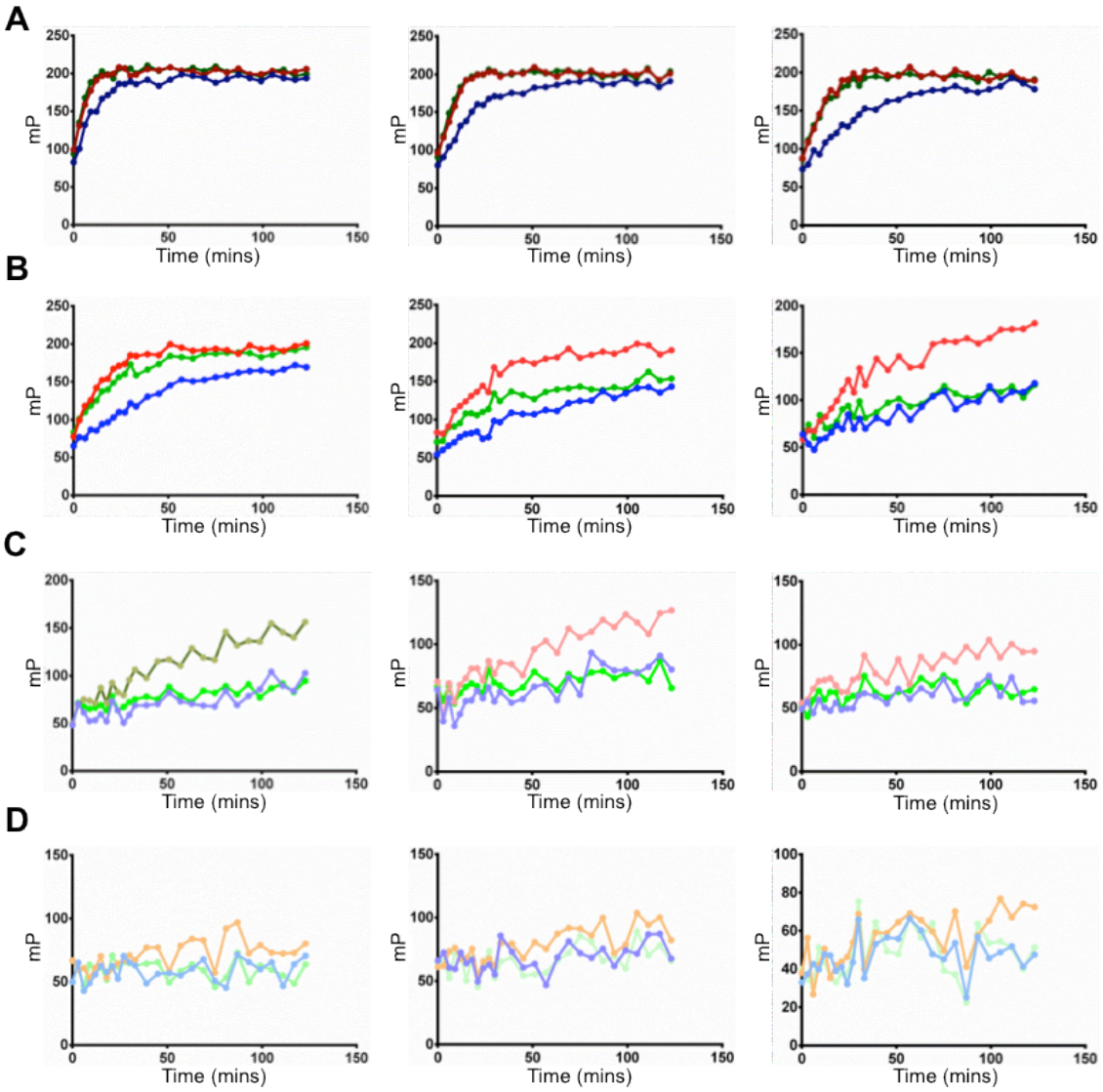


Figure 2.3 - Characterisation of MHC class I by fluorescence polarisation.
(A) Fluorescence polarisation data at low concentrations of ASNK*NMETM tracer peptide. H-2D^b/FL is shown in blue, H-2D^b/FA in red, and H-2D^b/NL in green, with ASNK*NMETM concentrations of 0.01, 0.02, and 0.04 µM left to right. **(B)** Fluorescence polarisation data at low intermediate concentrations of tracer peptide. H-2D^b/FL is shown in blue, H-2D^b/FA in red, and H-2D^b/NL in green, with ASNK*NMETM concentrations of 0.08, 0.16, and 0.32 µM left to right. **(C)** Fluorescence polarisation data at high intermediate concentrations of tracer peptide. H-2D^b/FL is shown in blue, H-2D^b/FA in dark green and red, and H-2D^b/NL in green, with ASNK*NMETM concentrations of 0.48, 0.72, and 1.08 µM left to right. **(D)** Fluorescence polarisation data at high concentrations of tracer peptide. H-2D^b/FL is shown in blue, H-2D^b/FA in orange, and H-2D^b/NL in green, with ASNK*NMETM concentrations of 1.62, 2.43, and 3.65 µM left to right. At higher concentrations of ASNK*NMETM the polarisation (mP) declines due to the increase in non-bound tracer which is rotating rapidly and thus decreasing observed polarisation values.

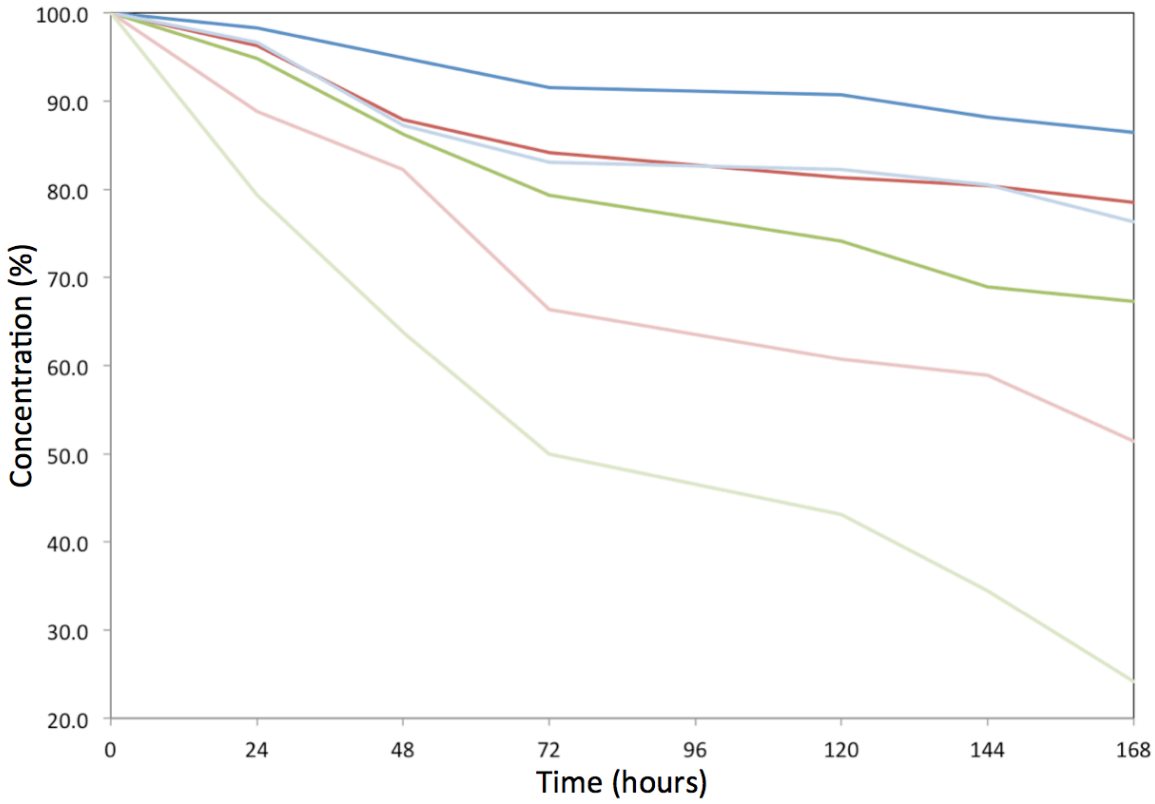


Figure 2.4 – Decrease in concentration of MHC class I complexes over time.
Concentration of H-2D^b/FL (blue), H-2D^b/FA (red), and H-2D^b/NL (green) is displayed as a percentage of the concentration at 0 hours. Concentration was measured at 24 hour intervals over seven days for protein solutions at 4°C (dark lines) and at RT (light lines). Aggregates were removed by centrifugation and concentration measured via Nanodrop.

MHC class I was also characterised using mass spectrometry along with hβ_2m, which was confirmed as the correct mass (Figure 2.5A). Analysis of H-2D^b/FL revealed all three components of the MHC class I complex to be present at the correct respective masses (Figure 2.5B). H-2D^b/FA did not appear to contain the FA peptide (Figure 2.5C). This suggests it has undergone degradation via loss of peptide; the fluorescence polarisation assay showing H-2D^b/FA with peptide selector function was conducted within 24 hours of the protein refolds whilst the mass spectrometry could only be undertaken after 5 days, during which time protein was stored at 4°C and the concentration assay revealed a 20% drop in soluble protein. H-2D^b/NL also did not appear as a trimeric MHC class I complex, with only hβ_2m present (Figure 2.5D). Although a 25% drop in soluble protein was recorded over 5 days, it is clear that there must be soluble as well as insoluble aggregates. It is likely that both H-2D^b/NL and H-2D^b/FA degraded via loss of peptide, resulting in dissociation of hβ_2m. For H-2D^b/NL this may have occurred more rapidly, thus allowing H-2D^b α-chain to aggregate and become insoluble. This explains the observation of both H-2D^b α-chain and hβ_2m for H-2D^b/FA but only hβ_2m for H-2D^b/NL.

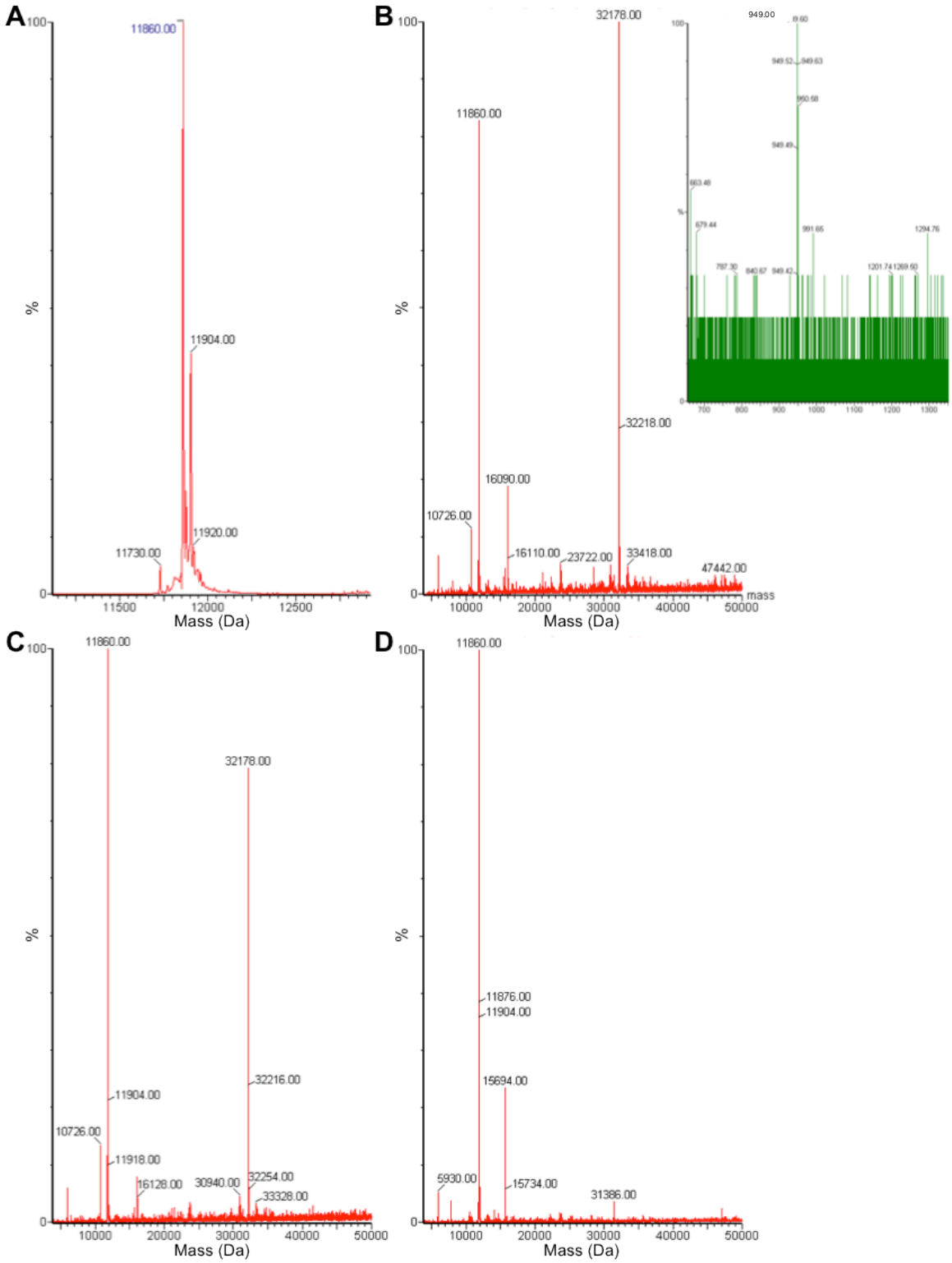


Figure 2.5 – Characterisation of hβ$_2$m and MHC class I by mass spectrometry.
(A) Electrospray ionisation mass spectrometry of hβ$_2$m. The expected molecular mass of hβ$_2$m is 11862 kDa, within 2 kDa error of the measured 11860 Da. **(B)** Electrospray ionisation mass spectrometry of H-2D^b/FL. A peak is seen for hβ$_2$m. The 32178 Da peak for H-2D^b α-chain peak matches the expected 32179 Da. Inset: FL peptide is also observed at 950 Da versus the expected 949 Da. **(C)** Electrospray ionisation mass spectrometry of H-2D^b/FA. Peaks are observed for hβ$_2$m and H-2D^b α-chain. FA peptide at the expected 907 Da could not be located. **(D)** Electrospray ionisation mass spectrometry of H-2D^b/NL. A peak is present for hβ$_2$m but there appears to be no H-2D^b α-chain present. The NL peptide at an expected 577 Da also could not be observed.

It is clear that H-2D^b/FL is a stable MHC class I complex and therefore one-dimensional NMR spectroscopy was performed on 600 µl, 600 µM samples of hβ_2m and H-2D^b/FL. Both proteins produced spectra containing sharp and distinct peaks. These occurred in those areas expected for the various protein constituents (hβ_2m; Figure 2.6A), and were dispersed beyond the range of values that would be produced by a random coil formation (H-2D^b/FL; Figure 2.6B). This indicates that protein is folded, and along with the mass spectrometry data, suggests that H-2D^b/FL is a trimeric MHC class I complex.

2.4 Summary

This chapter has described the successful expression of both hβ_2m and H-2D^b α-chain in *E. coli*, using LB media and M9 minimal media, as well as the development of a protocol for the expression of deuterated protein. The protein is expressed in inclusion bodies, which can be purified via homogenisation with detergent, before the protein is refolded. This method has produced isolated, monomeric hβ_2m, as well as H-2D^b/FL, H-2D^b/FA, and H-2D^b/NL MHC class I complexes.

This chapter also describes the initial characterisation of hβ_2m and the three MHC class I complexes using methods including fluorescence polarisation, mass spectrometry, and NMR spectroscopy. This has revealed that H-2D^b/FL is a stable protein complex that is suitable for further investigation. Further research in this thesis will focus solely on H-2D^b/FL MHC class I.

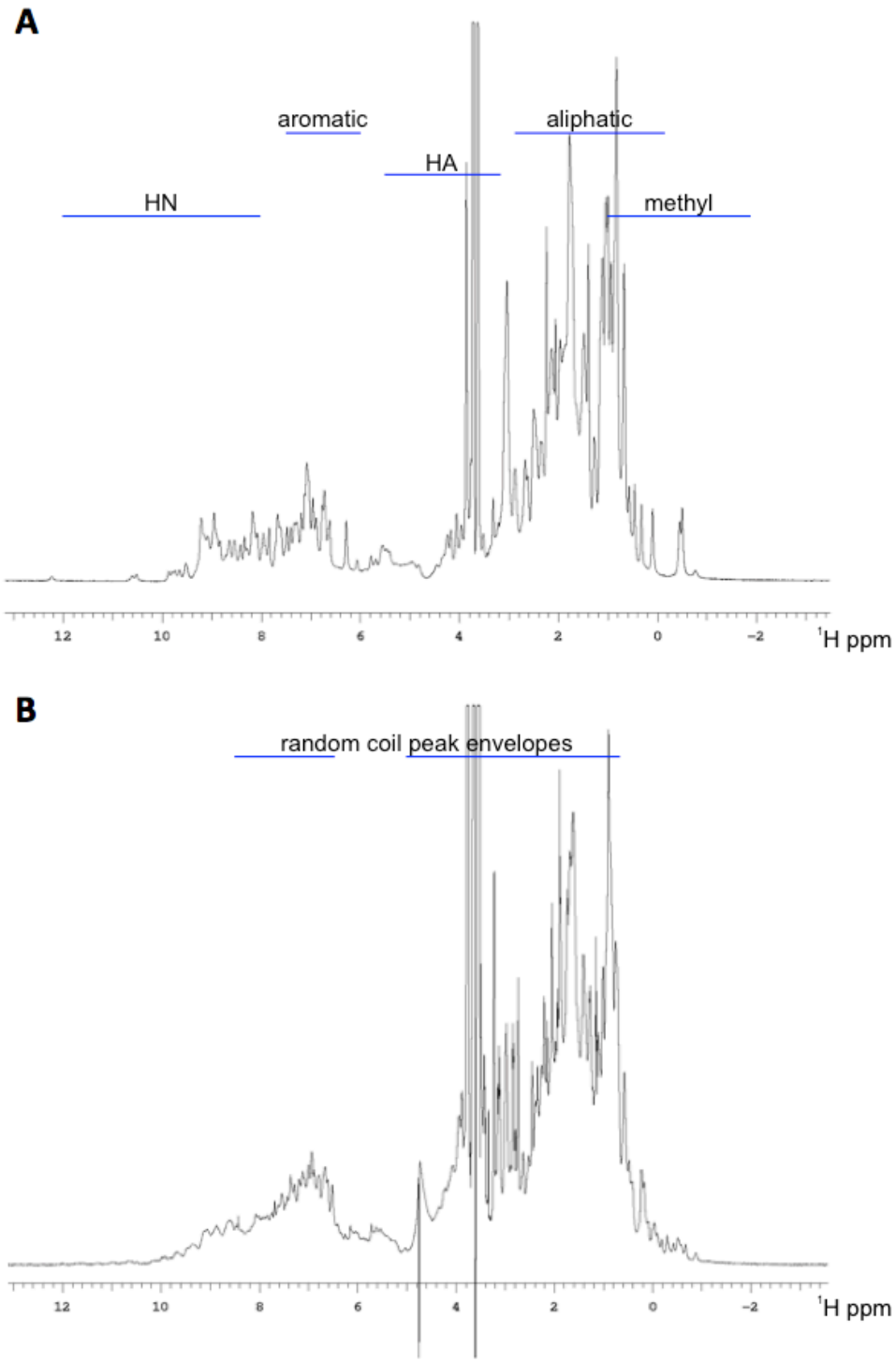


Figure 2.6 – Characterisation of hβ₂m and H-2Dᵇ/FL by 1D NMR spectroscopy.
(A) ^{1}H-NMR spectrum of hβ$_2$m. Location of different protein constituents indicated in blue. **(B)** ^{1}H-NMR spectrum of H-2Dᵇ/FL. Peaks are dispersed beyond random coil range (blue).

Chapter 3

3 Backbone resonance assignment for H-2D^b/FL

The process of signal assignment in NMR spectroscopy is the determination of the chemical shifts for various nuclei within a protein. Backbone resonance assignment resolves the chemical shifts along the peptide bond, providing hydrogen and nitrogen resonances for the amide group, as well as carbonyl and alpha carbon resonances (beta carbon chemical shifts are also commonly measured). The chemical shift values collected during the assignment process are one of the fundamental outputs of NMR spectroscopy.

This chapter describes the backbone resonance assignment of both the hβ$_2$m and α-chain components of H-2D^b/FL.

3.1 Introduction

The MHC class I complex is 45 kDa and is therefore challenging to investigate using NMR spectroscopy. As discussed in Section 1.5.3, there are several techniques that can be used during protein expression and during data collection to address this challenge. However, due to the number of residues within H-2D^b/FL (385, of which 363 are NMR-observable) it is likely that there will be areas of the spectrum with significant overlap between peaks. The decision was therefore made to reduce the number of peaks by not employing isotopic labelling on the FL peptide during data collection for the purpose of assignment.

It is possible to further deconstruct the problem using differential isotopic labelling to separate the hβ$_2$m and H-2D^b α-chain resonances; the HNCO pulse sequence is selective for ^{13}C and hence if one part contains ^{13}C but the other does not, only one will be visible. Using this technique, it is logical to begin the assignment process with hβ$_2$m because it has only 94 NMR-observable residues compared to 261 in the H-2D^b α-chain. In addition, crystal structures suggest little change to the secondary structure of hβ$_2$m between the isolated and MHC-bound forms (Figure 1.4); it is likely that there will be only minor alterations to the majority of chemical shifts, thus making the assignment easier because the HSQC spectrum has previously been assigned for isolated hβ$_2$m.

3.2 Resonance assignment for β_2m

The backbone residues of isolated hβ_2m have been 87% assigned by Dr Garrick Taylor (University of Southampton)[188]. For the 100-residue construct (identical to that used in this thesis; Appendix A), 94 backbone NH resonances were expected in a ^{1}H-^{15}N-HSQC spectrum since the five proline residues and the N-terminal methionine are not detected. 88 peaks were observed, of which 68 were assigned. The missing residues were 1-4, 12-13, 28-31, 33-36, 55-65, and 88, the majority of which belong to the B-C and D-E loops. The HSQC spectrum had an excellent signal to noise ratio, suggesting these cross peaks were absent due to exchange broadening.

3.2.1 Materials and methods

H-2D^{b}/FL was expressed in *E. coli* with a labelling scheme of triple labelled (^{2}H^{13}C^{15}N) hβ_2m and dually labelled (^{2}H^{15}N) H-2D^{b} α-chain, and purified as described in Chapter 2 to produce a 600 µl, 600 µM sample for NMR spectroscopy. Three spectra were collected; 3D HNCA and both 2D and 3D HNCO (the differential isotopic labelling meant that only the β_2m component of H-2D^{b}/FL would be observed on the 2D HNCO). Experiments were performed on a Varian Inova 800 MHz spectrometer equipped with a 5 mm triple resonance (HCN) PFG cryogenic probe by Dr Geoff Kelly at the MRC Biomedical NMR Centre, National Institute for Medical Research, London. This data was collected and analysed by myself at the University of Southampton whilst working toward the MBioSci Master of Biomedical Sciences qualification.

3.2.1.1 NMR spectroscopy

The HNCA is a triple resonance, three-dimensional experiment that correlates the backbone NH group with the alpha carbon of the same and previous residues (Figure 1.7A). It is so named because it passes magnetisation from the amide H^{N}, to the nitrogen, and then to the alpha carbon; H-N-Cα, before it returns along the same pathway. It can be visualised as a series of ^{1}H-^{13}C two-dimensional spectra taken at a number of ^{15}N shifts, in which the linkage of the H^{N} nucleus with the two Cα signals results in two cross peaks correlated to a single NH group. For the i residue (i.e. the same residue as the nitrogen nucleus) there is a direct bond to the Cα and the signal will appear more intense than that of the indirectly bonded i-1 Cα. For a discussion of the pulse sequence and a description of the magnetisation transfer pathway using product operator notation, see Kay, Ikura, Tschudin, and Bax[189]. A TROSY-modified pulse sequence[190] was used due to the size-related slow tumbling of the protein.

The HNCO is similar to the HNCA experiment but correlates the backbone NH with the carbonyl of the previous residue, passing magnetisation from H^N to the nitrogen of residue i, and then to the carbonyl of i-1 (H-N-C^O). Although the usual choice would be a HN(CA)CO experiment, this is more complex (Figure 1.7C) than the HNCO (Figure 1.7D) and therefore the HNCO is better suited for large proteins when obtaining the Co chemical shifts. For a discussion of pulse sequence and operator-based magnetisation transfer pathway for the HNCO experiment, see Kay, Ikura, Tschudin, and Bax[189]. To further improve sensitivity due to the size of MHC class I, a TROSY-modified HNCO sequence was used[190]. The HNCO experiment produces three-dimensional data that can be visualised as a number of ^{1}H-^{13}C two-dimensional spectra taken at varying ^{15}N resonances. Acquiring a HSQC or TROSY spectrum is not desirable in this situation since the aim is to identify only those peaks belonging to hβ_2m and these experiments would produce a fingerprint spectrum containing peaks from both β_2m and H-2D^b α-chain. In comparison, a ^{1}H-^{15}N HNCO plane will contain signal only from β_2m residues that are ^{13}C-labelled.

3.2.1.2 Analysis of secondary structure

It is possible to use chemical shift data to predict the secondary structure of a protein by comparing the experimental NMR data to random coil chemical shifts, which are the values obtained for short peptides in which a residue is free to access all sterically allowed regions of conformational space. This can be achieved by measuring the chemical shift of residue X in a Gly-Gly-X-Y-Gly-Gly sequence, which is short enough to remain unstructured. Residue Y can be varied to investigate nearest-neighbour effects[191]. Chemical Shift Index 2.0 software (available via the web server)[192] utilises this data and outputs a decision on whether the residue in question is in α-helix, β-strand, or coil conformation.

Chemical shift data can also be used to infer the backbone conformation via the phi and psi angles. The backbone conformation of a protein is specified by the angles of internal rotation about the polypeptide backbone; for each residue within a protein, rotation about the N-Cα-C bond describes the phi angle and rotation about the Cα-C bond describes the psi angle. The phi and psi backbone torsion angle were inferred using TALOS-N[193].

3.2.2 Results

The ^{1}H-^{15}N plane of the trHNCO experiment for hβ_2m bound by H-2D^{b} α-chain has well-dispersed peaks, indicating a correctly folded protein. The three-dimensional trHNCO and trHNCA experiments yielded spectra with a good signal to noise ratio that had sufficient resolution for the assignment process.

3.2.2.1 Backbone resonance assignment for hβ_2m

The trHNCO plane contained 92 of the expected 95 signals (there are 5 prolines and the N-terminal methionine is not observable) and 91 of these were assigned (Figure 3.1, chemical shifts listed in Appendix C). The three unassigned residues are Ile1, Phe56, and Ser88; the otherwise excellent signal to noise ratio makes it likely that these peaks are absent due to exchange broadening.

The three dimensional trHNCA spectrum was arranged into two dimensional ^{1}H-^{13}C sections at every ^{15}N resonance (strip plots). In this visualisation, each strip contains two peaks, which permitted the sequential linking of strips by stepping from strong peaks (residue i Cα) to weaker peaks (residue i-1 Cα in the next strip); Figure 3.2A. This also permitted linking of the corresponding strips from the three-dimensional trHNCO spectrum (Figure 3.2B). The assignment of isolated hβ_2m was used as a guide during this linking process, and although some peaks have migrated on the fingerprint ^{1}H-^{15}N HNCO plane for H-2D^{b} α-chain-bound hβ_2m with respect to the HSQC of isolated β_2m, most remained in very similar locations, permitting use of the assigned isolated spectrum to suggest residues that may produce a link in the HNCA spectrum. Despite the short magnetisation pathway of the HNCA experiment producing strips on which one peak is clearly stronger than the other, some ambiguity can still arise during cross peak assignment: the chemical shifts of Cα nuclei are not unique for each amino acid. For example Gln and Glu both have an average shift of around 57 ppm, with further uncertainty arising due to variation caused by the local chemical environment. As a result, several trHNCA strips must be linked together before a conclusive assignment can be made, although residues with distinctive Cα chemical shifts such as Gly (low at around 45 ppm) are of great help.

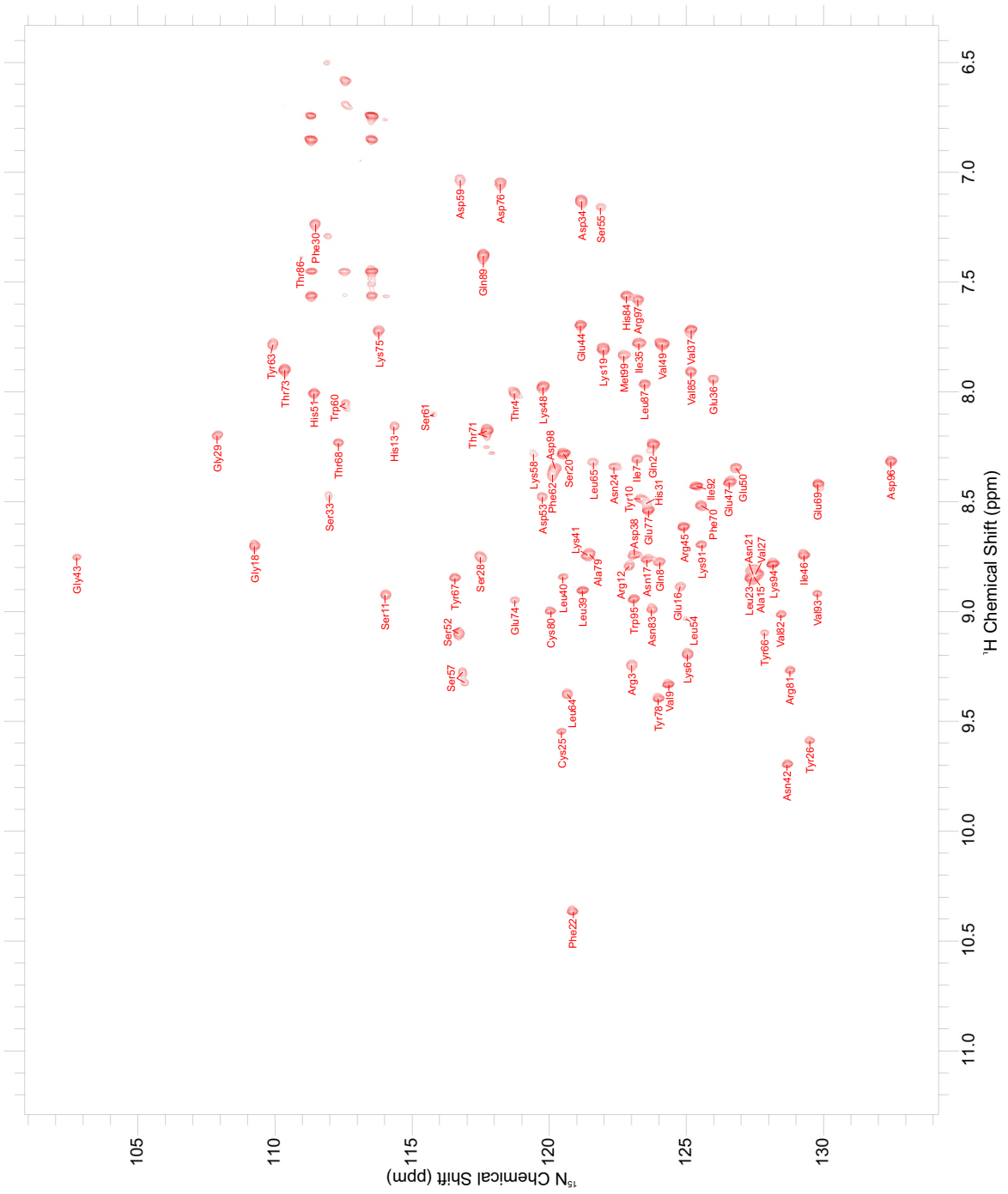


Figure 3.1 – Backbone resonance assignment for the hβ$_2$m component of H-2D^b/FL.
This trHNCO plane selects only the peaks of β$_2$m since this moiety is the only part of the complex that is ^{13}C-labelled. Neither NH_2 groups (top right) nor indole NH groups (bottom left) are assigned.

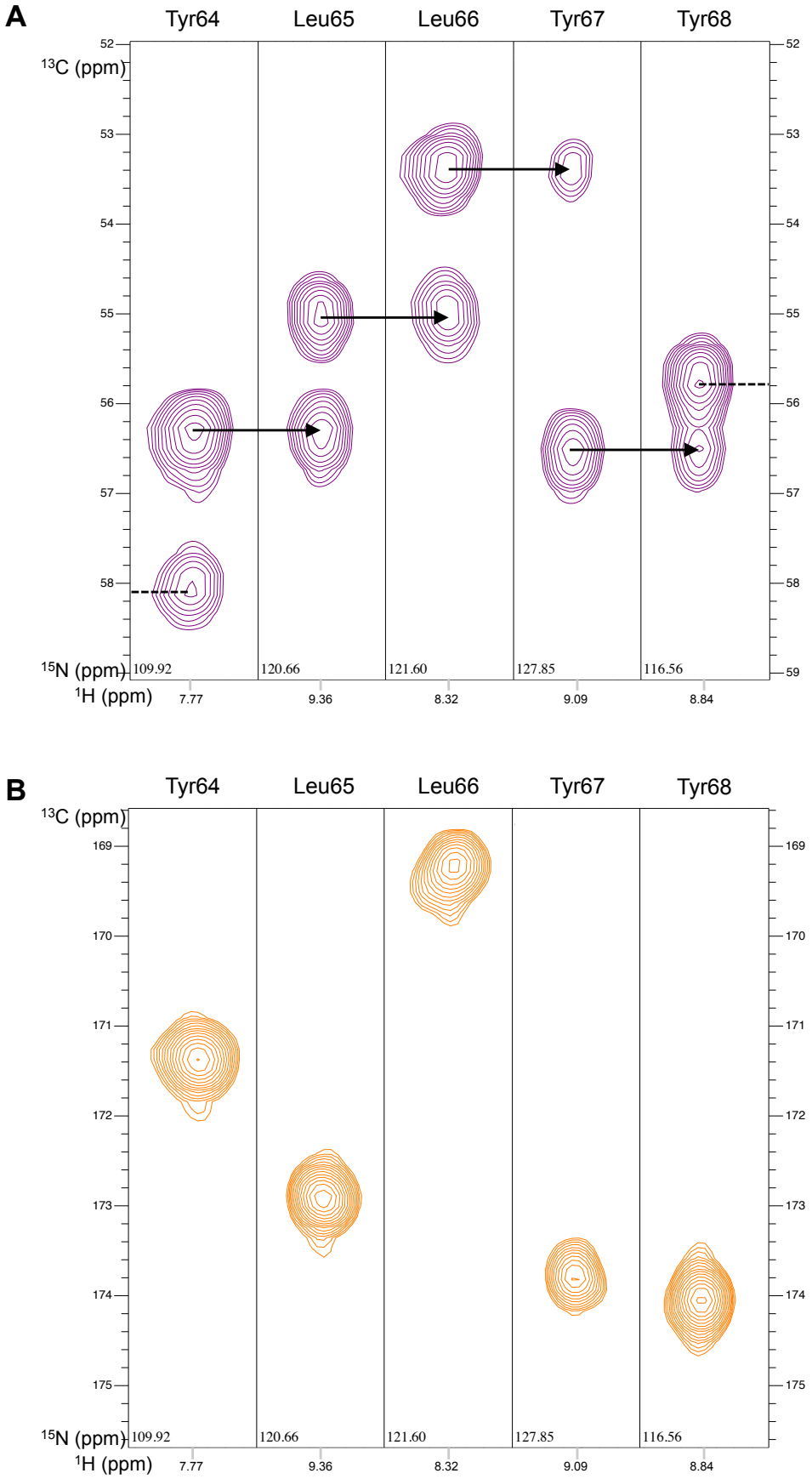


Figure 3.2 – Resonance assignment of backbone carbon nuclei of the hβ$_2$m component of H-2D^b/FL.
(A) Representative trHNCA strips for sequential residues Tyr64 through Tyr68. Unbroken arrows indicate peak linkages, broken lines the continuation of sequential assignment. **(B)** Representative trHNCO strips for sequential residues Tyr64 through Tyr68. These contain only the carbonyl resonance of the previous residue.

3.2.2.2 Prediction of secondary structure

The experimentally derived chemical shifts of $h\beta_2m$ bound by H-2D^b α-chain were analysed using Chemical Shift Index 2.0 software (Figure 3.3A). The prediction shows the expected β-sheet structure consisting of seven β-strands, albeit with an extended E strand. This is likely due to β-strand being very similar to the extended coil conformation, both in chemical shift and amino acid composition. Indeed, the authors of CSI 2.0 found ambiguity between coil and β-strand populations[192], with misidentification occurring mainly at the borders of β-strand regions. For this reason, the prediction of secondary structure acts only as a guide, although it is interesting that the misidentification occurs in the D-E loop that sits directly beneath secondary structure within the H-2D^b α-chain. This prediction suggests that the $h\beta_2m$ bound by H-2D^b α-chain has folded into the correct conformation i.e. it is consistent with X-ray crystallography structures of bound $h\beta_2m$.

The second method of secondary structure analysis is via prediction of backbone torsion angles using TALOS-N software (Figure 3.3B). This predicts the majority of residues in the β-sheet region of the Ramachandran. Outlying values (those in the α-helical regions around psi -45°) belong to loop residues that could be expected to return erroneous estimates due to their lack of secondary structure. An alternative explanation is perturbation due to the presence of structural dynamics compromising the TALOS-N predictions (the software assumes a chemical shift originating from a predominant conformation); this will be investigated further in later chapters.

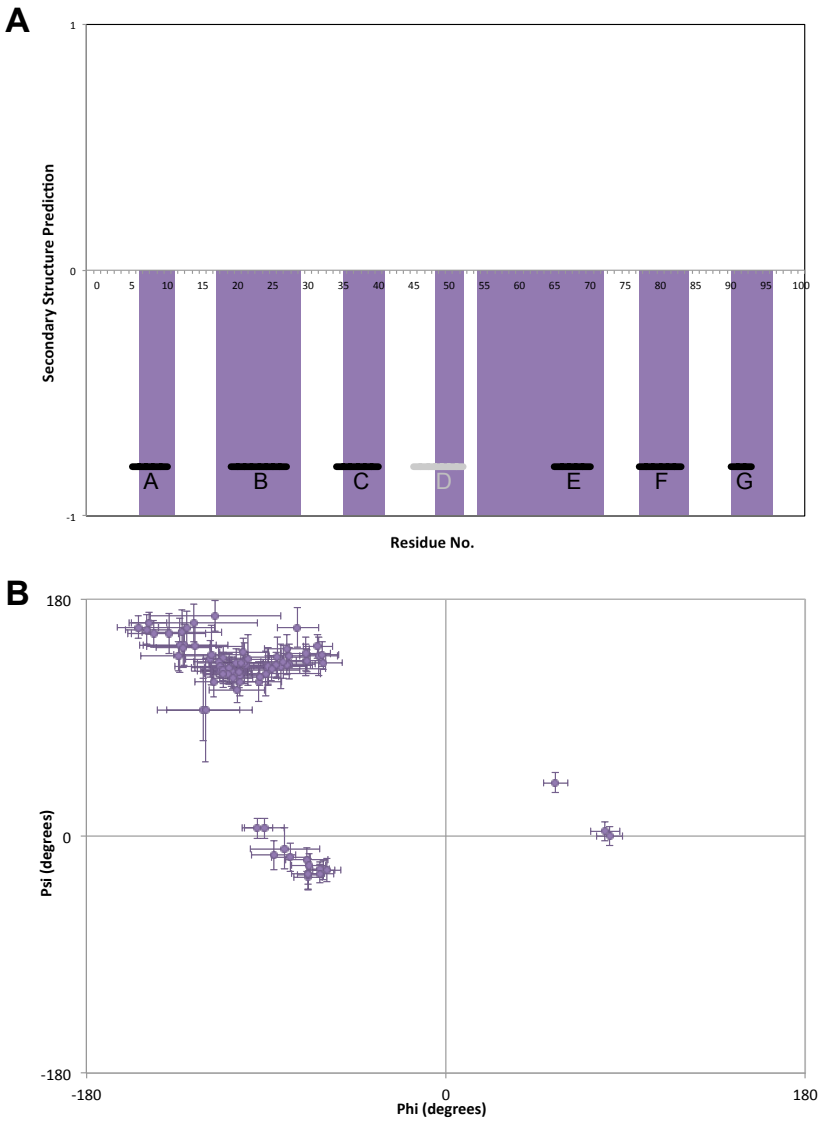


Figure 3.3 – Prediction of secondary structure for the hβ$_2$m component of H-2D^b/FL.
(A) Assigned Cα, CO, N, and HN chemical shift values are compared to random coil values using CSI 2.0; a value of +1 predicts α-helix and -1 predicts β-strand, whilst a value of 0 indicates no secondary structure. Chemical shifts are corrected for the presence of deuterium[194]. Crystal structure β-strands are indicated by solid black lines labelled with strand number, with the non-canonical D strand in grey. **(B)** Prediction of phi and psi angles from chemical shift data using TALOS-N; indicated errors are standard deviation.

3.3 Resonance assignment for H-2D^b α-chain

Following the successful backbone resonance assignment of hβ$_2$m, the H-2D^b α-chain was targeted. The process was simplified by the ability to use a high-resolution TROSY fingerprint spectra that could be streamlined by the immediate classification of 92 peaks as belonging to hβ$_2$m, rather than the lower-resolution trHNCO plane as used for the hβ$_2$m assignment.

3.3.1 Materials and methods

H-2D^b/FL was expressed in *E. coli* with a labelling scheme of triple labelled (^{2}H^{13}C^{15}N) H-2D^b α-chain and dually labelled (^{2}H^{15}N) hβ$_2$m, and purified as described in Chapter 2 to produce 600 µl, 600 µM samples for NMR spectroscopy. This sample was used to acquire TROSY, trHNCA, and trHNCACB spectra. The TROSY spectra of the H-2D^b α-chain of MHC class I had well-dispersed cross peaks, indicating correctly folded protein. The three-dimensional trHNCA and trHNCACB experiments yielded spectra with a good signal to noise ratio that had sufficient resolution for the assignment process. Both the TROSY and trHNCACB experiments were performed on a Varian Inova 800 MHz spectrometer equipped with a 5 mm triple resonance (HCN) PFG cryogenic probe by Dr Geoff Kelly at the MRC Biomedical NMR Centre, National Institute for Medical Research, London. The trHNCA experiment was performed at 900 MHz on a Varian spectrometer equipped with a 5 mm HCN z-PFG cryogenic probe by Dr Sara Whittaker at the Henry Wellcome Building for Biomolecular NMR Spectroscopy at the University of Birmingham.

Additional versions of isotopically labelled H-2D^b/FL were expressed; in addition to the collection of 3D triple resonance experiments it is possible to further the assignment of a protein using selective amino acid labelling, in which only a single residue is labelled with a NMR-observable isotope. For reasons of cost, selective un-labelling can be used to achieve a similar result, producing a spectrum in which all peaks are observed except those originating from the chosen amino acid. This is particularly useful for larger proteins such as MHC class I, but the biosynthesis pathway of amino acids in *E. coli* can limit the selection of residues. In order to minimise the dilution of label into other amino acids (scrambling), it is preferable to use amino acids that lie at the end of a non-reversible pathway; limiting the selection to Arg, Cys, His, Ile, Leu, Lys, Met, Phe, Pro, Trp, and Tyr[195]. Despite this, some scrambling can occur if the culture growth continues for a significant length of time. For this reason, the unlabelled amino acid is added to the growth medium at the point

of induction of protein expression, and the time until protein harvest is limited to less than five hours. Two TROSY spectra were collected for 200 μM and 300 μM H-2D^{b}/FL containing ^{15}N-H-2D^{b} α-chain (unlabelled hβ_2m and peptide) selectively unlabelled for Lys and Tyr respectively. Experiments were performed at 700 MHz on a Bruker spectrometer equipped with a 5 mm HCN PFG cryoprobe with a cooled ^{13}C preamplifier by Dr Alain Oregioni at the MRC Biomedical NMR Centre, National Institute for Medical Research, London.

Additional spectra that had been collected by previous members of the research group were also used. A three-dimensional ^{1}H-^{1}H ^{15}N-edited NOESY spectrum was collected at 900 MHz on a Varian spectrometer using a 5 mm HCN z-PFG cryogenic probe at the Henry Wellcome Building for Biomolecular NMR Spectroscopy at the University of Birmingham, using 600 μM H-2D^{b}/FL containing ^{2}H^{15}N-hβ_2m and ^{2}H^{13}C^{15}N-H-2D^{b} α-chain, produced by Sina Reckel (University of Southampton).

An additional TROSY spectrum for 640 μM H-2D^{b}/FL (^{15}N-H-2D^{b} α-chain selectively unlabelled for Arg) was collected at 800 MHz on a Varian spectrometer using a 5 mm HCN z-PFG cryogenic probe at the Henry Wellcome Building for Biomolecular NMR Spectroscopy at the University of Birmingham, protein sample produced by Ilona Goralyczyk (University of Southampton).

3.3.1.1 NMR spectroscopy

The HNCACB is a triple resonance, three-dimensional experiment that correlates the backbone NH group with the α and β carbons of the same and previous residues. It is so named because it passes magnetisation from the amide H^{N} to the nitrogen, and then to the alpha carbons and beta carbons; H-N-Cα-Cβ, before it returns along the same pathway. The spectrum is arranged into two dimensional ^{1}H-^{13}C strip plots at every NH cross peak, in which the linkage of the H^{N} nucleus with the two Cα and two Cβ results in four peaks correlated to a single nitrogen. For the i residue the Cα and Cβ signals will appear relatively more intense than the indirectly bonded i-1 Cα and Cβ. Furthermore, when magnetisation runs along the same pathway for residue i-1, the strip produced will contain a Cα for the i-2 residue but also a Cα for i-1. This i-1 Cα will have the same carbon shift as the original strip for the i residue, thus providing a sequential link. The experiment is not as sensitive as the HNCA experiment due to the additional magnetisation transfer steps; however the acquisition of Cβ chemical shifts is extremely useful during backbone resonance assignment due to their dispersion and dependence on amino acid type. A TROSY-

modified HNCACB pulse sequence was used[190] to improve the sensitivity. For a discussion of the pulse sequence and a brief description of magnetisation transfer using product operators, see Yamazaki *et al*[196].

The ^{1}H-^{1}H ^{15}N-edited NOESY experiment can be thought of as a combination of two dimensional nuclear Overhauser effect and ^{15}N HSQC experiments. In the former, cross peaks are due to nuclei that interact through dipolar couplings. The signal is therefore strongly dependent on the internuclear (through-space) separation of two nuclei. A ^{1}H-^{1}H NOESY transfers magnetisation between all hydrogen nuclei within about 5 Å separation. The ^{1}H-^{15}N HSQC transfer then filters (and edits) the signal through the ^{1}H-^{15}N moieties such that only NOEs that are transferred to the NH groups are detected. In the case of a fully deuterated protein the scarcity of protons leads to an increase of the distance for which NOEs can be detected. The ^{1}H-^{1}H NOESY experiment can be visualised as a series of ^{1}H-^{1}H two-dimensional spectra taken at a number of ^{15}N shifts. Sequential residues will produce a strong cross-peak, in which the proton signal on the diagonal for residue i will match a peak in the i-1 strip and vice versa. A TROSY-modified H-H NOESY pulse sequence was used to improve the sensitivity for the three dimensional data collected; for a discussion of the pulse sequence and magnetisation transfer, see Meissner and Sorenson[197].

3.3.1.2 Analysis of secondary structure

As for the hβ_2m of H-2D^{b}/FL, the secondary structure of H-2D^{b} α-chain was predicted using both comparison to random coil chemical shifts via CSI 2.0, and prediction of backbone torsion angles via TALOS-N.

3.3.2 Results

The various TROSY spectra collected for H-2D^b/FL, containing peaks from both the H-2D^b α-chain and hβ$_2$m components, had well-dispersed peaks to indicate a correctly folded protein, whilst the three-dimensional trHNCA, trHNCACB, and ^{1}H-^{1}H NOESY spectra had a signal to noise ratio that had sufficient resolution for the assignment process.

3.3.2.1 Backbone resonance assignment for H-2D^b α-chain

The fingerprint TROSY spectra collected using ^{2}H^{13}C^{15}N-H-2D^b α-chain and ^{2}H^{15}N-hβ$_2$m contained significantly fewer than the expected 261 signals (there are 15 Pro) after discounting the signals arising from hβ$_2$m, suggesting that a number of residues are affected by exchange broadening. Although not all cross peaks could be unambiguously assigned, 224 residues were assigned (Figure 3.4, chemical shifts listed in Appendix C), some of which gave rise to multiple resonances (since this relates to the presence of internal dynamics, discussion of this is found in Section 5.3). Several cross peaks on the fingerprint TROSY spectrum had correlated peaks in the trHNCA, trHNCACB, and NOESY spectra but could not be unambiguously assigned to the sequence with sufficient reliability. Missing and unassigned residues are largely located in the PBD of H-2D^b α-chain (Figure 3.5A) although some loop regions of the α$_3$ domain are absent. Within the PBD (Figure 3.5B) the unassigned residues are not evenly spread but are also not clustered in any particular region, occurring in several loop regions as well as on both α-helices and the platform itself. The α$_2$ helix, particularly the α$_{2\text{-}2}$ helix around the kink induced by Gly162, is more affected than the α$_1$ helix. On the β-strands of the platform, two notable sections that could not be assigned are Met5-Phe8 and Gln96-Asp102, which sit above the D-E loop of hβ$_2$m. The linker region between the PBD and the α$_3$ domain is commonly considered to consist of residues 176 to 180, and of these Ala177, Thr178, and Leu179 are unassigned.

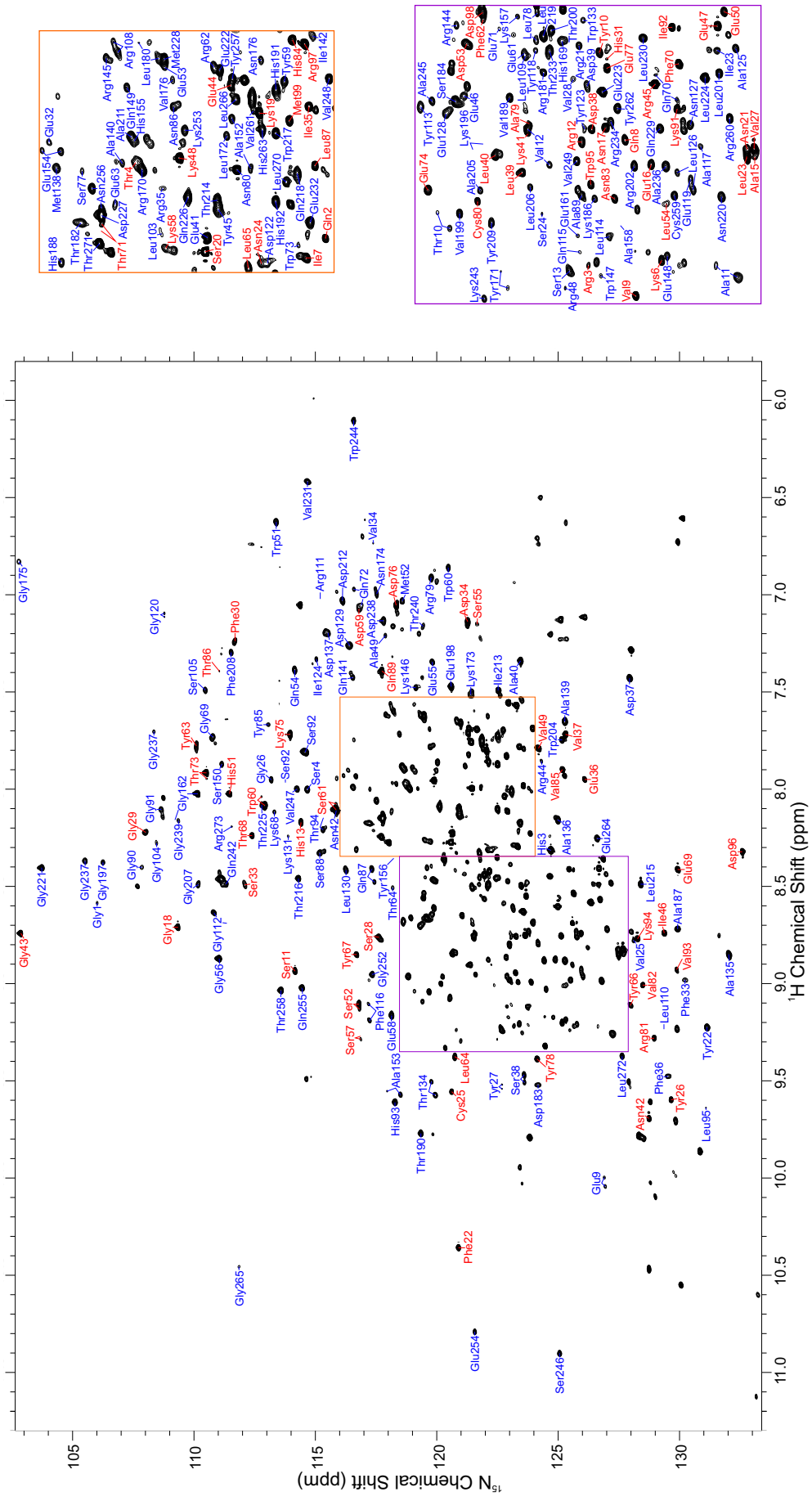


Figure 3.4 – Backbone resonance assignment for the H-2D^{b} α-chain of H-2D^{b}/FL.
Neither NH_2 groups (top right) nor indole NH groups (bottom left) are assigned on this TROSY spectrum. For clarity, previously assigned β_2m peaks (red) are labelled in addition to those originating from H-2D^{b} (blue).

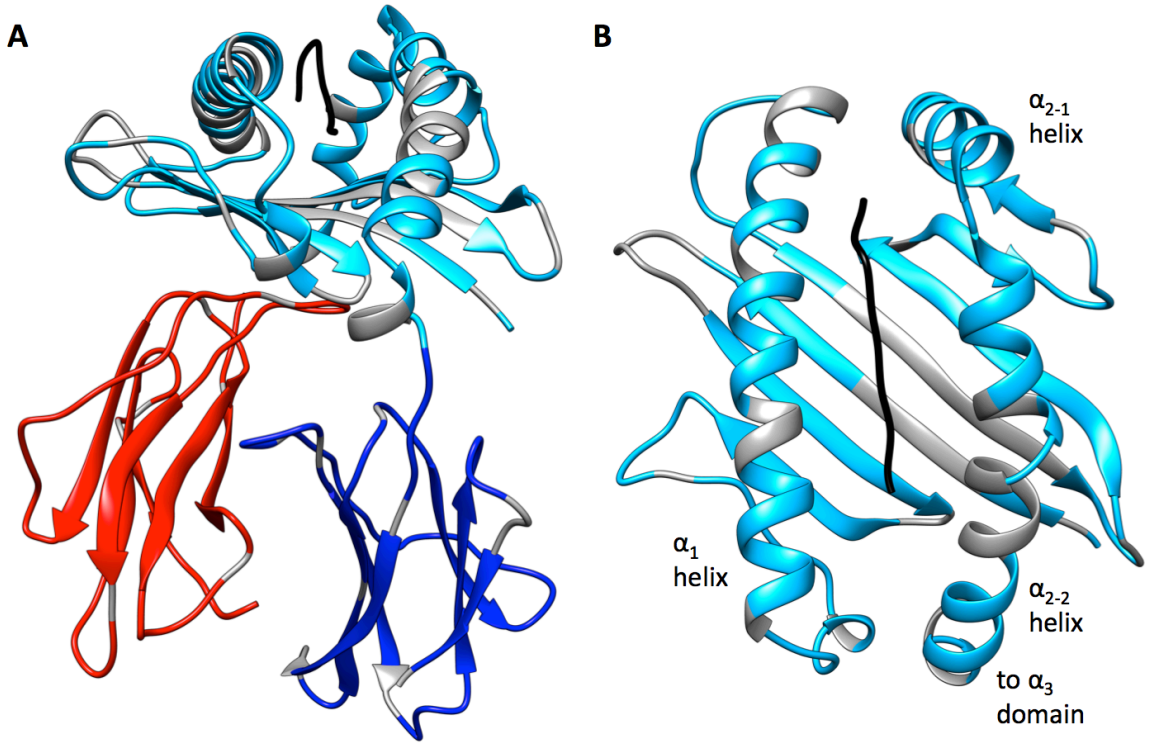


Figure 3.5 – Unassigned residues of H-2D^b/FL.
(A) Side view of H-2D^b/FL (PDB: 1CE6[98]). Unassigned residues on hβ_2m (red), H-2D^b α-chain PBD (light blue), and H-2D^b α-chain α_3 domain (dark blue) are coloured grey. Peptide (unlabelled and therefore unassigned) is shown in black. **(B)** The PBD of H-2D^b α-chain (PDB: 1CE6[98]).

Due to the long magnetisation transfer pathway, the poor signal to noise in the three-dimensional HNCACB spectrum (Figure 3.6) led to some ambiguity regarding which was the stronger of a peak pair within a strip and therefore the sequential assignment was aided by a HNCA spectrum was also collected, which had significantly better signal to noise ratio and therefore significantly less ambiguity (Figure 3.6). Arranging both spectra into strip plots, each containing two overlaid Cα peaks in both spectra and two Cβ peaks in the trHNCACB spectrum, permits the sequential linking of strips by stepping from strong peaks (residue i Cα and Cβ) to weaker peaks (residue i-1 Cα and Cβ).

