Structural and Functional Studies of Cell Surface Receptors

Ellen C. Border

Trinity College

Division of Structural Biology

University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of *Doctor of Philosophy*

Trinity Term 2012

Structural and Functional Studies of Cell Surface Receptors

Ellen C. Border

Trinity College and Division of Structural Biology

A thesis submitted in partial fulfilment of the requirements for the degree of *Doctor of Philosophy*

University of Oxford, Trinity Term 2012

Abstract

Receptor proteins on the surfaces of cells equip them to communicate with each other and to sense and interact with their environment. One receptor family, the αβ T-cell receptors (TCRs), allow T lymphocytes to detect and respond to pathogens *via* interactions with antigen-presenting major histocompatibility complex (MHC) molecules on target cells. A degree of TCR cross-reactivity (e.g. through structural similarity between peptide-MHC (pMHC) complexes) is essential to account for all possible pathogens, but can also lead to the misinterpretation of self antigens as foreign, and thereby elicit an autoimmune response, resulting in diseases such as multiple sclerosis (MS). Structural studies of pMHC and TCR-pMHC complexes have been key to developing of an understanding of the molecular basis of TCR cross-reactivity, and the first strand of this thesis describes attempts to express and purify a highly cross-reactive MS patient-derived TCR for structural characterisation. The formation, purification and crystallisation of a TCR-self pMHC complex including another autoreactive TCR is also described.

Another family of receptors, the fibronectin leucine-rich transmembrane proteins (FLRTs), has been implicated in roles in embryonic development including cell sorting and adhesion. In the second strand of this thesis, the nature of homotypic interactions between FLRTs, which may underlie adhesion between FLRT-transfected cells, is investigated. Biophysical analyses demonstrate that these interactions may be mediated by the extracellular leucine-rich repeat (LRR) domain, and crystal structures of all three FLRT LRR domains suggest how interactions between them may underlie FLRT self-association at the cell surface. Residues which contribute to these interactions are conserved across different members of the FLRT family and different species. These findings confirm that FLRTs induce homotypic cell-cell adhesion, and suggest that this behaviour is mediated by self-association at the cell surface *via* the LRR domain.

Acknowledgements

Firstly, I would like to sincerely thank my supervisors, Maria Harkiolaki and Yvonne Jones, for the opportunity to study in Strubi and to develop as a scientist. Their support and encouragement have been invaluable at all stages of my doctoral studies. I would like to thank Maria especially for her patience and practical expertise, and Yvonne for all her suggestions and guidance.

I have also been privileged to be able to work with some fantastic people within Strubi and elsewhere. Thanks go especially to Radu Aricescu for some very interesting discussions, and to Elena Seiradake for her patience and help, and for sharing my excitement over MALS results! I am particularly indebted to Jeff McIlhinney for his expertise with confocal microscopy and FRET, and for his enthusiasm for the FLRT project. Thanks must also go to Margaret Jones, for running the lab so smoothly, to Weixian Lu for being a friendly face during hours of tissue culture, and to Robert Gilbert, Geoff Sutton, Tom Walter and Karl Harlos for readily sharing their wealth of knowledge. There are too many people to name who made the lab such a great place to work, but thanks especially to Amber, Ben, Charlotte and Tomas for countless small favours and some comic relief!

I wouldn't be where I am without the unending support of my family and friends. Special thanks must go to my Mum, for always cheering me on, for picking me up when I'm down, and for being so interested in my work. Similarly, I want to thank my Dad, who incredibly has read every word of this thesis, for being so supportive and not too critical with his corrections! Thank you also to my wonderful friends, especially Babette for oodles of *hygge*, and to Mel for her valiant attempts at keeping me sane for a decade, as both a biochemist and a bride.

Finally, a special thank you is saved for my husband David, who has had to put up with a lot over the last four years and has borne it all with grace, positivity and unwavering faith and pride. I couldn't have done it without you.

"Nobody said it was easy

No-one ever said it would be so hard."

Coldplay - "The Scientist"

Table of Contents

Abbreviations

A_{280nm}	Light absorbance at 280 nm
AcMNPV	*Autographa californica* nucleopolyhedrovirus
AMIGO	Amphoterin-induced gene and ORF
APBS	Adaptive Poisson-Boltzmann solver
APC	Antigen presenting cell
AU	Absorbance units
AUC	Analytical ultracentrifugation
Bis-Tris	1,3-bis(tris(hydroxymethyl)methylamino)propane
BSA	Bovine serum albumin
C (domain)	Constant domain
CAPS	3-(cyclohexylamino)-1-propane sulphonic acid
CCD	Charge-coupled device
CCP4	Collaborative computational project, number 4
CD	Cytoplasmic domain
CD2/45	Cluster of differentiation-2/45
cDNA	Complementary deoxyribonucleic acid
CDR	Complementarity-determining region
CLIP	Class II-associated invariant chain peptide
CNS	Central nervous system
CTL	Cytotoxic T lymphocyte
CV	Column volumes
DLS	Diamond Light Source, Oxfordshire
DM	HLA-DM
DMEM	Dulbecco's modified Eagle's medium

DNA	Deoxyribonucleic acid
DQ6	HLA-DQ6 (DQB1*0602)
DR2a	HLA-DR2 (DRB5*0101)
DR2b	HLA-DR2 (DRB1*1501)
DTT	Dithiothreitol
EAE	Experimental autoimmune encephalomyelitis
EBV	Epstein-Barr virus
ECD	Extracellular domain
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetra-acetic acid
EMBL-EBI	European Molecular Biology Laboratory - European Bioinformatics Institute
ER	Endoplasmic reticulum
ERK	Extracellular regulated kinase
ESRF	European Synchrotron Radiation Facility, France
Fab	Fragment antigen binding (antibody fragment)
FACS	Fluorescence-activated cell sorting
FCS	Foetal calf serum
FGF(R)	Fibroblast growth factor (receptor)
FLRT	Fibronectin leucine-rich transmembrane protein
$FLRT_{ECD}$	FLRT extracellular domain (construct)
$FLRT_{LRR}$	FLRT LRR domain (construct)
FNIII	Fibronectin type III (domain)
FPLC	Fast protein liquid chromatography

FRET	Förster (or fluorescence) resonance energy transfer
GlcNAc	*N*-acetyl-glucosamine
GnTI	*N*-acetylglucosaminyltransferase I enzyme
HA	(Influenza) haemagglutinin (N-terminal tag)
HEK 293S	Human embryonic kidney 293 cells deficient in GnTI
HEK 293T	Human embryonic kidney 293 cells expressing SV40 type T antigen
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLA	Human leukocyte antigen
Hy	Hy.2E11 TCR
Ig	Immunoglobulin (-like domain)
IMAC	Immobilised metal affinity chromatography
IMGT	Immunogenetics online information system
IPTG	Isopropyl β-D-1 thiogalactopyranoside
K_d	Dissociation constant
LB	Luria-Bertani medium
LRR	Leucine-rich repeat
MALS	Multi-angle light scattering
Man	Mannose
MAP(K)	Mitogen-activated protein (kinase)
MBP	Myelin basic protein
mCerulean	monoCerulean C-terminal fluorescent tag
MCS	Multiple cloning site
MES	2-(*N*-morpholino)ethanesulfonic acid
MHC	Major histocompatibility complex
mRNA	Messenger ribonucleic acid
MS	Multiple sclerosis

MS49	MS49 D3 TCR
mVenus	monoVenus C-terminal fluorescent tag
Mw	Molecular weight
N-CAM	Neural cell adhesion molecule
NCS	Non-crystallographic symmetry
NGL	Netrin-G ligand
NP-40	Nonyl phenoxylpolyethoxylethanol
Ob	Ob.1A12 TCR
$OD_{(600nm)}$	Optical density (at 600 nm)
ORF	Open reading frame
PAPC	Paraxial protocadherin
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEI	Polyethylenimine
pI	Isoelectric point
PIPES	Piperazine-*N*,*N′*-bis(2-ethanesulfonic acid)
PISA	Protein interfaces, surfaces and assemblies (server)
pMHC	Peptide-MHC complex
PNGase F	Peptide: *N*-glycosidase F
PNS	Peripheral nervous system
rbs	Ribosome binding site
(C_α) rmsd	(Alpha carbon) root mean squared deviation
ROI	Region of interest (in FRET experiments)
RONN	Regional order neural network (server)

(k)rpm	(Thousand) revolutions per minute
RPTP	Receptor protein tyrosine phosphatase
S75/S200	Superdex-75/-200 (column)
SALM	Synaptic adhesion-like molecule
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC-MALS	Size exclusion chromatography-MALS
Sf9	*Spodoptera frugiperda* (cells)
SHP	Structural homology program
SOC	Super optimal broth
SP	Signal peptide
TAP	Transporter associated with antigen processing
TCR	T cell receptor
T_H	$CD4^+$ T helper cells
TLR	Toll-like receptor
TM	Transmembrane (domain)
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
TrkC	Tyrosine receptor kinase-C
V (domain)	Variable domain
v/v	Volume by volume
w/v	Weight by volume

Chapter 1 Introduction

Chapter 1 Introduction

1.1 Background - the interactions of cells with their environment

From single-celled prokaryotes to complex multi-cellular organisms such as humans, the ability of cells to sense and interact with their environment is essential for sustaining life. The cell membrane acts as an interface between the cell and the exterior environment, and the receptor proteins associated with it are able to recognise a multitude of biochemical signals and pass them from the cell surface, across the membrane, and into the cell. Once these messages, ranging from the direction of a bacterial food source to a nerve impulse transmitted across a synapse, have been relayed into the cytoplasm, a series of signal transduction pathways enables the cell to integrate and process them and to determine an appropriate response.

Elaborate communication mechanisms are essential between the cells of a multicellular organism if they are to be able to govern their behaviour for the benefit of the organism as a whole. Due to their ubiquity in complex and essential processes, many of the cell surface receptors and signalling complexes that mediate these mechanisms represent proven and potential targets for therapeutic intervention in a range of diseases. Two examples of cell surface receptors which are involved in immunity and development are the focus of this thesis.

T-cell receptors (TCRs) on the surface of T lymphocytes of the adaptive immune system recognise antigen fragments of pathogenic origin, presented in the context of MHC molecules on the surface of target cells, in order to detect infection and activate the body's defences. However, although T cells are trained to ignore antigens originating from the host organism, the necessity of a degree of cross-reactivity in the

TCR repertoire (Mason, 1998) means that a non-self antigen displaying significant similarity to a self antigen can potentially trigger an immune response that will ultimately target healthy host tissue. This can lead to autoimmune pathologies such as multiple sclerosis (MS), rheumatoid arthritis and Type I diabetes (Oldstone, 1987; Jones et al., 2006).

The fibronectin leucine-rich transmembrane (FLRT) protein family are plasma membrane glycoproteins with reported roles in homotypic cell adhesion and fibroblast growth factor- (FGF-) mediated signalling (Lacy et al., 1999; Karaulanov et al., 2006). Studies in knockout mice, showing that embryos lacking members of this family suffer developmental defects (Maretto et al., 2008; Müller et al., 2011), point to a role for FLRTs in early development where changes in the way cells adhere to, and interact with, each other are essential to the development of an embryo into a functional organism.

1.2 The adaptive immune system and autoimmunity

The immune system comprises the cells and molecules which protect from, and respond to, bacterial, viral, fungal and parasitic infections, in order to maintain the health of the organism. In common with many higher eukaryotes, the mammalian immune system is broadly divided into the innate immune system, the adaptive immune system and the complement system, which each have different purposes. However, these components are intricately linked to one another, and all function to recognise and eliminate different pathogens. An excellent introduction to the principles of immunology can be found in (Murphy et al., 2012).

The focus of the research detailed in one strand of this thesis is the adaptive immune system, which provides highly antigen-specific recognition, and allows the

establishment of immunological memory, where exposure to a foreign antigen enhances the ability of the organism to respond to that antigen again. This protection lasts for the lifetime of the organism, and subsequent responses are usually more rapid and more aggressive. The adaptive immune response is carried out jointly by the twin arms of humoral and cellular immunity, with the humoral arm mediated by antibodies produced by B lymphocytes, which circulate in the blood and bind to extracellular antigens. To access intracellular bacteria and viruses, however, cellular immunity is required to recognise and destroy infected cells. T lymphocytes have highly variable membrane-bound antigen receptors called T cell receptors (TCRs), which can recognise their cognate antigen presented on the surface of infected cells, or dedicated antigen presenting cells (APCs), in the context of Major Histocompatibility Complex (MHC) molecules. The structural basis of this recognition event, and how inappropriate recognition of self antigens can lead to autoimmune pathologies, is the focus of one strand of this thesis.

1.2.1 MHC class II molecules and antigen presentation

The MHC region of the genome (chromosome 6 in humans, chromosome 17 in mice) is highly genetically diverse, and contains loci for MHC class I, class II and class III alleles. Class I and II genes encode two classes of MHC molecules which present antigen-derived peptides, acquired *via* distinct intracellular pathways, to two subsets of αβ T lymphocytes (see Section 1.2.2), and class III genes encode other immune proteins including cytokines and components of the complement response. Genes encoding the MHC class II molecules are located in the human leukocyte antigen (HLA)-D region, and are grouped into HLA-DM, HLA-DP, HLA-DQ and HLA-DR loci in humans, and I-A and I-E loci in mice.

MHC class I molecules are found on all nucleated cells, and are loaded in the endoplasmic reticulum (ER) with peptides derived from proteasomal degradation of cytosolic proteins. Peptides are transported into the ER compartment by the Transporter associated with Antigen Processing (TAP), and complete peptide-MHC class I complexes are transported to the cell surface, allowing surveillance for cytosolic pathogens such as viruses. In contrast, MHC class II molecules are only expressed on the surface of professional APCs including dendritic cells, B lymphocytes and macrophages, and present peptide fragments deriving from extracellular antigens taken up through the endocytic pathway. Endocytosed antigenic proteins are delivered to endosomes, which become increasingly acidic as they mature and progress into the interior of the cell, eventually fusing with lysosomes; this decrease in pH activates acid proteases within the vesicles, which degrade antigens enzymatically into peptide fragments. Folded empty MHC class II molecules (Stern and Wiley, 1992) initially bind the invariant chain in the ER, before this is proteolytically cleaved to the class II-associated invariant chain peptide (CLIP) in endosomal compartments. Nascent CLIP-class II complexes then pass through acidified compartments containing the protease-digested peptide fragments *en route* to the cell surface, and CLIP is exchanged for an antigenic peptide.

This peptide exchange is catalysed by HLA-DM (Trombetta and Mellman, 2005), a specialised class II-like molecule which binds preferentially to empty MHC class II molecules, promoting dissociation of CLIP (or weakly-bound antigenic peptides) from the peptide binding groove. This permits formation of more stable antigenic peptide-MHC (pMHC) complexes which are more resistant to HLA-DM action and which, once transported to the cell surface, can persist long enough to activate circulating T cells. Recognition of pMHC complexes on the surface of dendritic cells

by naive $CD4^+$ T lymphocytes (see Section 1.2.2) activates the T cell to a mature cytokine-secreting state, in which it can activate B cells and macrophages to trigger responses aimed at eliminating the source of the antigen. B cells are activated to produce soluble antibodies which will bind any circulating antigen such as bacterial toxins, while macrophages are activated to destroy any phagocytosed microbes.

Class II molecules are made up of an α chain and a β chain; the human MHC class II locus includes at least 6 α and 10 β chains. These assemble non-covalently to form a membrane-distal extracellular peptide-binding groove between two α-helices with an antiparallel β-sheet base, and two membrane-proximal Ig-like domains (see Figure 1.1). Residues within the peptide binding groove are highly polymorphic, and molecules encoded by a single MHC allele can bind up to 2000 different antigenic peptides (Hunt et al., 1992). The class II binding groove is open-ended, enabling it to bind longer peptides (10-30 amino acid residues) than the closed-ended class I binding groove (Chicz et al., 1992; Hunt et al., 1992), and is composed of a series of pockets which bind amino acid side chains to govern the peptide specificity, as determined by allele-specific polymorphisms localised to these pockets. The peptide main chain atoms are also bound into the groove along the length of the peptide by hydrogen bond interactions with non-polymorphic MHC residues (Brown et al., 1993; Stern et al., 1994). The highly polymorphic nature of MHC class II molecules determines the subset of peptides that an individual's APCs can present for inspection by $CD4^+$ T cells (see Section 1.2.2), and clear associations between certain MHC class II molecules and susceptibility to, or protection from, various autoimmune diseases have been established (Hillert and Olerup, 1993; Jones et al., 2006).

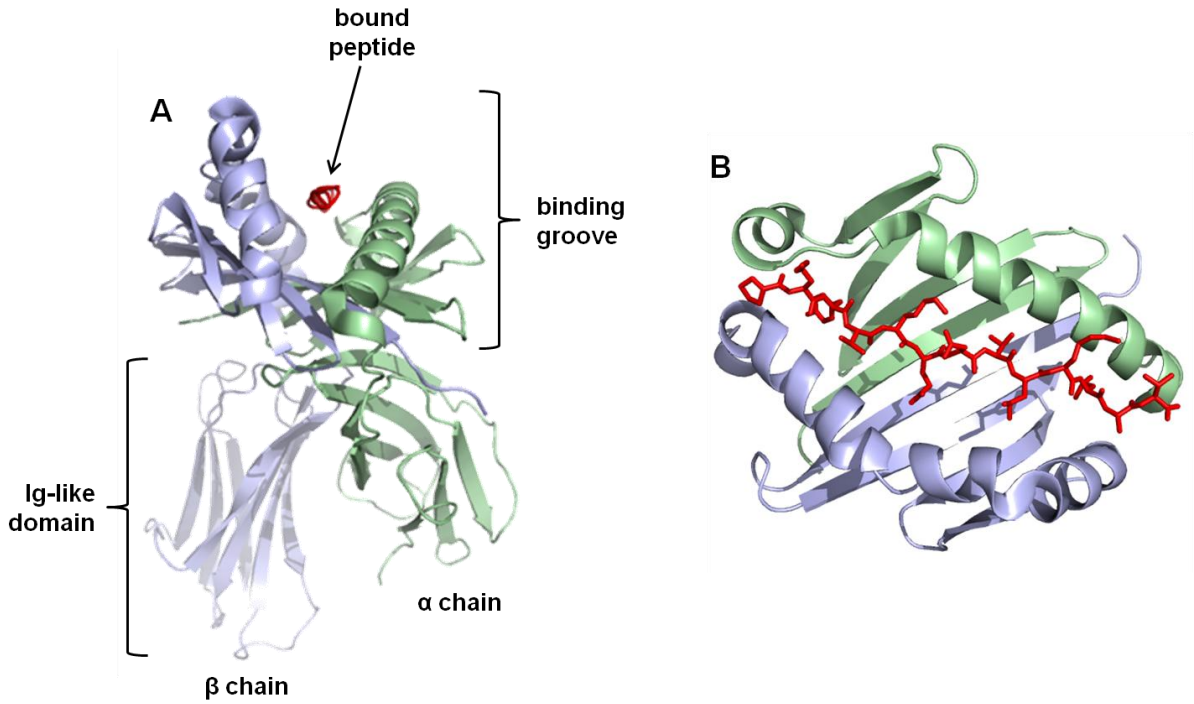


Figure 1.1 MHC class II molecule structure and peptide binding groove, as exemplified by HLA-DR1 (PDB code 1DLH (Stern et al., 1994)); the α chain is shown in green and the β chain in blue, while the bound influenza virus peptide is shown in red. (A) The MHC class II molecule viewed from the side, showing the peptide bound in the binding groove; (B) a close up of the binding groove and peptide viewed from above, with the peptide N-terminus on the left and C-terminus on the right.

1.2.2 T cell receptors and antigen recognition

There are two types of TCR - αβ and γδ TCRs; both assemble with the CD3 complex on the surface of T lymphocytes to form a TCR complex capable of signalling a recognition event across the membrane to the inside of the cell. αβ T cells can be classified into two major populations by the association of the αβ TCR with either the CD4 or CD8 co-receptors. $CD8^+$ cytotoxic T lymphocytes (CTLs) only recognise antigen in the context of MHC class I molecules, and upon recognition, eliminate the presenting cells by cell-mediated cytotoxicity and release of cytotoxic vesicles. $CD4^+$ helper T (T_H) cells are stimulated to secrete cytokines by recognition of antigenic peptides in the context of MHC class II molecules (Section 1.2.1); these cytokines

stimulate proliferation and differentiation of T and B lymphocytes, macrophages and other leukocytes. T_H cells can be further subdivided by their cytokine profile and role into a number of subsets.

αβ TCRs are heterodimeric, consisting of one α and one β chain, covalently liked by an inter-chain disulphide bond. In this way, the molecular structure of TCRs closely resembles that of the Fragment Antigen Binding (Fab) fragment of antibodies produced by B cells, which comprise a complete light chain paired with the N-terminal portion of a heavy chain *via* a disulphide bond. This similarity is despite the different antigen recognition strategies of the two types of molecules; antibodies recognise native antigen directly, while TCRs recognise cell-surface bound composite pMHC ligands. However, while Fab fragments (and indeed antibodies) are soluble, TCRs are localised to the cell membrane; each TCR α or β chain has a single transmembrane region and a short intracellular tail, with membrane-distal variable and membrane-proximal constant Ig-like domains in the extracellular region (see Figure 1.2). Fab fragments (which, like TCRs, have a single antigen recognition site) and unfragmented antibodies (which have two) bind antigen using the variable regions of the light and heavy chains, and similarly, the TCR antigen recognition site is made up of the variable domains (V_α and V_β) of each chain. The variable regions of TCRs and antibodies each have hypervariable loops (three per α/β or light/heavy chain) deriving from the complementarity-determining regions (CDRs) where variability between different antibody and TCR molecules is concentrated.

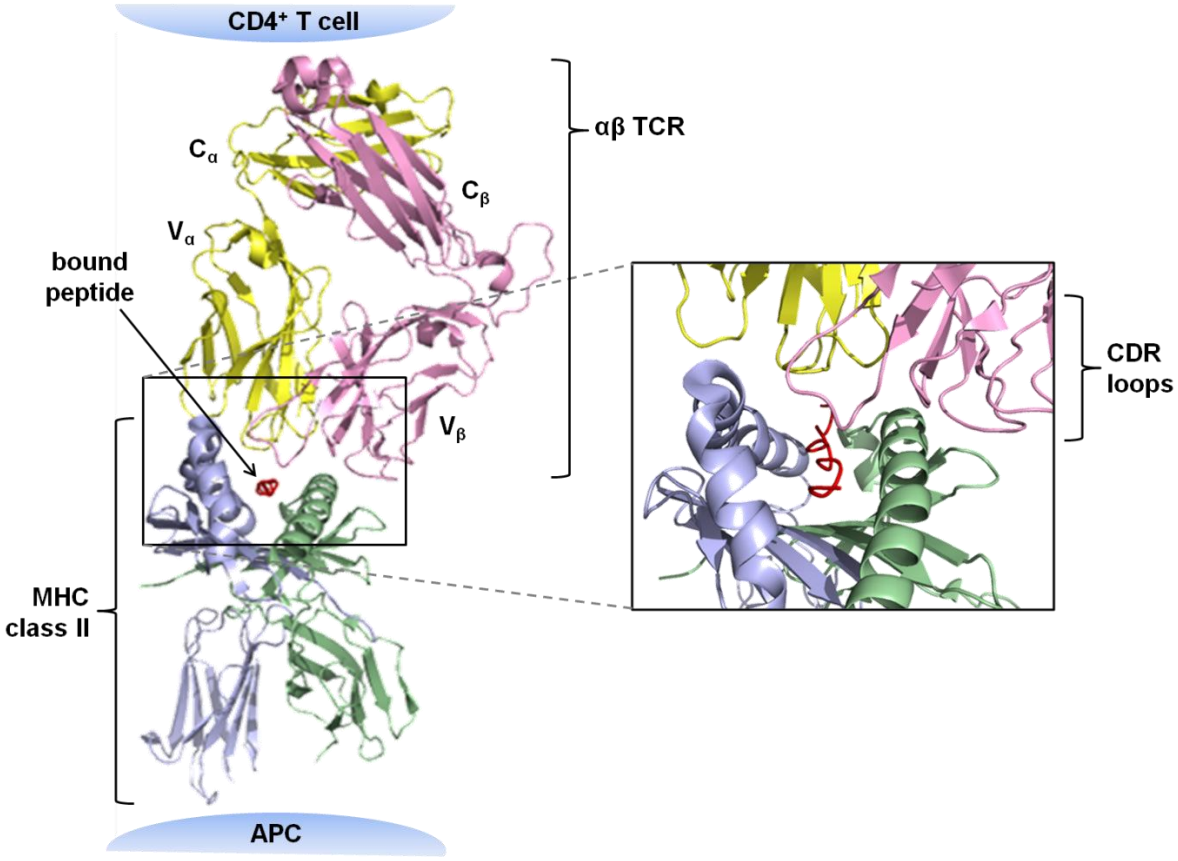


Figure 1.2 αβ TCR recognition of a pMHC class II complex (PDB code 1FYT (Hennecke et al., 2000)); the MHC molecule is HLA-DR1, as in Figure 1.1, presenting an influenza peptide. The colouring of pMHC is as for Figure 1.1, the TCR α chain is shown in yellow and the TCR β chain in pink; $V_{\alpha/\beta}$ and $C_{\alpha/\beta}$ are the variable and constant domains respectively. A close-up of the MHC binding groove (right panel) indicates the importance of the TCR CDR loops in specificity, as they are the primary parts of the TCR to interact with the compound antigenic pMHC ligand.

The generation of considerable diversity in the antibody and TCR repertoires is necessary to allow every potential pathogen to be recognised, and this is achieved *via* a variety of mechanisms, including somatic recombination of homologous variable (V), diversity (D), joining (J) and constant (C) gene segments. While antibody genes contain less diversity at junctions between gene segments, they are further diversified in the variable region after rearrangement through somatic hypermutation. TCR genes do not undergo this process, but recombination produces greater junction diversity, resulting in molecules with the highest level of diversity at

CDR3. Crystal structures of αβ TCRs in complex with pMHC complexes confirm the importance of this diversity to receptor-specific interactions; typically, the CDR3 loops form the majority of interactions with the peptide, while the other loops bind the helices bordering the MHC binding groove. However, as indicated in Figure 1.2, the variable domains of the TCR recognise the presented surface of the pMHC antigen overall, rather than either the peptide or the MHC alone, so that the combination of both and the way they associate to form a compound ligand determines whether they are recognised by a TCR.

1.2.3 Autoimmunity and TCR cross-reactivity

It is important for the immune system to be able to discriminate between self and foreign antigens, so that invading pathogens can be ferociously attacked while healthy host tissue is recognised as "safe", and remains unharmed. This is immunological tolerance, which is established during negative thymic selection in the developing immune system; T cells which bind self pMHC complexes presented on thymic epithelial cells too strongly are deleted to ensure self-reactivity is within safe levels (Hogquist et al., 2005). This centrally established tolerance is maintained in the mature immune system by various regulatory mechanisms (Walker and Abbas, 2002). A loss of this self tolerance can lead to an aberrant autoimmune response against healthy host tissue, typically resulting in inflammation and disease of the targeted tissue.

Through the intricate series of DNA recombination events required to produce a TCR during the development of the immune system (mentioned in Section 1.2.2), a receptor can be produced to defend the host against any pathogen it might encounter in its lifetime. However, due to the size limits of the T cell repertoire, a high level of cross-reactivity with multiple antigenic peptides (e.g. through structural similarity and

molecular mimicry) is essential to account for all possible pathogens (Mason, 1998). Unfortunately, this cross-reactivity might also lead to misinterpretation of innocuous self antigens as structurally-similar pathogenic ones, and thereby elicit a pathogenic autoimmune response, as in the case of diseases including multiple sclerosis (MS), rheumatoid arthritis and type I diabetes (Oldstone, 1987). The connection between molecular mimicry and autoimmune pathogenesis has been particularly well-described in the case of acute enteric *Campylobacter jejuni* infection and Guillain-Barré syndrome (Wim Ang et al., 2004). It appears that ganglioside-like structures on the outer wall of the bacteria trigger a cross-reactive antibody response against host neuronal gangliosides, leading to destruction of peripheral nerves; it may be that other autoimmune diseases, such as MS, can be triggered in a similar way.

The mechanisms for the proposed involvement of molecular mimicry in autoimmune pathogenesis remain enigmatic (Benoist and Mathis, 2001), though several hypotheses exist. One proposes that the peripheral tolerance mechanisms which usually maintain any autoreactive T cells in a quiescent state (Walker and Abbas, 2002) may simply fail in the highly inflammatory environment of a microbial infection (Anderton and Wraith, 2002), allowing activation of the autoreactive T cells and initiation of an autoimmune response. Another posits that recognition of so-called pathogen associated molecular patterns on the surface of pathogens by Toll-like receptors (TLRs) "licenses" APCs, which can then present pathogenic antigens along with inflammatory signals. This could result in high avidity interactions between the licensed APC and any autoreactive T cells, eventually allowing them to break tolerance and react to the mimetic self antigen (Goverman, 2009). Yet another implicates enhanced peptide presentation, as a result of up-regulation of MHC

molecules by interferon-γ produced by $CD4^+$ T cells as part of the initial immune response (Benoist and Mathis, 2001). This may augment and propagate the inflammatory response, resulting in eventual loss of tolerance to the self antigen, and initiation of an autoimmune reaction.

1.2.4 Multiple sclerosis

MS is a common chronic inflammatory, and degenerative, autoimmune disease of the CNS affecting approximately 1-2 million people worldwide (Sospedra and Martin, 2005). It is characterised clinically by inflammatory lesions in the brain accompanied by neuronal demyelination and eventual neuronal and axonal degeneration, caused by aberrant T cell-mediated attack against components of the myelin sheath (e.g. myelin basic protein (MBP)) which encases neuronal axons (Sospedra and Martin, 2005). Resulting symptoms can include perturbation of motor, sensory, neurocognitive and autonomic function, and although the disease is not usually fatal, there is a high socioeconomic cost associated with the necessary palliative patient care. There is, at present, no cure for MS.

The aetiology of MS remains incompletely understood, though what is clear is that there are two main contributing factors required to initiate the disease - genetic susceptibility and environmental triggers. Studies combining genetics and epidemiology have established clear associations between certain MHC class II alleles and susceptibility to MS, particularly the DR2 haplotype (DRB1*1501, DRB5*0101 and DQB1*0602 alleles), which accounts for just under 50 % of the total genetic basis for MS susceptibility (Hillert and Olerup, 1993; Haines et al., 1998; Sawcer et al., 2005). Strong linkage disequilibrium in the DR2 region has meant assigning the relative contributions of each allele to MS susceptibility has proved challenging, though humanised mouse models of the MS-like disease experimental

autoimmune encephalomyelitis (EAE) indicate that DRB1*1501 (DR2b) confers the most risk, with DRB5*0101 (DR2a) appearing to play a modifier role (Gregersen et al., 2006). The contribution of DQB1*0602 (DQ6) to MS risk remains unclear.

Despite the strong genetic contribution, concordance of MS incidence between monozygotic twins is only 20-35 % (Willer et al., 2003), indicating that there is a significant, albeit poorly characterised, environmental contribution to the development of MS (Marrie, 2004; Ebers, 2008). The most avidly investigated environmental trigger has been infection, for example with Epstein Barr virus (EBV), Herpes simplex virus and cytomegalovirus. However, despite extensive study, the mechanisms by which infections can trigger MS are unclear (Sibley et al., 1985).

One hypothesis is that molecular mimicry causes TCR cross-reactivity, so that recognition of a microbial antigen leads to aberrant targeting of a structurally-similar self antigen, and thereby an autoimmune response (see Section 1.2.3 and (Benoist and Mathis, 2001)). This hypothesis requires autoreactive T cells, and MBP-specific CD4+ T cells can indeed be found in MS patients (Olsson et al., 1990). Mice expressing both HLA-DR alleles and an MBP-reactive patient-derived TCR develop MS-like EAE either spontaneously, or upon administration of MBP peptide with adjuvant (Madsen et al., 1999); EAE can also be induced by adoptive transfer of myelin-specific T cells into naive mice (Pettinelli and McFarlin, 1981). However, the requirement for cross-reactivity in the T cell repertoire means all individuals may harbour self-reactive T cells, yet only a small minority develop autoimmune diseases such as MS. This is because under most circumstances, any autoimmune T cells are prevented from initiating an autoimmune response by strict peripheral tolerance mechanisms (Walker and Abbas, 2002); it is in the rare cases where these fail (some

possible mechanisms for this are discussed in Section 1.2.3) and the autoreactive T cells are activated that an autoimmune reaction can occur.

1.2.5 Existing structures of autoimmune pMHC and TCR/pMHC structures

As mentioned in Section 1.2.4, the contribution of certain MHC alleles to the genetic component of risk of MS (and autoimmune disease more widely) is well-established. This may occur *via* slight changes in presentation of self antigens in the developing thymus, influencing central T cell selection, as well as presentation in the periphery, influencing T cell activation. Structural studies, particularly by x-ray crystallography, of both pMHC and TCR-pMHC complexes have been instrumental in gaining an appreciation of the molecular basis of these mechanisms (Jones et al., 2006), and it is possible that this structural understanding could aid design of therapeutic agents to treat autoimmune conditions such as MS. For example, knowledge of the structural basis of aberrant TCR-pMHC recognition might be used to develop agents to prevent the recognition event; in fact, such an approach has already been used to design peptide 15-mers which inhibit binding of the immunodominant epitope of MBP to HLA-DR2 and suppress EAE in humanised mice (Stern et al., 2005). However, a larger database of TCR-pMHC structures is essential to develop and improve such approaches.

Crystal structures of a number of MHC class II molecules which either confer protection from, or susceptibility to, a number of autoimmune diseases have revealed the importance of slight changes in specificity and affinity of peptide binding to the MHC binding groove in determining the diversity of self peptides that can be presented to T cells during thymic selection. For example, the promiscuous DQB1*0602 (DQ6) allele is thought to confer protection from insulin dependent diabetes mellitus by enhancing deletion of more autoreactive T cells during negative

selection, or favouring the positive selection of regulatory T cells, which act to modulate the immune response (Ettinger and Kwok, 1998). Low-affinity peptide binding, meanwhile, can mean the peptide is not MHC-bound for long enough to cause negative selection of autoreactive T cells, as shown for the mouse MHC class II molecule I-A^u bound to an MBP peptide which predisposes animals to EAE (He et al., 2002). Structures of the MS-associated DR2a and DR2b molecules, bound to peptides from EBV (foreign) and MBP (self) respectively, have also provided insight into the potential role of molecular mimicry in autoimmunity; these complexes are structurally similar and are both recognised by an autoreactive T cell (Lang et al., 2002).

Such recognition events between autoreactive T cells and pMHC complexes have been the focus of further crystallographic studies, and have been particularly relevant to our understanding of the structural mechanisms of autoimmunity in MS. Several structures have been determined, involving four human TCRs from MS patients (3A6, Ob.1A12 (Ob), Hy.1B11 and MS2-3C8) and three murine TCRs from the EAE model of MS (Maynard et al., 2005; Feng et al., 2007); details of these structures are summarised in Table 1.1. The 3A6 and Ob TCRs are each restricted by an allele of the HLA-DR2 haplotype (DR2a for 3A6 and DR2b for Ob), while Hy.1B11 is restricted by HLA-DQ1 and MS2-3C8 by HLA-DR4. Ob and Hy.1B11 recognise the immunodominant epitope (residues 85-99) of MBP (Hahn et al., 2005; Sethi et al., 2011), while 3A6 recognises residues 89-101 and MS2-3C8 recognises residues 111-129 (Li et al., 2005; Yin et al., 2011). In all cases except MS2-3C8, the complex structures have shown the human TCRs interacting with pMHC in a non-canonical fashion (see Table 1.1), leading to impaired TCR-peptide binding which could underlie cross-reactivity with foreign and self peptides. In contrast to the

classical centrally-positioned diagonal binding footprints seen in non-autoimmune TCR-pMHC complexes (Hennecke et al., 2000), 3A6 and Ob have binding footprints shifted significantly towards the peptide N-terminus (and in the case of Ob, more orthogonal to the MHC), while the highly tilted binding mode of Hy.1B11 means only one of its CDR loops binds the helices of the MHC binding groove. While the murine TCRs bind with normal topology, the binding groove is only partially occupied by peptide, so the TCR-peptide interaction is still perturbed, as for the human TCRs. This could result in low affinity interactions between autoreactive TCRs and self pMHC complexes - indeed, affinity of 3A6 for MBP-DR2a is estimated to be in the micromolar range, which represents weak binding amongst TCR-pMHC interactions (Li et al., 2005). In the case of MS2-3C8, the low affinity interaction is instead between the MBP peptide and HLA-DR4 (Muraro et al., 1997; Yin et al., 2011). Weak interactions, whether between TCR and pMHC or between peptide and MHC, will destabilise the TCR-peptide-MHC recognition unit, and could result in escape of autoreactive TCRs from negative thymic selection (Wucherpfennig et al., 2009). These TCRs might later be triggered by foreign mimetic antigens to launch an autoimmune response against the self antigen.

Crystal structures have also highlighted the structural underpinnings of autoimmune TCR cross-reactivity. The binding degeneracy of 3A6 results in it being highly cross-reactive, and it can bind a range of peptides varying widely from the immunodominant MBP_{85-99} epitope, including MBP_{89-101}, with which it has been crystallised (Li et al., 2005). Ob TCR has also been crystallised in complex with its cognate MHC class II molecule, DR2b, binding both self (MBP_{85-99}) and foreign (*E. coli*) peptides with low affinity (Hahn et al., 2005; Harkiolaki et al., 2009). The *E. coli* peptide, which has limited sequence homology to MBP, can provoke EAE in a

humanised mice model, providing proof-of-principle that low-affinity TCR-pMHC interactions and structural mimicry may underlie TCR cross-reactivity and initiation of autoimmunity.

Description of TCR-pMHC complex (TCR:MHC(peptide))	PDB code (reference)	Disease involvement	Position of TCR binding footprint on pMHC	TCR binding orientation on pMHC	Comments on pMHC binding	Comments on affinity of overall TCR-pMHC interaction	Comments on TCR cross-reactivity
HA1.7:HLA-DR1 (influenza haemagglutinin$_{306-318}$)	1FYT (Hennecke et al., 2000); see Figure 1.2	Anti-microbial recognition of influenza virus	Central; centred over peptide position 5	Diagonal; TCR-pMHC angle of 70°	-	High affinity; 9 peptide residues (positions -1 to 8) recognsied by TCR	Recognises same peptide in context of both HLA-DR1 and HLA-DR4
3A6:HLA-DR2a(MBP$_{89-101}$)	1ZGL (Li et al., 2005)	From human MS patient	Shifted towards N-terminus; centred over peptide position 2	Diagonal; TCR-pMHC angle of 65°	-	Very low affinity; K_D>200μM	Highly degenerate peptide binding
Ob.1A12:DR2b(MBP$_{85-99}$)	1YMM (Hahn et al., 2005)	From human MS patient	Shifted towards N-terminus; centred over peptide position 2	Orthogonal; TCR-pMHC angle of 110°	-	Low affinity; K_D~40μM	Structure of complex with DR2b binding an *E.coli* peptide also available
Hy.1B11:HLA-DQ1(MBP$_{85-99}$)	3PL6 (Sethi et al., 2011)	From human MS patient	Central; centred over peptide position 5	Highly tilted binding mode (14° compared to HA1.7) in addition to small TCR-pMHC angle of 40°	-	Low affinity; K_D~15μM	Reactive to Herpes simplex virus, adenovirus and *Pseudomonas aeruginosa* peptides.
MS2-3C8:HLA-DR4(MBP$_{111-129}$)	3O6F (Yin et al., 2011)	From human MS patient	Central; centred over peptide position 5	Diagonal; TCR-pMHC angle of 65°	Low affinity peptide-MHC interaction; ~2μM	Relatively high affinity; K_D~5μM However, this was measured with a DR4-MBP$_{111-129}$ ligand engineered using peptide tethering to enhance the affinity, so the true affinity is unlikely to be as high	-
172.10:I-A^u(MBP Ac$_{1-11}$)	1U3H (Maynard et al., 2005)	Murine EAE	Central; centred over peptide position 3 (which is in the location of position 5 in canonical structures)	Diagonal	Partial occupancy of binding groove; N-terminal portion is empty, and only 6 peptide residues (positions 3 to 8) recognised by TCR	K_D~5μM; However, this was measured with an I-A^u-MBP Ac$_{1-11}$ ligand engineered using peptide tethering to enhance the affinity, so the true affinity is unlikely to be as high	-
1934.4:I-A^u(MBP Ac$_{1-11}$)	2PXY (Feng et al., 2007)	Murine EAE	Central; centred over peptide position 3 (which is in the location of position 5 in canonical structures)	Diagonal	Partial occupancy of binding groove; N-terminal portion is empty, and only 6 peptide residues (positions 3 to 8) recognised by TCR	K_D~5μM; However, this was measured with an I-A^u-MBP Ac$_{1-11}$ ligand engineered using peptide tethering to enhance the affinity, so the true affinity is unlikely to be as high	-
cl19:I-A^u(MBP Ac$_{1-11}$)	2Z31 (Feng et al., 2007)	Murine EAE	Central; centred over peptide position 3 (which is in the location of position 5 in canonical structures)	Diagonal	Partial occupancy of binding groove; N-terminal portion is empty, and only 6 peptide residues (positions 3 to 8) recognised by TCR	K_D~5μM; However, this was measured with an I-A^u-MBP Ac$_{1-11}$ ligand engineered using peptide tethering to enhance the affinity, so the true affinity is unlikely to be as high	-

Table 1.1 Summarised details of crystal structures of autoreactive TCR-pMHC complexes solved to date, highlighting differences between these and a canonical anti-microbial TCR-pMHC complex (shaded in blue), which might underlie weak interactions within the complex, allowing escape of the autoreactive TCR from thymic deletion. Information compiled from references in table, as well as (Wucherpfennig and Strominger, 1995; Hennecke and Wiley, 2002; Harkiolaki et al., 2009; Wucherpfennig et al., 2009).