Although the dispersion of Cα chemical shifts is relatively small and can lead to doubt over the exact residue type being observed, Cβ chemical shifts have a greater dispersion and therefore are more distinctive. For example, only Gly has a characteristic Cα chemical shift whilst Ala, Ser, and Thr have distinctive Cβ shifts and Gly does not have a Cβ at all. In addition, there are two further 'bands' of Cβ shifts; Arg, Gln, Glu, His, Lys, Met, Pro, Trp, and Val occur between 26 and 36 ppm, with Asn, Asp, Ile, Leu, Phe, and Tyr between 35 and 45 ppm. Cys is a special case since the

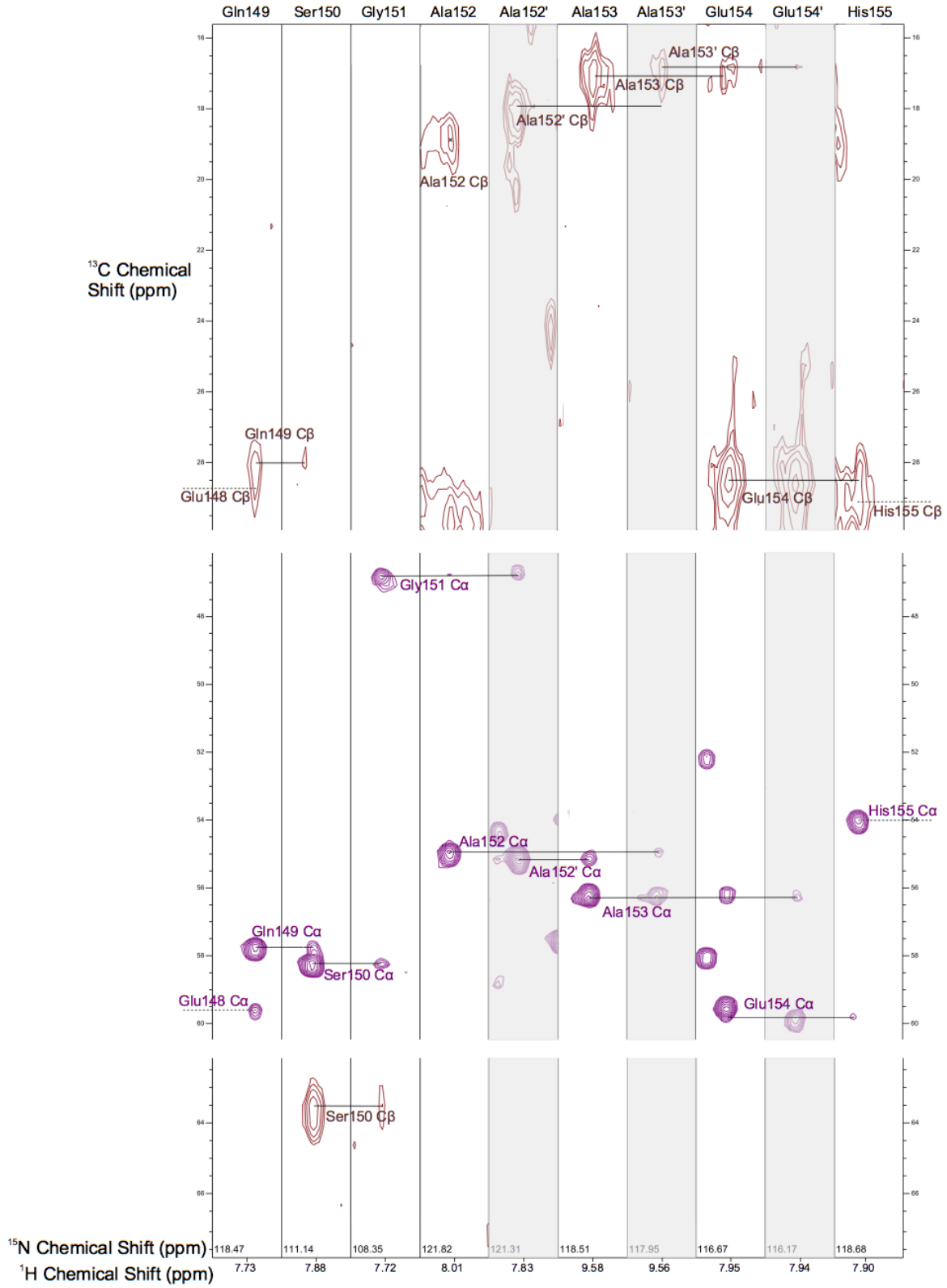


Figure 3.6 – Resonance assignment of Cα and Cβ nuclei for the H-2D^b α-chain component of H-2D^b/FL.
Representative trHNCACB strips (positive Cα peaks green, negative Cβ peaks brown) for sequential residues Gln149 through His155, overlaid with HNCA (purple). Unbroken arrows indicate peak linkages; two overlaid linkages are indicated by a thicker arrow. The second peak of a doublet is labelled residue i' and shown in a shaded strip.

chemical shift of its Cβ depends on the oxidation state of the side chain; reduced Cys Cβ has a shift of around 28 ppm whilst oxidised Cys (present in disulphide bonds) has a shift of around 41 ppm. Only the oxidised state of Cys has been observed, further suggesting that MHC class I is correctly folded.

The scarcity of HN-HN cross-peaks in the ^{1}H-^{1}H ^{15}N-edited NOESY spectrum (Figure 3.7) meant that it could not be used to sequentially link residues independently of other data. Instead, it was used to confirm links observed in the HNCA and HNCACB spectra and acted as additional proof of putative sequential assignments.

The amino acid selective unlabelling approach was used both to identify NH-^{15}N resonances of Arg, Lys, and Tyr residues on the TROSY and to confirm residues thought to be Arg, Lys, or Tyr through the sequential linking of HNCA, HNCACB, and NOESY strips (Figure 3.8). Due to shorter acquisition times, these TROSY spectra were not as well resolved as the high resolution TROSY spectrum used as a reference for assignment. Nevertheless, these spectra were extremely useful during the

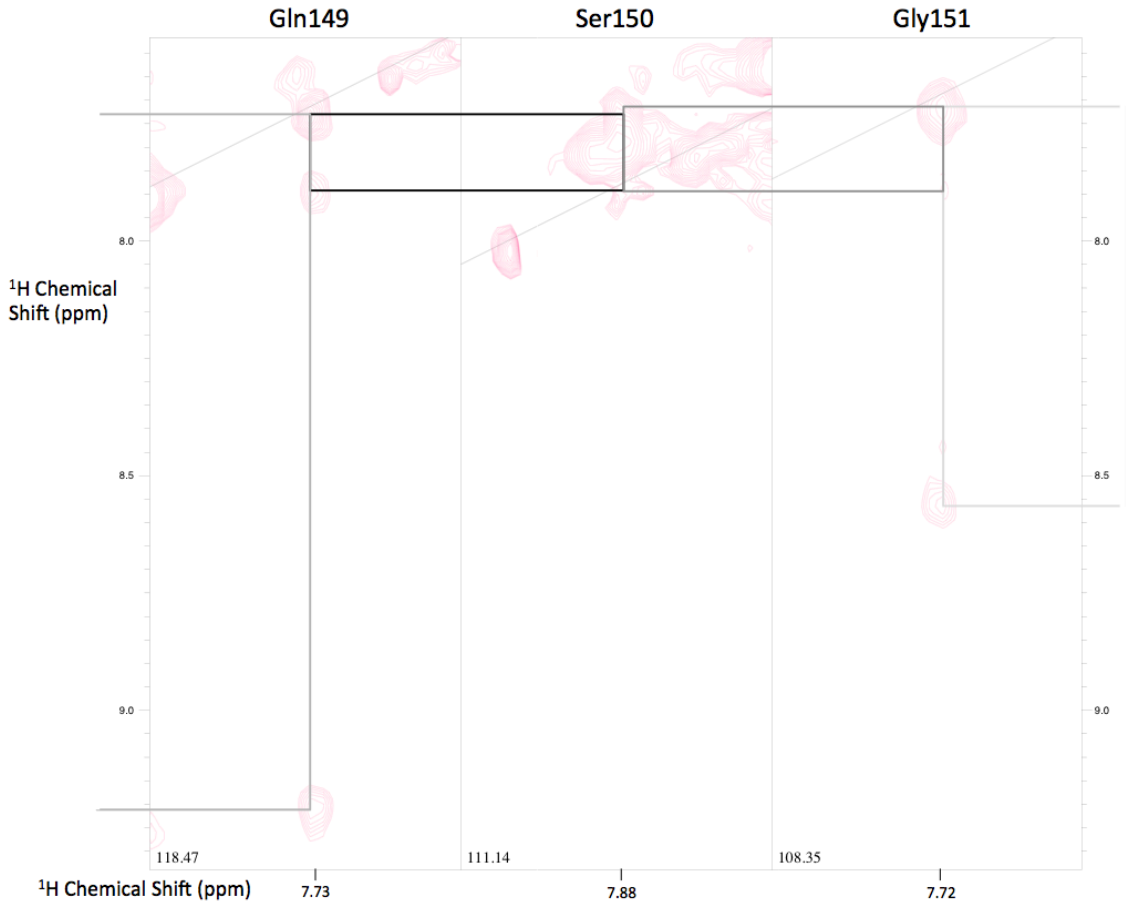


Figure 3.7 – Sequential peaks linkages for the H-2D^{b} α-chain component of H-2D^{b}/FL using a ^{1}H-^{1}H ^{15}N-edited NOESY spectrum.
Representative H-H ^{15}N-edited NOESY strips for sequential residues Gln149 through Gly151. In each strip the ^{15}N chemical shift (ppm) is indicated in the lower left corner. Unbroken lines indicate the linkage of cross-peaks.

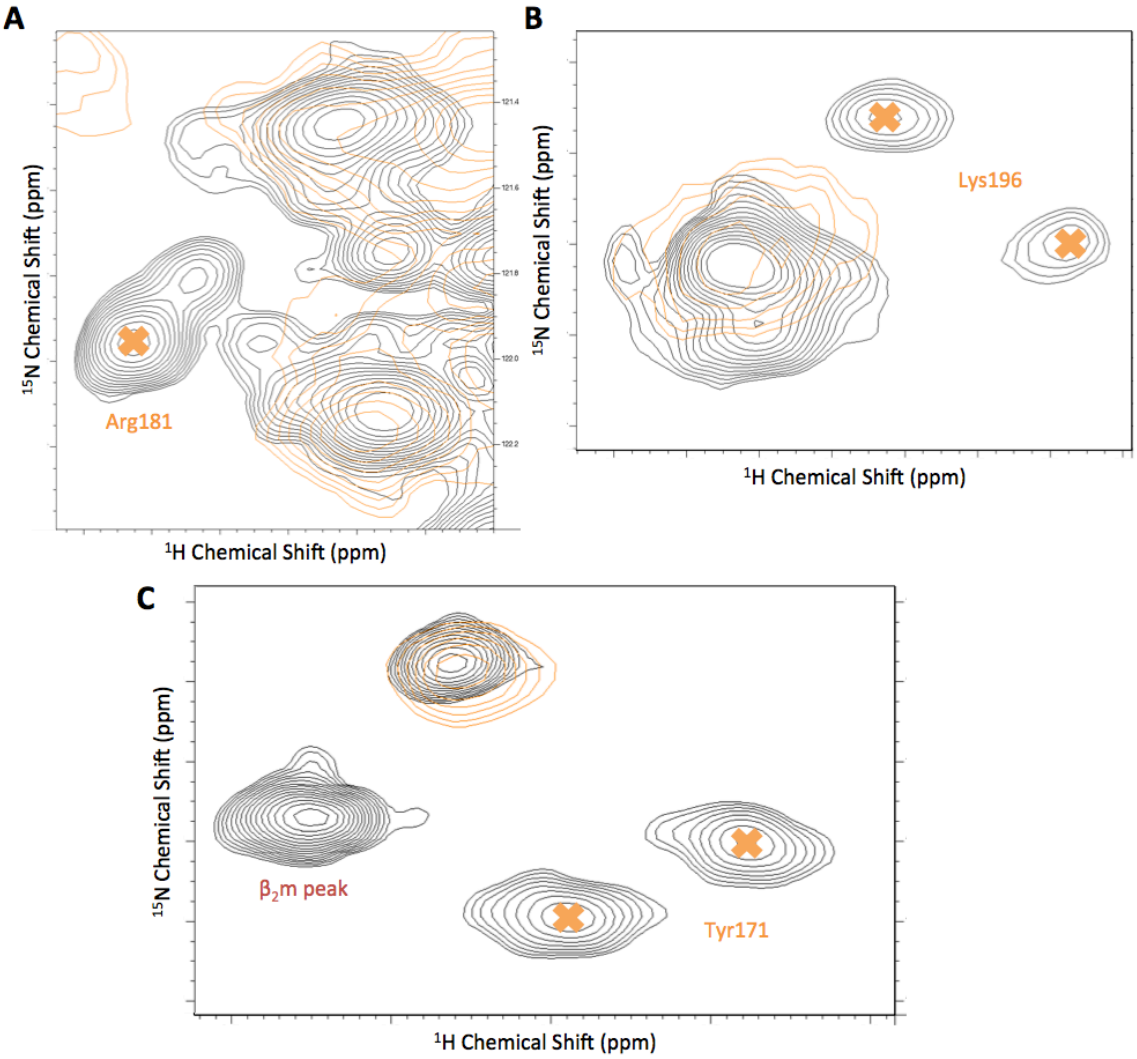


Figure 3.8 – Identification of Arg, Lys, and Tyr residues of the H-2Dᵇ α-chain component of H-2Dᵇ/FL on TROSY spectra using amino acid selective unlabelling.
(A) Amino acid selective labelling identifies peaks originating from Arg residues. Arg residue peaks are absent on a TROSY (orange) compared to the complete TROSY (black), for example Arg181. **(B)** Amino acid selective labelling identifies peaks originating from Lys residues; Lys196 is a doubled peak. **(C)** Amino acid selective labelling identifies peaks originating from Tyr residues; Tyr171 is a doubled peak. The third peak appearing only in the uniformly-labelled TROSY originates from hβ_2m and can be identified when the trHNCO plane is overlaid.

assignment process when a set of linked strips was placed in the protein sequence; for example the Cα and Cβ chemical shifts of Arg are very similar to those of His and Trp, whilst Lys shifts are similar to Met, and Tyr to Phe. The ability to make this distinction between residues enabled the confident assignment of sections of protein that otherwise would have been difficult.

3.3.2.2 Prediction of secondary structure

Experimental chemical shifts were compared to those expected in a random coil structure for the H-2D^b α-chain component of H-2D^b/FL using Chemical Shift Index 2.0 software (Figure 3.9A). In the PBD there are two main areas of β-strands both followed by regions of α-helix, suggesting that the correct secondary structure consisting of a β-sheet with two α-helices is present, whilst in the α_3 domain there are six regions of β-strand with a gap for the D strand (as seen in the X-ray crystallography structure; thought to be due to the presence of internal dynamics

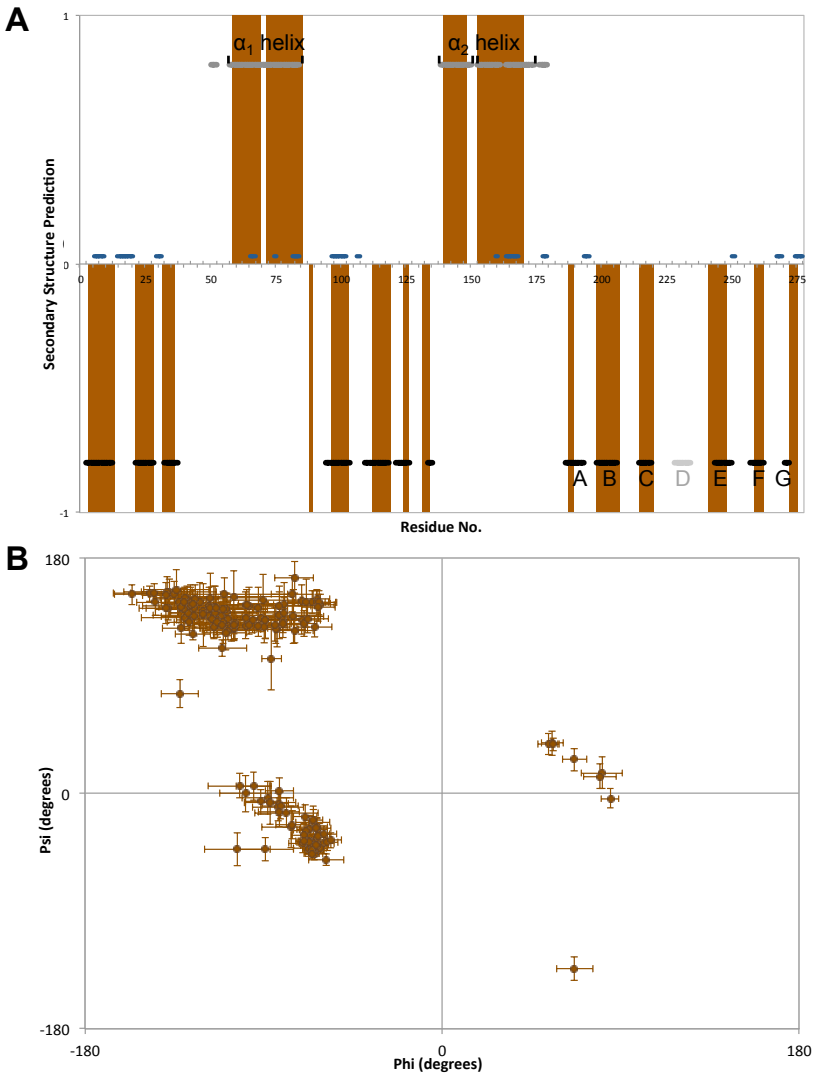


Figure 3.9 – Prediction of secondary structure for the H-2D^b α-chain component of H-2D^b/FL.
(A) Assigned Cα, CO, N, and HN chemical shift values are compared to random coil values using CSI 2.0; a value of +1 predicts α-helix and -1 predicts β-strand, whilst a value of 0 indicates no secondary structure. Chemical shifts are corrected for the presence of deuterium[194]. Crystal structure α-helices are indicated by grey lines labelled with helix number and β-strands are indicated by solid black lines labelled with strand number, with the α_3 non-canonical D strand in light grey. Residues lacking any chemical shift data are indicated by dark blue line. **(B)** Prediction of phi and psi angles from chemical shift data using TALOS-N; indicated errors are standard deviation.

disrupting the β-sheet hydrogen bonding network). Similar to the observations for the β_2m prediction, there is some ambiguity between coil and β-strand regions due to the similarity in chemical shift and composition. The level of ambiguity between α-helix and coil regions is much lower[192], so the absence of the expected helical regions preceding the α_1-helix and proceeding the α_2-helix may be significant and indicate that these regions are in coil conformation for H-2D^b/FL in solution.

The backbone torsion angles predicted by the TALOS-N software show the majority of residues for which data are available lie in the expected regions of the Ramachandran plot (Figure 3.9B). Residues that are categorised as α-helix or β-strand in the Chemical Shift Index prediction are present in the α-helical or β-strand sections of the Ramachandran plot. As with hβ_2m, the outlying residues are present in the loop regions between the secondary structure elements of the initial prediction, either due this lack of regular secondary structure or due to structural plasticity.

3.4 Comparison with existing structures

When $h\beta_2m$ is bound by any MHC class I α-chain, there are several changes that are expected to occur to its secondary structure due to the proximity of the α-chain and accompanying spatial restrictions for side chains. These are mostly based around the D-strand bulge region (see Figure 1.4 for changes in the secondary structure when bound by the H-$2D^b$ α-chain), and should be observed via changes in chemical shift from the isolated form (as assigned by Dr Taylor) and the H-$2D^b$/FL-bound form as assigned in this thesis.

A second comparison can be made for the entire MHC class I complex; H-$2D^b$/FL has a X-ray crystallograpic structure solved at 2.9 Å[98]. This enables comparison of a number of features, including the regions of secondary structure and backbone torsion angles.

H-$2D^b$/FL has also been characterised by Dr Alistair Bailey (University of Southampton) using molecular dynamics (MD) simulations. Briefly, these simulations employ the Amber99SB-ILDN force field and TIP3P water model within the GROMACS molecular dynamics package[164]. The crystal structure is used to provide the starting atoms and coordinates for MHC class I, with the protein placed into a simulated box of water with periodic boundary conditions and heated to standard temperature and pressure. Following an equilibrium process the simulation output of energies, velocities, and positions of each atom as a function of time are collected for further analysis[164]. The analysis presented here derives from a simulation extending over a total timescale of 0.5 μs and hence overlaps with the timescales observable via NMR spectroscopy.

3.4.1 Materials and methods

To compare the HN ^{15}N chemical shifts of isolated and H-$2D^b$/FL-bound $h\beta_2m$, the HSQC spectrum of the former was overlaid with the trHNCO plane of the latter. The Cα shifts of $h\beta_2m$ were also compared in both states.

There are three items that were compared between the crystal structure and NMR experimental data for H-$2D^b$/FL; the location of secondary structure, the chemical shifts, and the backbone torsion angles. Chemical shifts were predicted from the crystal structure using SHIFTX2[198] and following comparison of data mismatched residues were categorised as those beyond three times SHIFTX2 RMS of 0.4412ppm

(applying the 68-95-99.7 statistical rule). Phi and psi angles were inferred from the crystal structure using UCSF Chimera[187].

For the comparisons between the MD simulations and the NMR experimental data, the chemical shifts were calculated from the co-ordinates taken from a 560 ns simulation of H-2D^b/FL using SHIFTX2, resulting in chemical shift distributions[164]. The Cα shift distributions were analysed with regard to both NMR and X-ray crystallography-derived data.

3.4.2 Results

3.4.2.1 Comparison with isolated hβ$_2$m

The assignment of isolated β$_2$m is incomplete, with twelve residues unassigned on the ^{1}H-^{15}N fingerprint spectrum (Figure 3.10A). The majority of these residues belong to the loop region between strands D and E of the immunoglobulin fold and are likely absent due to exchange broadening, which suggests that this region of the protein is affected by internal dynamics, weak protein-protein interactions, or by solvent exchange[135]. When hβ$_2$m as a component of H-2D^b/FL was assigned only three residues were missing (Figure 3.10B). The unassigned residues in hβ$_2$m bound to the H-2D^b α-chain of H-2D^b/FL are Ile1, Phe56, and Ser88; the excellent signal to noise

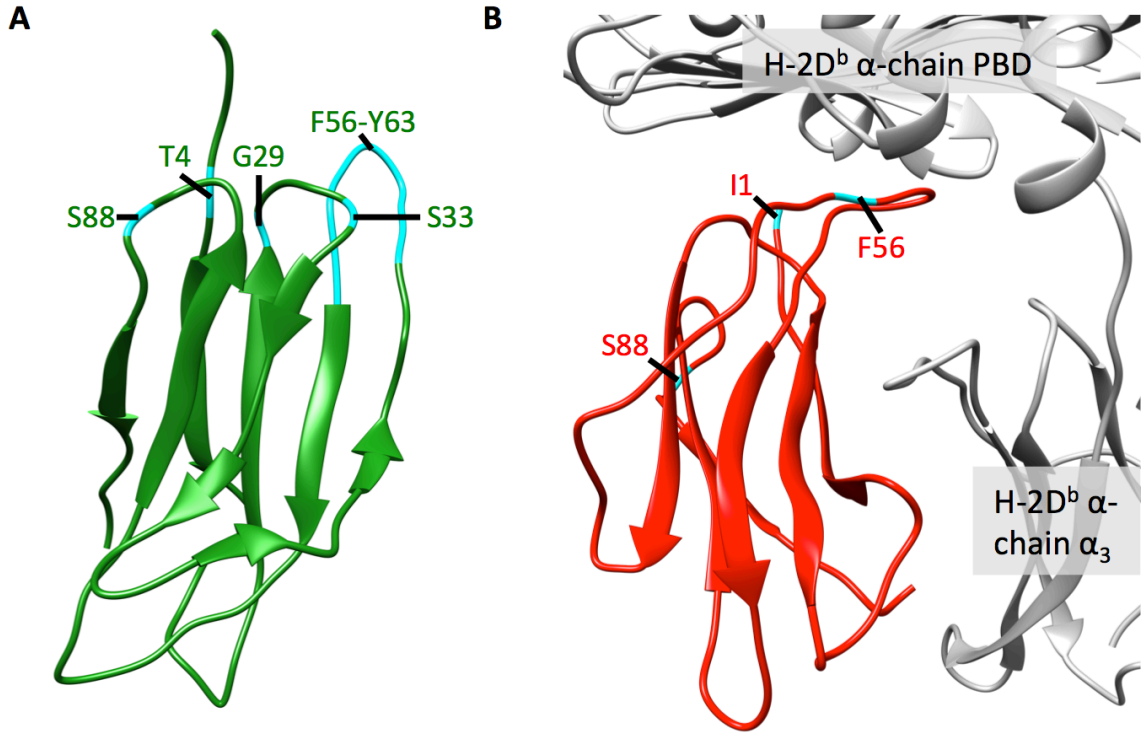


Figure 3.10 – Unassigned residues on isolated and H-2D^b/FL-bound hβ$_2$m.
(A) Unassigned residues on isolated hβ$_2$m. Residues represented on the solution-state NMR structure (PDB: 1JNJ[134]). **(B)** Unassigned residues on the hβ$_2$m component of H-2D^b/FL. Residues represented on the X-ray crystallography structure (PDB: 1CE6[98]).

ratio and lack of unassigned peaks on the ^{1}H-^{15}N plane makes it likely that these resonances are affected by exchange broadening. Ile1 is assigned in isolated β$_2$m but could not be located in the bound form. Ser88 could not be assigned in either isolated or bound hβ$_2$m, belonging to the loop between the F and G strands of the immunoglobulin fold. Phe56 is part of the D-E loop that was unassigned in isolated hβ$_2$m, but the remainder of this loop (residues 57-63) has been assigned for H-2D^b α-chain-bound hβ$_2$m. This argues that the D-E loop is partially stabilised by α-chain

binding. Indeed, Trp60, Phe62, and Tyr63 are all known closely interact with the α-chain of MHC class I via their side chains (Table 1.1).

When the assigned ^{1}H-^{15}N spectra of isolated and H-2D^{b} α-chain-bound β_2m are compared, there are several peaks that have significantly altered chemical shifts that have changed more than 1 ppm in the ^{15}N dimension and/or more than 0.15 ppm in the ^{1}H dimension (Figure 3.11A). For the majority of these residues, the change may be attributed to the presence of the H-2D^{b} α-chain (Figure 3.11B). Both Gln8 and Ser11 make contact with the α_3 domain of α-chain (Table 1.1); likely to also affect the local chemical environment of Val9. Although Asn24 does not make direct contact with α-chain, the side chain does point toward the D-E loop of the α_3 immunoglobulin fold. Val27 is on a strand of hβ_2m facing towards the α_3 domain and is in close proximity to Val37, suggesting a route of propagation of the observed chemical shift perturbation. His31, part of the B-C loop, sits beneath the PBD with its side chain pointing upward toward the PBG. Val49, Glu50, Leu54, and Ser55 are all part of the D strand and hence not in proximity to any part of the α-chain; it is conceivable that the apparent stabilisation of the D-E loop by the α-chain has a secondary impact on the D strand, perhaps explaining how the β bulge region could be affected. Similarly, Trp95, Arg97, Asp98, and Met99 may be affected by the C-terminus of β_2m facing toward the α_3 domain, possibly influencing the conformational sampling or average conformation of this part of β_2m.

3.4.2.2 Comparison with crystal structure

Using the chemical shift index predictions of secondary structure (hβ_2m Figure 3.3A, H-2D^{b} α-chain Figure 3.9A) the regions of α-helix and β-sheet are in broad agreement with those found in the crystal structure (PDB ID: 1CE6[98]). In the PBD of H-2D^{b} α-chain there are two regions of β-strands each followed by α-helix, whilst both the α_3 and hβ_2m contain β-strands in a pattern that would match the seven-stranded β-sheet immunoglobulin fold observed in the crystal structure.

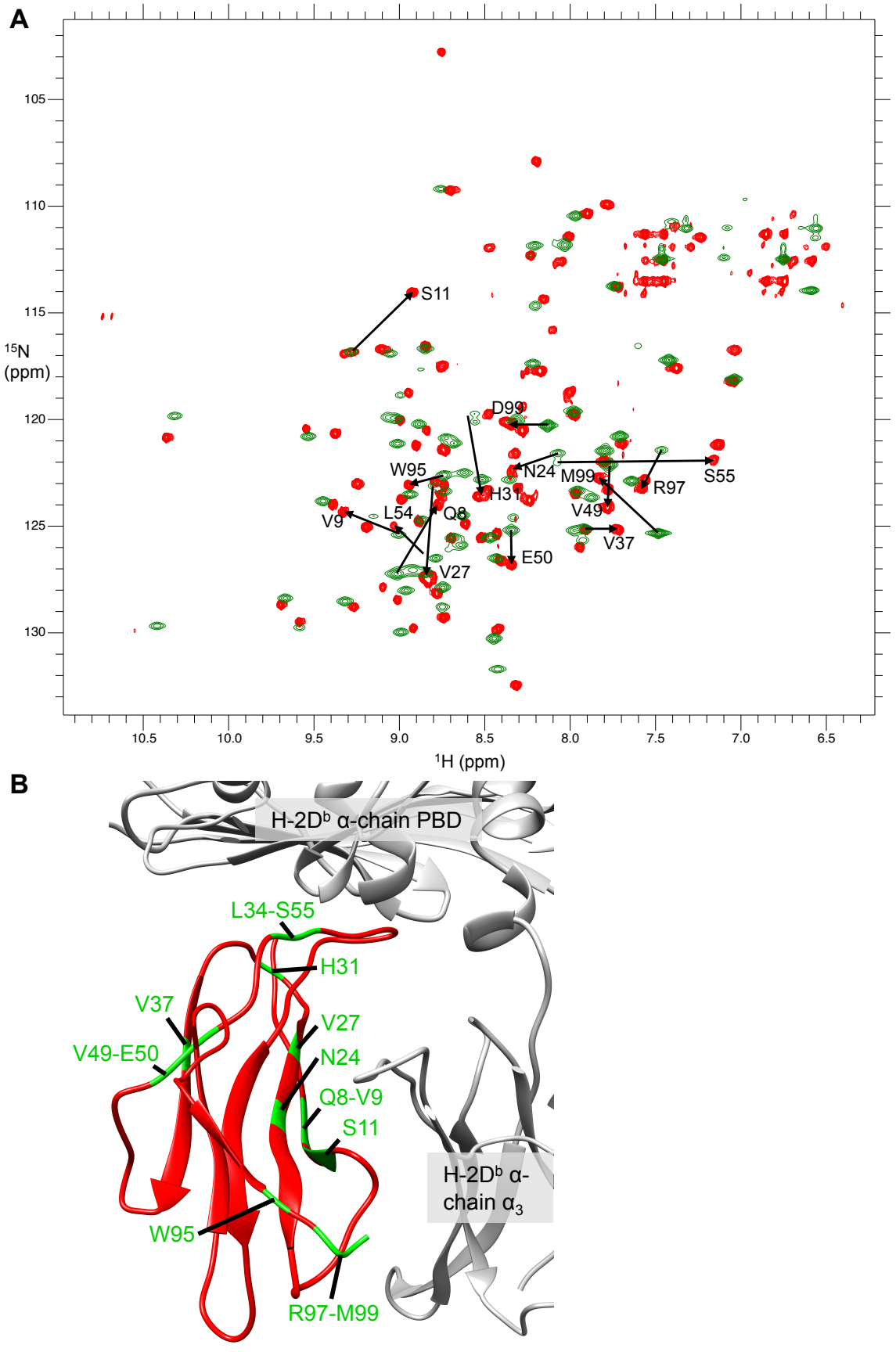


Figure 3.11 – Altered chemical shifts for H-2D^b/FL-bound hβ$_2$m in comparison to isolated hβ$_2$m.
(A) ^{1}H-^{15}N spectra for isolated (HSQC, green) and H-2D^b/FL-bound hβ$_2$m (trHNCO plane, red). Arrows indicate difference between the isolated and bound states. **(B)** Location of altered residues on H-2D^b/FL-bound hβ$_2$m, represented on the X-ray crystallography structure (PDB: 1CE6[98]).

For a direct comparison between the crystal structure and that observed in solution via NMR spectroscopy, chemical shifts were predicted from the former. Of the chemical shifts obtained from both NMR and SHIFTX2, the Cα chemical shift dataset was selected for comparison because it is particularly sensitive to changes in backbone conformation (Figure 3.12A). This reveals some variation in chemical shift value, mostly within the predicted uncertainty for data obtained using SHIFTX2. Residues for which the difference between NMR and crystallography chemical shift falls outside this uncertainty range are mostly located at the termini of secondary structure elements. However, 79% of outlying residues are in close proximity (within 6 Å) of aromatic residues and therefore tertiary effects may be influencing the chemical shift. The outliers not affected by aromatics are mostly located in loop regions or at breaks within the α helices. Interestingly, the $\alpha_{2\text{-}2}$ helix of the H-2D^b α-chain contains a large number of residues with a significant difference in chemical shift, suggesting that the crystal structure might not accurately reflect the arrangement of this helix in the solution state. The TPN binding region in the α_3 domain contains a number of residues with significant differences in chemical shift, and Leu54 of the β_2m D strand β bulge region is another outlier.

Another direct comparison that can be made between the crystal structure and the NMR experimental data is between the backbone torsion angles of phi and psi. Although both angles can be obtained from the crystal structure, for the sake of brevity the analysis here is restricted to the phi angle (Figure 3.12B). Similar to the chemical shifts, the majority of residues are in good agreement but there are outlying residues. These outliers are mostly located in loop regions, particularly between β-strands in the α_3 domain; Gly207 and Phe 208 in the B-C loop, Asn220-Glu222 and Asp228 in the C-D loop, and Gly239-Thr240 in the D-E loop. Interestingly hβ_2m His31 is an outlier; movement of this side chain is implicated in the formation of the β-bulge within the D strand, indicating that this region may retain the ability to form the alternative bulged conformation generally observed only in isolated β_2m. It is worth noting that the majority (74%) of outlying residues from the chemical shift comparison have differences in local structure as shown by this phi angle comparison, and all the outlying residues identified through the phi comparison are near (in primary sequence) to residues with differences in chemical shift comparison.

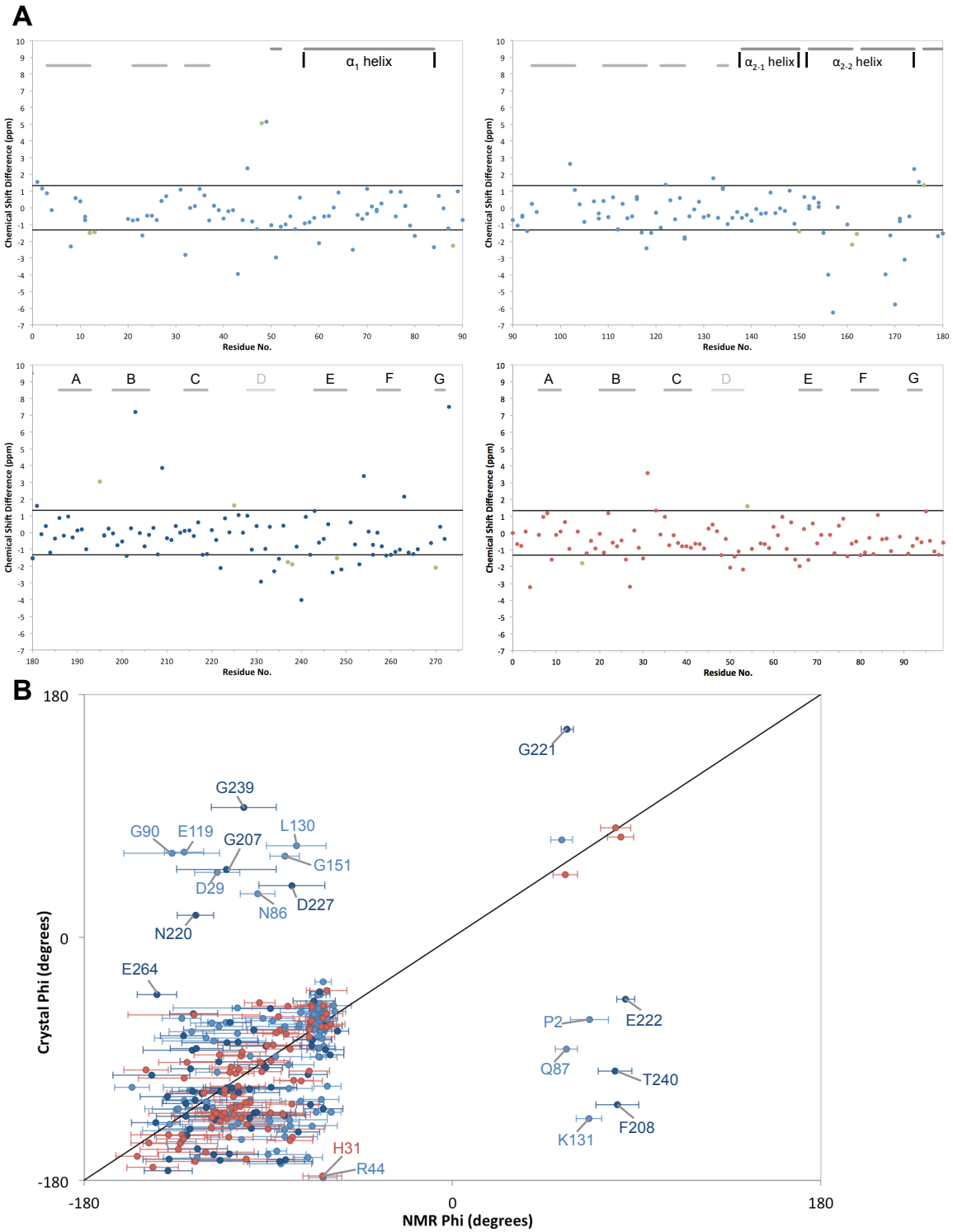


Figure 3.12 – Comparison between NMR and X-ray crystallography data.
(A) Comparison of chemical shift data. Assigned Cα chemical shifts are compared to those predicted from the crystal structure of H-2D^b/FL (PDB: 1CE6[98]) (H-2D^b α-chain α_1 domain top left, α_2 domain top right, α_3 domain bottom left, and hβ_2m bottom right). SHIFTX2 error (±1.32ppm) is indicated as black lines. Outlying residues not affected by nearby aromatics shown in green. Crystal structure α-helices are indicated by a dark grey line labelled with helix number. β-strands are indicated by light grey lines labelled with strand number, with the non-canonical D strands of both H-2D^b α-chain α_3 domain and hβ_2m in a lighter grey. **(B)** Comparison of backbone conformation. The backbone torsion phi angle is used as representative of the backbone conformation for H-2D^b/FL. Data shown is for H-2D^b α-chain PBD (light blue) and α_3 domain (dark blue), along with hβ_2m (red) versus the crystal structure (PDB: 1CE6[98]). Outlying residues are labelled.

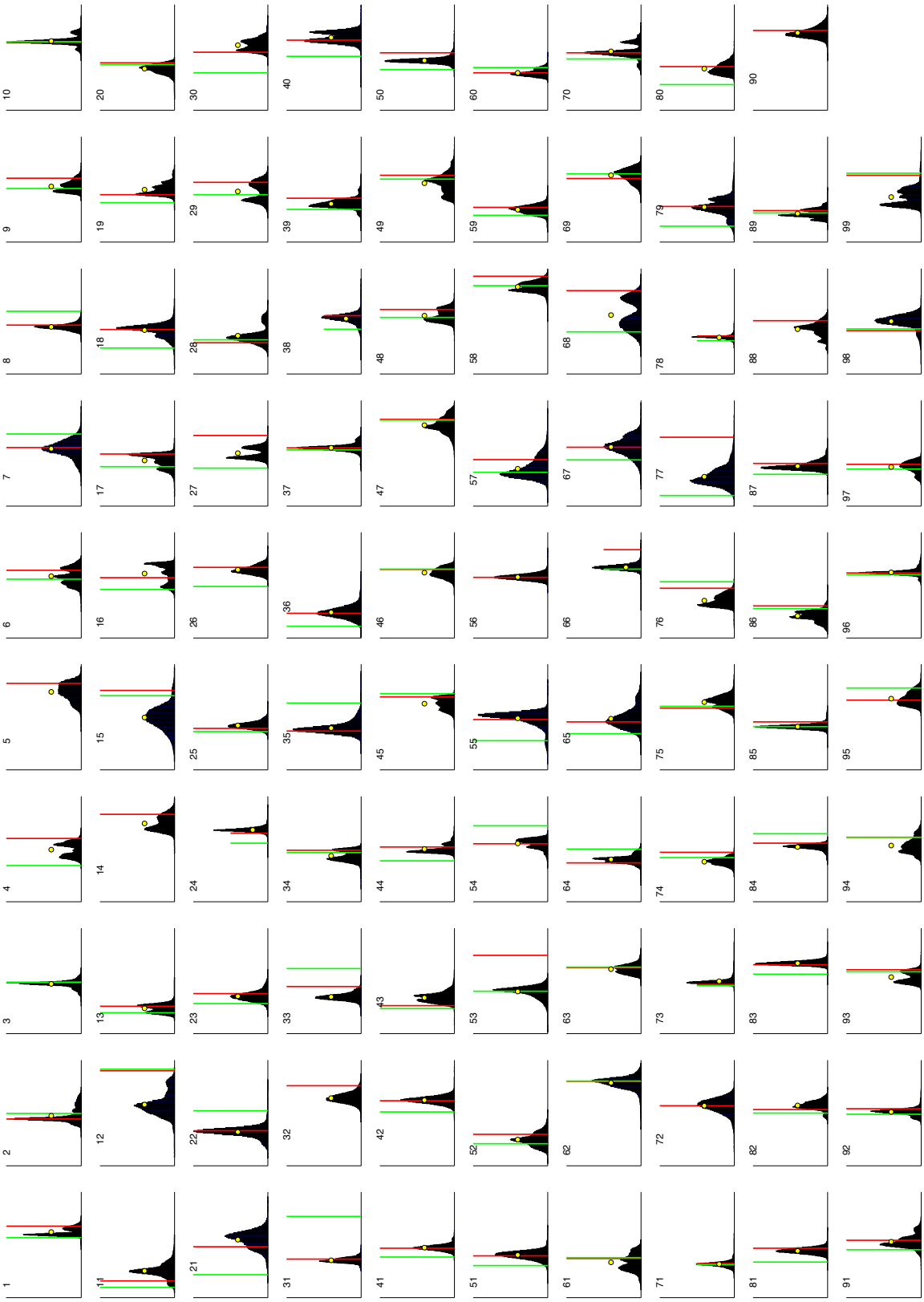
3.4.2.3 Comparison with molecular dynamics simulation

The MD simulation uses the coordinates of the crystal structure as a starting point, and the secondary structure is maintained throughout the simulation with minor changes at the termini of elements. Thus all three structural studies (NMR spectroscopy, X-ray crystallography, and MD) agree on the secondary structure of H-2D^b/FL.

Chemical shifts were predicted for each snapshot within the 560 ns MD simulation of H-2D^b/FL to produce a histogram of Cα chemical shifts for each residue (Figure 3.13). This analysis revealed that the majority of residues experienced a range of chemical shifts, which is indicative of dynamics throughout the protein backbone and side chains. Several residues in both the H-2D^b α-chain and hβ_2m components of H-2D^b/FL experience at least two major populations of chemical shift. For example, Thr68 of β_2m appears to exist between two populations with an equal amount of time spent in each one whilst the MD suggests that Asn176 of H-2D^b α-chain is present in three unequal populations such that the time spent in each one is not evenly distributed. It should be noted that this analysis does not inform us of the rate of exchange between states; there may have only been one slow conversion between the two states at the halfway point of the simulation, or there may have been multiple conversions. In general, residues with a large range of chemical shifts, and hence possible multiple conformational states (the movement of a side chain containing a indole ring will affect the surrounding chemical shifts), are located in regions of H-2D^b/FL that could be predicted to be more dynamic. For example, many are located in loop regions, with the non-canonical D-strand of H-2D^b α-chain showing greater variation than the remainder of the β-sheet residues and the residues around the PBD-α_3 domain linker also showing motion. Both TPN binding regions (residues 128-136 and 222-229) of H-2D^b α-chain are observed to have multiple conformations, suggesting that these regions are capable of motion and adding weight to the notion that TPN binding may be influenced by the dynamics of MHC class I. Also of note is that the α_1 helix (residues 138-150) experiences fewer conformational states that the α_2 helix (152-174), which may mean that the latter is capable of greater motion than the former.

There are multiple populations observed in the MD trajectory in many of the regions of the H-2D^b α-chain where backbone resonances could not be assigned. A good example of this is residues 96-101, where there are several residues with equally weighted populations; these multiple populations would decrease the intensity of the peaks produced on the TROSY spectrum and the peaks might also suffer from line

A



B

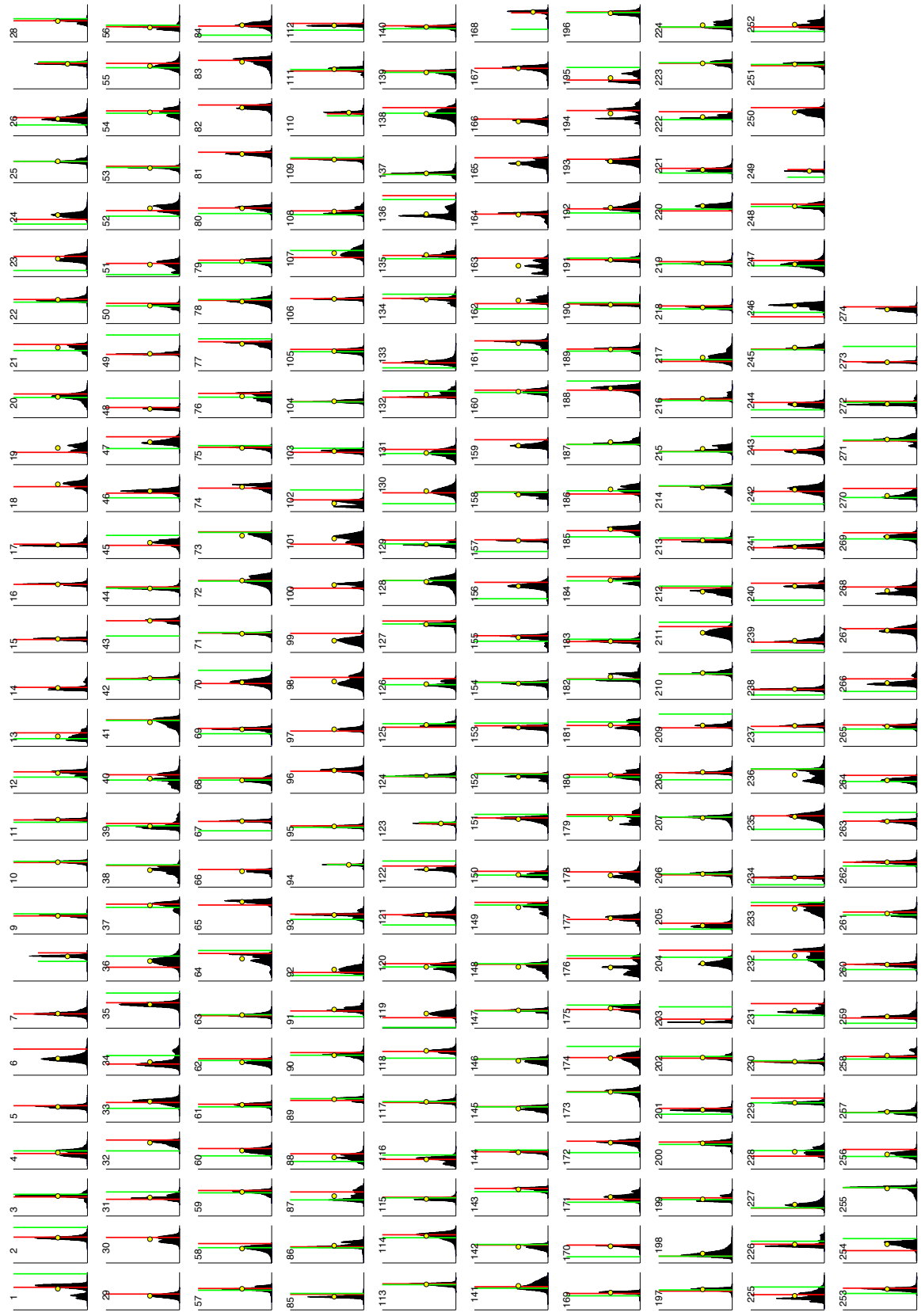


Figure 3.13 – Chemical shift predictions from MD simulation of H-2D^b/FL.
Cα chemical shifts for each residue from a 560 ns MD simulation are shown as a histogram in black (MD population against chemical shift) for **(A)** the hβ$_2$m component of H-2D^b/FL and **(B)** the H-2D^b α-chain. The average MD value is shown as a yellow circle; variation in MD prediction means that the MD-average is not necessarily representative. The chemical shift predicted from the X-ray structure as per Section 3.4.2.2 is shown as a red line, with the experimental NMR chemical shift as a green line.

broadening due to chemical exchange. However, the MD simulation for unassigned residues in other regions (e.g. residues 14-19) suggests that there is not significant mobility and the failure to assign these residues may be a result of spectral crowding. The mismatch between MD and NMR data suggests that the NMR experiments that have been carried out for the purpose of backbone resonance assignment have been unable to observe the full range of dynamic states that occur in H-2D^b/FL when in solution; further investigation is required.

In general, Figure 3.13 shows that the MD simulation has greater agreement with the NMR experimental chemical shifts than with those predicted from the crystal structure of H-2D^b/FL; the overlap between NMR and MD is greater than that between crystal structure and MD. Where there are multiple chemical shift populations observed in the MD, only one of these populations matches the NMR-observed or crystal-observed one. For example, the NMR chemical shift matches one MD-population of β_2m Thr68 with the other matched by the crystal structure, whilst both the crystal and NMR chemical shifts match the same one of the three H-2D^b α-chain Asn176 MD populations. Previous experimentation with smaller (and easier to analyse) proteins has also found that MD simulations produce data closer to NMR chemical shifts than to the crystal structure, with the note that poor MD prediction of chemical shift values may identify spurious conformations arising from insufficient sampling such that a necessary 'correction' from the crystal input structure weights the population towards an inaccurate conformer[199]. Current computational resources mean that MD simulations beyond the microsecond timescale presented here are not yet generally available. However, continuous improvements of the force fields and computing power suggest that MD will become an increasingly useful tool for the interpretation of NMR and structural data. It is encouraging that the available simulations yield chemical shift information that on average agrees with NMR and X-ray crystallography sources. However, the approximation made in the derivation of chemical shifts from structures, as well as the approximate nature of the MD force fields and the limited timescale for simulations, requires a judicious use of this information.

Molecular dynamics data can be used to further investigate the differences identified through the NMR and crystallographic comparison of the previous section. There are a number of possible explanations for a difference in NMR and crystallographic chemical shift values, the most significant of which is structure. For the majority of outlier residues (Figure 3.12A) it is not possible to rule out the effect of nearby

aromatic rings, although many of these residues also have differences in backbone conformation as predicted by phi angle comparison. In addition half of these residues are also identified through the MD data as having a range of observed chemical shifts, such that dynamics may also be a contributing factor to the observed difference. For the subset of residues not affected by local aromatic rings for which there is a clear difference in NMR and crystal structure chemical shift value this difference is likely due to dynamic averaging of the chemical shift due to fluctuations in the structure; the MD data shows that 75% of these identified outliers possess a distribution of Cα shifts that is broader than average. The simulation appears to capture much of this dynamic averaging, as evidenced by the average MD chemical shift improving on the crystallographic data for 81% of the outlier residues originally identified. For the remaining 19% of residues, it is possible that the structural dynamics fall outside the MD timescale such that the effect on chemical shift is not wholly captured. Although it is important to consider that chemical shift calculation is based on semi-empirical approximations that have inherent errors, the use of chemical shift data to examine protein structure in this way is valid because overall the crystallographic and MD chemical shifts are in close agreement with the NMR data, thus suggesting there is no systematic error in calculation. For the remaining residues where there is unexplained disagreement between the three datasets, this is likely due to the crystal structure not being an accurate representation of all local aspects of H-2D^{b}/FL structure in solution, and in terms of MD data this may have an effect on the simulation trajectory.

3.5 Summary

This chapter has described the successful backbone resonance assignment for H-2Db/FL, which employed ^{2}H and ^{15}N isotopic labelling with differential ^{13}C labelling. The hβ_2m component was 91% assigned using a trHNCO plane in place of the more usual HSQC or TROSY fingerprint spectrum, along with three dimensional trHNCO and trHNCA experiments. The H-2D^{b} α-chain has been 81% assigned on a TROSY spectrum using trHNCA, trHNCACB, and ^{1}H-^{1}H ^{15}N-edited NOESY experiments, with additional TROSY spectra collected for H-2D^{b}/FL with amino acid selective unlabelling. Unassigned residues are mostly located in the PBD, in loop regions, with the $\alpha_{2\text{-}2}$ helix affected more than other helical regions. An area of the PBD platform above the D-E loop of hβ_2m is also affected; this may be due to the presence of ring currents induced by the side chain of β_2m Trp60.

Comparing data collected for isolated hβ_2m to this assignment suggests that some of the internal dynamics of isolated hβ_2m are reduced upon binding to H-2D^{b} α-chain, and the observed changes in chemical shifts are largely due to the proximity of the α-chain.

When data from the assignment of H-2D^{b}/FL is compared to the crystal structure and a MD simulation there is broad agreement between datasets; secondary structure elements are in the same locations and a comparison between backbone torsion angles predicted from the NMR and crystal data reveals differences only in loop regions. From the MD data it is clear that there are residues within H-2D^{b}/FL that experience a number of conformations, which are not consistently captured by either the NMR or crystallography techniques. These are located in loops as well as the PBD-α_3 domain linker and the TPN-binding regions. The MD suggests that the majority of unassigned residues have multiple conformers and this potential exchange broadening may well explain the inability to complete the resonance assignment – this requires further investigation, initially via analysis of the linewidths from the TROSY spectra.

Chapter 4

4 Solution-state domain arrangement of H-2D^b/FL

NMR spectroscopy can be used to obtain information on molecular orientation via the collection of residual dipolar coupling (RDC) data, which involves restricting the tumbling of a molecule. These data are commonly used to supplement structure calculations or to observe the relative orientations of different regions of a molecule and hence follow structural changes due to protein folding or ligand binding. In the case of MHC class I, it may be possible to use RDC's to gain insight into the relative orientations of the four constituent domains.

This chapter describes the collection of residual dipolar coupling data and the characterisation of domain orientations within H-2D^b/FL.

4.1 Introduction

Standard NMR experiments measure isotropic data for which there is an even distribution of orientation, whilst RDC experiments utilise weakly orientated liquids in which molecules have a slight preference for a particular orientation. When a molecule experiences orientational sampling, strong first-order interactions including dipolar coupling are no longer averaged to zero – giving rise to *residual* dipolar couplings. However, it is important that the alignment is only partial; for high-resolution NMR, spectra must remain relatively simple without the complexity induced by the presence of large dipolar couplings.

RDC data is often analysed in tandem with other NMR measurements, and where it is analysed in the absence of other quantitative indicators the focus is frequently on domain orientations, most notably applied for the 81.4 kDa enzyme malate synthase G[200]. However, it is possible to calculate a high-resolution structure solely from RDCs provided that an approximate structural model is available and that sufficient data is collected[178], as demonstrated by studies on ubiquitin[201] and cytochrome c[202].

4.1.1 Residual dipolar couplings

Structure calculations using NMR spectroscopic data typically rely on proton-proton distance restraints obtained from NOE data. However, due to spin diffusion it is not always possible or practical to record these for large proteins[200] such as the 45 kDa MHC class I. An alternative is to employ the RDC, a parameter that is not affected by

spin diffusion and that reports a time-averaged orientation for an internuclear vector. Instead of identifying a unique orientation, each individual RDC measurement represents a continuum of potential orientations[203] such that additional measurements must be made in order to identify the actual orientation. Common consensus is that five measurements are required, whether the same dipolar coupling measured in five different orienting media or five different couplings, although due to experimental uncertainties even then a single vector orientation may not be able to uniquely satisfy all the recorded couplings[204,205].

An additional consideration is how many RDC measurements are required to accurately calculate a structure because a limited set of RDCs can be satisfied to within experimental error by an incorrect structure; a dipolar coupling does not uniquely describe a vector orientation and hence the fact that a calculated structure satisfies a limited set of RDCs is not proof of its accuracy (in the case of structure refinement using only RDC data, deliberately incorrect assignment of RDC measurements can nevertheless result in an apparently valid structure)[178]. Experimentation has shown that the number of RDC measurements must be larger than the number of torsional degrees of freedom (over 2N couplings for a N-residue protein) in order to obtain a reliable structure[178]. In addition to structure quality factors, it is common practice to cross-validate any structure by withholding 10% of the data for later comparison[206].

The most significant structural applications of RDC data are to probe domain-domain orientations, assessing motion, or changes upon ligand binding. Domain level structure examination removes the requirement for five measurements and >2N couplings; as few as five couplings per domain are needed for an accurate assessment[178], although alignment tensors are more accurate where more than ten couplings are available[207].

4.1.2 Orienting media

The collection of RDCs requires orienting media to induce partial alignment. RDC data has not previously been collected for any MHC class I allele and it is therefore important to consider the choice of media carefully; RDC experiments require isotopically labelled protein that is costly and time-consuming to produce, and some types of media can interact with some proteins in an unhelpful manner.

The first type of orienting media is a polyacrylamide gel; the gel is typically between 3 and 10% acrylamide and the gel pH can match the protein sample. Gel preparation is

similar to that for SDS-PAGE, but polymerisation occurs within narrow diameter tubing. Once the gel is set, it is extracted, cut to size, and dried, before protein solution is used to rehydrate the gel in a NMR tube. The diameter of the gel is wider than the NMR tube to induce compression; the pores are not isotropic and therefore protein tumbling is sterically restricted[208].

Lipid bicelles also induce a steric alignment, usually employing $C_{12}E_5$ PEG with variable pH to match the protein buffer. Additionally, CTAB (cetyltrimethyl ammonium bromide) or SOS (sodium octyl sulphate) can be added for positively or negatively charged proteins respectively. 16% PEG is mixed with hexanol before protein is added such that the final PEG concentration is 4.2%[209].

The use of Pf1 phage orienting media is significantly more expensive, but sample preparation is relatively simple, with 10-20 mg/ml Pf1 phage required and a pH range of 6-8, although NaCl is often required as a stabiliser in the buffer solution. Bacteriophages have a net negative surface charge (positively charged proteins may interact too strongly) and hence the principle basis for alignment is electrostatic[210].

Both polyacrylamide gels and lipid bicelles are used as initial media because they have a high success rate, with Pf1 phage or others (Otting media, paramagnetic tags) only used if problems are encountered or to obtain additional datasets. The degree of alignment is an important consideration; too large an alignment results in complex spectra that are difficult to interpret and too small an alignment produces data that may not be statistically significant.

4.2 Materials and methods

H-2D^b/FL was expressed in *E. coli* with a labelling scheme of dually labelled ($^2H^{15}N$) hβ_2m and triple labelled ($^2H^{13}C^{15}N$) H-2D^b α-chain, and purified as described in Chapter 2 to produce a 600 μM sample for collection of data in a 5% polyacrylamide gel (see Appendix B for full recipe). A second RDC dataset was collected using an isotropic sample of 480 μM H-2D^b/FL with the labelling scheme of $^2H^{15}N$ for both the hβ_2m and H-2D^b α-chain components, whilst anisotropic data collection employed a Pf1 phage alignment medium (ASLA Biotech) such that the final sample conditions were 330 μM H-2D^b/FL, 15 mg/ml Pf1, and 200 mM KCl.

Experiments were performed on a Varian Inova 800 MHz spectrometer equipped with a 5 mm triple resonance (HCN) PFG cryogenic probe, at the MRC Biomedical NMR Centre, National Institute for Medical Research, London. $^1J_{NH}$ splittings were determined from the 600 μM H-2D^b/FL sample using a J-modulation experiment[211], and anisotropic datasets were obtained using an ARTSY experiment[212]. Briefly, the J-modulation experiment uses a range of time delays to modulate the signal where couplings are active. The coupling constants can then be extracted from a curve fitted to the peak intensities for these various time delays. The ARTSY experiment provides both reference and attenuated anisotropic data, the ratio between which can be combined with the coupling constant already obtained to calculate the RDC value.

4.2.1 Measurement of $^1J_{NH}$

A J-modulation experiment to measure $^1J_{NH}$ was selected due to the significant improvement in accuracy when compared to the more traditional method of measuring doublet splittings on a HN spectral plane[211]. The production of a doublet due to differential relaxation during evolution of ^{15}N transverse magnetisation is prevented because the pulse sequence is designed such that the magnetisation is equally split between the $|a\rangle$ and $|b\rangle$ states (angular momentum of + or – ½ along the z-axis), therefore eliminating the cross-correlation between the ^{15}N-^1H dipolar and ^{15}N chemical shift anisotropy relaxation mechanisms. For more detail on the pulse sequence see Tjandra, Grzesiek, and Bax[211].

A series of twelve $^1J_{NH}$-modulated spectra were recorded using dephasing delays of 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24 milliseconds. During analysis in CCPN it became apparent that the 16 ms delay produced data of an intensity so close to zero it could not be distinguished from noise, and this spectrum was excluded. Data was

exported to Excel for manipulation. XMGrace and MATLAB were used to plot peak intensity against delay time and fit the following curve;

$$f(x) = a\left(-b + cos\left(\pi \cdot 2 \cdot {}^{1}J_{NH} \cdot x\right)\right) exp\left(-2(x/T_2)\right)$$

Equation 4.1

where x is the dephasing delay, a is the initial intensity of the peak as assessed in CCPN Analysis, b accounts for the imperfection of π pulses during the experiment such that it varies for each residue ($0<b<0.05$), and ${}^{1}J_{NH}$ is approximately −92Hz

4.2.2 Measurement of anisotropic attenuation

The study of a large protein such as MHC class I by NMR can be complicated by broadened linewidths, such that traditional HSQC-based methods cannot produce spectra with sufficient resolution to distinguish neighbouring peaks. The ARTSY method is TROSY-based and therefore retains a high degree of resolution. For details of the pulse sequence see Fitzkee and Bax[212].

The ARTSY experiment is carried out in two steps. A reference spectrum is collected such that $^{1}J_{NH}$ dephasing is active during half of the time during which $^{1}H^{N}$ chemical shift anisotropy/^{15}N-^{1}H dipolar coupling relaxation interference is active (the interval T of approximately 10.7 ms). A second, attenuated spectrum is collected such that $^{1}J_{NH}$ dephasing is active during the entire T interval. The intensity ratio between the two spectra can be obtained by analysis in CCPN Analysis.

4.2.3 Measurement of RDC

The $^{1}J_{NH}$ and ARTSY data can be used to calculate the RDC value;

$${}^{1}D_{NH} = \frac{-1}{T} - {}^{1}J_{NH} + \left(\frac{2}{\pi T}\right) sin^{-1}\left(\frac{Q}{2}\right)$$

Equation 4.2

where T is 10.7 ms (variable dependent on precise experimental set-up) and Q is the ARTSY intensity ratio

The error in the RDC value can be assessed using error propagation from the ARTSY experiment. This simplifies to:

$$\sigma_{RDC} = \frac{2}{\pi T (I_R/N)} \sqrt{\frac{1+Q^2}{4-Q^2}}$$

Equation 4.3

where I_R is the peak intensity in the reference ARTSY spectrum, and N is the noise level of either of the ARTSY spectra (these should contain the same noise level since they are collected for identical lengths of time). The contribution to σ_{RDC} by the J-modulation experiment is insignificant in comparison to that arising from the ARTSY set of data, and has not been included.

4.2.4 Comparison with crystal structure

RDC values were fitted to the crystal structure (PDB 1CE6[98]) using Module 1.0[186] to define domain orientations. This software generates an alignment tensor that is characterised by axial (A^a) and rhombic components (A^r) to indicate the strength of the alignment, and by three rotation angles (Euler angles α, β, and γ) that give the orientation of the tensor with respect to the molecular frame provided by the existing structure. Molecules can be treated as one domain, or multiple domains can be manually defined.

Module 1.0 compares experimental RDC values with those back-calculated from the crystal structure. Residues that contribute greatly to internal dynamics appeared as outliers from this RDC correlation and were removed from the system. This removal was based on an error function, χ^2, which describes the agreement between experimental and calculated RDCs; this was minimized to find the tensor parameters that best reproduce the experimental results. Subsequently, five hundred Monte Carlo simulations were carried out to estimate uncertainties in the alignment tensor parameters. The agreement between the data and the structure was evaluated using a quality factor Q[213]:

$$Q = \sqrt{\sum (rdc^{measured} - rdc^{calculated})^2} \Big/ \sqrt{\sum (rdc^{measured})^2}$$

Equation 4.4

Independently from the Module1.0 analysis, experimental data was compared to the crystal structure using REFMAC5, which permits the combination of X-ray crystallographic data and NMR RDC or pseudocontact shift data during the structure refinement process. REFMAC5 is a new addition to the CCP4 suite of software that is widely used during the analysis of crystal data; the raw crystallographic reflection data is first refined to obtain reference R factor and R_{free} values before the NMR data is used to drive a further refinement. If the structures present in the crystal and solution states are in agreement there will be an improvement in these structure validation parameters as well as a Q factor of <0.2[213,214]. Following addition of NMR restraints, the relative weightings of the NMR and crystallography data are varied iteratively as described by Rinaldelli *et al*[214]. If the two structures do not agree, it will not be possible to improve the R factor and R_{free} values.

The 2.90 Å crystal structure that is used in the remainder of this thesis was published by Glithero *et al,* in 1999[98], at which point the PDB (Protein DataBank) portal did not

mandate deposition of the raw crystallographic data. It was not possible to obtain this data in a timely fashion. Therefore, an alternative pre-existing 3.92 Å dataset was used; crystal reflection data obtained by Dr Halina Mikolajek (University of Southampton) for H-2D^b/FL composed of H-2D^b α-chain refolded with hβ_2m and a modified FL peptide such that the penultimate and ultimate amino acids contained a carbonyl sulphur in place of the oxygen atom: FL-CSNH. The Bragg's diffraction data sets were collected at the i24 Microfocus MX beamline at the Diamond Light Source (Didcot, UK).

4.2.5 Structure refinement

RDC data was used to drive a structure refinement via the WeNMR web portal[215], using the crystal structure (PDB 1CE6[98]) as an initial set of atomic coordinates. The AMPS-NMR service utilises the Amber99SB force field with a 10 Å solvent box created using the TIP3P water model. The established protocol carries out an initial fitting of the RDC tensor to generate an alignment tensor, before an energy minimisation step. The molecular system is heated to 300 K and a 30 ps MD simulation performed incorporating the uploaded RDC restraint list. This is followed by cooling to 0 K and a final energy minimisation, and produces a structure output in PDB format.

4.3 Results

Analysis of the results from the RDC experiments occurred in several stages, the first of which was to obtain the 'raw' RDC values. These were then used to drive a structure refinement, and to separately delineate the domains of H-2D^b/FL and then define the interdomain orientations. The results were then compared to the existing crystal structure and to independent MD simulations.

4.3.1 Measurement of RDC

Peak intensities for each residue of both hβ_2m and the H-2D^b α-chain were exported from CCPN Analysis for each of the eleven J-modulated spectra; residues that produced peaks in close proximity or that overlaid on the NH spectral plane (for example Phe62 and Asp98 of β_2m; Figure 3.1) could not be picked. An example fit graph of the J-modulated data is shown in Figure 4.1, confirming that the 16 ms delay produces data extremely close to zero.

NMR data recorded on aligned samples revealed negligible chemical shift changes when compared with the isotropic data, indicating the effects of the gel matrix on the protein structure to be minimal. Although TROSY-based, both the resolution and noise level of the ARTSY spectra proved to be the limiting factor; a number of residues that could be selected in the J-modulated spectra could not be picked in the ARTSY. Despite this, for the polyacrylamide gel 212 RDC's were measured and for the phage media 219 RDC's were measured.

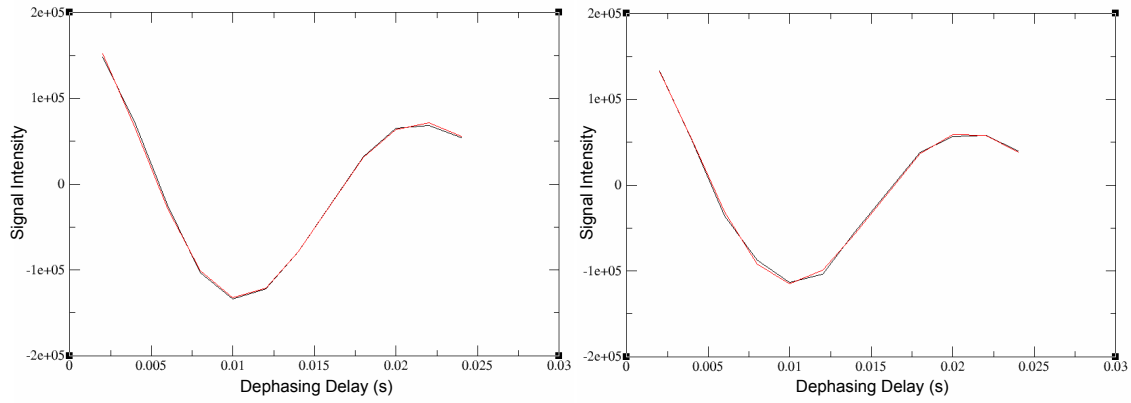


Figure 4.1 – Calculation of $^1J_{NH}$ using a J-modulation experiment.
Peak intensity on the NH plane plotted against J-modulated dephasing delay for Phe22 of MHC class I-bound hβ_2m (left) and Trp244 of H-2D^b α-chain (right). Graphs created using XMGrace. Equation of fitted curve as per Equation 4.1; f(x)=a0*(-a1+cos(π*2*a2*x))*exp(-2*x/a3).

4.3.2 Definition of domains I: polyacrylamide gel alignment media

The J-modulation and ARTSY experiments provided a dataset of 213 H-N RDC values spread relatively evenly within H-2D^b/FL. Using Module1.0 these data were used to model the solution state against the crystal structure (PDB 1CE6[98]). If these structures are in agreement there should be a good correlation between the experimental data and those back-calculated by Module1.0 from the crystal structure: when H-2D^b/FL was modelled as a single domain there was a poor correlation and an extremely high χ^2 error (Table 4.1). This high χ^2 error shows that H-2D^b/FL is in a different conformation in solution to that shown in the crystal structure. The many outliers on the correlation plot are scattered throughout the structure, suggesting that this difference is not simply due to movement of residues that are in loop regions but that H-2D^b/FL consists of domains that may have different orientations with respect to each other to what is observed in the crystal structure.

The next step is therefore to divide H-2D^b/FL into constituent domains; the most obvious first step is the separation of the hβ_2m component from the H-2D^b α-chain and thus create a two-module analysis. This results in a significantly improved

Table 4.1 – Definition of domains within H-2D^b/FL using RDC couplings collected in the polyacrylamide gel alignment media.

No. of Modules	Domain	α (°)	β (°)	γ (°)	A^a ($x10^{-4}$)	A^r ($x10^{-4}$)	$\chi2$ per coupling
1	–	-138	146	-92	6.6	3.1	364
2	α-chain	-114	121	-87	2.5	1.5	132
	hβ_2m	-141	141	-87	11.3	7.1	436
3	PBD	-16	85	5	-3.9	-2.1	105
	α_3	-139	138	-97	7.8	3.9	85
	hβ_2m	-141	141	-87	11.3	7.1	436
5	Platform	-128	138	-77	10.6	5.7	58
	α_1 helix	-105	31	71	-52.9	-34.8	25
	α_2 helix	-20	81	-46	19.8	4.5	27
	α_3	-139	138	-97	7.8	3.9	85
	hβ_2m	-141	141	-87	11.3	7.1	436
6	Platform	-128	138	-77	10.6	5.7	58
	α_1 helix	-105	31	71	-52.9	-34.8	25
	$\alpha_{2\text{-}1}$ helix	143	20	-153	-87.3	-48.6	3
	$\alpha_{2\text{-}2}$ helix	130	105	139	12.8	3.1	10
	α_3	-139	138	-97	7.8	3.9	85
	hβ_2m	-141	141	-87	11.3	7.1	436

correlation for both α-chain and β_2m. The large χ^2 error for hβ_2m arises because β_2m produces substantially stronger signals than H-2D^b α-chain residues such that the signal to noise ratio is considerably increased. The error in RDC measurement (Equation 4.3) for hβ_2m is therefore extremely low, resulting in a low tolerance of any deviation from a precise correlation and hence a high χ^2 value. Examining the correlation plot for the hβ_2m module, outlier residues are located mostly in loop regions or are suggested to have significant internal dynamics due to the presence of doubled peaks on the trHNCO plane; hβ_2m is therefore a domain of H-2D^b/FL but the data require further refinement prior to calculation of a final alignment tensor.

A poor correlation within the H-2D^b α-chain required further domains to be delineated; it was divided into the α_3 domain and the PBD (residues 1-180). This improved the correlation of the data significantly, and the Euler angles show that the α_3 and β_2m domains have similar orientations (Table 4.1). However, there were a number of outlying residues, all of which belonged to the α_1 and α_2 α-helices such that the χ^2 error could be further minimised by redefining the PBD as the α_1 helix (residues 57-84), the α_2 helix (residues 138-180), and the β-sheet platform (residues 1-56 and 85-137). This brought the platform into an approximate alignment with the α_3 and β_2m domains. The χ^2 error within the α_2 helix could be improved by defining the $\alpha_{2\text{-}1}$ (residues 138-151) and $\alpha_{2\text{-}2}$ (residues 152-180) helices as separate modules. H-2D^b/FL therefore can be said to consist of six domains; the H-2D^b α-chain platform, α_1 helix, $\alpha_{2\text{-}1}$ helix, $\alpha_{2\text{-}2}$ helix, and α_3 domain along with hβ_2m.

4.3.3 Definition of domains II: Pf1 phage alignment media

An identical domain definition was carried out using the 219 H-N RDC couplings obtained using the Pf1 phage alignment media; a single-domain model of H-2D^b/FL did not describe the data effectively and the protein was divided into six domains (Table 4.2). Again, the hβ_2m domain contained few significant outliers but had a high χ^2 due to intense peaks in the ARTSY spectrum. The Euler angles show that in this initial domain definition the α_3 and hβ_2m domains do not have similar orientations as observed previously. The β_2m and platform domains are comparable, suggesting that at this stage of the analysis there is disagreement between the two datasets regarding the α_3 domain. In both datasets it appears that the α_1 helix is not described by sufficient data to produce a reliable result.

Table 4.2 - Definition of domains within H-2D^b/FL using RDC couplings collected in the Pf1 phage alignment media.

No. of Modules	Domain	α (°)	β (°)	γ (°)	A^a (x10^{-4})	A^r (x10^{-4})	χ2 per coupling
1	–	-44	122	41	-7.5	-0.9	356
2	α-chain	-67	120	39	-7.8	-1.7	336
	hβ$_2$m	-12	123	45	-8.0	-1.0	354
3	PBD	-55	107	37	-7.1	-3.7	238
	α$_3$	75	49	-143	-8.0	-0.6	195
	hβ$_2$m	-12	123	45	-8.0	-1.0	354
5	Platform	-51	108	38	-7.6	-4.6	209
	α$_1$ helix	-42	110	-48	213.7	133.4	63
	α$_2$ helix	118	39	149	-7.6	-4.2	90
	α$_3$	75	49	-143	-8.0	-0.6	195
	hβ$_2$m	-12	123	45	-8.0	-1.0	354
6	Platform	-51	108	38	-7.6	-4.6	209
	α$_1$ helix	-42	110	-48	213.7	133.4	63
	α$_{2\text{-}1}$ helix	15	28	-20	-42.8	-25.8	13
	α$_{2\text{-}2}$ helix	-101	67	40	14.7	1.5	-[a]
	α$_3$	75	49	-143	-8.0	-0.6	195
	hβ$_2$m	-12	123	45	-8.0	-1.0	354

[a] Calculation of χ^2 not possible due to an insufficient number of couplings

4.3.4 Comparison with crystal structure I: Module1.0

Following domain definition, the χ^2 error was minimised via identification of residues with a high RDC error from spectral analysis and of outlying residues on the RDC correlation graphs. These residues were then investigated; the majority were found to be located in loop regions that could be expected to experience mobility in solution, and the couplings removed from the dataset if residues nearby in the primary structure also produced outlying results. In addition, RDC couplings were removed if they originated from residues for which two peaks were observed on a ^{1}H-^{15}N plane during backbone resonance assignment. Removal of these couplings did not have a significant effect on the alignment tensors, but produced a substantial improvement in χ^2 value. This refinement process resulted in 150 and 160 H-N RDC couplings for the polyacrylamide gel and Pf1 phage alignment media datasets respectively (Appendix D).

The refined alignment tensors are given in Table 4.3, with errors calculated via 500 Monte Carlo simulations. It is immediately clear that the α_1, $\alpha_{2\text{-}1}$, and $\alpha_{2\text{-}2}$ helical domains produce questionable data, particularly when the strength of domain alignments (the A^a and A^r values) are compared to those from the remaining domains. These helices are not reliable due to insufficient data: although these couplings have a good back-correlation (Figure 4.2), the inherent errors are high because the NH bond vectors are effectively co-linear such that not all orientations are being sampled and therefore an unrealistic tensor is obtained; Figure 4.3. These domains must therefore be disregarded. However, it is important to note that for both alignment media the Euler angles of the helical domains differ significantly from the platform and from each other, as demonstrated in Figure 4.2 in which the alignment tensors for each of the helical domains is fixed to that of the platform to illustrate the justification in defining them as separate domains. The crystal structure therefore appears to contain inaccurate relative orientations within the PBD in the solution state of H-2D^b/FL.

The platform, α_3, and hβ_2m domains are also require rotation by Module1.0 to produce RDC values in agreement with those back-calculated from the crystal structure (Figure 4.2). The α_3 and hβ_2m domain alignment tensors are in close agreement with one another, with the platform rotated by approximately 20° with respect to them (Table 4.3, Figure 4.3). This is observed for both alignment media, although for the Pf1 phage dataset the β_2m alignment tensor requires a 90° rotation to obtain directly comparable values (the A_{xx} and A_{yy} axis of the rotation are of very similar size and Module1.0 automatically denotes the shortest axis as A_{xx}); these datasets are independent and thus provide verification of the observations.

Table 4.3 – Alignment tensors of domains within H-2D^b/FL.
Molecular frame from PBD 1CE6[98].

	A^a (10^{-4})	A^r (10^{-4})	α (°)	β (°)	γ (°)
Polyacrylamide Gel Alignment Media					
Platform	10.5 ± 0.2	6.7 ± 0.2	-128.6 ± 0.9	129.9 ± 0.9	-73.4 ± 0.5
α_1 helix	125.9 ± 12.5	83.8 ± 8.2	-44.3 ± 1.2	57.6 ± 1.2	-1.8 ± 1.3
$\alpha_{2\text{-}1}$ helix	-76.6 ± 5.4	-42.8 ± 3.7	105.2 ± 12.5	10.0 ± 1.1	-108.8 ± 13.3
$\alpha_{2\text{-}2}$ helix	19.8 ± 2.5	3.2 ± 1.7	165.9 ± 10.7	110.8 ± 3.9	-99.8 ± 6.2
α_3	7.5 ± 0.2	4.7 ± 0.1	-148.4 ± 1.4	137.1 ± 0.6	-100.2 ± 0.9
β_2m	11.5 ± 0.1	7.3 ± 0.1	-141.6 ± 0.3	143.7 ± 0.2	-87.4 ± 0.3
Pf1 Phage Alignment Media					
Platform	-7.9 ± 0.1	-2.7 ± 0.1	-42.7 ± 0.7	107.9 ± 0.3	46.2 ± 0.3
α_1 helix	94.7 ± 11.9	54.7 ± 8.0	-41.0 ± 1.3	108.5 ± 1.6	-46.8 ± 1.6
$\alpha_{2\text{-}1}$ helix	-42.8 ± 2.3	-25.8 ± 1.5	14.75 ± 0.9	27.9 ± 0.5	-19.8 ± 0.8
$\alpha_{2\text{-}2}$ helix	14.7 ± 1.1	1.5 ± 0.8	-100.6 ± 8.9	67.2 ± 1.2	40.4 ± 5.0
α_3	-8.3 ± 0.1	-0.5 ± 0.1	-53.0 ± 3.0	134.0 ± 0.2	37.6 ± 0.2
β_2m	-8.7 ± 0.1	-1.2 ± 0.1	19.4 ± 1.0	125.2 ± 0.1	38.7 ± 0.2

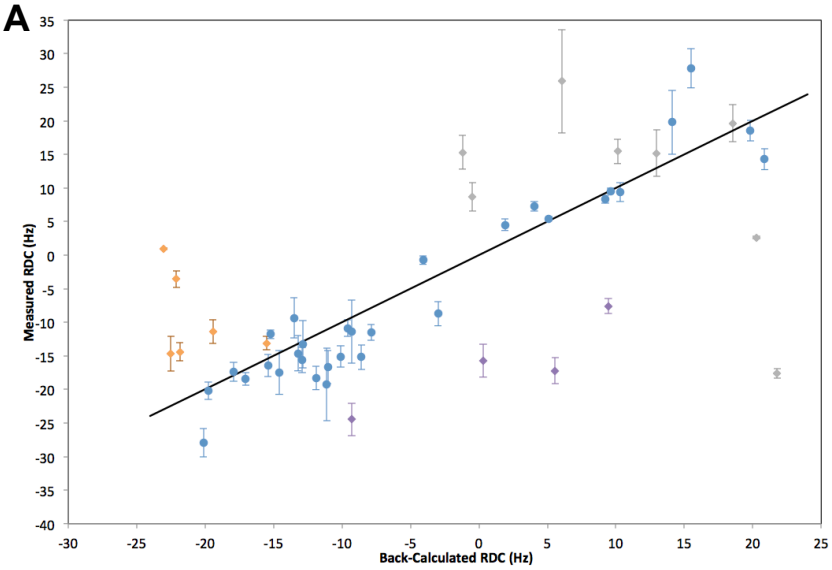
	No. of couplings	**Q[a]**	**χ^2**	**[χ^2][b]**
Polyacrylamide Gel Alignment Media				
Platform	31	0.30	215.3 ± 7.2	6.95
α_1 helix	6	0.37	50.0 ± 1.5	8.33
$\alpha_{2\text{-}1}$ helix	8	0.17	5.4 ± 2.6	0.68
$\alpha_{2\text{-}2}$ helix	4	-[c]	-[c]	-[c]
α_3	50	0.53	1360.6 ± 9.8	27.21
β_2m	51	0.38	3875.4 ± 9.5	75.99
Pf1 Phage Alignment Media				
Platform	33	0.55	5826.6 ± 7.5	176.6
α_1 helix	7	0.02	1.1 ± 1.9	0.2
$\alpha_{2\text{-}1}$ helix	9	0.24	120.7 ± 2.7	13.4
$\alpha_{2\text{-}2}$ helix	5	-[c]	-[c]	-[c]
α_3	50	0.28	4614.8 ± 8.9	92.3
β_2m	56	0.28	4578.0 ± 10.0	81.8

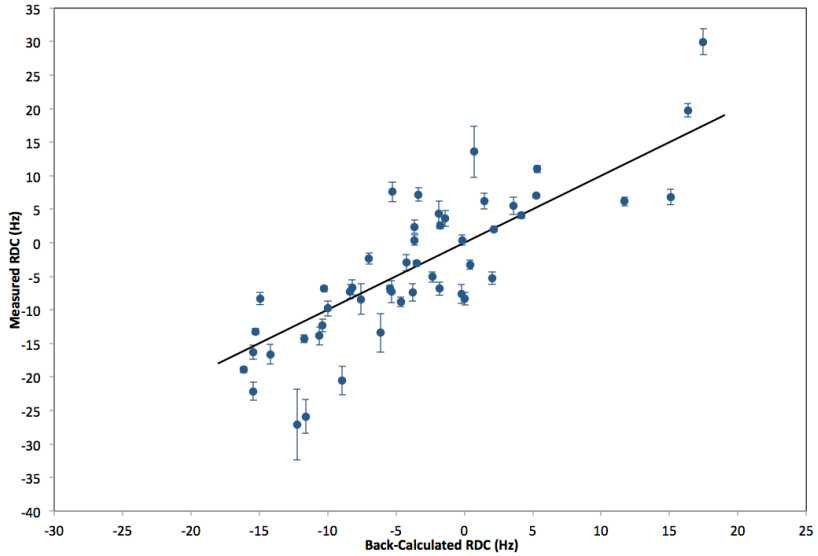
[a] The quality factor Q measures the agreement between the data and the crystal structure

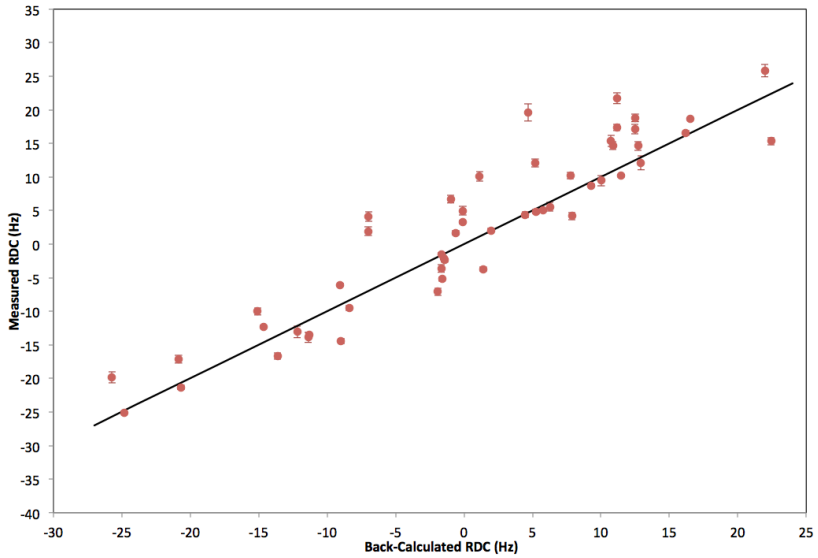
[b] [χ^2] values are the average for each named domain

[c] Calculation not possible due to an insufficient number of couplings

A


Measured RDC (Hz)
Back-Calculated RDC (Hz)


Measured RDC (Hz)
Back-Calculated RDC (Hz)


Measured RDC (Hz)
Back-Calculated RDC (Hz)

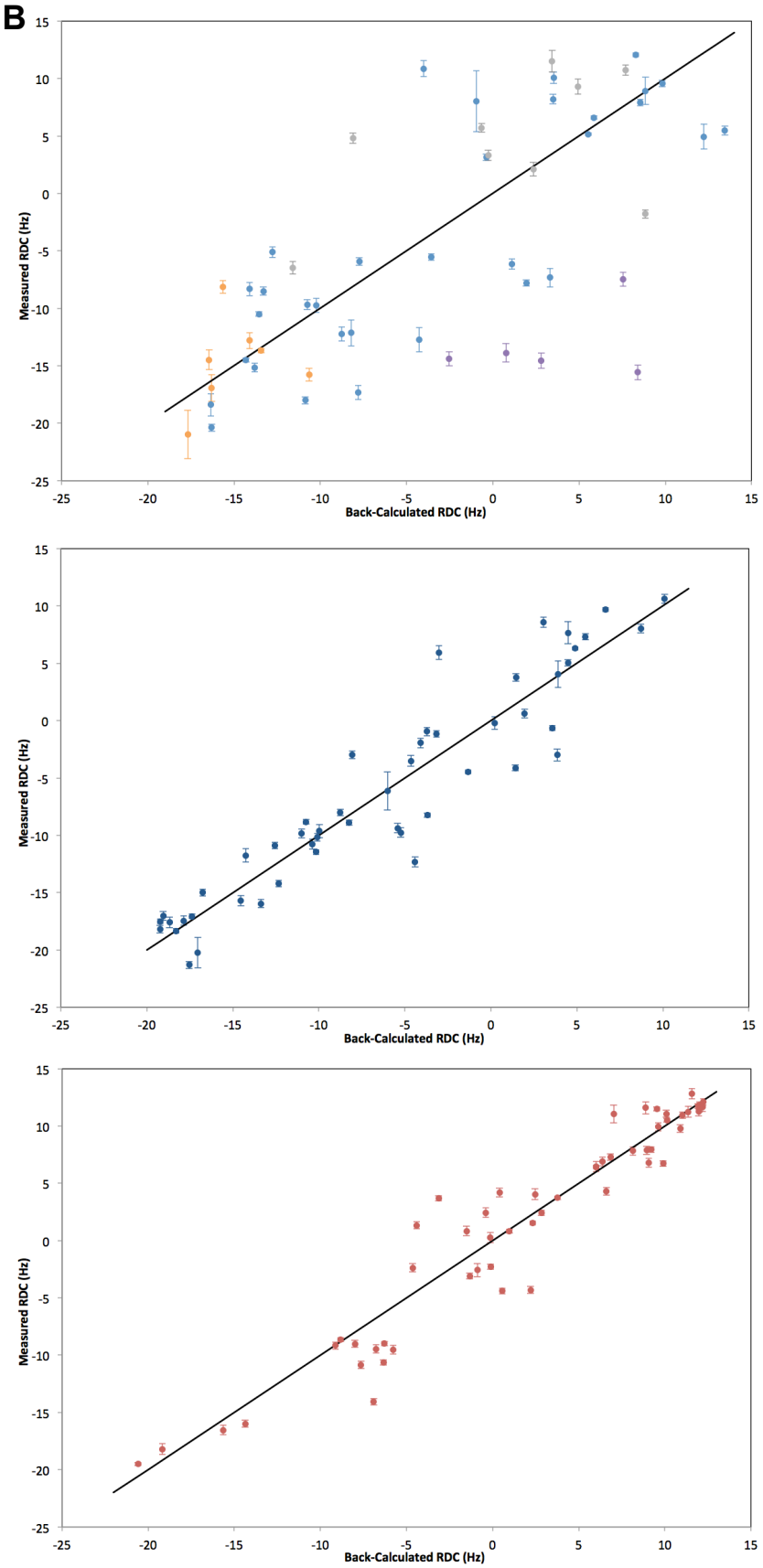


Figure 4.2 – Correlation and linear regression between measured and back-calculated H-N dipolar couplings for the six domains of H-2D^b/FL.
For both the polyacrylamide gel **(A)** and Pf1 phage **(B)** alignment media the six domains of H-2D^b/FL, as defined by RDC data, are; platform (top, light blue), α_1 helix (top, orange), $\alpha_{2\text{-}1}$ helix (top, grey), $\alpha_{2\text{-}2}$ helix (top, purple), α_3 (middle, dark blue), and β_2m (bottom, red). Insufficient data is present for calculation of alignment tensor for the helical domains and for these data the back-calculation is carried out using the platform alignment tensor.

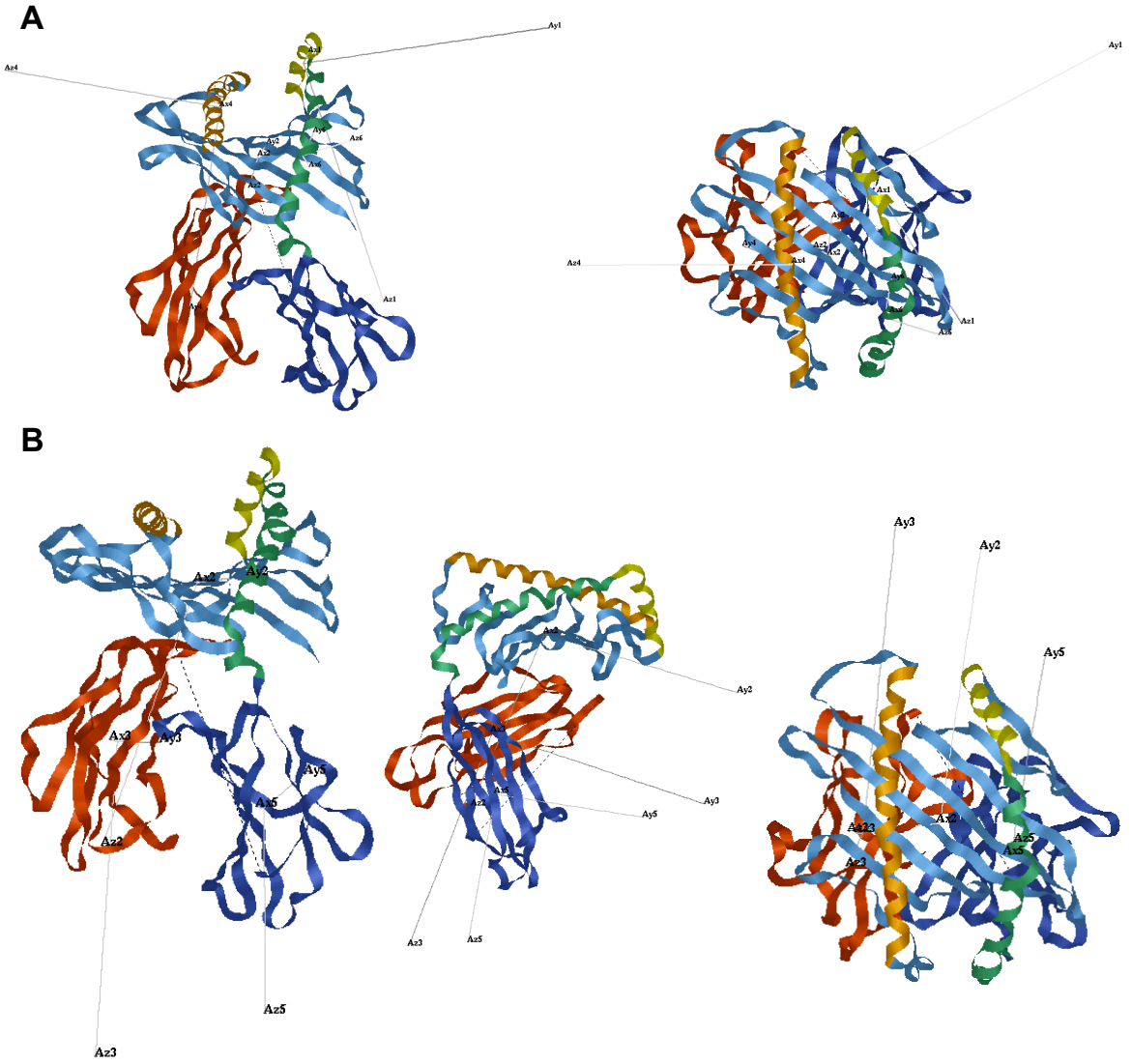


Figure 4.3 – Residual dipolar couplings reveal domain alignment in the solution state of H-2D^b/FL.
Data shown is for the polyacrylamide gel alignment dataset as representative of both results (Table 4.3). **(A)** Alignment within the PBD: α_1 helix (orange, numbered 4), $\alpha_{2\text{-}1}$ helix (yellow, 1), $\alpha_{2\text{-}2}$ helix (green, 6), and platform (light blue, 2). **(B)** Alignment between the platform, α_3 (dark blue, 5), and hβ_2m (red, 3) domains.

4.3.5 Comparison with crystal structure II: REFMAC5

The crystallographic reflection data obtained for H-2D^b/FL-CSNH was expected to be very similar to the Glithero *et al* structure[98] and therefore automated molecular replacement was carried out using MolRep[216] as part of the CCP4 suite[217]. This resulted in poor R factors indicating that the space groups of the two datasets differ; this was resolved using the Phaser[218] molecular replacement tool. A reference set of quality measures (R factor and R_{free}) was created and compared to those produced when the RDC data was included, see Table 4.4.

Table 4.4 – Quality of refined structures produced using REFMAC5.

	X-ray	Polyacrylamide Gel +RDC	X-ray	Pf1 Phage +RDC
Resolution (Å)	3.92		3.92	
R factor	0.154	0.156	0.154	0.156
R_{free}	0.294	0.302	0.294	0.301
Q	0.623	0.358	0.564	0.378
No. of violating RDC data[a]	96/150	36 / 150	38/160	23 / 160
Backbone r.m.s.d (Å)[b]	-	0.081	-	0.079
Ramachandran (%)[c]				
Core	90.8	89.4	90.8	88.7
Allowed	8.2	9.2	8.2	9.9
Disallowed	1.0	1.4	1.0	1.4
Weight on pep1/pep2[d]	-	1.5	-	1.0
Weight of RDCs	-	0.1	-	0.06

[a] The tolerance for RDCs was set to 5 Hz.

[b] Backbone r.m.s.d. to X-ray structure (PDB 1CE6[98]) measured for H, N, Cα, and CO atoms.

[c] Evaluated using MolProbity[219].

[d] Pep1 and pep2 are torsion angle restraints to restrain the planarity of the O_i-C_i-N_{i+1}-$C\alpha_{i+1}$ and C_{i-1}-N_i-$C\alpha_i$-H_i torsions.

During the refinement process it was apparent that the R factor and R_{free} quality measures could not be improved when the NMR data was added, and increasing the relative weight of the RDC data significantly worsened the result. Although the structures generated during the REFMAC5 refinement process could be validated, this inability to improve the data quality statistics is suggestive of a difference between the crystallographic and solution structures[214]. The most important factor to consider here is the Q factor (Equation 4.4) that acts as a direct measure of the correlation between experimental RDC data and that back-calculated from the structure. Initially, the Q factor is extremely poor for both alignment media, with 96/150 and 38/160 RDC values lying outside the generous tolerance of 5 Hz for the polyacrylamide gel and Pf1 phage respectively. Although there is an improvement in Q factor following joint refinement it does not approach Q<0.2 to indicate a good

correlation[214]. The fact that the Q factor remains high shows that there is a significant difference between the crystal structure and the solution state of H-2D^b/FL.

4.3.6 Structure refinement

A test simulation was used to evaluate the effect of the AMBER99SB force field on the structure of H-2D^b/FL in which no RDC restraints were included and the structure simply heated and cooled following the same protocol to be used in combination with the RDC data. This resulted in minor changes to the secondary structure and a small amount of twisting between the PBD and α_3/hβ_2m domains (Figure 4.4A), indicating that the crystal structure may not be at an energetic minimum for a solution state. However, the minor changes observed during this 'no-RDC' simulation did not result in a significant improvement in correlation between experimental RDC values and those back calculated from the no-RDC structure by Module1.0, suggesting that the structure does require additional refinement using the RDC restraints.

The RDC simulation produced a structure that was validated (Table 4.5) and cross-checked against the experimental data using Module1.0 (Figure 4.4B), producing an excellent correlation when run as a single module; this simulation has produced a structure in agreement with the experimental RDCs. The structures of the individual domains remain the same and have the same secondary structure elements (Figure 4.4C), with the exception of the C-D loop of α_3 where the kink around Asn220 is removed in the RDC-restrained simulation; Asn220 in the crystal structure has unfavourable phi/psi angles but in this simulation is within the favoured region of the Ramachandran plot. A large twist between the PBD and α_3/hβ_2m domains is observed and when this is compared to the average structure obtained from the 0.5 μs GROMACS MD simulation carried out by Dr Bailey[164] this twist is away from that observed in the 'GROMACS structure'. However, within the PBD the α-helices are in closer agreement with the 'GROMACS structure' than with the crystal structure.

Experimental uncertainty in the data was accounted for by creating an ensemble of twenty structures, randomly removing 10% of the RDCs from each dataset, with each structure validated and cross-checked. The degree of twist observed varied significantly, as did the orientation of the α-helices, particularly at the carboxyl-terminus of peptide. The pattern of hydrogen bonding observed in the crystal structure was maintained, both within the peptide binding groove and in the H-2D^b α-chain/hβ_2m interface.

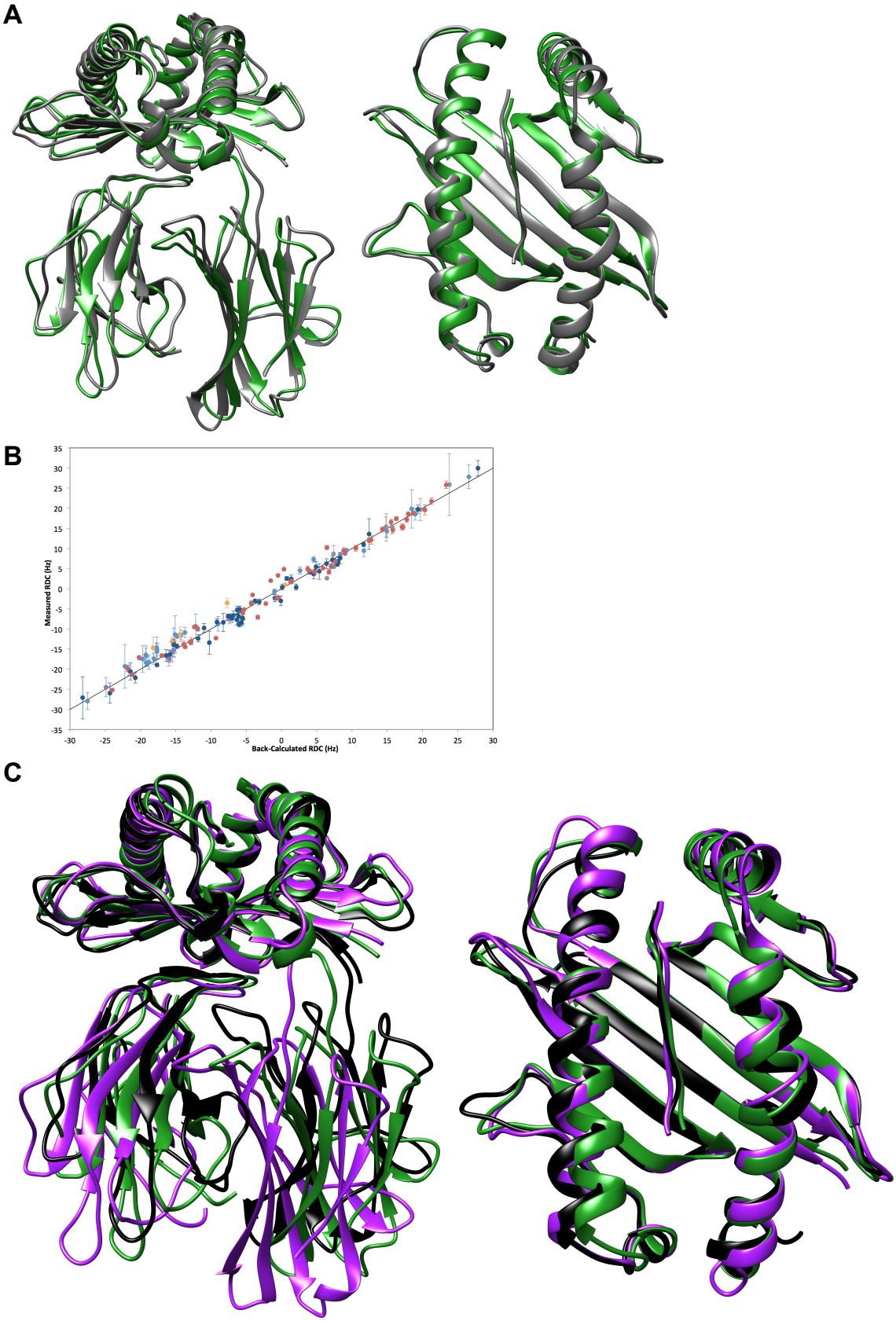


Figure 4.4 – Refinement of crystal structure using RDC restraints.
(A) Simulation in absence of RDC restraints. Overlay of no-RDC Amber-simulated structure (grey) with the crystal structure (PDB 1CE6[98], green) showing relative domain orientations (left) and changes within the PBD (right). **(B)** Correlation between experimental RDC values and those back-calculated from the RDC-refined structure using Module1.0 (colours as per Figure 4.2). **(C)** Simulation including RDC restraints. Overlay of RDC-refined structure (black) with the crystal structure (green) and the average GROMACS-simulated structure (purple), showing relative domain orientations (left) and changes within the PBD (right).

Table 4.5 – Quality of refined structures produced using MD simulation.

	X-ray	Amber simulation in absence of RDC restraints	Amber simulation with RDC restraints
Ramachandran (%)[a]			
Core	90.8	96.4	94.4
Allowed	8.2	2.5	4.3
Disallowed	1.0	1.1	1.3
Backbone r.m.s.d. (Å)[b]	-	2.24	3.21

[a] Backbone r.m.s.d. to X-ray structure (PDB 1CE6[98]) measured for H, N, Cα, and CO atoms.

[b] Evaluated using MolProbity[219].

An individual RDC provides information on the orientation of the N-H vector, restraining it to the surface of a distorted cone about a principal axis of the alignment tensor[203,204]. Since each RDC thus describes a continuum of possible orientations, multiple datasets are required to identify the unique orientation; a base requirement is a 2N dataset such that each residue is characterised by two RDC values[204,178]. For H-2D^b/FL, the polyacrylamide gel and Pf1 phage datasets provide two separate alignment tensors (Table 4.3), thus characterising two N-H vector orientation cones but not fulfilling the 2N requirement[178]. In cases such as this where the dataset does not reach 2N, it is possible to over-interpret the data and obtain an incorrect structure that nonetheless satisfies the RDC restraints[178] and it is therefore important to consider observations of local change as potentially inaccurate. However, this refinement of the existing crystal structure is extremely visually informative in terms of the domain-domain relationship in the solution state (as few as five RDC restraints can be used to inform domain orientations[178] and this dataset far exceeds that requirement).

4.4 Summary

This chapter has described the characterisation of the domain orientations for the solution state of H-2D^b/FL. Residual dipolar couplings were collected via J-modulation and ARTSY experiments using polyacrylamide gel and Pf1 phage alignment media, producing datasets of 150 and 160 RDCs respectively.

Independent analysis of both datasets with Module1.0 resulted in the definition of six domains within H-2D^b/FL; the H-2D^b α-chain platform, α_1 helix, $\alpha_{2\text{-}1}$ helix, $\alpha_{2\text{-}2}$ helix, and α_3 domain along with hβ_2m. These domains had different orientations to what is observed in the structure obtained via X-ray crystallography, the most significant being a 20° rotation between the PBD and the α_3/hβ_2m domains. This apparent difference between the crystal structure and the solution state of H-2D^b/FL was confirmed using REFMAC5 analysis.

The RDC data was also used to drive a structure refinement using a molecular dynamics simulation with the Amber99SB force field, in which a significant reorientation between the PBD and the α_3/hβ_2m domains was also observed. Although there may be over-interpretation of the available RDC data[178] by employing this type of structure refinement, a key observation is that the domain reorientation observed in the MD simulation is of approximately the same amplitude as that observed in the Module1.0 analysis. However, this domain-domain twist does not match that observed when an alternative MD force field (GROMACS) is used[164], but it is clear from both the Module1.0 and REMAC5 analysis that such a twist is very likely to occur in solution.

Chapter 5

5 Solution-state dynamics of H-2D^b/FL

There are a variety of NMR spectroscopic techniques that can be used to obtain information on solution state dynamic motions. This includes measurement of heteronuclear relaxation rates for individual residues, as well as observing the rate of hydrogen exchange and other factors such as spectral lineshape. Previous chapters have suggested that MHC class I may exhibit internal dynamics, and the collection of the above datasets will allow the characterisation of motions on a microsecond to millisecond timescale.

This chapter describes the collection and analysis of lineshape, heteronuclear relaxation, and hydrogen exchange data, leading to the characterisation of the solution state dynamics within free hβ$_2$m and H-2D^b/FL.

5.1 Introduction

Of the different analyses within this chapter, qualitative lineshape analysis does not require further data collection since it requires only an assigned ^{1}H-^{15}N spectral plane. Analysis of heteronuclear relaxation rates is carried out via collection of new data; multiple datasets recorded at multiple spectrometer field strengths enables the characterisation of models of motion. Observation of hydrogen exchange rates is a developing field within protein NMR spectroscopy and various techniques are available.

5.1.1 Lineshape analysis

Qualitative lineshape analysis relies on chemical or conformational exchange (see Section 1.5.4.1); the process whereby a nucleus exchanges between two or more environments between which the NMR-observable parameters differ. These processes may be intermolecular or intramolecular, but for H-2D^b/FL there are none of the latter events and any chemical exchange that is observed can therefore be characterised as due to backbone or side chain motion, or conformational exchange. The off-rates of both the immunodominant FL peptide and hβ$_2$m are so slow they would not significantly contribute to line broadening and therefore NH resonance broadening is predominantly due to conformational exchange or fast hydrogen exchange. Protein side chain motion is expected for any protein in solution and

chemical exchange within H-2D^b/FL will be observed via changes in chemical shift because this type of motion has a small effect on this parameter. The exception to this is the movement of aromatic side chains; the delocalised electrons within aromatic rings precess in a magnetic field and can have a significant effect on nearby residues. The final and most significant intramolecular cause of chemical exchange is conformational motion, separate from the low-level side chains fluctuation, in which the protein backbone explores the available conformational landscape and may switch between different conformational states that represent different energetic minima on this landscape. In H-2D^b/FL such line broadening will not be due to an equilibrium between peptide-bound and peptide-free states because the FL peptide off-rate is slow (hours) and thus will not be observed on the millisecond timescale.

5.1.2 Heteronuclear relaxation

The mechanism of nuclear relaxation relies on magnetic interactions such as the fluctuation of the dipole-dipole coupling that causes nuclei to experience varying local magnetic fields. These are time-dependent due to the rotational motion of the protein. Heteronuclear relaxation of a ^{15}N nucleus is a convenient probe of protein motion as its relaxation is dominated by the bound ^{1}H, allowing the measurement of three factors; longitudinal (^{15}N T_1) and transverse (^{15}N T_2) relaxation parameters, along with the heteronuclear ^{1}H-^{15}N NOE (for details see Section 1.5.4.2).

The simplest set of experiments to obtain heteronuclear relaxation data employ an inversion recovery (for T_1), a spin-echo (for T_2), and saturation of target nuclei (for the heteronuclear NOE). In practice, these pulse sequences are optimised for sensitivity, with TROSY-based sequences used for large proteins. For T_1 and T_2 experiments, the choice of τ delays is determined on a case-by-case basis by obtaining low-sensitivity, low-resolution spectra immediately prior to full data collection. For a full description of the pulse sequences used for collection of the data discussed in this chapter see Zhu, Xia, Nicholson, and Sze[220].

It is possible to predict the expected ^{15}N relaxation rates for a given protein using HYDRONMR software. This utilises a computational procedure to calculate the rotational diffusion tensor and other properties from an atomic-level structure, as is available for both free hβ_2m and H-2D^b/FL. It is based on the creation of a bead model in which the hydrodynamic properties of arbitrarily shaped particles can be determined. For details of the processes employed, see de la Torre, Huertas, and Carrasco[221].

5.1.2.1 The Lipari-Szabo model-free method

Heteronuclear relaxation analysis initially yields T_1, T_2, and NOE values for each residue of a protein. This data can be further analysed to reveal information on protein dynamics by applying the Lipari-Szabo model-free approach. The use of dynamic models can be useful but it is easy to over-interpret the data, and the chosen model cannot be proven, only eliminated from the analysis. G. Lipari and A. Szabo introduced the model-free approach in 1982, both in theory[222] and in practice[223]. In brief, the motion of a protein is regarded as being separable into overall tumbling (either isotropic or anisotropic) and internal motions. The fast internal motions (that are of particular interest for MHC class I) can then be characterised by two model-independent variables. A generalised order parameter, S, describes the spatial restriction of motion, and an effective correlation time, τ_e, describes the timescale of this motion. These variables are treated as parameters to be optimised; they have a simple relationship to spectral density (Equation 5.1) such that they can be adjusted to fit relaxation data via a least-squares fit[222].

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_M}{1+(\tau_M\omega)^2} + \frac{(1-S^2)\tau}{1+(\tau\omega)^2}\right)$$

Equation 5.1

where $J(\omega)$ is the spectral density, S^2 the generalised order parameter, τ_M the correlation time describing the overall motion with the relationship to the effective internal correlation time, τ_e, of: $\tau^{-1} = \tau_M^{-1} + \tau_e^{-1}$.

This formalism was extended by Clore *et al*[224] to include fast motions on two timescales with order parameters S^{2f} and S^{2f} for the fast and slow motions respectively. In this extended model the overall order parameter S^2 is the product of the fast and slow parameters. ModelFree[225,226] is a program that implements the model-free approach using a least-squares fit to the measured relaxation data, assuming dipolar and chemical shift anisotropy relaxation mechanisms for T_1, T_2, and NOE as well as including an additional term (R_{ex}) to account for the contribution to T_2 of chemical exchange broadening. The first step in analysis of a dataset is to estimate τ_M; this can be done within ModelFree. Next, an isotropic model is applied to the data in which D_{II} and $D\perp$ (Figure 5.1) are identical such that the molecule is assumed to be spherical. If necessary, the anisotropic model can be applied in which D_{II} and $D\perp$ are independently varied. Model selection involves a choice between five alternative models for internal motion; Model 1 is based on S^{2s} (the generalised order parameter S^2 can be split into both slow (s) and fast (f) components); Model 2, S^{2s} and τ_e; Model 3, S^{2s} and R_{ex}; Model 4, S^{2s}; τ_e, and R_{ex}; and Model 5, S^{2f}, S^{2s}, and τ_e. Choice of model is

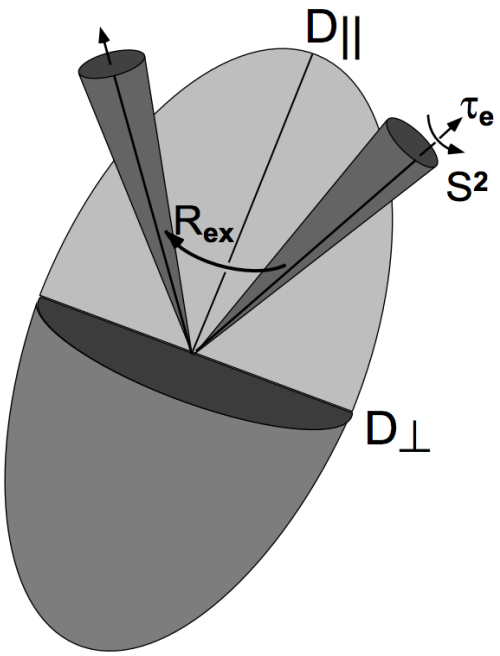


Figure 5.1 – ModelFree parameters
D_{II} and D⊥ describe molecular anisotropy, S^2 is spatial restriction of motion, τ_e describes the timescale of this motion, and R_{ex} permits two-site chemical exchange.

made on a per-residue basis using the F-statistic (Equation 5.2). Finally, the overall rotational diffusion (τ_M) and the chosen internal parameters for each residue are simultaneously optimised. Thus, ModelFree enables analysis of both global and local motions.

$$F = [d_2/(d_1 - d_2)]\,[SSE_1 - SSE_2]/SSE_2$$

Equation 5.2

where SSE_1 and SSE_2 are the sum of squared errors of prediction for models 1 and 2 with d_1 and d_2 degrees of freedom ($d_1 > d_2$)

5.1.3 Hydrogen exchange

The interaction between a protein and its solvent is a useful method to probe both protein structure and protein dynamics. For a region tightly organised in regular secondary structure less exchange is expected than for a residue at the protein-water interface, thus informing on the protein structure. The exception to this relationship is when an area of a protein undergoes motion, thus exposing residues to exchange that otherwise might be buried within the protein core.

There are a number of interaction between protein and solvent, the first of which is the physical exchange of backbone amide protons with solvent protons. Supplying deuterated solvent enables hydrogen-deuterium exchange to be followed using NMR spectroscopy or mass spectrometry. Rapidly exchanging protons can be investigated

using the selective application of magnetisation to solvent protons; the millisecond transfer of magnetisation to amide protons can easily be followed using NMR.

Hydrogen-deuterium exchange

Information about the dynamics of protein can be obtained from hydrogen-deuterium exchange (HDX). Hydrogen attached to the nitrogen of a peptide bond is able to exchange with solvent such that supplying a deuterated solvent will result in the incorporation of deuterium. This exchange occurs at physiological pH and lowering the pH below the minimum required for exchange halts the reaction. Following protein digestion, peptides are separated by HPLC to prevent back-exchange, after which the location of deuteration can be determined by electrospray ionisation mass spectrometry. The rate of deuteration is dependent on how accessible a proton is to solvent; in a structured (folded) protein this accessibility is governed by the degree of flexibility such that a fast HDX rate for a residue located within regular secondary structure is indicative of dynamics in its vicinity that make the protein effectively locally porous to solvent.

Hydrogen exchange

Hydrogen exchange (HX) refers to the process by which protons of the protein backbone amide are interchanged with those of the solvent. The rate of exchange is strongly dependent on pH and temperature[227]; for consistency experiments are generally conducted at pH 7 and 25°C. The degree of hydrogen bonding is the largest limiting factor in HX exchange because strongly bonded residues are protected from exchange. A further consideration is that residues buried within the protein structure are better protected than those that are surface exposed, such that a structural change is required to release protected protons. Measurement of HX exchange rates coupled with knowledge of the protein secondary structure therefore give insight into protein flexibility[227].

There are a number of kinetic factors that must be taken into consideration when measuring HX exchange rates, including T_1 relaxation and NOE peak attenuation[227]. The basic premise of a hydrogen exchange experiment is the saturation of water magnetisation such that no magnetisation resides on protein. After a variable delay an HSQC spectrum can be collected to measure the transfer of magnetisation to amide protons; as the delay time increases peak intensity will also increase. There are multiple variations of this basic method, the most common of which is the CLEANEX-PM approach; for a discussion of the pulse sequence see Hwang, van Zijl, and Mori[228].

In this method the exchange rates are extracted from change in peak intensity over time[229];

$$I = I_{\infty}(I - e^{-h_x t_m})$$

Equation 5.3

$$I = I_{\infty} e^{-\tau_m / T_1} (1 - e^{-h_x \tau_m})$$

Equation 5.4

where I is peak intensity, I_{∞} is maximum peak intensity, h_x is the exchange rate, and τ_M the correlation time

Equation 5.3 is applied in cases where the maximum intensity has not yet been reached for the time delays used during data collection, whilst Equation 5.4 is applicable where the peak intensity has begun to decrease due to T_1 relaxation.

5.2 Dynamics of hβ$_2$m

A significant amount of data has been collected regarding the dynamics of H-2D^b/FL, necessitating the definition of sub-topics that are more manageable in size. The first of these is the dynamics of hβ$_2$m, with data characterising the behaviour of both hβ$_2$m when present both isolated in homogeneous solution and as a domain within H-2D^b/FL. Choosing to split the analysis of H-2D^b/FL into α-chain and hβ$_2$m permits the ready examination of changes to the dynamics of hβ$_2$m when it binds to MHC class I.

5.2.1 Materials and methods

Both H-2D^b/FL-bound hβ$_2$m and isolated hβ$_2$m have been characterised using qualitative lineshape analysis, heteronuclear relaxation, and hydrogen exchange.