1.3 Overview of the Fibronectin Leucine-Rich Transmembrane family

The three members of the FLRT family, FLRT1, FLRT2 and FLRT3, were initially identified serendipitously while screening a cDNA library for novel extracellular matrix (ECM) proteins expressed in human skeletal muscle (Lacy et al., 1999). FLRT1 was seen to have an mRNA expression pattern restricted to the human brain and kidney, that is maintained from early development through to adulthood, while FLRT2 was found to be abundantly expressed in adult pancreas, and to a lesser extent in skeletal muscle, brain and heart. FLRT3 had the broadest expression pattern, and FLRT3 mRNA was found in multiple tissues including kidney, skeletal muscle, lung and brain (Lacy et al., 1999).

From their amino acid sequence, FLRTs resemble small leucine-rich proteoglycans found in the ECM, some of which act as receptors, signal transducers and participants in cell-cell contact. Hydropathy and sequence analysis indicate the presence of an N-terminal signal sequence (Petersen et al., 2011), as well as a putative membrane-spanning region towards the C-terminus, implying that FLRTs are single-pass transmembrane proteins (Lacy et al., 1999). FLRT3 was found to be a type I membrane protein (Tsuji et al., 2004), and localised to the cell surface (Robinson et al., 2004), with the prediction that FLRT1 and FLRT2 can be similarly characterised. Within the putative extracellular portion, sequence analysis indicates the FLRTs share a common domain structure of 10 leucine-rich repeats bordered by cysteine-rich caps (LRR domain) and a fibronectin type III (FNIII) domain, as shown in Figure 1.3. The LRR is a widespread repeating protein motif which is frequently involved in protein-protein interactions, and is described further in Section 1.5; FNIII domains have an evolutionarily conserved beta sandwich structure approximately

100 amino acids in length, and are found in a wide variety of extracellular proteins, many of which are involved in cell surface binding. The extracellular regions of each FLRT contain multiple potential sites for N-linked glycosylation (Gupta et al., in preparation), and indeed were shown to be glycosylated when recombinantly expressed in mammalian cells (Lacy et al., 1999). The relatively short (approximately 100 residues) C-terminal intracellular tail is devoid of known structural motifs, but does contain possible sites for tyrosine phosphorylation (Blom et al., 1999), and FLRT1 has been shown to be phosphorylated by FGF receptor 1 (FGFR1) (Wheldon et al., 2010).

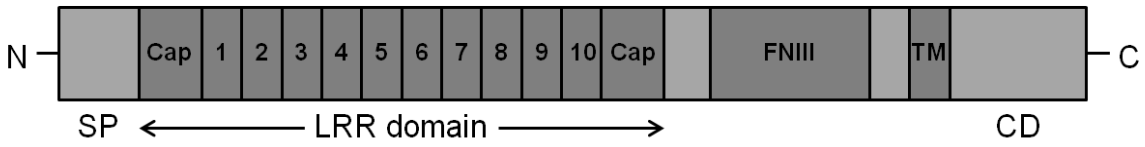


Figure 1.3 Generalised schematic representation of the domain organisation of members of the FLRT family. SP, signal peptide; Cap, N- and C-terminal cap regions; LRR, leucine rich repeat domain; FNIII, fibronectin domain type III; TM, transmembrane domain; CD, cytoplasmic domain. Figure adapted from (Lacy et al., 1999).

Although the three FLRT proteins "have no obvious homologues in *Drosophila melanogaster* or *Caenorhabditis elegans*" (Bottcher et al., 2004), they are generally highly conserved across higher vertebrates (though, interestingly, no avian FLRT1 homologue has yet been identified). The high degree of homology between the three FLRTs, as well as their overlapping expression pattern, initially led researchers to believe that they might be functionally redundant (Lacy et al., 1999), but a body of evidence now indicates that they have distinct and essential roles.

1.4 Roles of FLRTs

1.4.1 Neurite outgrowth and neural development

Unlike neurons in the central nervous system (CNS), which only exhibit minimal regeneration following injury due in part to the inhibitory environment (Fournier and Strittmatter, 2001; Filbin, 2003), neurons in the peripheral nervous system (PNS) are capable of axon regeneration, even after complete nerve transection (Fawcett and Keynes, 1990). In the PNS, neurotrophic factors and adhesion molecules are expressed by the distal segment of the injured nerve, and promote neurite outgrowth and regeneration (Bixby et al., 1988). These pro-regenerative molecules have hence been the source of considerable interest, as it is hoped that they might be exploited therapeutically to confer regenerative capacity to neurons within the CNS.

Microarray analyses by two different groups identified FLRT3 as being up-regulated following transection of the sciatic nerve in mice. (Tanabe et al., 2003) showed that FLRT3 expression is up-regulated almost ten-fold between dorsal root ganglion neurons before and after axotomy, and up-regulation was also observed in the distal stumps of axotomised sciatic neurons (Tsuji et al., 2004). Work by (Robinson et al., 2004) further confirmed up-regulation of FLRT3 (at the mRNA and protein levels) in regenerating rat dorsal root ganglions, from where it is transported along the injured nerve, and showed that it is specific to peripheral sensory nerve injury; up-regulation was not observed after induction of peripheral inflammation.

FLRT3 was also shown to promote outgrowth of neurites from cerebellar granule cells *in vitro* (Tsuji et al., 2004), and its role in affecting neurite outgrowth was confirmed by knockdown and overexpression (Robinson et al., 2004), which reduce and promote outgrowth, respectively. This led to speculation that it might act by modulating FGF signalling (see Section 1.4.2), or by generating adhesive cell-cell

contacts between the neurite growth cone and Schwann cells recruited to provide pathways for regrowth. The latter has been observed for other adhesion molecules including neural cell adhesion molecule (N-CAM) and N-cadherin (Fawcett and Keynes, 1990; Ide, 1996), both of which also interact physically with FGFRs and promote their signal transduction (Cavallaro et al., 2001; Williams et al., 2001).

In fact, the importance of the involvement of FGF signalling in neurite outgrowth was confirmed in studies showing that FLRT1, as well as FLRT3, can promote neurite outgrowth (Wheldon et al., 2010). Expression of FLRT1 alone in SH-SY5Y cells (a neuroblastoma cell line which responds to FGF1 overexpression by producing neuritic extensions (Raguenez et al., 1999)) results in changes in dendritic architecture which may be due to alterations in cell-cell adhesion, but co-expression of FLRT1 with FGFR1 also results in increases in the length and complexity of the dendritic extensions. The involvement of FGF-mitogen activated protein kinase (MAPK) signalling was further confirmed by complete abrogation of neurite outgrowth in the presence of inhibitors of FGFR1 and the MAPK pathway (Wheldon et al., 2010).

Early studies on the effect of FLRTs on neurite outgrowth often refer to the hypothesis that regeneration following injury requires activation, or recapitulation, of processes involved in development (Tanabe et al., 2003; Robinson et al., 2004; Tsuji et al., 2004). This led researchers to study whether FLRTs might also be involved in neuronal development, as well as regeneration in the adult organism following injury. All three FLRTs are widely expressed at early stages of brain development in rodents (Lacy et al., 1999; Robinson et al., 2004; Yamagishi et al., 2011); for example, FLRT3 is strongly expressed in developing mice dorsal root ganglions (Tanabe et al., 2003). However, expression is not universal (Lein et al., 2007) and can alter during

development, indicating the importance of its regulation. A heterogeneous expression pattern in adult rat brain suggested that FLRT3 expression is regulated during neuronal development (Tsuji et al., 2004), and indeed, FLRT3 mRNA expression was shown to decline rapidly postnatally (Robinson et al., 2004), while the location of FLRT2 expression in the cerebral cortex also changes prior to birth (Yamagishi et al., 2011).

More recent studies have also implicated FLRTs in neuronal development in new and exciting ways. The interaction of FLRT3 with the repulsive axon guidance receptor Unc5B has previously been shown to regulate cell adhesion (Karaulanov et al., 2009) (see also Section 1.4.3), and there is now evidence that FLRTs act as repulsive guidance cues which interact specifically with Unc5 receptors to affect neuron growth and migration *in vitro* and *in vivo*. Soluble FLRT2 and FLRT3 ectodomains (ECDs), which may be proteolytically cleaved and shed from cells to act as diffusible ligands *in vivo*, induce growth cone collapse of cortical neurons in an Unc5-dependent manner and repel hippocampal neurons; further, FLRT2 appears to interact with Unc5D to delay radial migration of cortical neurons (Yamagishi et al., 2011). In additional to a role in repulsive axon guidance, FLRT3 has also been recently implicated in excitatory synapse development as a post-synaptic ligand for latrophilin-3 (O'Sullivan et al., 2012); FLRT3 positively regulates synapse number in combination with latrophilin-3, and also helps to determine synapse function and connectivity. In this role, it joins an emerging group of LRR proteins implicated in synapse formation, differentiation, maintenance and plasticity (Ko and Kim, 2007).

FLRTs therefore appear to affect neuronal development and repair processes *via* a variety of mechanisms, which may rely on their interactions with FGF signalling and

their involvement in cell adhesion. It is these characteristics, and their involvement in many other biological processes, that will be discussed in the following sections.

1.4.2 Fibroblast Growth Factor mediated signalling and early development

FGFs and FGFRs constitute a complex signalling system which plays crucial roles in many vertebrate developmental and neural repair processes (Grothe et al., 2001; Romero et al., 2001; Thisse and Thisse, 2005). Briefly, FGFs bind the different transmembrane FGFRs (1-4), which have three extracellular Ig-like domains, at the cell surface, causing FGFR dimerisation and activation of the cytoplasmic tyrosine kinase domain, leading to transautophosphorylation (Lemmon and Schlessinger, 1994; Eswarakumar et al., 2005). Phosphorylated FGFR subsequently activates a variety of cellular signalling pathways, including the MAPK pathway, which signal the FGF-FGFR binding event at the cell surface into the cell nucleus to affect target gene expression (Kouhara et al., 1997).

Multiple partners, some of which have LRR domains (Söllner and Wright, 2009; Kim et al., 2012), have been shown to regulate the FGF-MAPK signalling pathway at different levels to affect the final output of the pathway, suggesting that precise adjustment of FGF signalling is critical to development (Tsang and Dawid, 2004). The first evidence for FLRTs as modulators of FGF-MAPK signalling implicated FLRT3 (Bottcher et al., 2004), which has an expression pattern closely resembling that of FGF8, and is itself regulated by FGF signalling (as are other modulators of FGF signalling (Niehrs and Meinhardt, 2002)) in early *Xenopus* development. The FLRT3 FNIII domain appears to mediate a physical interaction with FGFRs (as for the interaction of N-CAM with FGFR1 (Kiselyov et al., 2003)), though the ability to promote FGF signalling is dependent on the membrane-localised C-terminus; it may be that there is some phosphorylation state-dependent signalling activity in the short

FLRT intracellular domain (Wheldon et al., 2010), or it may act to recruit necessary cofactors to the membrane. It was noted in this initial investigation that significant sequence conservation between the three FLRTs could indicate that FLRTs 1 and 2 might also play a role in regulation of FGF signalling, and indeed, both FLRT3 and FLRT2 can phenocopy FGF signalling in gain- and loss-of-function experiments (Bottcher et al., 2004). Physical interactions between FGFR1 and all three FLRTs were confirmed (Haines et al., 2006; Karaulanov et al., 2006), and FLRT2 was also shown by yeast two-hybrid analysis to bind FGFR2 *in vivo* and *in vitro* (Wei et al., 2011). The interaction between FLRT1 and FGFR1 has also been reported to result in phosphorylation of the FLRT1 intracellular domain, which appears to negatively regulate FGFR1-MAPK signalling (Wheldon et al., 2010). Though the FLRT3 LRR domain was shown to be dispensable for FGFR interaction (Bottcher et al., 2004), the FLRT2 LRR domain appears to be critical for its interaction with FGFR2, and an interaction between the intracellular portions of FLRT2 and FGFR2 also contributes (Wei et al., 2011), raising the possibility that different FLRT/FGFR combinations interact in different ways, which may be important in the necessarily tight regulation of FGFR signalling in development.

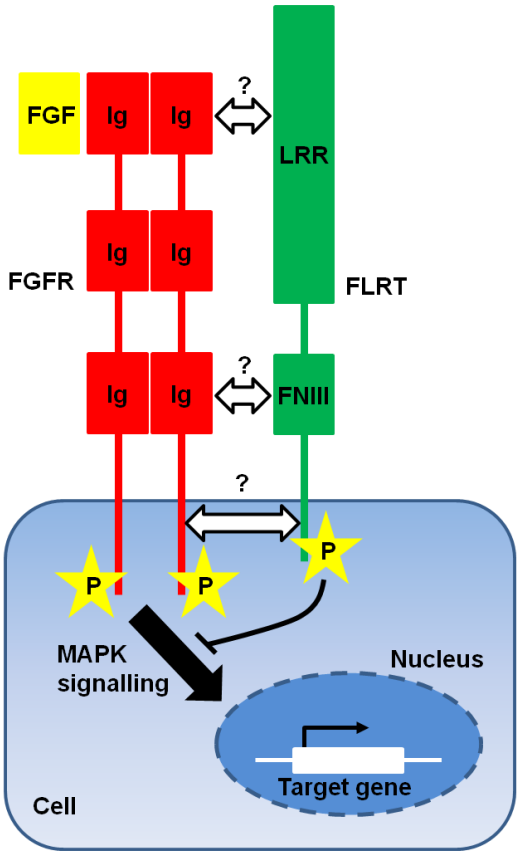


Figure 1.4 A summary model of FGFR-FLRT interactions at the cell surface, and how these may affect MAPK signalling. "?" indicates regions of FLRTs (green) that have been shown to interact with FGFRs (red); "P" in a yellow star indicates one or more phosphorylation event.

Interactions between FLRTs and FGFRs have been implicated in the early development of vertebrate embryos. FLRT expression patterns are tightly and differentially regulated during development (Mühlfriedel et al., 2007; Gong et al., 2009), and coincide closely with regions known to express FGF signalling components, and where FGF signalling is known to be important (Haines et al., 2006; Smith and Tickle, 2006; Tomás et al., 2011), as mentioned for FLRT3 above (Bottcher et al., 2004). One report notes additionally that FLRT3 expression during chick limb development is confined to regions where extracellular regulated kinase (ERK, a component of some MAPK signalling pathways) is phosphorylated, an

indication that the FGFR-MAPK signalling pathway may be activated (Tomás et al., 2011). Interdependence in the expression of the two protein families has been demonstrated in different organisms; injection of FGF8 mRNA increased FLRT3 expression in *Xenopus* animal caps (Bottcher et al., 2004), while mutant mice embryos carrying a tissue-specific inactivation of the FGFR1 gene, exhibited completely ablated FLRT3 expression in the midbrain-hindbrain boundary (Jukkola et al., 2006). Manipulation of FLRT expression also affects FGFRs, so that silencing of FLRT3 expression in the developing chick limb phenocopies pharmacological inhibition of FGF or MAPK signalling, while FLRT3 overexpression causes a broadening of the FGF8 expression domain.

Because of the well-demonstrated requirement for functional FGF signalling in vertebrate development, and the apparent interactions between the FLRT and FGFR families, disruption of FLRT expression might be expected to lead to many problems in the developing embryo. Indeed, knockdown of FLRT2 in certain progenitor cells slows cell proliferation (Xu et al., 2011), and knockout of FLRT2 and FLRT3 in mouse embryos has been shown to result in multiple developmental defects, including defects in ventral closure, headfold fusion, endoderm migration, disorganisation of the basement membrane and cardiac insufficiency (Egea et al., 2008; Maretto et al., 2008; Müller et al., 2011). It came as a surprise, therefore, that FLRT3 knockout did not affect FGF target gene expression in these embryos, and that inhibition of FGF signalling does not perturb FLRT2 or FLRT3 expression in normal embryos (Maretto et al., 2008; Müller et al., 2011). Further, *FLRT2*$^{-/-}$ and *FLRT3*$^{-/-}$ cells do not display enhanced levels of MAPK signalling pathway components upon FGF stimulation (Egea et al., 2008; Müller et al., 2011). These studies suggest that FLRT functions during early development may be independent of FGF signalling, though (due to the

early lethality of the null mutants) do not exclude a function for FLRTs in modulating FGF signalling in later stages of development or in adulthood. Rather, they suggest that the defects described are due to the involvement of FLRTs in cell-cell adhesion, which will be explored in the next section.

1.4.3 Cell sorting and cell-cell adhesion

Recognition between cells and regulation of their adhesive properties is vital, both to neurite outgrowth and developmental processes more generally (Hammerschmidt and Wedlich, 2008). There are precedents for proteins similar to FLRTs displaying dual function by signalling *via* receptor tyrosine kinase signalling pathways (including the FGF signalling pathway (Hansen et al., 2008)) (Cavallaro and Christofori, 2004), and having autonomous cell adhesion function, and it appears that FLRTs should join this group. As mentioned above, studies in FLRT knockout mice indicate that the multiple developmental defects observed are not caused by defects in FGF signalling (Maretto et al., 2008; Müller et al., 2011), and cell deadhesion caused by FLRT3 overexpression was also shown to be FGF signalling-independent (Ogata et al., 2007). Additionally, FLRT-mediated homotypic cell sorting was shown to be LRR domain-dependent (the LRR domain is dispensable for FGF signalling), but independent of the cytoplasmic domain (which is required for signalling), excluding the possibility of an indirect adhesion effect *via* FGFR signalling (Karaulanov et al., 2006).

These observations could indicate that a direct interaction between FLRT molecules on opposing cells underlies any homotypic cell-cell adhesion or cell sorting behaviour. Such behaviour was initially described in HEK 293T cells (Karaulanov et al., 2006; Tossell et al., 2011), and shown to be LRR domain-dependent, along with evidence that FLRT molecules can interact with each other. FLRTs were also shown

to localise to cell-cell contacts between Cos cells, along with the cytoskeletal attachment protein vinculin, identifying the contacts as focal adhesions (Haines et al., 2006). Other reports did not find evidence for homophilic interactions between FLRTs (Robinson et al., 2004; Tsuji et al., 2004; Yamagishi et al., 2011), but these observations could be explained if other partner cell surface molecules which are necessary for an interaction were not present, e.g. for secreted FLRT constructs in solution.

The small GTPase Rnd1, and the netrin receptor Unc5B have been identified as such FLRT binding partners in *Xenopus* (Ogata et al., 2007; Karaulanov et al., 2009). FLRT3 binds Unc5B *via* its LRR domains, so could theoretically bind in *trans* between two neighbouring cells, e.g. when acting as a repulsive guidance molecule (Yamagishi et al., 2011)), while Rnd1 interacts with FLRT3 *via* the FLRT3 cytoplasmic tail. Rnd1 is coexpressed with FLRT3 in gastrula-stage embryos, and embryos lacking Rnd1 display significant gastrulation defects. It displays strong anti-adhesive activity (Nobes et al., 1998; Wunnenberg-Stapleton et al., 1999), which it appears to mediate by controlling cell surface availability of cadherin through a dynamin- and Rab5-dependent endocytosis pathway (see Figure 1.5). Loss of expression of either Rnd1 or FLRT3 increases cadherin-mediated cell adhesion, while overexpression of Rnd1, Unc5B or FLRT3 has the opposite effect (Ogata et al., 2007; Karaulanov et al., 2009); the same has also been shown for FLRT2 (Xu et al., 2011). Paraxial protocadherin (PAPC) has been reported to restrain the deadhesion activity of FLRT3 by inhibiting the recruitment of Rnd1 to the intracellular tail (Chen et al., 2009). Mouse FLRT2 also appears to bind Unc5D (as well as Unc5B to a lesser extent) (Yamagishi et al., 2011), while zebrafish FLRT1b and FLRT3 also bind FLRT3 (Söllner and Wright, 2009).

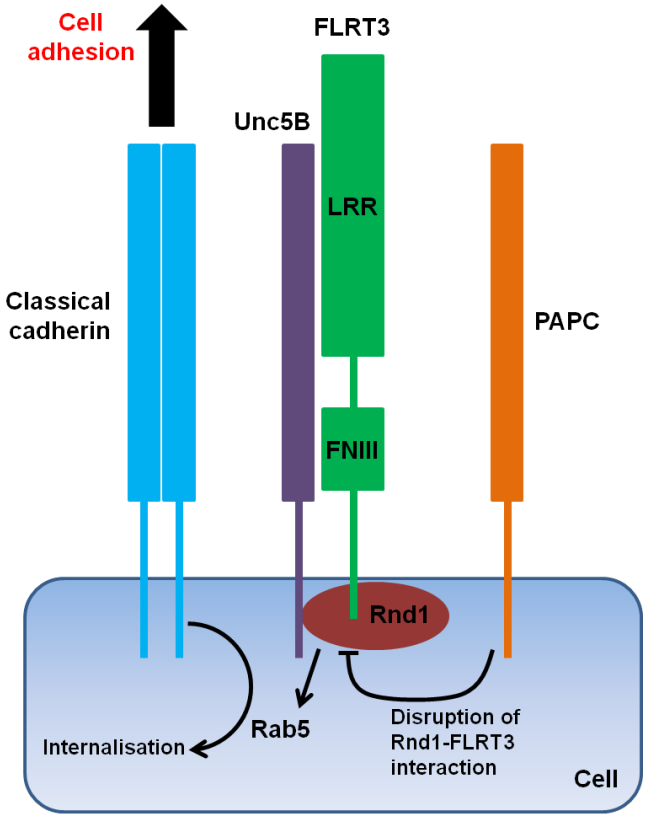


Figure 1.5 A summary of interactions of FLRT3 which regulate cadherin-dependent cell-cell adhesion. Figure adapted from (Hammerschmidt and Wedlich, 2008).

Evidence of FLRT-dependent deadhesion behaviour in *Xenopus* (Ogata et al., 2007) may appear to contradict data describing the role of FLRTs in promoting homotypic cell adhesion in mice (Egea et al., 2008; Maretto et al., 2008). The forming and releasing of adhesive interactions are both important processes in development; the correct regulation of the latter (e.g. by PAPC (Chen et al., 2009) is necessary for cells to move through a tissue and slide past each other without sacrificing tissue integrity. These two actions can be reconciled within FLRT3 by functional decoupling between the extracellular and intracellular domains. Homotypic cell sorting and adhesion is promoted *via* the extracellular LRR domain (Karaulanov et al., 2006), while cell adhesion is regulated *via* the intracellular tail, and its interaction with other protein partners (Ogata et al., 2007; Chen et al., 2009; Karaulanov et al., 2009). It is

possible that similar mechanisms may exist for the FLRT1 and FLRT2. Tight spatial and temporal regulation of such bifunctionality during development (e.g. by interaction partners (Chen et al., 2009)) is likely to be crucial, for example, to maintain homotypic cell groups while loosening adhesion in gastrulating cells, allowing them to move while preventing tissue damage.

1.5 The Leucine Rich Repeat

1.5.1 Introduction to LRR domains and their biological versatility

Proteins containing segments with repeating amino acid motifs are a common way for evolution to produce a class of stable structural units that can be modified for involvement in diverse processes (Kobe and Kajava, 2001; Mosavi et al., 2004; Schmitz-Linneweber and Small, 2008). Consensus motifs provide a regular, stable scaffold, while non-consensus residues can be changed to alter ligand specificity and thereby biological function (Helft et al., 2011).

The LRR is a widespread repeating protein motif, found in many intracellular, membrane-associated and extracellular proteins of viral, prokaryotic and eukaryotic origin (Kobe and Kajava, 2001). Figure 1.6 shows schematic representations of a selection of LRR-containing proteins mentioned in this Chapter, indicating the diversity in number of repeat motifs and other non-LRR domains which permit the involvement of this family of proteins in diverse biological roles. LRRs are able to interact with a range of substrates including small molecule hormones, sugars and nucleic acids (Price et al., 1998; Napier, 2004; Han et al., 2008), but in many cases, the LRR appears to function as a versatile structural framework for protein-protein recognition processes. In this way, LRR proteins participate in diverse processes including enzyme inhibition (Maulik et al., 2012), regulation of gene expression (Linhoff et al., 2001) and platelet aggregation (Andrews et al., 2004). The benefit of

the repetitive structure in rapid generation of new variants has been well studied in plant and animal immunity (Nürnberger et al., 2004; Herrin and Cooper, 2010; Ng et al., 2011), and in bacterial virulence (Evdokimov et al., 2001).

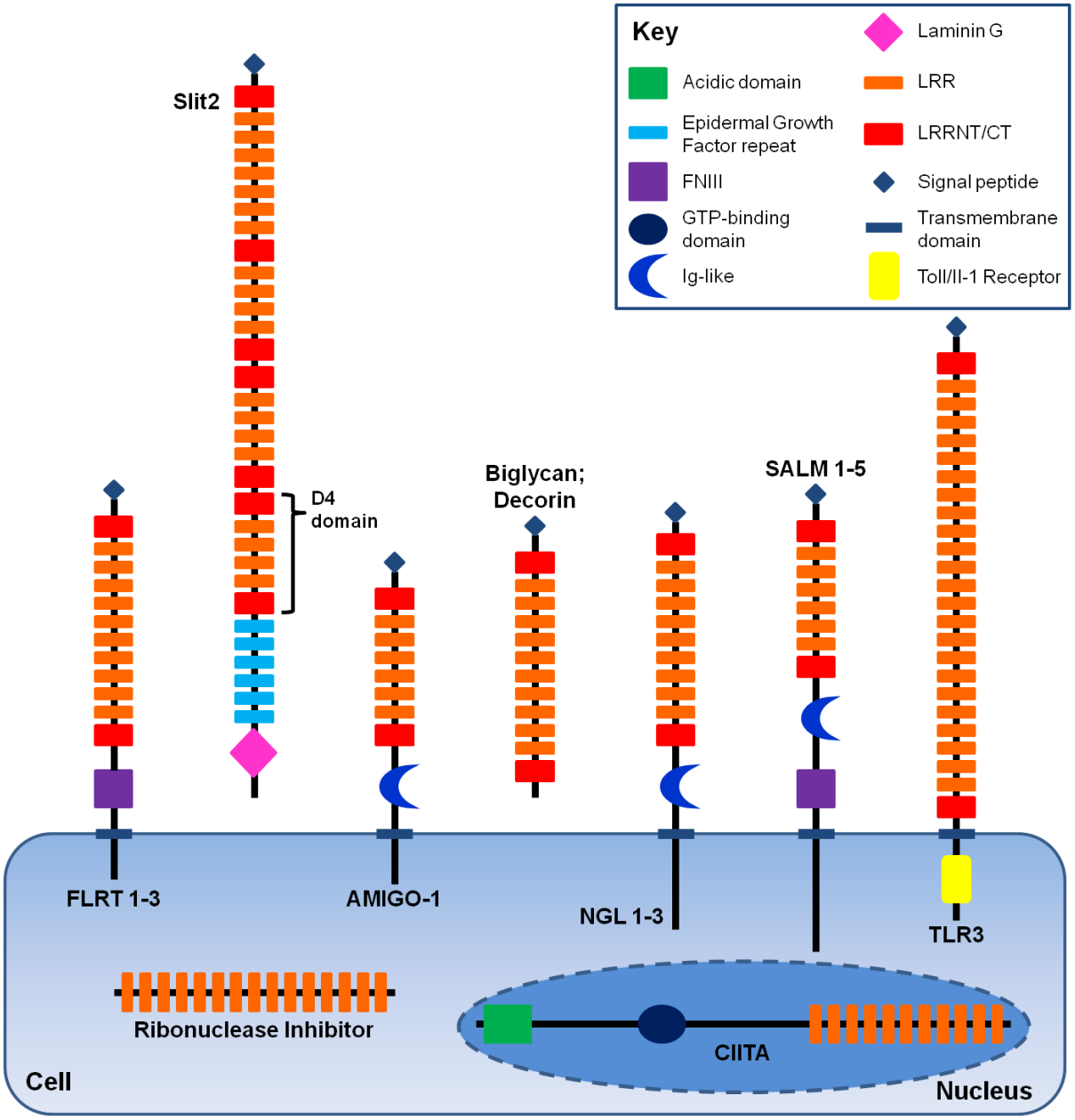


Figure 1.6 Schematic representation of several LRR-containing proteins which are mentioned in this thesis, illustrating diversity in numbers of LRR repeat units and other domains, allowing functional versatility across the family; see Key for further information. Figure adapted from (Dolan et al., 2007) with further information from (Kobe and Deisenhofer, 1993) and (Linhoff et al., 2001).

As well as interacting with other proteins, the homodimerisation of proteins *via* an LRR domain has also been shown to be important in biological function. In CIITA, which regulates MHC class II gene expression (see Section 1.2.1 for a summary of the function of MHC class II molecules in the adaptive immune system), an LRR domain allows self-association, interacting both with itself and with the GTP-binding domain elsewhere in the protein (Linhoff et al., 2001), which may be key to correct recruitment of transcription factors. Dimerisation of Slit2 *via* its D4 LRR domain has also been shown to be important for correct function of Slit2 in neuronal development (Seiradake et al., 2009). Amphoterin-induced gene and ORF (AMIGO-)-1, which has, like FLRTs, been implicated in cell adhesion and in neurite outgrowth (Kuja-Panula et al., 2003), appears to be an obligate homodimer; cell surface expression does not occur properly without dimerisation, which may mean that protein folding in the endoplasmic reticulum (ER) is dimerisation-dependent (Kajander et al., 2011). The same phenomenon has been observed for the small leucine-rich repeat proteoglycans biglycan and decorin (Scott et al., 2006), where the dimerisation interface may also be involved in ligand binding to collagen (Orgel et al., 2009). Although high-resolution structures have shown that there is some variability in the way these LRR proteins dimerise, in terms of angles between the two molecules and parts of the domain involved, the dimerisation interface involves the inside concave face of the LRR domain (see Section 1.5.3) in all cases.

1.5.2 LRR proteins in cell surface recognition events in development

There are many examples of LRR domain-containing proteins which, like FLRTs, have been shown to have a role in development. As mentioned above, there is recent evidence (O'Sullivan et al., 2012) linking FLRTs to other LRR proteins, including the netrin-G ligand (NGL) and synaptic adhesion-like molecule (SALM)

families, which have been implicated in regulation of synapse structure and function (Ko and Kim, 2007). More classically, however, LRR proteins affect neuronal development or repair *via* an effect on neurite outgrowth (see Section 1.4.1 for a summary of the evidence for FLRT involvement); they may promote neurite extension, or inhibit it (Chen et al., 2006).

Critical steps in the development of an organism are dependent on cellular interactions mediated by adhesion molecules, including dynamic processes such as cell migration and axonal guidance, as well as architectural purposes such as cell sorting and boundary formation. In addition to facilitating neurite outgrowth, the involvement of LRR domains in cell adhesion and cell sorting phenomena (Krantz and Zipursky, 1990) therefore also implicates them in development more generally, and there are many examples of cell-surface LRR-containing proteins which have been shown to be involved in development in several capacities (e.g. boundary formation (Milán et al., 2001; Andreae et al., 2007) and neuronal growth cone guidance (Nose et al., 1997; Kuja-Panula et al., 2003)). Several of these examples promote homotypic adhesion (Krantz and Zipursky, 1990; Nose et al., 1992), while others are capable of homophilic and heterophilic binding within their protein family (Kuja-Panula et al., 2003; Seabold et al., 2008), in *cis* on the same cell surface and in *trans* between cells (Kajander et al., 2011); see Figure 1.7. This suggests that the LRR domain may be involved in regulating the affinity of like cells for each other, and in fact, the LRR domain was shown to be essential for the homotypic cell sorting behaviour of cells overexpressing FLRT3 (Karaulanov et al., 2006). In this thesis I propose that, as has also been suggested for AMIGO-1 (Kajander et al., 2011), a direct interaction between the LRR domains of FLRTs at the cell surface might underlie the physical interactions (homotypic and heterotypic) previously observed

between FLRTs (Bottcher et al., 2004; Karaulanov et al., 2006), and thereby their biological function.

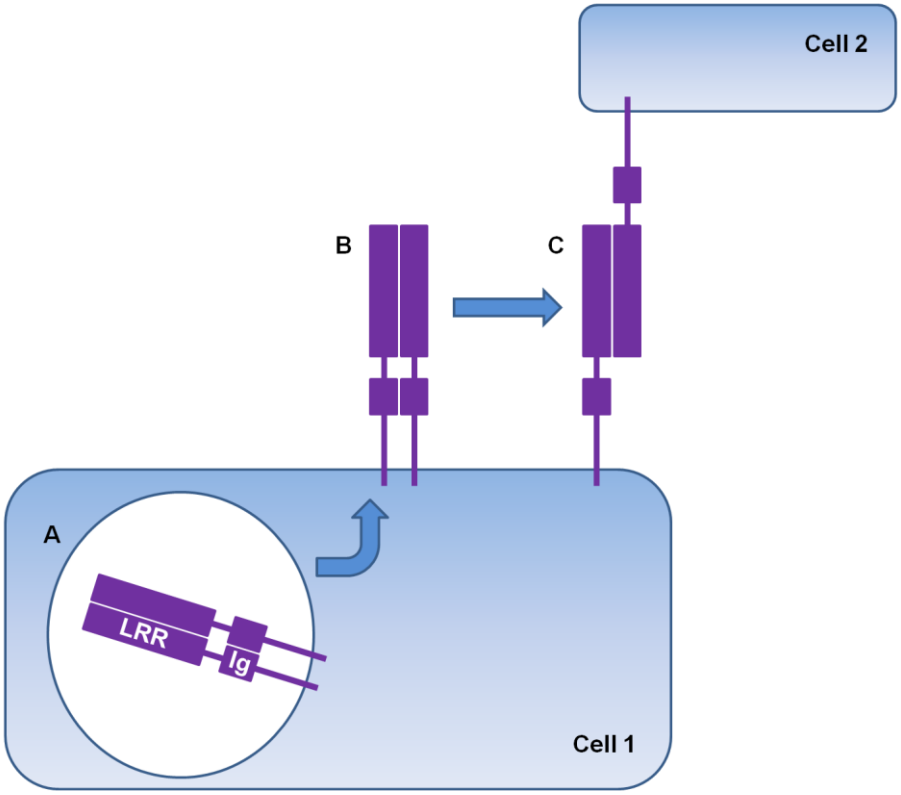


Figure 1.7 AMIGO-1 as a model for *cis* and *trans* homodimerisation mediated by LRR domains. (A) Homodimer formation is essential for correct folding and transport to the cell surface, suggesting dimerisation occurs in the ER (Kajander et al., 2011); (B) because of this, AMIGO-1 is expressed at the cell surface as an obligate dimer; (C) AMIGO-1 can make homomeric *trans* interactions between surfaces (Kuja-Panula et al., 2003), and it is suggested that the same interface that is used for *cis* dimerisation is also employed for *trans* dimerisation (Kajander et al., 2011). LRR, LRR domain; Ig, immunoglobulin domain.

1.5.3 LRR domain structure

LRR domains vary in length by the number of repeat units they contain, from two repeats up to fifty two; these repeats pack against each other to form the body of the LRR domain. Each repeat unit is typically 19-29 amino acid residues long, and includes a conserved segment with the consensus sequence LxxLxLxx(N/C)xL (where x can be any amino acid; L positions can also be occupied by valine,

isoleucine or phenylalanine) (Kobe and Kajava, 2001). The first structural data obtained for an LRR-containing protein was from the crystal structures of porcine ribonuclease inhibitor, alone (Kobe and Deisenhofer, 1993) and in complex with protein ligands (Kobe and Deisenhofer, 1995; Papageorgiou et al., 1997). These show how each repeat unit forms a helix (α-helix, 3_{10} helix, or polyproline II helix) connected by loops to a β-strand, which in the domain context align to form a continuous hydrogen-bonded parallel β-sheet composed of one strand from each repeat (Bella et al., 2008). In this way, each repeat forms a turn of the canonical curved solenoid structure, with the β-sheet on the concave face, and other secondary structural elements on the convex face (Figure 1.8); the differences in cross-section between these two sides result in the curvature of the structure (Enkhbayar et al., 2004). While the solvent-exposed portions of the structure are dominated by hydrophilic residues which determine ligand specificity, the small, hydrophobic leucine residues which make up the consensus leucine-rich motif are ideally suited to pack together in the sterically-restricted space within the solenoid, forming the hydrophobic core of the domain.

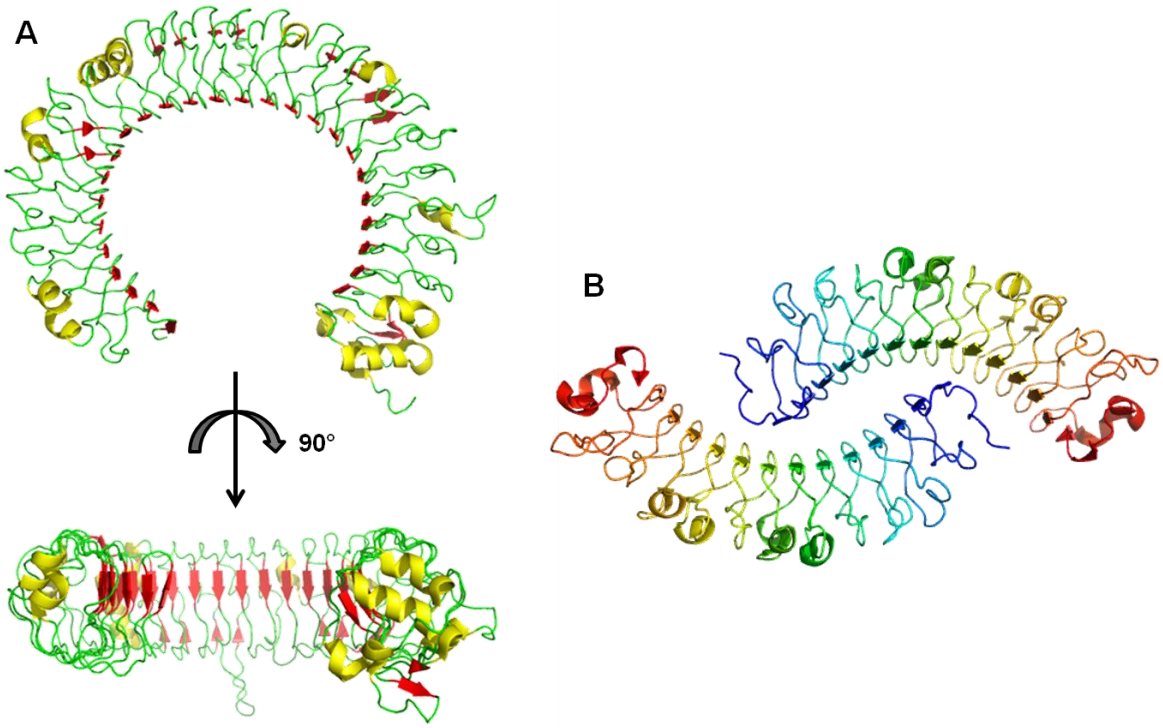


Figure 1.8 Structures of LRR domains, and the role of the concave face in protein-protein interactions including homodimerisation. (A) The structure of Toll-like receptor 3 (TLR3; PDB code 2A0Z), showing the characteristic solenoid fold. The concave face is made up of a parallel β-sheet (coloured red), while the convex side shows a variety of secondary structures; each LRR repeat unit contributes one turn of the solenoid. The structure is viewed from the side of the solenoid (top), and rotated by 90° to view the parallel β-sheet in the concave face (bottom). (B) The structure of the decorin homodimer (PDB code 1XKU), as an example of how the concave face is used in antiparallel LRR dimerisation interfaces. Decorin has been shown to be dimeric in solution (Scott et al., 2003), with dimerisation essential to correct folding and secretion (Scott et al., 2006). Figure adapted from (Bella et al., 2008).

The concave face is generally thought to be the site for protein-protein interactions in ligand binding and dimerisation (Kobe and Deisenhofer, 1995), and this has been shown in several LRR crystal structures (e.g. decorin in Figure 1.8; (Scott et al., 2004; Morlot et al., 2007)), though there are other examples which employ alternative surfaces (Jin et al., 2007). Of the structures of membrane-associated LRR proteins known to function as homodimers (decorin (Scott et al., 2004), biglycan (Scott et al.,

2006), the Slit2 D4 domain (Seiradake et al., 2009) and AMIGO-1 (Kajander et al., 2011)), there is some variation in the way the monomers interact, but the interface is on the concave face in all cases and all the dimers are antiparallel, i.e. the chains run in opposite directions. The conserved hydrophobic residues are usually buried in the core of the solenoid, stabilising the domain by optimising van der Waals interactions between the closely-packed side chains. Residues at the variable positions, meanwhile, are predominantly solvent-exposed; this helps to explain that whilst there is a high degree of structural conservation between LRR domains of different proteins, they bind their protein partners in a highly specific fashion (Lin et al., 2003). Extracellular and membrane-associated LRR domains are typically flanked at their N- and C-termini by cysteine-rich, disulphide-bonded cap motifs (LRRNT and LRRCT) that protect the hydrophobic core of the first and last repeat units.

1.6 Objectives of this thesis

As discussed above, cell surface receptors are vital for organisms to sense and interact with their environment, for example to recognise and mount an immune response against pathogens, or to organise cell populations during development. The twin aims of this thesis are the characterisation of two such receptor families: cross-reactive TCRs and the FLRTs.