5.2.1.1 Qualitative lineshape analysis

Dr Garrick Taylor kindly provided the assigned HSQC spectrum for isolated hβ2m. H-2D^b/FL-bound hβ$_2$m linewidth analysis utilised the TROSY spectrum previously used in the assignment of the H-2D^b α-chain (Figure 3.4).

The assignment of additional peaks for a residue (those arising from chemical exchange; Section 1.5.4.1, Figure 1.8) relied on the presence of alpha carbons at extremely similar chemical shifts as for the first peak (the sequential linking of residues using trHNCA strips must still apply), and on the presence of an extremely similar carbonyl chemical shift. To rule out any contribution of protein unfolding to the observation of doubled peaks, control fast-HSQC spectra were collected before and after recording the main dataset.

It should be noted that the linewidths discussed in this thesis are *apparent* linewidths due to the use of a window function during data processing. This processing is necessary to obtain sufficient data quality and does not impact on the qualitative nature of the data analysis employed.

5.2.1.2 Collection of heteronuclear relaxation data

Relaxation data for 500 μM $^{13}C^{15}N$ isolated hβ$_2$m was collected at 700 MHz on a Bruker spectrometer using a 5 mm HCN PFG cryoprobe with a cooled ^{13}C preamplifier, at the at the MRC Biomedical NMR Centre, National Institute for Medical Research, London. To obtain T_1 values, seven two-dimensional ($^1H^{15}N$) spectra were collected, with time delays of 0.02, 0.06, 0.12, 0.25, 0.40, 0.80, and 1.20 seconds. Nine

spectra were collected for T_2 measurement, with delays of 8, 16, 24, 32, 40, 48, 72, 96, and 144 milliseconds. ^{15}N NOE data was also collected on this protein sample.

Relaxation data for 600 μM $^{2}H^{13}C^{15}N$ H-2D^{b}/FL-bound hβ_2m ($^{2}H^{15}N$ H-2D^{b} α-chain and unlabelled FAPGNYPAL peptide) was collected at 700 MHz (NIMR, London). To obtain T_1 values, seven two-dimensional spectra were collected, with delays of 0.01, 0.04, 0.10, 0.30, 0.60, 1.20, and 1.40 seconds. Seven spectra were also collected for T_2 measurement, with delays of 4, 8, 12, 20, 28, 40, and 60 milliseconds. ^{15}N NOE data was also collected.

Relaxation data for 600 μM $^{2}H^{15}N$ H-2D^{b}/FL-bound hβ_2m ($^{2}H^{13}C^{15}N$ H-2D^{b} α-chain and unlabelled FAPGNYPAL peptide) was also collected at 950 MHz on an Oxford Instruments spectrometer using a HCN PFG cryoprobe at the University of Oxford. To obtain T_1 values, seven spectra were collected with τ delays of 0.02013, 0.10063, 0.30190, 0.60380, 1.20761, 2.41522, and 4.83043 seconds. To obtain T_2 values another seven spectra were collected, with delays of 4.976, 9.952, 14.928, 19.904, 24.880, 29.856, and 74.640 milliseconds. NOE data were not collected.

Nominal T_1 and T_2 values for different correlation times and S^2 values were calculated based on the equations set out by Lipari and Szabo[222,223], using a custom Linux script created by Dr Joern Werner, to create 'Lipari-Szabo graphs'. This script is also capable of calculating theoretical values for ^{15}N T_1, T_2, and NOE as a function of an isotropic correlation time τ_m, with internal dynamics of $S^2<1$, $\tau_e>0$. The mechanism of relaxation is assumed to be a combination of ^{1}H-^{15}N dipolar relaxation and ^{15}N chemical shift anisotropy. HYDRONMR[221] was used to predict ^{15}N T_1 and T_2 values for isolated hβ_2m at 700 MHz and for H-2D^{b}/FL-bound hβ_2m at 950 and 700 MHz using the water viscosity at 298K; beads of radius 3.2 Å were placed at Cα positions using the structure PDB 1CE6[98].

5.2.1.3 Hydrogen-deuterium exchange

HDX analysis of both isolated (21 μM) and H-2D^{b}/FL-bound hβ_2m (9 μM) was carried out using a Synapt G2S mass spectrometer in HDMSe mode (Waters, Manchester). 15-20 fold excess D_2O was added to samples in H_2O solvent, pH 7.0, at 20°C. Following protein labelling, the pH was reduced to 2.5 to quench the reaction; time points of 0.0, 0.5, 1.0, and 2.0 minutes were used prior to pepsin digestion and HDMSe analysis. Disulphide bridges were reduced using TCEP during the quench step prior to digestion (i.e. performed after labelling with no effect on protein accessibility to deuterium). The deuterium uptake evidenced in the data was monitored using the

Waters ProteinLynx Global Server to identify the peptide products. Waters DynamX software was used to interpret the results and plot the deuterium uptake.

5.2.1.4 Hydrogen exchange

HX data was collected for 100 μM $^{2}H^{15}N$ isolated β_2m at 600 MHz (University of Southampton). Nine two-dimensional spectra were collected, with mix times of 0.02, 0.05, 0.12, 0.20, 0.30, 0.50, 0.70, 1.00, and 1.5 seconds. During data analysis both methods for estimating exchange rate were applied, with the final exchange rates determined using Equation 5.3; inclusion of the T_1 parameter did not significantly affect the fitted exchange rate constant and the introduction of this additional variable unnecessarily increased the error of the calculation, and therefore Equation 5.4 was not applied. To exclude the effect of T_1 relaxation on peak intensity the exchange rates were calculated using the time points up to and including 0.5 seconds; this did not significantly alter the predicted rate.

A similar HX approach was applied to H-2D^b/FL-bound hβ_2m; H-to-D exchange was measured for 450 μM ^{15}N-labelled H-2D^b/FL-bound hβ_2m via a CLEANEX-PM experiment[228] on a Varian Inova 800 MHz spectrometer equipped with a 5 mm triple resonance (HCN) PFG cryogenic probe, at the MRC Biomedical NMR Centre, National Institute for Medical Research, London. Exchange delays of 0.025, 0.050, 0.075, 0.100, 0.125, 0.150, 0.175, 0.200, 0.225, and 0.250 milliseconds were used.

In addition to the CLEANEX-PM method, D-to-H exchange was measured for 400 μM ^{15}N-labelled H-2D^b/FL-bound hβ_2m at 950 MHz (University of Oxford). Following collection of a reference TROSY spectrum, H_2O was buffer exchanged for D_2O before a second spectrum was collected.

5.2.2 Results

5.2.2.1 Qualitative lineshape analysis

No peak position changes were observed in two control HSQC spectra for either isolated or H-2D^b/FL-bound hβ_2m so no protein unfolding occurred. The average apparent proton linewidth for isolated hβ_2m was 36 Hz (Figure 5.2A), with broad linewidths defined as those greater than 40 Hz (Table 5.1). The β-strands do not contain broad linewidths, with the exception of Leu32 and Val27 in the B strand. The HSQC spectrum did not contain sufficient peaks for full assignment (88 peaks instead of 94) and residues with broad linewidths are mostly adjacent to unassigned areas. Although 68 residues were assigned with 20 unassigned peaks in the spectrum, it is reasonable to predict that the residues missing due to chemical exchange broadening are located in either the B-C or D-E loops, because these unassigned regions are bordered by particularly broadened peaks.

The average apparent proton linewidth for MHC class I-bound hβ_2m was 21.6 Hz (Figure 5.2), with broad linewidths defined as greater than 30 Hz. In addition to several broadened peaks, six doubled peaks were observed (Table 5.1). The average linewidth of bound β_2m is significantly smaller than that of isolated hβ_2m since the TROSY experiment selects only the sharpest component of the ^{1}H-^{15}N doublet; as such the linewidth values are not directly comparable. Peaks affected by exchange are not located within β-strands with the exception of Asn21, Leu23, and Val27 in the B strand. These three residues are also broadened in isolated β_2m, and therefore have not been stabilised by the presence of the H-2D^b α-chain despite the fact that Asn24 and Tyr26 are both present at the interface (Table 5.1, Figure 5.3). Thr86 in the F-G loop also remains broadened, with Thr71 and Thr73 in the E-F loop now doubled; both areas are distant from the H-2D^b α-chain. The B-C loop underneath the PBD is broadened, and the D-E loop underneath the PBG is both broadened and doubled, indicating that although these regions have been stabilised by the presence of H-2D^b α-chain sufficient for assignment they remain affected by conformational dynamics. It is worth noting that motion in the D-E loop may be disproportionately represented due to the presence of Trp60, whose side chain may induce significant ring current shifts.

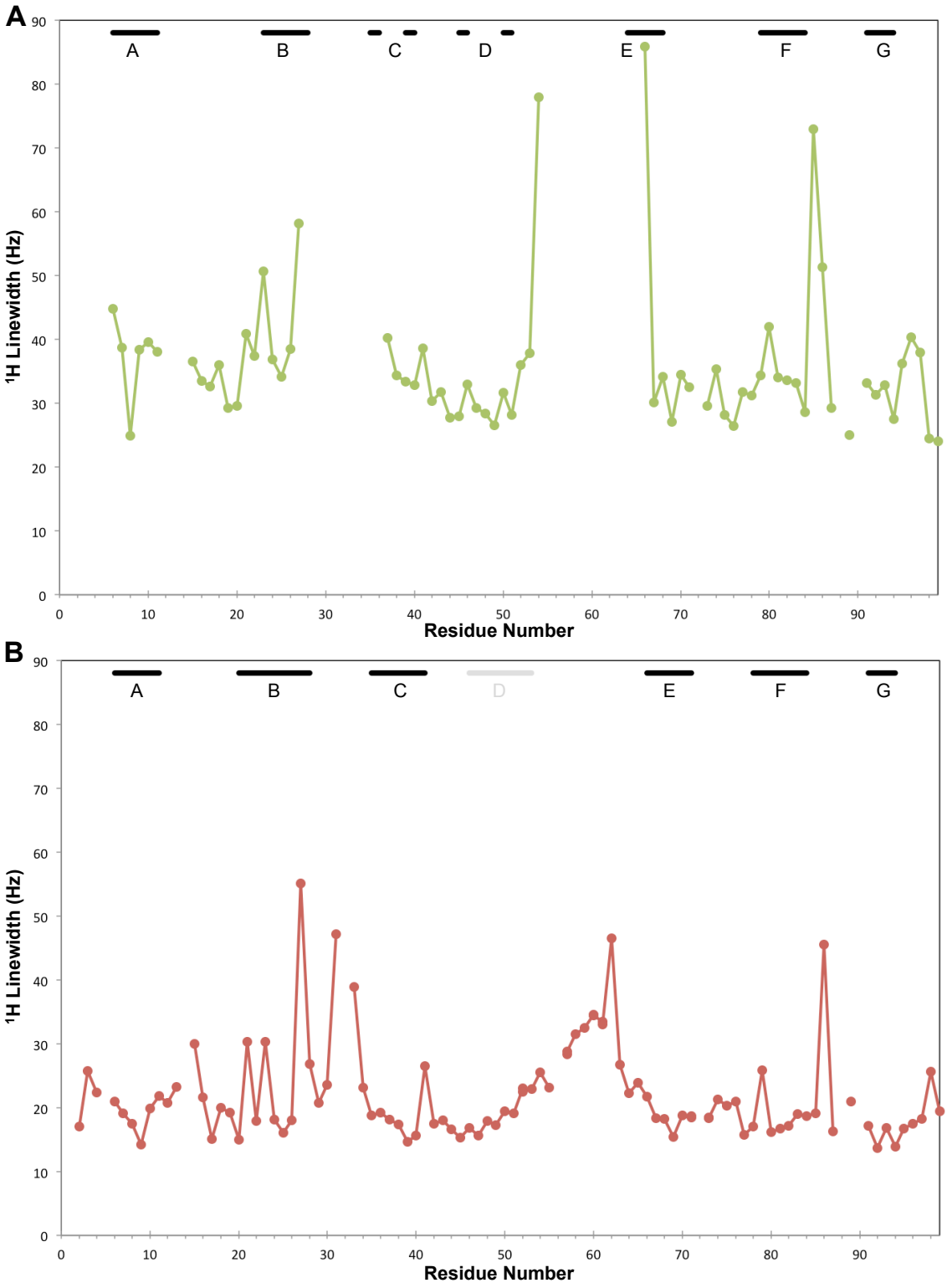


Figure 5.2 – ^{1}H linewidths for isolated and H-2D^b/FL-bound hβ_2m.
Location of β-strands in **(A)** isolated hβ_2m (PDB:1JNJ[134]) and **(B)** H-2D^b/FL-bound hβ_2m (PDB: 1CE6[98]) are indicated by solid black lines, labelled A-G. Doubled peaks indicated by presence of two linewidth values.

Table 5.1 – Location of broad linewidths and doubled peaks in isolated and H-2D^b/FL-bound hβ_2m.

	Broad Linewidth	Doubled Peak
Isolated hβ_2m	Lys6, Asn21, Leu23, Val27, Val37, Leu54, Tyr66, Cys80, Val85, Thr86, Asp96	
H-2D^b/FL-bound hβ_2m	Asn21, Leu23, Val27, His31, Ser33, Lys58, Asp59, Trp60, Ser61, Phe62, Thr86	Ser52, Ser57, Trp60, Ser61, Thr71, Thr73

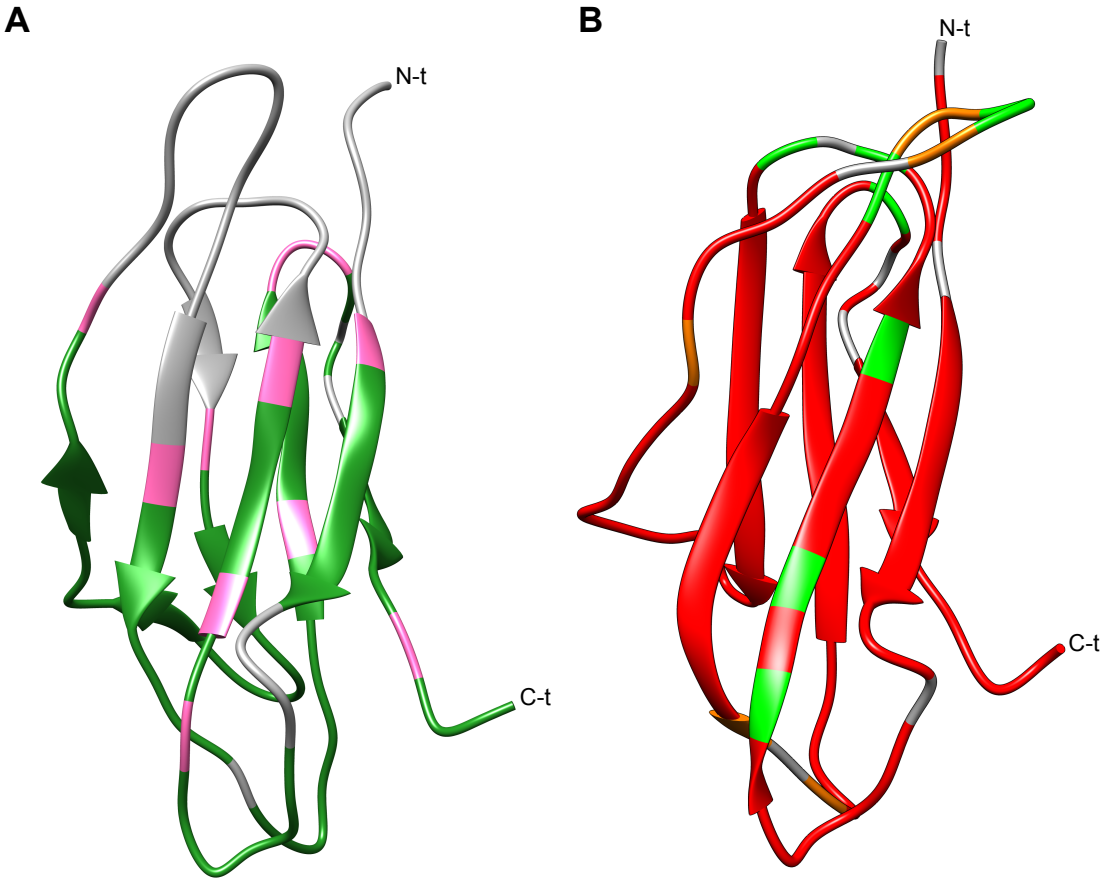


Figure 5.3 – Location of broad linewidths and doubled peaks in isolated and H-2D^b/FL-bound hβ_2m.
(A) Isolated β_2m (green) (PDB: 1JNJ[134]). Unassigned residues are shown in grey (including Pro), broad linewidths in pink. **(B)** MHC class I-bound β_2m (red) (PDB: 1CE6[98]). The H-2D^b α-chain is omitted for clarity; the PBD lies above the N-terminus whilst the α_3 domain lies between the viewer and β_2m. Unassigned residues are shown in grey (including Pro), broad linewidths in light green, and doubled peaks in orange.

5.2.2.2 Heteronuclear relaxation of isolated hβ$_2$m

The average ^{15}N T_1 relaxation for 500 μM isolated hβ$_2$m at 700 MHz is 875 ms, the average T_2 relaxation 65 ms, whilst the average ^{15}N NOE is 0.793 (Figure 5.4). A low T_1 or a high T_2 indicates additional internal mobility on a fast timescale, and thus this data agrees with the qualitative lineshape analysis. Unassigned areas are mostly bordered by low T_1 and high T_2 values, whilst the mobility of the B-C loop observed in linewidth analysis is confirmed by both T_1 and T_2. However, mobile residues do not border the unassigned D-E loop (although there are decreased ^{1}H-^{15}N NOE values) suggesting that the broad linewidths here might be due to the presence of a ring current induced by Trp60. The E-F loop appears the most mobile region of isolated hβ$_2$m, with the exception of the expected plasticity observed in the C-terminal region. The D strand appears to be notably stable in the T_1 analysis, but less so in the T_2 and NOE data; some degree of flexibility is anticipated here due to the β-bulge destabilisation.

Relaxation values correlate with molecular weight ($\tau_c \propto M^w$; Figure 1.9) such that a T_2 value of approximately 100 ms is expected for a protein of 11.8 kDa and the observed 65 ms is more consistent with a 20 kDa protein. There are a number of possible causes for this, one of which is the presence of transient self-association introduced by the relatively elevated 500 μM protein concentration. Permanent aggregation can be discounted by the presence of well-dispersed peaks, with no changes observed between a set of control HSQC spectra taken before and after collection of relaxation data. In addition to transient self-association, slow internal dynamic processes may contribute to line broadening, which can be investigated by examining the available data in the context of a Lipari-Szabo graph and by comparing it to values predicted by HYDRONMR[221] (Figure 5.5).

Following the principles set out by Lipari and Szabo[222,223] the graph is created such that ^{15}N-T_1 and -T_2 values for various protein models are presented. The first of these, the S^2=1 model, is the prediction for a rigid structure in which there is no internal motion. The S^2=0.9 line represents a model in which the protein is mostly rigid with a small element of fast motion permitted; with decreasing S^2 values the level of fast motion modelled increases.

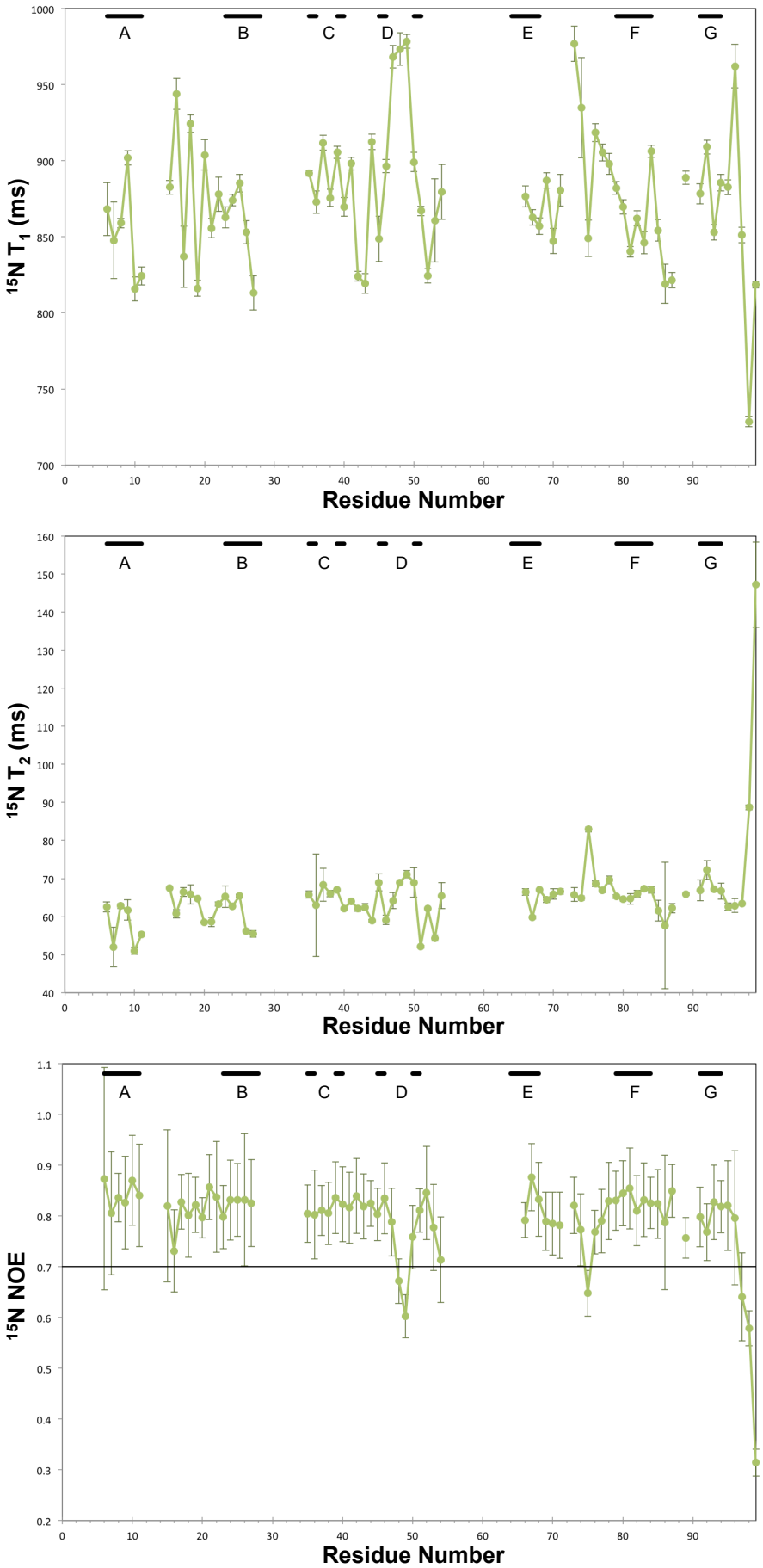


Figure 5.4 – ^{15}N heteronuclear relaxation data for isolated hβ$_2$m at 700 MHz.
^{15}N T_1 (top), ^{15}N T_2 (middle) and ^{1}H-^{15}N NOE (bottom) values are shown. Location of β-strands in PDB:1JNJ[134] are indicated by solid black lines, labelled A-G. For NOE data, residues whose values fall below the 0.7 cut-off line are considered notably flexible.

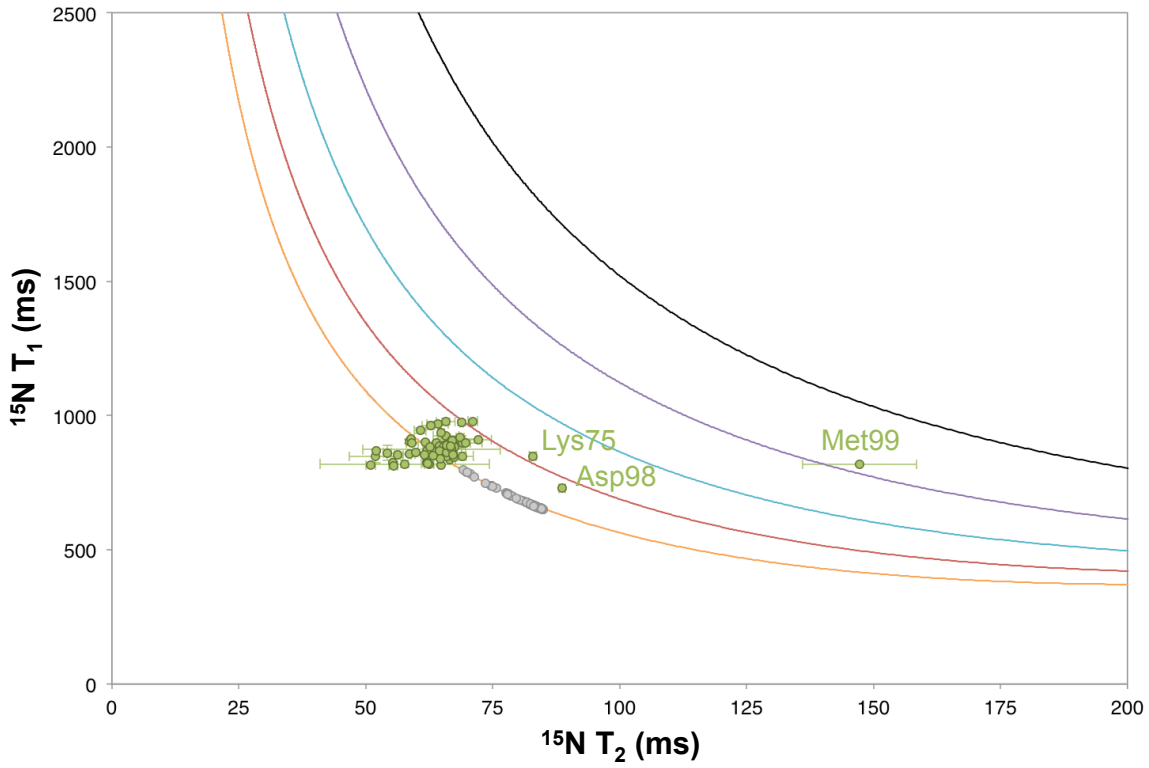


Figure 5.5 – Lipari-Szabo graph for 500 µM isolated hβ$_2$m at 700 MHz.
Lipari-Szabo lines represent various models of motion; S^2=1 (orange), S^2=0.9 (red), S^2=0.8 (blue), S^2=0.7 (purple), S^2=0.6 (black). Notable outlying residues are labelled. HYDRONMR data is depicted in grey.

500 µM isolated hβ$_2$m produces relaxation data that largely fits the Lipari-Szabo rigid model (S^2=1, Figure 5.5) with the majority of residues located between S^2=1 and S^2=0.9. There is variation due to fast motion that effects several residues, particularly in the C-terminus, likely due to the lack of regular secondary structure in this region. A significant minority of residues fall below the S^2=1 line and following the above explanation this appears to suggest they experience additional mobility on top of the overall tumbling on a slow timescale. This is consistent with both transient self-association or slow conformational exchange on a microsecond to millisecond timescale[223].

HYDRONMR utilises a rigid model and therefore data sits on the S^2=1 line of the Lipari-Szabo graph (Figure 5.5). There is a spread of data due to protein anisotropy. It is clear that the experimental data is not in agreement with values predicted by HYDRONMR for isolated hβ$_2$m, largely due to increased T_1 values. Again, this is consistent with an increased overall tumbling time.

5.2.2.3 Hydrogen deuterium exchange for isolated hβ$_2$m

Hydrogen-deuterium exchange was carried out on isolated hβ$_2$m (Figure 5.6) using mass spectrometry. It was not possible to generate sufficient peptides for complete overlapping coverage, particularly in the D-E loop (residues 55-64). There is fast uptake indicating easy accessibility to solvent at the N-terminus but not the C-terminus, which is at odds with the heteronuclear relaxation data that suggests the latter is flexible and therefore unstructured. There is also fast uptake in peptides corresponding to the D-E and E-F loops. Regions with slower deuterium uptake are the B and C strands and the B-C loop, as well as the G strand. This largely corresponds to expectations that residues in regular secondary structure will be slower to exchange.

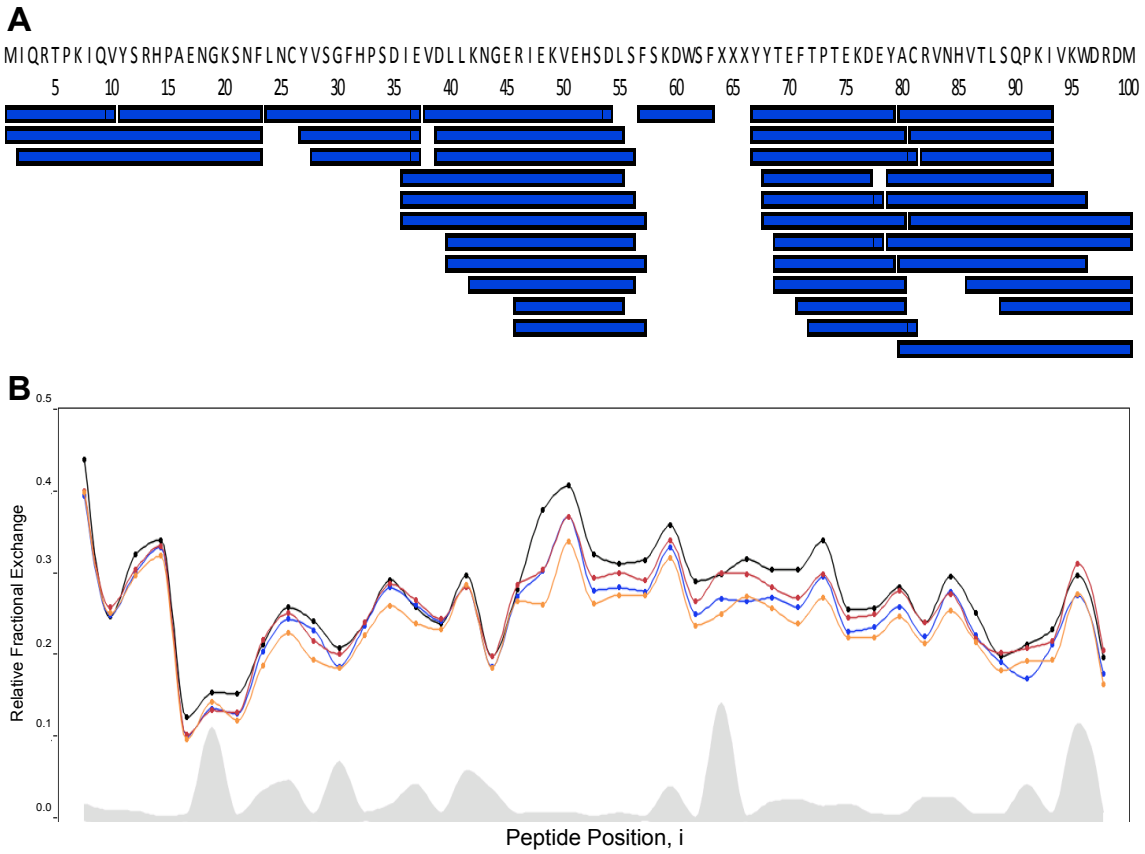


Figure 5.6 – Hydrogen-deuterium exchange for isolated hβ$_2$m.
(A) Peptide coverage map. 97% of the protein sequence was covered, although multiple overlapping peptides for the region 55-64 could not be obtained. **(B)** Deuterium uptake for time points 0.5 minutes (yellow), 1.0 minutes (blue), 2.0 minutes (red), and 5.0 minutes (black). Peptide position 'i' is determined by arranging the peptide midpoints in linear numerical order. A slow uptake is indicated by decreased values. Each point represents a peptide corresponding with (A). The shaded grey region is the estimated error in exchange rate.

5.2.2.4 Hydrogen exchange for isolated hβ$_2$m

Hydrogen exchange data for isolated β$_2$m shows that there is significant variation in the exchange rates (Figure 5.7A). The exchange requires conformational motion to enable hydrogens to be released from the densely packed protein structure, such that regions containing many hydrogen bonds are less susceptible to motion and therefore expected to have slower exchange rates (Figure 5.7B). Nonetheless, it is clear that some loop regions such as the E-F loop (residues 69-77) have extremely rapid exchange, and some unassigned loop regions are bordered by residues with rapid exchange, in agreement with the heteronuclear relaxation data.

Two notable areas of β-sheet that possess dynamics are the A strand, and the D strand immediately around the β-bulge. These strands are both at the edge of the immunoglobulin fold secondary structure, and hence the network of hydrogen bonds does not protect the alternate, outward-facing side chains. This explains the fast exchange observed for some residues within these strands. A number of assigned residues were not observable in the spectra recorded during this experiment, suggesting their exchange timescale is slower than 1.5 seconds. This includes Gln8 of the A strand, which is likely strongly protected from exchange by the presence of two hydrogen bonds (to hβ$_2$m Tyr26). This is also true of Tyr10 (two hydrogen bonds to Asn24) and Lys6 (one hydrogen bond to Ser28).

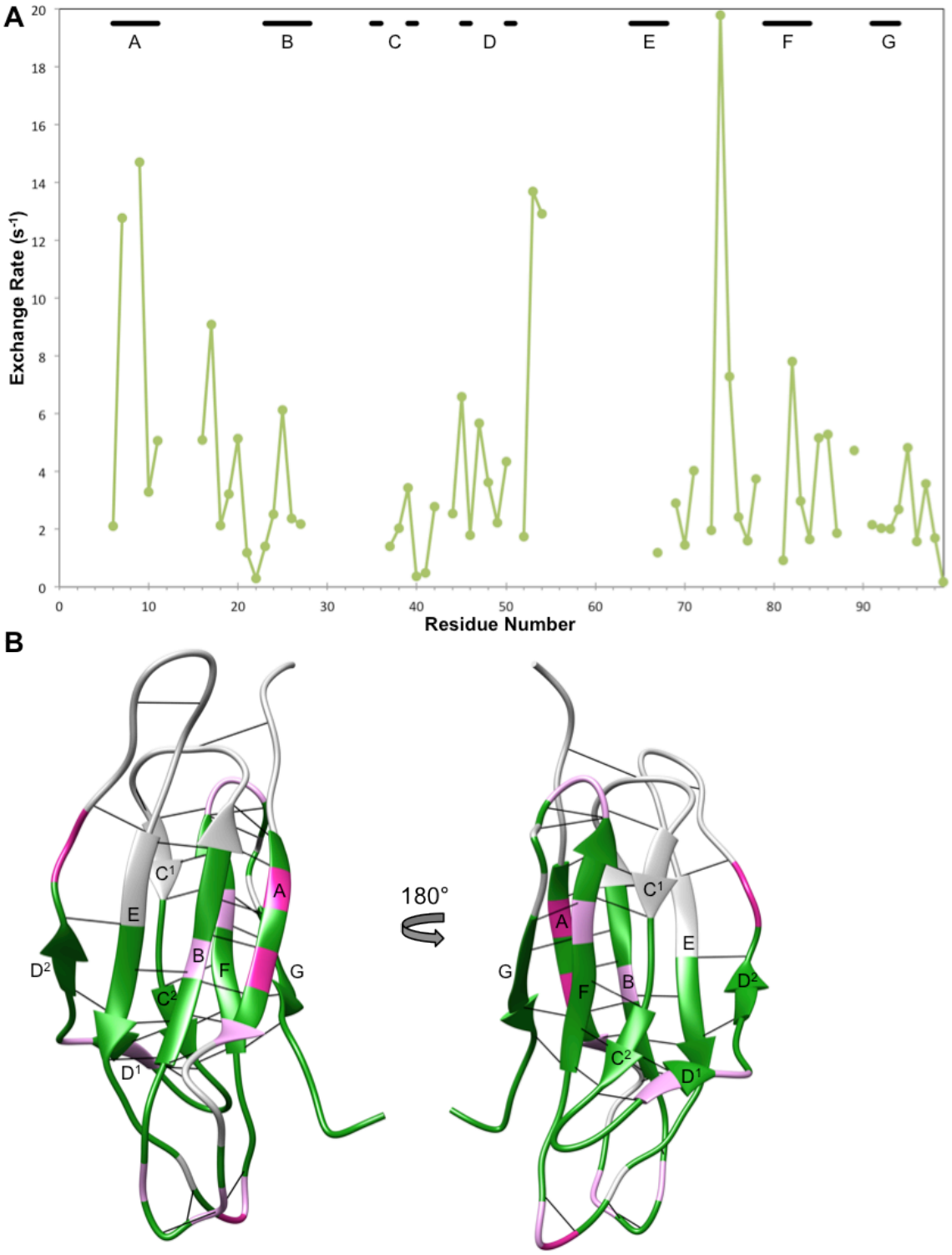


Figure 5.7 – Hydrogen exchange for isolated hβ$_2$m.
(A) Hydrogen exchange rates. Secondary structure (conforming to PDB 1JNJ[134]) is shown above in solid black lines, labelled A-G. **(B)** Location of rapid hydrogen exchange on isolated hβ$_2$m (PDB: 1JNJ[134]). Rates of 5-10 s^{-1} are shown in light pink and rates above 10 s^{-1} in dark pink, whilst unassigned residues (including Pro) are shown in grey, with hydrogen bonds as black lines.

5.2.2.5 Heteronuclear relaxation of MHC class I-bound hβ$_2$m

^{15}N relaxation data has been collected at two field strengths for H-2D^b/FL-bound hβ$_2$m, with the average values given in Table 5.2. When the isolated and H-2D^b/FL-bound data is compared there is higher T_1 and lower T_2 for the former as anticipated for a larger protein (Figure 1.9). In addition, qualitatively the expected field dependence, particularly for ^{15}N T_2 data, is observed between the 700 and 950 MHz datasets. This is both reassuring for the data quality and is an important source for the data interpretation.

At 950 MHz there is a significant uncertainty apparent in the measurements (Figure 5.8,

Table 6.6) such that only general observations can be made. Ile7, and to a lesser extent Ile35, are outliers despite not being affected by interference from other peaks in close proximity on a $^{1}H^{15}N$ plane (Figure 3.1, Figure 3.4); with a lower T_1 and higher T_2 both these residues appear exceptionally flexible. The C-terminus is less flexible relative to the remainder of β_2m than for the isolated form, suggestive of a stabilisation upon H-2D^b α-chain binding. The loop regions are more flexible than the β-sheets; both the T_1 and T_2 data show the B-C and E-F loops as the most flexible regions, with the D-E loop highlighted in the T_1 but less so in the T_2 data.

At 700 MHz (Figure 5.9, Table 6.7) Ile35 no longer produces an outlying value and Ile7 is less exaggerated. As for the 950 MHz data, the loop regions are more mobile than the secondary structure elements, with the D-E, E-F, and F-G loops possessing the greatest flexibility in all three datasets. In addition, the NOE data shows the D strand has notable dynamics between Ile46 and Glu50; this is in the region of the β-bulge (Figure 1.4) and suggests inherent flexibility in this region.

Table 5.2 – Average ^{15}N heteronuclear relaxation data for isolated and H-2D^b/FL-bound hβ_2m.

	Field Strength (MHz)	**^{15}N T_1 (ms)**	**^{15}N T_2 (ms)**	**^{1}H-^{15}N NOE**
Isolated hβ_2m	700	875 ± 44	65 ± 12	0.79 ± 0.08
H-2D^b/FL-bound hβ_2m	950	2403 ± 409	21 ± 9	-[a]
	700	1837 ± 116	27 ± 1	0.82 ± 0.12

[a] NOE experimental pulse sequence unavailable on 950 MHz instrument.

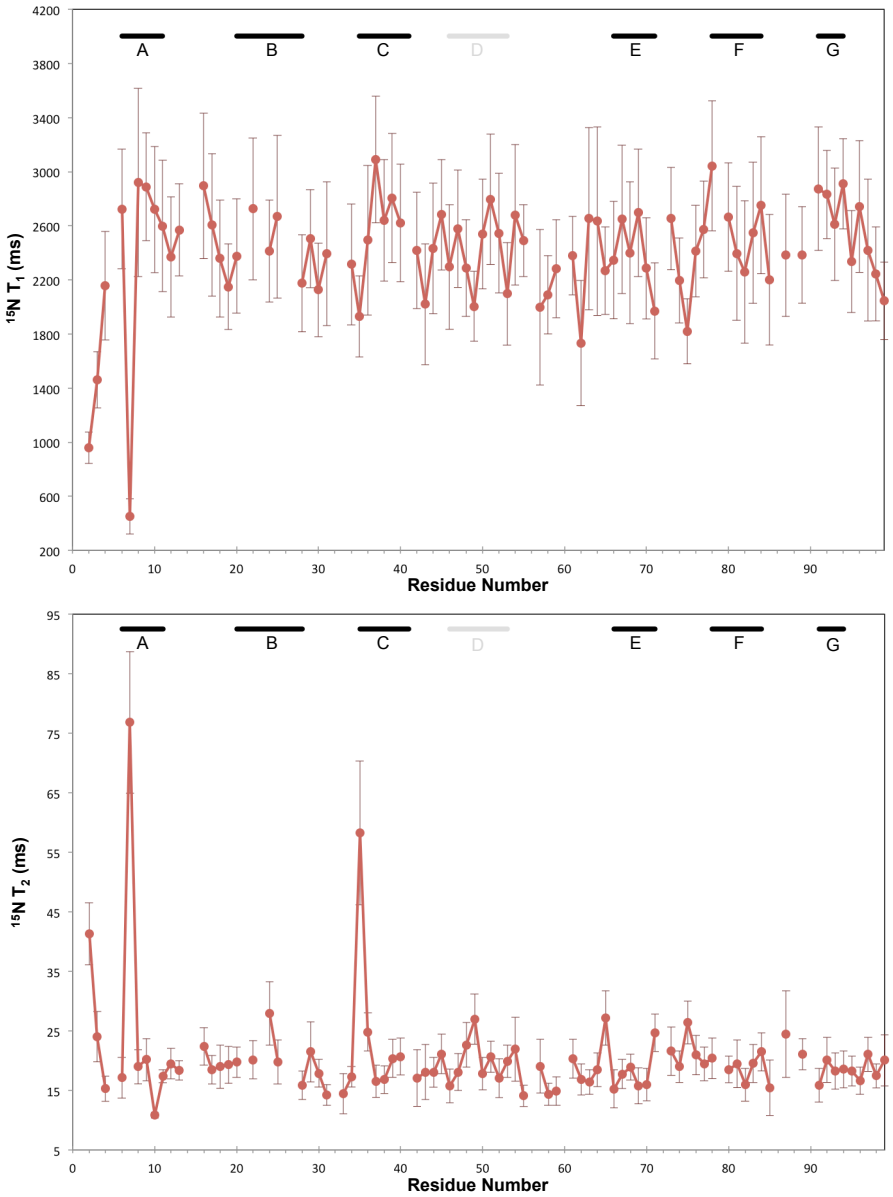


Figure 5.8 – ^{15}N heteronuclear relaxation data for hβ$_2$m bound by H-2D^b/FL at 950 MHz.
^{15}N T_1 (top) and ^{15}N T_2 (bottom) values are shown. Location of β-strands in PDB:1CE6[98] are indicated by solid black lines, labelled A-G.

Both the 700 and 950 MHz data for H-2D^b/FL-bound hβ$_2$m can be represented on a Lipari-Szabo graph and compared to predictions for a rigid protein (Figure 5.10). Similar to isolated hβ$_2$m there is a clear disagreement between predicted and experimental data with the majority of residues having apparent T_2 values shorter than those predicted by overall protein rotation and falling beneath the S^2=1 line. This may suggest the presence of a slow dynamic process caused by transient self-association or generic pervasive conformational flexibility. There are also several outlier residues dominated by fast motion for both the 700 and 950 MHz data; these belong to the N-terminus (already identified as notably flexible) as well as the D-E (Leu64) and F-G (Leu87) loop regions. Glu36 on the C strand, and Ile92 on the G strand also have fast dynamics.

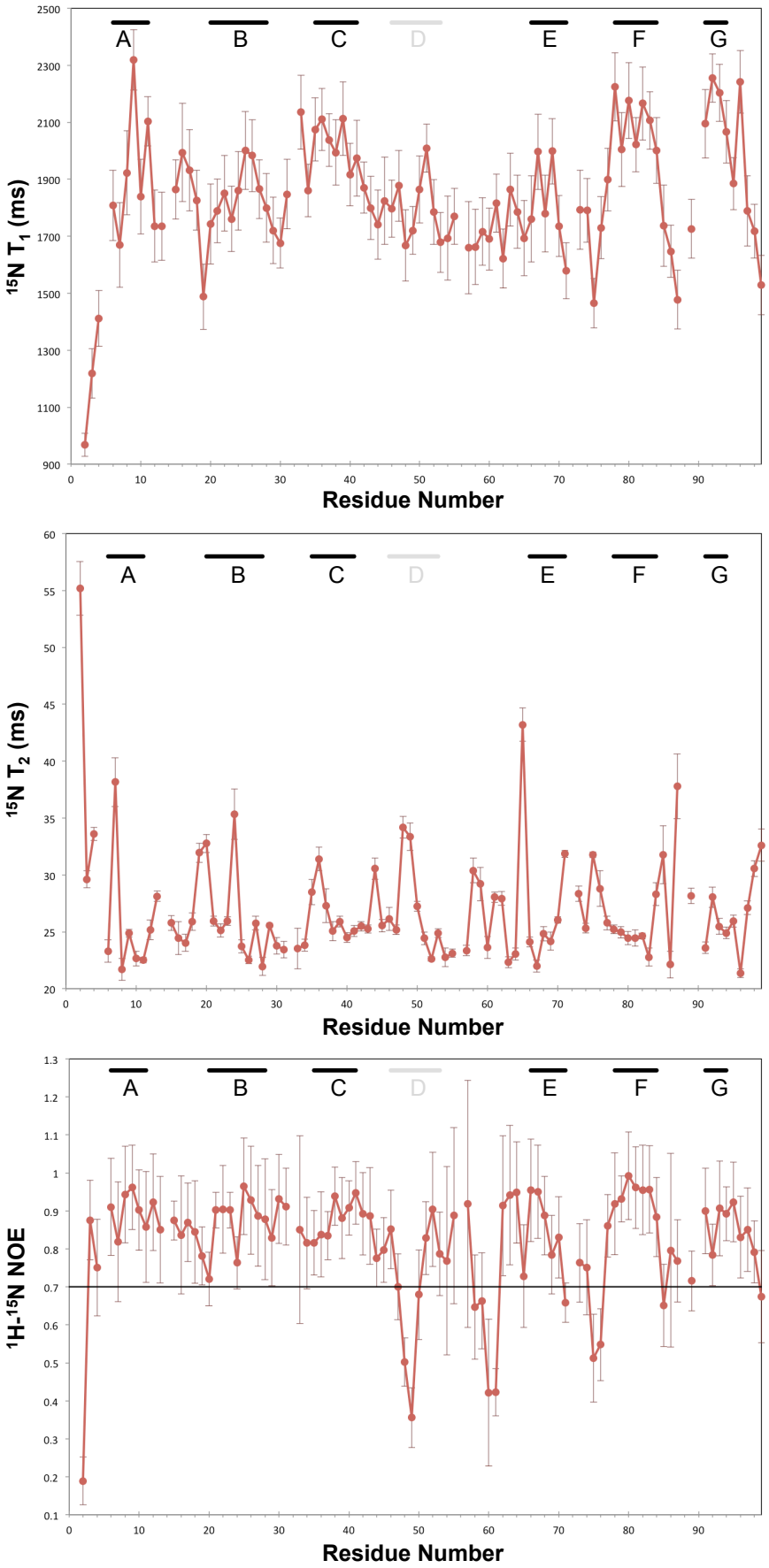


Figure 5.9 – ^{15}N heteronuclear relaxation data for hβ$_2$m bound by H-2D^b/FL at 700 MHz.

^{15}N T_1 (top), ^{15}N T_2 (middle) and ^{1}H-^{15}N NOE (bottom) values are shown. Location of β-strands in PDB:1CE6[98] are indicated by solid black lines, labelled A-G. For NOE data, residues whose values fall below the 0.7 cut-off line are considered notably flexible.

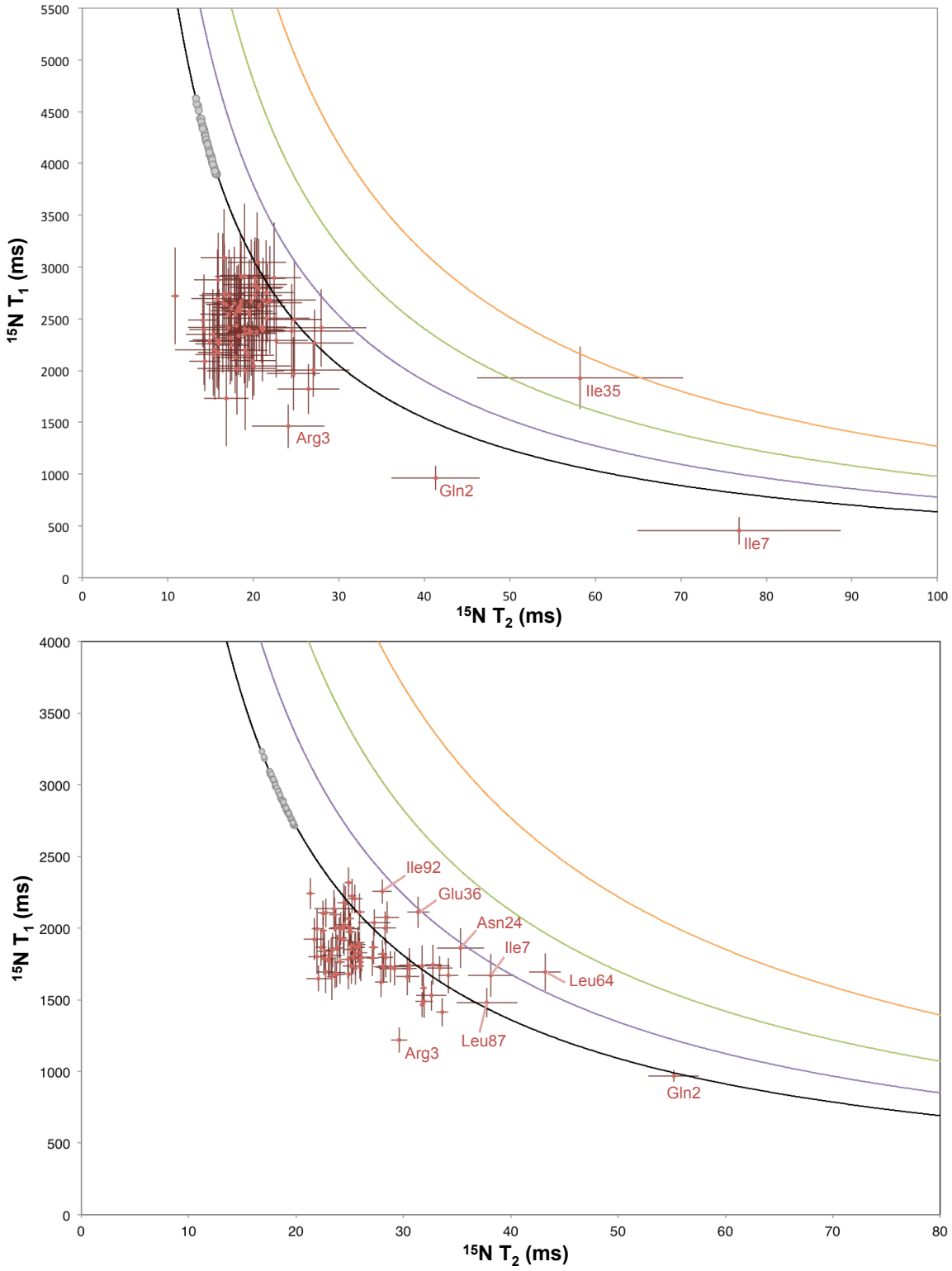


Figure 5.10 – Lipari-Szabo graphs for hβ$_2$m bound by H-2D^b/FL.
Data for 950 MHz (top) and 700 MHz (bottom) is shown. The S^2=1 line (black) denotes a rigid protein; decreasing S^2 values (0.9 purple, 0.8 red, 0.7 orange) permit increasing amounts of fast internal motion. Experimental values below S^2=1 might be explained by slow conformational exchange. Notable outlier residues are labelled. HYDRONMR predictions are depicted in grey.

5.2.2.6 Hydrogen deuterium exchange for MHC class I-bound hβ$_2$m

When the rate of deuterium uptake for H-2D^b/FL-bound β$_2$m is analysed (Figure 5.11A) it is consistent with the heteronuclear relaxation data. Areas of secondary structure are slower to exchange and hence are less flexible than loop regions, with the exception of fast uptake in the E-F strand. As for isolated hβ$_2$m, there is slow uptake in peptides corresponding to the B and C strands, and to the B-C loop. However, here there is also slow exchange in the D-E loop. Indeed, when the rate of hydrogen-deuterium exchange is compared between isolated and H-2D^b/FL-bound

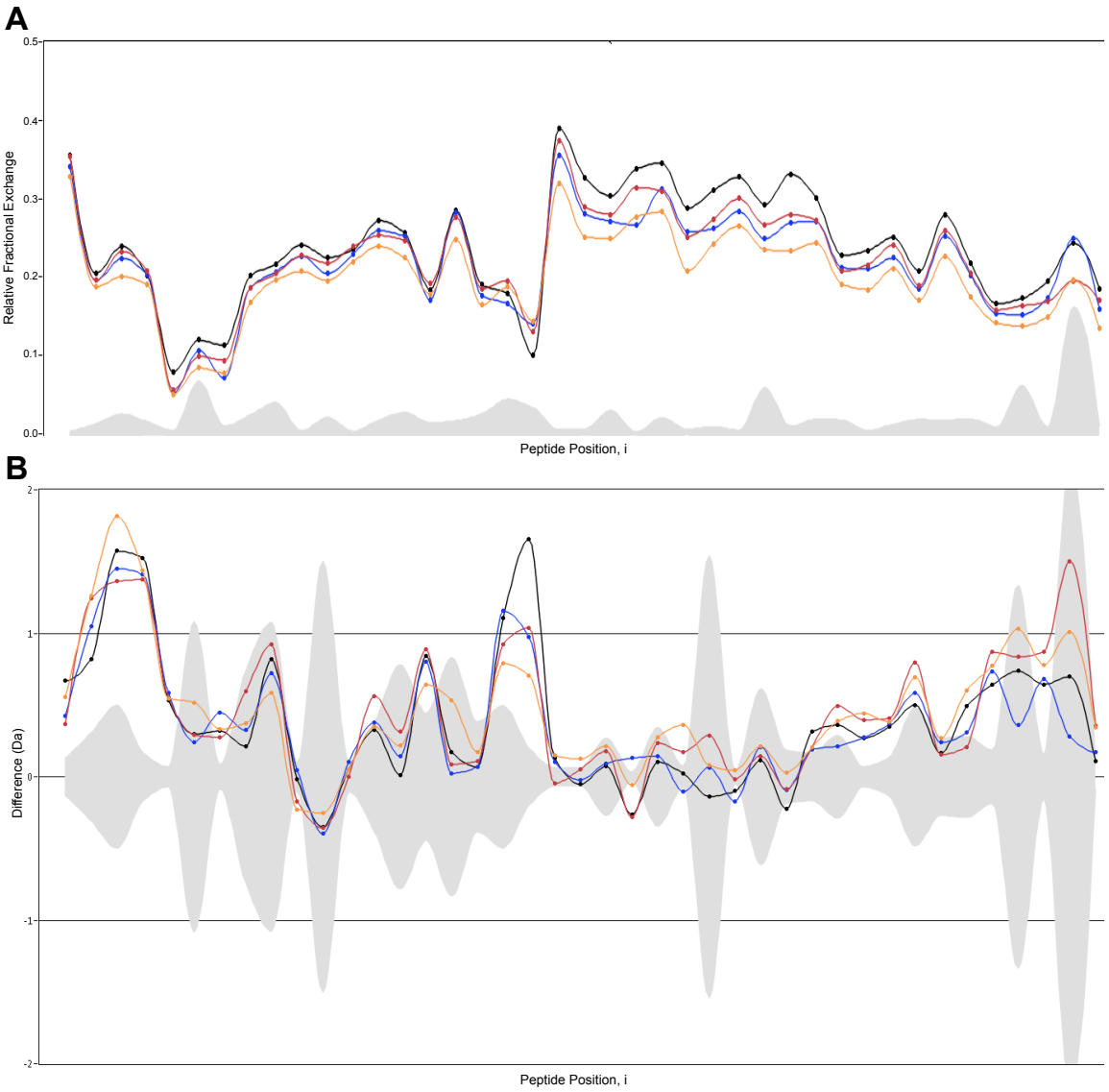


Figure 5.11 – Hydrogen-deuterium exchange for hβ$_2$m bound by H-2D^b/FL.
(A) Deuterium uptake for time points 0.5 minutes (yellow), 1.0 minutes (blue), 2.0 minutes (red), and 5.0 minutes (black). Peptide position 'i' is determined by arranging the peptide midpoints in linear numerical order. A slow uptake is indicated by decreased values. Each point represents a peptide corresponding with (A). The shaded grey region is the estimated error in exchange rate. A slow uptake is indicated by decreased values. The peptide coverage is the same as for isolated β$_2$m. **(B)** Difference between deuterium uptake for isolated and H-2D^b/FL-bound hβ$_2$m for the four time points of (A), with estimated error in grey. Positive values indicate a decrease in deuterium uptake time for H-2D^b/FL-bound hβ$_2$m compared to isolated hβ$_2$m.

hβ_2m, there are three regions of difference in which uptake is faster for isolated β_2m (Figure 5.11B). The N-terminus has become slower to exchange, as have the D strand, D-E loop, and C-terminal regions. Overall, the rate of exchange has decreased, which suggests the presence of the H-2D^b α chain has had a stabilising effect. The key observation is that there has been a large decrease D-E loop region, indicating that this loop has been protected from exchange by hydrogen bonding with the α chain platform above. Although exchange on the D strand has slowed, elevated uptake rates are observed in this region in comparison to the remainder of bound hβ_2m, such that dynamics are still likely to be present.

5.2.2.7 Hydrogen exchange for MHC class I-bound hβ_2m

Although a CLEANEX experiment was carried out, the H to D exchange was too slow (the experimental maximum time was 0.25 milliseconds) to be captured. Conversely, D to H exchange occurred too rapidly to be captured through collection of TROSY spectra (the experimental minimum time was 150 minutes; buffer exchange of solvent for D_2O required approximately 60 minutes, whilst collection of a high-quality TROSY spectra required 90 minutes).

5.3 Dynamics of the H-2D^b α-chain

Following on from hβ$_2$m, the second sub-topic dealing with the dynamics of H-2D^b/FL is the dynamics of the H-2D^b α-chain.

5.3.1 Materials and methods

The H-2D^b α-chain has been characterised using lineshape analysis, heteronuclear relaxation data, and hydrogen exchange. Hydrogen-deuterium exchange (using the method applied for isolated and H-2D^b/FL-bound hβ$_2$m) is not applicable to a protein of this length because the margin of error (five residues for β$_2$m) would be too large. With regards to hydrogen exchange data, attempts to apply the CLEANEX-PM approach to H-2D^b/FL-bound hβ$_2$m were unsuccessful.

5.3.1.1 Qualitative lineshape analysis

Analysis of H-2D^b α-chain is based on the TROSY spectrum of the backbone resonance assignment (Figure 3.4). Similar to hβ$_2$m, the assignment of additional peaks for a residue relied on the presence of alpha carbons at extremely similar chemical shifts as for the first peak (the sequential linking of residues using trHNCACB strips must still apply), and on the presence of an extremely similar chemical shift in the ^{1}H-^{1}H NOESY spectrum. To rule out any contribution of protein unfolding to the observation of doubled peaks, control fast-HSQC spectra were collected before and after recording the main dataset.

5.3.1.2 Collection of heteronuclear relaxation data

Relaxation data for 600 μM H-2D^b/FL composed of ^{2}H^{13}C^{15}N H-2D^b α-chain, ^{2}H^{15}N hβ$_2$m and unlabelled FAPGNYPAL peptide was collected at 950 MHz on an Oxford Instruments spectrometer using a HCN PFG room temperature probe at the University of Oxford. To obtain T_1 values, seven spectra were collected with delays of 0.02013, 0.10063, 0.30190, 0.60380, 1.20761, 2.41522, and 4.83043 seconds. To measure T_2 values another seven spectra were collected, with delays of 4.976, 9.952, 14.928, 19.904, 24.880, 29.856, and 74.640 milliseconds. NOE data was not collected.

Relaxation data for 600 μM H-2D^b/FL composed of ^{2}H^{15}N H-2D^b α-chain, ^{2}H^{13}C^{15}N hβ$_2$m and unlabelled FAPGNYPAL peptide was also collected at 700 MHz (NIMR, London). To obtain T_1 values, seven two-dimensional spectra were collected, with delays of 0.01, 0.04, 0.10, 0.30, 0.60, 1.20, and 1.40 seconds. Seven spectra were also

collected for T_2 measurement, with delays of 4, 8, 12, 20, 28, 40, and 60 milliseconds. ^{1}H-^{15}N NOE data was collected.

Lipari-Szabo graphs were created using Dr Werner's Linux script, based on the principles set out by Lipari and Szabo[222,223]. HYDRONMR[221] was used to predict ^{15}N T_1 and T_2 values for the H-2D^{b} α-chain of H-2D^{b}/FL at 950 and 700 MHz at 298K; beads of radius 3.2 Å were placed at Cα positions of the PDB 1CE6[98] coordinates.

5.3.2 Results

5.3.2.1 Qualitative lineshape analysis

The before and after control HSQC spectra showed no alterations and thus no protein unfolding contributed to observation of chemical exchange. The average proton linewidth for the H-2D^b α-chain was 24.2Hz (Figure 5.12), with broad linewidths defined as those greater than 35Hz. A large number of doubled peaks were also observed (Table 5.3); a tripled peak was located in the Gly area of the TROSY spectrum but could not be confidently assigned.

Many of the unassigned regions are bordered by dynamic resides and it is clear that the α_3 domain is relatively rigid in comparison to the PBD because it has few residues with a broad linewidth or doubled peak (Figure 5.12). Dynamic residues in the α_3 domain are located in loop regions with the exception of Thr214 and Val261, present on the C and F strands respectively. These residues face away from both the PBD and hβ_2m (Figure 5.13). The α_2-α_3 linker region appears mobile as would be expected for an unstructured region tethering two sections of unrelated secondary structure.

Within the PBD, mobile residues were not evenly distributed, but also were not clustered in any particular binding pocket (with only Glu9 in the B and C pockets identified by this analysis). Several residues above the D-E loop of hβ_2m are line broadened; this may represent true conformation mobility or be due to the ring current induced by the motion of β_2m Trp60. The long β-strand (Leu109 to Tyr118) preceding the α_2 helix appears particularly mobile, containing four plastic residues. Indeed, the first section of the $\alpha_{2\text{-}2}$ helix (before the Gly162 kink) that sits above this strand also appears mobile. The α_2 helix is more affected than the α_1 helix, as has been suggested by various *in vivo* and *in vitro* experiments (Section 1.4.3). Interestingly, although the $\alpha_{2\text{-}1}$ helix is thought to flex during peptide binding, in the peptide-bound state the average linewidths indicate it does not appear to be capable of significant motion.

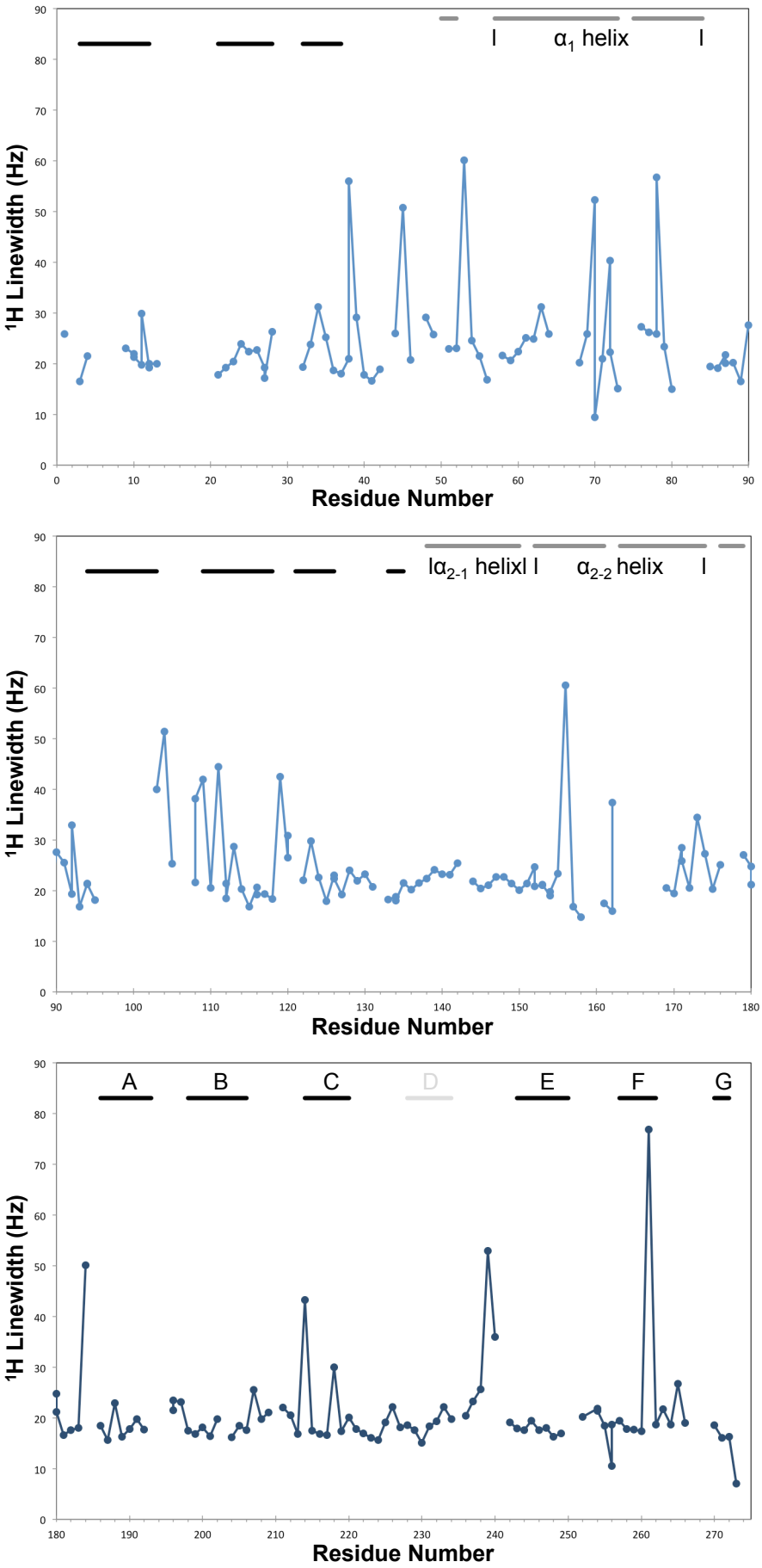


Figure 5.12 – ^{1}H linewidths for the H-2D^b α-chain of H-2D^b/FL.
Location of β-strands in PDB:1CE6[98] is indicated by solid black lines, those comprising the immunoglobulin fold labelled A-G with the D strand of α_3 in light grey. α-helices are indicated by a solid dark grey line, labelled with helix number.

Table 5.3 – Location of broad linewidths and doubled peaks in H-2D^b α-chain of H-2D^b/FL.

Domain	Broad Linewidth	Doubled Peak
α_1	Ser38, Tyr45, Glu53, Gln70, Gln72, Leu78	Glu9, Thr10, Ala11, Val12, Tyr27, Ser38, Gln70, Gln72, Leu78, Gln87
α_2	Ser92, Leu103, Gly104, Arg108, Leu109, Arg111, Glu119, Tyr156, Gly162	Ser92, Thr94, Arg108, Gly112, Phe116, Gly120, Leu126, Thr 134, Ala152, Ala153, Glu154, Gly162, Tyr171, Leu180
α_3	Ser184, Thr214, Gly239, Val261	Lys196, Asn256

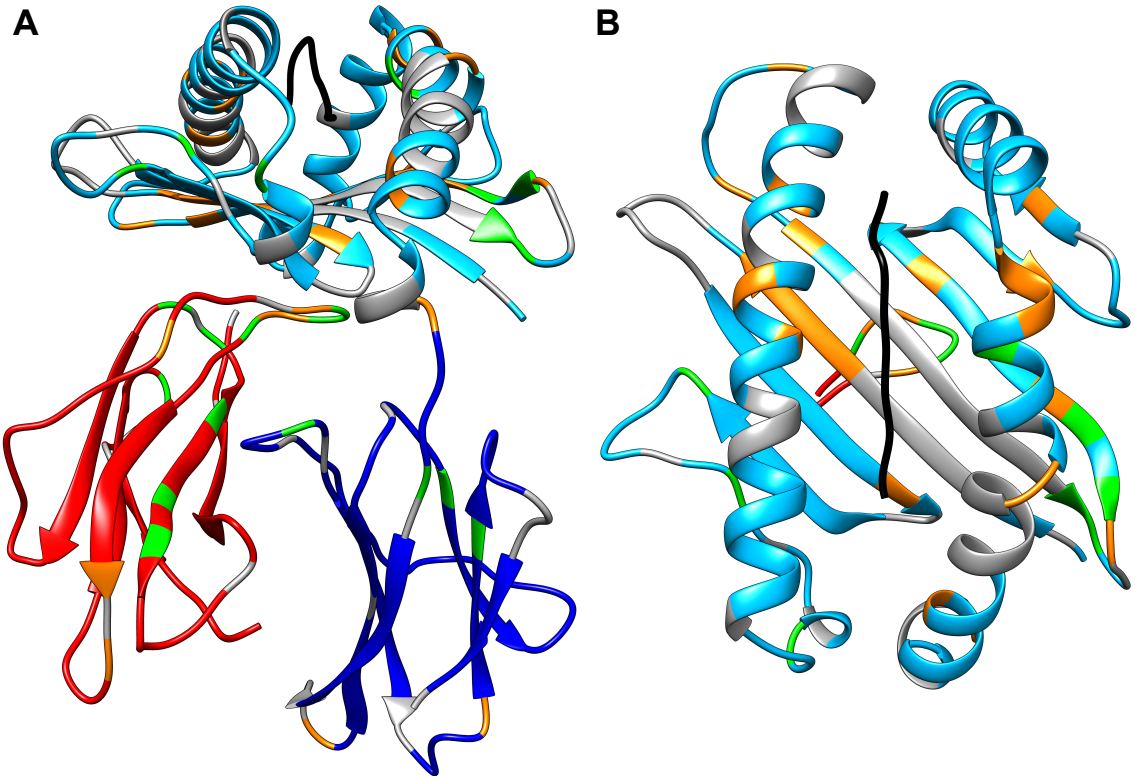


Figure 5.13 – Location of broad linewidths and doubled peaks in H-2D^b/FL.
(A) Side view of H-2D^b/FL (PDB: 1CE6[98]). Unassigned residues on hβ_2m (red), H-2D^b PBD (light blue), and H-2D^b α_3 domain (dark blue) are coloured grey, residues with broad linewidths in green, and residues with doubled peaks in orange. **(B)** The PBD of H-2D^b α-chain (PDB: 1CE6[98]).

5.3.2.2 Heteronuclear relaxation of H-2D^b α-chain

^{15}N heteronuclear relaxation data was collected at two field strengths for the H-2D^b α-chain with average values given in Table 5.4. Subtle variations are observed between domains within the H-2D^b α-chain, with difference average values between the PBD and the α_3 domains.

Similar to the H-2D^b/FL-bound hβ_2m dataset the 950 MHz spectra had poor signal to noise, which caused a significant uncertainty in the measurements (Figures 5.14 and 5.15, Table 6.8). The general observations that can be made are that the protein termini are not notably flexible and the secondary structure is generally stable with dynamics concentrated in the loop regions. Both the T_1 and T_2 data show that immediately prior to the α_1 helix there is prominent flexibility, as well as around Val76 and Asn80 in the α_1 helix itself. The $\alpha_{2\text{-}1}$ helix is stable, whilst the $\alpha_{2\text{-}2}$ contains flexible areas at the Gly162 kink and around Tyr171 (T_1 Figure 5.14) and Leu172 (T_2 Figure 5.15); these are areas already identified through the linewidth analysis. The linker region between the PBD and the α_3 domain is flexible, although not as much as might be expected for such an unstructured region. The D strand of α_3 appears to possess significant mobility, particularly around Thr225.

These general observations match the 700 MHz data (Figures 5.16-5.18), which has a superior noise level that permits more precise observations. The $\alpha_{2\text{-}1}$ helix is stable, but both the α_1 and $\alpha_{2\text{-}2}$ helices contain regions that alter conformation. In the α_1 helix this includes Lys68 (T_1, Figure 5.16), Trp73 (T_1, NOE, Figure 5.18), Val76 (T_2, Figure 5.17), Leu78 (T_2, NOE, identified as a doubled peak in the linewidth analysis), and Asn80 (NOE); none of these contribute to a binding pocket but Trp73 forms part of the hydrophobic ridge. In the $\alpha_{2\text{-}2}$ helix there are dynamics around the kink induced

Table 5.4 - Average ^{15}N heteronuclear relaxation data for H-2D^b α-chain.

Field Strength (MHz)	H-2D^b α-chain domain	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
950	Full	2316 ± 399	22 ± 5	-[a]
	PBD	2345 ± 427	21 ± 5	-[a]
	α_3	2271 ± 355	22 ± 5	-[a]
700	Full	1924 ± 251	29 ± 1	0.79 ± 0.16
	PBD	1915 ± 258	28 ± 1	0.80 ± 0.19
	α_3	1941 ± 241	29 ± 1	0.77 ± 0.11

[a] NOE experimental pulse sequence unavailable on 950 MHz instrument.

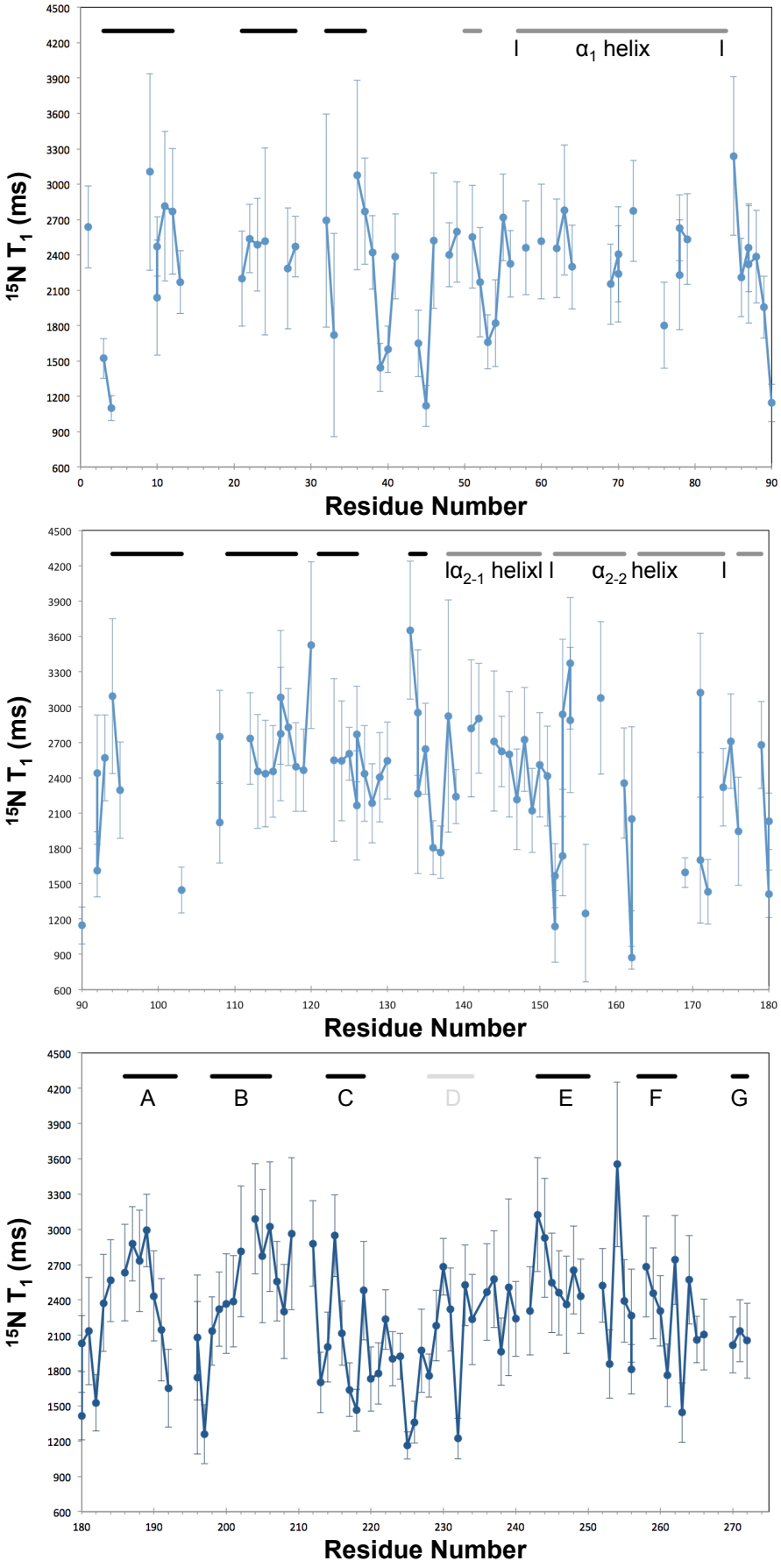


Figure 5.14 – ^{15}N T_1 heteronuclear relaxation data for H-2D^b α-chain at 950 MHz.
Location of β-strands in PDB:1CE6[98] is indicated by solid black lines, those comprising the immunoglobulin fold labelled A-G with the D strand of α_3 in light grey. α-helices are indicated by a solid dark grey line, labelled with helix number.

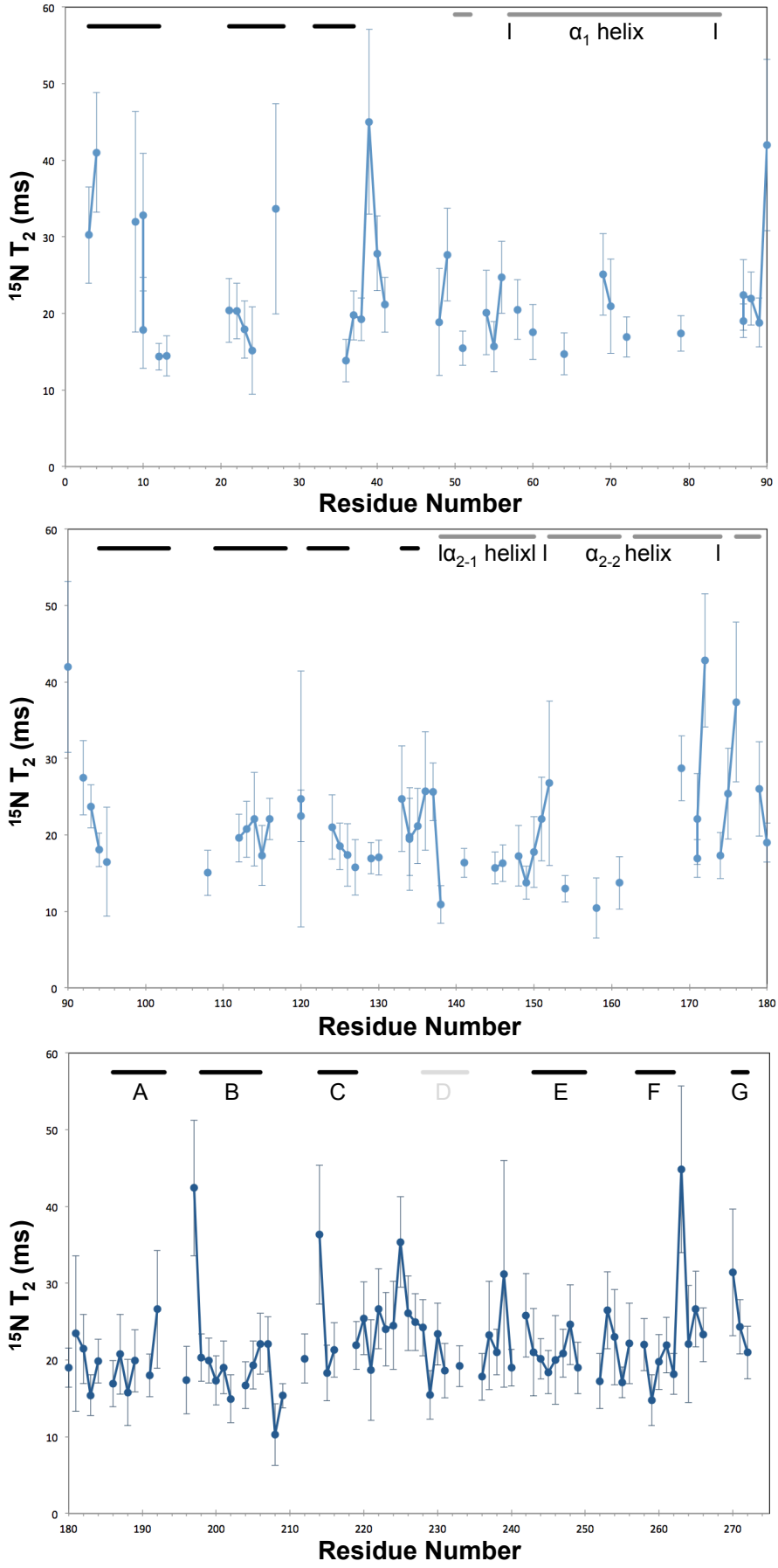


Figure 5.15 – ^{15}N T_2 heteronuclear relaxation data for H-2D^b α-chain at 950 MHz.
Location of β-strands in PDB:1CE6[98] is indicated by solid black lines, those comprising the immunoglobulin fold labelled A-G with the D strand of α_3 in light grey. α-helices are indicated by a solid dark grey line, labelled with helix number.

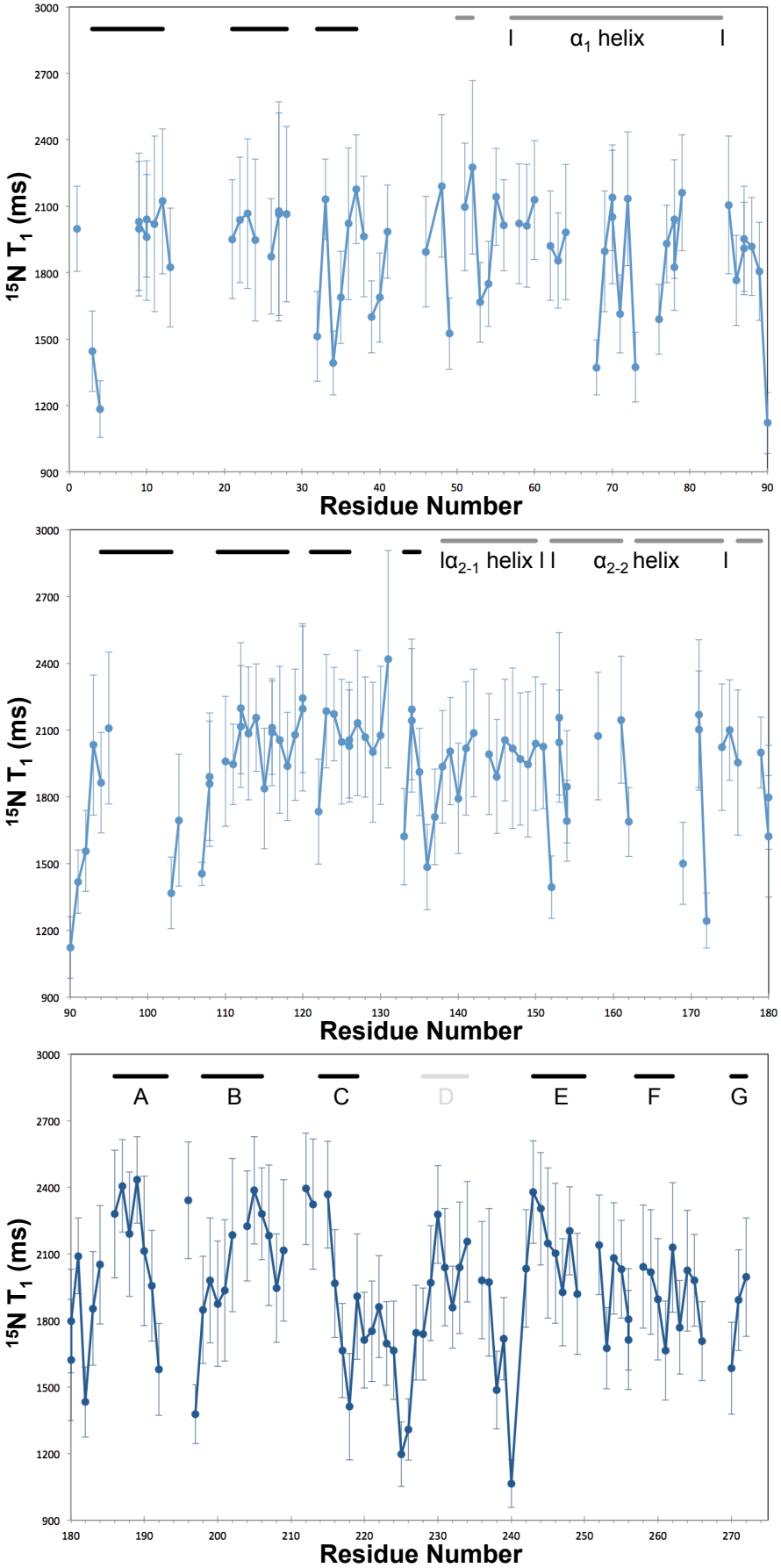


Figure 5.16 – ^{15}N T_1 heteronuclear relaxation data for H-2D^b α-chain at 700 MHz.
Location of β-strands in PDB:1CE6[98] is indicated by solid black lines, those comprising the immunoglobulin fold labelled A-G with the D strand of α_3 in light grey. α-helices are indicated by a solid dark grey line, labelled with helix number.

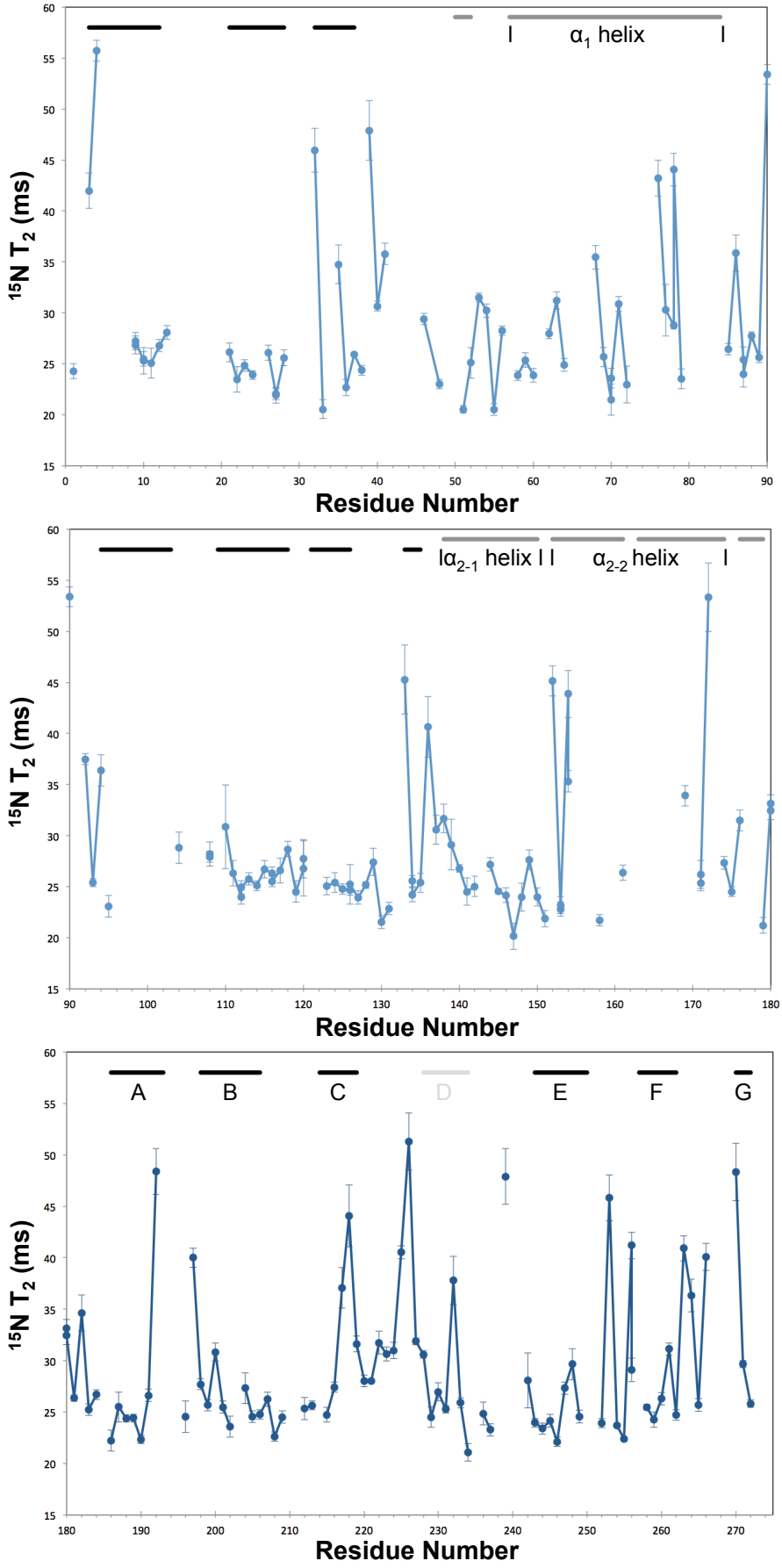


Figure 5.17 – ^{15}N T_2 heteronuclear relaxation data for H-2D^b α-chain at 700 MHz.
Location of β-strands in PDB:1CE6[98] is indicated by solid black lines, those comprising the immunoglobulin fold labelled A-G with the D strand of α_3 in light grey. α-helices are indicated by a solid dark grey line, labelled with helix number.

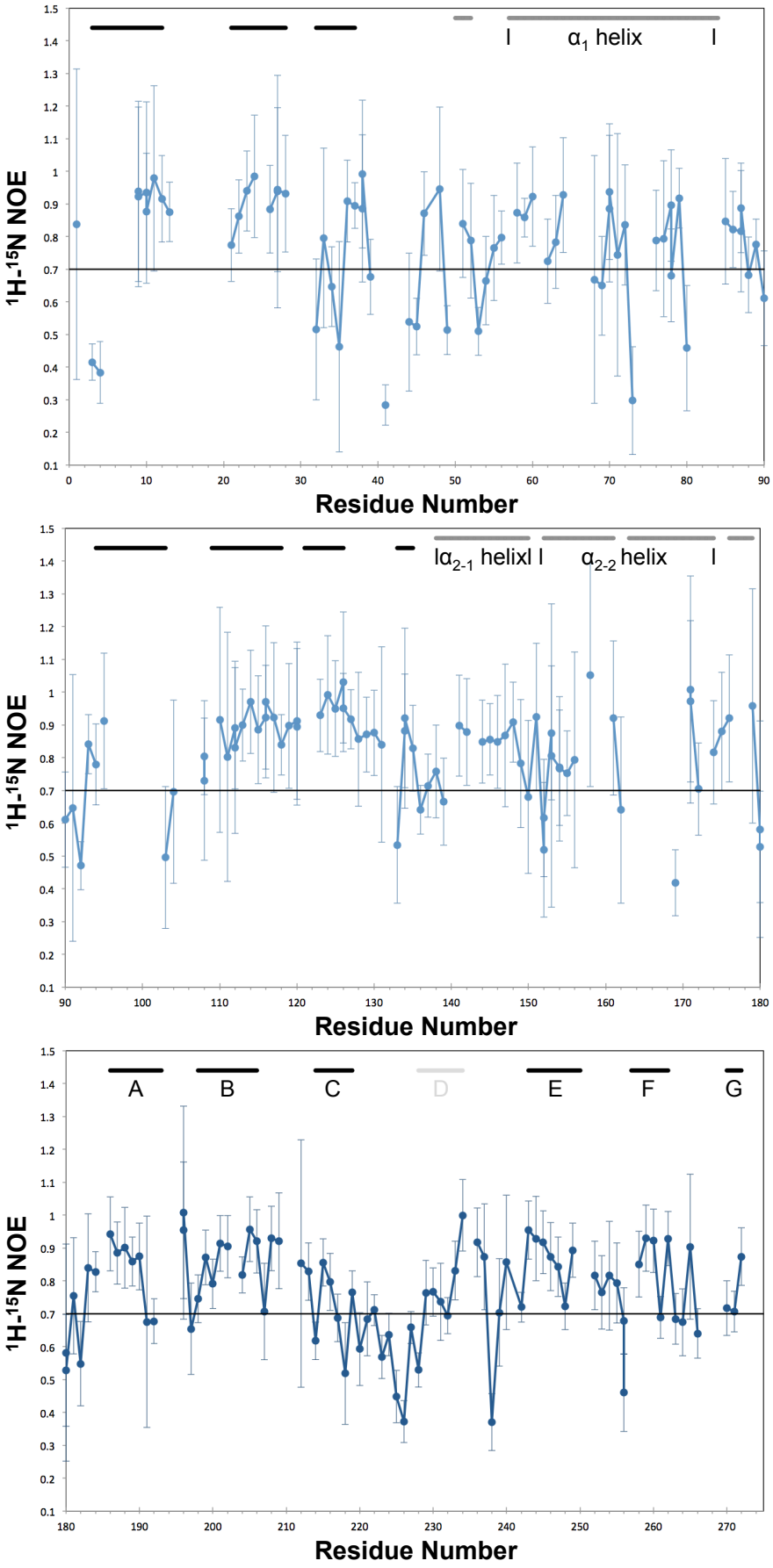


Figure 5.18 – $^{1}H^{15}N$ heteronuclear NOE data for H-2D^{b} α-chain at 700 MHz.
Location of β-strands in PDB:1CE6[98] is indicated by solid black lines, those comprising the immunoglobulin fold labelled A-G with the D strand of α_3 in light grey. α-helices are indicated by a solid dark grey line, labelled with helix number. Residues whose values fall below the 0.7 cut-off line are considered notably flexible.

by Gly162, as well as for Leu172 (Tyr171 is a doubled peak). Of the five residues contributing to the N-terminal closure of the PBG three are unassigned (Tyr7, Tyr159, and Trp167) which may suggest they are dynamic and have a broadened linewidth, whilst Tyr59 is stable throughout all analyses. The identification of the regions around the fifth residue, Tyr171, as being dynamic implies the N-terminal end of peptide may not be as securely bound as the C-terminus (Tyr84 is unassigned but both Tyr123 and Thr143 are stable in both the linewidth and relaxation analysis). This correlates with research into the formation of MHC class I with shortened versions of the FAPGNYPAL peptide[79] which found that the PC anchor is vital for formation of a mature complex. There is significant motion within the D strand of the α_3 domain, particularly around Thr225 and Gln226, suggesting that when TPN binds to the 222-229 region it might stabilise one conformer. Indeed, MD simulations by Dr Bailey[164] show that preventing motion in this region results in a widening of the PBG to enable peptide optimisation, with transmission via an unknown mechanism. The main TPN binding site of residues 128-136 also shows dynamics in the heteronuclear relaxation data.

Both the 700 and 950 MHz data for H-2D^b α-chain can be represented on a Lipari-Szabo graph and compared to predictions for a rigid protein (Figure 5.19). This reveals that the majority of residues fall beneath the $S^2=1$ line and are likely affected by a pervasive slow dynamic process such as conformational exchange. Many of the remaining residues lie extremely close to this line and therefore do not have much internal motion. The difference in the spread of data of the two datasets is likely due to the considerable uncertainty bounds for the 950 MHz data, since there are well-dispersed peaks and a lack of differences between before and after control HSQC spectra. This absence of aggregation also means that the apparent disagreement between experimental data and HYDRONMR-predictions is due to motion; either the molecule is rotating quickly or there is a slow tumbling time and a fast internal motion. The significant decrease in experimental ^{15}N T_1 values suggests the latter.

Outlier residues dominated by fast motion are labelled; these are largely located in loop regions including the pre-α_1 helix loop and the α_3 domain F-G loop, and several have previously been discussed in light of the NOE and raw T_1 and T_2 data. Residues of particular interest in the PBD are Glu32, on the floor of the PBG at the N-terminus of peptide and which contributes to the notion that there are dynamics in this region of the PBD despite the presence of a high affinity peptide, and Glu154, which is located on the $\alpha_{2\text{-}2}$ helix and which has previously been identified as mobile through

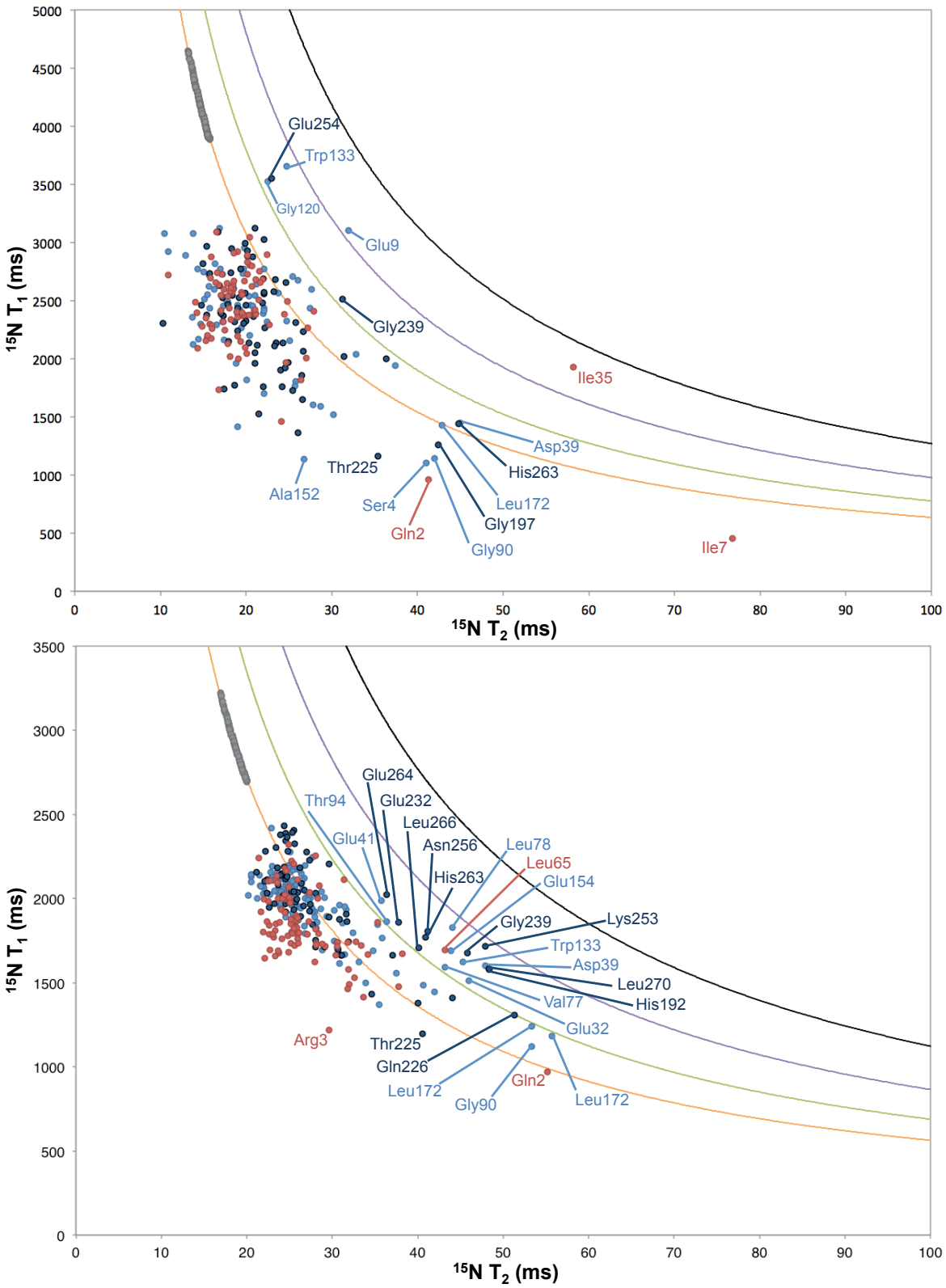


Figure 5.19 – Lipari-Szabo graphs for the H-2D^b α-chain.
Data for 950 MHz (top) and 700 MHz (bottom) is shown. For clarity, error bars are omitted. Data pertaining to the PBD of the H-2D^b α-chain is shown in light blue, the α_3 domain in dark blue, with hβ_2m shown in red for reference. The S^2=1 line (orange) denotes a rigid protein; decreasing S^2 values (0.9 green, 0.8 purple, 0.7 black) permit increasing amounts of fast internal motion. Experimental values below S^2=1 might be explained by slow conformational exchange. Notable outlier residues are labelled. HYDRONMR predictions are depicted in grey.

qualitative lineshape analysis. Apart from residues located towards the C-terminus of the α chain and those of the F-G loop, outliers in the α_3 domain belong to the D strand TPN binding region.

Field Dependence of ^{15}N T_1 and T_2 data

In order to fully address why so many data points are outside the $S^2=1$ line on the Lipari-Szabo graph the field dependence of the T_1 and T_2 must be analysed. There are two possible explanations; firstly that T_2 relaxation is too short when compared to the Lipari-Szabo rigid model due to pervasive conformational exchange, or secondly due to internal motion in addition to the overall tumbling. The ^{15}N T_2 parameter is strongly attenuated by exchange, such that if exchange were present the T_2 values at 950 MHz would be significantly smaller than at 700 MHz. However, this is not the case; when the experimental and HYDRONMR average ^{15}N T_2 values are compared, the percentage change between the two field strengths are similar (Table 5.5; -24.5% and -21.6% respectively for α-chain and -24.0% and -21.6% for β_2m). The HYDRONMR calculation does not include any allowance for exchange factors and hence the field-dependent differences in T_2 are almost entirely explained by overall tumbling. This suggests that H-2D^b/FL is not affected by exchange.

The presence of data outside the $S^2=1$ Lipari-Szabo limit must still be explained so a second scenario must be considered. Examining the field dependence of the ^{15}N T_1 relaxation data, the T_1 gets longer at a higher field, but the experimental values do not

Table 5.5 – Field dependence of heteronuclear relaxation data.

	Experimental		**HYDRONMR**	
	Average ^{15}N T_1 (ms)	**Average ^{15}N T_2 (ms)**	**Average ^{15}N T_1 (ms)**	**Average ^{15}N T_2 (ms)**
H-2D^b α-chain, 700 MHz	1924 ± 251	28.7 ± 1.0	2856	19.0
H-2D^b α-chain, 950 MHz	2316 ± 399	21.7 ± 4.7	4121	14.9
Percentage change	**20.4%**	**-24.5%**	**44.3%**	**-21.6%**
H-2D^b/FL-bound β_2m, 700 MHz	1837 ± 115	27.0 ± 0.8	2860	19.0
H-2D^b/FL-bound β_2m, 950 MHz	2403 ± 422	20.5 ± 3.4	4126	14.9
Percentage change	**30.8%**	**-24.0%**	**44.3%**	**-21.6%**

increase as much as is predicted by HYDRONMR (Table 5.5; 20.4% to 44.3% respectively for α-chain and 30.8% to 44.3% for β_2m). This is a significant difference, suggesting there must be additional factors to consider. Therefore, whilst HYDRONMR predicts a tumbling time of 33.8 ns and an increasein T_1 values of 44% between 700 MHz and 950 MHz, this is not observed and can possibly be explained by the presence of additional internal motion affecting the majority of residues on a low nanosecond timescale. For example, this could be a domain-domain rearrangement.

If heteronuclear relaxation data had been measured only at a single field the logical conclusion would have been that the data was affected by exchange on the microsecond to millisecond timescale. However, analysing data from two fields reveals that this is not the correct interpretation and instead the motion is fast (ns). The explanation of domain-domain motion causing this discrepancy also fits with the Lipari-Szabo graph S^2=1 anomaly; the calculation of Lipari-Szabo T_1 and T_2 curves assume a single overall tumbling time and only model internal dynamics by different S^2 values on a localised basis[222,223], without making provision for large-scale internal motions such as domain-domain shifting. Including pervasive internal motion on a nanosecond timescale in the Lipari-Szabo model would result in curves with a slower rate of T_1 at low T_2 values. The shift of the curve to the left would likely result in the experimental relaxation data falling within the S^2=1 limit.

5.3.2.3 Hydrogen exchange for H-2D^b α-chain

As for the analysis of data pertaining to the hβ_2m domain, the CLEANEX experiment showed that H to D exchange was too slow (the experimental maximum time was 0.25 milliseconds) to be captured whilst D to H exchange occurred too rapidly to be observed through collection of TROSY spectra (the experimental minimum time was 150 minutes; buffer exchange of solvent for D_2O required approximately 60 minutes, whilst collection of a high-quality TROSY spectra required 90 minutes).

5.4 Dynamics of H-2D^b/FL

In addition to considering the domains of H-2D^b/FL separately, it is important to analyse the dynamics of the entire H-2D^b/FL molecule.

5.4.1 Materials and methods

The Lipari-Szabo model-free approach[222,223] was used to characterise the dynamics of H-2D^b/FL. The overall rotational diffusion was predicted using HYDRONMR[221], whilst ModelFree software[225,226] was used to characterise internal fast motions.

5.4.1.1 HYDRONMR

HYDRONMR[221] has multiple potential applications; in addition to the previous prediction of ^{15}N T_1 and T_2 values for a rigid protein, it can be used to predict overall rotational diffusion (τ_M). It also provides data on anisotropy via predictions of D_Z, D_Y, and D_X.

HYDRONMR was used to predict both τ_M and diffusion anisotropy for H-2D^b/FL at 950 and 700 MHz at 298 K; beads of radius 3.2 Å were placed at Cα positions using the coordinates of PDB 1CE6[98].

5.4.1.2 ModelFree

ModelFree[226] was run as described in Section 5.1.2.2 with 500 Monte Carlo simulations, using relaxation data collected at both 950 and 700 MHz as described in Sections 5.2 and 5.3. For inclusion in the ModelFree fit, residues were required to possess a minimum of three measured relaxation observables (T_1, T_2, and ^{1}H-^{15}N NOE) at 700 MHz, with additional T_1 and/or T_2 values at 950 MHz included to increase the reliability of the fit.

Model selection was made based on the F-statistic[225] (Figure 5.20A) although the inclusion of additional data from a second field strength necessitated some modifications to this scheme (Figure 5.20B). Model 5 involves fitting three model-free parameters such that a dataset comprising three data points will always produce a perfect fit; the post-Model 2/Model 3 decision that is required to avoid unnecessarily using Model 5 can be removed when more than three data points are present.

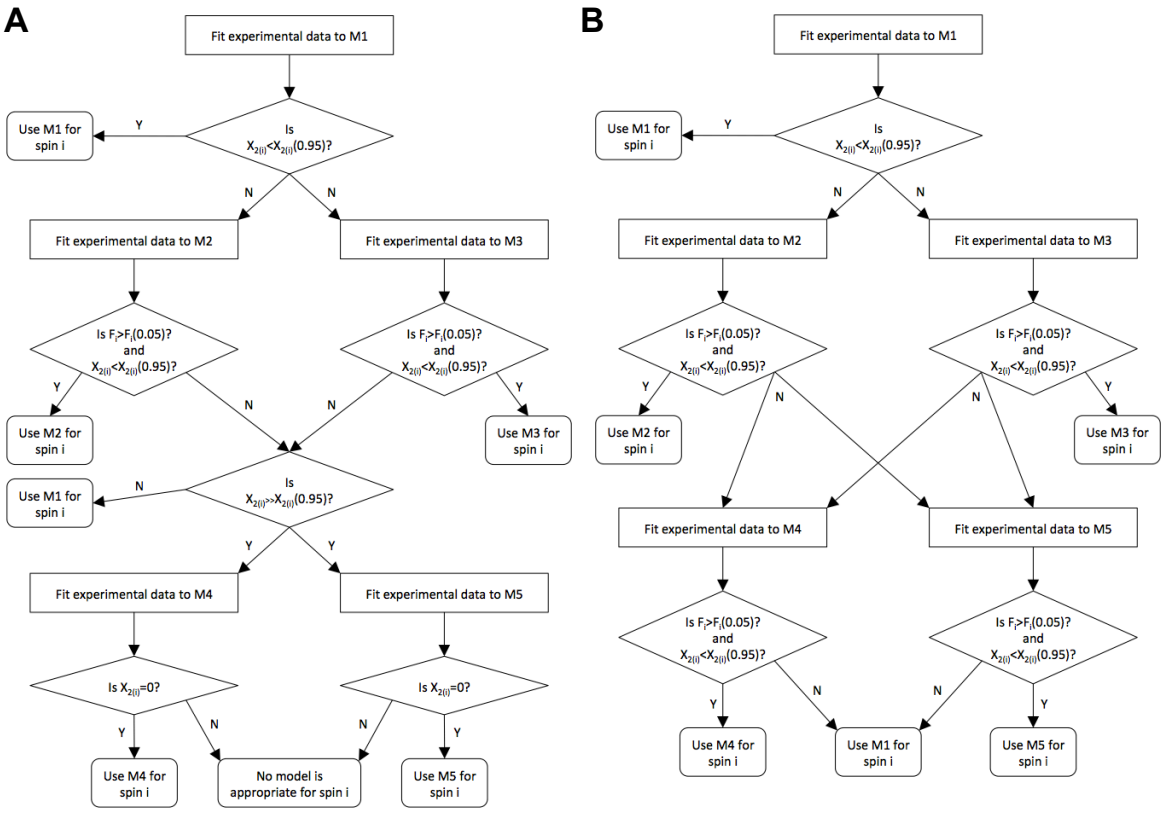


Figure 5.20 – ModelFree model selection scheme.
Models 1-5 are indicated by M1-5, the F-statistical test by F, and the chi-square test by X_2. **(A)** ModelFree model selection scheme for three experimental data points. Scheme taken from Mandel *et al*[225]. **(B)** ModelFree model selection scheme for four or more experimental data points. Scheme adapted from Mandel *et al*[225].

5.4.2 Results

5.4.2.1 Selection of τ_M

When using ModelFree[226] to estimate an initial rotational correlation time (to be optimised towards the end of the ModelFree process) the typical method is to calculate the R_2:R_1 (T_1/T_2) ratio. For molecules known to have significant internal motion, this calculation can be carried out by selecting a subset of residues for which chemical exchange is absent and internal motions are restricted. Based on all previous analyses, a subset of such residues for H-2D^b/FL was chosen and a τ_M estimate of 23.0 ns obtained. This value is significantly lower than expected for a protein of ~45 kDa, and therefore two separate calculations were run; at 700 MHz the preedicted τ_M was 23.9 ns and at 950 MHz the predicted τ_M was 21.5 ns.

Although both these values are smaller than expected for a protein of the size of H-2D^b/FL, there is another problem. The T_2 parameter has field dependence such that at a higher field it is smaller, and thus we expect the R_2:R_1 ratio to yield a larger value at 950 MHz compared to 700 MHz. As previously discussed, aggregation is not a factor in these data and the evidence from analysis of the field dependence suggests a pervasive fast (ns) internal motion. An alternative method of estimating τ_M is therefore required. HYDRONMR[221], based on a hydrodynamics approach, predicted a τ_M of 33.8 ns. This is significantly closer to what is expected of a ~45 kDa protein and provides a reliable base for the initial ModelFree calculations. It should be noted that HYDRONMR is based solely on the X-ray crystallography structure and thus does not account for internal motion. However, this rotational correlation time will be optimised later in the ModelFree method to incorporate protein dynamics.

5.4.2.2 Isotropic analysis

An isotropic ModelFree analysis did not fit to the data; increasing the number of parameters in Models 4 and 5 did not decrease the Chi-squared metric to an acceptable level, and a significant number of residues could not be assigned to a model when the F-statistic was applied.

5.4.2.3 Selection of anisotropic models

Anisotropic diffusion analysis employs an asymmetric model in which several new correlation times are required. D_{ratio} is the ratio $D_{II}/D\perp$ for the diffusion tensor and thus controls the degree of anisotropy (whether the molecule is prolate or oblate and by how much it thus deviates from a sphere). Theta (Θ) and phi (Φ) are the polar

angles for the symmetry axis of the diffusion tensor in the coordinate frame of the supplied molecule.

HYDRONMR[221] analysis provides a guide to each of these factors by providing D_Z, D_Y, and D_X, as well as eigenvectors for each:

$D_Z = 6.068 \times 10^6\ s^{-1}$

$$= \begin{pmatrix} 0.5176 \\ -0.0642 \\ 0.8532 \end{pmatrix}$$

$D_Y = 4.317 \times 10^6\ s^{-1}$

$$= \begin{pmatrix} 0.8237 \\ -0.2325 \\ -0.5171 \end{pmatrix}$$

$D_X = 4.422 \times 10^6\ s^{-1}$

$$= \begin{pmatrix} 0.2316 \\ 0.9705 \\ -0.0675 \end{pmatrix}$$

It can immediately be seen that a ratio of D_Y and D_X is near to 1, and therefore an axially symmetric model rather than a fully asymmetric model can be used.

The first anisotropic factor that can be identified is D_{ratio}:

$$D_{II} = D_Z$$

$$D\perp = (D_Y + D_X)/2$$

$$D_{ratio} = D_{II}/D_\perp = 1.39$$

Projecting a molecular frame based on the eigenvectors and using basic trigonometric rules can be used to obtain theta and phi. This is demonstrated below using the D_Z eigenvector:

$$\cos\Theta = 0.8532 \div 1,\ \Theta = 31°$$

$$\tan\Phi = -0.061 \div 0.5176,\ \Phi = 353°$$

HYDRONMR is a structure-based prediction of relaxation rates, τ_M, and anisotropy, and for H-2D^b/FL previous chapters have shown the presence of dynamic processes that are not present in the X-ray crystallography structure that the HYDRONMR simulation is based upon. ModelFree is able to investigate protein motion, and therefore will produce more reliable measurements of these three anisotropic factors but the HYDRONMR data provides an excellent starting point. A comparison between predicted parameters was therefore made; a ModelFree fitting, including 500 Monte Carlo simulations, was run only on data from residues in the secondary structure of H-2D^b/FL to avoid the dynamics known to be present in loop regions. The results are

shown in Table 5.6; D_{ratio} is very similar to that predicted by HYDRONMR, but theta shows more of a change. Neither method is especially sensitive to phi because D_Y and D_X are similar. Moving from the isotropic model to the axially symmetric anisotropic model run with the HYDRONMR data does not result in a decrease in Chi-square. However, the more accurate ModelFree data results in a statistically significant decrease in Chi-square; these values for the anisotropic factors D_{ratio}, theta, and phi can be accepted for the purpose of model selection prior to final optimisation.

5.4.2.4 Model selection

Models 1 through 5 were assessed for all residues (Tables 6.10 and 6.11), and residues were assigned to a model using the Chi-square factor and an F-test, as per Figure 5.20. As shown in Table 5.7, and in agreement with previous analyses, it is clear that H-2D^b/FL consists of several domains because there are clear differences in assignment of a Model between H-2D^b α-chain and hβ$_2$m, as well as between the PBD and α_3 domains of H-2D^b α-chain.

Of 276 residues, only one (Val28) could not be fitted to a model. However, a further four residues (one dataset of Glu9, one dataset of Tyr171, Glu254, and Leu361) produced τ_e in excess of τ_M when assigned to Model 5 and thus were also excluded from further analysis. Several further residues assigned to Model 5 produced τ_e in excess of 10% of τ_M (Tables 6.10 and 6.11); these results are unlikely and can be expected to reduce when τ_M is permitted to vary.

Table 5.6 – Selection of anisotropic factors for ModelFree analysis.

ModelFree analysis type		**D_{ratio}**	**Theta**	**Phi**	**X_2**	**p-value** [a]
Isotropic		-	-	-	4179	-
Axially symmetric	HYDRONMR	1.39	31	353	4320	- [a]
	ModelFree	1.25	151	209	3499	0.0001

[a] p-value refers to the probability of obtaining the observed decrease in Chi-square by chance; $p<0.01$ is considered statistically significant

[a] p-value not applicable due to increase in Chi-square measure

Table 5.7 – Model selection for H-2D^b/FL domains.

H-2D^b/FL domain	**M1 (%)**	**M2 (%)**	**M3 (%)**	**M4 (%)**	**M5 (%)**
PBD	75.6	0.8	0.0	0.0	23.6
α_3	77.3	0.0	0.0	0.0	22.7
hβ$_2$m	31.3	3.8	0.0	0.0	64.9

During model selection, no residues were assigned to either Model 3 or Model 4 (the only Models with the R_{ex} term). During linewidth analysis some residues were found to produce broadened peaks (suggesting the presence of dynamics) but in this ModelFree analysis these residues do not produce R_{ex} values in Models 3 and 4. The ModelFree analysis relies on the heteronuclear relaxation data, but when examining ^{15}N-T_2 relaxation values (the variable attenuated by R_{ex}) for affected residues it is clear that they only partially relay the known linewidth broadening: residues identified through linewidth analysis are significantly more likely to have immeasurable ^{15}N-T_2 relaxation compared to the remainder of assigned residues (at 950 MHz this is 40% for broadened peaks and 71% for unaffected residues). It should be noted that broadened residues with measurable T_2 data are not significantly broader than the average, likely because T_2 data collection employs a pulse sequence incorporating a CPMG segment of 500 μs that will quench slower motions. Data arising from residues identified as linewidth broadened must therefore be monitored closely when the internal parameters are optimised.