One strand of this thesis concerns the structural basis of TCR cross-reactivity in the autoimmune disease MS, which I aim to investigate using macromolecular crystallography. I aim firstly to clone, express and purify a highly cross-reactive TCR derived from the same patient as Hy.1B11 (Hy.2E11 (Hy)), and then characterise its interactions with multiple MHC class II-presented peptides. Hy recognises the self peptide MBP_{85-99} in the context of DR2b, as well as virally derived peptides in the contexts of DR2a and HLA-DR4 ((Lang et al., 2002) and personal communication,

Lars Fugger (Human Immunology Unit, Weatherall Institute of Molecular Medicine (HIU, WIMM), University of Oxford)). This TCR therefore provides an opportunity to study the structural basis of molecular mimicry in TCR cross-reactivity, should it prove possible to solve the structures of some of these complexes. Also as part of this strand of the thesis, I aim to add to the database of autoreactive TCRs in complex with self pMHC ligands by extending the work described in (McMahon, 2009) on solving the structure of the MS patient-derived TCR MS49 in complex with HLA-DQ6 and an MBP_{30-44} peptide. As the data on the contribution of the third allele of the DR2 haplotype (DQB1*0602) to MS susceptibility remains inconclusive, this first structure of a DQ6-restricted TCR complex, if solved, would hopefully provide insight into the role of DQ6 in autoreactive TCR recognition in MS.

While interesting data continue to accumulate on the functions and interactions of FLRTs in nerve regeneration and early development, there is very little structural or biophysical information available for these interactions to date. In the other strand of this thesis, I aim to provide further descriptions of the effects of FLRT molecules on cellular morphology and homotypic cell-cell interactions. The extracellular LRR domain has been shown to be essential for this homotypic cell adhesion behaviour (Karaulanov et al., 2006), and in this thesis, I will provide evidence for a physical interaction between FLRT LRR domains which underlies FLRT homodimerisation. Using high-resolution models obtained using x-ray crystallography, combined with biophysical and cellular studies, I will attempt to present a model for the functional relevance of such dimerisation at the cell surface, and discuss how this might underlie homotypic cell adhesion events during early development and nerve regeneration.

Chapter 2 Methods and Materials

Chapter 2 Methods and Materials

This chapter outlines the procedures used for the design, production and characterisation of various constructs used in the course of this thesis. These constructs, including those based on FLRTs and others based on the Hy and MS49 TCRs, can be broadly divided into soluble constructs used for biophysical characterisation and crystallisation, and transmembrane constructs used to investigate the behaviour of FLRTs at the cell surface.

2.1 Molecular biology

2.1.1 cDNA and expression vectors

cDNA clones encoding mouse FLRT1, FLRT2 and FLRT3 were a kind gift from Liz Bikoff and Liz Robertson (Sir William Dunn School of Pathology, University of Oxford). Residue numbering refers to: mouse FLRT1 (GenBank NP_958813; cDNA NM_201411), mouse FLRT2 (GenBank NP_958926; cDNA NM_201518) and mouse FLRT3 (GenBank NP_001165631; cDNA NM_178382). DNA clones encoding the human Hy.2E11 TCR (hereafter Hy) α and β chains were kindly provided by Pia Svendsen (Clinical Institute, Aarhus University Hospital, Denmark).

Constructs for expression in mammalian cell lines were cloned into the pHLsec vector (see Appendix A), which is based upon the pLEXm backbone (Aricescu et al., 2006). For expression using the native N-terminal secretion signal sequence of the target protein, constructs were inserted using the *EcoRI* and *KpnI* restriction sites, which introduces a C-terminal hexahistidine tag; where constructs contained internal *EcoRI* and/or *KpnI* sites, they were digested using other restriction enzymes to generate compatible ends (see Appendices A and B). When the native N-terminus of the construct was not present, either for insertion of an N-terminal sequence (e.g. a

FLAG or HA tag, or a linked antigenic peptide), or in the case of single domain ($FLRT_{LRR}$) constructs, the *AgeI* site was used in place of the *EcoRI* site, which introduces a 5' end Kozak sequence and N-terminal secretion signal from the vector, in addition to the C-terminal hexahistidine tag. Where constructs contained internal *AgeI* sites, they were digested with *NgoMIV* to produce compatible ends (see Appendices A and B). For microscopy and FRET assays, constructs were cloned into additional pHLsec vectors (pHL-mVenus and pHL-mCerulean) constructed by Amber Clayton and Charlotte Coles respectively (both Division of Structural Biology, University of Oxford (Strubi)); these contain insertions between the *KpnI* and *XhoI* restriction sites of a 3C protease cleavage site followed by residues 1-239 monoVenus (GenBank DQ092360) or residues 1-239 monoCerulean (GenBank ACO48272), prior to a C-terminal octahistidine tag. These fluorescent proteins have been selectively mutated to be monomeric and to have increased stability, and therefore offer enhanced efficiency of maturation compared to the more commonly used green fluorescent protein (Nagai et al., 2002). For expression in *E. coli* cells, constructs were cloned into the pET22b(+) vector (Invitrogen; Appendix A) using the *NdeI* and *XhoI* restriction sites, which introduced a C-terminal hexahistidine tag (Appendix A). For expression in Sf9 insect cells, inserts including the N-terminal signal sequence and C-terminal fluorescent and hexahistidine tags were cut from sequence verified pHL-mVenus constructs with *EcoRI* and *XhoI* and cloned into pBacPAK9 (Clontech). The single exception to this approach was $FLRT3_{LONG}$ which includes an internal restriction *EcoRI* site, so was cloned into pBacPAK9 using *BamHI* and *SacI* restriction sites, with the mVenus tag cloned into the same vector using the *KpnI* and *XhoI* restriction sites downstream of the $FLRT3_{LONG}$ insertion site (Appendix A).

2.1.2 Construct design and cloning

Details of the final constructs used in this thesis are listed in Appendix B; primers used in PCR reactions (MWG Operon) are listed with restriction sites shown in red. Deletion constructs encoding soluble FLRT ectodomains ($FLRT_{ECD}$) and LRR domains ($FLRT_{LRR}$) were designed and produced by Radu Aricescu and Maria Harkiolaki (both Strubi); transmembrane FLRT molecules with truncated cytoplasmic domains ($FLRT_{LONG}$) were designed by Radu Aricescu. Two constructs encoding Hy TCR α and β chains for expression in mammalian cells were designed and produced by Maria Harkiolaki; other Hy constructs designed by Maria Harkiolaki and Pia Svendsen (Clinical Institute, Aarhus University Hospital, Denmark) are listed in Table 3.1, though these (and all other Hy constructs) were produced by me. Constructs encoding transmembrane FLRT proteins for cellular studies, and further Hy α and β chain constructs for expression in mammalian and bacterial systems, were designed using domain boundaries listed on the ImMunoGeneTics (IMGT) online database, the N-terminal signal sequence prediction server SignalP (Petersen et al., 2011) and the transmembrane helix/disorder prediction server RONN (Yang et al., 2005). In some instances, inserts contain internal *EcoRI*, *KpnI* or *AgeI* restriction sites; in these cases, primers were designed with restriction sites that produce compatible ends to sites in the vector multiple cloning site e.g. *MfeI* replacing *EcoRI* (see Appendix B). For co-expression of Hy α and β chains in mammalian cells, the vector-derived C-terminal hexahistidine tag was removed from one of the chains by insertion of a stop codon prior to the *KpnI* restriction site.

PCR reactions were performed in a volume of 50 µl containing: 1x *Pyrobest*™ buffer II (containing 1 mM Mg^{2+}), 0.4 mM dNTPs, template DNA (approximately 2 ng/µl), 0.2 pmol/µl each of the forward and reverse primers, and 0.1 units

Pyrobest™ DNA polymerase (Takara Co.); this thermostable polymerase can achieve high-fidelity amplification due to 3'-5' exonuclease activity. PCR reactions were performed using a Peltier Thermal Cycler-225 (MJ Research), as follows:

1. 96 °C for 2 min for initial melting of the template DNA
2. 96 °C for 45 s to separate the strands of the template DNA double helix
3. 55 °C for 45 s for primers to anneal to the single-stranded template
4. 72 °C for 3 min for the polymerase to synthesise (elongate) new DNA
5. Repeat steps 2-4 29 times for sufficient amplification
6. 72 °C for 10 min to ensure the final elongation step is complete.

Site-directed mutagenesis was performed by first producing N- and C-terminal fragments in separate initial PCR reactions, then joining the fragments in a second reaction. In the first reaction, the terminal forward primer for the wild-type construct was used with an internal reverse primer, and the reverse primer for the wild-type construct combined with an internal forward primer; following the reaction, the fragments were purified using the QIAquick PCR purification kit (Qiagen). The internal primers contain altered codons encoding the desired mutations (shown in blue in Appendix B), and are complementary to each other, allowing fragment pairs to be joined at the point of the mutation, and amplified by a second PCR reaction. This reaction proceeded for three cycles before terminal forward and reverse primers were added for a further 27 reaction cycles to define the full-length mutated construct. To create constructs with several mutations which are too far separated in the sequence to be introduced by a single set of internal primers, as in the case of the Arg225Glu, Arg226Glu, Ala321S FLRT1 mutant, the mutations were introduced sequentially using the sequence verified construct containing the first mutation as the template for the second mutagenesis step.

2.1.3 Restriction digest, ligation, transformation and amplification

To remove impurities prior to the restriction digest of the cloned insert, all PCR mixtures were purified using the QIAquick PCR purification kit (Qiagen). The purified PCR products and 5 µg quantities of the appropriate vector were digested for two hours with the restriction enzymes (New England BioLabs) indicated in Appendix B, using the recommended buffer conditions and temperature for each enzyme. Where the recommended conditions for the two enzymes digesting each end of an insert or vector were incompatible, the digests were performed sequentially with the second digest following purification of the products of the first digest using the QIAquick PCR purification kit (Qiagen). The doubly-digested inserts and vectors were purified by agarose gel electrophoresis, and bands of appropriate size were visualised using SYBRSafe stain (Invitrogen), cut and purified using the QIAquick gel extraction kit (Qiagen).

Ligation reactions were set up for 10 min and 25 °C in a total volume of 20 µl using the Quick Ligation Kit (New England BioLabs), with purified insert and vector in approximately a 3:1 ratio. 2 µl of this ligation mix was used to transform 40 µl DH5α library efficiency competent cells (Invitrogen); the cells were incubated with the ligation mix on ice for 30 min prior to a heat shock at 42 °C for 45 s and a further 2 min on ice. 100 µl SOC medium was added before incubation at 37 °C with 200 rpm shaking for 1 hour. Following this recovery stage, cells were spread on LB-agar plates containing 100 µg/ml ampicillin and incubated overnight at 37 °C. Three colonies for each construct were picked and used to inoculate 3 ml overnight cultures of LB containing 100 µg/ml ampicillin for preparation of plasmid DNA using the QIAprep spin Miniprep kit (Qiagen). 1 µg purified DNA was digested with the appropriate restriction enzymes in the recommended buffer and temperature

conditions for 30 min and analysed by agarose gel electrophoresis. Where an insert of the correct size was observed, clones were sequenced for final verification (Source Bioscience). External sequencing primers for the vectors used in this thesis are:

pHLsec:	pHLsec_fwd	5' - gctggttgttgtgctgtctcatc - 3'
	pHLsec_rev	5' - caccagccaccaccttctgatag - 3'
pHL-mVenus/mCerulean:	pHLsec_fwd	5' - gctggttgttgtgctgtctcatc - 3'
	Venus_rev	5' - gccctggaacagcacctccag - 3'
pBacPAK9:	Bac1_fwd	5' - aaccatctcgcaaataaata - 3'
	Bac2_rev	5' - acgcacagaatctagcgctt - 3'
pET22b(+):	T7_fwd	5' - taatacgactcactataggg - 3'
	T7_rev	5' - gctagttattgctcagcgg - 3'

As expression in mammalian cells requires a significant amount of plasmid DNA (in the range of 5 µg/transfection for small scale expression and microscopy applications, and 2 mg/litre for large scale soluble expression), amplification was frequently required to produce sufficient quantities. Approximately 300 ng sequence verified plasmid DNA was used to transform 35 µl DH5α cloning efficiency competent cells (Invitrogen) as described above for library efficiency cells, but omitting the recovery step so that colonies on the plate would be sufficiently separated. A single colony was picked and used to inoculate 500 ml or 2.5 litre overnight cultures of LB containing 100 µg/ml ampicillin for preparation of isopropanol-extracted plasmid DNA using Endotoxin-free Plasmid Mega or Giga kits (Qiagen).

2.2 Expression and purification of soluble constructs

2.2.1 Small-scale protein expression in mammalian cells

Soluble FLRT constructs and certain Hy TCR α and β chain constructs were expressed in human embryonic kidney (HEK) 293T and HEK 293S GnTI$^-$ (HEK 293S) mammalian cell lines. These were maintained by Weixian Lu (Strubi) at 37 °C, 5 % CO_2, in Dulbecco's Modified Eagle Medium (DMEM, Sigma) supplemented with 10 % fetal calf serum (FCS, Sigma), L-glutamine and non-essential amino acids (Gibco) (Aricescu et al., 2006). Prior to larger scale expression for purification from the cell media, the soluble expression of certain constructs was tested in small-scale trials in HEK 293T cells. Approximately $3x10^5$ cells were seeded into 6-well plates in DMEM containing 10 % FBS and grown to approximately 95 % confluence before media exchange into DMEM containing 2 % FBS (1.5 ml per well). For transfection, 250 µl serum-free DMEM containing 4 µg sequence verified plasmid DNA was gently mixed with 250 µl serum-free DMEM containing 5 µg Lipofectamine^{TM} 2000 transfection reagent (Invitrogen) and incubated at room temperature for 20 min, then added to each well.

Three days post-transfection, the cells and cell media were collected and prepared for Western blot analysis (Aricescu et al., 2006). The media was carefully removed from each well and centrifuged to remove cell debris, and the cells were resuspended in 1 ml Phosphate Buffered Saline (PBS, Sigma). 20 µl samples of each were boiled in 10 µl reducing loading buffer for 10 min, and loaded onto a sodium dodecyl sulphate-polyacrylamide gel which separates proteins in the samples according to size by electrophoresis (SDS-PAGE). Benchmark hexahistidine-tagged and pre-stained protein ladders (Invitrogen) were also loaded to act as a molecular weight marker and to check that transfer from the SDS-PAGE gel to the membrane

had been successful. Following separation, proteins were transferred from the gel to a nitrocellulose membrane using either an XCell SureLock Mini-Cell (Invitrogen) with 10 % methanol in NuPAGE® transfer buffer (Invitrogen), or a Trans-Blot SD semi-dry transfer cell (Bio-Rad) with 20 % ethanol, 25 mM Tris, 200 mM glycine, for 1 hour at 30 V. Unbound sites on the membrane were blocked with 5 % (w/v) dried milk in PBS for 1 hour at room temperature, before the membrane was probed with a monoclonal mouse PentaHis primary antibody (1:1,000 dilution; Qiagen), washed three times with 0.1 % Tween-20 in PBS for 5 min each to reduce non-specific binding), and probed again with a goat anti-mouse secondary antibody conjugated to horseradish peroxidase (1:2,000; Sigma). The membrane was washed three times as before, and covered with an enhanced chemiluminescence reagent (ECL kit, GE Healthcare) for 1 min and exposed to Hyperfilm ECL (GE Healthcare). Finally, to visualise the signal, the film was developed using a Compact X4 machine (Xograph Healthcare).

2.2.2 Protein expression in mammalian cells for large-scale purification

Constructs that proved to be soluble were carried forward for large-scale transient expression, which was typically carried out by Weixian Lu (Strubi). Adherent HEK 293T or HEK 293S cells were grown to approximately 90 % confluence in expanded surface polystyrene roller bottles (2125 cm^2, Greiner Bio-One) and transfected with 2 mg isopropanol-extracted DNA and 4 mg polyethylenimine (PEI) per litre of cell media (Aricescu et al., 2006). After transfection, the roller bottles were incubated at 37 °C with 5 % CO_2 with gentle rolling to circulate the media over the secreting cells; the media was typically collected 7-10 days post-transfection by centrifugation at 3,000 g for 20 min. The supernatant was sterile filtered using Steritop^{TM} filter units (Millipore), supplemented with 10 mM Tris pH 7.2, and stored at

4 °C. In some cases, large-scale transient expression was instead performed by Yuguang Zhao (Strubi) using a TAP CompacT SelecT tissue culture robot, and either Nunclon triple flasks (500 cm^2, Thermo Scientific) or HYPERflasks (1720 cm^2, Corning).

Although mammalian cells, including HEK 293T cells, are ideal for producing correctly folded recombinant secreted and cell surface proteins of mammalian origin due to efficient post-translational processing in the secretory pathway, this processing can include the addition of complex and diverse *N*-linked oligosaccharides. These are problematic for crystallisation studies as the inherent flexibility and heterogeneity of the glycans contributes to disorder within the crystal lattice, preventing formation of well-diffracting crystals. To produce protein with as homogenous as possible a population of N-linked glycans for biophysical studies and crystallisation, the HEK 293S cell line was used, or a specific inhibitor of α-mannosidase I (kifunensine) was added to HEK 293T cell growth media. HEK 293S cells are deficient in the *N*-acetylglucosaminyltransferase I (GnTI) enzyme, meaning the *N*-linked glycosylation pathway in these cells only proceeds to a simplified $Man_5GlcNAc_2$ unit (Reeves et al., 2002). HEK 293T cells do not lack any enzymes in this pathway, but stalling at the high mannose glycan $Man_9GlcNAc_2$ can be induced by addition of 5 μM kifunensine (Toronto Research Chemicals) immediately after transfection (Elbein et al., 1990).

2.2.3 Expression of TCR constructs in E. coli *cells*

Constructs encoding the α and β chains of the MS49 TCR were expressed as inclusion bodies in BL21 (DE3) pLysS *E. coli* cells (Novagen), by Laura Millican (HIU, WIMM, Oxford). The expression of Hy TCR α and β chain constructs was trialled in small-scale cultures of Rosetta pLysS (Novagen) cells; Rosetta cells express tRNAs

to allow translation from codons found in eukaryotic sequences that are not commonly used in the prokaryotic host. 40 µl cells were transformed with approximately 10 ng plasmid DNA encoding either an α chain or β chain construct by incubating together on ice for 30 min, heat shocking for 30 s at 42 °C, then returning to ice for 2 min. 250 µl SOC media was added and the cells were incubated at 37 °C for 1 hour with 250 rpm shaking, then spread onto an LB-agar plate containing 100 µg/ml ampicillin and incubated at 37 °C overnight. Single colonies were selected and used to inoculate a 2 ml starter suspension culture of LB media containing 100 µg/ml ampicillin, which was incubated for approximately 6 hours at 37 °C with 250 rpm shaking. 300 ml LB media containing 100 µg/ml ampicillin was inoculated with the starter culture and grown at 37 °C with 250 rpm shaking overnight; 50 ml of this overnight culture was then added to each of six flasks of 500 ml LB media containing 100 µg/ml ampicillin and grown at 37 °C with 250 rpm shaking to an optical density of 0.8-1.0 at 600 nm, then induced with 1 mM isopropyl β-D-1 thiogalactopyranoside (IPTG; Melford Labs). The cultures were grown for 5 hours post-induction at 37 °C with 250 rpm shaking, then the bacteria were harvested by centrifugation at 4 krpm for 15 min at 4 °C; the cell pellet was stored at -80 °C.

2.2.4 Protein expression in stably-transfected insect cells

HLA-DQ6 and HLA-DM proteins were expressed in stably transfected *Drosophila melanogaster* S2 cells under control of the metallothionein promoter (Siebold et al., 2004). Attempts to express DQ6 in a prokaryotic system in a manner analogous to that described here for the TCRs (or elsewhere for other MHC class II molecules (Frayser et al., 1999; Li et al., 2000)) produced highly heterogeneous samples which were not suitable for crystallisation (McMahon, 2009), and the simpler *N*-linked

glycosylation pathway of insect cells can make them a better expression host than mammalian cells for producing samples for crystallisation (Chang et al., 2007). DQ6 protein was initially produced by Bjarke Hansen (Department of Clinical Immunology, Copenhagen University Hospital, Denmark) while DM was initially produced by Lise Torp Jensen at the Department of Clinical Immunology, Aarhus University, Denmark. Both DQ6 and DM were latterly produced by me, when the stably transfected cells were a kind gift from Bjarke Hansen and Lise Torp Jensen respectively.

The S2 cells were maintained in an adherent state in Sf900II media (Invitrogen) supplemented with 5 % FCS and hygromycin B (for DQ6; Invitrogen) or Geneticin® (for DM; Invitrogen), then expanded in suspension in Sf900II media supplemented with 5 % FCS; cells were seeded into media at approximately 0.8×10^6 cells/ml. Cells were incubated at 27 °C with shaking at 130 rpm, and induced with 700 µM $CuSO_4$ 1 day after seeding. The media was typically collected 4-7 days post-transfection by centrifugation at 3,000 g for 20 min, sterile filtered using Steritop^{TM} filter units, supplemented with a protease inhibitor cocktail (Sigma) and stored at 4 °C.

2.2.5 Protein purification from mammalian cell media

$FLRT2_{LRR}$ and $FLRT3_{LRR}$ protein were purified by me for the purposes of biophysical characterisation, and by Elena Seiradake (Strubi) for crystallisation; all other proteins described in this thesis that were produced using this method were purified by me.

Because the mammalian expression media contains traces of ethylenediamine tetra-acetic acid (EDTA) which would chelate the metal ions from the columns used in the initial stage of the purification, it was necessary to first thoroughly dialyse the media against 10 mM Tris pH 7.2, with the dialysate being refreshed at least 4 times

over 3-4 days. After dialysis, the media was subjected to immobilised metal affinity chromatography (IMAC) in order to purify the proteins using their hexahistidine tag. 2 ml of a 50 % slurry of Talon™ resin (Takara Bio) was added per litre of media, and incubated at 16 °C for 1-2 hours with gentle (100 rpm) shaking, before the beads were collected on a gravity flow column. The beads were washed thoroughly with 3 × 50 column volumes 20 mM Tris pH 7.2, 100 mM NaCl, followed by 30 column volumes 5 mM imidazole in 20 mM Tris pH 7.2, 100 mM NaCl; the protein was then eluted with 2 × 2 column volumes of 500 mM imidazole in 20 mM Tris pH 7.2, 100 mM NaCl. In some cases, this stage of the purification was instead performed by loading the media onto a 5 ml Ni-containing HisTrap™ HP column (GE Healthcare), washing with 20 mM Tris pH 7.2, 300 mM NaCl, 10 mM imidazole, and eluting with 20 mM Tris pH 7.2, 300 mM NaCl, 500 mM imidazole.

Fractions eluted either from the gravity flow or HisTrap™ columns were analysed by SDS-PAGE; those observed to contain the expressed construct (by comparison to a Benchmark protein ladder also loaded on the same gel) were pooled, and desalted into 20 mM Tris pH 7.5, 150 mM NaCl using a PD-10 desalting column (GE Healthcare) or an Amicon-Ultra centrifugal filter unit (Millipore). The sample was then concentrated down to approximately 2 ml for injection onto a Superdex S200 16/60 size exclusion column (GE Healthcare) equilibrated in 20 mM Tris pH 7.5, 150 mM NaCl, which acted as a second purification step to remove higher molecular weight contaminants and small quantities of aggregated protein. Fractions from the main protein peak were analysed by SDS-PAGE and those that appeared to be of sufficient purity (> 95 %) were pooled and concentrated as appropriate for either biochemical analysis or crystallisation. $FLRT2_{LRR}$ and $FLRT3_{LRR}$ protein prepared for crystallisation was supplemented with 25 mM ($FLRT2_{LRR}$) or 100 mM ($FLRT3_{LRR}$)

non-detergent sulphobetaine 256 (Hampton) to prevent protein aggregation during concentration; 10 µg/ml endoglycosidase F1 was added to these proteins once concentrated to remove any *N*-glycans prior to crystallisation (Grueninger-Leitch et al., 1996; Chang et al., 2007).

2.2.6 Solubilisation, in vitro *refolding and purification of Hy and MS49 TCR protein*

Purification of Hy constructs from inclusion bodies was performed by me; purification of MS49 TCR constructs was performed by Laura Millican (HIU, WIMM, Oxford). Frozen cell pellets were thawed and resuspended in PBS with protease inhibitors (Roche) and DNaseI (Sigma) to obtain a homogenous suspension. The cells were lysed by sonication (on power setting 8, in pulses of 30 s at 1 min intervals for a total time of 5 min) on ice, and crude inclusion bodies were isolated from the resulting homogenate by centrifugation at 15 krpm for 10 min at 4 °C. To wash the inclusion bodies, they were resuspended in 0.5 % Triton X-100, 50 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA to obtain a homogenous suspension, and pelleted by centrifugation at 15 krpm for 10 min at 4 °C. This step was repeated four times until the pellet was white and crumbly, when another wash step in 50 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT) was used to remove any residual Triton X-100 detergent; inclusion bodies were pelleted as before, and this step was repeated once. After washing, the inclusion bodies were solubilised by incubation in 6 M guanidine-HCl, 5 mM DTT, 100 mM Tris pH 8.0 at 4 °C with gentle agitation overnight. The guanidine acts as a chaotrope i.e. an agent which disrupts protein structure by destabilising salt bridge and hydrogen bond interactions, and by taxing the ability of solvent water molecules to effectively solvate all the chaotrope and protein molecules, leading to disordering of water molecules around the protein surface. This decreases the net hydrophobic effect which stabilises tertiary protein

folding and hydrophobic regions become solubilised, thereby denaturing the protein and allowing DTT (which is present to ensure any cysteine side chains are reduced so that no non-native intra- or interchain disulphide bonds can be formed) to fully reduce all cysteines, which might otherwise be shielded within a folded protein structure. Together, guanidine (or other chaotropes) and DTT ensure the proteins are fully denatured and reduced so that they enter the refolding process without any non-native (or native) tertiary structure or disulphide bonds; this aids optimisation of the yield of correctly refolded protein at the end of the process. Any remaining insoluble material was pelleted by centrifugation at 15 krpm for 30 min at 4 °C, and the supernatant containing solubilised protein was sterile filtered and stored in 1 ml aliquots at -80 °C.

The concentration of solubilised α and β chain proteins was assessed using a Nanodrop machine, or by comparison to known quantities of bovine serum albumin (BSA) on an SDS-PAGE gel. The proteins were mixed at a 1:1 ratio (25 mg each for Hy TCR; 90 mg each for MS49) at room temperature with 1 mM DTT for 15 min, then flash diluted by adding dropwise into 1 litre refolding buffer (for Hy: 4.5 M urea, 550 mM L-arginine, 100 mM Tris pH 8.0, 1 mM reduced glutathione, 100 µM oxidised glutathione; for MS49: 3 M urea, 200 mM L-arginine, 150 mM Tris pH 8.0, 1.5 mM reduced glutathione, 150 µM oxidised glutathione, 0.5 mM EDTA) at 4 °C, under strong vortex. At high concentration, the urea acts as a chaotrope to maintain the proteins in a denatured state, but also prevents association of refolding intermediates or unfolded species to suppress aggregation. L-arginine also increases refolded protein yield by suppressing non-specific aggregation, though the mechanism of this action remains incompletely understood. The reduced and oxidised forms of glutathione act as a redox couple, which can reduce and oxidise cysteine thiol

groups, allowing the protein to sample different refolding intermediate states with different intra- and interchain disulphide bond configurations so that ideally, a significant proportion of protein eventually reaches the most stable, native, arrangement. After 30-60 min (for MS49) or 40 hours (for Hy), the mixture was dialysed (using visking membrane, 8 kDa Mw cut off (Fisher)) at 4 °C against 15 litres 10 mM Tris pH 8.5; the dialysate was refreshed at least 6 times over 6-7 days, during which time the denaturing and reducing agents present in the refolding mixture are gradually removed, allowing the protein to refold into αβ dimers with native-like disulphide bond configuration and tertiary structure.

Following dialysis, refolded TCR was filtered on a 0.22 μm filter membrane (Appleton Woods) and loaded onto a 5 ml HiTrap™ Q HP anion exchange column (GE Healthcare) pre-equilibrated with 10 mM Tris-HCl pH 8.5, 10 mM NaCl; the protein was eluted with a 50 column volume gradient from 0 to 1 M NaCl in 10 mM Tris-HCl pH 8.5. Next, an IMAC purification step was performed to remove ββ homodimers by virtue of the C-terminal hexahistidine tag of the α chain, which binds αβ heterodimers to the IMAC column; αα homodimers also bind, but more strongly than αβ heterodimers due to the presence of two hexahistidine tags, so are eluted at a higher imidazole concentration. Imidazole was added to the pooled fractions from anion exchange to a final concentration of 10 mM; NaCl was also added to 500 mM. The pooled protein was then loaded onto a 1 ml HisTrap™ FF column (GE Healthcare) pre-equilibrated in 10 mM imidazole pH 8.0, 500 mM NaCl in PBS; protein was eluted with a 5 column volume gradient from 10 to 400 mM imidazole in 500 mM NaCl, PBS. Eluted fractions containing MS49 αβ TCR were pooled and desalted into 20 mM Tris pH 8.0, 100 mM NaCl using a PD-10 column (GE Healthcare) and concentrated to approximately 0.5 ml using an Amicon-Ultra

centrifugal filter unit. Because of the relatively low yields of this procedure and of complex formation with DQ6, several batches of refolded protein were necessary to obtain sufficient quantities of the MS49 TCR, so each batch was frozen in 50-60 % glycerol and stored at -20 °C.

2.2.7 Protein purification from insect cell media

For purification of DQ6, an affinity column was prepared by conjugation of the anti-DQ6 monoclonal SPVL3 antibody, prepared and kindly supplied by Ruth Etzensperger (HIU, WIMM, Oxford), to CnBr Sepharose (GE Healthcare) according to the recommended protocol; this column was stored in 0.1 % sodium azide, PBS at 4 °C and used for multiple batches of protein. The stored insect cell media was passed multiple (3-5) times over the beads in a gravity flow column to increase the yield, before the column was washed in three consecutive steps using:

i) 10 column volumes 0.5 % nonyl phenoxylpolyethoxylethanol (NP-40), PBS

ii) 10 column volumes 0.05 % NP-40, PBS

iii) 10 column volumes PBS.

DQ6 was eluted from the column using 3×2 column volumes 150 mM NaCl, 50 mM diethanolamine pH 11.5; for each fraction, the column was closed and incubated with the eluting buffer with gentle agitation for 15 mins in order to elute the maximum amount of protein, before being opened so that the eluate ran directly into 1/20 volume 2 M Tris-HCl pH 6.3 to neutralise the solution. Fractions from the elution were analysed by SDS-PAGE and those seen to contain DQ6 were pooled, buffer exchanged into 50 mM Tris pH 6.3, 100 mM NaCl using an Amicon-Ultra centrifugal filter unit and concentrated. As for MS49, several batches were stored in 60 % glycerol at -20 °C until sufficient quantities for complex formation and crystallisation had been accrued.

Soluble DM was immunoaffinity purified using the FLAG M2 epitope tag attached in place of the α chain transmembrane region (the β chain had a KT3 tag attached in the same manner), according to the methods described in (Sloan et al., 1995) and (Busch et al., 1998). Insect cell media was passed over a column of anti-M2 agarose (Sigma), equilibrated with 10 mM Tris pH 7.4, 150 mM NaCl. The column was then washed with 10 column volumes 10 mM Tris pH 7.4, 150 mM NaCl, and DM was eluted using 4×1 column volume 100 μM FLAG M2 peptide (Sigma) in 10 mM Tris pH 7.4, 150 mM NaCl. Elution fractions were pooled, concentrated and stored in 60 % glycerol at -20 °C.

2.3 Biophysical characterisation of soluble FLRT constructs

2.3.1 Characterisation of Hy TCR α1 produced in mammalian cells

To confirm whether multiple bands visible on SDS-PAGE gels of Hy α1 expressed in HEK 293T cells correspond to differentially glycosylated species of the protein, purified α1 was treated with Peptide: *N*-Glycosidase F (PNGase F; New England Biolabs), which removes all *N*-linked glycans. Digestion was performed according to the recommended protocol, under denaturing conditions to ensure that the protein was linearised and PNGase F could access all the sites of *N*-linked glycosylation. The digested protein was analysed by SDS-PAGE.

Having shown that it is possible to remove *N*-glycans from Hy α1, a sample of deglycosylated protein was submitted for intact mass spectrometry to confirm its identity and purity. A 20 μM sample of Hy α1 in 20 mM Tris pH 8.0, 100 mM NaCl was boiled at 100 °C for 10 mins, then treated with PNGase F at 37 °C overnight to allow complete deglycosylation. The sample was analysed by Joanne Nettleship (Oxford Protein Production Facility, Research Complex at Harwell) by liquid

chromatography-electrospray ionisation mass spectrometry using a quadrupole time-of-flight Micro mass spectrometer (Waters).

Analysis of α1 using intact mass spectrometry was not successful, apparently due to sample heterogeneity, and the alternative approach of N-terminal (Edman) sequencing was used. Purified α1 was run on an SDS-PAGE gel, and transferred to PVDF membrane using 10 mM 3-(cyclohexylamino)-1-propane sulphonic acid (CAPS), pH 11, 10 % (v/v) methanol as the blotting buffer; the electroblotting process was otherwise as described for semi-dry Western blotting in Section 2.2.1. After blotting, the membrane was washed, briefly stained with Coomassie Blue, rinsed and dried, and bands of interest were excised from the blot and thereafter analysed by Jeffrey Keen (Proteomics Facility, University of Leeds) using a liquid-pulse peptide sequencer.

Hy α1 did not appear to be amenable to Edman degradation, possibly due to exposure of an N-terminal glutamine residue following removal of the signal peptide during secretion, and so a final attempt at confirming the protein identity was made using mass spectrometry of a trypsin-digested sample. A band containing approximately 2 µg α1 protein was excised from an SDS-PAGE gel and analysed by analysed by Jeffrey Keen (Proteomics Facility, University of Leeds) using a MicroMass TOFSpec laser desorption machine (Waters).

2.3.2 Analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation (AUC) experiments were carried out with the assistance of Robert Gilbert (Strubi) using an Optima XL-I analytical ultracentrifuge (Beckman Instruments) at 20 °C, using a scanning absorbance of 280 nm and interference optics. For these analyses, proteins were purified as

described in Section 2.2.5, and $FLRT_{ECD}$ constructs were concentrated to 0.6 mg/ml in 20 mM Tris pH 8.0, 100 mM NaCl, while the $FLRT1_{LRR}$ construct was concentrated to 0.2 mg/ml in two buffers used to simulate the distinct pH values of the different conditions under which this construct crystallised: (i) 20 mM Tris pH 8.0, 100 mM NaCl and (ii) 20 mM Tris pH 5.0, 100 mM NaCl. Samples were held in 12 mm Epon sector-shaped 2-channel centrepieces and equilibrated at 20 °C in the rotor before the velocity was increased to 40 krpm, with 50 scans taken at 6 min intervals. SedFit software (Schuck, 2000) was used to analyse the results by fitting Gaussian curves to the c(s) plots obtained, and to describe the distribution of species in the solution environment.

2.3.3 Multi-angle light scattering

Multi-angle light scattering (MALS) experiments were carried out on a Wyatt MALS/AFFFF System (Wyatt Technologies) with the assistance of Geoff Sutton and Elena Seiradake (both Strubi). Proteins were purified as described in Section 2.2.5, then $FLRT_{ECD}$ constructs were concentrated to 5.1 mg/ml ($FLRT1_{ECD}$), 7.1 mg/ml ($FLRT2_{ECD}$) and 9.7 mg/ml ($FLRT3_{ECD}$) while $FLRT_{LRR}$ constructs were concentrated to approximately 2.9 mg/ml; separate samples of $FLRT1_{LRR}$ were concentrated to 4.7 mg/ml and 40.2 mg/ml, all in 20 mM Tris pH 7.5, 150 mM NaCl. Separation for MALS by size exclusion chromatography was carried out on an analytical Superdex 200 10/30 column (GE Healthcare) in 20 mM Tris pH 7.5, 150 mM NaCl, and the eluate was passed through online static light scattering (DAWN HELEOS II, Wyatt Technology), differential refractive index (Optilab rEX, Wyatt Technology) and Agilent 1200 UV (Agilent Technologies) detectors. Data were analysed using the ASTRA software package (Wyatt Technology).

2.4 Crystallisation

2.4.1 Formation of an MS49 TCR-MBP/DQ6 complex

The conditions (pH, temperature and ratio of components) used for formation of a complex of the MS49 TCR (including an MBP_{30-44} peptide attached to the β chain by a flexible glycine linker) with DQ6 were informed by work described in (McMahon, 2009).

To compensate for the absence of their transmembrane regions and permit expression in an insect cell system, the ectodomains of the DQ6 α and β chains were fused at their C-termini to the leucine zipper dimerisation domains of the Fos and Jun transcription factors, respectively (Kalandadze et al., 1996), *via* a glycine-rich linker. The zipper domains each form an α helix, with regularly-spaced leucine residues arranged at the interface between the two zippers allowing them to form a stable interaction and assemble as a coiled coil, whilst charged residues on the domain surface contribute to solubility. This is advantageous for expression and purification, but the inherent flexibility of the zippers (including the linker attaching them to the α and β chains) is not conducive to successful crystallisation. They were therefore removed prior to complex formation by digestion with endoproteinase Glu-C (V8 protease, Roche), which cleaves at an aspartic residue at the N-terminus of each zipper domain. Glycerol was removed from stored DQ6 by buffer exchange in an Amicon-Ultra centrifugal filter unit into 20 mM Tris pH 7.4, 50 mM NaCl and incubated at a ratio of 60:1 (μg/μg) with Glu-C overnight at room temperature with gentle agitation. Removal of the zipper domains results in a loss of solubility, and significant insoluble precipitate was removed by centrifugation at 10 krpm for 10 min at 4 °C. As Glu-C also acts on an internal site in the MS49 TCR, it was necessary to remove it prior to complex formation, and this was done by anion exchange chromatography,

which also removed the digested zippers. The clarified supernatant was loaded onto a 1 ml MonoQ column (GE Healthcare), which was washed with 20 mM Tris pH 7.4, 100 mM NaCl before the bound DQ6 was eluted with a 15 column volume gradient from 100 mM to 1 M NaCl in 20 mM Tris pH 7.4. Finally, the eluate was buffer exchanged into the complex formation buffer of 50 mM MES pH 5.6, 100 mM NaCl.

HLA-DM, which is included to mediate exchange of the TCR-linked peptide into the binding groove of DQ6 forming a stabilised complex, and MS49 were removed from storage glycerol by buffer exchange, using an Amicon-Ultra centrifugal filter unit, into 50 mM MES pH 5.6, 100 mM NaCl (a low pH buffer being beneficial for the peptide exchange activity of DM (Sloan et al., 1995)). MS49, DQ6 and DM were then combined at a molar ratio of 9: 3: 1 (actual masses for the attempt resulting in crystal formation were 13.8: 3.76: 1.22 mg) representing an excess of TCR which pushes the reaction towards occupancy of the limiting empty DQ6 binding grooves. The mixture was concentrated to a final volume of 1-2 ml and incubated for 15 hours at 24 °C, this temperature having been found to be a compromise between efficiency of complex formation and loss of protein, particularly the TCR, by precipitation. However, there was still a significant amount of precipitate, which was removed by centrifugation at 10 krpm for 10 min at 4 °C. The supernatant containing soluble protein, including the MS49-DQ6 complex, was injected onto an S200 16/60 gel filtration column pre-equilibrated with 20 mM Tris pH 7.4, 100 mM NaCl. Fractions from the much smaller peak shown to correspond to the complex (McMahon, 2009) were pooled and concentrated for crystallisation using a centrifugal filter unit that had previously been used to buffer exchange and concentrate MS49, in order to avoid protein loss by binding to the filter.

2.4.2 Crystallisation screens

Crystallisation screens for $FLRT2_{LRR}$ and $FLRT3_{LRR}$ were set up by Elena Seiradake (Strubi); other trials described below were set up by me.

Of the protein prepared as described in sections 2.2.5 and 2.4.1, only $FLRT1_{LRR}$ was produced with a high enough yield to submit it to a pre-crystallisation test (Hampton Research) to determine an appropriate protein concentration to proceed to crystallisation condition screening; the yield also meant a broader screen of conditions could be tested. Other proteins (the Hy TCR α1 construct and the MS49-DQ6 complex) were simply concentrated as far as possible (down to a minimum volume of 17 µl, as this was the lowest possible volume for the operation of a Cartesian machine) and immediately set for crystallisation trials, the number of which were limited by the amount of protein available. Protein concentration was monitored by absorbance of light by the protein solution at 280 nm using a Nanodrop machine.

Trials were set up in 96-well Greiner plates using a Cartesian Technologies Microsys Mic4000 robot, in drops of 100 nl protein plus 100 nl mother liquor, using the sitting drop vapour diffusion method (Walter et al., 2005). Plates were stored either at 20.5 °C in a TAP Homebase storage vault and imaged *via* a Veeco visualisation system (Mayo et al., 2005) or equilibrated at 4 °C before being stored and imaged in a Formulatrix Rock Imager 1000, at 6.5 °C. The standard commercial screens tested by me for various protein constructs described in this thesis are listed in Table 2.1, with conditions that produced crystals that were tested at a synchrotron detailed in Table 2.2.