5.4.2.5 Optimisation of parameters

As a result of model selection it was possible to create a single ModelFree input file in which the appropriate model was assigned (Appendix F), along with the relevant S^{2s}, τ_e, and S^{2f} terms. The parameters τ_M, D_{ratio}, theta, and phi were then permitted to simultaneously vary, using wide boundaries to eliminate the possibility of an artefact result (for example τ_M was able to vary between 20 and 40 ns). The result was that τ_M decreased to 25.5 ns, D_{ratio} remained similar at 1.35, Theta changed to 27°, whilst phi altered to -14°. There is a statistically significant decrease in Chi-square, but the result remains outside the 95% confidence limits and thus is not an accurate representation of the data. From the D_{ratio} it is clear that the anisotropic nature of H-2D^b/FL was predicted correctly, but the decrease in τ_M means the standard ModelFree[226] approach is not suitable to accurately account for the internal dynamics of the molecule. The field dependence of T_1 suggests that global internal motions on a ns timescale should be considered. Therefore a more sophisticated model of the internal dynamics would yield a better description of the data.

It can be concluded that there is anisotropic diffusional motion within H-2D^b/FL, and that internal motions are present. However, ModelFree[226] drives τ_M to an incorrect value (as discussed in Section 5.4.2.1) and thus analysis cannot be completed; model selection is highly dependent on the selected τ_M and such a large discrepancy between the realistic τ_M of 33.8 ns (for the size of H-2D^b/FL) and the predicted τ_M of

25.5 ns is likely to have a significant effect. The absence of R_{ex} terms during model selection shows that ModelFree is unable to describe the dynamics accurately.

ModelFree assigns internal parameters according to the difference from the average and is sensitive to the selected τ_M. Its analysis therefore implies that the overall molecule is rigid such that obtained data is averaged so as to be free of dynamic effects such as exchange. However, in reality the data is reporting on a range of dynamics, and for H-2D^b/FL these dynamics are best observed via other methods.

5.5 Summary

This chapter has described the global and internal solution state dynamics of H-2D^b/FL, using qualitative lineshape analysis, heteronuclear relaxation, and hydrogen exchange data.

Qualitative lineshape analysis, using the ^{1}H-^{15}N spectral planes assigned in Chapter 3, showed that several areas of H-2D^b/FL are likely to possess conformational dynamics. For some residues the chemical exchange is so slow that two peaks can be observed for a single NH (doubled peaks). Broadened linewidths were concentrated in loop regions, although some regions of secondary structure were also affected. Comparing isolated and H-2D^b/FL-bound hβ_2m, the presence of H-2D^b/FL α-chain has resulted in the stabilisation of some regions, in agreement with other studies[136]. For example, the B-C loop beneath the PBD and the D-E loop under the PBG are unassigned in isolated β_2m but broadened in H-2D^b/FL-bound β_2m, such that the presence of the α-chain has apparently decreased the dynamics sufficient for assignment. In the H-2D^b/FL α-chain the PBD is more dynamic compared to the α_3 domain, although both contain broadened peaks.

Heteronuclear relaxation data for isolated hβ_2m is in agreement with lineshape analysis, and reveals further evidence of internal dynamics from a Lipari-Szabo graph on which the C-terminus appears particularly mobile. For many residues there is evidence of slow conformational exchange on a millisecond to microsecond timescale. The same observation was made for H-2D^b/FL-bound hβ_2m, where these internal dynamics result in a disagreement between experimental and HYDRONMR-predicted data. Due to additional data from field dependence, for bound hβ_2m this mismatch is due to internal fast motion (ns) rather than slow conformational exchange (μs-ms). Further experimentation with isolated β_2m at a second field strength would reveal whether the internal dynamics are fast or slow. In both isolated and bound β_2m there is evidence of motion in the D strand β-bulge region; hydrogen-deuterium exchange showed that there is a loss of flexibility when the α-chain is present but that this area retains dynamics compared to the remainder of the protein. For H-2D^b/FL α-chain, the relaxation data highlights mobility in loop regions, as well as in the $\alpha_{2\text{-}2}$ helix and the TPN binding residues 128-136 in the PBD and the α_3 D strand. Interestingly, data suggests that the amino-terminal of antigenic peptide may be less securely bound, with this region containing one dynamic, one stable, and three unassigned residues whilst the carboxyl-terminal PBG closure is made by one unassigned and two stable

residues. When all residues of H-2D^b/FL are plotted on a Lipari-Szabo graph it is clear that a significant proportion are affected by motion, suggesting a slowly tumbling molecule with fast internal motion, perhaps a domain-domain motion on a low nanosecond timescale.

The heteronuclear relaxation data was further analysed using ModelFree to characterise the internal motions. The significance of the internal dynamics of H-2D^b/FL is clear when ModelFree estimated a τ_M of 23.0 ns, likely closer to the effective correlation time. τ_M was therefore predicted using HYDRONMR to give 33.8 ns, closer to what is expected for a 45 kDa molecule. Calculations showed that H-2D^b/FL is an anisotropic protein, with the anisotropic factors obtained from ModelFree analysis verified by comparison with those from a HYDRONMR calculation (D_{ratio} is approximately 1.25). Model selection patterns revealed clear disparities between domains, with the PBD, α_3, and hβ_2m again identified as possessing different dynamics. Analysis resulted in an absence of R_{ex} factors. The usual final step of ModelFree analysis of optimising parameters by varying τ_M to further minimise the overall χ^2 resulted in a τ_M of 25.5 ns; ModelFree cannot accurately account for the overall dynamics of the protein. ModelFree assigns internal parameters according to the difference of relaxation data from the average values and is sensitive to the selected τ_M. Its analysis therefore implies that the overall molecule is rigid such that the average is free of dynamic effects such as exchange. However, in the case of H-2D^b/FL the average T_1 and T_2 is reporting on overall motion and a range of internal dynamics, which would be better described by more complex models.

Finally, the hydrogen exchange from amide sites data was in agreement with the other analyses presented in this chapter for isolated hβ_2m; CLEANEX-PM experiments showed that internal mobility was likely concentrated in loop regions, the borders of unassigned regions, and the D strand β-bulge. However, HX data for the trimeric complex could not be obtained using this experiment.

Chapter 6

6 Discussion

6.1 Technological advances in NMR spectroscopy have enabled analysis of large proteins

MHC class I is a large protein complex at 45 kDa, which stretches the limits for characterisation by NMR. Despite this, NMR provides a powerful means to study the suspected dynamics of this protein at atomic resolution. The two major alternative techniques are X-ray crystallography and MD simulations; the former only provides a snapshot of a complex whilst the latter only predicts the mobility. A key factor in understanding how MHC class I binds and selects immunodominant peptides is the mobility in the peptide binding groove, meaning that crystallography is unable to provide sufficient insight into dynamics in this region of protein, whilst MD simulations require experimental validation. Therefore, in this thesis both crystallography and MD simulations have supplemented the information gained from NMR data. Combining data from multiple techniques in this way generates a comprehensive insight into protein structure, with the high-resolution static atomic structure available from crystallography complementing the dynamics data obtained through NMR, and with the molecular dynamics data enabling further insight into the potential dynamic coupling of different areas of MHC class I.

6.1.1 Technical advances in NMR

Investigation of a protein of 45 kDa requires implementation of the most recent developments in NMR because of its size. The decreased tumbling time of large proteins causes the NMR signal to decay quicker, resulting in broad resonances. In addition, the increased number of residues in large proteins results in spectral crowding. The availability of high field spectrometers significantly increases spectral resolution. The additional development of TROSY-based pulse sequences means that NMR study of large proteins is a growing field[230].

NMR research requires the collection of numerous spectra in a variety of 2D and 3D combinations (data from in excess of one hundred spectra is presented in this thesis) and these spectra have to be reliably collated in specialist software. Although assignment is not yet fully automated, software such as CCPN Analysis ensures that

cross-referencing the different spectra is a less time-consuming process than would otherwise be the case. Software with the ability to compare data from different techniques (e.g. NMR versus crystallographic structural restraints) is another recent advance.

Developments in both software and hardware have increased the availability of MD simulations (e.g. via the open-access WeNMR internet portal) and permit ever-increasing lengths of MD simulations to be performed such that MD timescales of 500 ns are possible and potentially more comparable with the timescale of the motions relevant for NMR observables. This means that the protein motion observed in MD can be verified using NMR; a significant development in the investigation of protein dynamics in solution. Although the focus of this thesis has not been to verify independently conducted MD simulations, it is clear that the MD simulations of H-2D^{b}/FL by Dr Bailey[22,164] broadly agree with NMR data. Careful comparison of MD data and NMR data is required because the simulations contain many simplifying assumptions. When the data is compared between the two techniques and found to agree, molecular dynamics provides an insight into atomic motion in which the effect of structural perturbations (e.g. point mutations) can be examined with much greater ease than through generating mutated protein for NMR study, and thus MD is a useful tool for exploring potential new experiments.

A further advance that permits feasible investigation of large proteins using NMR is the wide availability of various growth media and isotopic labels. Protein production is dependent on bacterial expression, and the improvement in NMR signal obtained from the use of particular isotopes (Section 1.5.3.1) cannot be ignored. For example, the decreased cost of deuterated water and glucose has allowed them to be used as a supplement for large-scale bacterial growth; a single 600 µl 600 µM sample of H-2D^{b}/FL may require six litres of *E. coli* culture. The high cost of such media and labels mean that it is vital for the work in this thesis to establish a reliable bacterial growth protocol; as a result of the experiments presented here this is now in place for H-2D^{b}/FL. The same protocol has been used for other proteins with a high success rate.

6.1.2 Application of new techniques

All of these improved techniques have been used in this thesis to investigate H-2D^{b}/FL. During the resonance assignment, isotopic labelling combined with careful selection of NMR experiment permitted differentiation between protein subunits. Selective amino acid unlabelling (for Lys, Tyr, and Arg) was used to further the

completeness of the assignment; e.g. collection of data where all visible peaks originate from Arg residues made it very easy to differentiate Arg from the similar chemical shifts of Glu, His, and Trp. The use of TROSY-modified pulse sequences rather than the standard HSQC-basis has increased the spectral quality; the large number of peaks produced by large proteins can result in spectral crowding that prevents sets of chemical shifts being identified. When recording data, magnetisation pathways were kept as short as possible whilst still collecting complementary spectra such as the HNCA and HNCACB pair. Combined, these techniques have resulted in the most complete backbone resonance assignment of an MHC class I protein complex to date, with 88% of backbone HN resonances assigned. This permits further investigation into the properties of H-2D^b/FL, providing data which has previously only been implied by molecular biology approaches or which has only been observed via molecular dynamics simulations. Significantly, 32 residues were identified as occupying multiple resonance states, suggesting that H-2D^b/FL is not a rigid protein and leading into research around its mobility in solution.

When determining the solution state domain arrangement of H-2D^b/FL, ^{1}H-^{15}N RDC data was used rather than H-H distance restraints. A large set of data was collected through the use of multiple alignment media, the more sensitive J-modulation experiment in place of the doublet splitting measurements from an H-N plane[211], and via the use of the TROSY-based ARTSY pulse sequence instead of the more traditional HSQC-based approach[212]. The use of TROSY-based pulse sequences has also permitted the collection of heteronuclear relaxation data[220] for the characterisation of the backbone dynamics of H-2D^b/FL.

6.2 H-2D^b/FL MHC class I is an intrinsically dynamic protein

Various studies into MHC class I have suggested that it is capable of undergoing some internal motion; these include the variation in crystal structures obtained (across various α-chain alleles), both *in vivo* and *in vitro* experiments, and molecular dynamics simulations (for detail see Section 1.4.3). In brief, studies involving TPN suggest that there is widening of the peptide binding groove to produce a peptide-receptive conformation. The backbones of the PBD α-helices, as well as residues forming the F pocket, are implicated through *in vitro* peptide binding rate, FRET, and antibody binding studies, and through molecular dynamics simulation.

The results presented in this thesis demonstrate conclusively that H-2D^b/FL MHC class I is also intrinsically dynamic. The dynamics of proteins are between a liquid (where diffusion describes the relationship between displacement and distance travelled) and a solid (where the atoms vibrate around fixed positions). In a folded protein the backbone and side chain atoms are dynamically distinct, such that side chain motions are more liquid-like because in the backbone secondary structure elements the atoms are stabilised by hydrogen bonds. NMR is able to measure local (predominantly backbone) motions on a sub-nanosecond timescale using heteronuclear relaxation, with residues not in regular secondary structure elements expected to possess larger amplitude motion due to the decreased hydrogen bonding. For such loop regions MDS can be employed to investigate the cause of increased motion and may be able to link it to either a domain reorientation or a collective secondary structure rearrangement such as the motion in the α-helices. NMR can also investigate microsecond to millisecond motion through T_2 data, and although analysis is usually restricted to a simplistic assumption of a two state interconversion, valuable timescale information can be obtained. Here MDS may be used to suggest the structures that give rise to these states. NMR hydrogen exchange experiments provide information on a third timescale of milliseconds to hours by examining hydrogen bond breaking.

6.2.1 Initial analysis

During resonance assignment, it became apparent that several peaks were missing from the H-N spectrum; for hβ$_2$m three residues were apparently absent, whilst H-2D^b α-chain backbone could only be 81% assigned (Figure 3.5) with insufficient remaining peaks for a complete assignment to be possible. This is suggestive of peak broadening due to chemical exchange in which residues experience two or more

different local chemical environments i.e. two or more different conformations. These residues are located mostly in loop regions between elements of regular secondary structure, although interestingly both the α-helices and the platform of the PBD are affected (Figure 3.5, Appendix C).

MD simulations based on the crystal structure of H-2D^b/FL were carried out to predict chemical shifts, and this 560 ns simulation showed significant movements within the protein structure particularly in loop regions and in the TPN-binding regions. When the predicted chemical shifts were examined for NMR-unassigned residues, there were multiple populations in the MD simulations for some residues (e.g. H-2D^b α-chain residues 96-101) supporting the notion that some residues might be missing due to chemical exchange. However, these residues with predicted multiple populations do not always match the unassigned residues (e.g. H-2D^b α-chain residues 14-19) so it cannot be assumed that *all* unassigned residues are affected by protein dynamics. Some peaks on the assignment spectra were not sufficiently conclusive to be assigned. Additionally, some residues will be affected by ring currents e.g. an α-chain unassigned area above the hβ$_2$m D-E loop is likely to be due to the aromatic ring of hβ$_2$m Trp60 although other causes such as mobility cannot be ruled out.

Taken together, the many NMR-unassigned residues that match multiple populations in the MD simulations show that H-2D^b/FL MHC class I is subject to fluctuating protein conformations. These occur mostly in the PBD but not only in loop regions.

6.2.2 Dynamic residues are distributed throughout H-2D^b/FL

Further investigation of H-2D^b/FL MHC class I protein dynamics began with linewidth analysis and included heteronuclear relaxation and hydrogen exchange data. The hβ$_2$m domain contains both linewidth-broadened and peak doubled areas, with heteronuclear relaxation data confirming that this is mostly confined to loop regions, particularly the B-C and D-E loops that face towards the PBD (Figure 5.3). The α$_3$ domain appears to possess the most stable conformation of all four H-2D^b/FL domains, with relatively few broadened or doubled peaks. The loop regions facing away from the PBD are highlighted in all analyses as most dynamic, which could be expected since they are not confined by nearby structures. The relaxation data suggests that the D strand TPN-binding region also experiences mobility (Section 5.3.2.2, Figure 5.18); this correlates with observations of flexibility at the second recognition site on the PBD platform at residues 128-136. The PBD-α$_3$ linker region

was also highlighted in these three datasets and when this data was compared to MD simulation it appears that this motion is not simply a local perturbation but part of a larger hinge motion between the peptide binding and α_3 domains[155,164].

The PBD possesses the most evidence of protein flexibility; lineshape analysis and heteronuclear relaxation highlighted that the helices are affected more than the platform, and that the α_2 helix is more mobile than the α_1 helix. When the relaxation data is further interpreted using the Lipari-Szabo method, it is clear that the majority of residues throughout H-2D^b/FL experience fast motion with a timescale in the low nanoseconds. When this data is compared to values predicted from the crystal structure, there is a significant discrepancy due to internal motion; due to the depressed ^{15}N-T_1 values in the NMR data, it is possible to conclude that in addition to the overall tumbling of the whole complex, the backbone NH data reports global internal motion on a nanosecond timescale as well as fast local internal motion. Investigating this further by examining MDS data, these lower than expected T_1 values may instead be the result of inter-domain motion on a nanosecond timescale; MD data shows the dominant principal component of collective motion in H-2D^b/FL is inter-domain motion[155,164] (Figure 6.1).

hβ_2m is stabilised upon binding H-2D^b/FL α-chain

There are a number of comparisons that can be made between free hβ_2m and that bound by H-2D^b/FL α-chain. The immunoglobulin fold is present in both, with regular secondary structure elements found in the same locations. During the assignment of the hβ_2m domain, a pre-existing resonance assignment for free hβ_2m was used as a guide; several residues have very similar chemical shifts although a perturbation for some side chain signals is apparent. Due to the close contact with the H-2D^b α-chain the number of unassigned residues decreases from twelve to three. This is likely due to a decrease in exchange broadening; in both forms broadened linewidths occur in loop regions or are directly adjacent to unassigned residues. The B-C and D-E loops are both very mobile: these regions are stabilised by the presence of α-chain because in isolated hβ_2m they were too broad to be assigned at all. All data collected here (linewidth, heteronuclear relaxation, and hydrogen exchange) shows that the degree of flexibility appears to decrease in H-2D^b/FL-bound hβ_2m, both in the aforementioned loop regions and in the majority of the β-strands (noticeably the A and D strands). In both isolated and bound β_2m the B strand is affected by broadened ^{1}H linewidths, although ^{15}N relaxation data shows that the C-terminus is stabilised by the presence of H-2D^b α-chain.

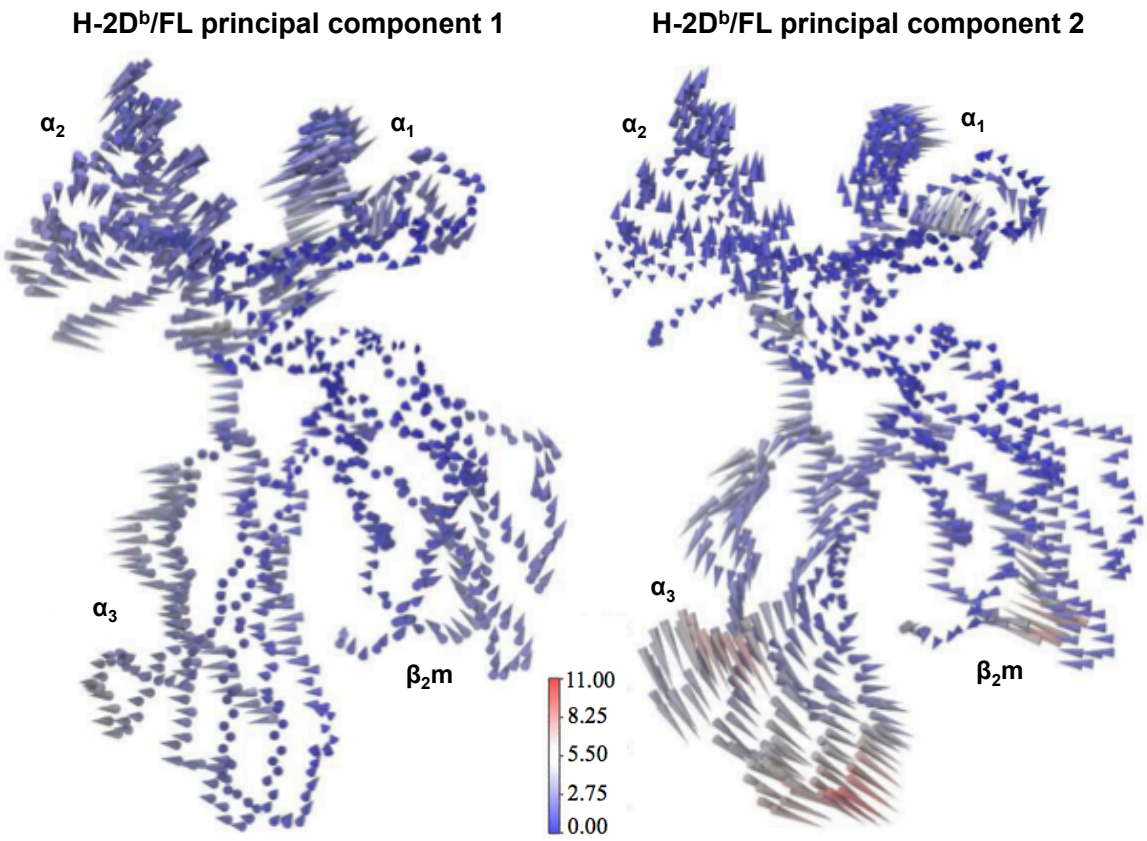


Figure 6.1 – H-2D^b/FL MDS principal component analysis.
Principal component analysis describes the collective atomic motions, with principal components 1 and 2 accounting for approximately 50% of the total motion in a combined MD trajectory of 850 ns. The basis for the analysis is a covariance analysis carried out using the backbone atoms of the protein (N-Cα-C). Porcupine plot generated using cones attached to each backbone atom to indicate the amplitude and direction of the atoms motion as a liner interpolation between its extreme conformations when transformed into Cartesian coordinates. The colour bar indicates the magnitude between extremes in Å. The peptide was included in the MD trajectory but not in this analysis. Principal component 1 contains domain-domain twisting between the PBD and α_3/β_2m domains, and principal component 2 is characterised by the flexing together and apart of the α_3 and hβ_2m domains. Figure adapted from A. Bailey[164].

In existing structures of both forms of hβ_2m, the D strand of the immunoglobulin fold is not canonical and is broken into two sections. Intriguingly, both hydrogen exchange and the relaxation data suggest that although the D strand is stabilised in the protein complex it retains more flexibility compared to the other β-strands. Analysis of predicted backbone torsion angles shows that His31 may retain the ability to form the alternate conformation. This β-bulge region is traditionally thought to form an alternative conformation only in isolated hβ_2m; data presented here (in agreement with other NMR studies[136]) suggest it retains structural mobility whilst in complex.

6.2.3 Impact of internal mobility on characterisation of dynamics

The presence of dynamics throughout the protein complex significantly complicated the NMR-assisted refinement of the high-resolution X-ray crystal structure. Despite removing residues already identified as flexible the ModelFree analysis was unable to estimate τ_m. It had already been identified that the T_1 values are systematically too small due to either internal fast motion or domain-domain motion, and therefore the software was unable to generate a suitable starting point for the estimation of the overall dynamics. As a consequence, fitting the internal dynamics at the same time as the overall dynamics was unsuccessful because even a small error in τ_m significantly changes the model of internal motion. This inability to further characterise the dynamics of H-2D^b/FL provides further evidence that there is widespread internal motion.

6.3 In the case of H-2D^b/FL MHC class I, there are significant differences between the crystal and solution state structures

H-2D^b/FL has been investigated using a range of techniques, including X-ray crystallography, *in vivo* and *in vitro* studies, MD simulations, and in this thesis using NMR. H-2D^b α-chain has been refolded with a variety of peptide ligands to produce a number of published crystal structures, and when these structures are compared it is clear that the elements of secondary structure are in agreement. However, when the PBD domains of these various structures are aligned the comparative structures show variation in the relative alignments of the remaining α_3 and β_2m domains (Figure 6.2). Thus, it appears that the available crystal structures may provide 'snapshots' of the solution structure ensemble. Specific to H-2Db/FL, when the structural results from crystallography and NMR are compared, there are significant differences. In addition, evidence from both MDS and NMR suggest that the α-helices of the PBD are capable of motion relative to the platform base.

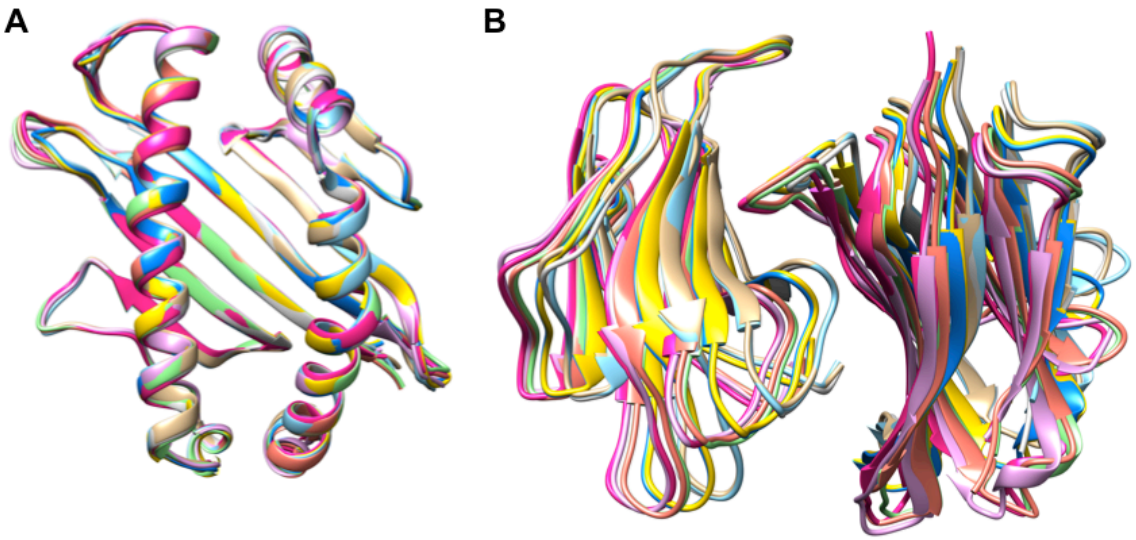


Figure 6.2 – Comparison of published crystal structures for H-2D^b MHC class I in complex with a range of peptide ligands.
Eight structures are compared with the H-2D^b/FL crystal structure, PDB 1CE6[98] (tan). The remaining eight structures are: PDB 2CII[79] (light blue), PDB 1YN6 (dark blue) and 1YN7 (yellow)[231], PDB 1FFN[121] (purple), PDB 1INQ (green) and 1JUF (orange)[232], PDB 1QLF[98] (grey), and PDB 1BZ9[233] (pink). **(A)** Structures are similar when aligned to the PBD. **(B)** Alignment in the PBD leads to differing relative orientations of the α_3 (right) and β_2m (left) domains.

6.3.1 Differences in chemical shifts

Chemical shift values predicted from MD simulations and those predicted from the crystal structure of H-2D^b/FL were compared to those obtained by NMR. This showed that, despite the averaging effect, NMR is a better representation of the solution-state structure as assessed by MD; the crystal structure is normally capable of capturing one possible conformation at most, and the chemical environment represented is possibly not the most common conformation in solution. Indeed, in the $\alpha_{2\text{-}2}$ helix there is a significant difference in the chemical shifts from the crystal structure, suggesting that this region takes a different conformation in the solid state than in solution state. This may represent an arrangement that is produced during the crystallisation process. However, it must be noted that the crystal structure does agree with the solution-state structure in many ways; for example, secondary structure elements predicted from NMR and seen in MD simulations match those in the solid state.

Overall, the NMR data is in better agreement with the MD simulations than the crystal data, although the MD data does suggest that not all conformations are captured in the NMR. The fact that on average the Cα chemical shifts from the MD simulation are closer to the observed NMR shifts than the X-ray structure (Figure 3.13) suggests that the ensemble is a better representation of the solution state than the X-ray structure. However, for some residues the X-ray structure fits the data better than the MD average, suggesting that further analysis is needed to wholly validate this approach. It is worth noting that there are very few residues for which the MD-predicted range of chemical shifts does not contain the NMR observed value.

6.3.2 Differences in relative domain orientation

During the analysis of RDC data, it again became apparent that the crystal structure is not an accurate representation of the solution state because when NMR data was compared to couplings back-calculated using Module1.0 from the crystal structure there were regions of significant difference. These were widespread and not confined to loop regions lacking regular secondary structure. In addition, there were different relative domain orientations; the NMR data was split into six domains (the PBD platform, α_1, $\alpha_{2\text{-}1}$, and $\alpha_{2\text{-}2}$ helices along with the α_3 domain and hβ_2m; insufficient sampling of the NH bond orientation meant that the helical domains could not be individually analysed but a robust relative orientation could be calculated for the PBD platform, α_3, and hβ_2m domains) to ensure a fit to the crystal structure model. This

occurred for both alignment media employed and hence is a significant finding, meaning that these six domains have different relative orientations to those shown in the crystal structure. Although it was not possible to collect sufficient RDC data to calculate a high-resolution NMR structure it was possible to investigate these relative domain orientations. There is a 20° twist of the α_3 and hβ_2m domains relative to the platform, slightly larger than the twisting shown in the ensemble of crystal structures in Figure 6.2.

A refinement of the crystal structure using the NMR RDC data was not possible with the available data, but did show a statistically significant difference between crystal and solution state. MD simulations originating from the crystal structure using the RDC data to inform the motion revealed that the crystal structure is not necessarily at an energetic minimum; lending credence to the argument that the crystal structure is partly shaped by the requirements of lattice formation during crystallisation. The twist between platform and α_3/β_2m was observed from the MD (Figure 4.4C), although the amount of twist varied between simulations. However, this twisting was more significant here than in other MD simulations[150,152,158,164] so although it clearly exists it is not possible to be conclusive as to the extent and direction. The majority of MDS investigations focus on dynamics solely within the PBD, with evidence from this thesis demonstrating that motion is more widespread throughout the entire trimeric complex. Studies examining correlated dynamics between the domains of MHC class I are in agreement that the platform-α_3/β_2m twist is a 'principal component' of overall dynamics[164] (Figure 6.1). This significant twist would appear to be in agreement with the heteronuclear relaxation data analysis showing pervasive internal motion. Overall, it is clear that the protein complex is dynamic in solution with a large component being a PBD-α_3/β_2m twisting motion.

6.4 MHC class I internal dynamics has significant implications

H-2D^b/FL can be assumed to be an accurate model for classical MHC class I alleles; structure and function are essentially identical, with complex maturation involving a conformational change[74] and observed to occur *in vivo* by association with the PLC[20] as for other alleles. Peptide binding hierarchies are present, and although there is a ridge within the PBG (Figure 1.2) this does not affect how antigen is bound[107] or presented[234], with peptide anchor residues and buried termini still evident. The use of hβ_2m in place of the murine version has no effect on either structure[98,106] or function[183]. It is important to consider the possible effects of a dynamic solution state structure on the function of MHC class I.

6.4.1 For the presentation of endogenous antigen

In the adaptive immune system it is antigen presentation to T lymphocytes that is vital for recognition of pathogens. Focusing in this thesis on MHC class I presenting endogenous peptides to CD8$^+$ T cells, in order to understand how the immune system targets a particular pathogen (or indeed how a pathogen may evade detection) it is vital to first understand how one peptide is selected from the pool of available processed antigens within the ER. Association of peptide with MHC class I occurs within the peptide loading complex where a number of chaperone proteins (most notably tapasin) facilitate both binding and optimisation, leading to the existence of immunodominance in which a subset of high-affinity peptides are preferentially bound by MHC.

X-ray crystallography has been used to determine the structures of several class I alleles in complex with immunodominant peptides, leading to observation of an apparent impediment to peptide optimisation: in the mature complex the peptide termini are buried at either end of the peptide binding groove (Figure 1.2) and hence there must be a structural change within the MHC α-chain to permit exchange. The existence of immunodominance[64] means that peptide exchange does occur[65], with numerous molecular dynamics studies suggesting that the PBG is capable of flexibility in solution[164]. Data presented here conclusively confirms the presence of dynamics within the mature structure of MHC class I, with several datasets (e.g. lineshape analysis and heteronuclear relaxation; Section 5.5) suggesting motion in both PC and PN pockets. This correlates with the fact that motion within the α-chain surrounding both ends of peptide will be required for peptide exchange. Studies involving partial peptide epitopes have shown that occupation of the PC pocket is vital for stability

whilst peptides lacking the PC anchor are unable to facilitate complex maturation[79]. This can be explained using the NMR data, which (in agreement with MDS results, Section 3.4.2.3) suggests that residues surrounding PN are more dynamic than those forming the PC pocket; the α-chain dynamics will increase peptide motion, decreasing affinity and increasing peptide off-rate.

For the peptide termini to be exposed the peptide binding groove must widen, as has been hinted at by antibody binding and FRET studies, and described in simulation by a large number of MD studies (see Section 1.4.3.2). It is clear from the NMR data that the PBD is inherently dynamic and that it is possible for the α-helices to attain alternative conformations relative to the α-chain platform when compared to the 'closed-groove' X-ray crystallography structures (Section 4.4). It is clear that the helices are capable of motion in the solution state and it is likely that this motion correlates with that observed via MDS[164] and relates to a widening of the PBG. Given that in the mature, stable structure the peptide termini are always buried, this widened groove means that the MHC complex is less stable, adding weight to the theory that *in vivo* the tapasin chaperone acts as a stabilising element for peptide-empty open-groove MHC during peptide optimisation within the ER. The presence of a widened PBG will also have the effect of universally decreasing peptide affinities, thus explaining the observation that peptide dissociation rates are faster inside the ER than on the cell surface[73] and that peptide binding half-lives are longer than the amount of time available for peptide optimisation to occur[57,72]. Interestingly, the presence of such an intermediate state during peptide binding mirrors enzyme studies in which the transition state intermediate structure is the most important factor in the selectivity of the enzyme[235].

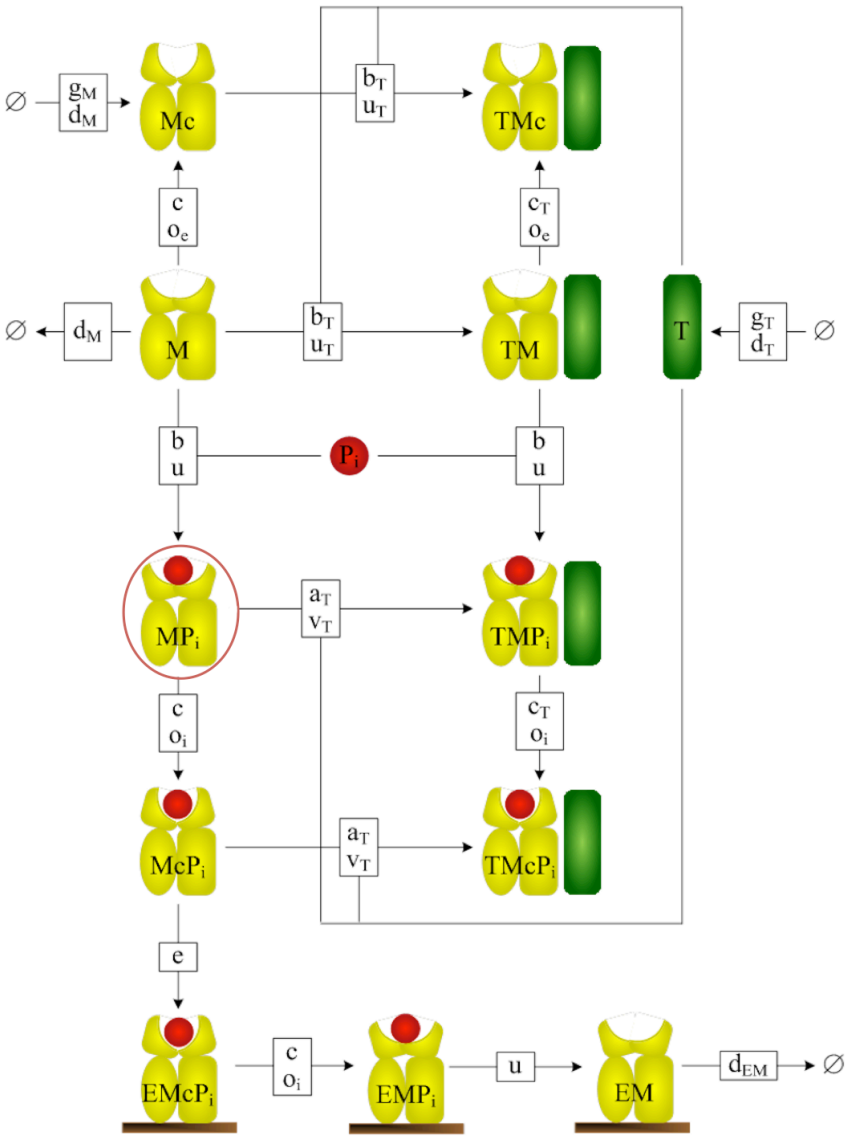
For the majority of MHC class I alleles, TPN is the major requirement for peptide optimisation (Section 1.3.2). However, although TPN contacts both the α_2 and α_3 domains it appears to lack direct contact with the PBG. It is therefore unclear how TPN is able to effect a widening of the PBG to increase peptide off-rate. Examining the TPN binding sites themselves, these α-chain residues possess dynamic properties (both prior to the $\alpha_{2\text{-}1}$ helix in the PBD and in the D strand of α_3) that would be affected by TPN binding but from the NMR data alone there is no clear mechanism for a transmission of a TPN-induced conformational change from the α_3 binding site to the PBD and thence to the PBG, despite extensive evidence from MD studies that such transmission might occur. The picture is made clearer by taking into account the NMR evidence for a domain-domain twist of the PBD relative to the α_3/hβ_2m domains

(Section 4.4) and detailed MD analysis of H-2D^b/FL by A. Bailey[164]. MDS shows widespread strong 'correlation' (coupled motions) between residues throughout the α-chain, such that motion in one location effects motion at another. In addition, the MD also observes a significant domain-domain twist, with principal component analysis revealing that this twisting is coupled to the opening and closing of the PBG[164]. This leads to a possible mechanistic explanation for the effects of TPN; TPN contacts the D strand of α_3 and a region in the α_2 domain, such that binding causes a domain-domain twist of the platform relative to α_3 (and hence β_2m), thus effecting a widening of the PBG and permitting peptide optimisation. When a peptide of sufficiently high affinity to close the PBG binds, it results in a domain-domain twist in the opposite direction, signalling to TPN that peptide optimisation has occurred and permitting TPN dissociation and MHC egress to the cell surface. Alternatively, it may be that TPN binds to empty MHC class I and in doing to allosterically induces a dynamic change in the PBD characterised by the exploration of more conformation space and a rapid interconversion between open-groove intermediate and closed structures irrespective of peptide. This permits the sampling of peptide binding strength in a closed-groove state such that lower affinity peptides dissociate quicker; thus determining the editing efficiency of MHC class I. In either case, the NMR data presented in this thesis is in agreement with *in vivo* and *in vitro* data for both BF2*15 and BF2*19 (chicken)[76,161] and B*4402 and B*4405 (human) alleles[155], which also conclude that there is dynamic coupling between the α_3 and peptide binding domains. To uncover the exact mechanism will require further data from both biophysics studies and computational simulations.

The role of TPN in antigen presentation on MHC class I closely mirrors the role of HLA-DM in antigen presentation on MHC class II. HLA-DM controls antigen binding to the binding groove and induces dissociation to create a peptide repertoire skewed toward high-affinity binding[236]. Peptide-empty MHC class II is stabilised by HLA-DM in a peptide-receptive conformation[236], with clear comparisons to the intermediate empty, peptide-receptive MHC class I state.

In summary, in solution there are significant structural fluctuations within MHC class I that permit peptide optimisation. For peptide exchange the PBG widens to expose the buried peptide termini, a process often facilitated by the chaperone TPN. However, the overall complex is potentially not sufficiently stable to permit exit from the ER when the PBG is 'opened' by outward motion of the α-helices. Therefore it could well be that there is an initial peptide-empty MHC complex, followed by a

peptide-receptive open-groove intermediate conformation, before a high-affinity peptide matures the complex to attain the final, closed-groove stable conformation. The molecular dynamics simulations have been used to independently formulate a 'dynamical systems model' incorporating this intermediate structure[22] (Figure 6.3). Given the strong agreement between NMR and MDS data for H-2D^b/FL it is likely that this model best represents the current understanding of the presentation of endogenous antigen by MHC class I. When this model was tested with *in vivo* biochemical data[155] it showed that allelic differences in peptide selection efficiency could be explained by differences in their ability to undergo conformational changes.



Components:		Kinetic Rates:	
M	MHC	g	generation
Mc	closed MHC	d	degradation
MPi	MHC-peptide complex	b/a	binding
McPi	closed MHC-peptide complex	u/v	unbinding
T	Tapasin	c	closing
TM	TPN-MHC complex	o	opening
TMPi	TPN-MHC-peptide complex	e	egress
TMcPi	TPN-closed MHC-peptide complex		
EM	egressed MHC		
EMPi	egressed MHC-peptide complex		
EMcPi	egressed closed MHC-peptide complex		

Figure 6.3 – Proposed model of peptide editing by MHC class I.
Each shape represents a molecular species and each box a reaction. The boxes are labelled with reaction rates; a single rate denotes an irreversible reaction, and two rates a reversible reaction. The rate of the forward reaction is on top. In this model, closed MHC that is peptide-empty (Mc) can become open (M). Peptide can then bind to form MPi (circled in red), the key intermediate state in which peptide is bound but the groove is open. The groove must close (McPi) to allow egress (e) to the cell surface (EMcPi), where the complex maintains plasticity with reversible groove opening (EMcPi ↔ EMPi) before eventual peptide unbinding (u), leading to degradation (d_{EM}). Within the cell, TPN can associate (or dissociate) at any stage to alter the kinetic rates. Figure is updated from Dalchau *et al*[22] in collaboration with A. Bailey[164], and is tested through use of *in vivo* biochemical data in Bailey *et al*[155].

6.4.2 For recognition by T-cell receptors

The generation of an effective immune response to pathogen relies on immunodominance in terms of both the number of MHC class I complexes found on antigen presenting cells and the numbers of T lymphocytes with complementary TCRs: more than 90% of immunodominance is generated by the selection of high affinity peptide by MHC class I to ensure that a sufficient number of complexes are available to activate the $CD8^+$ T cells[49]. Immunodominance within MHC class I antigen binding is vital to ensure the complex is stable for sufficient time on the cell surface for the interaction with the TCR complex[50]. The interaction between MHC class I and a TCR heterodimer involves several co-receptors that are necessary for the initiation of the signalling cascade for CTL activation, but initial contact is solely between the PBD of MHC and the CDRs of the TCR (Figure 1.3). The presented antigen is then surveyed by the TCR, with the initial diagonal angle of engagement likely modified to maximise peptide contact with the CDRs[51]. Binding geometry appears to be significant, with the strength of the immune response linked to matching structural fluctuations in the peptide and TCR[122,123]. Given that NMR data has shown that MHC class I is inherently dynamic, especially in the PBD (even in an immunodominant structure), the motion within the PBG will be transmitted to the peptide, creating peptide dynamics that may be important in instigating an immune response.

The effect of MHC class I dynamics on TCR recognition is an area that requires further investigation; given the logistical issues involved in generating a stable MHC-TCR complex for NMR analysis (e.g. reliable high-yield protein expression, and successful secondary structure generation incorporating a stable cross-link to boost the relatively low-affinity interaction) and the protein size limitation of NMR itself, molecular dynamics simulations will be vital. Several research groups are making progress in this area. Information regarding H-$2D^b$/FL dynamics obtained using MDS is in remarkable agreement with data generated using NMR; acting as validation of MD as a technique for investigating the dynamics of MHC class I.

6.5 Conclusions

The results of this thesis demonstrate that MHC class I molecules are intrinsically dynamic in the solution state. In Chapter 2 protein production methods were developed to allow reliable generation of a mature H-2D^b/FL MHC complex, deuterated to make it suitable for NMR analysis. Chapter 3 follows the backbone resonance assignment for this complex, resulting in 85% of residues being located on the NMR spectrum. Chapter 4 described the collection of residual dipolar coupling data and the characterisation of relative orientation between the PBD, α_3, and β_2m domains, whilst in Chapter 5 the dynamics of the complex were probed using hydrogen exchange, qualitative lineshape analysis, and heteronuclear relaxation data. This work revealed MHC class I to have a fast internal motion.

MHC class I is inherently dynamic, experiencing motion on a range of timescales from slow (microsecond-millisecond) to fast (nanosecond) in regions throughout the structure. Greater mobility is found in the PBD than in either α_3 or β_2m. In regard to β_2m, the structure is stabilised upon binding to the α-chain but the D strand (the location of the β-bulge implicated in amyloid formation) remains flexible. In the α_3 region it is notable that the TPN binding region is dynamic in all NMR analyses, as is the secondary TPN binding region in the PBD. The α-helices bordering the PBG contain a number of residues with characteristics indicating flexibility, suggesting they have the ability to move relative to the PBD platform, in agreement with MD simulations[155,164]. The peptide C-terminus is likely more secure the than N-terminal, in agreement with findings that the C-terminal alone, but not the N-terminal alone, is sufficient to generate stable mature complex[79].

This solution state mobility means that structures obtained using X-ray crystallography are not necessarily wholly accurate representations of MHC class I, acting instead as snapshots of stable conformers. This is particularly true when relating structure and function; for example, the key observation of a twist of the PBD relative to α_3/β_2m obtained via NMR RDC analysis is not captured in crystallographic structures but is found in MD simulations. Indeed, comparing MD and NMR data provides the first strong validation of these simulations. Recent technological advances in NMR have made this investigation possible, but the presence of such widespread mobility has precluded the use of some established data analysis techniques, and it is clear that MD studies will prove vital in future investigations.

The dynamics of MHC class I have significant implications in our understanding of how antigen is bound, optimised, and presented. The data and analyses presented in this thesis are consistent with the presence of an intermediate conformation during peptide optimisation, in which the PBD is peptide-receptive and has an open binding groove, and that is different from the mature peptide-bound closed-groove state.

Appendices

Appendix A

Protein Sequences

Human β2m

```
          9          19          29          39          49
MIQRTPKIQV YSRHPAENGK SNFLNCYVSG FHPSDIEVDL LKNGERIEKV

         59          69          79          89          99
EHSDLSFSKD WSFYLLYYTE FTPTEKDEYA CRVNHVTLSQ PKIVKWDRDM
```

Murine H-2D^{b}

```
          9          19          29          39          49
MGPHSMRYFE TAVSRPGLEE PRYISVGYVD NKEFVRFDSD AENPRYEPRA

         59          69          79          89          99
PWMEQEGPEY WERETQKAKG QEQWFRVSLR NLLGYYNQSA GGSHTLQQMS

        109         119         129         139         149
GCDLGSDWRL LRGYLQFAYE GRDYIALNED LKTWTAADMA AQITRRKWEQ

        159         169         179         189         199
SGAAEHYKAY LEGECVEWLH RYLKNGNATL LRTDSPKAHV THHPRSKGEV

        209         219         229         239         249
TLRCWALGFY PADITLTWQL NGEELTQDME LVETRPAGDG TFQKWASVVV

        259         269
PLGKEQNYTC RVYHEGLPEP LTLRWEP
```

FL Peptide

FAPGNYPAL

FA Peptide

FAPGNYPAA

NL Peptide

NYPAL

Appendix B

Protein production buffer and media recipes

Reagents were obtained from Sigma-Aldrich unless otherwise stated.

LB Media (1 l)

20 g/l pre-mixed pellets (Melford Laboratories)
After autoclaving:
100 µg/ml Ampicillin

Minimal M9 Media (1 l)

6 g Na_2HPO_4 pH 7.4
3 g KH_2PO_4
0.5 g NaCl
1 ml 1M $MgSO_4$
0.3 ml 1M $CaCl_2$

After autoclaving:

	Desired protein isotopic labelling			
	$^{2}H^{15}N$	**$^{2}H^{13}C^{15}N$**	**$^{13}C^{15}N$**	**^{15}N**
Ammonium chloride[a]	1 g $^{15}ND_4Cl$	1 g $^{15}ND_4Cl$	1 g $^{15}NH_4Cl$	1 g $^{15}NH_4Cl$
Glucose[a]	3 g D_7-glucose	3 g ^{13}C-D_7-glucose	3 g ^{13}C-glucose	3 g
Biotin (1 mg/ml)	1 ml (in D_2O)	1 ml (in D_2O)	1 ml (in H_2O)	1 ml (in H_2O)
Thiamine (0.1 mg/ml)	1 ml (in D_2O)	1 ml (in D_2O)	1 ml (in H_2O)	1 ml (in H_2O)
Trace elements[b]			10 ml	10 ml
Rich media[c] (OD1)	3%	3%		
Ampicillin (100 mg/ml)	1 ml (in D_2O)	1 ml (in D_2O)	1 ml (in H_2O)	1 ml (in H_2O)

[a] Obtained from Goss Scientific

[b] Trace elements per litre:
5 g Na_2EDTA, 0.5 g $FeCl_3$, 0.05 g $ZnCl_2$, 0.01 g $CuCl_2$, 0.01 g $CoCl_2.6H_2O$, 0.01 g H_3BO_3, 1.6 g $MnCl_2.6H_2O$

[c] Obtained from Silantes GmbH

LB/M9 Agar

As per LB or M9 recipe or a combination LB/M9, plus 3% Micro-Agar.

SDS-PAGE Gel Loading Buffer

2.5 ml 1M Tris pH 6.8
4 ml 20% SDS
40mg Bromophenol blue
2 ml Glycerol
1 ml dH_2O

SDS-PAGE 12% Gel

	Stacking Gel	**Running Gel**
Protogel 33% Acrylamide	0.66 ml	3.33 ml
dH_2O	2.9 ml	3.33 ml
0.3% SDS, 3M Tris pH 8.45	1.24 ml	3.33 ml
TEMED	10 µl	10 µl
Ammonium persulphate	25 µl	50 µl

SDS-PAGE Running Buffers

Cathode Buffer:

100 mM Tris pH 8.2
100 mM Tricine
0.1% SDS

Anode Buffer:

200 mM Tris pH 8.9

SDS-PAGE Stain

Coomassie:

0.2% Brilliant blue dissolved in MeOH
40% Methanol
10% Acetic acid
50% dH_2O

Destain:

10% Methanol
50% Acetic acid (glacial)
40% dH_2O

Triton Buffer

50 mM Tris pH 8
100 mM NaCl
0.5% Triton X-100

Resolubilisation Buffer

8M Urea pH 6.5
50 mM MES
0.5 mM EDTA
2 mM DTT

Refolding Buffer

100 mM Tris pH 8
400 mM L-arginine
2 mM EDTA
5 mM reduced glutathione
0.5 mM oxidised glutathione

PBS Buffer

Phosphate buffered saline (as per manufacturers instructions)
0.1% sodium azide

NMR HEPES Buffer

50 mM HEPES pH 7
0.05% sodium azide

RDC 5% Polyacrylamide Gel

5% acrylamide
0.1% ammonium persulphate
0.5% TEMED

6 mm tubing for polymerisation
1 l H_2O for washing
Gel trimmed to 2.1 cm

Appendix C

Chemical shifts for H-2D^b/FL

Table 6.1 – Chemical shifts of the hβ$_2$m component of H-2D^b/FL.
Residues affected by peak doubling in the ^{1}H-^{15}N dimension are listed twice, shaded blue. Resonances that could not be assigned (Pro ^{1}H and ^{15}N, and Pro-1 carbonyls) are shaded grey. Missing assignments are shaded red.

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Ile	1			174.99	60.71
Gln	2	8.24	123.74	175.30	54.96
Arg	3	9.24	123.01	173.78	54.91
Thr	4	8.00	118.78		58.44
Pro	5			176.88	62.56
Lys	6	9.19	125.03	177.43	55.58
Ile	7	8.30	123.21	175.99	60.73
Gln	8	8.77	124.01	171.09	55.95
Val	9	9.32	124.33	172.90	60.11
Tyr	10	8.48	123.32	168.50	56.51
Ser	11	8.92	114.03	174.85	55.99
Arg	12	8.78	122.89	171.87	59.57
His	13	8.15	114.34		52.00
Pro	14			177.95	63.71
Ala	15	8.82	127.56	177.67	52.81
Glu	16	8.88	124.69	173.92	54.60
Asn	17	8.76	123.56	179.11	53.76
Gly	18	8.69	109.25	170.18	44.99
Lys	19	7.80	121.95	174.40	53.76
Ser	20	8.27	120.51	170.59	59.33
Asn	21	8.84	127.41	171.50	50.87
Phe	22	10.36	120.83	169.85	57.61
Leu	23	8.84	127.32	168.55	52.86
Asn	24	8.34	122.46	171.33	50.57
Cys	25	9.54	120.44	167.35	53.63
Tyr	26	9.58	129.50	173.05	54.88
Val	27	8.81	127.40	171.87	58.61
Ser	28	8.75	117.50	172.54	56.63
Gly	29	8.19	107.88	171.36	45.80
Phe	30	7.24	111.46	172.23	53.54
His	31	8.50	123.62		57.71
Pro	32			177.27	62.80
Ser	33	8.47	111.94	171.78	61.19

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Asp	34	7.12	121.15	172.13	54.67
Ile	35	7.77	123.25	166.78	60.53
Glu	36	7.94	125.97	172.16	54.45
Val	37	7.71	125.16	170.64	60.33
Asp	38	8.74	123.09	172.36	52.13
Leu	39	8.90	121.21	172.92	53.12
Leu	40	8.84	120.49	174.59	52.56
Lys	41	8.74	121.38	176.76	53.86
Asn	42	9.69	128.68	175.05	54.05
Gly	43	8.75	102.77	169.72	44.96
Glu	44	7.69	121.13	175.31	53.89
Arg	45	8.61	124.88	176.51	57.22
Ile	46	8.74	129.27	175.99	61.52
Glu	47	8.40	126.59	178.44	58.14
Lys	48	7.97	119.79	172.72	55.37
Val	49	7.77	124.10	174.00	60.98
Glu	50	8.34	126.79	171.82	53.86
His	51	8.00	111.42	173.91	53.04
Ser	52a	9.10	116.71	171.25	57.10
Ser	52b	9.06	116.75	171.25	57.10
Asp	53	8.47	119.76	177.47	54.04
Leu	54	9.03	124.96	176.74	57.22
Ser	55	7.15	121.89		56.38
Phe	56			171.09	55.21
Ser	57a	9.27	116.83	174.65	56.61
Ser	57b	9.33	116.92	174.65	56.65
Lys	58	8.28	119.43	172.41	58.40
Asp	59	7.03	116.74	176.42	52.55
Trp	60a	8.06	112.68	174.06	59.04
Trp	60b	8.09	112.72	174.06	59.04
Ser	61a	8.09	115.78	169.12	58.81
Ser	61b	8.11	115.81	169.12	58.81
Phe	62	8.37	120.09	171.37	58.05
Tyr	63	7.78	109.92	172.92	56.32
Leu	64	9.37	120.64	169.22	55.02
Leu	65	8.32	121.60	173.79	53.38
Tyr	66	9.09	127.86	174.05	56.53
Tyr	67	8.84	116.56	170.90	55.75
Thr	68	8.22	112.29	166.70	59.60
Glu	69	8.41	129.82	175.27	56.50
Phe	70	8.51	125.54	166.98	54.87

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Thr	71a	8.17	117.72		58.13
Thr	71b	8.20	117.72		58.13
Pro	72			173.04	62.69
Thr	73a	7.89	110.34	174.24	59.56
Thr	73b	7.92	110.36	174.24	59.56
Glu	74	8.94	118.74	180.23	58.38
Lys	75	7.72	113.78	178.00	56.02
Asp	76	7.04	118.21	174.07	55.34
Glu	77	8.53	123.59	173.61	54.41
Tyr	78	9.39	123.98	174.68	56.34
Ala	79	8.73	121.45	170.94	50.61
Cys	80	8.99	120.03	174.28	52.43
Arg	81	9.26	128.79	171.59	53.38
Val	82	9.01	128.45	170.27	60.01
Asn	83	8.98	123.74	168.92	50.66
His	84	7.56	122.81	174.08	56.48
Val	85	7.9	125.16	174.61	64.11
Thr	86	7.39	110.82	173.12	62.79
Leu	87	7.96	123.49		53.02
Ser	88			171.88	59.83
Gln	89	7.37	117.58		52.79
Pro	90			175.68	63.07
Lys	91	8.69	125.59	173.06	54.55
Ile	92	8.42	125.36	175.59	59.90
Val	93	8.91	129.76	174.24	60.88
Lys	94	8.78	128.13	176.64	56.12
Trp	95	8.93	123.06	171.53	57.75
Asp	96	8.31	132.45	175.82	51.99
Arg	97	7.57	123.22	175.06	57.03
Asp	98	8.34	120.19	171.15	54.29
Met	99	7.83	122.72		56.54

Table 6.2 – Chemical shifts of the H-2D^b α-chain of H-2D^b/FL.
Residues affected by peak doubling in the ^{1}H-^{15}N dimension are listed twice, shaded blue. Resonances that could not be assigned (Pro ^{1}H and ^{15}N and Gly Cβ) are shaded grey. Missing assignments are shaded red.