Protein	Master block screen	Wells	Commercial screen	Manufacturer
$FLRT1_{LRR}$	Block 1	A1-D12	Crystal Screen	Hampton Research
		E1-H12	Crystal Screen 2	Hampton Research
	Block 2	A1-D12	Wizard 1	Emerald Biosystems
		E1-H12	Wizard 2	Emerald Biosystems
	Index	A1-H12	Index	Hampton Research
	Silver Bullets (pH 5.8, 6.8 and 7.8)	A1-H12	Silver Bullets	Hampton Research
Hy TCR α1	Block 1	As above		
	Index	As above		
	Limited Protein	A1-C10	Crystal Screen	Hampton Research
		C11-D12	Crystal Screen 2	Hampton Research
		E1-H12	PEG/Ion Screen	Hampton Research
	PACTpremier	A1-H12	PACTpremier	Molecular Dimensions
MS49-DQ6 complex	Block 1	As above		
	Block 2	A1-D12	Wizard 1	Emerald Biosystems
		E1-H12	Wizard 2	Emerald Biosystems
	Block 3	A1-D12	PEG/Ion Screen	Hampton Research
		E1-F12	Grid Screen PEG 6000	Hampton Research
		G1-H12	Grid Screen Ammonium Sulphate	Hampton Research
	Block 4	A1-D12	Natrix Screen	Hampton Research
		E1-H12	Crystal Screen Cryo	Hampton Research
	Block 5	A1-B12	Grid Screen PEG/LiCl	Hampton Research
		C1-D12	Grid Screen NaCl	Hampton Research
		E1-F12	Grid Screen MPD	Hampton Research
		G1-H12	Quik Screen	Hampton Research
	Index	As above		
	Limited Protein	As above		

Table 2.1 Crystallisation screens tested for the protein constructs described in this thesis that were submitted for crystallisation trials by me. Screens that yielded crystals that were tested at synchrotron beamlines are highlighted in yellow (details of conditions in Table 2.2).

2.4.3 Cryo-protection and crystal mounting

Crystals were mounted by Elena Seiradake (Strubi), or with the assistance of Karl Harlos and Maria Harkiolaki (both Strubi). They were removed from the Greiner plates using Hampton CryoLoops of a size appropriate to that of the crystal. Cryoprotectant solutions of 10-25 % glycerol in reservoir solution were added to well platforms adjacent to the crystallisation drops in the Greiner plates so that the crystals could be briefly immersed in them then immediately vitrified at 100 K in liquid nitrogen; in the case of MS49-DQ6 needle crystals, cryoprotectant was added to the crystallisation drop directly to allow the crystals to be fished more easily.

Protein construct	**Purification procedure**	**Protein concentration (mg/ml)**	**Crystallisation condition**	**Temperature (°C)**	**Freezing**	**Synchrotron and beamline**
$FLRT1_{LRR}$	As described in Section 2.2.5, but only up to the point of affinity chromatography and desalted into 20 mM Tris pH 8.0, 100 mM NaCl	13.4	30 % (w/v) PEG monomethyl ether 2000, 0.1 M sodium acetate pH 4.6, 0.2 M ammonium sulphate (Hampton Crystal Screen 2)	20.5	20 % glycerol	ESRF ID23.2
	As for row above	13.4	8 % (v/v) ethylene glycol, 10 % (w/v) PEG 8000, 0.1 M HEPES pH 7.5	20.5	10 % glycerol	ESRF ID23.2
	As described in Section 2.2.5, except gel filtration performed in 20 mM Tris pH 8.0, 100 mM NaCl	13.2	60 mM MES monohydrate, 60 mM PIPES, 0.33 % (w/v) hexamminecobalt(III) chloride, 20 mM HEPES pH 6.8, final pH 7.8 (Silver Bullet screen, Hampton)	20.5	20 % glycerol	DLS I24
	As for row above	13.2	0.2 % (w/v) 1,4-diaminobutane,	20.5	20 % glycerol	DLS I24

			0.2 % cystamine dihydrochloride, 0.2 % diloxanide furoate, 0.2 % sarcosine, 0.2 % spermine, 20 mM HEPES pH 6.8, final pH 7.8 (Silver Bullet screen, Hampton)			
	As described in Section 2.2.5	16.6	60 mM MES monohydrate, 60 mM PIPES, 0.33 % (w/v) hexamminecobalt(III) chloride, 20 mM HEPES pH 6.8, final pH 7.8 (Silver Bullet screen, Hampton)	20.5	20 % glycerol	DLS I24
$FLRT2_{LRR}$	As described in Section 2.2.5, except $FLRT2_{LRR}$ combined with Unc5D in a 1:1 ratio prior to crystallisation	6.0	19 % PEG 4000, 19 % 2-propanol, 5 % glycerol, 95 mM tri-sodium citrate pH 5.6 (Crystal Screen Cryo, Hampton)	20.5	20 % glycerol	DLS I24
$FLRT3_{LRR}$	As described in Section 2.2.5	2.0	20 % PEG 3350, 0.2 M ammonium nitrate, final pH 7.4 (PEG/Ion Screen, Hampton)	20.5	25 % glycerol	DLS I03
MS49 - DQ6	As described in Sections 2.2.6, 2.2.7 and 2.4.1	8.6	20 % PEG 1000, 0.2 M NaCl, 0.1 M sodium/potassium phosphate pH 6.2, final pH 6.5 (Wizard Screen 2, Emerald)	20.5	25 % glycerol (added to drop)	DLS I24

Table 2.2 Crystallisation and freezing conditions for crystals tested at a synchrotron (ESRF: European Synchrotron Radiation Facility, France; DLS: Diamond Light Source, Oxfordshire). Those that gave diffraction leading to crystal structures presented in this thesis are highlighted in green.

2.5 Structure solution for $FLRT_{LRR}$ constructs

2.5.1 Data processing

Data for $FLRT2_{LRR}$ and $FLRT3_{LRR}$ were processed by Elena Seiradake (Strubi). Data for $FLRT1_{LRR}$ in crystal form 1 ("antiparallel") were processed by Maria Harkiolaki (Strubi); data for $FLRT1_{LRR}$ in crystal form 2 ("parallel") were processed by me with assistance from Maria Harkiolaki. Diffraction images were indexed to determine the point group and unit cell dimensions which describe the arrangement of molecules within the crystal lattice, prior to integration, merging and scaling of the reflection intensities using either the HKL2000 (Otwinowski and Minor, 1997), xia2 (Winter, 2010) or XDS (Kabsch, 2010) data processing suites. Data quality was assessed according to criteria including R_{merge}, I/σI, completeness and redundancy, statistics for which can be found in Tables 5.2 and 5.3; formulae for the calculation of these criteria can be found in Appendix C.

2.5.2 Molecular replacement

All $FLRT_{LRR}$ structures could be solved by molecular replacement; the structures of $FLRT2_{LRR}$ and $FLRT3_{LRR}$ were solved by Elena Seiradake (Strubi), the structure of $FLRT1_{LRR}$ in crystal form 1 ("antiparallel") was solved by Maria Harkiolaki (Strubi), and the structure of $FLRT1_{LRR}$ in crystal form 2 ("parallel") was solved by me with assistance from Maria Harkiolaki.

Calculation of the Matthew's coefficient by the matthews_coef programme in the CCP4 suite (Collaborative Computational Project 4, 1994; Kantardjieff and Rupp, 2003) provided an estimate of the number of protein molecules in the asymmetric unit, by identifying the number of monomers equating to a solvent content in the 40-60 % range typically observed for protein crystals (Matthews, 1968). The asymmetric unit is the smallest unit the crystal may be divided into using the

crystallographic symmetry operations inherent to the space group; multiple copies of the asymmetric unit are arranged within the unit cell, which then stacks in three-dimensional space to produce the crystal. The structure of $FLRT1_{LRR}$ in crystal form 1 (1 molecule per asymmetric unit) was solved in AMoRe (Navaza, 1994) using the 2.80 Å resolution structure of the third LRR domain of *Drosophila* Slit 2 ((Howitt et al., 2004); PDB ID 1W8A). As the structure of $FLRT1_{LRR}$ in crystal form 2 (2 molecules per asymmetric unit) was solved after the structure in crystal form 1 had been solved, this "antiparallel" model (stripped of solvent and glycans) was used as a search model for crystal form 2 in Phaser (McCoy et al., 2007). $FLRT2_{LRR}$ (1 molecule per asymmetric unit) and $FLRT3_{LRR}$ (2 molecules per asymmetric unit) were also solved in this way, using $FLRT1_{LRR}$ in crystal form 1 as a search model for $FLRT3_{LRR}$, and subsequently $FLRT3_{LRR}$ as a search model for $FLRT2_{LRR}$. Validating the solutions, unmodelled features, such as extra amino acid residues at the N- and C-termini of the model, could be observed in the electron density map, and subsequent rigid body refinement in Phenix (Adams et al., 2010) further improved the map.

2.5.3 Refinement

The $FLRT2_{LRR}$ and $FLRT3_{LRR}$ models were refined by Elena Seiradake (Strubi). Both $FLRT1_{LRR}$ structures were refined by me with assistance from Maria Harkiolaki (Strubi). From the initial solutions obtained by molecular replacement, the structures were first built automatically using either the Phenix Autobuild interface (Terwilliger et al., 2008) or the Rosetta implementation in Phenix (DiMaio et al., 2009; DiMaio et al., 2011), which corrected the amino acid sequences and extended the models beyond the original solutions. This was followed by multiple rounds of manual inspection and adjustment, with localised real-space and geometry-restrained refinement using Coot

(Emsley and Cowtan, 2004), interspersed by global restrained refinement of the atomic positions and B-factors against the observed structure factors using Phenix (Adams et al., 2010) and autoBuster (Blanc et al., 2004). The agreement of the model structure factors with the data structure factors was monitored throughout using the R_{work} and R_{free} parameters (for statistics see Tables 5.2 and 5.3; see also Appendix C for formulae for their calculation), where decreasing of both values and a convergence between the two indicates an improvement in model quality. Early in the refinement of $FLRT1_{LRR}$ crystal form 2, which contains 2 molecules in the asymmetric unit, and $FLRT3_{LRR}$, which contains 2 molecules in the asymmetric unit, tight global non-crystallographic symmetry (NCS) was imposed using Phenix (Adams et al., 2010), in order to reduce the number of refinement parameters; these restraints were relaxed later in the refinement.

Water molecules were added to the higher resolution $FLRT1_{LRR}$ crystal form 1 model using Phenix (Adams et al., 2010) and manually inspected using Coot (Emsley and Cowtan, 2004). Idealised co-ordinate and restraint files for the *N*-linked glycans added to this model were obtained from the HIC-Up online server (Kleywegt, 2007).

2.5.4 Structure validation

The final refined models were validated by assessment of their stereochemical quality using Coot (Emsley and Cowtan, 2004), MolProbity (Davis et al., 2007) and the RAMPAGE server (Lovell et al., 2003); for statistics, see Chapter 5. Firstly, the model geometry was assessed by comparison of main chain dihedral angles to a Ramachandran plot; all residues bar a single glycine in the $FLRT3_{LRR}$ model were within an "allowed area", with at least 92.5 % falling within a "favoured" region. Secondly, the probability of the orientation of each side chain compared to a library

compiled from high resolution structures was used to correct any unlikely rotamers that were not supported by the electron density.

2.5.5 Structure analysis

The buried surface area for the various $FLRT_{LRR}$ homotypic interactions suggested by the crystal structures was calculated by the online European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) protein interfaces, surfaces and assemblies (PISA) server (Krissinel and Henrick, 2007). This also allowed calculation of the free energy change associated with formation of each interface. SHP (Stuart et al., 1979) was used to superpose models, based on C_α positions, to compare the dimensions and shapes of the different LRR domains, and to calculate the root mean squared deviation of the distance between equivalent C_α positions.

2.6 Fluorescence microscopy

2.6.1 HEK 293 cell fluorescence microscopy

For imaging experiments, HEK 293 cells were cultured by Jeff McIlhinney (Medical Research Council (MRC) Anatomical Neuropharmacology Unit, Department of Pharmacology, University of Oxford) at 37 °C, 5 % CO_2 in DMEM (Sigma) supplemented with 10 % (v/v) foetal calf serum (FCS, Biosera), L-glutamine and penicillin/streptomycin (Gibco). 24 hours prior to transfection, cells were subcultured onto 22 mm diameter borosilicate glass coverslips that had been flame treated and subsequently coated with poly-D-lysine. Transfection was performed using Lipofectamine 2000 (Invitrogen) using the recommended protocol, with 2.5 µg plasmid DNA and 10 µl Lipofectamine in a final solution of 400 µl Optimem (Gibco). Of this solution, 200 µl was added to each well of a six-well plate.

Cells were assayed by Jeff McIlhinney (MRC Anatomical Neuropharmacology Unit, Oxford) at 24 and 48 hours after transfection. Cells were washed twice with PBS before being fixed with 4 % (w/v) *p*-formaldehyde in PBS for 5 min at room temperature. They were then washed with 50 mM Tris-HCl pH 7.4, 100 mM NaCl, and incubated for 5 min in the same buffer at room temperature. The glass coverslips were mounted on glass slides in Vectashield (Vector Laboratories, Peterborough, UK) and imaging performed using a Zeiss LSM 510 confocal microscope. All images were taken using a X40 parafocal oil immersion lens, using optical slices of 1 µm. Z-stack images were taken at the same resolution using the LSM Imaging and ImageJ software packages for acquisition and processing.

In order to acquire quantitative data on relative levels of fluorescence at cell-cell interfaces and other non-interfacing regions, images were analysed using ImageJ. While measurements of average fluorescence density in a given area of membrane would have been preferable, all the requisite three-dimensional information for this was not available from the images, so data was measured from long, narrow regions selected to delineate cell-cell interface and non-interfacing regions; regions were 1 µm wide and 7-15 µm long depending on the image being sampled. Measurements of the mean grey value, a numerical measurement of the fluorescence brightness and therefore amount of tagged FLRT or CD45 protein, were taken with regions of equal size in each of three parts of the image: a cell-cell interface, a region of cell membrane not engaged in a cell-cell interaction, and a region of the image devoid of any cells to allow for subtraction of background noise. Measurements were taken from 10 images for each construct, corrected by subtraction of the background signal, and an average value calculated for interfacing and non-interfacing regions from each group.

2.6.2 Sf9 cell clustering assay

Sequence verified $FLRT_{LONG}$ constructs, containing the ectodomain, transmembrane helix and 20 intracellular residues with C-terminal mVenus tag (Nagai et al., 2002) were sub-cloned from pHLsec to pBacPAK9 as described in Section 2.1.1 for transfection into Sf9 insect cells. Sf9 cells were maintained in a suspension state by Weixian Lu (Strubi), in Sf900II media (Invitrogen) supplemented with penicillin and streptomycin (Sigma) at 27 °C with 130 rpm shaking.

Recombinant baculovirus was produced by co-transfection of the cells with linearised baculovirus DNA and the pBacPAK9 transfer vector. Approximately 6×10^5 cells were seeded into six-well plates in Sf900II media. For transfection, 100 µl serum-free Sf900II media containing 0.5 µg viral DNA and 2 µg plasmid DNA was gently mixed with 100 µl serum-free Sf900II media containing 3 µl Fugene (Roche) or 8 µl CellfectinII (Invitrogen) transfection reagent, and incubated at room temperature for 20 min. 800 µl Sf900II media was added to these transfection mixes, and the resulting 1 ml used to replace the media from each well. The plates were incubated at 27 °C with shaking at 130 rpm for 24 hours before another 1 ml of Sf900II media was added to each well, and the virus was left to replicate for approximately 7 days. Some of this initial stock was diluted 1: 5 in fresh media and the virus left to amplify further for four days, when this dilution step was repeated. 2 days after the second dilution, 2 ml cells with each virus construct were transferred into wells of a six-well plate, and incubated at 27 °C for at least 1 hour to allow clustering.

Images were acquired using a wide-field inverted Nikon TE2000U fluorescence microscope system equipped with a monochrome CCD camera (Hamamatsu), captured and processed using IPLab imaging software, and subsequently pseudocoloured.

2.7 Investigation of cell surface expression of FLRTs

2.7.1 Fluorescence activated cell sorting

To confirm surface expression of epitope-tagged (with N-terminal FLAG M2 or HA peptides) transmembrane FLRT, and control, constructs, HEK 293T cells were analysed by extracellular staining for flow cytometry (FACS ECS), where antibody staining of the cell surface is detected by a cell sorting machine; this was done with the assistance of Veronica Chang (WIMM, Oxford). HEK 293T cells were cultured and transfected as described in Section 2.2.1, then pelleted in FACS tubes 24 hours after transfection, washed by resuspending in PBS containing 5 % FCS, 1 % BSA, 0.05 % sodium azide, and re-pelleted. The cells were fixed in PBS containing 1 % *p*-formaldehyde for 15 min, and washed in PBS containing 0.05 % sodium azide before being re-pelleted. They were then incubated with mouse anti-FLAG M2 (Sigma) or mouse anti-HA (Sigma) monoclonal antibody, as appropriate to the construct being tested, diluted to 5 µg/ml in PBS containing 0.05 % sodium azide, for 40 min at 4 °C. Cells were washed in PBS containing 0.05 % sodium azide and pelleted, and this wash step was repeated a further two times to prevent any non-specific binding of the primary antibody. They were then incubated with an Alexa-647 donkey anti-mouse antibody (Invitrogen), diluted to 5 µg/ml in PBS containing 0.05 % sodium azide, for 30 min at 4 °C. The cells were then washed and pelleted three times as before, and finally resuspended in PBS containing 2 % *p*-formaldehyde. 60,000 cells were collected in a CyAn FACS machine (Dakocytomation, Glostrup, Denmark), and the data were analysed using the FlowJo software package (Treestar, Ashland, OR).

2.7.2 Immunocytochemistry

For additional confirmation of surface expression, the surface of Cos-7 cells transfected with epitope-tagged transmembrane $FLRT_{LONG}$ constructs was stained using antibodies against the N-terminal epitope tags (FLAG M2 or HA peptides), and then with fluorescent secondary antibodies which could be observed by microscopy. Cos-7 cells were cultured at 37 °C, 5 % CO_2 in DMEM supplemented with 10 % (v/v) FCS (Sigma), and subcultured onto 22 mm diameter borosilicate glass coverslips coated in-house with poly-L-lysine (Sigma) 24 hours prior to transfection. They were transfected as described for HEK 293 cells in Section 2.2.1. 48 hours after transfection, the medium was aspirated and the cells carefully washed in PBS, then fixed with PBS containing 4 % (w/v) *p*-formaldehyde for 15 min at room temperature. The fixed cells were washed in 50 mM Tris pH 7.4, 150 mM NaCl for 10 min to decrease auto-fluorescence in the final sample, and then incubated in PBS containing 1 % (w/v) BSA for 15 min to block non-specific binding sites. Each coverslip was then overlaid with 1 ml primary antibody (mouse anti-FLAG M2 (Sigma) or mouse anti-HA (Sigma)) diluted to 5 µg/ml in PBS containing 3 % BSA. After incubating for at least 1 hour in a moist chamber, coverslips were gently washed three times with PBS containing 0.1 % BSA for 5 mins each, then incubated with an Alexa-568 goat anti-mouse fluorescent secondary antibody (Invitrogen) diluted to 5 µg/ml in PBS containing 3 % BSA for 1 hour, in the dark. Coverslips were then washed three times in PBS containing 0.1 % BSA for 5 min each, rinsed with distilled water and mounted on microscope slides in Vectashield mounting medium. They were then examined using either a wide-field inverted Nikon TE2000U fluorescence microscope or a Zeiss LSM 510 confocal microscope, as described in Sections 2.6.2 and 2.6.1 respectively.

2.8 Investigation of cell surface homodimerisation of FLRTs

2.8.1 Cell surface co-patching assay

To investigate whether transmembrane FLRT, and control protein, constructs associate in a homotypic manner at the cell surface, an assay was employed where Cos-7 cells were co-transfected with two constructs differing in their N-terminal epitope tags (FLAG M2 or HA peptides), then stained live firstly with an antibody against one of the tags, and then with a secondary antibody to pull protein molecules with that tag into patches on the cell membrane. Cells were fixed to preserve the arrangement of proteins in the membrane and bound antibodies, and then stained with an antibody against the other tag (or control protein), followed by another fluorescent secondary antibody with a different coloured dye to distinguish the two cell surface proteins. Where the different dyes are seen to "co-patch" in the same areas of the cell membrane, this may be evidence for an association between the two constructs at the cell surface, as the second protein must be pulled into patches with the first by an intermolecular association.

Cos-7 cells were cultured for this purpose by Jeff McIlhinney (MRC Anatomical Neuropharmacology Unit, Oxford), as described in Section 2.7.2. They were transfected as described in Section 2.7.2, with a 1:1 mixture of sequence verified plasmid DNA encoding transmembrane FLRT constructs with HA and FLAG M2 N-terminal tags, or with a 1:1 mixture of plasmid DNA encoding one tagged FLRT construct and a control protein (mouse CD2). 24 hours post-transfection, the medium was aspirated and the coverslips were gently washed with PBS, then incubated with PBS containing 1 % BSA for 15 min at room temperature. The plates were then put on ice, and the cells incubated with either mouse anti-FLAG M2, or mouse anti-HA antibodies diluted to 1 µg/ml in cold PBS containing 1 % BSA, for

30 min with rocking. The cells were washed three times with cold PBS containing 0.1 % BSA, for 5 min each time, and then removed from ice to a 37 °C incubator. They were incubated with an Alexa-488 goat anti-mouse antibody (Invitrogen) diluted to 1 µg/ml in room temperature PBS containing 1 % BSA, in the dark; this staining without fixing should allow proteins with the appropriate tag to be brought together in patches on the cell membrane, as the antibodies can bind multiple epitopes. The coverslips were then washed three times with room temperature PBS, for 5 min each time, and the cells were fixed using PBS containing 4 % *p*-formaldehyde for 5 min to fix any patches of protein in place on the cell surface. The fixed cells were washed with 50 mM Tris pH 7.4, 150 mM NaCl for 5 min at room temperature, and incubated with rabbit anti-HA or rabbit anti-CD2 antibodies diluted to 1 µg/ml in PBS containing 1 % BSA, to bind the tag on the construct that had not already been stained. The cells were washed three times in PBS containing 0.1 % BSA for 5 min each time, then incubated with an Alexa-568 anti-rabbit antibody diluted to 1 µg/ml in PBS containing 1 % BSA for 15 min, in the dark. This effectively stained the second tag a different colour (red) to the first tag, which had been stained green prior to fixing; where the red stain co-localises with the patches of green stain, this is evidence for an association between the two tagged constructs. Finally, coverslips were washed three times with PBS for 5 min each, rinsed with distilled water and mounted onto microscope slides with Vectashield mounting medium. They were examined using a Zeiss LSM 510 confocal microscope as described in Section 2.6.1, but with a X63 oil immersion lens, and the fluorescent images were superposed using the LSM Imaging and ImageJ software packages.

In order to acquire quantitative data on the degree of co-localisation of red and green staining, images were analysed using ImageJ with the Co-localisation Indices plug-in

(Nakamura et al., 2007). This allows calculation of the Pearson's correlation coefficient, a measure of the strength of correlation between green and red fluorescence which varies from -1 to 1; values close to zero imply a poor correlation and low co-localisation of the fluorescence markers, while values closer to -1 or 1 indicate strong negative or positive correlations respectively. This analysis was applied to whole merged images, with measurements taken from five images for each set of conditions; an average Pearson's coefficient value was then calculated for each group.

2.8.2 Fluorescence resonance energy transfer

Fluorescence resonance energy transfer (FRET) can be used as a "molecular ruler" to probe the distances between two molecules which each include a different member of a "FRET pair" of fluorophores, such as monoVenus and monoCerulean; a fuller explanation of this phenomenon can be found in Chapter 5. Briefly, measuring FRET by acceptor photobleaching involves monitoring fluorescence from the "donor" fluorophore before and after the "acceptor" fluorophore is selectively photobleached, and an increase can be taken as evidence of close proximity between the FRET pair of fluorophores, possibly owing to an association between the protein molecules to which they are attached. In this way, the degree of FRET occurring between transmembrane FLRT constructs and control proteins with C-terminal mVenus and mCerulean tags (mCerulean and mVenus are a FRET pair, mCerulean being the donor and mVenus the acceptor) could be measured as a way to study whether the proteins associate in a homotypic manner at the cell surface.

FRET experiments were performed in Cos-7 cells with the assistance of Jeff McIlhinney (MRC Anatomical Neuropharmacology Unit, Oxford). Cos-7 cells were cultured as described in Section 2.7.2, and transfected as described in Section 2.7.2

with a 1:1 mixture of sequence verified plasmid DNA encoding transmembrane FLRT constructs with mVenus and mCerulean C-terminal tags, or with a 1:1 mixture of plasmid DNA encoding an mVenus-tagged control protein and an mCerulean-tagged FLRT construct. After 24 hours, they were washed with PBS and fixed with PBS containing 4 % (w/v) *p*-formaldehyde for 15 min. The fixed cells were washed in 50 mM Tris pH 7.4, 150 mM NaCl for 10 min, and rinsed with PBS and distilled water before being mounted onto microscope slides using Vectashield mounting medium. The fixed cells were imaged on a Zeiss LSM510 scanning confocal microscope using a 40x oil immersion lens. The mVenus and mCerulean were excited using the 514 and 458 nm lines of the He-Ne laser respectively, and the bleaching of the mVenus carried out by repetitive scanning (30 iterations) of the region of interest (ROI) of the cells using 90% of full laser power at 514 nm. The mVenus emission was collected through a long pass LP530 nm filter and the mCerulean fluorescence through a band pass filter BP480-520 nm. ROIs were defined on the cell surface membrane, avoiding the Golgi and ER, using a pinhole of 160 µm. After taking three 1 s scans across the ROI, it was bleached and three additional scans were collected. The FRET signal was calculated as the fractional increase in the mCerulean fluorescence signal in the ROI after the bleach thus:

$$[(\text{mCerulean}_{\text{post-bleach}} - \text{mCerulean}_{\text{pre-bleach}})/\text{mCerulean}_{\text{pre-bleach}}] \times 100$$

Data were collected from the number of cells shown in Table 5.1, during three independent experiments.

2.9 Figures

Sequence alignments were performed using ClustalW (Larkin et al., 2007) and formatted using ESPript (Gouet et al., 1999). These alignments were also used for amino acid sequence-based conservation analysis, which was performed using the ConSurf server (Ashkenazy et al., 2010). Data from the FPLC software Unicorn™ (ÄKTAFPLC™, GE Healthcare) was transferred into Microsoft Excel to produce figures for chromatography experiments. Crystallography figures were generated using Coot (Emsley and Cowtan, 2004) and PyMOL (DeLano, 2008), and when appropriate, in combination with electrostatic potential data calculated using the adaptive Poisson-Boltzmann solver (APBS; (Baker et al., 2001)) and conservation data calculated using the ConSurf server. Other illustrations were prepared using Microsoft PowerPoint.

Chapter 3 TCR and MHC Class II Production and Complex Formation

Chapter 3 TCR and MHC Class II Production and Complex Formation

3.1 Introduction

Crystal structures of cross-reactive TCRs in complex with pMHC have made a crucial contribution to our understanding of molecular mimicry and the structural mechanisms of autoimmunity, particularly in the case of MS (see Section 1.2). For example, the Hy.1B11 TCR, isolated from a relapsing-remitting MS patient, was crystallised in complex with the immunodominant epitope (residues 85-99) of MBP, presented by HLA-DQ1 (Sethi et al., 2011), but the TCR clone is also activated by four microbial peptides with limited sequence similarity to MBP_{85-99} (Wucherpfennig and Strominger, 1995). Two further *in vivo* expanded T cell clones, specific for MBP_{85-99}, were also isolated from the same MS patient, and the recombinant production of the TCR expressed by of these, Hy.2E11 (Hy), for structural and biophysical characterisation, is the subject of the first part of this chapter. Hy TCR is highly cross-reactive and recognises several viral peptides, including an EBV DNA polymerase peptide, $EBV_{627-641}$, in the context of DR2a and in the context of DR4 ((Lang et al., 2002) and personal communication, Lars Fugger (HIU, WIMM, Oxford)), as well as MBP_{85-99} in the context of DR2b. Because of this extreme cross-reactivity, which could explain how viral infection(s) might trigger MS by stimulating MBP-reactive T cells, Hy is an ideal model system in which to assess the contribution of structural mimicry to cross-reactivity. Comparison of the structures of $EBV_{627-641}$-DR2a (Lang et al., 2002) and MBP_{85-99}-DR2b (Smith et al., 1998), complexes with no TCR bound, suggest conservation of TCR contacts, but no Hy complexes exist to date due to poor protein yields of the TCR in standard bacterial expression systems. To address this problem, constructs were designed for transient

expression in mammalian cells; the process of construct design, as well as the results of expression trials, are reported in this Chapter.

The relatively small number of structures of autoreactive TCRs in complex with self and foreign pMHC ligands currently limits our understanding of the structural basis of how molecular mimicry could underlie autoimmune pathologies, including MS. To extend this database, in addition to producing sufficient quantities of Hy TCR for structural studies, I continued work described in (McMahon, 2009) on another TCR isolated from an MS patient. The MS49 D3 (MS49) T cell clone was isolated from a patient with a secondary progressive disease course (Mazza et al., 2002); it is restricted by DQ6 and specific for MBP_{30-44}. Briefly, single-chain constructs were designed for expression in *E. coli*, including the MBP_{30-44} peptide linked to the N-terminus of the TCR β chain by a flexible glycine linker, and C-termini engineered to enhance productive *in vitro* refolding of the soluble TCR αβ heterodimer. Expression and purification strategies were optimised, and optimal conditions determined for complex formation with DQ6, which along with HLA-DM (which catalyses peptide exchange into the DQ6 binding groove) was produced in stably transfected S2 cells. An initial attempt at large scale complex formation confirmed the formation of an MS49-DQ6 complex, but due to the low yield there were insufficient quantities to attempt crystallisation. To continue the work on this project as part of this thesis, further quantities of the complex components were produced, combined under the conditions determined to be optimal for complex formation, and used in crystallisation trials.

To summarise, in the first part of this chapter, I aim to successfully produce recombinant Hy TCR protein, which had previously proved resistant to attempts to express it *via* the conventional route of refolding of the α and β chains from bacterial

inclusion bodies. I will describe the rationale for the design of α and β chain constructs for expression in HEK 293 cells, and report the results of small-scale and, in one case, large-scale expression trials for these constructs. I will then outline the design of further constructs for expression in *E. coli* cells, and discuss the results of attempts to express, refold and purify Hy TCR protein using this method. In the second part of the chapter, I aim to extend previous work (described in (McMahon, 2009)) to form a complex of MS49 TCR with HLA-DQ6 and an MBP_{30-44} peptide. I will outline the production of components required for such a complex, describe the conditions used in complex formation, and present the results of attempts to crystallise the complex to obtain structural information on the interactions of this cross-reactive TCR.

3.2 Production of Hy TCR in mammalian cells

3.2.1 Construct design

Prior to my involvement in this project, initial attempts to produce Hy TCR using standard bacterial systems (as described for MS49 TCR in Section 3.4.1) had proved unsuccessful and had resulted in poor yields of refolded protein. It was decided to address this *via* transient expression in mammalian cells, as used for soluble FLRT constructs (see Section 2.2). Expression constructs designed (by Pia Svendsen (Clinical Institute, Aarhus University Hospital, Denmark)) for bacterial expression, encoding the ectodomains of each chain of the TCR, were therefore adapted for mammalian expression. The original α chain construct extends from the native signal sequence and includes TRAV17*01 and TRAJ6*01 variable and joining segments (identified using the ImMunoGeneTics information alignment software, IMGT/V-QUEST (Giudicelli et al., 2004; Brochet et al., 2008)) and the extracellular portion (up to residue Pro 90, as defined by numbering used by the IMGT information

system, http://imgt.cines.fr/) of the constant region. The original β chain construct also includes the native signal sequence, followed by TRBV29-1*01 variable, TRBJ1-2*01 joining and TRBD1*01 diversity segments, and the extracellular portion (up to residue Asp 129) of the constant region.

A number of alterations were made to these initial constructs to maximise the chance of soluble expression of the αβ heterodimer (Table 3.1). Each construct discussed was cloned into the pHLsec vector (see Appendix A) and made in two versions, with and without the vector-derived C-terminal His tag; this was intended to aid with assessment of small-scale expression trials, and with purification of singly-tagged αβ heterodimers from untagged homodimers. Firstly (constructs α1 and β1), the C-termini of both chains were engineered to carry an extension (ENDGGGCKLE and QDRGGGCD on the α and β chains, respectively) by PCR using extended constant domain specific primers. These extensions are intended to encourage productive association of the α and β chains in the cell by favouring the formation of an interchain disulphide bond. The extensions reposition the C-terminal cysteine residue by three amino acids in the N-terminal direction, and introduce flanking residues which increase flexibility and modify the electrostatic properties to maximise complementarity between the two chains (Stewart-Jones et al., 2003). A shorter extension (RGGGCD) which would reposition the terminal cysteine even further N-terminally was also trialled for the β chain, in construct β2.

In each of the α1 and β1 constructs, a single residue was also substituted relative to the original constructs described above. A point mutation between the joining and constant regions of the alpha chain resulted in substitution of an asparagine residue by a tyrosine. In construct α2, this residue was returned to the native asparagine by site directed mutagenesis. In the constant region of the beta chain, an unpaired

cysteine residue was substituted by a serine. Cysteine residues not involved in either the C-terminal interchain disulphide bridge, or in intrachain disulphide bonds which stabilise the Ig folds of the variable and constant domains, have been observed to inhibit *in vitro* refolding by forming non-native disulphide bonds with other cysteine residues (personal communication, Maria Harkiolaki (Strubi)). Removal of these "unpaired" cysteines can therefore help to ensure correct refolding of recombinant TCRs e.g. MS49 (see Section 3.4.1).

In a second group of constructs, the C-terminus of each construct was returned to its native sequence and extended to the end of the extracellular region, preserving the native position of the unpaired cysteine for interchain disulphide bond formation. In construct α3, the C-terminus was extended to Lys 97 of the constant region (according to numbering used by the IMGT information system), though a slightly altered sequence was used relative to the native sequence; the sequence introduced after Pro 90 was ESCDSVK as opposed to the native ESSCDVK (UniProt accession number A0N8L7). Two candidate "native-like" C-terminal sequences were suggested for the β chain, the first of which (in construct β3) was suggested by Pia Svendsen (Clinical Institute, Aarhus University Hospital, Denmark), and was (following Asp 129) CGEQKLIEEDLDG. The second (used in construct β4) uses the sequence of the β chain constant region recorded in the UniProt database (accession number A0A5B9), which after Asp 129 is CGFTSESYQQG. This latter native extension was also modified in construct β5, to CGFTSNSYQQG, with the intention of introducing a glycosylation site into the C-terminus which might aid in successful secretion of the β chain.

A further set of β chain constructs were created to reconstitute the native unpaired cysteine in the constant region, that had been substituted by a serine residue in

constructs β1-β5 (see above). Removal of this cysteine can improve yields of protein produced by *in vitro* refolding of bacterial inclusion bodies, but as these constructs were intended for soluble expression from mammalian cells, constructs were also designed with the native cysteine in place in case it is required for correct folding and association of the chains in the mammalian system. Accordingly, construct β1 had the native cysteine restored to become construct β6, construct β3 was similarly substituted to create construct β7 and construct β4 was used to create construct β8.

The next set of constructs were designed to encode the variable domain only, as more compact single-domain alternatives to the full-length constructs if these failed to express. The variable and joining regions (TRAV17*01 and TRAJ6*01 respectively) from construct α1, plus a single residue from the constant region and a final C-terminal glycine residue to increase stability, make up construct α4. A further α chain variable domain construct (construct α5) was also designed to reconstitute the native asparagine residue from the N-terminus of the constant region that was substituted by a tyrosine residue in constructs α1 and α3 due to a single point mutation (this approach has also been used in construct α2, see above). Similarly, for the β chain variable domain construct (β9), the variable, diversity and joining regions from construct β1 were used (TRBV29-1*01, TRBD1*01 and TRBJ1-2*01 respectively), with two residues at the C-terminus (glutamate and aspartate) from the N-terminus of the constant region.

The final set of constructs designed for mammalian cell expression use different versions of the N-terminal secretion sequence. All constructs described above use the full native signal sequence, so different variations of this were designed for both chains to use the pHLsec vector-derived secretion leader sequence (see Appendix A), and maximise the chances of identifying constructs that are correctly

targeted to the ER for secretion. Three α chain constructs were designed, all based on construct α1, with different lengths of the native N-terminal sequence removed. Construct α6 lacks the 21 residues of the native signal sequence (as identified by SignalP (Petersen et al., 2011)), so when translated should have the vector-derived signal sequence followed by residue glutamine-22 (numbering includes the native signal sequence, i.e. the initial methionine is methionine-1). It was thought the two glutamine residues at the N-terminus in construct α6 might cause problems with ER targeting and secretion due to the highly negative charge of this region, so in construct α7, these two residues were removed so that the protein sequence begins at glycine-24. Construct α8 is the most truncated of this set of constructs, and begins at proline-28, which corresponds to the most N-terminal structured residue in the crystal structure of the related Ob TCR (Harkiolaki et al., 2009). Three β chain constructs based on construct β1 were designed using a similar rationale; submission of the native N-terminal sequence of the β chain to the SignalP server resulted in two candidate signal peptide cleavage sites, so both of these were tested as potential starting points. In construct β10, 16 residues were removed up to the first candidate cleavage site, so that the vector-derived signal sequence would be followed by residue alanine-17. In construct β11, a further four residues were removed from the N-terminus up to the second candidate cleavage site, so that the first residue after the signal sequence is glutamine-21. Finally, construct β12 was designed to include the pHLsec signal sequence (originally from protein tyrosine phosphate μ) followed by the entire β chain native signal sequence. In contrast to the Ob α chain (see above), the N-terminus of the Ob β chain is structurally resolved up to the very first residue, so no further truncations of the Hy β chain were required. However, construct β1 already encodes the full native secretion signal (without the pHLsec sequence), so this final construct (beginning at valine-2; numbering includes

the native signal sequence) includes both potential signal sequences to determine whether inclusion of the pHLsec sequence affects mammalian expression and secretion of these Hy constructs.

3.2.2 Mammalian expression results

Each of the constructs described in Section 3.2.1 (and illustrated in Table 3.1) was used to transfect HEK 293T cells for small-scale expression trials as described in Section 2.2.1. As these trials are performed in six-well plates, a single plate could be used to test all possible combinations of a pair of α and β chain constructs, expressed both separately and together, as follows.

Well 1: Untransfected cells

Well 2: Hexahistidine-tagged α chain construct alone

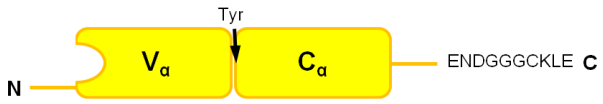
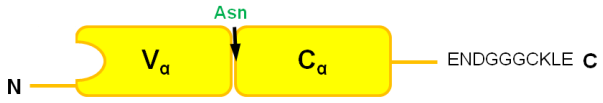
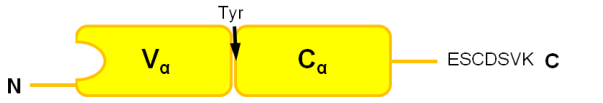
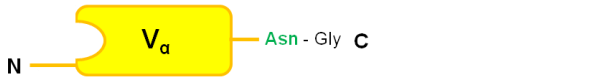
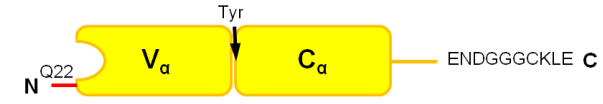
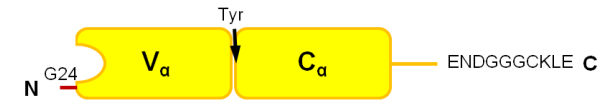
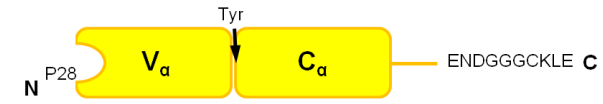
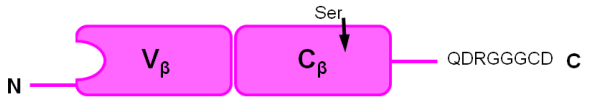
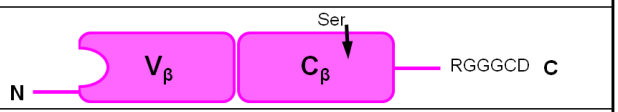
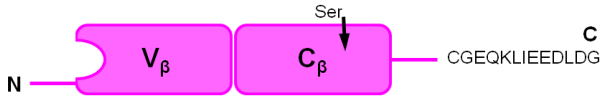
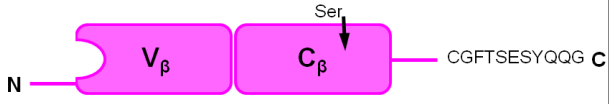
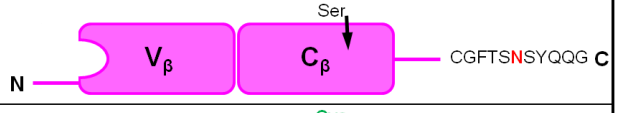
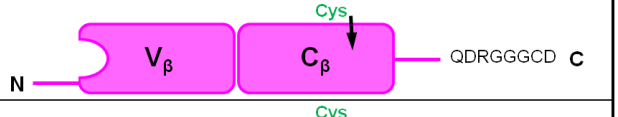
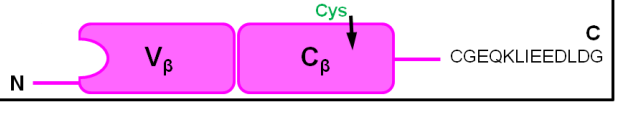
Well 3: Hexahistidine-tagged β chain construct alone

Well 4: Hexahistidine-tagged α chain with hexahistidine-tagged β chain

Well 5: Untagged α chain with hexahistidine-tagged β chain

Well 6: Hexahistidine-tagged α chain with untagged β chain

A number of different pair combinations of α and β chain constructs were trialled - these combinations are summarised in Table 3.2. Where cells were transfected with both α and β chain constructs, in order to test whether both could be successfully expressed and secreted together (ideally as a soluble αβ heterodimer), a 1:1 ratio of plasmid DNA was used with the total amount equalling that used for a single construct transfection.

Construct name	Designed by	Summary of sequence features	Schematic illustration
α1	Maria Harkiolaki (Strubi, Oxford)	Native signal sequence TRAV17*01 variable region TRAJ6*01 joining region Asn to Tyr substitution Constant region to Pro 90 ENDGGGCKLE extension	
α2	The author	As for α1, except: Tyr residue is returned to native Asn	
α3	Maria Harkiolaki (Strubi, Oxford)	As for α1, except: Constant region extended with native-like ESCDSVK sequence after Pro 90	
α4	The author	Native signal sequence TRAV17*01 variable region TRAJ6*01 joining region Asn to Tyr substitution C-terminal Gly	
α5	The author	As for α4, except: Tyr residue is returned to native Asn	
α6	The author	As for α1, except: Native N-terminus truncated to Gln 22	
α7	The author	As for α1, except: Native N-terminus truncated to Gly 24	
α8	The author	As for α1, except: Native N-terminus truncated to Pro 28	
β1	The author	Native signal sequence TRBV29-1*01 variable rgn. TRBD1*01 diversity region TRBJ1-2*01 joining region TRBC2 const rgn to Asp 129 Cys to Ser substitution QDRGGGCD extension	
β2	Maria Harkiolaki (Strubi, Oxford)	As for β1, except: Shorter RGGGCD C-terminal extension	
β3	Pia Svendsen (Aarhus University, Denmark) / Maria Harkiolaki (Strubi, Oxford)	As for β1, except: Constant region extended with suggested native-like CGEQKLIEEDLDG sequence after Asp 129	
β4	Maria Harkiolaki (Strubi, Oxford)	As for β1, except: Constant region extended with native CGFTSESYQQG sequence after Asp 129	
β5	Maria Harkiolaki (Strubi, Oxford)	As for β4, except: C-terminal extension modified to create a glycosylation site (CGFTSNSYQQG)	
β6	The author	As for β1, except: Ser residue is returned to native unpaired Cys	
β7	The author	As for β3, except: Ser residue is returned to native unpaired Cys	

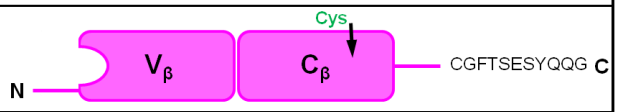
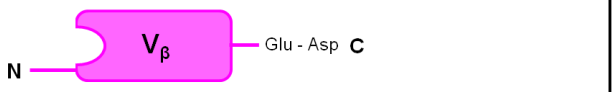
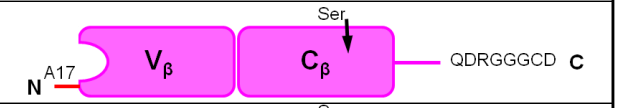
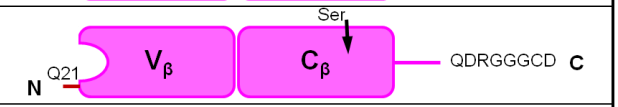
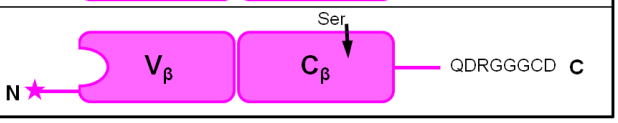
Construct name	Designed by	Summary of sequence features	Schematic illustration
β8	The author	As for β4, except: Ser residue is returned to native unpaired Cys	
β9	The author	Native signal sequence TRBV29-1*01 variable rgn. TRBD1*01 diversity region TRBJ1-2*01 joining region Glu-Asp from constant rgn.	
β10	The author	As for β1, except: Native N-terminus truncated to Ala 17	
β11	The author	As for β1, except: Native N-terminus truncated to Gln 21	
β12	The author	As for β1, except: pHLsec signal sequence included in addition to full native N-terminus	

Table 3.1 Hy α and β chain constructs designed for transient expression in HEK 293 cells. Each construct shown was produced both with and without the C-terminal hexahistidine tag from the pHLsec vector. Some of these constructs were designed prior to my involvement with this project, though all constructs apart from α1, β2 (produced by Maria Harkiolaki (Strubi)) were cloned by me. All constructs were sequence verified.

Following transfection, cells were left to express for three days, after which the cell media and cells were harvested separately so that both soluble, secreted expression (i.e. in the cell media) and expression of protein retained within the cell (i.e. in the cell fraction) could be identified. This created two samples (media and cells) for each well of the six-well plate, which were then subjected to SDS-PAGE, and analysed by Western blotting using a primary antibody against the C-terminal hexahistidine tag present on some of the constructs, and a secondary antibody conjugated to horseradish peroxidase to allow visualisation using an enhanced chemiluminescence reagent. Only the constructs that have a hexahistidine tag will be visualised, but the effects of co-expression of a pair of constructs can be inferred by comparison of samples where either the α or β chain construct, or both, are hexahistidine tagged.

α chain construct	β chain construct	Summary of results of small-scale expression trial
α1	β1	Shown in Figure 3.1 (A) α1 (transfected alone) expressed and secreted β1 (transfected alone) expressed but not secreted When co-transfected, both α1 and β1 expressed, but α1 no longer secreted
α1	β2	As for α1 with β1
α1	β5	As for α1 with β1
α1	β6	As for α1 with β1
α1	β10	As for α1 with β1
α1	β11	As for α1 with β1
α1	β12	As for α1 with β1
α2	β1	As for α1 with β1
α2	β6	As for α1 with β1
α3	β3	As for α1 with β1
α3	β4	As for α1 with β1
α3	β7	Shown in Figure 3.1 (C) α3 (transfected alone) expressed and secreted β7 (transfected alone) expressed and partly secreted When co-transfected, both α3 and β7 expressed, with evidence of some secretion of both constructs
α3	β8	As for α1 with β1
α3	β9	As for α1 with β1
α4	β9	As for α1 with β1
α5	β9	Shown in Figure 3.1 (B) As for α1 with β1
α6	β1	As for α1 with β1
α7	β1	As for α1 with β1
α8	β1	As for α1 with β1

Table 3.2 Combinations of Hy TCR α and β chain constructs used for small scale expression trials in HEK 293T cells, and a summary of the results (see also Figure 3.1). In general (though not exclusively), α and β constructs were paired if designed with a similar rationale. For example, α1, α2, α6, α7 and α8 which have a C-terminus engineered to favour the formation of an interchain disulphide bond were paired with β chain constructs with an engineered C-terminus. Similarly, α chain constructs with a native-like C-terminus were paired with β chain constructs with a native-like C-terminus. The only pairing (α3 with β7) that appears to result in secretion of both constructs when they are co-expressed is highlighted in green.

Representative Western blots from these trials are shown in Figure 3.1 (A and B). The results for the combination of constructs α1 and β1 are shown in Figure 3.1 (A), though results similar to this were replicated for all combinations of full ectodomain constructs (see Table 3.2). In all cases, the α chain construct appears to be expressed, and a significant proportion is secreted into the cell media fraction, as indicated by the green box in Figure 3.1 (A). However, the β chain construct, though expressed, seems not to be secreted at all, and is entirely trapped within the cell. When the α and β chain constructs are co-expressed together, both constructs appear to be expressed (by comparison to results of the single construct expression experiments, where the α and β chains have different molecular weights). Interestingly, however, the α chain construct is no longer secreted into the cell media, as indicated by the red boxes in Figure 3.1 (A). This implies that there is an interaction between the α and β chains within the cell, and that the β chain construct is trapping the α chain so that neither is secreted. The fact that there appears to be some kind of association interaction between the two chains was initially encouraging in terms of the potential for producing an αβ heterodimer using this expression system, but in all cases bar one (see below), it appeared that this could not be exploited because of the lack of soluble expression.

The results for variable domain-only constructs, represented by the Western for the small scale expression trial of constructs α5 and β9 in Figure 3.1 (B), were similar to those described above for full ectodomain constructs. Again, α chain constructs (α4 or α5) were secreted when expressed alone (green box), the β chain construct (β9) was expressed but not secreted, but when co-expressed, the quantity of α chain secreted into the cell media was significantly reduced (red boxes). This pattern was

also repeated for the only combination of a full ectodomain α chain construct (α3) and a variable domain only β chain (β9).

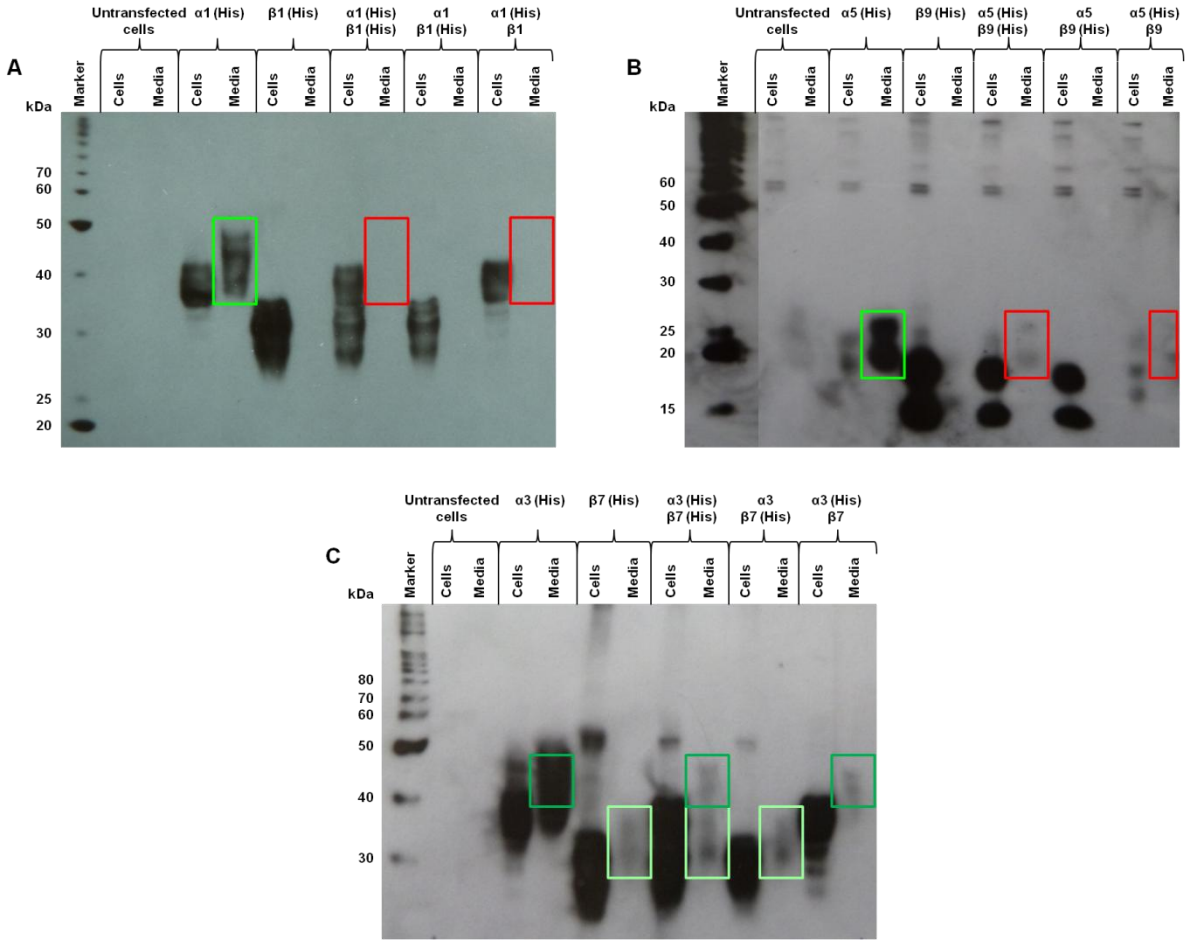


Figure 3.1 Western blots representing the results of the small-scale expression trials in HEK 293T cells. (A) Trial of constructs α1 and β1; α1 alone was expressed and secreted (green box), β1 alone was expressed but not secreted, and appears to trap the α1 construct in the cell when co-expressed (red boxes). This pattern of results was also seen for all other constructs, apart from α3 and β7, illustrated in (C). (B) Trial of constructs α5 and β9, which are variable domain only constructs, but still replicate the pattern seen in (A). (C) Trial of constructs α3 and β7; unlike all other construct pairs, there is evidence that a small amount of the β chain construct is secreted (light green boxes) and that the α chain construct is also secreted when the two constructs are co-expressed (dark green boxes). The smearing of the bands is due to the presence of multiple glycosylation states in non-kifunensine treated HEK 293T cells. Protein samples were prepared under reducing conditions.

However, one pair of constructs did appear more promising. Although the amount of protein secreted into the media appears low compared to overall expression, construct β7, which has a native-like C-terminus and unpaired cysteine, does appear to be partly secreted (bright green boxes in Figure 3.1 (C)). Construct α3, which also includes a native-like C-terminus, appears to be secreted like other α chain constructs when expressed alone (dark green boxes in Figure 3.1 (C)). Uniquely, there appears to be a low level of secretion of both constructs when they are co-expressed, making them a target for further investigation.

Although these initial trials were promising, larger quantities of secreted and correctly folded αβ heterodimer would be required for structural studies. To determine the best HEK cell system in which to express the α3/β7 pairing, α3 and β7 constructs (both hexahistidine tagged) were co-transfected at a 1:1 ratio into HEK 293S cells and HEK 293T cells in the absence and presence of kifunensine as described in Section 2.2.2. For this medium-scale experiment, a single roller bottle (equating to 250 ml harvested media) was used for each expression condition. After harvesting, the filtered media was subjected to batch IMAC with Talon^{TM} resin and samples of the beads, which should bind both constructs due to their hexahistidine tags, and the unbound media were prepared for SDS-PAGE and subsequent Western blot analysis (Figure 3.2).

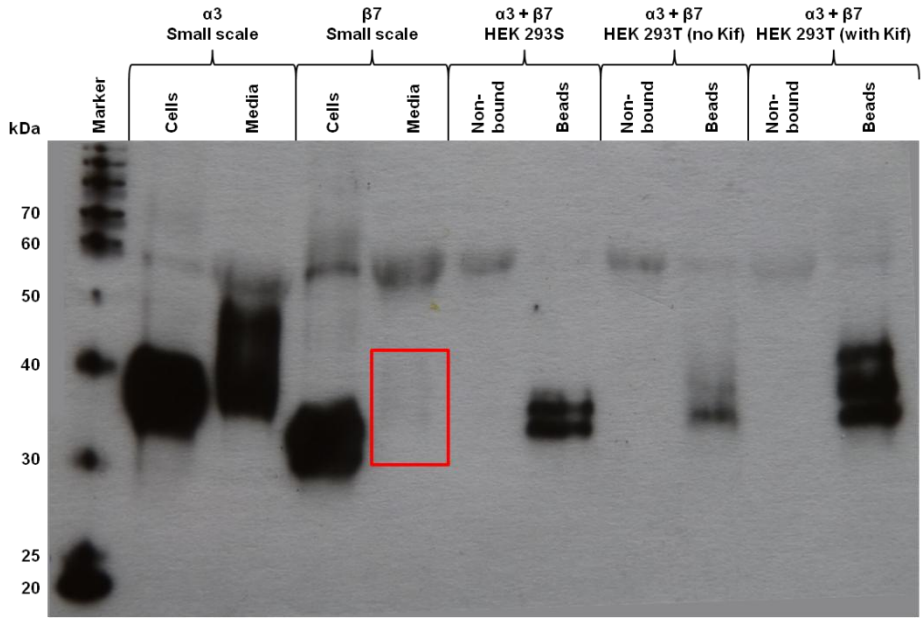


Figure 3.2 Western blot representing results of medium scale co-expression of constructs α3 and β7 in different mammalian cell systems. Small scale expression of each construct (in HEK 293T cells, without kifunensine (Kif)) was repeated for comparison, but unlike previous experiments (Figure 3.1), β7 did not appear to be secreted into the cell media (red box). There was a degree of soluble expression in all cell types tested, with the highest expression/secretion level in HEK 293T cells with kifunensine. However, it is not clear whether both constructs are indeed being secreted, or whether the observed bands represent construct α3 only, and not β7. Protein samples were prepared under reducing conditions.

Unfortunately, the results of small-scale expression trials (which were repeated to aid with identification of bands from the medium-scale trials) were not replicated, and very little to no soluble expression of β7 could be observed (Figure 3.2, red box). This led to concern that the promising results from the small-scale screen (Figure 3.1 (C)) might not be replicated, and that the protein expressed, secreted and bound by Talon^{TM} beads in the medium-scale expression experiments might be composed only of α3, and not include β7.

In the hope that soluble α3 and β7 expression might be replicated at a larger scale, the most promising expression system from the medium scale trials (HEK 293T cells

with kifunensine) was scaled up to produce 3 litres of media. As soluble expression of α3 had been reproducibly good, the ratio of amounts of plasmid DNA encoding the α3 and β7 constructs was altered to 1: 1.2, to increase successful transfection (and expression) of the β7 construct. Following expression, the media was harvested, diafiltrated against 20 mM Tris pH 8.0, 500 mM NaCl using a QuixStand benchtop system (GE Healthcare) and purified by batch IMAC using nickel-charged resin. Despite extensive washing, it proved challenging to separate the relatively small quantities of Hy construct(s) from higher molecular weight contaminants, thought to be BSA originating from FBS added to the cell expression media (see Section 2.2.1). As shown in Figure 3.3, these contaminants co-elute with bands which appear to correspond (by comparison with previous results) to the α3 constructs, though a lower molecular weight band corresponding to β7 cannot be observed. Nonetheless, the fractions indicated in the green box were pooled and concentrated for further purification by gel filtration, in the hope that any β7 would form a complex with the more visible α3 and be co-purified with it, and that this step would remove the higher molecular weight contaminants. However, gel filtration only resulted in a very low absorbance, broad protein peak (maximum absorbance 25 AU), indicating that effective separation had not been achieved. At this stage it was decided not to continue with attempts to form an Hy αβ complex in a mammalian cell system, as the soluble expression of construct β7 that had been observed in small-scale trials did not appear to be reproducible, and both constructs appeared to be produced solubly in such small quantities that effective purification from contaminants proved too challenging.

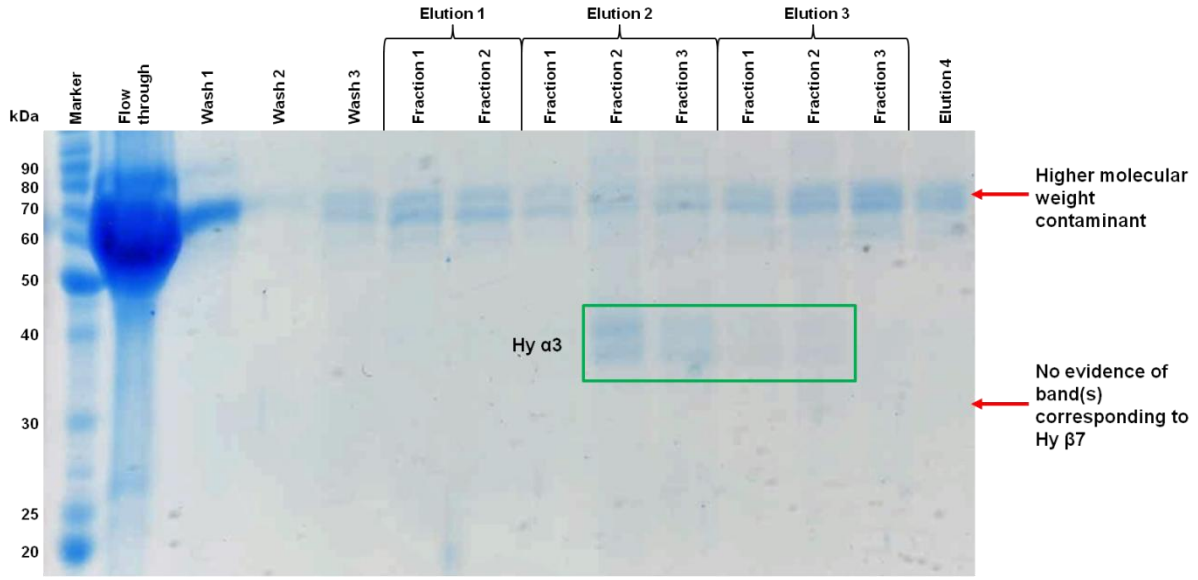


Figure 3.3 12% SDS-PAGE gel representing IMAC purification of media from large-scale expression of Hy constructs α3 and β7 in HEK 293T cells. α3 (~40 kDa) proved difficult to separate from a prevalent ~75 kDa contaminant, thought to be BSA, and β7 (expected to be at ~30 kDa from Figure 3.1) did not appear to be present. Wash 1, 250 ml Tris buffer (20 mM Tris pH 8.0, 500 mM NaCl); Wash 2, 250 ml Tris buffer with 5 mM imidazole; Wash 3, 20 ml Tris buffer with 15 mM imidazole; Elution 1, 2 x 2 ml Tris buffer with 25 mM imidazole (fractions 1 and 2 collected separately); Elution 2, 3 x 2 ml Tris buffer with 100 mM imidazole; Elution 3, 3 x 2 ml Tris buffer with 250 mM imidazole; Elution 4, 2 ml Tris buffer with 500 mM imidazole. Protein samples were prepared under reducing conditions.

3.2.3 Large scale production and characterisation of Hy TCR α chain

Although no further attempts were made to produce soluble Hy TCR β chain constructs, or αβ heterodimers, using a mammalian cell system, small-scale trials had shown that α chain constructs were more suited to this system and could be produced in soluble form in significant quantities (see Figure 3.1). It was decided to assess whether sufficient quantities of a soluble Hy α chain construct could be produced for structural and/or biophysical studies, in case this might shed light on the structure or function of the intact Hy TCR.

Construct α1 was selected due to reproducible soluble expression in multiple small-scale expression trials (see Figure 3.1). Plasmid DNA encoding the construct was transfected into HEK 293T cells for large-scale expression with the addition of kifunensine, as described in Section 2.2.2. After harvesting, media was treated as described in Section 2.2.5, and was purified by batch IMAC either with Talon^{TM} resin (charged with cobalt) or with nickel-charged resin; the latter proved more specific and better at purifying α1 away from higher molecular weight contaminants (a representative nickel-resin purification is illustrated in Figure 3.4 (A)). To purify α1 further after IMAC, appropriate fractions were pooled and concentrated for size exclusion chromatography to remove any residual contaminants; a representative chromatogram and SDS-PAGE gel are shown in Figure 3.4 (B). The problems of low yield of the α chain construct, and difficulty in removing higher molecular weight contaminants encountered when attempting to purify α3/β7 (see Section 3.2.2) were replicated for α1. As shown in Figure 3.4 (B), the proteins eluted from a size exclusion column as a broad peak, making it challenging to resolve α1 from the contaminants; it is possible that a non-specific interaction exists between the proteins. Further purification steps, including anion exchange chromatography on a MonoQ column, and affinity chromatography on a HiTrap^{TM} heparin column (which had been found to bind contaminating BSA protein by co-workers purifying other proteins produced in HEK 293 cell systems) were attempted, but were not successful at achieving full separation of the small quantities of α1 from the contaminants.

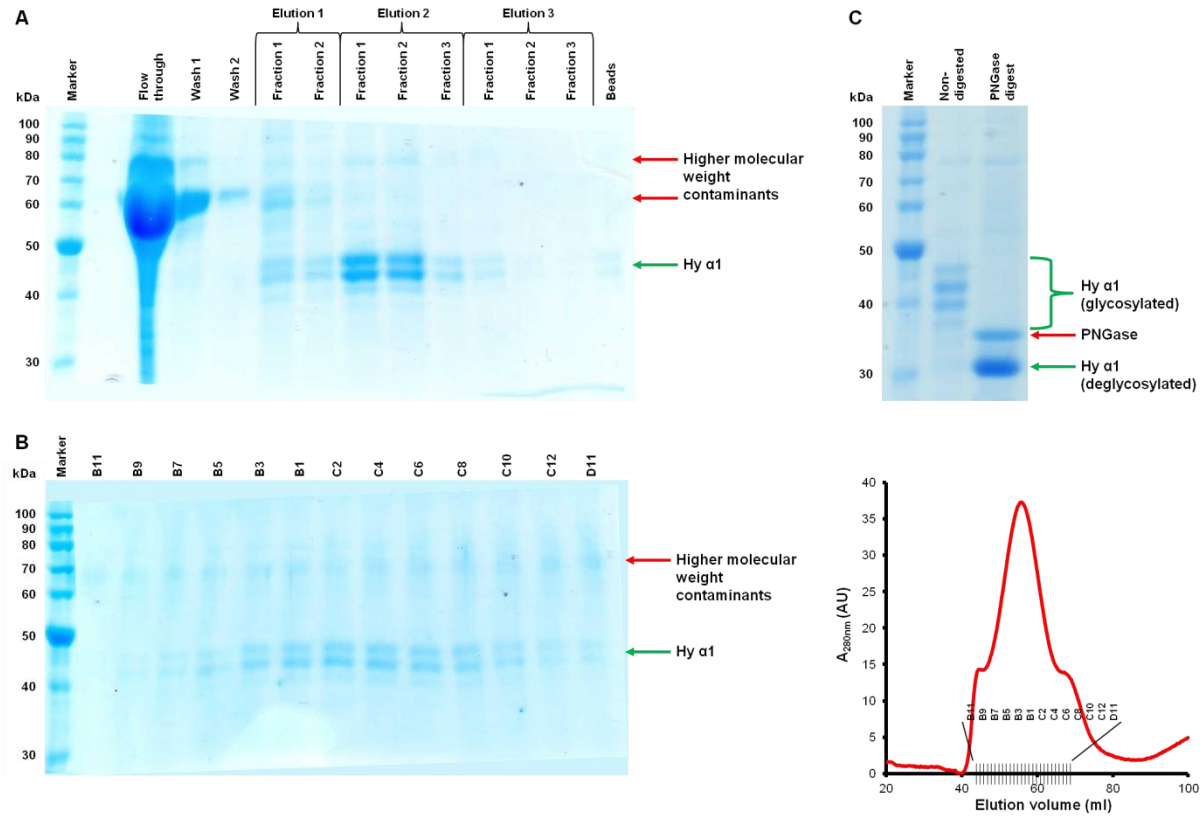


Figure 3.4 Purification of Hy α1 produced in HEK 293T cells. Media was first incubated with nickel-charged resin to purify α1 using its C-terminal hexahistidine tag (A). Fractions containing α1 were then subjected to size exclusion chromatography (B) to remove higher molecular weight contaminants, though this was not entirely successful. To confirm that multiple bands seen on SDS-PAGE correspond to differentially glycosylated forms of α1, purified α1 was digested with PNGase F (C), which resulted in the multiple bands collapsing to a single one at a lower molecular weight, confirming that α1 produced in this system is heavily glycosylated at a number of sites. Protein samples were prepared for SDS-PAGE analysis under reducing conditions.

It was noted that, as for α3 (see Figure 3.3), α1 appears as a doublet band on SDS-PAGE gels, and further bands are visible at higher concentrations (see Figure 3.4 (C)). Given that these bands run at positions corresponding to a higher molecular weight than that predicted from the sequence (24 kDa), and that the sequence contains 4 predicted *N*-glycosylation sites (NetNGlyc 1.0 server (Gupta et al., in preparation)), it was posited that the multiple bands correspond to differentially glycosylated species of α1. To ensure that this was the case, purified α1 was treated

with PNGase F (see Section 2.3.1), which cleaves between the innermost GlcNAc and asparagine residues of *N*-linked oligosaccharides (Tretter et al., 1991) and so removes all glycans added during expression in HEK 293T cells. As shown in Figure 3.4 (C), treatment of α1 with PNGase F did result in collapse of multiple bands to a single band at a lower molecular weight, indicating that the multiple bands correspond to populations of Hy α1 which have different numbers of their glycosylation sites occupied.

Having optimised the purification protocol for α1 over several rounds of production to achieve protein of reasonable purity (though still too low in yield for structural studies), attempts were made to confirm the identity of the protein. Firstly, a sample was submitted for intact mass spectrometry to confirm the identity (and purity) of the deglycosylated protein by comparison to the theoretical mass from the protein sequence. Hy α1 was deglycosylated using PNGase F, and thereafter analysed by Joanne Nettleship (Oxford Protein Production Facility, Research Complex at Harwell) by liquid chromatography-electrospray ionisation mass spectrometry using a quadrupole time-of-flight Micro mass spectrometer. Unfortunately, the sample appeared to be very heterogeneous, and the deconvoluted spectrum contained a large number of peaks, due to the presence of contaminants.

Following this result, an attempt was also made to confirm the presence of α1 in the protein sample using N-terminal (Edman) sequencing, as described in Section 2.3.1. The low-intensity data suggested that the protein may be N-terminally modified and thus not amenable to Edman degradation; this can occur *via* cyclisation if a signal peptide is removed to expose a glutamine residue at the N-terminus, as is the case for α1 (personal communication, Jeffrey Keen). As the protein did not appear to be amenable to N-terminal Edman sequencing, a further aliquot of purified α1 was

prepared for trypsin digest and mass spectrometry, in the hope that the masses of the digested peptide fragments would allow confirmation of the protein identity by comparison to the pattern (or "fingerprint") of predicted trypsin digestion sites in the protein sequence. An excised band of α1 protein from an SDS-PAGE gel was analysed by Jeffrey Keen (Proteomics Facility, University of Leeds) using a MicroMass TOFSpec laser desorption machine. The results revealed five peptides which can be mapped to the α1 sequence with ~30 % coverage, suggesting that α1 was indeed present, though not all predicted peptides were observed.

The results of these attempts at identity confirmation and characterisation seem to confirm the results of SDS-PAGE gels used to follow the purification of Hy α1 i.e. α1 is present, though only in relatively small quantities, and is significantly contaminated with several higher molecular weight contaminants, despite several attempts at purification on multiple different batches of protein. Any protein not used in characterisation was prepared for crystallisation, although the very low yields made it very challenging to prepare sufficient quantities, even for nanodrop crystallisation trials (see Section 2.4.2). In one case, a Fluidigm microfluidic chip was used to reduce the volume of protein required for a sparse matrix screen, though neither this nor standard Greiner plate crystallisation trials produced any crystals.

Given the challenges of producing soluble Hy α1 from HEK 293T cells in sufficient quantities, and of sufficient purity, for characterisation and crystallisation, it was decided prudent not to continue with efforts to make the protein in this system. Instead, the focus of efforts to produce Hy protein was returned to the canonical bacterial inclusion body system.

3.3 Small-scale production of Hy TCR from bacterial inclusion bodies

Although attempts to produce Hy αβ TCR *via* the conventional route of refolding protein from bacterial inclusion bodies had apparently failed prior to my involvement in this project, it was decided to attempt this again to determine whether this might be more successful than mammalian production (see Section 3.2). Constructs were designed using a similar rationale to that used in design of constructs for mammalian expression (see Table 3.1). In particular, the engineered C-termini designed to maximise electrostatic complementarity between α and β chains, and reposition the cysteine residue to favour the formation of an interchain disulphide bond during refolding, were conserved, as similar extensions have been proven to aid successful production of other αβ TCRs in bacterial systems (Stewart-Jones et al., 2003; Harkiolaki et al., 2009). However, the N-terminal signal sequence which is required for correct protein targeting and secretion in mammalian systems is not required in bacterial cells, and may interfere with refolding and (due to its innate flexibility) crystallisation, so should be removed. As discussed in Section 3.2.1, sequence analysis using the SignalP 4.0 server (CBS, Denmark (Petersen et al., 2011)) indicates that the Hy α chain sequence contains only one signal sequence cleavage site, so this was used to design a single α chain construct for bacterial expression, $\alpha_{Bact}1$ (Table 3.3), which is analogous to the mammalian expression construct α6. However, the N-terminus of the Hy β chain contains two possible signal peptide cleavage sites, which were both investigated in constructs $\beta_{Bact}1$ and $\beta_{Bact}2$ (analogous to mammalian constructs β10 and β11 respectively). The unpaired cysteine was mutated to a serine in these constructs to prevent incorrect refolding.

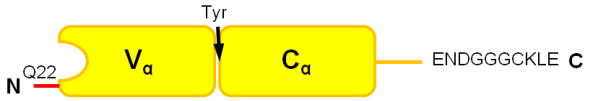
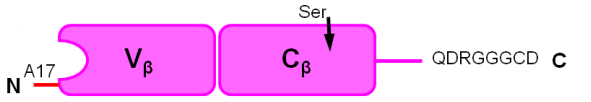
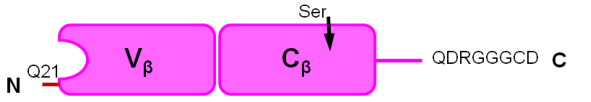
Construct name	Designed by	Summary of sequence features	Schematic illustration
$\alpha_{Bact}1$	The author	No signal sequence Native N-terminus truncated to Gln 22 TRAV17*01 variable region TRAJ6*01 joining region Asn to Tyr substitution Constant region to Pro 90 ENDGGGCKLE extension	
$\beta_{Bact}1$	The author	No signal sequence Native N-terminus truncated to Ala 17 TRBV29-1*01 variable rgn. TRBD1*01 diversity region TRBJ1-2*01 joining region TRBC2 const rgn to Asp 129 Cys to Ser substitution QDRGGGCD extension	
$\beta_{Bact}2$	The author	As for $\beta_{Bact}1$, except: Native N-terminus truncated to Gln 21	

Table 3.3 Hy TCR α and β chain constructs designed for expression in a bacterial cell system and subsequent refolding as an αβ heterodimer. Each construct shown was produced both with and without the C-terminal hexahistidine tag from the pET22b(+) vector. All constructs were sequence verified.

Constructs were cloned into pET22b(+) (see Appendix A) and transformed into competent Rosetta pLysS *E. coli* cells (Novagen), which are designed to enhance expression of eukaryotic proteins, such as the Hy constructs, by supplying tRNAs for codons rarely used in the *E. coli* genome. This strain also produces T7 lysozyme which inhibits transcription by the T7 RNA polymerase to reduce basal levels of expression of the gene of interest, and so allows high-level expression of genes that would otherwise be toxic to the cells before they were grown to a sufficient OD_{600nm} to be induced with IPTG.

To determine the optimum expression conditions for each bacterial Hy construct, small-scale trials were undertaken. Overnight starter cultures of cells expressing each of the $\alpha_{Bact}1$, $\beta_{Bact}1$ and $\beta_{Bact}2$ constructs were inoculated 1:100 into 200 ml fresh media, containing ampicillin to select for cells transfected with the Hy construct and chloramphenicol to conserve the plasmids encoding the rare tRNAs. The cells

were grown at 37 °C until an OD_{600nm} of 0.4-0.6 was reached, when the cultures were induced with 1 mM IPTG and split into two 100 ml cultures for trial expressions at 25 °C and at 37 °C, in case temperature should affect the expression of the Hy constructs. Samples were taken from each culture at four different time points spanning 43 hours from induction so that expression could be monitored over time. Analysis of SDS-PAGE gels and Western blots showed that Hy $\alpha_{Bact}1$ appeared to be expressed better at 25 °C, but that the two β chain constructs were expressed equally well, if not better, at the higher temperature of 37 °C, and that all constructs were expressed well if left over one night, so these were the conditions carried forward for larger scale protein production.

To scale up the expression of the three bacterial Hy constructs, an overnight starter culture of cells transformed with each construct was inoculated 1:100 into 3 litres media, grown until OD_{600nm} was 0.4-0.6 and induced overnight at 37 °C. Inclusion bodies were then isolated and solubilised as detailed in Section 2.2.6, and the amounts of each protein roughly quantified by comparison to known amounts of BSA on an SDS-PAGE gel. From this, it was deduced that there was sufficient $\alpha_{Bact}1$ protein for trial refolds with both β chain constructs, and $\alpha_{Bact}1$ was combined separately with $\beta_{Bact}1$ and $\beta_{Bact}2$ at a 1:1 ratio (25 mg each) and flash diluted into refolding buffer as described in Section 2.2.6.

After being left to refold, the $\alpha_{Bact}1/\beta_{Bact}1$ and $\alpha_{Bact}1/\beta_{Bact}2$ mixtures were dialysed, loaded onto a 5 ml HiTrap™ Q FF anion exchange column and eluted with a 10 CV gradient from 0 to 1 M NaCl, in a similar strategy to that used in the purification of refolded MS49 TCR (section 2.2.6). This step appeared to concentrate the Hy proteins, and Western blots of eluate fractions subjected to SDS-PAGE under non-reducing conditions indicated that dimeric complex(es) had been formed, but

SDS-PAGE analysis revealed that the fractions were still very impure, so that a further purification step would be required to prepare pure αβ complex (Figure 3.5). The presence of multiple bands between 50 and 60 kDa under non-reducing conditions indicates that multiple dimeric complexes, including αα and ββ homodimers as well as the desired αβ heterodimer, have formed, which has also been observed during purification of other refolded TCRs (e.g. MS49 (McMahon, 2009)). At this stage in the preparation of MS49 TCR, the αβ heterodimer is purified away from the homodimers (and other contaminants) *via* a batch IMAC step, where the α chain C-terminal hexahistidine tag is bound so that ββ homodimers remain in the flow though fraction and αα homodimers are eluted from the resin at a different imidazole concentration to αβ. Because all Hy TCR constructs used in this purification trial include a C-terminal hexahistidine tag (so that they could be visualised throughout using Western blots), all dimeric complexes include two tags and this IMAC step would not be effective here, though this would be appropriate in future Hy αβ TCR purifications where only one chain includes a hexahistidine tag, as is the case for MS49.

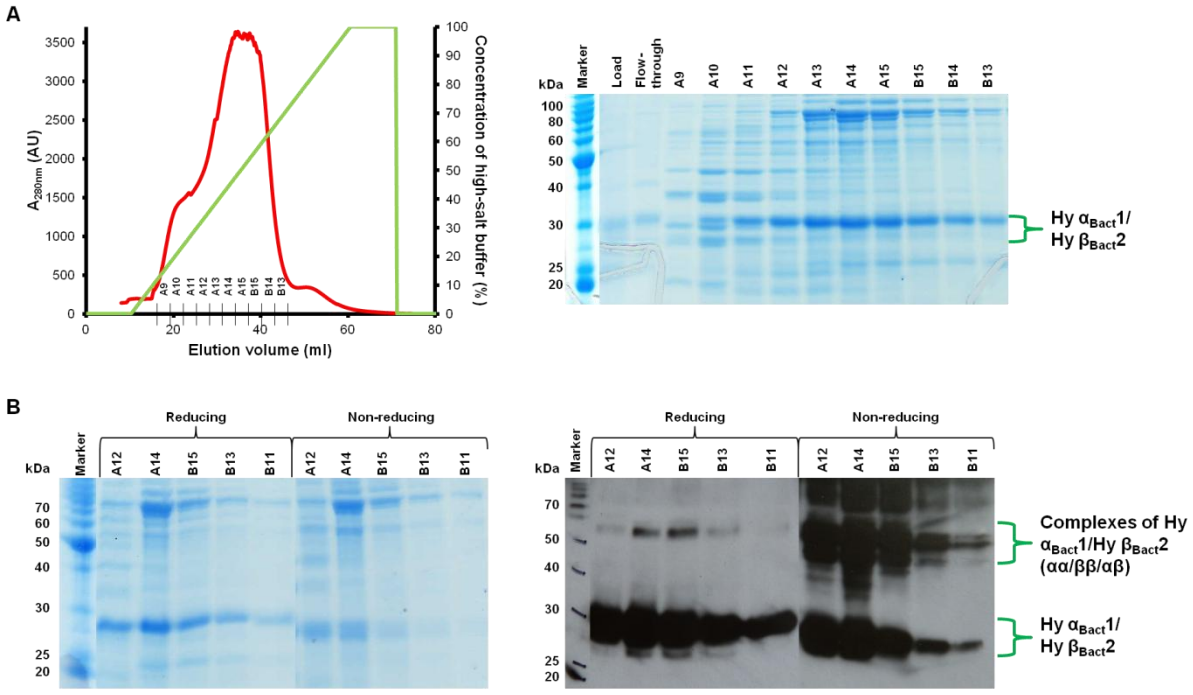


Figure 3.5 Purification trial of attempted refold of Hy $\alpha_{Bact}1$ and $\beta_{Bact}2$. Dialysed refolding mixture was purified by anion exchange chromatography (A) which appears to concentrate the Hy constructs, but leaves the mixture fairly impure, as shown by SDS-PAGE (A, right-hand panel; protein samples were prepared under reducing conditions). Further analysis of fractions from this step under reducing and non-reducing conditions, by SDS-PAGE (B, left-hand panel) and Western blot (B, right-hand panel), indicate that there is more than one population of dimeric complex at around the predicted molecular weight of 60 kDa, which may include αα and ββ homodimers, as well as the target αβ heterodimer. In future purifications, these could be resolved by only using one construct with a C-terminal hexahistidine tag (e.g. the α chain as in MS49), and following anion exchange chromatography with an IMAC step.