Residue Type	Residue Number	H Shift	N Shift	Cα Shift	Cβ Shift
Met	0			56.57	32.68
Gly	1	8.60	105.97	45.86	
Pro	2			63.80	31.45
His	3	8.33	124.72	55.84	31.98
Ser	4	8.01	114.66	57.43	63.50
Met	5				
Arg	6				
Tyr	7				
Phe	8			55.24	
Glu	9a	10.01	126.90	55.41	
Glu	9b	10.06	126.94	55.41	
Thr	10a	9.08	119.29	61.88	71.39
Thr	10b	9.04	119.09	61.89	71.42
Ala	11a	9.23	126.64	49.99	22.96
Ala	11b	9.27	127.26	49.79	23.00
Val	12a	8.78	121.53	59.37	34.29
Val	12b	8.88	122.11	59.37	34.29
Ser	13	9.24	122.66	57.04	64.08
Arg	14				
Pro	15				
Gly	16				
Leu	17				
Glu	18				
Glu	19				
Pro	20			62.18	30.80
Arg	21	8.48	122.92	55.44	30.46
Tyr	22	9.24	131.15	55.72	41.59
Ile	23	8.50	126.62	58.66	42.10
Ser	24	9.10	121.83	55.76	63.49
Val	25	8.78	128.11	60.77	33.74
Gly	26	7.96	113.19	43.17	
Tyr	27a	9.58	122.68	57.25	38.20
Tyr	27b	9.52	122.68	57.25	38.20
Val	28	8.59	122.75	61.18	34.49
Asp	29				
Asn	30				
Lys	31			56.46	30.18

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Glu	32	7.91	116.38	54.90	28.13
Phe	33	9.00	130.24	56.84	37.73
Val	34	6.72	117.47	60.10	37.04
Arg	35	8.04	119.38	55.76	34.27
Phe	36	9.48	129.54	56.67	41.48
Asp	37	7.44	127.97	52.17	42.68
Ser	38a	9.53	123.62	60.95	62.58
Ser	38b	9.49	123.62	60.95	62.58
Asp	39	8.62	123.14	53.83	39.92
Ala	40	7.35	123.43	51.41	19.04
Glu	41	8.24	120.65	58.44	28.67
Asn	42	8.12	115.89	50.32	38.21
Pro	43			59.39	30.37
Arg	44	7.87	124.31	54.68	29.50
Tyr	45	8.14	120.96	61.20	38.60
Glu	46	8.65	119.81	51.85	31.41
Pro	47			61.74	31.90
Arg	48	9.30	122.50	60.48	31.44
Ala	49	7.22	117.89	54.56	16.84
Pro	50			64.70	33.16
Trp	51	6.63	113.43	55.61	
Met	52	7.04	118.58	53.62	29.78
Glu	53	7.64	119.46	58.24	28.15
Gln	54	7.40	114.16	56.08	28.63
Glu	55	7.36	119.84	55.64	28.16
Gly	56	8.89	111.08	44.53	
Pro	57			65.29	31.38
Glu	58	9.17	118.13	59.12	27.46
Tyr	59	7.70	122.72	61.23	37.57
Trp	60	6.87	120.46	58.56	28.60
Glu	61	8.51	121.47	59.07	29.05
Arg	62	7.66	120.82	58.94	28.67
Glu	63	8.13	118.14	58.62	28.03
Thr	64	8.52	118.19	68.19	
Gln	65				
Lys	66				
Ala	67			52.57	
Lys	68	8.12	113.36	58.58	
Gly	69	7.74	110.80	46.70	
Gln	70a	8.57	125.04	57.63	30.31
Gln	70b	8.66	125.59	59.12	30.31

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Glu	71	8.44	120.77	60.32	28.82
Gln	72a	6.98	116.62	58.16	27.00
Gln	72b	7.06	116.91	58.25	26.99
Trp	73	8.26	123.17	61.99	29.44
Phe	74				
Arg	75			60.14	29.72
Val	76	7.70	119.69	65.52	30.85
Ser	77	8.00	117.25	63.55	61.81
Leu	78a	8.37	121.78	58.19	40.95
Leu	78b	8.47	121.77	58.21	40.90
Arg	79	6.92	119.82	58.74	28.39
Asn	80	8.01	121.40	54.31	38.60
Leu	81				
Leu	82				
Gly	83				
Tyr	84			56.84	37.17
Tyr	85	7.68	113.07	58.87	37.92
Asn	86	7.79	119.74	54.09	36.53
Gln	87a	8.42	117.32	54.38	30.94
Gln	87b	8.49	117.42	54.36	30.86
Ser	88	8.34	115.26	57.10	63.95
Ala	89	8.99	122.97	52.57	24.08
Gly	90	8.41	107.86	45.33	
Gly	91	8.11	108.64	43.14	
Ser	92a	7.82	114.57	56.46	63.46
Ser	92b	7.93	113.46	56.53	62.94
His	93	9.62	118.29	53.98	32.54
Thr	94a	8.10	115.08	62.15	70.80
Thr	94b	8.21	115.32	62.15	70.82
Leu	95	9.65	131.14	53.48	46.63
Gln	96				
Gln	97				
Met	98				
Ser	99				
Gly	100				
Cys	101				
Asp	102			55.83	40.60
Leu	103	8.22	119.32	54.43	39.89
Gly	104	8.28	108.48	44.26	
Ser	105	7.50	110.47	60.12	63.73
Asp	106				

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Trp	107			58.80	24.67
Arg	108a	7.60	118.25	54.37	30.15
Arg	108b	7.57	118.09	54.06	30.21
Leu	109	8.52	121.49	56.66	41.48
Leu	110	9.23	129.36	55.38	39.86
Arg	111	7.02	115.16	55.23	32.48
Gly	112a	8.64	110.85	43.22	
Gly	112b	8.69	111.34	43.19	
Tyr	113	8.69	118.63	56.77	40.91
Leu	114	9.14	123.75	54.70	42.09
Gln	115	9.18	122.94	54.89	
Phe	116a	9.20	117.22	57.01	43.70
Phe	116b	9.11	117.18	56.89	43.26
Ala	117	8.82	126.20	50.54	18.82
Tyr	118	8.51	121.76	54.54	40.57
Glu	119	8.94	125.93	56.10	25.28
Gly	120a	7.18	108.61	45.53	
Gly	120b	7.23	108.72	45.53	
Arg	121			52.99	32.04
Asp	122	8.32	122.18	57.03	40.54
Tyr	123	8.77	123.10	59.82	40.99
Ile	124	7.34	115.04	59.56	41.44
Ala	125	8.49	127.18	51.08	22.80
Leu	126a	8.92	125.61	53.85	43.26
Leu	126b	8.98	126.04	53.94	43.32
Asn	127	8.75	125.92	52.63	37.49
Glu	128	8.71	119.36	58.66	28.65
Asp	129	7.27	116.43	53.97	39.27
Leu	130	8.42	116.25	55.43	37.95
Lys	131	8.25	113.88	56.50	36.51
Thr	132			62.16	68.60
Trp	133	8.40	123.07	55.42	29.78
Thr	134a	9.51	119.77	62.73	69.70
Thr	134b	9.58	119.93	62.82	69.65
Ala	135	8.86	131.86	49.84	19.70
Ala	136	8.17	125.00	54.06	19.57
Asp	137	7.21	115.52	52.22	41.06
Met	138	8.00	116.58	58.02	30.85
Ala	139	7.66	125.27	54.41	16.90
Ala	140	7.98	118.32	53.56	18.73
Gln	141	7.43	116.53	57.92	26.34

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Ile	142	7.55	123.50	64.97	35.96
Thr	143			66.55	67.97
Arg	144	8.41	119.63	60.08	29.16
Arg	145	7.63	117.93	59.12	30.40
Lys	146	7.49	119.13	59.36	30.98
Trp	147	9.26	123.52	59.33	27.00
Glu	148	9.20	125.31	59.61	28.85
Gln	149	7.73	118.49	57.75	27.96
Ser	150	7.89	111.15	58.24	63.67
Gly	151	7.71	108.42	46.76	
Ala	152a	8.00	121.83	54.98	18.82
Ala	152b	7.82	121.34	55.14	17.82
Ala	153a	9.56	117.99	56.21	16.93
Ala	153b	9.59	118.52	56.24	16.92
Glu	154a	7.94	116.17	59.79	28.71
Glu	154b	7.94	116.67	59.56	28.54
His	155	7.91	118.71	58.00	29.21
Tyr	156	8.37	118.13	56.84	38.79
Lys	157	8.38	121.28	53.76	29.22
Ala	158	9.08	123.84	55.21	16.55
Tyr	159				
Leu	160			56.78	42.12
Glu	161	9.14	122.89	55.60	29.75
Gly	162a	8.03	110.17	43.95	
Gly	162b	8.14	109.12	43.94	
Glu	163				
Cys	164				
Val	165				
Glu	166				
Trp	167				
Leu	168			53.36	42.93
His	169	8.46	122.47	57.53	29.73
Arg	170	8.01	118.89	53.37	28.79
Tyr	171a	9.25	120.79	58.85	36.77
Tyr	171b	9.31	120.95	58.68	37.77
Leu	172	7.89	121.30	54.39	40.98
Lys	173	7.53	121.33	57.97	33.19
Asn	174	7.00	117.52	57.68	40.37
Gly	175	6.85	102.57	46.22	
Asn	176	7.63	122.09	56.77	36.25
Ala	177				

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Thr	178				
Leu	179			56.35	39.38
Leu	180a	7.58	118.71	54.37	40.58
Leu	180b	7.61	119.32	54.34	40.46
Arg	181	8.56	121.98	56.55	29.83
Thr	182	8.18	117.21	60.37	71.43
Asp	183	9.53	124.17	53.57	43.18
Ser	184	8.68	119.54	56.58	61.81
Pro	185			62.33	31.43
Lys	186	9.00	123.36	55.60	30.91
Ala	187	8.73	129.91	50.41	24.02
His	188	8.32	116.69	55.91	31.95
Val	189	8.65	121.05	60.30	34.24
Thr	190	9.78	119.35	60.59	69.39
His	191	7.72	122.51	54.99	33.07
His	192	8.10	122.50	51.80	32.59
Pro	193				
Arg	194				
Ser	195			60.76	62.71
Lys	196a	8.76	120.61	57.88	33.01
Lys	196b	8.81	120.37	57.89	33.03
Gly	197	8.38	106.25	45.34	
Glu	198	7.48	120.61	54.73	33.70
Val	199	9.05	119.68	59.39	34.65
Thr	200	8.42	122.18	61.62	68.77
Leu	201	8.57	126.72	52.12	42.63
Arg	202	8.90	124.44	55.03	31.41
Cys	203			61.49	41.04
Trp	204	7.76	125.17	56.55	30.80
Ala	205	8.97	120.25	50.55	21.91
Leu	206	8.96	121.58	53.55	44.92
Gly	207	8.50	110.19	46.49	
Phe	208	7.30	111.55	53.62	41.13
Tyr	209	9.09	120.55	58.66	41.62
Pro	210			61.87	34.25
Ala	211	7.90	118.58	54.04	16.86
Asp	212	7.03	116.14	55.20	40.23
Ile	213	7.51	122.56	60.48	39.41
Thr	214	8.10	120.92	61.92	70.70
Leu	215	8.51	128.40	53.47	44.30
Thr	216	8.47	114.32	59.61	71.10

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Trp	217	7.88	122.50	56.51	33.10
Gln	218	8.02	123.10	53.93	32.03
Leu	219	8.45	121.96	53.46	42.98
Asn	220	9.00	126.80	54.21	36.74
Gly	221	8.42	103.74	45.05	
Glu	222	7.67	121.00	54.35	30.78
Glu	223	8.64	123.57	56.99	29.22
Leu	224	8.60	126.36	54.04	40.97
Thr	225	8.10	112.93	63.07	69.10
Gln	226	8.10	120.14	55.85	29.19
Asp	227	8.16	117.79	54.88	40.49
Met	228	7.64	119.46	54.94	34.85
Glu	229	8.76	125.11	55.20	30.38
Leu	230	8.46	124.68	53.84	44.42
Val	231	6.43	114.71	58.45	32.03
Glu	232	8.19	123.43	56.28	28.66
Thr	233	8.51	122.14	64.22	68.09
Arg	234	8.74	123.79	50.68	29.74
Pro	235			61.17	30.91
Ala	236	8.89	125.28	53.28	19.09
Gly	237	8.38	105.55	43.78	
Asp	238	7.14	117.82	51.18	40.08
Gly	239	8.17	109.36	44.24	
Thr	240	7.17	119.43	56.47	
Phe	241			57.53	41.51
Gln	242	8.50	111.32	52.83	32.97
Lys	243	9.34	120.37	57.03	37.70
Trp	244	6.11	116.62	54.86	33.14
Ala	245	8.67	118.95	50.68	23.33
Ser	246	10.91	125.05	56.49	67.33
Val	247	8.01	114.27	59.16	34.48
Val	248	7.70	123.96	61.22	30.93
Val	249	8.88	122.88	56.54	32.02
Pro	250				
Leu	251			56.54	40.97
Gly	252	8.96	117.37	45.13	
Lys	253	7.87	120.05	54.41	32.03
Glu	254	10.81	121.56	61.55	29.16
Gln	255	9.04	114.47	57.65	27.00
Asn	256a	8.07	117.53	54.00	37.69
Asn	256b	8.12	117.76	53.42	37.79

Residue Type	Residue Number	H Shift	N Shift	CO Shift	Cα Shift
Tyr	257	7.70	121.31	57.53	39.83
Thr	258	9.04	113.62	59.30	71.93
Cys	259	8.99	125.51	51.35	40.60
Arg	260	8.73	127.02	53.07	31.51
Val	261	7.78	121.85	60.09	33.14
Tyr	262	8.70	123.97	55.65	38.63
His	263	7.87	122.15	57.47	35.46
Glu	264	8.27	126.62	58.20	29.17
Gly	265	10.47	111.86	44.62	
Leu	266	7.76	121.46	51.06	41.84
Pro	267				
Glu	268				
Pro	269			62.63	30.01
Leu	270	8.05	122.80	53.78	43.10
Thr	271	8.26	117.74	61.63	69.23
Leu	272	9.38	127.63	54.12	44.39
Arg	273	8.21	111.71	63.10	
Trp	274				
Glu	275				
Pro	276				
Met	277				

Appendix D

Residual dipolar couplings for H-2D^b/FL

Table 6.3 – RDC values for the hβ$_2$m component of H-2D^b/FL.

Residue No.	RDC value (Hz)	
	Polyacrylamide Gel	Pf1 Phage
6	14.6 ± 0.7	6.5 ± 0.4
7		-3.1 ± 0.2
8	17.1 ± 0.7	7.8 ± 0.4
9		3.7 ± 0.2
10	21.8 ± 0.8	9.8 ± 0.3
11	14.7 ± 0.6	11.2 ± 0.5
12	-3.6 ± 0.6	7.9 ± 0.4
13	5.0 ± 0.7	11.6 ± 0.5
16	15.3 ± 0.8	11.7 ± 0.4
18		-10.8 ± 0.3
19	-21.3 ± 0.3	
20	-25.1 ± 0.2	-19.5 ± 0.1
22	2.0 ± 0.3	11.1 ± 0.3
24	16.5 ± 0.3	12.1 ± 0.3
25	9.5 ± 0.8	4.3 ± 0.3
26	25.8 ± 0.9	12.8 ± 0.4
28	12.0 ± 0.6	4.1 ± 0.5
29		-2.4 ± 0.4
30	-19.8 ± 0.8	-18.2 ± 0.5
33		11.1 ± 0.8
34		-14.1 ± 0.3
35	-12.3 ± 0.2	-8.6 ± 0.2
36	-13.1 ± 0.9	-9.5 ± 0.4
37		1.6 ± 0.2
38		-10.6 ± 0.2
39	-5.2 ± 0.3	2.4 ± 0.2
42	3.3 ± 0.3	10.9 ± 0.3
43	1.7 ± 0.3	9.9 ± 0.4
45	-6.1 ± 0.2	±
46	-9.5 ± 0.4	-2.3 ± 0.2
47	18.6 ± 0.2	0.8 ± 0.2
49	10.2 ± 0.2	6.4 ± 0.2
51	5.1 ± 0.2	3.7 ± 0.2
53	-14.5 ± 0.4	-4.3 ± 0.3
54	4.1 ± 0.7	

Residue No.	RDC value (Hz)	
	Polyacrylamide Gel	Pf1 Phage
55	19.6 ± 1.3	
58	5.6 ± 0.6	-2.6 ± 0.6
59	12.1 ± 1.0	
63	17.4 ± 0.5	6.9 ± 0.4
64	10.1 ± 0.7	4.2 ± 0.4
65	4.2 ± 0.6	
67	18.8 ± 0.6	11.5 ± 0.3
68	10.2 ± 0.5	11.3 ± 0.4
69	8.7 ± 0.2	11.8 ± 0.2
70	-2.4 ± 0.3	6.7 ± 0.2
74	-16.6 ± 0.5	-16.5 ± 0.4
75	-17.2 ± 0.6	-16.0 ± 0.3
76		-9.2 ± 0.3
77		11.5 ± 0.2
78	4.8 ± 0.3	11.8 ± 0.3
80		0.3 ± 0.5
81	-10.0 ± 0.5	-9.0 ± 0.3
82	6.7 ± 0.6	2.4 ± 0.4
83	-13.9 ± 0.8	-9.5 ± 0.4
84	-3.7 ± 0.3	-4.4 ± 0.2
85	4.4 ± 0.4	
87		1.3 ± 0.3
89	-2.2 ± 0.1	10.5 ± 0.2
91	15.3 ± 0.6	6.8 ± 0.4
92	-13.4 ± 0.3	-9.0 ± 0.2
93	1.9 ± 0.7	0.8 ± 0.4
95	-1.5 ± 0.3	7.3 ± 0.3
97		8.0 ± 0.2
98	-7.1 ± 0.5	

Table 6.4 – RDC values for the H-2D^b α-chain of H-2D^b/FL.

Residue No.	RDC value (Hz)	
	Polyacrylamide Gel	Pf1 Phage
3	9.6 ± 0.4	12.1 ± 0.1
4	-0.7 ± 0.6	3.2 ± 0.3
13	-11.5 ± 1.2	-5.5 ± 0.3
21	-10.9 ± 1.3	-8.5 ± 0.3
22	-15.1 ± 1.6	-5.9 ± 0.3
23	-27.9 ± 2.1	
24	-9.3 ± 3.0	
26	-13.3 ± 3.5	-12.1 ± 1.1
28		4.9 ± 1.1
33		10.9 ± 0.7
34	-19.3 ± 5.4	
36	-14.6 ± 2.6	-8.3 ± 0.6
37		-14.5 ± 0.2
39	4.5 ± 0.8	10.1 ± 0.5
40	7.3 ± 0.7	6.6 ± 0.2
41		5.2 ± 0.1
46	-11.8 ± 0.6	-20.4 ± 0.3
52	14.3 ± 1.6	8.9 ± 1.2
55	19.8 ± 4.8	
56	8.3 ± 0.6	7.9 ± 0.3
58		-12.8 ± 0.7
59	-14.4 ± 1.4	-13.7 ± 0.2
60	-14.7 ± 2.6	
63	-3.6 ± 1.2	
64		-21.0 ± 2.1
69		-15.8 ± 0.6
73	1.0 ± 0.1	-16.9 ± 1.2
77		-8.1 ± 0.5
79	-13.1 ± 1.0	
80	-11.4 ± 1.8	-14.5 ± 0.9
85		-18.4 ± 1.0
86	18.5 ± 1.5	
88		9.6 ± 0.3
90		-7.3 ± 0.8
91		5.5 ± 0.4
93	-15.2 ± 1.8	-18.0 ± 0.3
95	-17.4 ± 3.3	-9.8 ± 0.6
113	-20.2 ± 1.3	
114	-18.3 ± 1.8	-9.7 ± 0.4
115	-15.6 ± 1.9	-17.3 ± 0.6

Residue No.	RDC value (Hz)	
	Polyacrylamide Gel	Pf1 Phage
117	-16.7 ± 2.5	-12.2 ± 0.6
119		-5.1 ± 0.5
122	5.5 ± 0.3	
124	-11.4 ± 4.7	-12.7 ± 1.0
125	-16.5 ± 1.6	
127	9.4 ± 1.4	
129	-17.4 ± 1.4	-15.2 ± 0.4
130	27.8 ± 2.9	8.2 ± 0.4
131		8.0 ± 2.7
133	-8.7 ± 1.8	-6.1 ± 0.5
135		-7.8 ± 0.3
136	-18.5 ± 0.9	-10.5 ± 0.2
138	15.5 ± 1.8	5.7 ± 0.4
139	2.6 ± 0.2	10.7 ± 0.4
140	-17.6 ± 0.7	-1.8 ± 0.3
141	8.7 ± 2.1	
142	15.2 ± 3.4	11.5 ± 0.9
144	15.3 ± 2.5	4.8 ± 0.5
145		3.3 ± 0.5
146	19.7 ± 2.8	
147	25.9 ± 7.7	
148		-6.5 ± 0.5
149		2.1 ± 0.6
150		9.3 ± 0.7
158		-13.9 ± 0.8
161	-17.2 ± 2.0	-7.5 ± 0.6
172	-7.6 ± 1.1	-15.6 ± 0.6
174	-24.5 ± 2.4	-14.4 ± 0.6
175	-15.7 ± 2.4	-14.5 ± 0.7
181	0.4 ± 0.7	8.6 ± 0.5
182	-22.2 ± 1.3	-17.6 ± 0.5
183	-27.1 ± 5.3	-20.2 ± 1.3
184		-21.3 ± 0.3
186	-7.4 ± 1.3	5.9 ± 0.6
187	-8.8 ± 0.7	-4.1 ± 0.3
188	-8.4 ± 2.2	-11.8 ± 0.6
189	0.4 ± 0.7	-8.8 ± 0.2
190	7.6 ± 1.4	-9.9 ± 0.4
191	-25.9 ± 2.5	
192		-0.7 ± 0.2
197	-6.6 ± 1.1	

Residue No.	RDC value (Hz)	
	Polyacrylamide Gel	Pf1 Phage
198	2.5 ± 0.5	9.7 ± 0.2
199	-8.4 ± 1.0	7.3 ± 0.3
201		-14.2 ± 0.3
202	6.2 ± 1.2	-9.8 ± 0.4
204		-8.3 ± 0.2
205		-9.4 ± 0.4
206	7.2 ± 1.0	-2.0 ± 0.4
207	-7.3 ± 1.7	-0.2 ± 0.5
209	-3.0 ± 1.2	-1.0 ± 0.4
212	19.8 ± 1.0	
213	-13.2 ± 0.4	-18.3 ± 0.2
214		-15.0 ± 0.3
215	-6.8 ± 0.5	-11.5 ± 0.2
216	-16.6 ± 1.5	-17.5 ± 0.4
217	-6.9 ± 0.4	
218	-3.0 ± 0.4	-3.0 ± 0.3
220	-5.1 ± 0.8	-3.0 ± 0.5
221	-5.3 ± 0.9	
222	11.0 ± 0.5	
223	7.0 ± 0.3	6.3 ± 0.1
225	4.1 ± 0.4	5.0 ± 0.3
226	2.1 ± 0.4	3.8 ± 0.3
229	6.8 ± 1.1	
230	-14.3 ± 0.6	-17.1 ± 0.2
231	-6.9 ± 1.0	-9.6 ± 0.6
232		-4.5 ± 0.2
236	5.5 ± 1.3	8.0 ± 0.4
237		4.1 ± 1.2
238		7.7 ± 1.0
239		-6.1 ± 1.7
243	3.6 ± 1.2	-10.2 ± 0.3
244	-2.4 ± 0.9	-10.9 ± 0.3
245	-13.4 ± 2.9	
246	4.3 ± 2.0	-10.8 ± 0.4
247	-7.7 ± 1.4	-12.3 ± 0.4
248	-3.2 ± 0.7	-8.9 ± 0.2
249	-7.3 ± 1.1	-1.2 ± 0.3
252	2.4 ± 1.0	0.6 ± 0.4
254	13.6 ± 3.8	
255	30.0 ± 1.9	10.6 ± 0.4
258	-13.9 ± 1.3	-16.0 ± 0.3

Residue No.	RDC value (Hz)	
	Polyacrylamide Gel	Pf1 Phage
259	-9.8 ± 1.1	-15.7 ± 0.4
260	-16.3 ± 1.0	-18.2 ± 0.3
262	-12.4 ± 0.9	-8.0 ± 0.3
264		-3.5 ± 0.5
266	-20.6 ± 2.1	
270	6.2 ± 0.7	
271	-18.9 ± 0.4	-17.5 ± 0.2
272	-8.3 ± 0.9	-17.0 ± 0.4

Appendix E

Heteronuclear relaxation data for isolated β_2m and H-2D^b/FL

Table 6.5 – Heteronuclear relaxation data for isolated hβ_2m at 700 MHz.

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
6	868 ± 17	62.5 ± 1.3	0.873 ± 0.219
7	848 ± 25	52.0 ± 5.2	0.805 ± 0.121
8	859 ± 3	62.8 ± 0.7	0.836 ± 0.048
9	902 ± 5	61.7 ± 2.7	0.826 ± 0.091
10	816 ± 8	51.0 ± 0.9	0.870 ± 0.089
11	824 ± 6	55.4 ± 0.4	0.841 ± 0.101
15	883 ± 5	67.5 ± 0.4	0.820 ± 0.149
16	944 ± 10	60.7 ± 1.1	0.731 ± 0.081
17	837 ± 20	66.5 ± 1.2	0.828 ± 0.054
18	924 ± 6	65.8 ± 2.5	0.801 ± 0.083
19	816 ± 5	64.8 ± 0.4	0.822 ± 0.054
20	904 ± 10	58.5 ± 0.4	0.797 ± 0.040
21	856 ± 6	58.6 ± 1.2	0.857 ± 0.064
22	878 ± 11	63.3 ± 0.6	0.837 ± 0.109
23	863 ± 7	65.2 ± 2.8	0.798 ± 0.062
24	874 ± 4	62.7 ± 0.3	0.832 ± 0.079
25	885 ± 6	65.4 ± 0.4	0.832 ± 0.072
26	853 ± 8	56.2 ± 0.6	0.832 ± 0.130
27	813 ± 11	55.5 ± 0.9	0.826 ± 0.086
35	892 ± 2	65.8 ± 0.9	0.805 ± 0.057
36	873 ± 7	63.0 ± 13.5	0.802 ± 0.087
37	912 ± 5	68.4 ± 4.3	0.811 ± 0.049
38	876 ± 6	66.1 ± 0.8	0.805 ± 0.062
39	906 ± 4	67.0 ± 0.5	0.836 ± 0.071
40	870 ± 6	62.1 ± 0.5	0.823 ± 0.074
41	898 ± 4	64.0 ± 0.5	0.816 ± 0.070
42	824 ± 3	62.0 ± 0.7	0.839 ± 0.074
43	819 ± 7	62.5 ± 0.9	0.852 ± 0.064
44	912 ± 5	58.9 ± 0.3	0.825 ± 0.045
45	849 ± 15	69.0 ± 2.2	0.803 ± 0.051
46	897 ± 4	59.1 ± 1.2	0.835 ± 0.070
47	968 ± 8	64.2 ± 2.1	0.788 ± 0.067
48	973 ± 11	68.9 ± 0.3	0.672 ± 0.044
49	978 ± 5	71.1 ± 1.0	0.602 ± 0.042
50	899 ± 6	69.0 ± 3.9	0.759 ± 0.063
51	867 ± 3	52.1 ± 0.4	0.811 ± 0.043

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
52	824 ± 5	62.0 ± 0.3	0.845 ± 0.092
53	861 ± 27	54.3 ± 0.8	0.777 ± 0.085
54	880 ± 18	65.5 ± 3.4	0.714 ± 0.084
66	877 ± 7	66.5 ± 0.9	0.792 ± 0.034
67	863 ± 5	59.8 ± 0.4	0.877 ± 0.066
68	857 ± 5	67.1 ± 0.3	0.833 ± 0.073
69	887 ± 5	64.5 ± 0.8	0.790 ± 0.057
70	847 ± 8	65.9 ± 1.4	0.785 ± 0.062
71	881 ± 10	66.6 ± 0.7	0.782 ± 0.066
73	977 ± 12	65.8 ± 1.8	0.821 ± 0.056
74	935 ± 33	64.9 ± 0.6	0.773 ± 0.071
75	849 ± 12	82.9 ± 0.6	0.648 ± 0.045
76	918 ± 6	68.6 ± 0.8	0.768 ± 0.043
77	905 ± 6	66.9 ± 0.6	0.790 ± 0.062
78	898 ± 7	69.6 ± 1.0	0.830 ± 0.076
79	882 ± 4	65.3 ± 0.5	0.830 ± 0.059
80	870 ± 5	64.6 ± 0.5	0.844 ± 0.064
81	840 ± 3	64.6 ± 1.4	0.854 ± 0.080
82	862 ± 5	66.0 ± 0.8	0.810 ± 0.069
83	846 ± 7	67.3 ± 0.5	0.832 ± 0.073
84	906 ± 4	67.0 ± 0.9	0.825 ± 0.051
85	854 ± 7	61.6 ± 2.7	0.824 ± 0.068
86	819 ± 13	57.7 ± 16.7	0.787 ± 0.132
89	889 ± 4	65.9 ± 0.3	0.757 ± 0.040
91	878 ± 7	66.9 ± 2.8	0.798 ± 0.059
92	909 ± 5	72.2 ± 2.5	0.768 ± 0.056
93	853 ± 5	67.1 ± 0.2	0.827 ± 0.074
94	886 ± 6	66.7 ± 2.1	0.819 ± 0.052
95	883 ± 5	62.6 ± 0.9	0.821 ± 0.088
96	962 ± 14	62.9 ± 1.8	0.796 ± 0.132
97	851 ± 5	63.3 ± 0.4	0.641 ± 0.087
98	729 ± 3	88.7 ± 0.6	0.579 ± 0.035
99	818 ± 2	147.3 ± 11.2	0.314 ± 0.026

Table 6.6 – Heteronuclear relaxation data for hβ$_2$m bound by H-2D^b/FL at 950 MHz.

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)
2	960 ± 117	41.3 ± 5.2
3	1461 ± 210	24.1 ± 4.3
4	2156 ± 402	15.3 ± 2.1
6	2725 ± 443	17.2 ± 3.4
7	452 ± 132	76.8 ± 11.9
8	2919 ± 695	19.0 ± 2.9
9	2889 ± 399	20.2 ± 3.6
10	2720 ± 467	10.9 ± 0.5
11	2598 ± 486	17.4 ± 1.1
12	2369 ± 446	19.5 ± 2.6
13	2568 ± 341	18.4 ± 1.6
16	2895 ± 537	22.4 ± 3.2
17	2606 ± 527	18.5 ± 2.4
18	2358 ± 433	19.0 ± 3.6
19	2148 ± 316	19.3 ± 3.1
20	2376 ± 424	19.8 ± 2.5
22	2725 ± 526	20.1 ± 3.2
24	2411 ± 377	27.9 ± 5.3
25	2667 ± 602	19.8 ± 3.7
28	2175 ± 359	15.9 ± 2.4
29	2504 ± 363	21.5 ± 5.1
30	2126 ± 347	17.9 ± 2.3
31	2395 ± 532	14.2 ± 1.7
34	2314 ± 447	17.3 ± 1.7
35	1929 ± 300	58.2 ± 12.1
36	2494 ± 554	24.8 ± 3.2
37	3091 ± 469	16.6 ± 2.7
38	2638 ± 449	16.8 ± 2.3
39	2804 ± 476	20.4 ± 3.2
40	2622 ± 435	20.7 ± 3.1
42	2418 ± 429	17.1 ± 4.7
43	2020 ± 447	18.1 ± 4.6
44	2434 ± 483	18.0 ± 2.5
45	2681 ± 409	21.1 ± 3.3
46	2295 ± 460	15.8 ± 2.9
47	2579 ± 435	18.1 ± 3.1
48	2288 ± 358	22.7 ± 3.7
49	2005 ± 259	27.0 ± 4.2
50	2539 ± 405	17.8 ± 2.8
51	2795 ± 481	20.7 ± 2.6
52	2545 ± 442	17.1 ± 3.3

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)
53	2097 ± 380	19.9 ± 2.7
54	2680 ± 519	21.9 ± 5.4
55	2489 ± 265	14.1 ± 1.8
57	1998 ± 575	19.1 ± 4.5
58	2090 ± 290	14.3 ± 1.9
59	2281 ± 362	14.9 ± 2.4
61	2378 ± 289	20.4 ± 3.3
62	1732 ± 462	16.8 ± 2.6
63	2653 ± 673	16.4 ± 2.1
64	2634 ± 697	18.5 ± 2.9
65	2267 ± 324	27.2 ± 4.6
66	2347 ± 435	15.3 ± 3.2
67	2648 ± 550	17.8 ± 2.5
68	2401 ± 524	18.9 ± 2.2
69	2696 ± 471	15.8 ± 3.0
70	2286 ± 374	16.0 ± 2.7
71	1970 ± 354	24.7 ± 3.1
73	2653 ± 379	21.6 ± 4.1
74	2195 ± 315	19.0 ± 2.7
75	1820 ± 241	26.4 ± 3.6
76	2412 ± 339	21.0 ± 3.3
77	2573 ± 358	19.4 ± 2.8
78	3043 ± 481	20.4 ± 3.4
80	2665 ± 402	18.5 ± 2.2
81	2395 ± 496	19.5 ± 4.0
82	2259 ± 527	16.0 ± 2.8
83	2549 ± 524	19.6 ± 3.2
84	2752 ± 506	21.5 ± 3.2
85	2200 ± 482	15.5 ± 4.7
87	2382 ± 454	24.5 ± 7.3
89	2385 ± 358	21.1 ± 2.7
91	2874 ± 459	15.9 ± 2.8
92	2831 ± 326	20.1 ± 3.8
93	2611 ± 415	18.3 ± 3.0
94	2911 ± 334	18.6 ± 3.1
95	2336 ± 379	18.2 ± 2.5
96	2742 ± 486	16.6 ± 2.3
97	2419 ± 524	21.1 ± 2.8
98	2245 ± 349	17.5 ± 2.1
99	2044 ± 286	20.1 ± 4.3

Table 6.7 – Heteronuclear relaxation data for hβ$_2$m bound by H-2D^b/FL at 700 MHz.

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
2	968 ± 41	55.2 ± 2.4	0.189 ± 0.063
3	1218 ± 87	29.6 ± 0.7	0.876 ± 0.104
4	1412 ± 98	33.6 ± 0.6	0.751 ± 0.127
6	1809 ± 124	23.3 ± 1.0	0.911 ± 0.127
7	1670 ± 149	38.2 ± 2.2	0.819 ± 0.157
8	1923 ± 147	21.7 ± 1.0	0.943 ± 0.128
9	2319 ± 106	24.9 ± 0.4	0.962 ± 0.111
10	1839 ± 131	22.7 ± 0.7	0.902 ± 0.106
11	2103 ± 86	22.5 ± 0.2	0.857 ± 0.146
12	1736 ± 127	25.2 ± 0.9	0.922 ± 0.127
13	1735 ± 120	28.1 ± 0.5	0.851 ± 0.141
15	1865 ± 104	25.8 ± 0.7	0.875 ± 0.052
16	1994 ± 172	24.4 ± 1.4	0.837 ± 0.156
17	1932 ± 143	24.0 ± 0.7	0.870 ± 0.103
18	1826 ± 105	25.9 ± 0.8	0.844 ± 0.135
19	1488 ± 114	31.9 ± 0.8	0.782 ± 0.077
20	1743 ± 140	32.8 ± 0.8	0.721 ± 0.071
21	1789 ± 112	26.0 ± 0.4	0.902 ± 0.047
22	1851 ± 133	25.1 ± 0.6	0.904 ± 0.116
23	1761 ± 115	26.0 ± 0.4	0.902 ± 0.047
24	1860 ± 138	35.3 ± 2.2	0.763 ± 0.069
25	2001 ± 137	23.7 ± 0.6	0.965 ± 0.127
26	1983 ± 126	22.5 ± 0.3	0.928 ± 0.142
27	1865 ± 104	25.8 ± 0.6	0.887 ± 0.132
28	1799 ± 120	22.0 ± 0.8	0.878 ± 0.159
29	1720 ± 116	25.6 ± 0.2	0.829 ± 0.126
30	1676 ± 88	23.8 ± 0.7	0.931 ± 0.117
31	1848 ± 122	23.4 ± 0.7	0.911 ± 0.101
33	2136 ± 130	23.5 ± 1.8	0.851 ± 0.247
34	1860 ± 92	23.8 ± 0.5	0.815 ± 0.120
35	2075 ± 111	28.5 ± 1.1	0.816 ± 0.085
36	2110 ± 109	31.4 ± 1.0	0.838 ± 0.112
37	2038 ± 93	27.3 ± 1.5	0.835 ± 0.063
38	1994 ± 115	25.1 ± 0.8	0.939 ± 0.076
39	2113 ± 129	25.9 ± 0.5	0.881 ± 0.106
40	1916 ± 109	24.5 ± 0.4	0.908 ± 0.072
41	1974 ± 133	25.1 ± 0.5	0.947 ± 0.082
42	1871 ± 90	25.5 ± 0.4	0.892 ± 0.108
43	1799 ± 111	25.3 ± 0.4	0.886 ± 0.127
44	1740 ± 122	30.6 ± 0.9	0.776 ± 0.075
45	1825 ± 154	25.5 ± 0.6	0.797 ± 0.085

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
46	1796 ± 100	26.1 ± 1.0	0.851 ± 0.103
47	1877 ± 124	25.2 ± 0.4	0.700 ± 0.087
48	1668 ± 125	34.2 ± 1.0	0.502 ± 0.064
49	1721 ± 84	33.4 ± 1.2	0.356 ± 0.079
50	1865 ± 118	27.2 ± 0.4	0.679 ± 0.118
51	2009 ± 85	24.4 ± 0.6	0.828 ± 0.096
52	1785 ± 114	22.6 ± 0.2	0.904 ± 0.150
53	1678 ± 106	24.9 ± 0.4	0.787 ± 0.110
54	1693 ± 148	22.8 ± 0.8	0.769 ± 0.248
55	1769 ± 98	23.1 ± 0.4	0.888 ± 0.232
57	1660 ± 162	23.4 ± 0.5	0.918 ± 0.325
58	1663 ± 133	30.4 ± 1.1	0.647 ± 0.137
59	1717 ± 118	29.2 ± 1.5	0.663 ± 0.127
60	1690 ± 108	23.6 ± 1.0	0.421 ± 0.193
61	1817 ± 102	28.0 ± 0.4	0.423 ± 0.062
62	1622 ± 103	27.9 ± 0.6	0.914 ± 0.184
63	1864 ± 128	22.3 ± 0.5	0.942 ± 0.184
64	1786 ± 129	23.0 ± 0.5	0.949 ± 0.133
65	1693 ± 132	43.2 ± 1.5	0.728 ± 0.135
66	1761 ± 151	24.1 ± 0.4	0.954 ± 0.135
67	1996 ± 132	22.0 ± 0.5	0.950 ± 0.122
68	1780 ± 135	24.8 ± 0.6	0.888 ± 0.104
69	1999 ± 115	24.2 ± 0.8	0.785 ± 0.103
70	1736 ± 107	26.0 ± 0.3	0.830 ± 0.107
71	1579 ± 98	31.9 ± 0.3	0.658 ± 0.052
73	1793 ± 139	28.4 ± 0.7	0.763 ± 0.104
74	1791 ± 112	25.3 ± 0.4	0.751 ± 0.125
75	1465 ± 87	31.8 ± 0.2	0.513 ± 0.116
76	1730 ± 109	28.8 ± 1.6	0.548 ± 0.095
77	1899 ± 110	25.8 ± 0.6	0.861 ± 0.082
78	2224 ± 119	25.2 ± 0.4	0.919 ± 0.134
79	2005 ± 130	25.0 ± 0.5	0.932 ± 0.061
80	2176 ± 133	24.4 ± 0.6	0.993 ± 0.115
81	2022 ± 96	24.5 ± 0.7	0.962 ± 0.108
82	2166 ± 129	24.6 ± 0.3	0.955 ± 0.118
83	2106 ± 101	22.8 ± 0.8	0.957 ± 0.115
84	2002 ± 116	28.3 ± 1.0	0.884 ± 0.104
85	1737 ± 143	31.8 ± 2.5	0.651 ± 0.108
86	1647 ± 91	22.1 ± 1.2	0.796 ± 0.255
87	1478 ± 103	37.8 ± 2.8	0.768 ± 0.108
89	1726 ± 103	28.2 ± 0.7	0.716 ± 0.079
91	2095 ± 120	23.6 ± 0.5	0.900 ± 0.112

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
92	2255 ± 85	28.0 ± 0.9	0.784 ± 0.081
93	2203 ± 100	25.5 ± 0.7	0.906 ± 0.124
94	2066 ± 109	24.9 ± 0.5	0.892 ± 0.071
95	1885 ± 92	25.9 ± 0.5	0.924 ± 0.105
96	2242 ± 110	21.4 ± 0.4	0.831 ± 0.108
97	1789 ± 123	27.1 ± 0.6	0.850 ± 0.110
98	1718 ± 94	30.6 ± 0.7	0.792 ± 0.081
99	1529 ± 105	32.6 ± 1.4	0.674 ± 0.121

Table 6.8 – Heteronuclear relaxation data for H-2D^b α-chain at 950 MHz.

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)
1	2636 ± 347	
3	1522 ± 171	30.2 ± 6.3
4	1101 ± 106	41.0 ± 7.8
9	3104 ± 834	32.0 ± 14.4
10	2037 ± 489	32.8 ± 8.1
10	2471 ± 251	17.9 ± 5.1
11	2813 ± 636	
12	2770 ± 534	14.4 ± 1.7
13	2171 ± 267	14.5 ± 2.6
21	2200 ± 403	20.4 ± 4.2
22	2538 ± 289	20.3 ± 3.6
23	2485 ± 393	17.9 ± 3.7
24	2514 ± 793	15.1 ± 5.7
27		33.6 ± 13.7
27	2286 ± 513	
28	2472 ± 257	
32	2691 ± 904	
33	1719 ± 862	
36	3077 ± 803	13.8 ± 2.8
37	2771 ± 450	19.7 ± 3.2
38	2422 ± 312	19.2 ± 2.8
38	1445 ± 205	45.0 ± 12.1
39	1600 ± 196	27.8 ± 4.9
40	2387 ± 361	21.2 ± 3.6
44	1118 ± 174	
45	2522 ± 575	
48	2402 ± 272	18.9 ± 7.0
49	2595 ± 428	27.7 ± 6.1
51	2554 ± 435	15.5 ± 2.3
52	2169 ± 465	

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)
53	1662 ± 228	
54	1821 ± 369	20.1 ± 5.5
55	2719 ± 367	15.7 ± 3.3
56	2325 ± 282	24.7 ± 4.7
58	2460 ± 397	20.5 ± 3.9
60	2515 ± 487	17.6 ± 3.6
62	2456 ± 418	
63	2781 ± 553	
64	2298 ± 357	14.7 ± 2.7
69	2152 ± 341	25.1 ± 5.3
70	2405 ± 403	20.9 ± 6.1
70	2239 ± 410	
72	2772 ± 428	16.9 ± 2.6
76	1803 ± 365	
78	2231 ± 467	
78	2629 ± 280	
79	2534 ± 383	17.4 ± 2.3
85	3239 ± 672	
86	2207 ± 332	
87	2461 ± 374	19.0 ± 2.2
87	2320 ± 497	22.4 ± 4.6
88	2387 ± 394	22.0 ± 3.5
89	1957 ± 264	18.8 ± 3.2
90	1145 ± 159	42.0 ± 11.2
92	2437 ± 496	27.5 ± 4.8
93	1611 ± 224	
94	2566 ± 364	23.7 ± 2.8
95	3093 ± 659	
96	2293 ± 409	16.5 ± 7.1
104	1445 ± 195	
108	2746 ± 395	15.1 ± 3.0
112	2732 ± 390	19.6 ± 3.1
113	2454 ± 485	20.8 ± 3.7
114	2435 ± 453	22.1 ± 6.1
115	2453 ± 391	17.3 ± 3.9
116	2770 ± 568	22.1 ± 2.7
116	3082 ± 568	
117	2829 ± 328	
118	2491 ± 378	
119	2462 ± 350	
120	3525 ± 859	22.5 ± 3.4
120		24.7 ± 16.7

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)
123	2549 ± 692	
124	2542 ± 508	21.0 ± 4.2
125	2602 ± 224	18.5 ± 3.0
126	2163 ± 464	17.4 ± 4.1
126	2770 ± 405	
127	2435 ± 405	15.8 ± 3.7
128	2183 ± 336	
129	2403 ± 380	17.0 ± 2.1
130	2544 ± 326	17.1 ± 2.3
133	3652 ± 887	24.8 ± 6.9
134	2952 ± 536	19.8 ± 5.1
134	2263 ± 678	19.5 ± 6.7
135	2644 ± 388	21.2 ± 4.9
136	1806 ± 228	25.8 ± 7.8
137	1767 ± 222	25.7 ± 3.8
138	2923 ± 987	10.9 ± 2.5
139	2238 ± 231	
141	2818 ± 582	16.4 ± 1.9
142	2904 ± 467	
144	2710 ± 595	
145	2621 ± 300	15.7 ± 2.1
146	2599 ± 534	16.3 ± 2.4
147	2216 ± 426	
148	2725 ± 443	17.3 ± 3.9
149	2120 ± 356	13.8 ± 2.2
150	2508 ± 442	17.8 ± 4.6
151	2413 ± 424	22.1 ± 5.5
152	1135 ± 304	26.8 ± 10.8
152	1568 ± 272	
153	1733 ± 335	
153	2938 ± 637	
154	3370 ± 558	
154	2889 ± 617	13.0 ± 1.7
156	1249 ± 584	
158	3078 ± 649	10.4 ± 3.9
161	2354 ± 468	13.7 ± 3.4
162	870 ± 97	
169	1593 ± 126	28.7 ± 4.3
171	3120 ± 507	16.9 ± 2.5
171	1699 ± 535	22.1 ± 6.0
172	1429 ± 275	42.8 ± 8.7
174	2317 ± 329	17.3 ± 3.0

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)
175	2710 ± 402	25.4 ± 5.9
176	1943 ± 459	37.4 ± 10.4
179	2677 ± 370	26.0 ± 6.2
180	1413 ± 202	19.0 ± 2.6
180	2029 ± 239	
181	2136 ± 457	23.5 ± 10.1
182	1526 ± 241	21.4 ± 4.5
183	2374 ± 411	15.4 ± 2.6
184	2565 ± 349	19.8 ± 2.8
186	2632 ± 409	16.9 ± 3.0
187	2877 ± 316	20.8 ± 5.2
188	2732 ± 430	15.8 ± 4.3
189	2991 ± 309	19.9 ± 4.0
190	2434 ± 382	
191	2148 ± 437	18.0 ± 2.8
192	1651 ± 332	26.6 ± 7.7
196	1738 ± 648	17.4 ± 4.4
196	2080 ± 530	
197	1258 ± 251	42.4 ± 8.8
198	2137 ± 291	20.3 ± 3.1
199	2321 ± 315	19.9 ± 2.9
200	2369 ± 425	17.3 ± 3.2
201	2389 ± 387	19.0 ± 3.4
202	2814 ± 556	14.9 ± 3.1
204	3091 ± 469	16.7 ± 3.0
205	2771 ± 566	19.3 ± 3.1
206	3023 ± 549	22.1 ± 4.0
207	2559 ± 339	22.1 ± 3.6
208	2302 ± 399	10.3 ± 4.0
209	2963 ± 645	15.4 ± 1.6
212	2881 ± 366	20.2 ± 3.2
213	1698 ± 256	
214	2000 ± 297	36.3 ± 9.1
215	2947 ± 347	18.3 ± 3.6
216	2118 ± 273	21.3 ± 3.6
217	1636 ± 227	
218	1463 ± 179	
219	2480 ± 421	21.9 ± 3.1
220	1728 ± 275	25.4 ± 4.7
221	1774 ± 260	18.7 ± 6.6
222	2234 ± 254	26.7 ± 5.2
223	1899 ± 231	24.0 ± 4.8

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)
224	1921 ± 195	24.5 ± 5.7
225	1162 ± 115	35.4 ± 5.9
226	1361 ± 178	26.1 ± 4.9
227	1969 ± 354	24.9 ± 3.7
228	1757 ± 183	24.2 ± 3.7
229	2181 ± 299	15.4 ± 3.2
230	2683 ± 243	23.4 ± 4.0
231	2320 ± 352	18.6 ± 3.6
232	1223 ± 172	
233	2525 ± 343	19.2 ± 2.7
234	2234 ± 386	
235	2467 ± 410	17.8 ± 3.1
236	2578 ± 411	23.2 ± 7.1
237	1962 ± 285	21.0 ± 3.0
238	2509 ± 752	31.2 ± 14.8
239	2239 ± 317	19.0 ± 2.4
242	2309 ± 376	25.8 ± 5.4
243	3125 ± 484	21.0 ± 5.7
244	2928 ± 505	20.2 ± 2.7
245	2546 ± 422	18.4 ± 2.8
246	2461 ± 357	20.0 ± 5.8
247	2360 ± 414	20.9 ± 3.1
248	2655 ± 375	24.6 ± 5.2
249	2432 ± 317	19.0 ± 3.4
252	2524 ± 314	17.2 ± 3.6
253	1856 ± 292	26.5 ± 5.0
254	3553 ± 699	23.0 ± 6.2
255	2391 ± 351	17.1 ± 2.0
256	2267 ± 395	22.2 ± 5.2
256	1811 ± 210	
258	2685 ± 430	22.0 ± 3.4
259	2458 ± 387	14.8 ± 3.3
260	2307 ± 299	19.7 ± 3.6
261	1761 ± 267	22.0 ± 3.6
262	2741 ± 378	18.2 ± 2.7
263	1442 ± 251	44.8 ± 10.8
264	2571 ± 376	22.1 ± 7.6
265	2064 ± 197	26.7 ± 4.9
266	2107 ± 302	23.3 ± 3.5
270	2018 ± 240	31.4 ± 8.3
271	2137 ± 263	24.3 ± 3.5
272	2054 ± 319	21.0 ± 3.4

Table 6.9 – Heteronuclear relaxation data for H-2D^b α-chain at 700 MHz.

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
1	1998 ± 192	24.3 ± 0.7	0.838 ± 0.476
3	1445 ± 182	42.0 ± 1.7	0.416 ± 0.056
4	1184 ± 129	55.7 ± 1.0	0.384 ± 0.095
9	1999 ± 304	26.8 ± 0.9	0.938 ± 0.276
9	2030 ± 311	27.2 ± 0.9	0.922 ± 0.276
10	1960 ± 285	25.3 ± 1.3	0.935 ± 0.279
10	2042 ± 262	25.5 ± 0.7	0.878 ± 0.177
11	2021 ± 397	25.1 ± 1.5	0.979 ± 0.284
12	2122 ± 328	26.8 ± 0.6	0.916 ± 0.133
13	1823 ± 269	28.1 ± 0.7	0.876 ± 0.091
21	1951 ± 268	26.1 ± 0.9	0.774 ± 0.112
22	2039 ± 283	23.5 ± 1.2	0.862 ± 0.113
23	2067 ± 338	24.8 ± 0.6	0.940 ± 0.122
24	1947 ± 366	23.9 ± 0.4	0.985 ± 0.188
26	1873 ± 260	26.1 ± 0.7	0.884 ± 0.134
27	2065 ± 458	22.0 ± 0.6	0.938 ± 0.356
27	2077 ± 495	21.9 ± 0.8	0.943 ± 0.251
28	2064 ± 396	25.6 ± 0.8	0.931 ± 0.179
32	1511 ± 203	46.0 ± 2.2	0.515 ± 0.216
33	2131 ± 181	20.5 ± 0.9	0.796 ± 0.275
34	1393 ± 145		0.646 ± 0.122
35	1689 ± 209	34.8 ± 1.9	0.462 ± 0.322
36	2021 ± 343	22.7 ± 0.8	0.909 ± 0.125
37	2176 ± 246	25.9 ± 0.1	0.895 ± 0.070
38	1963 ± 273	24.4 ± 0.5	0.886 ± 0.226
38			0.991 ± 0.226
39	1601 ± 163	47.9 ± 2.9	0.676 ± 0.115
40	1688 ± 202	30.7 ± 0.5	
41	1985 ± 211	35.8 ± 1.0	0.284 ± 0.062
44			0.538 ± 0.212
45			0.524 ± 0.087
46	1895 ± 249	29.4 ± 0.5	0.871 ± 0.128
48	2191 ± 322	23.0 ± 0.5	0.946 ± 0.251
49	1525 ± 162		0.514 ± 0.075
51	2097 ± 288	20.5 ± 0.4	0.840 ± 0.165
52	2276 ± 392	25.1 ± 1.5	0.788 ± 0.176
53	1666 ± 180	31.5 ± 0.4	0.509 ± 0.073
54	1749 ± 192	30.2 ± 0.7	0.665 ± 0.135
55	2142 ± 219	20.5 ± 0.6	0.765 ± 0.161
56	2015 ± 206	28.2 ± 0.5	0.797 ± 0.082
58	2021 ± 271	23.8 ± 0.5	0.873 ± 0.153

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
59	2012 ± 277	25.4 ± 0.7	0.858 ± 0.060
60	2129 ± 268	23.9 ± 0.7	0.923 ± 0.153
62	1922 ± 248	27.9 ± 0.5	0.725 ± 0.129
63	1855 ± 215	31.2 ± 0.9	0.784 ± 0.143
64	1983 ± 307	24.9 ± 0.6	0.927 ± 0.176
68	1371 ± 124	35.5 ± 1.2	0.668 ± 0.380
69	1897 ± 272	25.7 ± 0.9	0.649 ± 0.152
70	2139 ± 240	21.5 ± 1.5	0.886 ± 0.225
70	2051 ± 302	23.6 ± 1.0	0.938 ± 0.207
71	1614 ± 176	30.9 ± 0.7	0.744 ± 0.372
72	2134 ± 303	23.0 ± 1.8	0.836 ± 0.185
73	1373 ± 157		0.298 ± 0.166
76	1590 ± 158	43.2 ± 1.8	0.788 ± 0.154
77	1930 ± 176	30.3 ± 2.5	0.793 ± 0.239
78	2041 ± 268	28.8 ± 0.4	0.895 ± 0.171
78	1825 ± 196	44.0 ± 1.6	0.681 ± 0.142
79	2160 ± 262	23.5 ± 1.0	0.918 ± 0.092
85	2106 ± 311	26.4 ± 0.6	0.847 ± 0.193
86	1765 ± 204	35.8 ± 1.8	0.821 ± 0.117
87	1909 ± 209	25.4 ± 1.2	0.816 ± 0.186
87	1953 ± 238	24.0 ± 1.3	0.888 ± 0.138
88	1919 ± 222	27.7 ± 0.4	0.682 ± 0.116
89	1806 ± 222	25.6 ± 0.5	0.777 ± 0.078
90	1121 ± 138	53.4 ± 1.0	0.611 ± 0.145
91	1417 ± 142		0.647 ± 0.407
92	1556 ± 181	37.5 ± 0.5	0.471 ± 0.074
93	2032 ± 315	25.4 ± 0.4	0.841 ± 0.091
94	1863 ± 226	36.4 ± 1.5	0.780 ± 0.124
95	2108 ± 341	23.1 ± 1.1	0.913 ± 0.207
103	1368 ± 160		0.496 ± 0.217
104	1695 ± 296	28.8 ± 1.5	0.696 ± 0.279
108	1858 ± 282	28.2 ± 1.2	0.804 ± 0.117
108	1889 ± 286	27.9 ± 0.5	0.731 ± 0.244
110	1960 ± 293	30.9 ± 4.1	0.916 ± 0.344
111	1945 ± 181	26.3 ± 1.3	0.803 ± 0.380
112	2115 ± 274	24.0 ± 0.7	0.890 ± 0.186
112	2197 ± 295	24.9 ± 0.7	0.831 ± 0.262
113	2085 ± 299	25.8 ± 0.6	0.900 ± 0.110
114	2156 ± 241	25.1 ± 0.5	0.970 ± 0.157
115	1837 ± 271	26.7 ± 0.9	0.885 ± 0.165
116	2090 ± 241	26.3 ± 0.7	0.923 ± 0.158
116	2110 ± 210	25.5 ± 0.5	0.970 ± 0.232

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
117	2056 ± 332	26.6 ± 1.2	0.923 ± 0.228
118	1937 ± 244	28.7 ± 0.7	0.840 ± 0.092
119	2079 ± 294	24.5 ± 1.0	0.897 ± 0.191
120	2196 ± 371	26.8 ± 2.7	0.913 ± 0.240
120	2243 ± 335	27.7 ± 1.9	0.894 ± 0.240
123	2185 ± 255	25.1 ± 0.9	0.929 ± 0.111
124	2171 ± 209	25.4 ± 1.0	0.991 ± 0.181
125	2048 ± 280	24.8 ± 0.5	0.950 ± 0.146
126	2028 ± 252	24.6 ± 0.5	1.031 ± 0.213
126	2055 ± 260	25.2 ± 1.9	0.951 ± 0.107
127	2131 ± 326	23.9 ± 0.7	0.917 ± 0.090
128	2068 ± 270	25.2 ± 0.3	0.857 ± 0.205
129	2000 ± 314	27.4 ± 1.3	0.871 ± 0.114
130	2075 ± 310	21.5 ± 0.6	0.876 ± 0.130
131	2418 ± 488	22.9 ± 0.6	0.840 ± 0.298
133	1621 ± 217	45.3 ± 3.4	0.534 ± 0.178
134	2191 ± 316	25.5 ± 0.6	0.882 ± 0.174
134	2143 ± 323	24.2 ± 0.7	0.921 ± 0.275
135	1912 ± 196	25.4 ± 0.9	0.829 ± 0.131
136	1484 ± 192	40.6 ± 3.0	0.641 ± 0.074
137	1709 ± 215	30.6 ± 1.4	0.715 ± 0.096
138	1934 ± 253	31.7 ± 1.4	0.758 ± 0.142
139	2004 ± 241	29.1 ± 2.5	0.666 ± 0.133
140	1793 ± 248	26.8 ± 0.4	
141	2017 ± 299	24.5 ± 1.3	0.898 ± 0.154
142	2087 ± 287	25.0 ± 1.0	0.879 ± 0.163
144	1991 ± 272	27.2 ± 0.7	0.849 ± 0.127
145	1891 ± 257	24.6 ± 0.3	0.856 ± 0.109
146	2053 ± 274	24.2 ± 0.7	0.848 ± 0.141
147	2018 ± 361	20.1 ± 1.3	0.868 ± 0.218
148	1969 ± 297	24.0 ± 1.4	0.909 ± 0.122
149	1945 ± 326	27.6 ± 1.0	0.783 ± 0.196
150	2039 ± 301	24.0 ± 0.9	0.680 ± 0.233
151	2027 ± 281	21.9 ± 0.8	0.925 ± 0.224
152	1394 ± 141		0.518 ± 0.205
152		45.2 ± 1.5	0.616 ± 0.179
153	2157 ± 382	22.8 ± 0.7	0.876 ± 0.204
153	2044 ± 236	23.2 ± 0.8	0.806 ± 0.463
154	1691 ± 182	43.9 ± 2.3	0.770 ± 0.176
154	1844 ± 252	35.3 ± 1.1	0.766 ± 0.221
155			0.753 ± 0.130
156			0.793 ± 0.329

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
158	2072 ± 286	21.7 ± 0.5	1.052 ± 0.341
161	2145 ± 285	26.4 ± 0.7	0.922 ± 0.235
162	1688 ± 156		0.640 ± 0.285
169	1500 ± 184	33.9 ± 1.0	0.419 ± 0.101
171	2103 ± 261	25.3 ± 0.7	1.008 ± 0.347
171	2168 ± 339	26.2 ± 1.4	0.972 ± 0.246
172	1242 ± 123	53.3 ± 3.4	0.704 ± 0.141
174	2023 ± 285	27.3 ± 0.6	0.817 ± 0.158
175	2099 ± 227	24.5 ± 0.5	0.880 ± 0.180
176	1953 ± 326	31.5 ± 1.0	0.920 ± 0.194
179	1999 ± 160	21.2 ± 0.8	0.958 ± 0.358
180	1622 ± 273	33.1 ± 0.9	0.582 ± 0.330
180	1797 ± 234	32.4 ± 0.9	0.528 ± 0.170
181	2090 ± 170	26.4 ± 0.3	0.755 ± 0.177
182	1432 ± 158	34.6 ± 1.8	0.548 ± 0.129
183	1854 ± 257	25.2 ± 0.6	0.840 ± 0.165
184	2051 ± 268	26.7 ± 0.5	0.828 ± 0.061
186	2280 ± 288	22.2 ± 1.0	0.943 ± 0.112
187	2406 ± 208	25.5 ± 1.5	0.885 ± 0.095
188	2189 ± 281	24.4 ± 0.3	0.901 ± 0.123
189	2433 ± 195	24.4 ± 0.4	0.859 ± 0.074
190	2113 ± 336	22.3 ± 0.4	0.875 ± 0.102
191	1956 ± 249	26.6 ± 0.6	0.676 ± 0.321
192	1579 ± 207	48.4 ± 2.2	0.677 ± 0.068
196	2342 ± 263	24.6 ± 1.5	0.954 ± 0.208
196			1.007 ± 0.324
197	1377 ± 133	40.0 ± 0.9	0.654 ± 0.139
198	1847 ± 241	27.7 ± 0.5	0.746 ± 0.073
199	1981 ± 282	25.7 ± 0.5	0.871 ± 0.083
200	1875 ± 282	30.8 ± 0.9	0.792 ± 0.075
201	1936 ± 319	25.5 ± 0.6	0.914 ± 0.085
202	2184 ± 345	23.6 ± 1.0	0.905 ± 0.095
204	2225 ± 248	27.3 ± 1.5	0.818 ± 0.055
205	2385 ± 242	24.6 ± 0.6	0.956 ± 0.098
206	2280 ± 207	24.8 ± 0.5	0.920 ± 0.096
207	2183 ± 316	26.2 ± 0.7	0.707 ± 0.147
208	1946 ± 245	22.6 ± 0.5	0.929 ± 0.098
209	2115 ± 319	24.5 ± 0.6	0.922 ± 0.145
212	2393 ± 252	25.3 ± 1.1	0.853 ± 0.376
213	2324 ± 293	25.6 ± 0.4	0.829 ± 0.088
214			0.618 ± 0.057
215	2367 ± 241	24.7 ± 0.7	0.856 ± 0.072

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
216	1966 ± 242	27.4 ± 0.5	0.797 ± 0.086
217	1663 ± 213	37.1 ± 2.0	0.687 ± 0.072
218	1411 ± 240	44.1 ± 3.0	0.519 ± 0.155
219	1907 ± 283	31.6 ± 0.8	0.765 ± 0.065
220	1711 ± 216	28.0 ± 0.5	0.593 ± 0.111
221	1750 ± 226	28.0 ± 0.2	0.684 ± 0.112
222	1861 ± 230	31.7 ± 1.1	0.712 ± 0.047
223	1696 ± 189	30.6 ± 0.7	0.569 ± 0.065
224	1665 ± 222	31.0 ± 0.8	0.637 ± 0.064
225	1198 ± 146	40.5 ± 0.6	0.448 ± 0.080
226	1308 ± 138	51.3 ± 2.8	0.372 ± 0.064
227	1745 ± 214	31.9 ± 0.3	0.658 ± 0.049
228	1739 ± 208	30.6 ± 0.4	0.529 ± 0.052
229	1968 ± 260	24.5 ± 1.0	0.764 ± 0.098
230	2278 ± 221	27.0 ± 0.9	0.766 ± 0.074
231	2040 ± 265	25.3 ± 0.4	0.737 ± 0.117
232	1859 ± 185	37.8 ± 2.4	0.694 ± 0.055
233	2037 ± 296	25.9 ± 0.5	0.831 ± 0.089
234	2154 ± 271	21.1 ± 0.9	0.999 ± 0.109
236	1981 ± 265	24.8 ± 1.1	0.917 ± 0.105
237	1971 ± 332	23.3 ± 0.6	0.874 ± 0.161
238	1486 ± 176		0.371 ± 0.087
239	1717 ± 185	47.9 ± 2.7	0.704 ± 0.163
240	1065 ± 108		0.856 ± 0.204
242	2034 ± 265	28.1 ± 2.7	0.720 ± 0.046
243	2378 ± 231	24.0 ± 0.4	0.954 ± 0.088
244	2303 ± 253	23.4 ± 0.5	0.929 ± 0.128
245	2148 ± 338	24.1 ± 0.6	0.918 ± 0.096
246	2103 ± 316	22.1 ± 0.5	0.873 ± 0.104
247	1926 ± 241	27.3 ± 0.6	0.842 ± 0.090
248	2204 ± 198	29.6 ± 1.5	0.722 ± 0.071
249	1919 ± 273	24.6 ± 0.6	0.893 ± 0.082
252	2140 ± 224	23.9 ± 0.4	0.816 ± 0.104
253	1675 ± 183	45.8 ± 2.2	0.765 ± 0.112
254	2080 ± 251	23.7 ± 0.2	0.816 ± 0.166
255	2031 ± 221	22.4 ± 0.2	0.794 ± 0.122
256	1805 ± 229	41.2 ± 1.3	0.678 ± 0.101
256	1711 ± 222	29.1 ± 1.1	0.460 ± 0.118
258	2042 ± 277	25.5 ± 0.3	0.851 ± 0.101
259	2018 ± 281	24.3 ± 0.7	0.930 ± 0.101
260	1895 ± 273	26.3 ± 0.6	0.922 ± 0.097
261	1664 ± 223	31.1 ± 0.6	0.689 ± 0.065

Residue No.	^{15}N T_1 (ms)	^{15}N T_2 (ms)	^{15}N NOE
262	2128 ± 292	24.7 ± 0.5	0.929 ± 0.083
263	1768 ± 211	40.9 ± 1.2	0.684 ± 0.077
264	2024 ± 273	36.3 ± 1.6	0.674 ± 0.103
265	1980 ± 207	25.7 ± 0.6	0.904 ± 0.221
266	1707 ± 178	40.1 ± 1.3	0.640 ± 0.075
270	1585 ± 207	48.3 ± 2.8	0.717 ± 0.083
271	1891 ± 228	29.7 ± 0.3	0.707 ± 0.062
272	1995 ± 267	25.8 ± 0.4	0.873 ± 0.088

Appendix F

Anisotropic model selection for H-2D^{b}/FL

Table 6.10 - ModelFree model selection for the hβ$_2$m component of H-2D^{b}/FL.
Letters A or B next to residue number denotes a residue with doubled peaks. Residues marked with * are Model 5 with τ_e in excess of 10% of τ_M.