These results provide an early indication that, with further optimisation, it may be possible to produce refolded Hy αβ TCR from bacterial inclusion bodies. In preparation for future attempts, further constructs were designed including an alternative affinity (FLAG) tag on the α chain construct to aid with resolution of αβ heterodimers from the αα and ββ homodimers that appear to be formed during refolding (see Figure 3.5). As shown in Table 3.4, the new α chain construct ($\alpha_{Bact}2$) has a C-terminal FLAG peptide tag in place of the hexahistidine tag found in $\alpha_{Bact}1$,

enabling an αβ complex including this chain to be purified by anti-FLAG affinity chromatography as well as with IMAC to make use of the C-terminal hexahistidine tag still present on the β chain constructs. New β chain constructs were also designed based on $\beta_{Bact}1$, which was found to have an N-terminus that was more similar to published structures of other cross-reactive TCRs (Harkiolaki et al., 2009; Sethi et al., 2011) than $\beta_{Bact}2$. Comparison with other structures also revealed an unpaired cysteine residue at the N-terminus of the Hy β chain that is substituted by a serine in the published structures, so this substitution was replicated in the new β chain constructs. Three constructs were designed, each with a different antigenic peptide (plus flexible linker) added to the N-terminus, as seen in MS49 TCR (see Figure 3.6). Firstly, the MBP_{85-99} peptide (ENPVVHFFKNIVTRP) recognised by Hy TCR in the context of DR2b (Wucherpfennig and Strominger, 1995) was added and joined to the β chain N-terminus by the glycine-rich octapeptide linker (GGSGGGGG) used in MS49 as well as published TCR-pMHC structures (Harkiolaki et al., 2009; Sethi et al., 2011) to create $\beta_{Bact}3$. $\beta_{Bact}4$ was similarly created using the $EBV_{627-641}$ peptide (TGGVYHFVKKHVHES) from the viral DNA polymerase recognised by Hy in the context of DR2a and DR4 ((Lang et al., 2002) and personal communication, Lars Fugger (HIU, WIMM, Oxford)). Finally, $\beta_{Bact}5$ was created using an influenza type A virus peptide (YRNLVWFIKKNTRYP) that has also been shown to be recognised by Hy (Wucherpfennig and Strominger, 1995). These constructs are illustrated in Table 3.4; they have yet to be expressed in a bacterial system, but the initial results from attempted refolds and purifications of $\alpha_{Bact}1$ with $\beta_{Bact}1$ and $\beta_{Bact}2$ reported here are encouraging, and it is hoped that the new constructs may be used to optimise production of soluble Hy αβ TCR protein in the future.

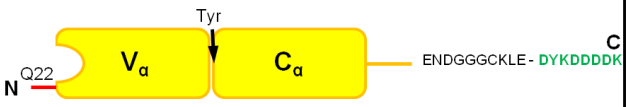
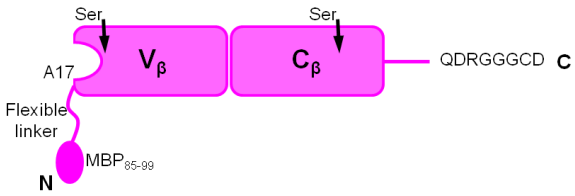
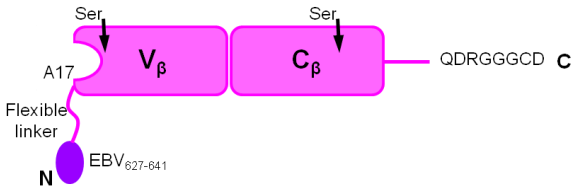
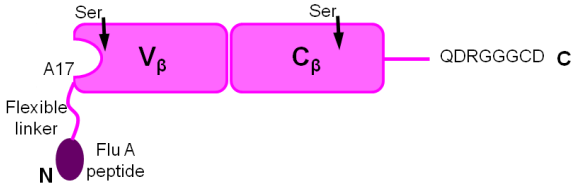
Construct name	Designed by	Summary of sequence features	Schematic illustration
$\alpha_{Bact}2$	The author	No signal sequence Native N-terminus truncated to Gln 22 TRAV17*01 variable region TRAJ6*01 joining region Asn to Tyr substitution Constant region to Pro 90 ENDGGGCKLE extension C-terminal FLAG tag	
$\beta_{Bact}3$	The author	No signal sequence MBP_{85-99} peptide plus flexible linker Native N-terminus truncated to Ala 17 TRBV29-1*01 variable rgn. Cys to Ser substitution TRBD1*01 diversity region TRBJ1-2*01 joining region TRBC2 const rgn to Asp 129 Cys to Ser substitution QDRGGGCD extension	
$\beta_{Bact}4$	The author	As for $\beta_{Bact}3$, except: $EBV_{627-641}$ peptide	
$\beta_{Bact}5$	The author	As for $\beta_{Bact}3$, except: Flu A peptide	

Table 3.4 Further Hy TCR α and β chain constructs designed for expression in a bacterial cell system, following early indications that refolding of $\alpha_{Bact}1$ with $\beta_{Bact}1$ and $\beta_{Bact}2$ was at least partially successful. All constructs were sequence verified.

3.4 Production of components for, and formation of an MS49 TCR/HLA-DQ6 complex

3.4.1 Production of MBP_{30-44}-MS49 TCR

The design of constructs encoding the α and β chains of the MS49 TCR is described in detail in (McMahon, 2009); see also Figure 3.6. Briefly, cDNA was produced from mRNA isolated from MS49 T cells, and sequences corresponding to the variable, (diversity) and joining regions and the ectodomains of the constant regions (excluding the transmembrane segments) were amplified. The C-termini of each chain was engineered to carry a short extension which should encourage formation of an inter-chain disulphide bridge during *in vitro* refolding. The MBP_{30-44} peptide, tethered *via* a glycine-rich flexible linker, was added to the N-terminus of the β chain construct, and an unpaired cysteine residue in the β chain was removed to prevent the formation of non-native disulphide bonds. Finally, the α and β chain constructs were cloned into the pET22b(+) vector (Invitrogen; see Appendix A) for bacterial expression; in the α chain, this was used to add a C-terminal hexahistidine tag to aid with purification of the refolded αβ heterodimer.

The details of expression and purification protocols can be found in Section 2.2 and (McMahon, 2009). Briefly, both chains were found to express well in B834 cells in small scale screens. Cells transformed with either the α chain or MBP-tagged β chain construct were cultured, induced and harvested, and the inclusion bodies containing the protein were isolated by centrifugation of the cell lysate. Inclusion bodies were solubilised and the soluble α and β chains mixed, flash diluted into a refolding buffer and dialysed. Finally, αβ TCR was purified by anion exchange, nickel ion affinity and size exclusion chromatography steps.

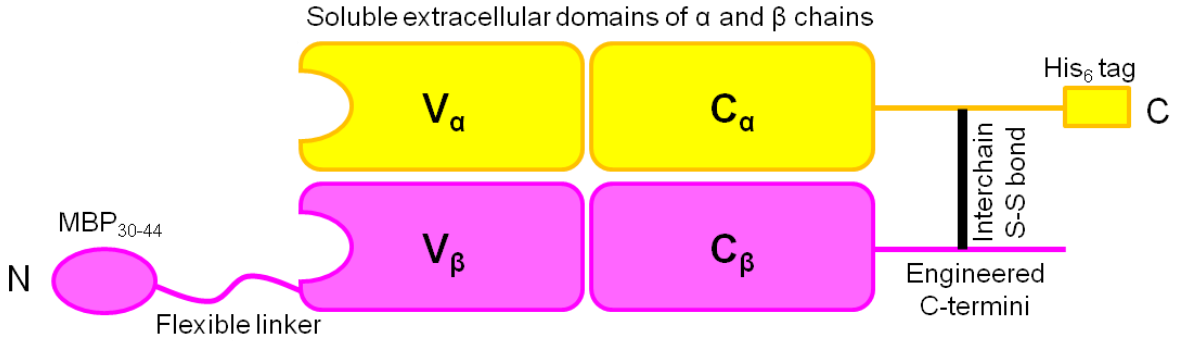


Figure 3.6 Schematic view of the MS49 TCR construct. The MBP_{30-44} peptide is covalently linked to the N-terminus (N) of the β chain by a glycine-rich flexible linker, to promote formation of a stable TCR-pMHC complex. The soluble ectodomains of the α and β chains (including variable ($V_{\alpha/\beta}$) and constant ($C_{\alpha/\beta}$) domains) were engineered at the C-terminus (C) to encourage the chains to associate during *in vitro* refolding, and to form an interchain disulphide (S-S) bond. A hexahistidine (His_6) tag is included at the C-terminus of the α chain to aid with purification. Figure adapted from (McMahon, 2009).

3.4.2 Production of HLA-DQ6 and HLA-DM

Attempts to produce DQ6 in a prokaryotic system, in a manner analogous to that described above for MS49 TCR production, proved unsuccessful (McMahon, 2009). It was hoped that this method would circumvent problems associated with the more established approach of expressing MHC class II molecules in insect cell systems, including large time requirements, low yield and sample heterogeneity due to glycosylation. However, extensive efforts at purification failed to resolve properly refolded DQ6 from a complex mixture of species of unconfirmed oligomerisation status, and DQ6 produced by this method failed to form detectable quantities of a complex with MS49. Production was therefore returned to the more established system of expression by stably transfected *Drosophila melanogaster* S2 cells (Siebold et al., 2004); this system was also used to produce HLA-DM.

For expression in insect cells, ectodomains of the DQ6 α and β chains were fused at their C-termini to the leucine zipper dimerisation domains of the Fos and Jun

transcription factors, respectively (see Figure 3.7 (Kalandadze et al., 1996)). The zippers form a stable interaction, mediated by regularly spaced leucine residues, and assemble in a soluble coiled coil structure, allowing assembly of the DQ6 heterodimer which would not otherwise occur in the absence of the chain transmembrane regions. Expression and purification protocols for DQ6 are described in Section 2.2. To summarise, cells were expanded in suspension and induced with $CuSO_4$, which acts at the metallothionein promoter. Harvested cell media was purified on a monoclonal anti-DQ6 (SPVL3) antibody column and buffer exchanged for storage at -20 °C. Prior to complex formation, the Fos and Jun domains were removed to aid crystallisation (along with the flexible linker joining them to the DQ6 chains, they are an unwelcome source of flexibility), by digestion with endoproteinase Glu-C, which cleaves at the N-terminus of each zipper domain. The digested zippers and any remaining protease (which also digests MS49 TCR) were removed by purification on a MonoQ column.

Soluble DM constructs were designed for expression in the same insect cell system as DQ6, with the transmembrane sections of the α and β chains replaced by FLAG and SV40 large T-antigen (KT3) epitope tags respectively (Figure 3.7 (Sloan et al., 1995)). Protein expression was performed as for DQ6 (and described, along with details of the purification protocol, in Section 2.2). Harvested culture media was affinity purified on a monoclonal anti-FLAG (M2) column and buffer exchanged for storage at -20 °C.

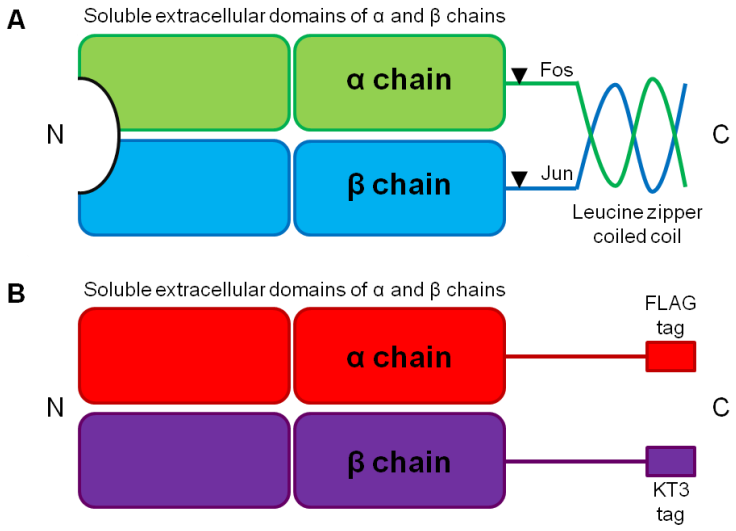


Figure 3.7 Schematic view of the (A) DQ6 and (B) DM constructs. (A) The C-termini of the DQ6 α and β chain constructs were replaced with the Fos and Jun leucine zipper domains, respectively, to promote formation of αβ heterodimers. These were removed prior to complex formation by digestion with endoproteinase Glu-C, which cleaves at an aspartic acid residue immediately N-terminal of each dimerisation domain, marked here by arrowheads. (B) FLAG M2 and KT3 epitope tags were added to the C-termini of the DM α and β chains respectively, to aid purification.

Representative SDS-PAGE gels from affinity purification of DQ6 and DM performed by me are shown in Figure 3.8, as well as a representative profile and SDS-PAGE gel from purification of Glu-C digested DQ6 by anion exchange chromatography. The yields of DM, and especially DQ6, were relatively low, such that several batches were prepared and stored at -20 °C until sufficient quantities had been accrued for complex formation. As larger quantities of DQ6 are required for complex formation (see Section 3.4.3), the yield from affinity purification was maximised by passing the insect cell media over the affinity column several times, and collecting and pooling the eluate each time.

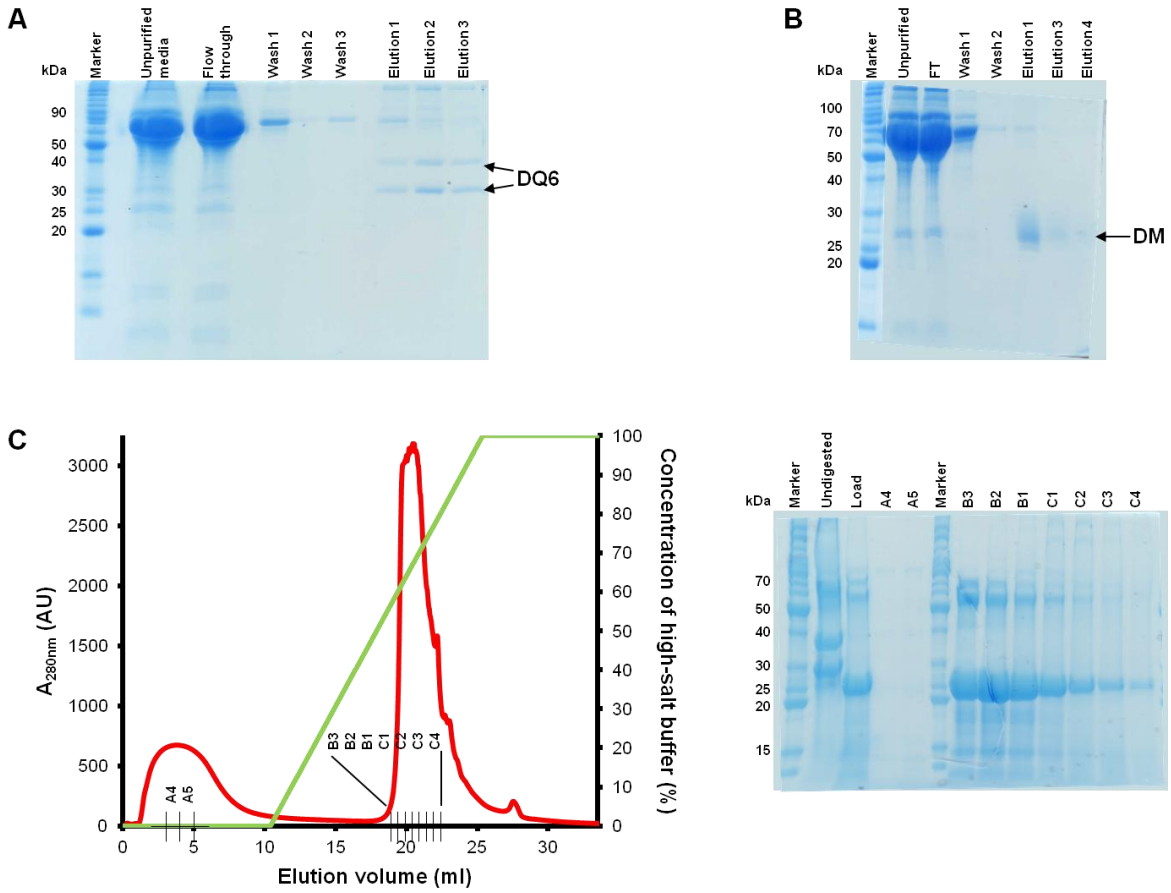


Figure 3.8 Representative purifications of soluble DQ6 and DM constructs for complex formation with MBP-MS49 TCR. (A) 15 % SDS-PAGE gel representing purification of DQ6 from insect cell media by affinity chromatography on a monoclonal antibody column. Bands corresponding to DQ6 are indicated - two bands are present because insect expressed recombinant DQ6 dissociates into its composite chains in the presence of SDS. Wash 1, 10 column volumes (CV) 0.5 % NP-40 in PBS; Wash 2, 10 CV 0.05 % NP-40 in PBS; Wash 3, 10 CV PBS; Elutions 1 and 2, 2 CV diethanolamine incubated with beads for 15 min and eluted into 0.1 CV 2 M Tris-HCl pH 6.3; Elution 3, repeat of Elutions 1 and 2 without incubation. (B) 12 % SDS-PAGE gel representing purification of DM from insect cell media by affinity chromatography on an anti-FLAG antibody column. The band corresponding to DM is indicated. FT, flow through; Washes 1 and 2, 10 CV 10 mM Tris pH 7.4, 150 mM NaCl; Elutions 1, 3 and 4, 2 CV wash buffer containing 100 µM FLAG peptide incubated briefly with beads. (C) MonoQ elution profile and 4-20 % precast Bis-Tris gel representing purification of Glu-C digested DQ6. Captured DQ6 was eluted over 15 CV with a linear gradient of high salt buffer (20 mM Tris pH 7.4, 1 M NaCl). Protein samples were prepared for SDS-PAGE analysis under reducing conditions.

3.4.3 Conditions for complex formation

The results of small-scale screens to identify optimal conditions for formation of an MS49-MBP_{30-44}-DQ6 complex for crystallisation and structural studies are described in (McMahon, 2009). Temperature, pH and the ratio of components in the reaction were identified as the key variables, due to the ability of higher temperatures to promote dynamic movement of the complex components, the pH dependence of DM activity and the dependence of complex formation on the affinity of the TCR-linked peptide for DQ6, respectively. The highest temperature assayed (30 °C) was found to be the most favourable for complex formation, but led to large quantities of protein being lost by degradation and precipitation, so that a temperature of 24 °C was chosen to be an optimum compromise. The lowest pH condition assayed (50 mM MES pH 5.6, 100 mM NaCl) was likewise found to be the optimum compromise between increased DM peptide exchange activity at more acidic pH values (Sloan et al., 1995), and increased protein instability and precipitation close to the pI of MS49 (McMahon, 2009). A peptide-TCR: MHC class II: DM molar ratio of 6: 3: 1 has previously been reported in the literature (Li et al., 2005; Harkiolaki et al., 2009), where an excess of peptide-linked TCR pushes the reaction towards occupancy of the limiting empty class II molecules, and DM is present in smaller quantities because of its catalytic role, rather than being a component of the complex. In the initial attempt at large-scale complex formation described in (McMahon, 2009), a molar ratio of 9: 3: 1 was used, which has an even greater excess of the TCR.

Details of attempts at large-scale complex formation performed by me can be found in Section 2.4.1. To summarise, MBP_{30-44}-TCR, Glu-C digested DQ6 and DM were combined at a molar ratio of 9: 3: 1 (actual masses for purification illustrated in

Figure 3.9 were 13.8: 3.76: 1.22 mg) in a pH 5.6 buffer, in a final volume of 1-2 ml. The reaction was incubated at 24 °C for 12-15 hours, after which time there was very heavy precipitate resulting from protein degradation in the acidic and elevated temperature conditions. After removal of the precipitate by centrifugation, the supernatant was immediately purified by size exclusion chromatography (representative gel filtration profile shown in Figure 3.9). As expected from results for this complex and another (Ob TCR-MBP-DR2b) discussed in (McMahon, 2009), the efficiency of complex formation was quite poor, and the earlier-eluting peak corresponding to the protein complex was significantly smaller than the doublet peak corresponding to non-complexed TCR and DQ6 protein. Fractions from the putative complex peak (the presence of full-length MS49 and DQ6, which should not occur at this elution volume if they were in a non-complexed form, was confirmed by immunoblotting (McMahon, 2009)) were pooled and concentrated to 6-10 mg/ml for crystallisation. Unfortunately, significant amounts of protein were lost at this final stage, and a compromise had to be made between maximising the concentration of complexed protein to favour crystallisation, and maximising the number of different conditions that were able to be screened using the small volumes that resulted. The purification illustrated in Figure 3.9 resulted in 40 µl concentrated protein (concentration estimated to be 8.6 mg/ml based on measurements made at lower concentration) that allowed screening of 240 crystallisation conditions.

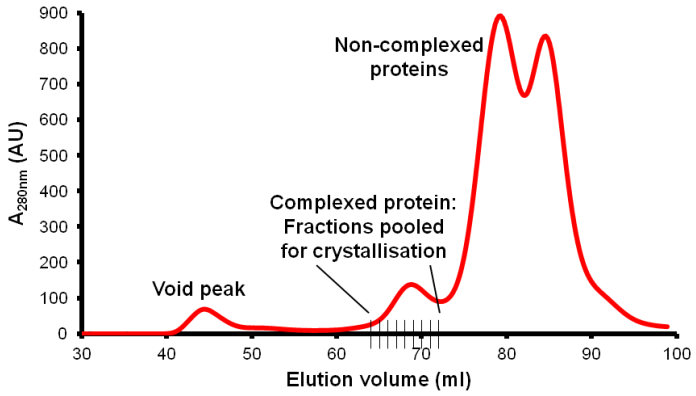


Figure 3.9 Representative purification of MS49-MBP$_{30-44}$-DQ6 complex by size exclusion chromatography.

3.4.4 Crystallisation

A review of the literature was used to identify conditions used to crystallise autoreactive TCR-MHC class II complexes, where the crystal(s) were then used to yield high resolution structural information. It was hoped this might guide selection of conditions to be screened for crystallisation of MS49-MBP$_{30-44}$-DQ6, as there was very limited material available. However, there was no obvious consensus between published structures, and crystallisation conditions varied widely. It was therefore decided to screen as broad a range of conditions as possible to maximise the chance of identifying candidate conditions which showed some evidence of crystal formation, and which could be optimised in future. A number of commercial sparse matrix screens were used for this purpose, including the Emerald Wizard I screen (Emerald BioSystems) which includes the condition in which the Ob TCR-*E. coli* peptide-DR2b was crystallised in the Division of Structural Biology (Harkiolaki et al., 2009).

Extremely fine needle crystals (see Figure 3.10) grew in 20 % (w/v) polyethylene glycol (PEG) 1000, 200 mM NaCl, 100 mM sodium/potassium phosphate pH 6.2, final pH 6.5 (Emerald BioSystems Wizard II condition 14), at 20.5 °C. This was for a protein complex aliquot estimated at 8.6 mg/ml, and the crystals took two weeks to

grow. The crystals were cryoprotected using 25 % glycerol (given the crystals were so delicate, freezing solution was added to the drop for ease of extraction into a loop), and due to their small size were tested at the I24 microfocus beamline at the Diamond Light Source. Grid scans were used to confirm the position of the crystals in the loops where this could not be confirmed visually. Two of the three needles did not diffract at all, and the third only diffracted very weakly, so that a few spots could be seen at low resolution; an attempt was made at data collection from this crystal, but no further diffraction spots were observed.

Although no structural information could be derived from these crystals, the lack of any strong spots at high resolution, and the observation of a few spots at low resolution in one crystal, indicates that the crystals are likely to be protein crystals, rather than salt. This means that the conditions under which these crystals formed can be used as a starting point for future optimisation trials to produce better diffracting crystals.

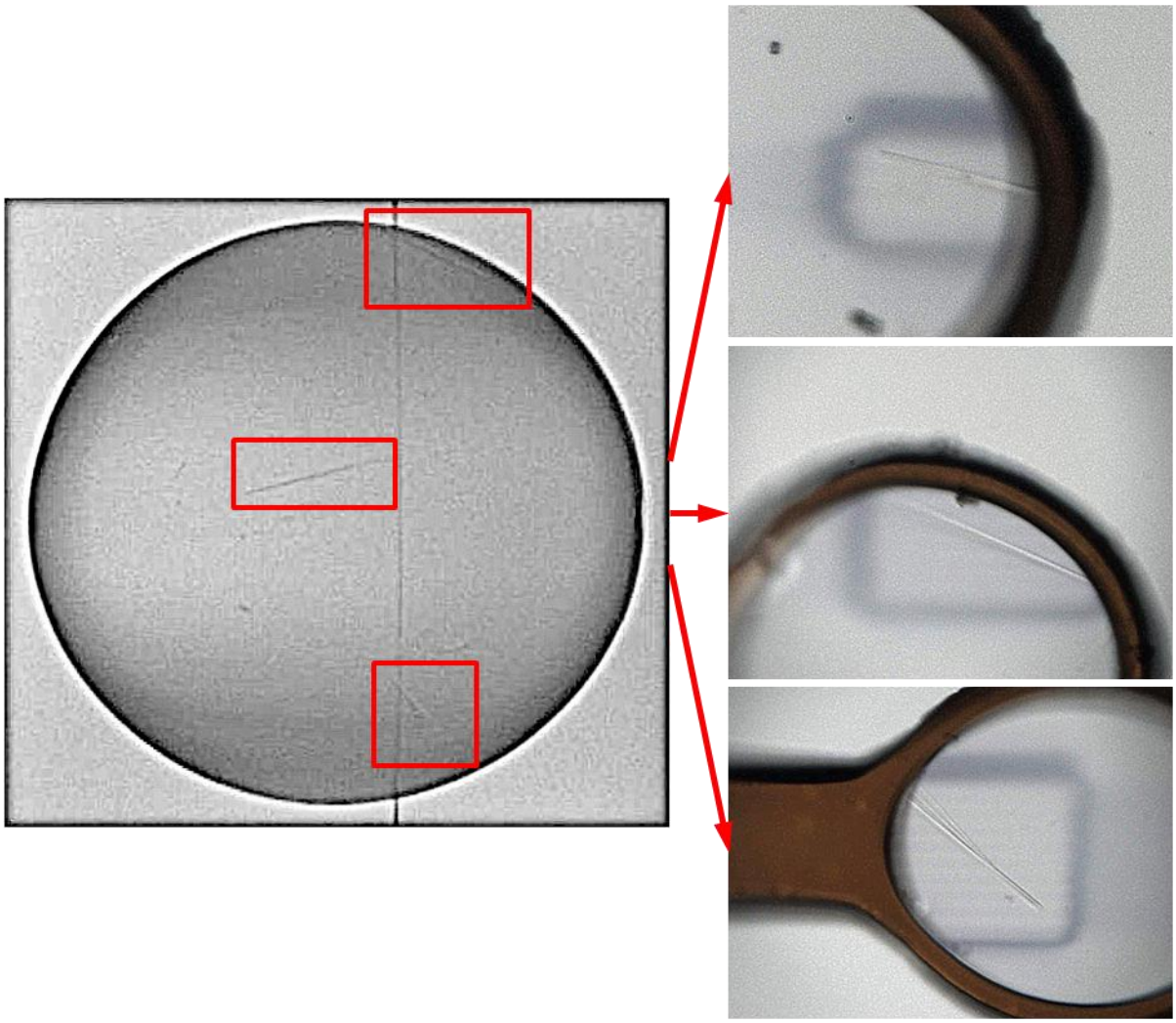


Figure 3.10 Three thin needle-like crystals, putatively of an MS49 TCR-MBP$_{30-44}$-DQ6 complex, formed under the same crystallisation conditions. These crystals diffracted poorly or not at all, but are likely to be protein rather than salt, so that these conditions can be used as the basis for optimisation screens in the future.

3.5 Conclusions and future perspectives

In this chapter, multiple different procedures for producing and purifying soluble recombinant TCR protein have been explored, focussing on two cross-reactive TCRs originally isolated from multiple sclerosis patients. Both TCRs, Hy and MS49, have presented distinct and significant challenges to producing protein in sufficient quantities for structural and/or biophysical studies, which could increase our understanding of the structural basis of molecular mimicry (Lang et al., 2002) and other mechanisms underlying TCR cross-reactivity in autoimmune pathologies.

Previous attempts to produce Hy TCR constructs in, and refold heterodimeric protein from, bacterial inclusion bodies had been unsuccessful, and the attempts to produce protein in an alternative mammalian cell system are documented here. It seems that although several α chain constructs can be produced in a soluble form using this system, it is challenging (and costly) to produce Hy α in sufficient quantities for structural studies of the α chain alone, due to the low yield and difficulty of separating it from higher molecular weight contaminants, despite several attempts at optimisation of the expression and purification protocols. Additionally, none of the Hy β chain constructs described here could reproducibly be expressed and purified in a secreted form, despite numerous efforts to modify the N-terminus (to affect protein targeting to the cell secretory system) and C-terminus (to affect interactions with the Hy α chain). Co-expression of α and β chain constructs within the same population appears to result in retention of the α chain within the cells, as well as the β chain, indicating that there may be an interaction between the α and β chains that could be exploited to favour formation of an αβ heterodimer. However, as neither protein appears to be secreted into the cell media when co-expressed, this could not be exploited, and it was decided prudent to repeat attempts to produce Hy TCR in the

canonical bacterial system, rather than continue with attempts to produce the protein in a soluble form from mammalian cells.

Further constructs, with alterations based on published crystal structures of TCRs produced by refolding of bacterial inclusion bodies, were therefore produced for expression in *E. coli* cells. Preliminary data showing the results of initial attempts at expression, refolding and purification are reported here, and provide an early indication that this may be a more suitable system for the production of soluble αβ heterodimer. However, optimisation of the purification protocol is clearly required to resolve αβ protein from αα and ββ homodimers also produced during the refolding process, as well as other contaminants, and to this end a further α chain construct was designed with a C-terminal FLAG peptide tag (as opposed to the hexahistidine tag found on the β chain constructs). Three further β chain constructs were also designed to include N-terminal antigenic peptides, attached using a flexible linker, which could be used to drive formation of Hy TCR-pMHC complexes during future structural studies. Although these constructs have not yet been produced in the bacterial cell system, they provide a good starting point for the next stage of this project, and it is hoped that they might be used to successfully produce soluble Hy protein from this system in the future.

The other TCR described in this chapter, MS49, has been successfully produced by refolding from bacterial inclusion bodies (as described in (McMahon, 2009)), although previous attempts to form a complex between it and HLA-DQ6 for crystallisation and structural studies have been thwarted by low yield of complex formation, as has also been observed for other autoreactive TCR-pMHC complexes (e.g. Ob-MBP$_{85-99}$-DR2b (Harkiolaki et al., 2009; McMahon, 2009)). This chapter details the protocols for production of the components required for formation of an

MS49-MBP_{30-44}-DQ6 complex, and for complex formation. Although the yields from this process were very low, perhaps owing to low-affinity interactions between the TCR and MBP_{30-44}-DQ6 which have been shown to play a part in the cross-reactivity of other TCRs (Harkiolaki et al., 2009), sufficient complex was produced for sparse matrix crystallisation trials. Very thin needle crystals were produced in one crystallisation condition, and although these did not diffract, it appears that they were protein rather than salt crystals, and therefore provide the basis for future optimisation screens.

The results of the research described in this chapter extend on work done previously on both of these challenging TCRs, which offer the promise of exciting insights into TCR cross-reactivity if high-resolution structural data on their interactions with appropriate pMHC complexes can be obtained. Although this has not yet been achieved, it is hoped that the results discussed here will provide the basis for further work in this area. Further work on production of the Hy TCR could result in the soluble refolded protein from bacterial inclusion bodies, and crystallisation of the MS49-MBP_{30-44}-DQ6 complex could be optimised to produce crystals which provide high-resolution structural data to further extend the database of available structures of cross-reactive TCR complexes that are relevant to human disease.

Chapter 4 Production and Characterisation of Recombinant FLRT Constructs

Chapter 4 Production and Characterisation of Recombinant FLRT Constructs

4.1 Introduction

As discussed in Chapter 1, FLRTs have previously been implicated in cell-cell adhesion and cell sorting, and there is evidence that these behaviours are dependent on the extracellular LRR domain. LRR domains have been shown to mediate functionally important homotypic interactions between proteins involved in cell surface recognition and adhesion events in development.

In this chapter, I aim firstly to investigate the effects of FLRT proteins on cellular morphology and homotypic cell-cell interactions in different cell lines. Although experiments in adherent cell lines have previously indicated that FLRTs have a positive effect on cell adhesion (Haines et al., 2006; Karaulanov et al., 2006), a suspension cell aggregation assay did not find a difference between FLRT3-expressing and control cultures (Robinson et al., 2004). I will attempt to resolve this discrepancy by studying the effects of FLRT cell surface expression on both adherent and suspension cell lines. Secondly, I aim to begin investigating the molecular basis of homotypic adhesion behaviour at the cellular level by assessing the oligomerisation state of soluble ectodomain and LRR domain constructs. Homotypic cell sorting of FLRT-transfected cells is LRR domain-dependent (Karaulanov et al., 2006), and LRR-mediated dimerisation has been shown to be functionally essential in other proteins (Scott et al., 2006; Kajander et al., 2011), so a central question is whether FLRT LRR domains dimerise in solution, and whether any association behaviour is mirrored by the ectodomain constructs.

4.2 FLRT-mediated cell-cell adhesion

4.2.1 FLRTs are expressed on the cell surface

FLRT3 has previously been shown, by two independent studies (Robinson et al., 2004; Tsuji et al., 2004), to be localised to the plasma membrane in CHO cells and cultured E15 rat cortical neurons. As different cell types are used in the experiments described in this thesis, to establish the effects of cell-surface FLRT expression on cellular morphology and interaction behaviour, it was necessary to confirm that all three FLRTs could be successfully expressed on the surface of these cells. Transmembrane FLRT constructs, containing the ectodomain, transmembrane region and a truncated cytoplasmic domain ($FLRT_{LONG}$; Figure 4.1) were transfected into HEK 293T and Cos-7 cells, and their expression on the cell surface assessed using FACS and immunocytochemical staining.

$FLRT_{LONG}$ constructs were produced for FLRT1, FLRT2 and FLRT3 and fused to either an N-terminal HA tag, or an N-terminal FLAG tag as described in Section 2.1 (HA- or FLAG-$FLRT_{LONG}$; Figure 4.1). The proteins were expressed in HEK 293T cells cultured and transfected as described in Section 2.2, then prepared for FACS analysis as described in Section 2.7. For unknown reasons, untransfected cells produced bimodal data, but cells transfected with HA-$FLRT_{LONG}$ or FLAG-$FLRT_{LONG}$ constructs exhibited significantly higher fluorescence than both peaks in the untransfected cell data (Figure 4.1), indicating that the HA and FLAG tags are located on the cell surface. The FACS data also indicate that the level of surface expression is similar for all three FLRTs, across each of the N-terminal tag types.

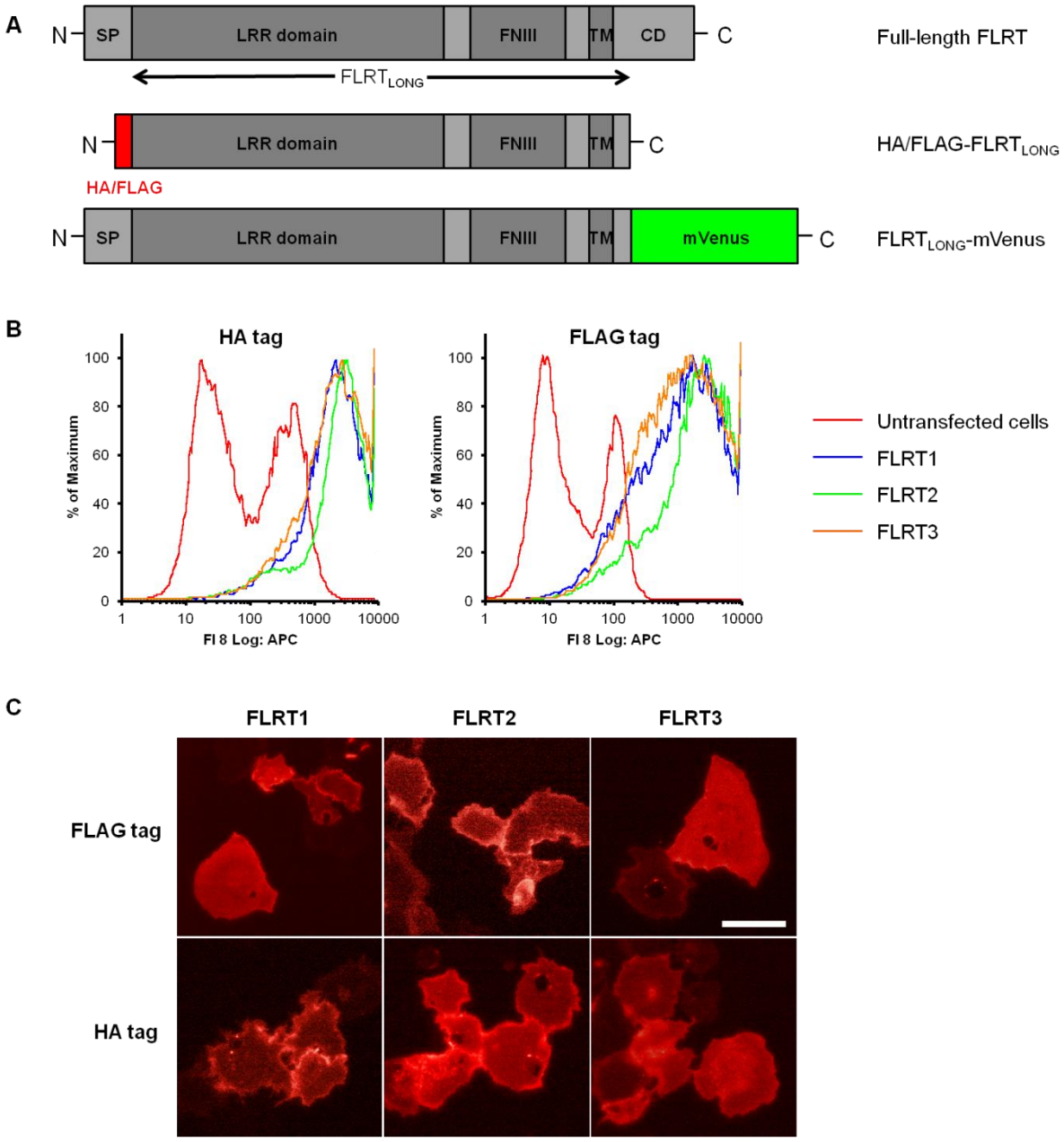


Figure 4.1 FLRTs are expressed at the cell surface. (A) Schematic representation of the structure of the HA-, FLAG- and mVenus-tagged FLRT$_{LONG}$ constructs used in this thesis; residue numbers for each construct are listed in Appendix B. (B) FACS data showing that HA- and FLAG-tagged constructs are expressed on the surface of HEK 293T cells, and that levels of surface expression are consistent across FLRTs 1, 2 and 3 for each N-terminal tag. This data was generated with the assistance of Veronica Chang (WIMM, Oxford). (C) HA- and FLAG-tagged FLRT$_{LONG}$ constructs were successfully immunostained, under non-permeabilising conditions, on the surface of Cos-7 cells; scale bar represents 50 µm. SP, signal peptide; LRR, leucine-rich repeat; FNIII, fibronectin domain type III; TM, transmembrane domain; CD, cytoplasmic domain; HA, haemagglutinin peptide tag; FLAG, FLAG M2 peptide tag; mVenus, monoVenus fluorescent tag.

For further confirmation of cell surface expression, HA- and FLAG-$FLRT_{LONG}$ constructs (Figure 4.1) were expressed in Cos-7 cells cultured as described in Section 2.7, transfected as described in Section 2.2, then prepared for immunocytochemistry analysis as described in Section 2.7. Staining for HA- and FLAG-tagged $FLRT_{LONG}$ constructs in the transfected Cos-7 cells, with anti-HA and anti-FLAG antibodies respectively, was successful for all three FLRTs (Figure 4.1). As the staining was performed under non-permeabilising conditions, it is possible to conclude that the tagged amino terminus of the FLRTs accumulates on the cell surface. Taken together, the immunostaining and FACS data indicate that mouse FLRT1, FLRT2 and FLRT3 constructs are indeed expressed on the surface of HEK 293T and Cos-7 cells, which are used for experiments elsewhere in this thesis, and that surface expression levels are consistent across the three FLRTs.

4.2.2 FLRTs accumulate at interfaces between HEK 293 cells

Adherent Cos-7 cells transfected with FLRT1, FLRT2 and FLRT3 show accumulation of protein at focal adhesion junctions between transfected cells, correlating with a decrease in protein levels across the rest of the cell membrane (Haines et al., 2006). To establish whether FLRT molecules are localised to sites of cell-cell adhesion in multiple cell types, an assay was employed in HEK 293 cells, which have been suggested as a sensitive system for studying cell sorting behaviour (Karaulanov et al., 2006). $FLRT_{LONG}$ constructs (Figure 4.1) were produced for mouse FLRT1, FLRT2 and FLRT3 and fused to a C-terminal monoVenus tag as described in Section 2.1 ($FLRT_{LONG}$-mVenus; Figure 4.1). These proteins were expressed in adherent HEK 293 cells, and imaged, as described in Section 2.6.1 (Figure 4.2). Cells transfected with FLRT1, FLRT2 and FLRT3 constructs demonstrate distinctive morphology within the confluent cell monolayer, defined by extended planarity of the

cell membrane between neighbouring cells. Fluorescence is evident around the whole cell surface, but is strongly concentrated at the cell-cell interface for all FLRTs (white arrowheads, Figure 4.2). In fact, quantitative analysis of levels of fluorescence in regions of the images corresponding to cell-cell interfaces and parts of the cell surface not engaged in a cell-cell interaction reveals that levels of fluorescence are significantly higher in cell-cell interfaces than at non-interfacing regions for all three FLRTs (Figure 4.2). Interfaces of this type were not observed in cells transfected with an mVenus-tagged CD45 construct, which demonstrated consistent fluorescence across the cell surface with no significant difference in mean fluorescence values between cell-cell interface and non-interface regions (negative control; red arrowheads, Figure 4.2). This suggests that FLRTs are capable of homotypic recognition and interaction in *trans*, between opposing cells.

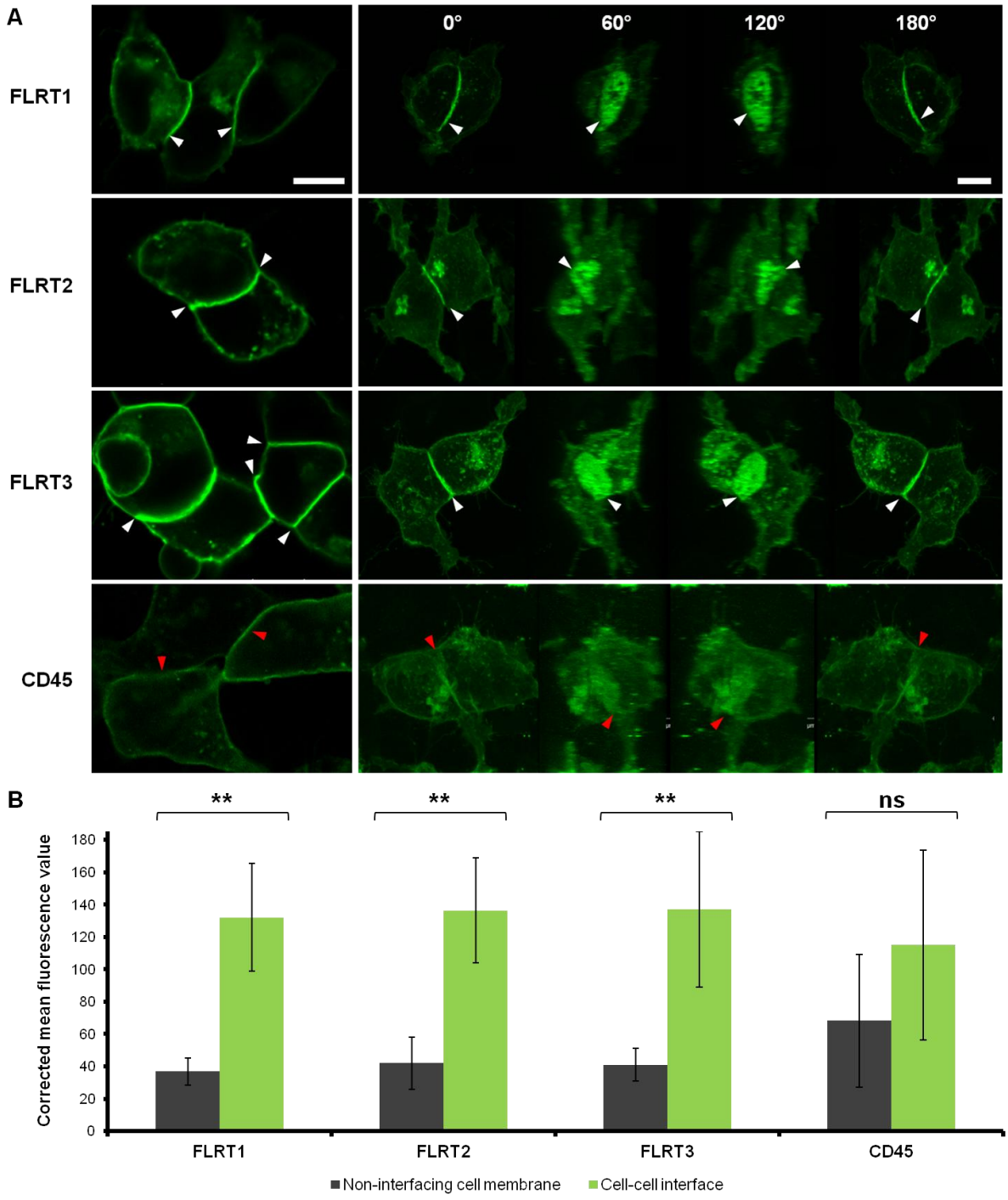


Figure 4.2 FLRT1, FLRT2 and FLRT3 expression is localised to cell-cell interfaces in HEK 293 cells. (A) The images in the left column are representative static images, while the images on the right are taken from a 180° rotation of a representative projection of the cells, so that the interface(s) can be viewed from multiple angles; scale bars represent 10 µm. Note the more intense fluorescence of the cell membrane at cell-cell interfaces than over the rest of the membrane in the case of FLRT-transfected cells, highlighted by white arrowheads. In the case of cells transfected with CD45, fluorescence is distributed evenly over the whole cell surface, and not concentrated to interfaces between cells, as highlighted by red

arrowheads. Images were taken by Jeff McIlhinney (MRC Anatomical Neuropharmacology Unit, Oxford). (B) These differences in fluorescence distribution are also indicated by quantitative analysis of areas of the images corresponding to cell-cell interface and non-interface regions of the cell membrane. These regions have significantly different levels of fluorescence for cells expressing FLRTs, but not those expressing CD45. Mean values are from 10 images for each construct; error bars indicate standard deviation. **, $P < 0.001$; ns, not significant at the $P < 0.05$ level (Student's *t* test).

4.2.3 FLRT2 and FLRT3 mediate homotypic cell-cell adhesion

FLRT2 and FLRT3 have previously been shown to promote homotypic cell sorting in HEK 293T cells, both when seeded on agarose and when in suspension (Karaulanov et al., 2006; Tossell et al., 2011). However, it was suggested that this effect might be cell specific, as while HEK 293 T and mouse P19 cells show FLRT-mediated sorting, CHO, L cells, HeLa, Cos1 and RKO cells do not (Karaulanov et al., 2006). To clarify whether the accumulation of FLRTs at cell-cell interfaces described in Section 4.2.2 can give rise to an adhesive effect sufficient to clump cells in suspension, $FLRT_{LONG}$-mVenus constructs (Figure 4.1) were transfected into non-adherent Sf9 insect cells by means of a baculovirus vector, as described in Section 2.6. After incubation, the cells were imaged as described in Section 2.6, and a clear clustering effect could be seen for cells expressing $FLRT2_{LONG}$-mVenus and $FLRT3_{LONG}$-mVenus (Figure 4.3). This suggests that FLRT2 and FLRT3 are capable of interacting in *trans* to induce homotypic cell adhesion in Sf9 cells, as has been previously observed in HEK 293T and P19 cells.

A $FLRT1_{LONG}$-mVenus construct was similarly tested, but no aggregation was observed between Sf9 cells expressing this construct (Figure 4.3). This discrepancy between FLRT1 and FLRT2/FLRT3 constructs, in terms of their ability to cause cell clumping in this system, is most likely to be due to differences in the affinity of the *trans* interaction between FLRT-expressing cells. Like FLRT2 and FLRT3, FLRT1

accumulates at interfaces between adherent cells (see Section 4.2.2), but we can speculate that the adhesion effect concomitant with such accumulation is not as strong for FLRT1 as for FLRT2 and FLRT3. Specifically, it appears not to be strong enough to clump suspension Sf9 cells, which are less restricted in their movement than adherent HEK 293 cells and may therefore require a stronger *trans* interaction to display an adhesive effect.

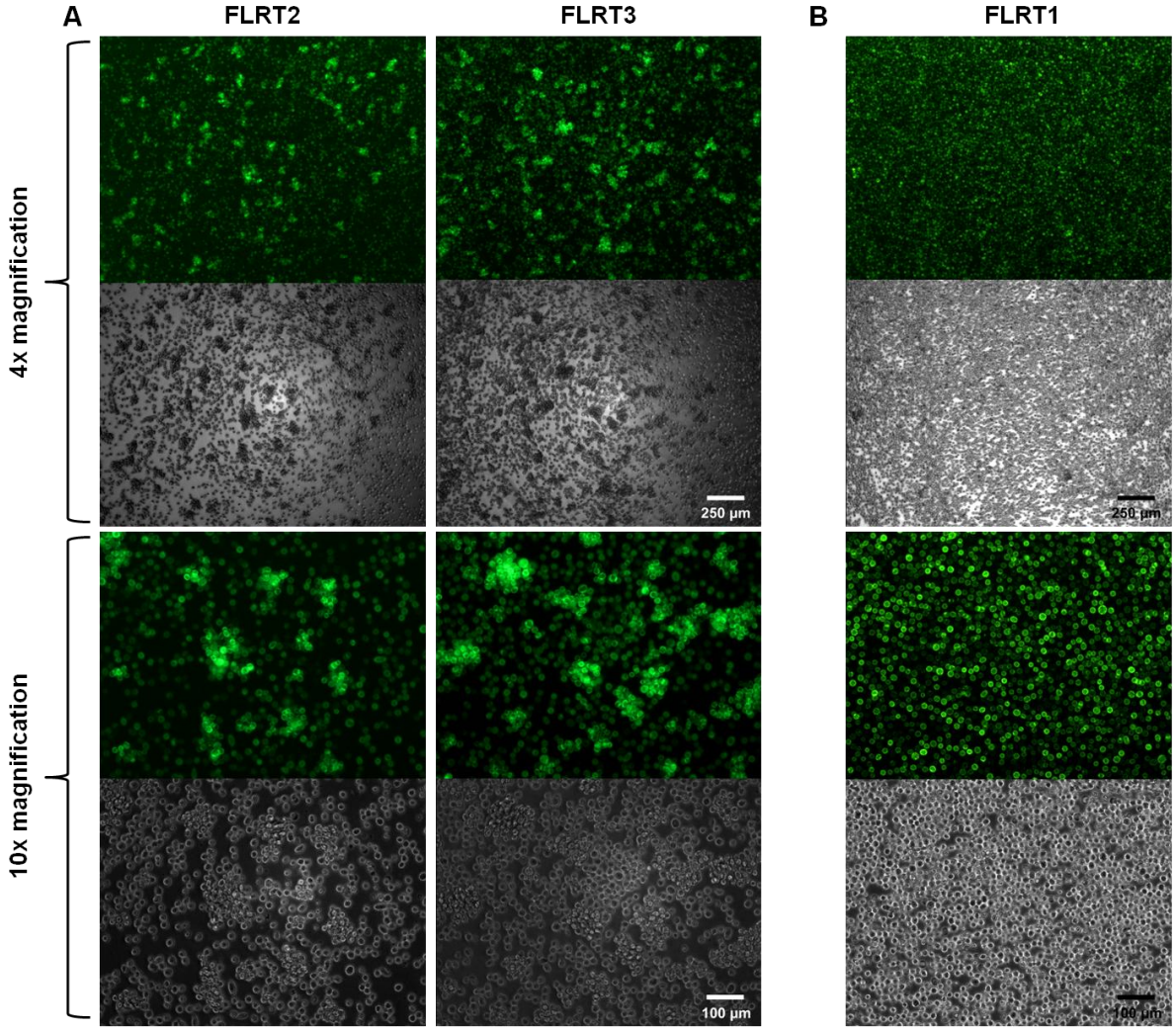


Figure 4.3 Effects of FLRTs on cell-cell adhesion. (A) FLRT2 and FLRT3 constructs can interact homotypically in *trans* to promote clustering of Sf9 cells; the resultant clumping is shown here at 4x (top) and 10x (bottom) magnification. (B) Cells expressing a FLRT1 construct form a monodisperse suspension, and no clustering is observed.

These observations appear to align with data from knock-out mice, which show a variety of developmental defects associated with failures in cell-cell adhesion when FLRT2 or FLRT3 is knocked out (see Section 1.4 and (Maretto et al., 2008; Müller et al., 2011)), but no obvious phenotype on knockout of FLRT1 (personal communication, Liz Robertson and Liz Bikoff (both Sir William Dunn School of Pathology, University of Oxford)). It may be that FLRT2 and FLRT3 mediate stronger cell-cell interactions, with the result that the knockout phenotype is more severe, while interactions between cells expressing FLRT1 are weaker and/or more transient, such that the knockout mouse does not display any significant developmental defects. However, it is possible that many other factors, including the timing and tissue distribution of expression of the different FLRTs, contribute to the phenotypes of the corresponding knockout mice, and any link between results seen in Sf9 cells and the severity of the knockout phenotype is speculative at this stage.

4.3 Expression and purification of FLRT ectodomains and LRR domains

Two sets of constructs were designed to probe the ability of parts of the FLRT proteins to self-associate in solution. The first set encodes the full ectodomain of mouse FLRT1, FLRT2 and FLRT3 ($FLRT_{ECD}$; see Figure 4.4), including the native signal sequence, LRR domain, FNIII domain and the remainder of the ectodomain up to 3-5 residues outside the predicted transmembrane region; residue numbers for each construct are listed in Appendix B. The second set encodes solely the extracellular LRR domain ($FLRT_{LRR}$; see Figure 4.4), including its N- and C-terminal cysteine-rich caps. The constructs were designed using the domain boundaries listed in Appendix B, and the cloning strategy used, including the pHLsec expression vector, PCR primers and restriction enzymes, is described in detail in Section 2.1.

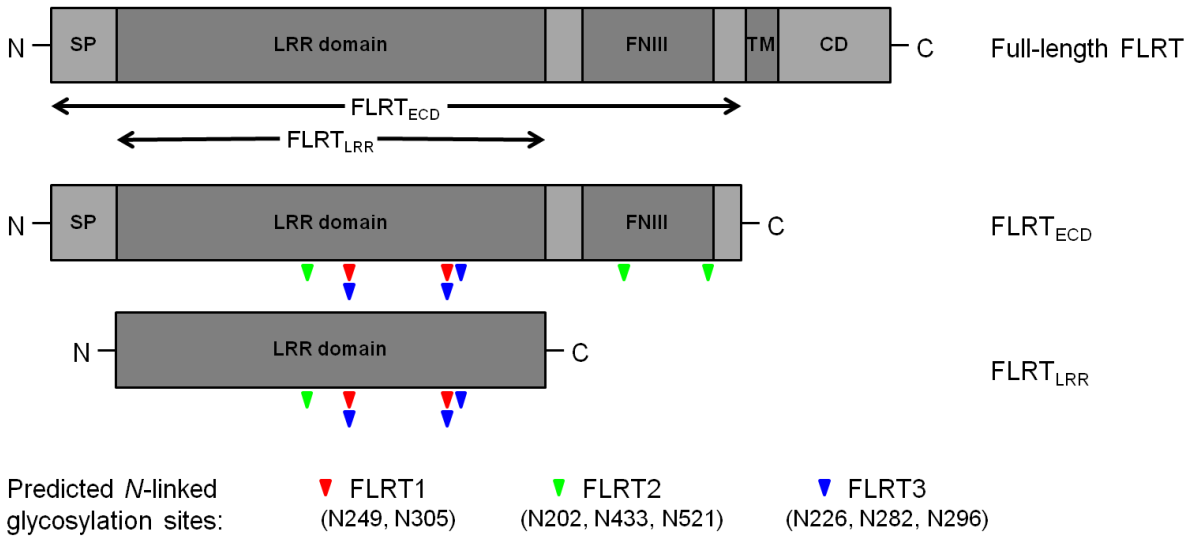


Figure 4.4 Schematic representation of the structure of the mouse FLRT$_{ECD}$ and FLRT$_{LRR}$ constructs used in this thesis. The positions of predicted *N*-linked glycosylation sites (according to the NetNGlyc 1.0 server (Gupta et al., in preparation)) are marked. Residue numbering corresponds to GenBank accession numbers NP_958813 (FLRT1), NP_958926 (FLRT2) and NP_001165631 (FLRT3); residue numbers for the boundaries of each construct are listed in Appendix B.

Each construct contains at least one predicted *N*-linked glycosylation site (using the NetNGlyc 1.0 Server, CBS, Denmark (Gupta et al., in preparation); see Figure 4.4), and large scale expression was initially performed in GnTI-deficient HEK 293S cells (see also Section 2.2.2); protein produced in this way was used for FLRT$_{LRR}$ crystallisation and AUC experiments. However, as the yields of most of the constructs were not sufficiently high to generate high concentration samples for MALS using this system, expression for this purpose was performed in HEK 293T cells, which typically produce higher yields of secreted protein than HEK 293S cells. In most cases where HEK 293T cells were used (and specifically for expression of FLRT$_{LRR}$ constructs for MALS), the *N*-glycosylation inhibitor kifunensine was added immediately post-transfection as described in Section 2.2.2 to produce a homogenous population of *N*-linked glycans. When expressing FLRT$_{ECD}$ constructs

for MALS, addition of kifunensine was erroneously omitted, but this has been taken into account in the analysis of the results (see Section 4.4.1).

A two-stage purification procedure, described in Section 2.2.5, comprising immobilised metal affinity and size exclusion chromatography steps, was used to purify the $FLRT_{ECD}$ and $FLRT_{LRR}$ proteins. Representative gel filtration profiles and SDS-PAGE gels from these purifications are shown in Figure 4.5.

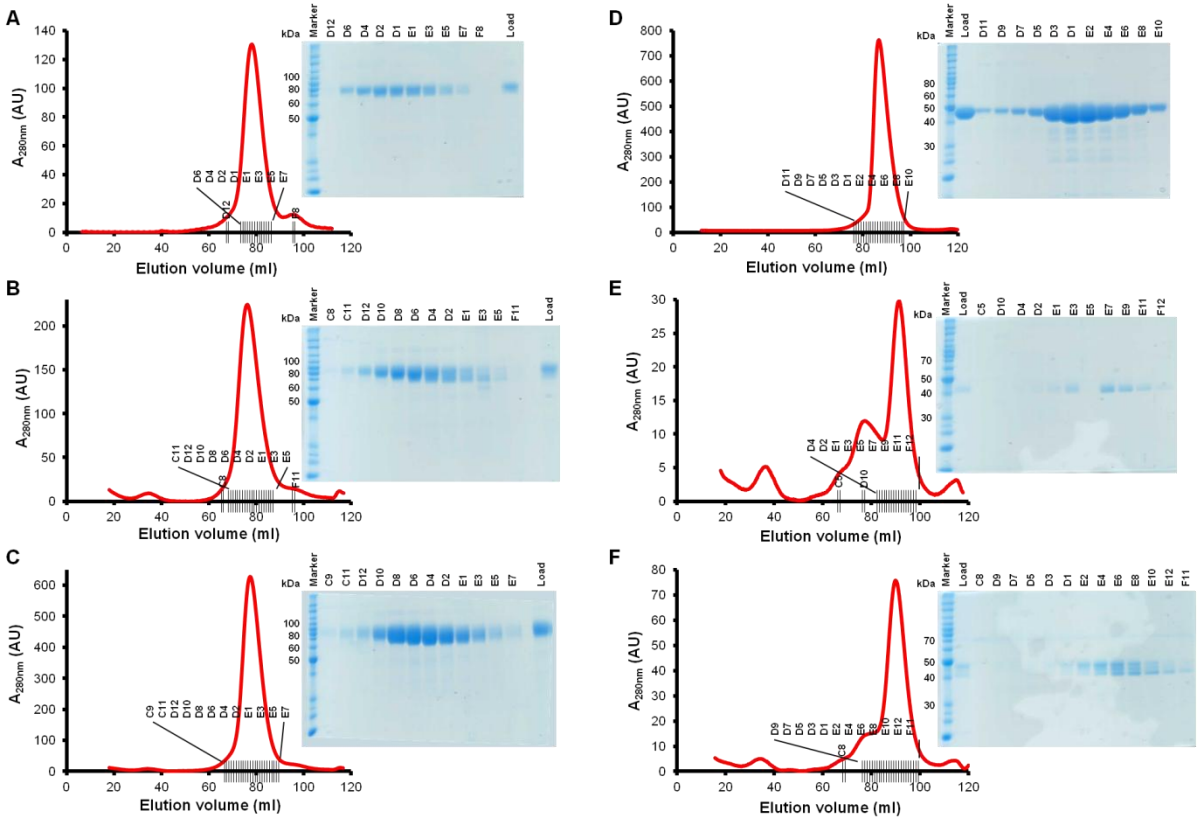


Figure 4.5 Representative purifications of $FLRT_{ECD}$ and $FLRT_{LRR}$ constructs. S200 gel filtration profiles and 12% SDS-PAGE gels for (A) $FLRT1_{ECD}$, (B) $FLRT2_{ECD}$, (C) $FLRT3_{ECD}$, (D) $FLRT1_{LRR}$, (E) $FLRT2_{LRR}$, (F) $FLRT3_{LRR}$. All proteins were expressed in HEK 293T cells. Protein samples were prepared for SDS-PAGE analysis under reducing conditions.

4.4 Investigation of FLRT construct oligomerisation

FLRTs have previously been shown to interact directly (Bottcher et al., 2004; Karaulanov et al., 2006; Söllner and Wright, 2009), and cell-cell adhesion experiments suggest that this interaction may be dependent on the LRR domain (Karaulanov et al., 2006; Tossell et al., 2011). As discussed in Section 1.5.3, LRR domains are frequently implicated in protein-protein interactions, including homodimerisation, which in some cases is necessary for the function of the protein (Scott et al., 2004; Scott et al., 2006; Seiradake et al., 2009; Kajander et al., 2011). To characterise the intermolecular interactions underlying the homotypic cell-cell interactions described in Section 4.2 and elsewhere (Karaulanov et al., 2006; Tossell et al., 2011), soluble $FLRT_{ECD}$ and $FLRT_{LRR}$ constructs (produced as described in Section 4.3) were subjected to analysis by SEC-MALS and AUC.

4.4.1 Multi-angle light scattering (MALS)

Protein samples were prepared as described in Section 4.3, and fractions from gel filtration were pooled to produce samples with the highest possible concentration (as the size exclusion chromatography component of the experiment dilutes the sample) and maximum purity, to exclude the possibility of a significant contribution to light scattering from contaminants, which would distort the data; this is especially true of high molecular weight contaminants as larger particles scatter more. MALS experiments were performed as described in Section 2.3.3, and the $FLRT_{ECD}$ proteins all eluted as a single main peak, corresponding to species of molar mass between 70 and 73 kDa (see Figure 4.7 and Table 4.1).

For interpretation of these values in the context of the possible oligomeric state of the proteins, they were compared to the predicted molar masses, taking into account that these proteins were produced in HEK 293T cells without the addition of kifunensine,

and are therefore likely to have been post-translationally modified with a heterogeneous population of complex glycans (Figure 4.6 and (Chang et al., 2007)). Analysis of the MALS data requires an estimate of the relative masses of protein and glycan in the analysed sample, but this heterogeneity makes it difficult to calculate the mass of glycan added to the proteins precisely, especially since the occupancy of glycosylation sites will not be complete. However, a working estimate for the purposes of this experiment was obtained from mass spectrometry of glycans attached to the Nipah virus glycoprotein when similarly expressed (Bowden et al., 2008). The mass/charge ratios of singly charged glycans released from this protein by in-gel PNGase F digestion range from approximately 1260 (corresponding to $Man_5GlcNAc_2$) to 2120 (corresponding to a low abundance complex glycan), and values in the middle of this range (i.e. approximately 1700) correspond to several complex type glycan species. A value of 1700 Da per glycan was therefore adopted to estimate the total predicted molar mass of the $FLRT_{ECD}$ samples.

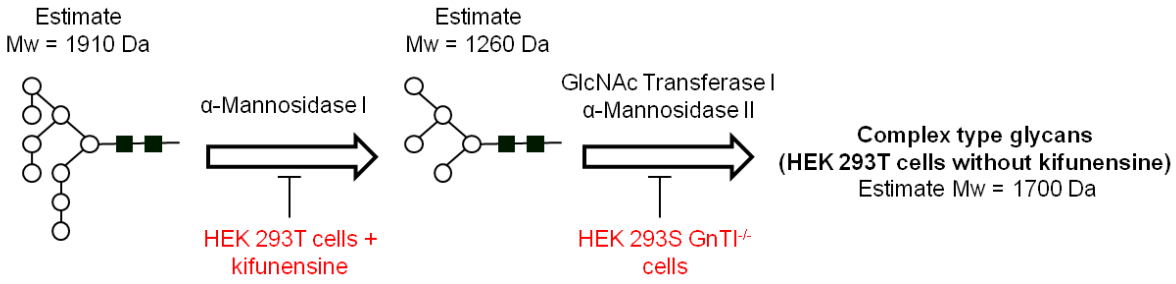


Figure 4.6 A simplified schematic view of a section of the *N*-glycosylation pathway in mammalian cells, where glycans are transferred to proteins at signature Asn-X-Ser/Thr motifs (where X is any amino acid except Pro). Enzymes which catalyse the key reactions relevant for calculation of added glycan mass for the purpose of analysing MALS data are shown, as well as the relevant mammalian expression systems used in this thesis. Variable site occupancy and glycan diversity contribute to a potentially complex mixture of glycans attached to cell surface proteins, but for the purposes of analysing MALS data, an estimate of glycan size (determined from analysis of other proteins expressed using these systems) is given for each expression system. Monosaccharide constituents are represented as follows: closed squares, GlcNAc; open circles, Mannose. Figure adapted from (Chang et al., 2007) with molecular weight estimates from (Bowden et al., 2008).

The molar mass values of the $FLRT_{ECD}$ proteins obtained by MALS were significantly higher than the predicted values for the molar mass of a monomer (Table 4.1), to an extent that cannot be fully explained by the addition of large glycans by HEK 293T cells; the values were, however, lower than would be predicted for a stable dimer. The values could be explained by the presence of a stable equilibrium between monomer and dimer forms of the $FLRT_{ECD}$ proteins in solution, where monomers interact weakly with a fast on/off rate that means the size exclusion chromatography step of the technique cannot resolve the two species into separate peaks. This means the elution volume of the protein peak, and associated molar mass, have values intermediate between what would be predicted for monomer and dimer forms, as the single peak is in fact an average for the two species present. Because of this, the value of the MALS average mass can also indicate the position of the equilibrium between monomer and dimer forms. For the $FLRT_{ECD}$ constructs, the MALS average

mass value is closer to the predicted mass of a monomer than that of a dimer, indicating that the equilibrium is positioned in favour of the monomer form.

$FLRT_{LRR}$ proteins, which were produced in HEK 293T cells with the addition of kifunensine, and therefore have a less diverse and better characterised *N*-glycan population, were analysed in the same way. These also all elute as a single main peak, corresponding to molar masses slightly higher than that predicted for glycosylated monomers (Figure 4.7, Table 4.1). This increase is less significant than that observed for the $FLRT_{ECD}$ proteins, especially for $FLRT2_{LRR}$ and $FLRT3_{LRR}$, which may indicate that any dimerisation interaction is very weak, and that almost all molecules in the population exist in a monomeric form. This might imply that for FLRT2 and FLRT3, the LRR domain contributes slightly to self-association, but other parts of the ectodomain are required for the monomer/dimer equilibrium behaviour observed for $FLRT2_{ECD}$ and $FLRT3_{ECD}$.

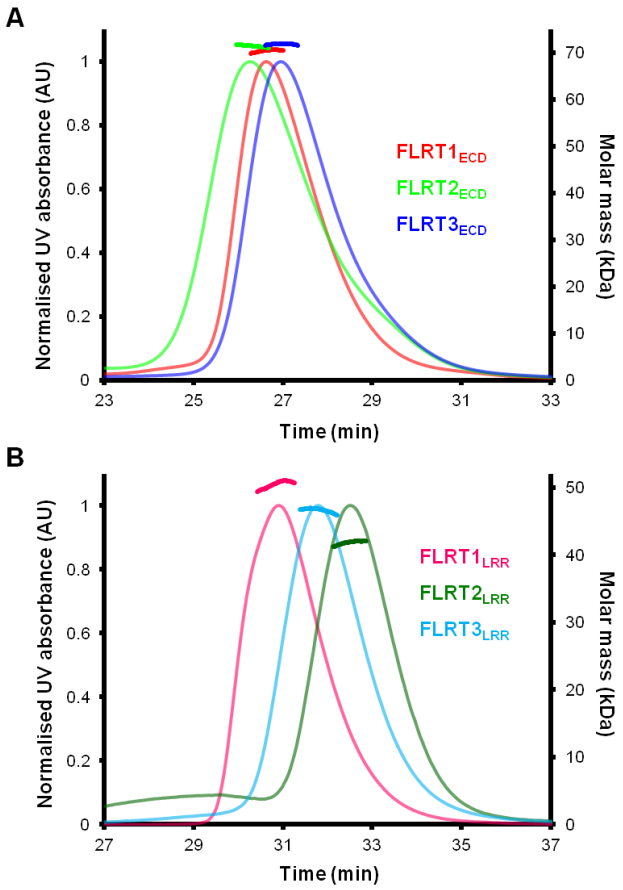


Figure 4.7 Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) was used to investigate the oligomerisation state of soluble mouse (A) $FLRT_{ECD}$ and (B) $FLRT_{LRR}$ constructs. (A) The measured molar masses appear to indicate (when compared to predicted masses, see Table 4.1) that all three $FLRT_{ECD}$s exist in a stable monomer/dimer equilibrium positioned towards the monomer form, with fast on/off binding between monomers. (B) Measured molar masses for $FLRT_{LRR}$ proteins are closer to values predicted for the monomer forms (Table 4.1), though are still higher than these values. Scaled UV absorbance traces are shown by fine lines, and the measured molar masses are shown by bold lines.

However, the difference between the predicted molar mass and average MALS mass was more significant for $FLRT1_{LRR}$ (Table 4.1), indicating that interaction between LRR domains may play a more significant role in dimerisation of this protein. In fact, on further investigation using samples of differing concentrations, the MALS-determined molar mass of $FLRT1_{LRR}$ was found to increase with concentration (Figure 4.8, Table 4.1). At the lowest concentration tested, the molar mass

determined by MALS was only slightly higher than the mass predicted for a monomer, but at the highest concentration, the peak retention time on the size exclusion column was reduced and the MALS estimate approached the mass predicted for a $FLRT1_{LRR}$ dimer. This indicates that the $FLRT1_{LRR}$ monomer/dimer equilibrium proposed above is concentration dependent, shifting from favouring the monomeric form of the protein at lower concentrations, to favouring the dimeric form at higher concentrations.

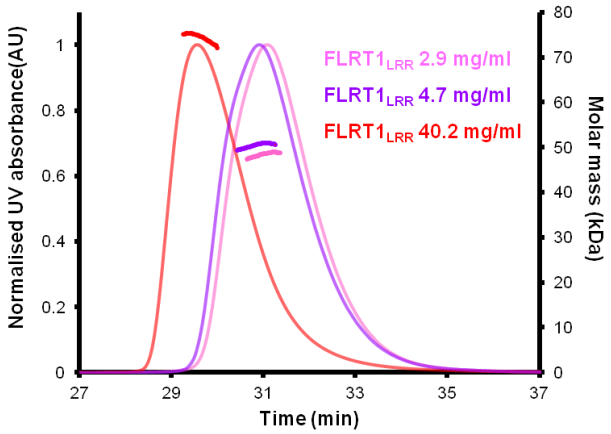


Figure 4.8 SEC-MALS results for a concentration series of $FLRT1_{LRR}$. The apparent molecular weight increases with concentration (see also table 4.1), which may indicate that a monomer-dimer equilibrium is shifted from favouring the monomer form at low concentration, to the dimeric form at high concentration. Scaled UV absorbance traces are shown by fine lines, and the measured molar masses are shown by bold lines

Construct	Expression system (mass of *N*-glyc. (kDa))	Conc. of protein sample (mg/ml)	No. *N*-glyc. sites	Theoretical Mw (inc *N*-glyc.) (kDa)	MALS peak apparent Mw (kDa)	Suggested oligomeric state
$FLRT1_{ECD}$	HEK 293T no kifunensine (1.70)	5.1	2	56.1 (59.5)	70.9	Monomer/ dimer
$FLRT2_{ECD}$	HEK 293T no kifunensine (1.70)	7.1	3	57.2 (62.3)	71.8	Monomer/ dimer
$FLRT3_{ECD}$	HEK 293T no kifunensine (1.70)	9.7	3	57.3 (62.4)	72.3	Monomer dimer
$FLRT1_{LRR}$	HEK 293T with kifunensine (1.91)	2.9	2	38.6 (42.4)	48.4	Mostly monomer
$FLRT1_{LRR}$	HEK 293T with kifunensine (1.91)	4.7	2	38.6 (42.4)	50.6	Monomer/ dimer
$FLRT1_{LRR}$	HEK 293T with kifunensine (1.91)	40.2	2	38.6 (42.4)	74.5	Mostly dimer
$FLRT2_{LRR}$	HEK 293T with kifunensine (1.91)	2.9	1	38.3 (40.2)	41.8	Mostly monomer
$FLRT3_{LRR}$	HEK 293T with kifunensine (1.91)	2.9	3	38.7 (44.5)	46.5	Mostly monomer

Table 4.1 Summary of MALS experiments on soluble FLRT ectodomain and LRR domain constructs, including the suggested oligomeric state deduced by comparison of the theoretical and MALS-derived values for molecular weight (Mw). The $FLRT1_{LRR}$ concentration series is highlighted in pink.

MALS data can also be used to calculate dissociation constant (K_d) values for the monomer-dimer interactions proposed above (see Formulae 1-6 in Appendix C and (Attri and Minton, 2005)), to provide information on the affinity of any homotypic interaction. Assuming a monomer-dimer equilibrium, with no higher oligomerisation states, and defining the limits of the protein concentration peak arbitrarily, K_d values were calculated for $FLRT1_{ECD}$ (at 5.1 mg/ml) and $FLRT1_{LRR}$ (at 4.7 mg/ml and

40.2 mg/ml) to permit comparison between an ectodomain and LRR domain-only construct. In line with the similarities observed between the apparent molecular weight distributions (and implied monomer-dimer equilibria), the calculated K_d values are relatively consistent between the two constructs. $FLRT1_{ECD}$ has an average K_d across the concentration peak of 102 µM (standard deviation 2.25 µM), while $FLRT1_{LRR}$ has an average K_d of 60 µM (standard deviation 19.5 µM) at 4.7 mg/ml, and 161 µM (standard deviation 22.4 µM) at 40.2 mg/ml (Figure 4.9). These values confirm that, as suggested by the single peak MALS profile, the interaction between monomers to form a dimer is fairly weak. They also agree with the suggestion that the oligomerisation behaviour is fairly similar for $FLRT1_{ECD}$ and $FLRT1_{LRR}$, which implies that homotypic interactions between FLRT1 ectodomains are predominantly mediated *via* the LRR domain.

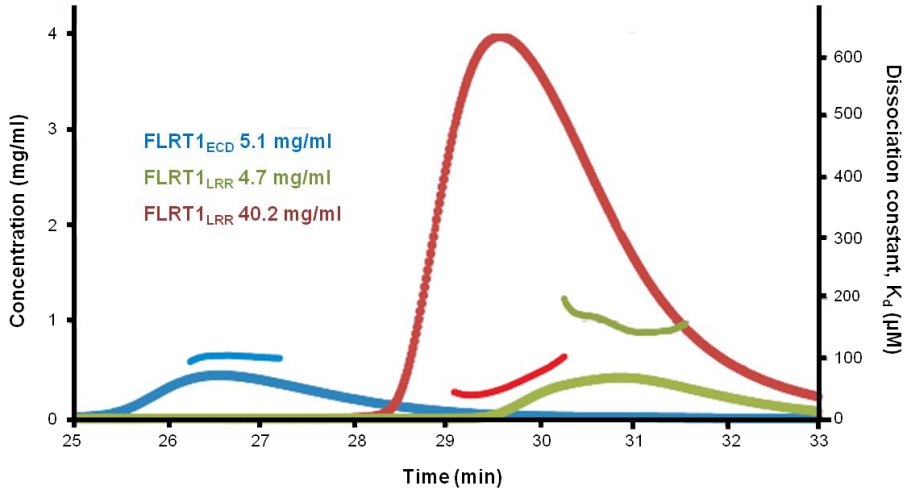


Figure 4.9 Dissociation constants, calculated from MALS data assuming a monomer-dimer equilibrium with no higher oligomerisation, for $FLRT1_{ECD}$ and $FLRT1_{LRR}$ at two concentrations (see Formulae 1-6 in Appendix C). The K_d values are relatively consistent between the two constructs. Protein concentration traces are shown by bold lines, and calculated K_d values are shown by fine lines. Figure produced with assistance from E. Seiradake (Strubi).

4.4.2 Analytical Ultracentrifugation (AUC)

Sedimentation velocity analytical ultracentrifugation experiments were also performed with mouse $FLRT_{ECD}$ and $FLRT1_{LRR}$ constructs as described in Section 2.3.2, to provide information on the number of oligomeric species in solution to supplement the data obtained from MALS (Figure 4.10). Different species may be identified by their sedimentation coefficients, which reflect their mass, density and shape; higher value coefficients indicate species which sediment faster, often due to their larger size, and may correspond to higher oligomeric association states. Species may also be characterised by c(S, f/f0), a probabilistic function of a species with a given sedimentation coefficient, S, having a certain frictional coefficient, f/f0; frictional coefficients express the deviation of a species' behaviour in solution from how it would behave if it were spherical, given its molecular weight. Peaks in the c(S, f/f0) probability function can therefore be used to indicate a rough estimate of the molecular weight of a species, though the accuracy of this can vary depending on how much the shape of the species deviates from that of a sphere.

$FLRT1_{ECD}$, $FLRT2_{ECD}$ and $FLRT3_{ECD}$ constructs were produced in HEK 293S cells, which resulted in relatively low yields compared to expression in HEK 293T cells. These protein batches also contained unidentified higher molecular weight contaminants which proved difficult to separate from the $FLRT_{ECD}$ proteins by methods including ion exchange and heparin affinity chromatography (data not shown), making purification to homogeneity in amounts suitable for AUC characterisation challenging. Nevertheless, sufficient quantities were prepared for each construct, and analysis indicated that each sample contained multiple species. With higher confidence in the purity of the samples, it might be possible to infer that

each peak corresponds to a different oligomeric species of the $FLRT_{ECD}$ under study, but this may not be possible for these samples.

For example, peaks in the sedimentation coefficient distribution (c(S)) in the $FLRT1_{ECD}$ and $FLRT3_{ECD}$ results where the sedimentation coefficient was less than 4.0 S were discounted as modelling of the maximum molecular weight of these species using SolPro software (University of Murcia, Spain) found that these species would be significantly smaller than the theoretical molecular mass of the $FLRT_{ECD}$ constructs under study; these peaks are likely to correspond to breakdown products or low molecular weight contaminants. Likewise, low abundance peaks corresponding to very high sedimentation coefficients were treated with suspicion. Although some of these peaks have non-trivial c(S, f/f0) solutions (indicated by off-diagonal peaks in the "temperature" scale in Figure 4.10) corresponding to molecular weight estimates, they were considered more likely to represent trace amounts of high molecular weight contaminants or non-specific aggregates rather than higher oligomeric species of $FLRT_{ECD}$s.

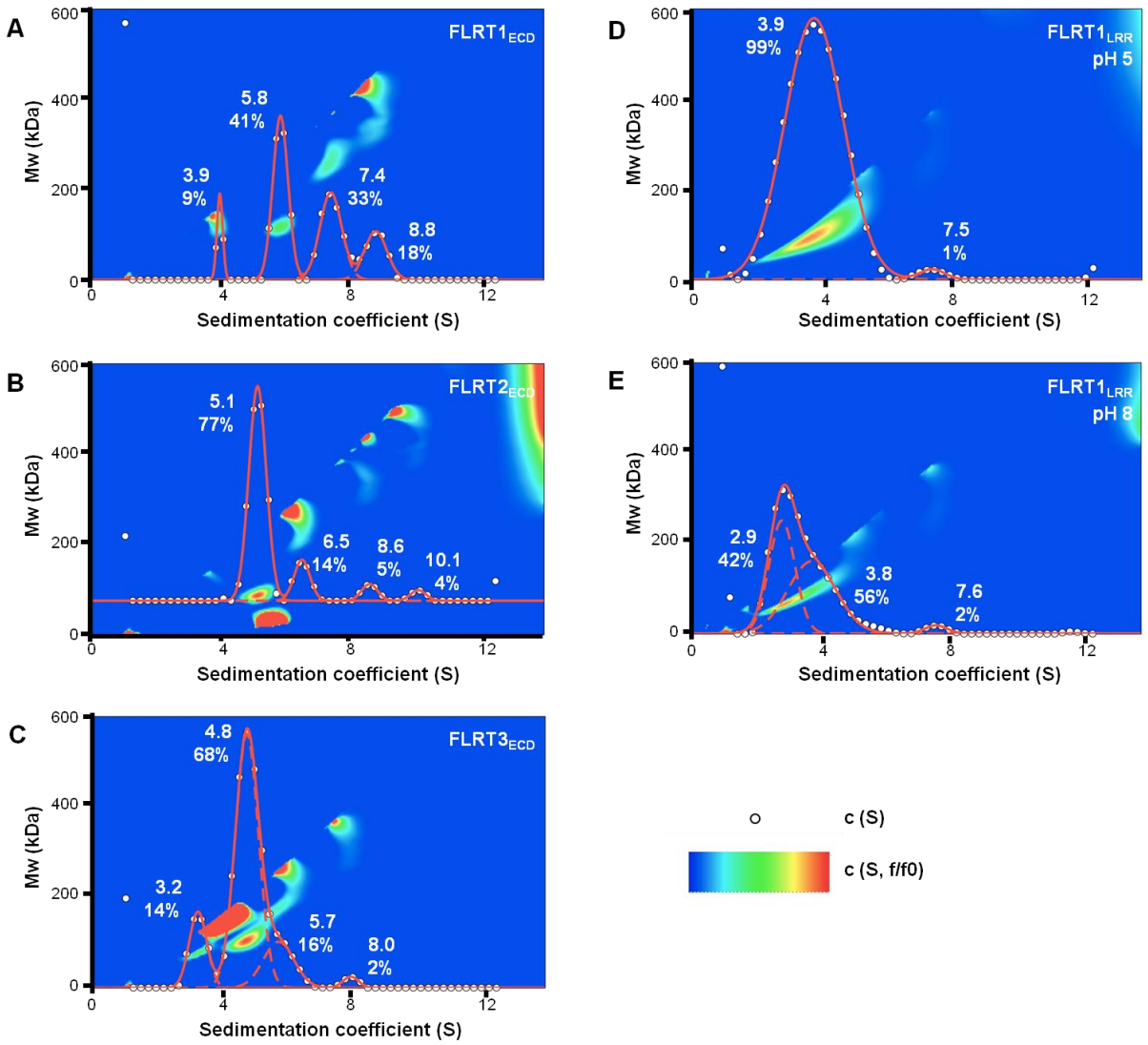


Figure 4.10 Sedimentation velocity AUC experiments for mouse $FLRT_{ECD}$ and $FLRT_{LRR}$ constructs expressed in HEK 293S cells. The multiple peaks in the sedimentation coefficient distribution, c(S), for (A) $FLRT1_{ECD}$, (B) $FLRT2_{ECD}$ and (C) $FLRT3_{ECD}$, all tested at pH 8, indicate the presence of multiple species in solution, though interpretation was made challenging by the presence of contaminants. The shape of c(S) for (D) $FLRT1_{LRR}$ at pH 5 and (E) $FLRT1_{LRR}$ at pH 8 indicate the presence of two main species in solution, which interconvert more rapidly at the lower pH such that the peaks for the two species cannot be resolved. Peaks in c(S), plotted along the x-axis, are labelled with corresponding sedimentation coefficient value (S) and relative abundance, determined using the areas under each peak. The probability function c(S, f/f0) is plotted on the x-y plane and is shown by a "temperature" scale (blue = low probability, red = high); non-trivial solutions corresponding to Mw estimates from the y-axis are shown by off-diagonal peaks. Figures produced using ProFit (QuantumSoft) by Robert Gilbert (Strubi).