Residue No.	Model	Theta (°)	S^2	S^{2f}	S^{2s}	τ_e	R_{ex}
2	5	95	0.34	0.79	0.43	914	
3*	5	80	0.63	0.88	0.72	3960	
6	1	46	0.88		0.88		
7	1	113	0.55		0.55		
8*	5	50	0.80	1.00	0.80	13036	
9*	5	111	0.77	0.84	0.92	4777	
10	1	55	0.87		0.87		
11	1	124	0.86		0.86		
12*	5	91	0.66		0.66	17904	
13*	5	104	0.40	1.00	0.40	32966	
16	1	55	0.85		0.85		
17	1	108	0.85		0.85		
18*	5	58	0.73	0.84	0.86	3666	
19	1	91	0.64		0.64		
20	1	116	0.61		0.61		
22*	5	105	0.60	1.00	0.60	26205	
24	1	134	0.61		0.61		
25*	5	53	0.63	1.00	0.63	28148	
26	1	146	0.82		0.82		
28	1	125	0.93		0.93		
29*	5	75	0.64	0.99	0.64	18021	
30	1	106	0.89		0.89		
31	1	124	0.87		0.87		
33	1	8	0.95		0.95		
34	5	22	0.74	0.85	0.87	2852	
35	1	78	0.69		0.69		
36	5	100	0.61	0.71	0.86	3080	
37	1	63	0.91		0.91		
38	1	101	0.85		0.85		
39*	5	70	0.73	0.84	0.87	5885	
42*	5	98	0.75	0.88	0.85	5644	
43*	5	80	0.76	0.88	0.87	3433	
44	5	132	0.61	0.74	0.82	3005	
45*	5	80	0.75	0.87	0.86	5039	

Residue No.	Model	Theta (°)	S^2	S^{2f}	S^{2s}	τ_e	R_{ex}
46*	5	78	0.74	0.86	0.86	4111	
47	5	158	0.70	0.81	0.87	1807	
48	5	150	0.52	0.69	0.76	1167	
49	5	39	0.54	0.73	0.74	930	
50	5	149	0.66	0.77	0.85	1897	
51*	5	56	0.76	0.87	0.88	4859	
52* A	5	129	0.77	1.00	0.77	12235	
53	5	91	0.78	0.90	0.87	2451	
54	2	116	0.87		0.87	909	
55	1	124	0.85		0.85		
57 A	2	97	0.84		0.84	947	
58	1	144	0.65		0.65		
59	5	164	0.61	0.75	0.82	1919	
60 A	2	100	0.84		0.84	762	
62	5	77	0.69	0.83	0.83	3247	
63	1	50	0.87		0.87		
64	1	121	0.85		0.85		
65	5	59	0.43	0.59	0.72	3225	
66*	5	96	0.74	1.00	0.74	14031	
67*	5	53	0.78	1.00	0.78	16415	
68*	5	117	0.64	1.00	0.64	19379	
69	1	59	0.86		0.86		
70*	5	87	0.74	0.87	0.86	3754	
71 A	5	90	0.61	0.77	0.79	1576	
73* A	5	50	0.64	0.80	0.79	5890	
74	5	130	0.74	0.85	0.87	2039	
75	5	105	0.60	0.80	0.76	1207	
76	5	108	0.68	0.81	0.83	1150	
77	5	104	0.75	0.85	0.88	3377	
78*	5	112	0.74	0.87	0.85	10309	
80*	5	104	0.79	0.88	0.89	6459	
81	1	84	0.88		0.88		
82*	5	117	0.77	0.86	0.89	5680	
83	1	73	0.94		0.94		
84*	5	121	0.59	0.86	0.69	15032	
85	5	56	0.60	0.74	0.81	1553	
89	5	90	0.69	0.81	0.86	1995	
91*	5	32	0.71	0.92	0.78	13417	
92	5	106	0.69	0.76	0.91	2169	
93	5	75	0.76	0.83	0.92	3057	
95*	5	105	0.74	0.86	0.85	5456	
96	1	23	0.88		0.88		

Residue No.	Model	Theta (°)	S^2	S^{2f}	S^{2s}	τ_e	R_{ex}
97*	5	107	0.69	0.85	0.82	5937	
98	5	103	0.63	0.77	0.83	3056	
99	5	69	0.59	0.74	0.79	2203	

Table 6.11 –ModelFree model selection for the H-2D^b α-chain of H-2D^b/FL.
Letters A or B next to residue number denotes a residue with doubled peaks. Residues marked with * are Model 5 with τ_e in excess of 10% of τ_M.

Residue No.	Model	Theta (°)	S^2	S^{2f}	S^{2s}	τ_e	R_{ex}
3	1	32	0.45		0.45		
4	1	135	0.34		0.34		
9 B	1	63	0.72		0.72		
10 A	1	100	0.77		0.77		
10 B	5	100	0.76	0.85	0.90	2474	
11 A	1	84	0.80		0.80		
12 A	1	100	0.75		0.75		
13	1	77	0.70		0.70		
21	5	104	0.74	0.84	0.88	1983	
22	1	85	0.89		0.89		
23	1	73	0.80		0.80		
24	1	91	0.83		0.83		
26	1	103	0.76		0.76		
27 A	1	34	0.83		0.83		
27 B	1	34	0.84		0.84		
32	1	145	0.41		0.41		
33	1	40	0.95		0.95		
35	1	84	0.59		0.59		
36	1	100	0.88		0.88		
37	1	80	0.76		0.76		
38 A	5	88	0.80	0.89	0.90	2548	
39	1	119	0.43		0.43		
41	5	155	0.50	0.66	0.75	833	
46	1	98	0.67		0.67		
48	5	50	0.81	0.89	0.91	1856	
51	1	116	0.95		0.95		
52	1	139	0.76		0.76		
53	5	151	0.57	0.75	0.76	1162	
54	5	125	0.62	0.76	0.81	1508	
55	1	134	0.94		0.94		
56	1	62	0.69		0.69		
58	5	108	0.81	0.89	0.91	2932	
59	1	117	0.77		0.77		
60*	5	119	0.80	0.88	0.91	3547	
62	1	107	0.70		0.70		
63	1	118	0.63		0.63		
64	5	107	0.78	0.87	0.89	2925	
68	5	96	0.54	0.74	0.73	1719	
69	5	110	0.75	0.86	0.87	1297	

Residue No.		Model	Theta (°)	S^2	S^{2f}	S^{2s}	τ_e	R_{ex}
70	A	2	121	0.93		0.93	1419	
70	B	1	121	0.83		0.83		
71		1	103	0.64		0.64		
72	A	1	92	0.94		0.94		
76		1	94	0.47		0.47		
77		1	95	0.77		0.77		
78	A	1	100	0.69		0.69		
78	B	1	100	0.46		0.46		
79		1	84	0.87		0.87		
85		1	105	0.74		0.74		
86		1	44	0.54		0.54		
87	A	1	73	0.76	0.86	0.89	2816	
88		1	153	0.66		0.66		
89		1	122	0.76		0.76		
90		1	49	0.35		0.35		
92	A	5	127	0.49	0.70	0.70	1137	
93		1	73	0.77		0.77		
94	A	1	85	0.56		0.56		
95		1	99	0.88		0.88		
104		1	26	0.64		0.64		
108	A	1	139	0.69		0.69		
108	B	1	139	0.67		0.67		
110		1	98	0.75		0.75		
111		1	127	0.76		0.76		
112*	A	5	70	0.80	0.88	0.91	3381	
112*	B	5	70	0.77	0.85	0.91	3979	
113		1	92	0.77		0.77		
114*		5	97	0.77	0.85	0.91	3546	
115		1	110	0.75		0.75		
116	A	1	50	0.73		0.73		
116	B	1	50	0.75		0.75		
117		1	86	0.76		0.76		
118		1	87	0.69		0.69		
119		1	82	0.82		0.82		
120	A	1	63	0.76		0.76		
120	B	1	63	0.73		0.73		
123		1	109	0.79		0.79		
124		1	128	0.78		0.78		
125*		5	62	0.76	0.87	0.88	4227	
126	A	5	117	0.77	0.87	0.89	3320	
126	B	1	117	0.83		0.83		
127		1	13	0.75		0.75		

Residue No.	Model	Theta (°)	S^2	S^{2f}	S^{2s}	τ_e	R_{ex}
128	1	114	0.77		0.77		
129	1	83	0.76		0.76		
130	1	133	0.89		0.89		
131	1	130	0.83		0.83		
133	1	32	0.44		0.44		
134* A	5	116	0.74	0.83	0.89	5341	
134 B	5	116	0.79	0.86	0.91	2419	
135*	5	50	0.72	0.85	0.85	5225	
136	1	95	0.54		0.54		
137	1	71	0.67		0.67		
138	1	126	0.62		0.62		
139	1	140	0.71		0.71		
141	1	111	0.84		0.84		
142	1	125	0.78		0.78		
144	1	113	0.72		0.72		
145	1	119	0.79		0.79		
146	1	134	0.79		0.79		
147	1	123	0.97		0.97		
148	1	90	0.87		0.87		
149	1	117	0.72		0.72		
150	5	133	0.78	0.86	0.90	1399	
151	1	96	0.92		0.92		
152 A	5	76	0.42	0.67	0.63	1240	
153 A	1	64	0.85		0.85		
153 B	1	64	0.85		0.85		
154 A	1	38	0.46		0.46		
154 B	1	38	0.53		0.53		
158	1	47	0.87		0.87		
161	5	26	0.68	0.77	0.88	2410	
169	1	18	0.54		0.54		
171 B	1	34	0.72		0.72		
172	1	23	0.35		0.35		
174	1	38	0.68		0.68		
175*	5	34	0.73	0.83	0.89	3577	
176	1	42	0.59		0.59		
180 A	5	63	0.57	0.79	0.73	1151	
180 B	1	63	0.60		0.60		
181	5	65	0.73	0.82	0.89	1564	
182	5	124	0.54	0.76	0.72	1290	
183	5	48	0.73	0.85	0.87	3350	
184	1	110	0.74		0.74		
186	1	60	0.90		0.90		

Residue No.	Model	Theta (°)	S^2	S^{2f}	S^{2s}	τ_e	R_{ex}
187	1	98	0.86		0.86		
188	1	38	0.76		0.76		
189	5	145	0.74	0.81	0.92	3106	
190	1	31	0.82		0.82		
191	5	128	0.71	0.82	0.86	1340	
192	1	76	0.42		0.42		
196 A	1	44	0.81		0.81		
197	1	34	0.46		0.46		
198	1	93	0.72		0.72		
199	1	54	0.75		0.75		
200	1	130	0.62		0.62		
201	1	48	0.74		0.74		
202	1	159	0.79		0.79		
204	1	158	0.72		0.72		
205	1	62	0.80		0.80		
206	1	125	0.78		0.78		
207	5	81	0.74	0.82	0.90	1407	
208	1	76	0.87		0.87		
209	1	120	0.80		0.80		
212	5	38	0.72	0.79	0.91	1751	
213	1	57	0.75		0.75		
215	1	45	0.78		0.78		
216	1	126	0.70		0.70		
217	1	40	0.52		0.52		
218	1	149	0.42		0.42		
219	1	34	0.59		0.59		
220	5	121	0.67	0.83	0.81	1256	
221	5	77	0.69	0.83	0.83	1509	
222	1	113	0.63		0.63		
223	1	36	0.61		0.61		
224	1	84	0.65		0.65		
225	1	112	0.48		0.48		
226	1	127	0.40		0.40		
227	1	41	0.58		0.58		
228	1	24	0.59		0.59		
229	5	166	0.72	0.83	0.87	1975	
230	1	41	0.72		0.72		
231	1	150	0.72		0.72		
232	1	92	0.57		0.57		
233	1	57	0.74		0.74		
234	1	146	0.89		0.89		
236	1	64	0.81		0.81		

Residue No.	Model	Theta (°)	S^2	S^{2f}	S^{2s}	τ_e	R_{ex}
237	1	45	0.81		0.81		
239	1	85	0.44		0.44		
242	1	42	0.74		0.74		
243	1	153	0.76		0.76		
244*	5	42	0.78	0.87	0.90	5826	
245	1	145	0.77		0.77		
246*	5	28	0.81	0.90	0.90	3363	
247	1	151	0.67		0.67		
248	5	19	0.59	0.69	0.86	1815	
249	1	101	0.81		0.81		
252	5	89	0.82	0.89	0.92	2089	
253	1	77	0.46		0.46		
255	5	153	0.80	0.89	0.90	2085	
256 A	1	106	0.49		0.49		
256 B	1	106	0.69		0.69		
258	1	43	0.73		0.73		
259	1	141	0.78		0.78		
260	1	53	0.73		0.73		
261	1	132	0.61		0.61		
262	1	71	0.80		0.80		
263	1	113	0.48		0.48		
264	1	56	0.55		0.55		
265	5	69	0.74	0.85	0.88	1681	
266	1	36	0.47		0.47		
270	1	38	0.41		0.41		
271	1	123	0.65		0.65		
272	1	49	0.73		0.73		

References

1. Wood, P. *Understanding Immunology*, (Pearson, 2006).
2. Kindt, T., Goldsby, R. & Osborne, B. *Kuby Immunology*, (W.H. Freeman and Company, 2007).
3. Wooldridge, L. et al. A single autoimmune T cell receptor recognises more than a million different peptides. *J. Biol. Chem.* **287**, 1168-1177 (2012).
4. Harty, J., Tvinnereim, A. & White, D. CD8(+) T cell effector mechanisms in resistance to infection. *Annu. Rev. Immunol.* **18**, 275-308 (2000).
5. Morrison, L., Braciale, V. & Braciale, T. Distinguishable pathways of viral-antigen presentation to lymphocytes-T. *Immunol. Res.* **5**, 294-304 (1986).
6. Madden, D.R. The 3-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol* **13**, 587-622 (1995).
7. Sidney, J. et al. Several HLA alleles share overlapping peptide specificities. *J. Immunol.* **154**, 247-259 (1995).
8. Groothius, T.A.M. & Neefjes, J. The many roads to cross-presentation. *J. Exp. Med.* **202**, 1313-1318 (2005).
9. Blum, J., Wearsch, P. & Cresswell, P. Pathways of antigen processing. *Ann. Rev. Immunol.* **31**, 443-473 (2013).
10. Yewdell, J.W., Reits, E. & Neefjes, J. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat. Rev. Immunol.* **3**, 952-961 (2003).
11. Reits, E., Vos, J., Gromme, M. & Neefjes, J. The major substrates for TAP in vivo are derived from newly synthesised proteins. *Nature* **404**, 774-778 (2000).
12. Qian, S.B. et al. Tight linkages between translation and MHC class I peptide ligand generation implies specialised antigen processing for defective ribosomal products. *J. Immunol.* **177**, 227-233 (2006).
13. Princiotta, M. et al. Quantitating protein synthesis, degradation, and endogenous antigen processing. *Immunity* **18**, 343-354 (2003).
14. Toes, R.E. et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J. Exp. Med.* **194**, 1-12 (2001).
15. Heink, S., Lubwig, D., Kloetzel, P.M. & Kruger, E. IFN-γ-induced immune adaptation of the proteasome system is an accelerated and transient response. *Proc Natl Acad Sci U S A* **102**, 9241-9246 (2005).
16. Yewdell, J.W. Immunoproteasomes: regulating the regulator. *Proc Natl Acad Sci U S A* **102**, 9089-9090 (2005).
17. York, I.A., Bhutani, N., Zendzian, S., Goldberg, A.L. & Rock, K.L. Tripeptidyl peptidase II is the major peptidase needed to trim long antigenic precursors, but is not required for most MHC class I antigen presentation. *J. Immunol.* **177**, 1434-1443 (2006).
18. Diedrich, G., Bangia, N., Pan, M. & Cresswell, P. A role for calnexin in the assembly of the MHC class I loading complex in the endoplasmic reticulum. *J. Immunol.* **166**, 1703-1709 (2001).
19. Garbi, N., Tanaka, S., van der Broek, M., Momburg, F. & Hammerling, G.J. Accessory molecules in the assembly of major histocompatibility complex class I/peptide complexes: how essential are they for CD8(+) T-cell immune responses? *Immunol. Rev.* **207**, 77-88 (2005).
20. Gao, B. et al. Assembly and antigen-presenting function of MHC class I molecules in cells lacking the ER chaperone calreticulin. *Immunity* **16**, 99-109 (2002).

21. Howe, C. et al. Calreticulin-dependent recycling in the early secretory pathway mediates optimal peptide loading of MHC class I molecules. *EMBO J.* **28**, 3730-3744 (2009).
22. Dalchau, N. et al. A peptide filtering relation quantifies MHC class I peptide optimisation. *PLOS Comput. Biol.* **7**, e1002144 (2011).
23. Garbi, N., Tanaka, S., Momburg, F. & Hammerling, G.J. Impaired assembly of the major histocompatibility complex class I peptide-loading complex in mice deficient in the oxidoreductase ERp57. *Nat. Immunol.* **7**, 93-102 (2006).
24. Leonhardt, R., Abrahimi, P., Mitchell, S. & Cresswell, P. Three Tapasin docking sites in TAP cooperate to facilitate transporter stabilization and heterodimerization. *J. Immunol.* **192**, 2480-2494 (2014).
25. Li, S., Paulsson, K., Chen, S., Sjogren, H. & Wang, P. Tapasin is required for efficient peptide binding to transporter associated with antigen processing. *J. Biol. Chem.* **275**, 1581-1586 (2000).
26. Geng, J., Pogozheva, I., Mosberg, H. & Raghavan, M. Use of functional polymorphisms to elucidate the peptide binding site of TAP complexes. *J. Immunol.* **195**, 3436-3448 (2015).
27. van Endert, P.M. et al. A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity* **1**, 491-500 (1994).
28. van Endert, P.M. et al. The peptide-binding motif for the human transporter associated with antigen processing. *J. Exp. Med.* **182**, 1883-1895 (1995).
29. Knittler, M., Alberts, P., Deverson, E. & Howard, L. Nucleotide binding by TAP mediates association with peptide and release of assembled MHC class I molecules. *Curr. Biol.* **9**, 999-1008 (1999).
30. Leonhardt, R., Keusekotten, K., Bekpen, C. & Knittler, M. Critical role for the tapasin-docking site of TAP2 in the functional integrity of the MHC class-I-peptide-loading complex. *J. Immunol.* **175**, 5104-5114 (2005).
31. Ortmann, B. et al. A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. *Science* **277**, 1306-1309 (1997).
32. Owen, B. & Pease, L. TAP association influences the conformation of nascent MHC class I molecules. *J. Immunol.* **162**, 4677-4684 (1999).
33. Elliott, T., Willis, A., Cerundolo, V. & Townsend, A. Processing of major histocompatibility class-I-restricted antigens in the endoplasmic-reticulum. *J. Exp. Med.* **181**, 1481-1491 (1995).
34. Chang, S.C., Momburg, F., Bhutani, N. & Goldberg, A.L. The ER aminopeptidase, ERAP1, trims percursors to lengths of MHC class I peptides by a 'molecular ruler' mechanism. *Proc. Natl. Acad. Sci. USA* **102**, 17107-17112 (2005).
35. Reeves, E., Edwards, C., Elliott, T. & James, E. Naturally occurring ERAP1 haplotypes encode functionally distinct alleles with fine substrate specificity. *J. Immunol.* **191**, 35-43 (2013).
36. Hammer, G.E., Gonzalez, F., Champsaur, M., Cado, D. & Shastri, N. The aminopeptidase ERAAP shapes the peptides repertoire displayed by major histocompatibility complex class I molecules. *Nat. Immunol.* **7**, 103-112 (2006).
37. Zhang, Y. et al. ERp57 does not require interactions with calnexin and calreticulin to promote assembly of class I histocompatibility molecules, and it enhances peptide loading independently of its redox activity. *J. Biol. Chem.* **284**, 10160-10173 (2009).
38. Peaper, D., Wearsch, P.A. & Cresswell, P. Tapasin and ERp57 form a stable disulfide-linked dimer within the MHC class I peptide-loading complex. *EMBO J.* **24**, 3613-3623 (2005).

39. Dong, G., Wearsch, P.A., Peaper, D.R., Cresswell, P. & Reinisch, K. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* **30**, 21-32 (2009).
40. Dick, T., Bangla, N., Peaper, D.R. & Cresswell, P. Disulfide bond isomerization and the assembly of MHC class I-peptide complexes. *Immunity* **16**, 89-98 (2002).
41. Wijeyesakere, S., Gagnon, J., Arora, K., Brooks, C. & Raghavan, M. Regulation of calreticulin-major histocompatibility complex (MHC) class I interactions by ATP. *Proc Natl Acad Sci U S A* **112**, E5608-E5617 (2015).
42. Chen, M. & Bouvier, M. Analysis of interactions in a tapasin/class I complex provides a mechanism for peptide selection. *EMBO J.* **26**, 1681-1690 (2007).
43. Wearsch, P.A., Peaper, D.R. & Cresswell, P. Essential glycan-dependent interactions optimizse MHC class I peptide loading. *Proc. Natl. Acad. Sci. USA* **108**, 4950-4955 (2011).
44. Moore, S. & Spiro, R. Inhibition of glucose trimming by castanospermine results in rapid degradation of unassembled major histocompatibility complex class-I molecules. *J. Biol. Chem.* **268**, 3809-3812 (1993).
45. Van Leeuwen, J. & Kearse, K. Deglucosylation of N-linked glycans is an important step in the dissociation of calreticulin-class I-TAP complexes. *Proc. Natl. Acad. Sci. USA* **93**, 13997-14001 (1996).
46. Ghanem, E. et al. The transporter associated with antigen processing (TAP) is active in a post-ER compartment. *J. Cell Sci.* **123**, 4271-4279 (2010).
47. Pentcheva, T. & Edidin, M. Clustering of peptide-loaded MHC class I molecules for endoplasmic reticulum export imaged by fluorescence resonance energy transfer. *J. Immunol.* **166**, 6625-6632 (2001).
48. Hermann, C., Trowsdale, J. & Boyle, L. TAPBPR: a new player in the MHC class I presentation pathway. *Tissue Antigens* **85**, 155-166 (2015).
49. Yewdell, J.W. Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. *Immunity* **25**, 533-543 (2006).
50. van der Burg, S.H., Visseren, M.J.W., Brandt, R.M.P., Kast, W.M. & Melief, C.J.M. Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *J. Immunol.* **156**, 3308-3314 (1996).
51. Rudolph, M., Stanfield, R. & Wilson, I. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol* **24**, 419-466 (2006).
52. Ding, Y., Baker, B., Garboczi, D., Biddison, W. & Wiley, D.C. Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. *Immunity* **11**, 45-56 (1999).
53. Hemmer, B. et al. Predictable TCR antigen recognition based on peptide scans leads to the identification of agonist ligands with no sequence homology. *J. Immunol.* **160**, 3631-3636 (1998).
54. Galea, I. et al. CD8+ T-cell cross-competition is governed by peptide-MHC class I stability. *Eur. J. Immunol.* **42**, 256-263 (2012).
55. Yewdell, J.W. & Bennink, J.R. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu. Rev. Immunol.* **17**, 51-88 (1999).
56. Hsu, V.W. et al. A recycling pathway between the endoplasmic-reticulum and the Golgi-apparatus for retention of unassembled MHC class I molecules. *Nature* **352**, 441-444 (1991).
57. Garstka, M. et al. Peptide-receptive major histocompatibility complex class I molecules cycle between endoplasmic reticulum and cis-Golgi in wild-type lymphocytes. *J. Biol. Chem.* **282**, 30680-30690 (2007).
58. Radcliffe, C. et al. Identification of specific glycoforms of major histocompatibility complex class I heavy chains suggests that class I peptide

loading is an adaptation of the quality control pathway involving calreticulin and ERp57. *J. Biol. Chem.* **277**, 46415-46423 (2002).
59. Leonhardt, R. et al. Post-endoplasmic reticulum rescue of unstable MHC class I requires proprotein convertase PC7. *J. Immunol.* **184**, 2985-2998 (2010).
60. Paulsson, K. et al. Association of tapasin and COPI provides a mechanism for the retrograde transport of major histocompatibility complex (MHC) class I molecules from the Golgi complex to the endoplasmic reticulum. *J. Biol. Chem.* **277**, 18266-18271 (2002).
61. Williams, A.P., Peh, C.A., Purcell, A.W., McCluskey, J. & Elliott, T. Optimization of the MHC class I peptide cargo is dependent on tapasin. *Immunity* **16**, 509-520 (2002).
62. Ellgaard, L., Molinari, M. & Helenius, A. Setting the standards: quality control in the secretory pathway. *Science* **286**, 1882-1888 (1999).
63. Purcell, A.W. et al. Quantitative and qualitative influences of tapasin on the class I peptide repertoire. *J. Immunol.* **166**, 1016-1027 (2001).
64. Thirdborough, S.M. et al. Tapasin shapes immunodominance hierarchies according to the kinetic stability of peptide – MHC class I complexes. *Eur. J. Immunol.* **38**, 364-369 (2008).
65. Lewis, J.W. & Elliott, T. Evidence for successive peptide binding and quality control stages during MHC class I assembly. *Curr. Biol.* **8**, 717-20 (1998).
66. Howarth, M., Williams, A., Tolstrup, A.B. & Elliott, T. Tapasin enhances MHC class I peptide presentation according to peptide half-life. *Proc. Natl. Acad. Sci. USA* **101**, 11737-11742 (2004).
67. Momburg, F. & Tan, P. Tapasin - the keystone of the loading complex optimising peptide binding by MHC class I molecules in the endoplasmic reticulum. *Mol. Immunol.* **39**, 217-233 (2002).
68. Lehner, P., Surman, M. & Cresswell, P. Soluble tapasin restores MHC class I expression and function in the tapasin-negative cell line .220. *Immunity* **8**, 221-231 (1998).
69. Paulsson, K. et al. Assembly of tapasin-associated MHC class I in the absence of the transporter associated with antigen processing (TAP). *Int. Immunol.* **13**, 23-29 (2001).
70. Park, B. & Ahn, K. An essential function of tapasin in quality control of HLA-G molecules. *J. Biol. Chem.* **278**, 14337-14345 (2003).
71. Lewis, J.W., Sewell, A., Price, D. & Elliott, T. HLA-A*0201 presents TAP-dependent peptide epitopes to cytotoxic T lymphocytes in the absence of tapasin. *Eur. J. Immunol.* **28**, 3214-3220 (1998).
72. Wearsch, P.A. & Cresswell, P. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. *Nat. Immunol.* **8**, 873-881 (2007).
73. Praveen, P.V.K., Yaneva, R., Kalbacher, H. & Springer, S. Tapasin edits peptides on MHC class I molecules by accelerating peptide exchange. *Eur. J. Immunol.* **40**, 214-224 (2010).
74. Springer, S., Doring, K., Skipper, J., Townsend, A. & Cerundolo, V. Fast association rates suggest a conformational change in the MHC class I molecule H-2D(b) upon peptide binding. *Biochemistry* **37**, 3001-3012 (1998).
75. Wright, C., Kozik, P., Zacharias, M. & Springer, S. Tapasin and other chaperones: models of the MHC class I loading complex. *Biol. Chem.* **385**, 763-778 (2004).
76. van Hateren, A., Bailey, A., Werner, J. & Elliott, T. Plasticity of empty major histocompatibility complex class I molecules determines peptide selector function. *Molecular Immunology* **68**, 98-101 (2015).

77. Cerundolo, V., Elliott, T., Elvin, J., Bastin, J. & Rammensee, H.G. The binding-affinity and dissociation rates of peptides for class-I major histocompatibility complex-molecules. *Eur. J. Immunol.* **21**, 2069-2075 (1991).
78. Sijts, A.J.A.M. & Pamer, E. Enhanced intracellular dissociation of major histocompatibility complex class I-associated peptides: A mechanism for optimising the spectrum of cell surface-presented cytotoxic T lymphocye epitopes. *J. Exp. Med.* **185**, 1403-1411 (1997).
79. Glithero, A. et al. The crystal structure of H-2Db complexed with a partial peptide epitope suggests a major histocompatibility complex class I assembly intermediate. *J. Biol. Chem.* **281**, 12699-12704 (2006).
80. Baker, B., Scott, D., Blevins, S. & Hawse, W. Structural and dynamic control of T-cell receptor specificity, cross-reactivity, and binding mechanism. *Immunol. Rev.* **250**, 10-31 (2012).
81. Klein, J. Seeds of time - 50 years ago Gorer, Peter, A. discovered the H-2-complex. *Immunogenetics* **24**, 331-338 (1986).
82. Snell, G. & Higgins, G. Alleles at the histocompatibility-2 locus in the mouse as determined by tumor transplantation. *Genetics* **36**, 306-310 (1951).
83. NobelMedia. The Nobel Prize in Physiology or Medicine 1980. *nobelprize.org/nobel_prizes/medicine/laureates/1980/* (2016).
84. Beck, S. et al. Complete sequencing and gene map of a human major histocompatibility complex. *Nature* **401**, 921-923 (1999).
85. Cresswell, P., Turner, M. & Strominger, J.L. Papain-solubilised HL-A antigens from cultured human lymphocytes contain 2 peptide fragments. *Proc. Natl. Acad. Sci. USA* **70**, 1603-1607 (1973).
86. Grey, H. et al. Small subunit of HL-A antigens is beta2-microglobulin. *J. Exp. Med.* **138**, 1608-1612 (1973).
87. Robinson, J. et al. The IMGT/HLA Database. *Nucleic Acids Res.* **41**, D1222-1227 (2013).
88. Lopez de Castro, J.A. et al. Complete amino-acid sequence of a papain-solubilized human histocompatibility antigen, HLA-B7. 1. Isolation and amino-acid composition of fragments and of tryptic and chymotryptic peptides. *Biochemistry* **18**, 5704-5711 (1979).
89. Orr, H., Lopez de Castro, J.A., Lancet, D. & Strominger, J.L. Complete amino-acid sequence of a papain-solubilized human histocompatibility antigen, HLA-B7. 2. Sequence determination and search for homologies. *Biochemistry* **18**, 5711-5720 (1979).
90. Orr, H., Lancet, D., Robb, R., Lopez de Castro, J.A. & Strominger, J.L. Heavy-chain of human histocompatibility antigen HLA-B7 contains an immunoglobulin-like region. *Nature* **282**, 266-270 (1979).
91. Springer, T. & Strominger, J.L. Detergent-soluble HLA antigens contain a hydrophilic region at COOH-terminus and a penultimate hydrophobic region. *Proc. Natl. Acad. Sci. USA* **73**, 2481-2485 (1976).
92. Dausset, J. Iso-leuco-anticorps. *Acta Haematol.* **20**, 156-166 (1958).
93. Bjorkman, P.J. et al. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506-512 (1987).
94. Saper, M., Bjorkman, P.J. & Wiley, D.C. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6A resolution. *J. Mol. Biol.* **219**, 277-319 (1991).
95. Cresswell, P. et al. Immunological identity of the small subunit of HL-A antigens and beta2-microglobulin and its turnover on the cell membrane. *Proc. Natl. Acad. Sci. USA* **71**, 2123-2127 (1974).

96. Garrett, T., Saper, M.A., Bjorkman, P.J., Strominger, J.L. & Wiley, D.C. Specificity pockets for the side-chains of peptide antigens in HLA-AW68. *Nature* **342**, 692-696 (1989).
97. Madden, D., Gorga, J., Strominger, J.L. & Wiley, D.C. The 3-dimensional structure of HLA-B27 at 2.1 Angstrom resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* **70**, 1035-1048 (1992).
98. Glithero, A. et al. Crystal structures of two H-2D(b)/glycopeptide complexes suggest a molecular basis for CTL cross-reactivity. *Immunity* **10**, 63-74 (1999).
99. Zhang, W., Young, A.C.M., Imari, M., Nathenson, S.G. & Sacchettini, J.C. Crystal structure of the major histocompatibility complex class I H-2Kb molecule containing a single viral peptide: Implications for peptide binding and T-cell receptor recognition. *Proc. Natl. Acad. Sci. USA* **89**, 8403-8407 (1992).
100. Madden, D.R., Gorga, J., Strominger, J.L. & Wiley, D.C. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature* **353**, 321-325 (1991).
101. Zhang, C., Anderson, A. & DeLisi, C. Structural principles that govern the peptide-binding motifs of class I MHC molecules. *J. Mol. Biol.* **281**, 929-947 (1998).
102. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H.G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290-296 (1991).
103. Fremont, D., Matsamura, M., Stura, E., Peterson, P. & Wilson, I. Crystal structures of 2 viral peptides in complex with murine MHC class-I H-2K(b). *Science* **257**, 919-927 (1992).
104. Tysoe-Calnon, V., Grundy, J. & Perkins, S. Molecular comparisons of the beta-2-microglobulin-binding site in class-I major-histocompatibility-complex alpha-chains and proteins of related sequences. *Biochem. J.* **277**, 359-369 (1991).
105. Young, A.C.M., Nathenson, S.G. & Sacchettini, J.C. Structural studies of class-I major histocompatibility complex proteins - insights into antigen presentation. *FASEB J.* **9**, 26-36 (1995).
106. Young, A.C.M., Zhang, W.G., Sacchettini, J.C. & Nathenson, S.G. The 3-dimensional structure of H-2D(b) at 2.4 angstrom resolution - implications for antigen-determinant selection. *Cell* **76**, 39-50 (1994).
107. Ciatto, C. et al. Zooming in on the hydrophobic ridge of H-2D(b): Implications for the conformational variability of bound peptides. *J. Mol. Biol.* **312**, 1059-1071 (2001).
108. Meng, W.S., vonGrafenstein, H. & Haworth, I.S. A model of water structure inside the HLA-A2 peptide binding groove. *Int. Immunol.* **9**, 1339-1346 (1997).
109. Petrone, P.M. & Garcia, A.E. MHC-peptide binding is assisted by bound water molecules. *J. Mol. Biol.* **338**, 419-435 (2004).
110. Nossner, E. & Parham, P. Species-specific differences in chaperone interaction of human and mouse major histocompatibility complex class-I molecules. *J. Exp. Med.* **181**, 327-337 (1995).
111. Bergeron, J. et al. The role of the lectin calnexin in conformation independent binding to N-linked glycoproteins and quality control. *Adv. Exp. Med. Biol.* **435**, 105-116 (1998).
112. Ribaudo, R. & Margulies, D. Independent and synergistic effects of disulfide bond formation, beta-2-microglobulin, and peptides on class-I MHC folding and assembly in an in vitro translation system. *J. Immunol.* **149**, 2935-2944 (1992).
113. Tector, M. & Salter, R. Calnexin influences folding of human class I histocompatibility proteins but not their assembly with beta 2-microglobulin. *J. Biol. Chem.* **270**, 19638-19642 (1995).

114. Farmery, M.R., Allen, S., Allen, A.J. & Bulleid, N.J. The role of ERp57 in disulfide bond formation during the assembly of major histocompatibility complex class I in a synchronized semipermeabilized cell translation system. *J. Biol. Chem.* **275**, 14933-14938 (2000).
115. Del Cid, N. et al. Modes of calreticulin recruitment to the major histocompatibility complex class I assembly pathway. *J. Biol. Chem.* **285**, 4520-4535 (2010).
116. Hughes, E., Hammond, C. & Cresswell, P. Misfolded major histocompatibility complex class I heavy chains are translocated into the cytoplasm and degraded by the proteasome. *Proc. Natl. Acad. Sci. USA* **94**, 1896-1901 (1997).
117. Borbulevych, O. et al. Structures of MART-1 (26/27-35) peptide/HLA-A2 complexes reveal a remarkable disconnect between antigen structural homology and T cell recognition. *J. Mol. Biol.* **372**, 1123-1136 (2007).
118. Wu, L., Tuot, D., Lyons, D., Garcia, K. & Davis, M. Two-step binding mechanism for T-cell receptor recognition of peptide-MHC. *Nature* **418**, 552-556 (2002).
119. Liu, Y. et al. The energetic basis underpinning T-cell receptor recognition of a super-bulged peptide bound to a major histocompatibility complex class I molecule. *J. Biol. Chem.* **287**, 12267-12276 (2012).
120. Degano, M. et al. A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. *Immunity* **12**, 251-261 (2000).
121. Wang, B. et al. Peptidic termini play a significant role in TCR recognition. *J. Immunol.* **169**, 3137-3145 (2002).
122. Hawse, W. et al. TCR scanning of peptide/MHC through complementary matching of receptor and ligand molecular flexibility. *J. Immunol.* **192**, 2885-2891 (2014).
123. Insaidoo, F.K. et al. Loss of T cell antigen recognition arising from changes in peptide and major histocompatibility complex protein flexibility - implications for vaccine design. *J. Biol. Chem.* **286**, 40163-40173 (2011).
124. Devine, L., Sun, J., Barr, M. & Kavathas, P. Orientation of the Ig domains of CD8 alpha beta relative to MHC class I. *J. Immunol.* **162**, 846-851 (1999).
125. Ingulli, E. Mechanism of cellular rejection in transplantation. *Pediatr. Nephrol.* **25**, 61-74 (2010).
126. Goulder, P. & Walker, B. HIV and HLA class I: An evolving relationship. *Immunity* **37**, 426-440 (2012).
127. Gough, S. & Simmonds, M. The HLA region and autoimmune disease: associations and mechanisms of action. *Curr. Genomics* **8**, 453-465 (2007).
128. Fogdell-Hahn, A., Ligers, A., Gronning, M., Hillert, J. & Olerup, O. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. *Tissue Antigens* **55**, 140-148 (2000).
129. Nejentsev, S. et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* **450**, 887-892 (2007).
130. Pohlmann, T. et al. Differential peptide dynamics is linked to major histocompatibility complex polymorphism. *J. Biol. Chem.* **279**, 28197-28201 (2004).
131. Berggard, I. & Bearn, A. Isolation and properties of a low molecular weight beta-2-globulin occuring in human biological fluids. *J. Biol. Chem.* **243**, 4095-4103 (1968).
132. Cunningham, B., Wang, J., Berggard, I. & Peterson, P. The complete amino acid sequence of beta 2-microglobulin. *Biochemistry* **12**, 4811-4822 (1973).
133. Okon, M., Bray, P. & Vucelic, D. 1H NMR assignments and secondary structure of human beta2-microglobulin in solution. *Biochemistry* **31**, 8906-8915 (1992).

134. Verdone, G. et al. The solution structure of human beta2-microglobulin reveals the prodromes of its amyloid transition. *Protein Sci.* **11**, 487-499 (2002).
135. Trinh, C.H., Smith, D.P., Kalverda, A.P., Phillips, S.E.V. & Radford, S.E. Crystal structure of monomeric human β-2-microglobulin reveals clues to its amyloidogenic properties. *Proc. Natl. Acad. Sci. USA* **99**, 9771-9776 (2002).
136. Hee, C.-S. et al. Dynamics of free versus complexed B2-microglobulin and the evolution of interfaces in MHC class I molecules. *Immunogenetics* **65**, 157-172 (2013).
137. Burdin, N. & Kronenberg, M. CD1-mediated immune responses to glycolipids. *Curr. Opin. Immunol.* **11**, 326-331 (1999).
138. Bjorck, L., Akerstrom, B. & Berggard, I. Occurrence of beta-2-microglobulin in mammalian lymphocytes and erythrocytes. *Eur. J. Immunol.* **9**, 486-490 (1979).
139. Karlsson, F., Wibell, L. & Evrin, P. Beta-2-microglobulin in clinical medicine. *Scand. J. Clin. Lab. Invest.* **40**, 27-37 (1980).
140. Chiti, F. & Dobson, C. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).
141. Divry, P. & Florkin, M. Sur les proprietes optiques de l'amyloide. *C. R. Soc. Biol.* **97**, 1808-1810 (1927).
142. Shirahama, T. & Cohen, A. High-resolution electron microscopic analysis of the amyloid fibril. *J. Cell Biol.* **33**, 679-708 (1967).
143. Sunde, M. et al. Common core structure of amyloid fibrils by synchotron X-ray diffraction. *J. Mol. Biol.* **273**, 729-739 (1997).
144. Tycko, R. Solid-state NMR studies of amyloid fibril structure. *Annu. Rev. Phys. Chem.* **62**, 279-299 (2011).
145. Heegaard, N.H.H. Beta(2)-microglobulin: from physiology to amyloidosis. *Amyloid* **16**, 151-173 (2009).
146. Yu, Y., Turnquist, H.R., Myers, N., Balendiran, G. & Hansen, T.H. An extensive region of an MHC class I alpha 2 domain loop influences interaction with the assembly complex. *J. Immunol.* **163**, 4427-4433 (1999).
147. Suh, W. et al. Interaction of murine MHC class I molecules with tapasin and TAP enhances peptide loading and involves the heavy chain alpha(3) domain. *J. Immunol.* **162**, 1530-1540 (1999).
148. Hildebrand, W.H. et al. HLA class I polymorphism has a dual impact on ligand binding and chaperone interaction. *Hum. Immunol.* **63**, 248-255 (2002).
149. Zernich, D. et al. Natural HLA class I polymorphism controls the pathway of antigen presentation and susceptibility to viral evasion. *J. Exp. Med.* **200**, 13-24 (2004).
150. Sieker, F., Springer, S. & Zacharias, M. Comparative molecular dynamics analysis of tapasin-dependent and -independent MHC class I alleles. *Protein Sci.* **16**, 299-308 (2007).
151. Zacharias, M. & Springer, S. Conformational flexibility of the MHC class I alpha(1)-alpha(2) domain in peptide bound and free states: A molecular dynamics simulation study. *Biophys. J.* **87**, 2203-2214 (2004).
152. Sieker, F., Straatsma, T., Springer, S. & Zacharias, M. Differential tapasin dependence of MHC class I molecules correlates with conformational changes upon peptide dissociation: a molecular dynamics simulation study. *Mol. Immunol.* **45**, 3714-3722 (2008).
153. Garstka, M.A. et al. Tapasin dependence of major histocompatibility complex class I molecules correlates with their conformational flexibility. *FASEB J.* **25**, 3989-3998 (2011).
154. Keinast, A., Preuss, M., Winkler, M. & Dick, T.P. Redox regulation of peptide receptivity of major histocompatibility complex class I molecules by ERp57 and tapasin. *Nat. Immunol.* **8**, 864-872 (2007).

155. Bailey, A. et al. Selector function of MHC I molecules is determined by protein plasticity. *Sci Rep* **5**(2015).
156. Fisette, O., Wingbermuhle, S., Tampe, R. & Schafer, L. Molecular mechanism of peptide editing in the tapasin-MHC I complex. *Sci Rep* **6**, 19085 (2016).
157. Fabian, H. et al. HLA-B27 subtypes differentially associated with disease exhibit conformational differences in solution. *J. Mol. Biol.* **376**, 798-810 (2008).
158. Narzi, D. et al. Dynamical characterisation of two differentially disease associated MHC class I proteins in complex with viral and self-peptides. *J. Mol. Biol.* **415**, 429-442 (2012).
159. Badrinath, S. et al. Position 156 influences the peptide repertoire and tapasin dependency of human leukocyte antigen B*44 allotypes. *Haematologica* **97**, 98-106 (2012).
160. Park, B., Lee, S., Kim, E. & Ahn, K. A single polymorphic residue within the peptide-binding cleft of MHC class I molecules determines spectrum of tapasin dependence. *J. Immunol.* **170**, 961-968 (2003).
161. Bailey, A., van Hateren, A., Elliott, T. & Werner, J. Two polymorphisms facilitate differences in plasticity between two chicken Major Histocompatibility Complex Class I proteins. *PLOS One* **9**, e89657 (2014).
162. Schneeweiss, C., Garstka, M., Smith, J., Hutt, M.-T. & Springer, S. The mechanism of action of tapasin in the peptide exchange on MHC class I molecules determined from kinetics simulation studies. *Mol. Immunol.* **46**, 2054-2063 (2009).
163. Elliott, T. & Williams, A. The optimization of peptide cargo bound to MHC class I molecules by the peptide-loading complex. *Immunol. Rev.* **207**, 89-99 (2005).
164. Bailey, A. University of Southampton (2013).
165. Mage, M. et al. A structural and molecular dynamics approach to understanding the peptide-receptive transition state of MHC-I molecules. *Mol. Immunol.* **55**, 123-125 (2013).
166. Simone, L., Tuli, A., Simone, P., Wang, X. & Solheim, J.C. Analysis of major histocompatibility complex class I folding: novel insights into intermediate forms. *Tissue Antigens* **79**, 249-262 (2012).
167. Catipovic, B., Dalporto, J., Mage, M., Johansen, T. & Schneck, J. Major histocompatiblity complex conformational epitopes are peptide specific. *J. Exp. Med.* **176**, 1611-1618 (1992).
168. Catipovic, B. et al. Analysis of the structure of empty and peptide-loaded major histocompatibility complex-molecules at the cell-surface. *J. Exp. Med.* **180**, 1753-1761 (1994).
169. Gakamsky, D., Davis, D., Strominger, J.L. & Pecht, I. Assembly and dissociation of human leukocyte antigen (HLA)-A2 studied by real-time fluorescence resonance energy transfer. *Biochemistry* **39**, 11163-11169 (2000).
170. Weaver, J., Chaves, F. & Sant, A. Abortive activation of CD4 T cell responses during competitive priming in vivo. *Proc. Natl. Acad. Sci. USA* **106**, 8647-8652 (2009).
171. van Hateren, A. et al. The cell biology of major histocompatibility complex class I assembly: towards a molecular understanding. *Tissue Antigens* **76**, 259-275 (2010).
172. Hore, P. *Nuclear Magnetic Resonance*, (Oxford University Press, 1995).
173. Hore, P., Jones, J. & Wimperis, S. *NMR: The Toolkit*, (Oxford University Press, 2000).
174. Levitt, M. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, (Wiley, 2008).
175. NobelMedia. The Nobel Prize in Physics 1952. *nobelprize.org/nobel_prizes/physics/laureates/1952* (2016).

176. Proctor, W. Reminiscences of the early days of nuclear magnetic resonance at Stanford University. (ed. Sykora, S.) (10.3247/SL1Nmr05.002, ~1967).
177. Wuthrich, K. NMR - this other method for protein and nucleic acid structure determination. *Acta Cryst.* **51**, 249-270 (1995).
178. Bax, A. & Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol.* **15**, 563-570 (2005).
179. Higman, V. Protein NMR - a practical guide. *protein-nmr.org.uk* (2012).
180. Schleucher, J. et al. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed-field gradients. *J. Biomol. NMR* **4**, 301-306 (1994).
181. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shifft anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366-12371 (1997).
182. Meiboom, S. & Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* **29**, 688-691 (1958).
183. Achour, A. et al. Structural basis of the differential stability and receptor specificity of H-2Db in complex with murine versus human b2-Microglobulin. *J. Mol. Biol.* **356**, 382-396 (2006).
184. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293 (1995).
185. Vranken, W. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696 (2005).
186. Dosset, P., Hus, J., Marion, D. & Blackledge, M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR* **20**, 223-231 (2001).
187. Petterson, E. et al. UCSF Chimera - a visualisation system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
188. Taylor, G. University of Southampton (2011).
189. Kay, L., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins *J. Magn. Reson.* **89**, 496-514 (1990).
190. Yang, D. & Kay, L. Improved 1HN-detected triple resonance TROSY-based experiments. *J. Biomol. NMR* **13**, 3-10 (1999).
191. Wishart, D., Bigam, C., Holm, A., Hodges, R. & Sykes, B. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbour effects. *J. Biomol. NMR* **5**, 67-81 (1995).
192. Noor, E. & Wishart, D. CSI 2.0: a significantly improved version of the Chemical Shift Index. *J. Biomol. NMR* **60**, 131-146 (2014).
193. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227-241 (2013).
194. Maltsev, A., Ying, J. & Bax, A. Deuterium isotope shifts for bacbone 1H, 15N and 13C nuclei in intrinsically disordered protein a-synuclein. *J. Biomol. NMR* **54**, 181-191 (2012).
195. Lian, L.-Y. & Middleton, D. Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Prog. Nucl. Mag. Res. Sp.* **39**, 171-190 (2001).
196. Yamazaki, T., Lee, W., Arrowsmith, C., Muhandiram, D. & Kay, L. A suite of triple resonance NMR experiments for the backbone assignment of 15N, 13C, 2H labeled proteins with high sensitivity. *J. Am. Chem. Soc* **116**, 11655-11666 (1994).

197. Meissner, A. & Sorensen, O. Three-dimensional protein NMR TROSY-type N-15-resolved (HHN)-H-1-H-N_1 NOESY spectra with diagonal peak suppression. *J. Magn. Reson.* **142**, 195-198 (2000).
198. Han, B., Liu, Y., Ginzinger, S. & Wishart, D. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43-57 (2011).
199. Robustelli, P., Stafford, K. & Palmer, A. Interpreting protein structural dynamics from NMR chemical shifts. *J. Am. Chem. Soc* **134**, 6365-6374 (2012).
200. Tugarinov, V. & Kay, L. Quantitative NMR Studies of High Molecular Weight Proteins: Application to Domain Orientation and Ligand Binding in the 723 Residue Enzyme Malate Synthase G. *J. Mol. Biol.* **327**, 1121-1133 (2003).
201. Delaglio, F., Kontaxis, G. & Bax, A. Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc* **122**, 2142-2143 (2000).
202. Hus, J., Marion, D. & Blackledge, M. De novo determination of protein structure by NMR using orientational and long-range order restraints. *J. Mol. Biol.* **298**, 927-936 (2000).
203. Tolman, J. & Ruan, K. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem. Rev.* **106**, 1720-1736 (2006).
204. Ulmer, T., Ramirez, B., Delaglio, F. & Bax, A. Evaluation of Backbone Proton Positions and Dynamics in a Small Protein by Liquid Crystal NMR Spectroscopy. *J. Am. Chem. Soc* **125**, 9179-9191 (2003).
205. Clore, G. & Schwieters, C. Concordance of residual dipolar couplings, backbone order parameters and crystallographic B-factors for a small a/b protein: a unified picture of high probability, fast atomic motions in proteins. *J. Mol. Biol.* **355**, 879-886 (2006).
206. Clore, G. & Garrett, D. R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures. *J. Am. Chem. Soc* **121**, 9008-9012 (1999).
207. Zweckstetter, M. & Bax, A. Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J. Biomol. NMR* **23**, 127-137 (2002).
208. Tycko, R., Blanco, F. & Ishii, Y. Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *J. Am. Chem. Soc* **122**, 9340-9341 (2000).
209. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-1114 (1997).
210. Zweckstetter, M. & Bax, A. Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J. Biomol. NMR* **20**, 365-377 (2001).
211. Tjandra, N., Grzesiek, S. & Bax, A. Magnetic field dependence of nitrogen-proton J splittings in 15N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc* **118**, 6264-6272 (1996).
212. Fitzkee, N. & Bax, A. Facile measurement of 1H-15N residual dipolar couplings in larger perdeuterated proteins. *J. Biomol. NMR* **48**, 65-70 (2010).
213. Cornilescu, G., Marquardt, J., Ottiger, M. & Bax, A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc* **120**, 6836-6837 (1998).
214. Rinaldelli, M. et al. Simultaneous use of solution NMR and X-ray data in REFMAC5 for joint refinement/detection of structure differences. *Acta. Cryst. D* **70**, 958-967 (2014).
215. Wassenaar, T. et al. WeNMR: Structural biology on the grid. *J. Grid Computing* **10**, 743-767 (2012).

216. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022-1025 (1997).
217. Winn, M. et al. Overview of the CCP4 suite and current developments. *Acta. Cryst. D* **67**, 235-242 (2011).
218. McCoy, A. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658-674 (2007).
219. Chen, V. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst.* **66**, 12-21 (2010).
220. Zhu, G., Xia, Y., Nicholson, L. & Sze, K. Protein dynamics measurements by TROSY-based NMR experiments. *J. Magn. Reson.* **143**, 423-426 (2000).
221. de la Torre, G., Huertas, M. & Carrasco, B. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138-146 (2000).
222. Lipari, G. & Szabo, A. Model-Free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc* **104**, 4546-4559 (1982).
223. Lipari, G. & Szabo, A. Model-Free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc* **104**, 4559-4570 (1982).
224. Clore, G. et al. Deviations from the simple two parameter model free approach to the interpretation of 15N nuclear magentic relaxation of proteins. *J Am Chem Soc* **112**, 4989-4991 (1990).
225. Mandel, A., Akke, M. & Palmer, A. Backbone dynamics of Escherichia-coli ribonuclease HI - correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**, 144-163 (1995).
226. Cole, R. & Loria, J.P. FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR* **26**, 203-213 (2003).
227. Dempsey, C. Hydrogen exchange in peptides and proteins using NMR spectroscopy. *Prog. Nucl. Mag. Res. Sp.* **39**, 135-170 (2001).
228. Hwang, T.-L., van Zijl, P. & Mori, S. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **11**, 221-226 (1998).
229. Pellecchia, M. et al. SEA-TROSY (solvent exposed amides with TROSY): A method to resolve the problem of spectral overlap in very large proteins. *J. Am. Chem. Soc* **123**, 4633-4634 (2001).
230. Kay, L. NMR studies of protein structure and dynamics. *J. Magn. Reson.* **173**, 193-207 (2005).
231. Turner, S. et al. Favourite flavors of surfaces. *Nat. Immunol.* **6**, 382-389 (2005).
232. Ostrov, D. et al. How H13 histocompatibility peptides differing by a single methyl group and lacking conventional MHC binding anchor motifs determine self-nonself discrimination. *J. Immunol.* **168**(2002).
233. Zhao, R., Loftus, D., Appella, E. & Collins, E. Structural evidence of T cell xeno-reactivity in the absence of molecular mimicry. *J. Exp. Med.* **189**, 359-370 (1999).
234. Meijers, R. et al. Crystal structures of murine MHC class I H-2D-b and K-b molecules in complex with CTL epitopes from influenza A virus: Implications for TCR repertoire selection and immunodominance. *J. Mol. Biol.* **345**, 1099-1110 (2005).
235. Walvoort, M., van der Marel, G., Overkleeft, H. & Codee, J. On the reactivity and selectivity of donor glycosides in glycochemistry and glycobiology: trapped covalent intermediates. *Chem Sci* **4**, 897-906 (2013).

236. Schulze, M. & Wucherpfennig, K. The mechanism of HLA-DM induced peptide exchange in the MHC class II antigen presentation pathway. *Curr. Opin. Immunol.* **24**, 105-111 (2012).