However, a degree of careful interpretation is possible. For example, if the two peaks in the $FLRT2_{ECD}$ sedimentation coefficient distribution corresponding to the highest sedimentation coefficient values are excluded as being likely to be due to contaminants, then an average of the theoretical molecular weights of the monomer and dimer forms, weighted by the relative abundance of the first and second peaks in the AUC spectrum, gives a very close approximation to the apparent molecular weight value obtained by MALS (Table 4.2). The same is true for $FLRT3_{ECD}$, if the first peak is excluded for being too small to correspond to a full-length $FLRT3_{ECD}$ protein, and the last peak is excluded as likely to correspond to a high molecular weight contaminant (Table 4.2). Overall, these results confirm the presence of multiple oligomeric forms of $FLRT_{ECD}$s in solution, though more in-depth analysis was hindered by the quality of the protein samples.

This AUC method was also used to analyse samples of $FLRT1_{LRR}$ protein, also expressed in HEK 293S cells (Figure 4.10). Two samples were analysed, one at pH 5 and one at pH8, to reflect the conditions under which $FLRT1_{LRR}$ was crystallised in distinct crystal forms (see Chapter 5), and to determine whether pH affects the homotypic association between $FLRT1_{LRR}$ monomers suggested by MALS analysis (see Section 4.4.1). Because $FLRT1_{LRR}$ expressed with a higher yield in this system than the $FLRT_{ECD}$s, it was easier to separate it from contaminants while retaining sufficient quantities for analysis, making the results somewhat easier to interpret. A very low abundance peak corresponding to a sedimentation coefficient of approximately 7.5 was observed in both sets of results (Figure 4.10), but this is the only evidence of a small amount of higher molecular weight contaminants in these samples.

Construct	Theoretical molecular weight from HEK 293S cells (dimer) (kDa)	MALS apparent molecular weight (Table 4.1) (kDa)	Peaks excluded from AUC results	Included peaks with suggested oligomeric form (adjusted relative abundance)	Theoretical molecular weight, averaged by relative abundance (kDa)
$FLRT2_{ECD}$	61.0 (122.0)	71.8	S=8.6 S=10.1 Likely contaminants	S=5.1, monomer (84 %) S=6.5, dimer (16 %)	70.8 84% monomer, 16% dimer
$FLRT3_{ECD}$	61.1 (122.2)	72.3	S=3.2 S=8.0 Likely breakdown products and/or contaminants	S=4.8, monomer (81 %) S=5.7, dimer (19 %)	72.7 81% monomer, 19% dimer

Table 4.2 Agreement between AUC and MALS data for $FLRT2_{ECD}$ and $FLRT3_{ECD}$, provided some AUC peaks are discarded for being likely to correspond to contaminants. Theoretical average molecular weights for these two constructs, assuming a monomer-dimer equilibrium of the type suggested by MALS, and weighted by relative abundance as observed by AUC, provide estimates that very closely approximate the average molecular weight values observed for the putative equilibria by MALS (both types of estimate are highlighted in green).

Excluding this peak from the analysis of data at pH 8, there are two overlapping peaks which may correspond to monomer (S = 2.9) and dimer (S = 3.8) forms of $FLRT1_{LRR}$ (Figure 4.10). At pH 5, these two peaks appear to have merged to form a single, broad peak (S = 3.9). As these two aliquots were obtained from the same sample of $FLRT1_{LRR}$, it is possible to exclude the possibility of gross inconsistencies in the content of the samples, so this peak merging effect is caused by the decrease in pH. Despite a significant overlap at pH 8, the technique is able to resolve the two peaks, but appears unable to do so at pH 5; one explanation for this is that the exchange between the monomer and dimer populations, which MALS has previously indicated to be relatively fast (see Section 4.4.2), becomes even faster at lower pH,

such that although both populations still exist, they can no longer be resolved by AUC.

4.5 Conclusions and future perspectives

The data presented in this Chapter provide evidence for the effects of membrane-bound FLRT expression on cell-cell adhesion properties, and for some details of the molecular interactions which may underlie this behaviour. Transmembrane $FLRT_{LONG}$ constructs are expressed on the surface of HEK 293 cells, and localise to the interfaces between opposing cells; they also induce homotypic cell-cell adhesion in Sf9 cells. Soluble constructs encoding FLRT ectodomains ($FLRT_{ECD}$) and LRR domains ($FLRT_{LRR}$) were expressed and purified for the purpose of biophysical analyses which it was hoped would shed light on the ability of FLRT molecules to associate homotypically, potentially *via* their LRR domains (which are frequently implicated in protein-protein interactions, including functionally relevant homodimerisation). Such interactions, if they occur in *trans* between opposing cells, might help to explain the homotypic adhesion behaviour seen in cells.

The soluble FLRT proteins were indeed shown to interact homotypically, by MALS and AUC. MALS provided evidence for a stable but low-affinity monomer-dimer equilibrium, positioned in favour of the monomer form, for all three $FLRT_{ECD}$ constructs. The existence of at least two oligomeric forms of the $FLRT_{ECD}$ constructs in solution was further confirmed by AUC, and although analysis was hampered by sample heterogeneity, relative proportions of monomer and dimer forms of $FLRT2_{ECD}$ and $FLRT3_{ECD}$ were in good agreement with MALS data. The apparent molecular weight value obtained by MALS for $FLRT2_{LRR}$ and $FLRT3_{LRR}$ constructs was only slightly higher than the theoretical value for the monomer forms, indicating that in

these proteins the LRR domain may be necessary (shown previously for FLRT3 in cellular studies (Karaulanov et al., 2006)) but not sufficient for a homotypic interaction.

The behaviour of $FLRT1_{LRR}$ in MALS, however, was very comparable to that of $FLRT1_{ECD}$; MALS indicated a monomer-dimer equilibrium, which was supported by AUC data showing two species which interconvert more rapidly at lower pH. The $FLRT1_{LRR}$ monomer-dimer equilibrium also appears to be concentration dependent, shifting to favouring the dimeric form at high concentrations such as might be found in a tightly-packed cell surface environment. Characterisation of the dissociation constants for a monomer-dimer interaction between $FLRT1_{ECD}$ and $FLRT1_{LRR}$ also revealed these values to be relatively consistent between the ectodomain and the LRR domain. This evidence together indicates that $FLRT1_{ECD}$ behaves similarly to $FLRT1_{LRR}$ in solution in terms of the ability of these constructs to weakly dimerise, which implies that in FLRT1, dimerisation between the ectodomains is mediated predominantly by the LRR domain.

Taking all these data into account, it is possible to hypothesise that homodimerisation of FLRT ectodomains (which appears for FLRT1 at least to be mediated *via* the LRR domain) observed using biophysical characterisation underlies the interactions observed at the cellular level. It is worth bearing in mind that in a cell surface environment, such as that probed here in HEK 293 and Sf9 cells, FLRT molecules are more restrained than constructs in solution due to their membrane localisation, and are tightly packed (especially at cell-cell interfaces; Figure 4.2). This means an equilibrium between monomeric and dimeric forms (e.g. in *trans* between two expressing cells) of FLRTs at the cell membrane is likely to be more finely balanced,

and due to the higher effective concentration, may be positioned more towards the dimeric form than in a lower concentration solution.

The results presented in this Chapter provide the basis for much of the research to be discussed in Chapter 5, which examines the molecular basis for FLRT LRR domain dimerisation using crystal structures, and the functional relevance of FLRT self-association at the cell surface.

Chapter 5 Structural Analysis of FLRT LRR domains

Chapter 5 Structural Analysis of FLRT LRR domains

5.1 Introduction

In chapter 4, the biophysical evidence for homotypic association of soluble FLRT ectodomain ($FLRT_{ECD}$) and LRR domain ($FLRT_{LRR}$) constructs was discussed. These data indicate that the ectodomains of all three FLRTs self-associate, and that this behaviour might best be characterised as a stable equilibrium between monomer and dimer populations, positioned in favour of the monomer form. In particular, the dimerisation behaviour exhibited by the ectodomain of FLRT1 is closely mirrored by that of the LRR domain of the same protein, indicating that dimerisation between ectodomains is mediated predominantly by the LRR domain.

As discussed in Section 1.5.1, LRR domains often participate in protein-protein recognition, and more specifically, functionally relevant homodimerisation (Scott et al., 2006; Seiradake et al., 2009; Kajander et al., 2011). Homotypic cell sorting of FLRT-transfected cells, which may underlie the roles of FLRT family members in neurite outgrowth and development, is dependent on the LRR domain (Karaulanov et al., 2006; Tossell et al., 2011), and it is possible that this behaviour is in fact reliant on LRR-mediated FLRT homodimerisation at the cell surface. However, the molecular basis for such homodimerisation is not yet clear.

The molecular architecture of LRR domains has been demonstrated by a number of crystal structures (Bella et al., 2008) which have a curved solenoid topology (see Figure 1.8). The concave face of the solenoid has frequently been shown to act as the site for protein-protein interactions including homodimerisation (Scott et al., 2004; Scott et al., 2006; Seiradake et al., 2009; Kajander et al., 2011), with variable solvent exposed residues on this surface determining the specificity of the interaction. In the

case of LRR-containing cell-surface proteins including AMIGO-1, biglycan and decorin, antiparallel homodimerisation (i.e. between chains running in opposite directions) mediated by the LRR concave face has been shown to be essential for correct cell surface expression (Scott et al., 2006; Kajander et al., 2011). Additionally, the AMIGO-1 homodimerisation interface that is required in *cis* has been suggested to underlie *trans* dimerisation between molecules on opposing cell surfaces to mediate cell adhesion, as illustrated in Figure 1.7.

The first main aim of this chapter is to determine whether the homodimerisation behaviour displayed by FLRT LRR domains and ectodomains in solution can also be observed at the cell surface. The second aim is to investigate the molecular basis of any self-association by determining the crystal structures of the FLRT LRR domains, and to assess the physiological relevance of these structures using cellular studies, mutagenesis and conservation across the FLRT family.

5.2 Studies of FLRT association state at the cell surface

5.2.1 Fluorescence Resonance Energy Transfer experiments

Having shown in chapter 4 that FLRT constructs derived from the ectodomain portion interact homotypically in solution, cellular studies were employed to determine whether transmembrane FLRT constructs also self-associate on the surface of expressing cells. The expression of $FLRT_{LONG}$ constructs on the surface of Cos-7 cells has been confirmed by immunocytochemical staining (Section 4.2.1, Figure 4.1), and this cell line was used in two assays to probe *in vivo* $FLRT_{LONG}$ self-association at the cell surface: Förster (or fluorescence) resonance energy transfer (FRET; all three FLRTs) and cell surface co-patching (FLRT1 only).

FRET can be used as a "molecular ruler" to probe the distances between two molecules (e.g. FLRT proteins) which each include a different member of a "FRET pair" of fluorophores, defined as fluorophores where the emission spectrum of the first ("donor") overlaps with the absorbance spectrum of the second ("acceptor"). This means that excitation of the donor fluorophore within a short distance of the acceptor can lead to energy transfer, and fluorescence of the acceptor fluorophore in the absence of exogenous excitation within its absorbance spectrum. Fluorescence from the acceptor upon excitation of the donor can therefore be interpreted as evidence that the two fluorophores are situated close together in space, and may indicate a direct interaction between the two protein molecules to which the fluorophores are attached.

As levels of fluorescence are dependent upon the levels of expression of the protein of interest, which can vary between different batches of cells, this should be taken into account when performing FRET experiments. To control for different levels of expression, FRET can be measured using an acceptor photobleaching experiment, in which donor fluorescence is monitored before and after the acceptor fluorophore is selectively photobleached (Nashmi et al., 2003; Liu et al., 2004). As photobleaching destroys the fluorescence of the acceptor, it precludes the ability of the donor to transfer energy even if the two fluorophores remain in close proximity, and therefore donor fluorescence increases, indicating that prior to photobleaching, FRET was occurring. Conversely, if no increase in donor fluorescence is observed after acceptor photobleaching, this is evidence that no FRET transfer is occurring and the proteins of interest are not sufficiently close. Because this method measures the change in donor fluorescence upon acceptor photobleaching relative to initial levels, it provides some control for variation in expression levels.

To apply the FRET approach to the question of whether FLRTs associate homotypically (e.g. as homodimers) at the cell surface, $FLRT_{LONG}$ (and control protein) constructs were designed including C-terminal mVenus and mCerulean tags, as illustrated for $FLRT_{LONG}$-mVenus in Figure 4.1; details of all constructs are listed in Appendix B. The mVenus and mCerulean tags are a FRET pair of fluorophores, with mCerulean acting as the donor fluorophore (which is monitored for increased fluorescence) and mVenus acting as the acceptor (which is photobleached). Cos-7 cells were transfected with a 1:1 mixture of DNA encoding $FLRT_{LONG}$-mVenus and $FLRT_{LONG}$-mCerulean constructs, or with a 1:1 mixture of mVenus-tagged non-FLRT cell surface control proteins (CD2-mVenus and TrkC-mVenus) and $FLRT_{LONG}$-mCerulean constructs. CD2, a cell adhesion molecule, and TrkC, a receptor tyrosine kinase, were selected as suitable controls as they have no documented association with the FLRT family so would not be expected to form direct interactions leading to FRET signal at the cell surface. A batch of cells were also transfected (as described in Section 2.7.2) with DNA encoding $FLRT_{LONG}$-mCerulean only, to confirm that FRET signal cannot be observed in the absence of the acceptor fluorophore. Cells were then fixed, mounted and imaged as described in Section 2.8.2.

The percentage increase in mCerulean fluorescence across the selected region of the cell membrane was calculated for each cell tested as described in Section 2.8.2; any negative values were corrected to zero to discount the effect of any donor fluorophore bleaching during the experiment. Results were then pooled and initial mean and standard deviation values calculated, and results further than two standard deviations away from the initial mean were excluded as outliers, before adjusted mean and standard deviation values were calculated using the remaining results.

Summarised results for each population of cells are illustrated in Figure 5.1, and tabulated in Table 5.1.

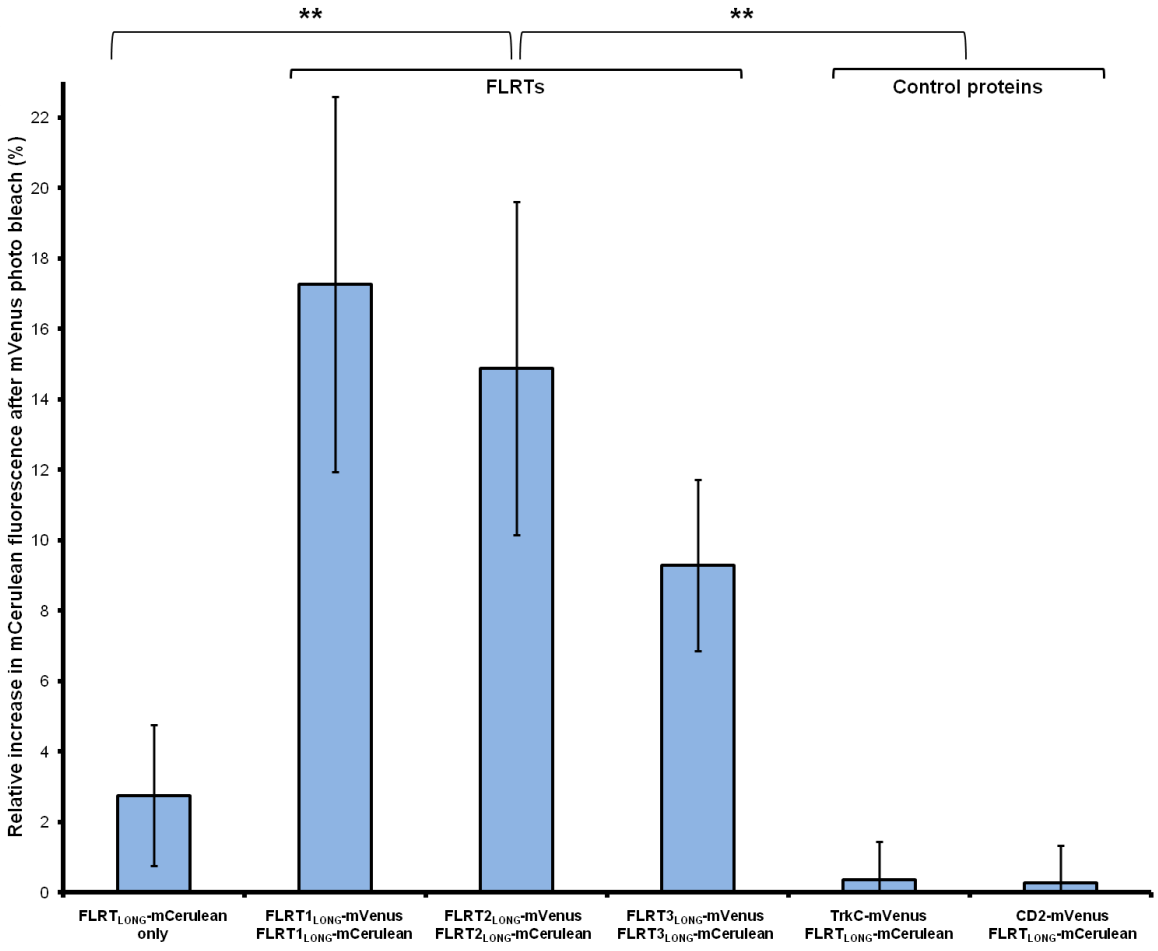


Figure 5.1 Results from FRET experiments between FLRT$_{LONG}$ proteins, as well as between FLRT$_{LONG}$ proteins and TrkC and CD2 control proteins and for FLRT$_{LONG}$-mCerulean proteins without an mVenus FRET pair partner. An increase in mCerulean fluorescence following photobleaching of the mVenus acceptor fluorophore indicates that the two fluorophores are in close proximity and that there may be a direct interaction between the proteins of interest. Data group values as listed in Table 5.1. **, $P < 0.001$ (Student's *t* test). Error bars indicate standard deviation.

mVenus construct (acceptor)	mCerulean construct (donor)	Number of cells	Initial mean relative increase in mCerulean fluorescence (%)	Initial s.d. of relative increase in mCerulean fluorescence (%)	Number of outlier cells excluded* (number of cells remaining)	Adjusted mean relative increase in mCerulean fluorescence (%)	Adjusted s.d. of relative increase in mCerulean fluorescence (%)
None	$FLRT1_{LONG}$/ $FLRT2_{LONG}$/ $FLRT3_{LONG}$	29	2.75	2.00	0 (29)	2.75	2.00
$FLRT1_{LONG}$	$FLRT1_{LONG}$	43	17.7	6.56	3 (40)	17.3	5.32
$FLRT2_{LONG}$	$FLRT2_{LONG}$	49	15.2	5.26	1 (48)	14.9	4.73
$FLRT3_{LONG}$	$FLRT3_{LONG}$	29	9.08	2.63	1 (28)	9.28	2.44
TrkC	$FLRT1_{LONG}$/ $FLRT2_{LONG}$/ $FLRT3_{LONG}$	60	1.22	3.17	5 (55)	0.359	1.08
CD2	$FLRT1_{LONG}$/ $FLRT2_{LONG}$/ $FLRT3_{LONG}$	60	0.738	2.46	3 (57)	0.281	1.05

Table 5.1 Summarised results of acceptor photobleaching FRET experiments to determine whether FLRT molecules self-associate at the cell surface. For experiments where $FLRT_{LONG}$-mCerulean proteins were transfected with either no mVenus tagged protein, or with control mVenus tagged proteins, results were pooled for cells expressing all three FLRT proteins as indicated. s.d., standard deviation. *Results were excluded from further analysis if the FRET value lay ≥ 2 initial standard deviations from the initial mean. Results shaded in blue are illustrated in Figure 5.1.

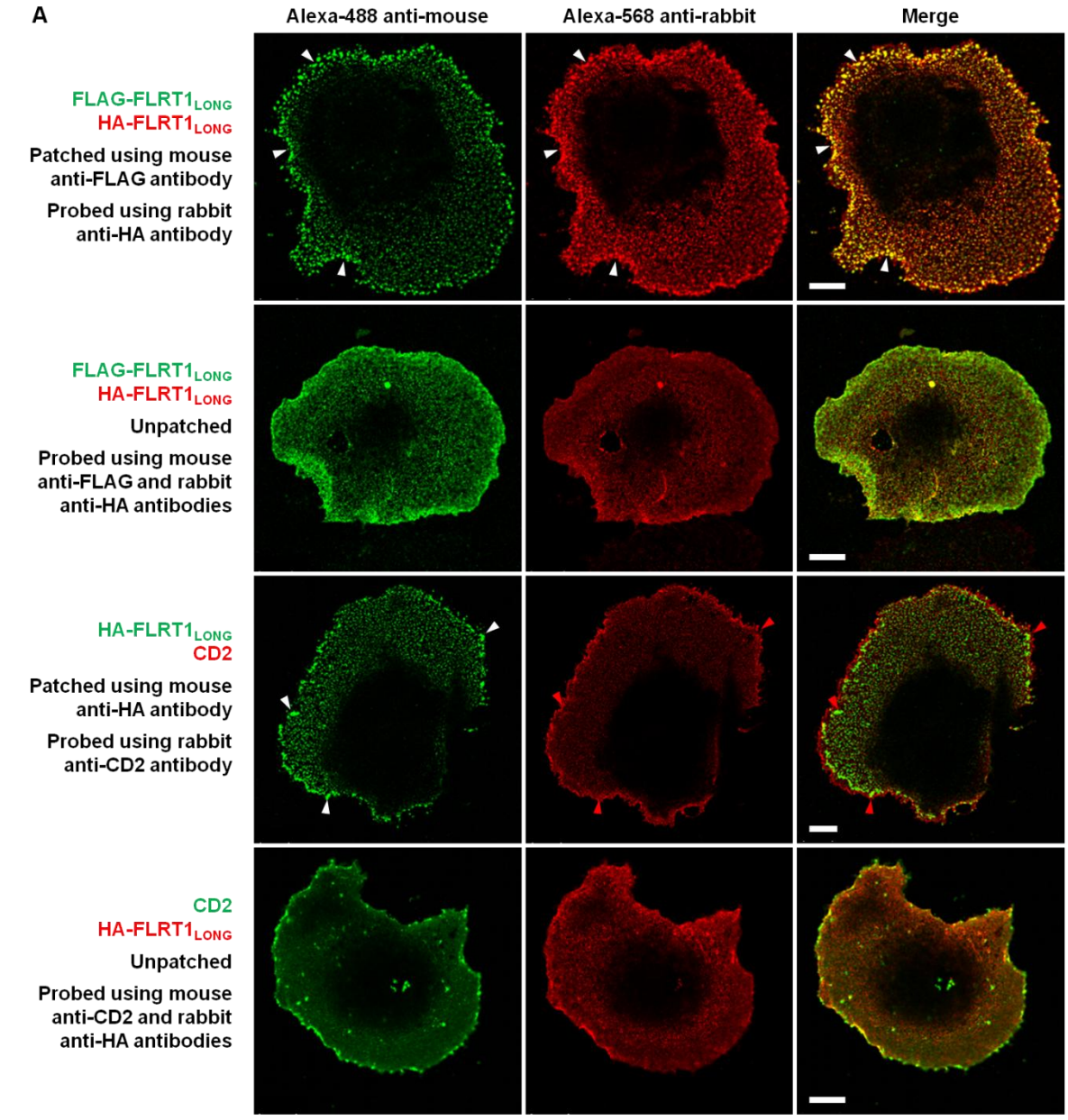
As illustrated in Figure 5.1, there were significant increases in mCerulean fluorescence following mVenus photobleaching for all three $FLRT_{LONG}$ FRET pairs. This indicates that the two fluorophore tags are in close proximity, and that there is likely to be a direct interaction between the proteins to which the tags are attached, to allow them to approach each other so closely. All three sets of FLRT results were significantly different, at the $P < 0.001$ level, to data from cells transfected only with $FLRT_{LONG}$-mCerulean (to control for background mCerulean fluorescence increase over the photobleaching period), and to data from cells transfected with $FLRT_{LONG}$-mCerulean and either TrkC-mVenus or CD2-mVenus control proteins. This indicates that FLRT proteins do not interact directly with either TrkC or CD2, and that the FRET signal observed is likely to be due solely to specific interactions between tagged proteins, and not to non-specific packing caused by overexpression at the cell surface. However, these results do indicate that there is a specific and direct homotypic interaction between all three FLRT molecules at the cell surface, and are consistent with FLRTs forming homodimers *in vivo*, as they were shown to do *in vitro* in Chapter 4.

5.2.2 Cell-surface co-patching of FLRT1

To complement results from the FRET experiments and investigate whether FLRT1 molecules associate at the cell surface using a different technique, another *in vivo* cell-based assay was employed using Cos-7 cells. Cells were co-transfected with a 1:1 mixture of DNA encoding two constructs, either HA-$FLRT1_{LONG}$ with FLAG-$FLRT1_{LONG}$ (see Figure 4.1) or HA-$FLRT1_{LONG}$ with CD2. After being left to express, "unpatched" controls were fixed and stained with mouse and rabbit primary antibodies as indicated in Figure 5.2, followed by Alexa 488-coupled anti-mouse and Alexa 568-coupled anti-rabbit fluorescent antibodies (see Section 2.8.1). Meanwhile,

cells to be "patched" (i.e. probed for protein self-association at the cell surface) were left unfixed and incubated on ice with mouse primary antibodies as indicated in Figure 5.2. Under these conditions, only the appropriately tagged molecules will interact with the antibodies, which label single or, maximally, pairs of molecules as they are monoclonal. Cells were then treated with an Alexa 488-coupled anti-mouse secondary antibody at 37 °C. Incubation at a higher temperature allows proteins to diffuse in the plane of the cell membrane, and because the secondary antibody is polyclonal, it can recognise multiple epitopes on the primary antibodies, and thereby cross-link proteins labelled in the first incubation step to form patches on the cell membrane (indicated by white arrowheads in the left-hand column of Figure 5.2). After this "patching" procedure, cells were fixed and stained with rabbit primary antibodies against the non-patched protein, followed by an Alexa 568-coupled anti-rabbit antibody. Cells were imaged, and the images analysed, as described in Section 2.8.1.

If proteins associate homotypically at the cell surface, they would be expected to co-localise to the same patches in the cell membrane (i.e. the same staining pattern should be observed using both secondary antibodies, indicated by white arrowheads in the right-hand column of Figure 5.2) as the "non-patched" protein should be pulled into patches with the "patched" protein *via* the intermolecular interaction between them. This could also be indicated by a Pearson's correlation coefficient (see Section 2.8.1) value close to 1, showing a strong positive association between green and red fluorescent staining. Conversely, if the proteins do not associate, the "non-patched" protein will not be pulled into patches with the "patched" protein; the staining pattern will not be the same (as indicated by red arrowheads in Figure 5.2) and the Pearson's correlation coefficient would be expected to be closer to zero.



See overleaf for caption

Figure 5.2 Co-patching experiments to investigate whether FLRT1 associates homotypically at the cell surface. In unpatched experiments, proteins exhibit a relatively uniform pattern of expression across the cell surface. However, in patched experiments, FLRT1 was pulled into large puncta using mouse primary antibodies and an Alexa 488-coupled anti-mouse antibody (indicated by white arrowheads in left-hand column of (A)). In cells expressing FLAG-FLRT1$_{LONG}$ and HA-FLRT1$_{LONG}$, staining for the HA tag revealed co-localisation of the two FLRT proteins in the same membrane puncta (indicated by white arrowheads on the top row of (A) and in the corresponding close-up view in (B)). However, in cells expressing HA-FLRT1$_{LONG}$ and CD2, staining for CD2 remained uniform across the cell surface and was not co-localised in fluorescent puncta with HA-FLRT1$_{LONG}$ (indicated by red arrowheads). (B) shows enlarged views of merged images of cells treated as shown in (A), with average Pearson's correlation coefficient values from five cells. Images were taken by Jeff McIlhinney (MRC Anatomical Neuropharmacology Unit, Oxford). Scale bars represent 10 μm.

Results from these experiments, illustrated in Figure 5.2, indicate that FLRT1, which (like CD2) is expressed uniformly across the cell surface in unpatched controls, can be pulled into punctate patches on the cell surface using mouse primary antibodies against either HA or FLAG N-terminal tags, followed by anti-mouse fluorescent secondary antibodies. Whilst these doubly fluorescent puncta were most prominent at the cell periphery, similar patches were also observed in the membrane more generally. CD2, which FRET experiments indicate does not interact with FLRT proteins, fails to co-patch with these FLRT cell surface patches, and CD2 cell-surface staining remains non-punctate; the Pearson's correlation coefficient value is relatively low, at 0.50. However, when HA- and FLAG-tagged FLRT1$_{LONG}$ were co-expressed, staining of the HA-tagged version co-localised significantly with that of the patched FLAG-tagged construct, indicating that FLAG-FLRT1$_{LONG}$ molecules must associate with HA-FLRT1$_{LONG}$ molecules, in order to aggregate them into the same puncta in the cell membrane. The high degree of co-localisation is reflected in the higher positive value of the Pearson's correlation coefficient for this group of images: 0.84. These results therefore agree with those obtained using FRET, and indicate that

FLRT1 self-associates at the cell membrane, possibly as a homodimer as observed in solution.

5.3 Crystallisation and structural analysis of the FLRT1 LRR domain

5.3.1 Crystallisation and structure determination of FLRT1$_{LRR}$ in two crystal forms

Prior to my involvement in this project, attempts had been made to crystallise FLRT$_{ECD}$s (see Figure 4.4), though these had proved unsuccessful. This may be because of flexibility in the region linking the LRR domain to the FNIII domain, and in the region between the FNIII domain and the C-terminus of the construct, which might not favour long-range order in a protein crystal; these regions were confirmed to have a high likelihood of disorder using the Regional Order Neural Network (RONN) server (Yang et al., 2005). However, it was hoped that the absence of these linker regions in the FLRT$_{LRR}$ constructs (see Figure 4.4) might favour crystallisation of FLRT$_{LRR}$ proteins to enable structural study.

Having optimised expression and purification protocols for the three FLRT$_{LRR}$ constructs, a number of commercial sparse matrix crystallisation screens were used in attempts to crystallise them. Purified FLRT1$_{LRR}$ was crystallised (Figure 5.3) under two distinct sets of conditions as detailed in Table 2.2, and crystals were frozen by briefly immersing in mother liquor supplemented with 20 % glycerol as a cryoprotectant, before flash freezing as described in Section 2.4.3.

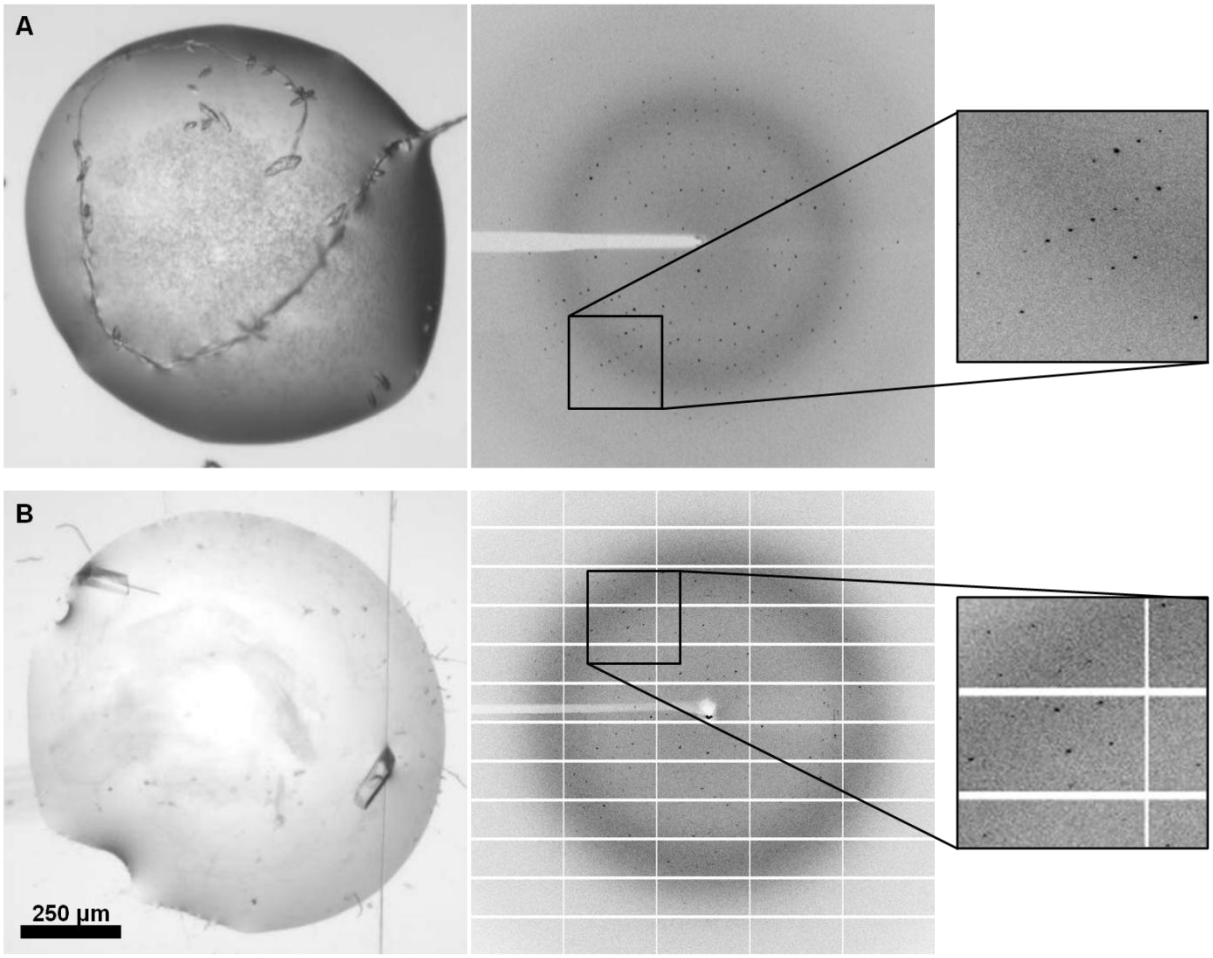


Figure 5.3 X-ray diffraction from two crystals of $FLRT1_{LRR}$, formed under distinct conditions. (A) Crystal 1, formed under acidic conditions (pH 4.6), diffracted to 1.96 Å. (B) Crystal 2, formed under more neutral conditions (pH 7.8), diffracted to 3.23 Å.

X-ray diffraction data were collected at the ESRF microfocus beamline ID23.2 (crystal 1; Figure 5.3 (A)) or at the Diamond Light Source microfocus beamline I24 (crystal 2; Figure 5.3 (B)). Data processing and structure solution proceeded as described in Section 2.5; statistics can be found in Table 5.2. To summarise, the 2.8 Å model of the third LRR domain of *Drosophila* Slit (Howitt et al., 2004) was used as a search model for molecular replacement to phase data from crystal 1 using Phaser (McCoy et al., 2007), which revealed a model consisting of a single copy of the FLRT1 LRR domain in the asymmetric unit. This was initially built using Phenix Autobuild (Terwilliger et al., 2008), and inspection of the electron density maps after

an initial round of restrained refinement using Phenix (Adams et al., 2010) indicated that much of the structure had been built correctly, but that some regions including the N- and C-termini were still to be modelled; these were added manually using Coot (Emsley and Cowtan, 2004). This was followed by iterative rounds of refinement using Phenix and model building using Coot. The quality of the final refined model, containing residues 53-379 from the mouse FLRT1 sequence (GenBank accession number NP_958813; see also Appendix B) as well as an addition N-terminal glycine residue from the pHLsec expression vector (Aricescu et al., 2006), was assessed using Coot and MolProbity (Davis et al., 2007), with statistics shown in Table 5.2. Discussion of all $FLRT_{LRR}$ structures in this chapter will use residue numbering corresponding to the sequence of mouse FLRT1 for consistency; residue numbering for each construct according to sequences in the GenBank database are listed in Appendix B.

When crystal 2 was subsequently obtained (Figure 5.3) and diffraction data collected (see Table 5.1), the model obtained from crystal 1 was successfully used as a search model for molecular replacement in Phaser (McCoy et al., 2007). The resultant model consists of two molecules in the asymmetric unit, and was refined as described above, with the inclusion of NCS restraints early in the refinement process as described in Section 2.5.4. The final refined model contains residues 54-369 (molecule A) and 57-360 (molecule B) from FLRT1; statistics obtained from the final round of refinement are listed in Table 5.2.

	Crystal 1 (acidic pH)	Crystal 2 (neutral pH)
Crystal data		
Crystallisation conditions	30 % w/v PEG monomethyl ether 2000 0.1M sodium acetate pH 4.6 0.2M ammonium sulphate	0.06M MES monohydrate 0.06M PIPES 0.33% w/v hexamminecobalt(III) chloride 0.02M HEPES pH6.8 Final pH7.8
Space group	*C2*	*C2*
Cell dimensions (Å)	*a* = 118.23, *b* = 52.94, *c* = 64.12 α = γ = 90°, β = 117.47°	*a* = 108.44, *b* = 52.82, *c* = 117.73 α = γ = 90°, β = 97.20°
Monomers per asymmetric unit	1	2
Solvent content (%)	46.7	43.2
Data collection		
X-ray source	ID23-2 (ESRF)	I24 (Diamond)
Image plate system	225 CCD (MAR)	Pilatus 600
Wavelength (Å)	0.8726	0.9778
All data		
Resolution (Å)*	33.13 - 1.96 (2.01 - 1.96)	51.37 - 3.23 (3.31 - 3.23)
No. reflections*	25310 (1865)	10750 (800)
R_{merge} (%)*	10.6 (65.3)	17.2 (66.3)
R_{pim} (%)*	6.4 (39.0)	13.6 (50.2)
$I/\sigma I$*	9.5 (2.2)	5.3 (1.9)
Completeness (%)*	99.7 (99.8)	99.0 (99.2)
Redundancy*	3.7 (3.8)	3.2 (3.3)

	Crystal 1 (acidic pH)	Crystal 2 (neutral pH)
Refinement		
R_{work}/R_{free} (%)	18.10/22.30	20.11/26.79
No. atoms protein	2601	4931
No. atoms glycan	28	0
No. atoms water	223	0
Average B factor main/side chain (Å^2)	19.9/28.2	40.6/48.1
r.m.s.d. bond lengths (Å)	0.010	0.010
r.m.s.d. bond angles (°)	1.10	1.22
Model quality (Ramachandran plot)		
Most favoured residues (%)	313 (96.0)	570 (92.5)
Additional allowed residues (%)	13 (4.0)	46 (7.5)
Disallowed residues (%)	0 (0.0)	0 (0.0)

Table 5.2 Crystallisation, data collection, refinement and model quality statistics for $FLRT1_{LRR}$ crystals 1 and 2; data was collected, processed and refined with the assistance of Maria Harkiolaki (Strubi). *Values in parentheses correspond to the highest resolution shell. Pertinent formulae can be found in Appendix C.

These figures indicate that the structural information obtained from crystal 1 is of higher quality than that obtained from crystal 2, though the crystal 2 values are still acceptable for producing a model of the protein structure. The higher level of resolution (1.96 Å compared to 3.23 Å) means that features such as side chain conformations may be resolved in the crystal 1 dataset, while these features are not clear in the electron density map from crystal 2. The greater number of reflections recorded from crystal 1 has led to a better degree of completeness of the data (99.7 % compared to 99.0 %) and a greater degree of redundancy i.e. the sampling of a single reflection multiple times (3.7 compared to 3.2). The R_{merge} and R_{pim} values, which measure agreement among multiple measurements of the same reflections (R_{pim} includes a term to correct for over-sampling; see Appendix C), are significantly higher for crystal 2 (R_{merge} is 17.2 % compared to 10.6 %), indicating the lower quality of the data. This is also reflected in $I/\sigma I$ (the signal-to-noise ratio), which falls below 2 in the highest resolution shell of crystal 2 (it is 2.2 for crystal 1); this value is lower than is ideal, making it difficult to distinguish meaningful high-resolution information from the noise. The higher quality of the crystal 1 data is also reflected in the refinement statistics (R_{work} and R_{free}, for the majority of the intensity data and a "test set" not used in refinement, respectively), which are slightly lower than those for crystal 2, indicating a greater degree of agreement between the refined model and original diffraction data. Although these values are higher for crystal 2 (20.11 % compared to 18.10 %), they are acceptable in the context of a relatively low resolution structure such as this.

5.3.2 Key features of the two FLRT1$_{LRR}$ crystal structures

Automated structural comparison against the Protein Data Bank (Dali server (Holm and Rosenström, 2010) reveals that in the models obtained from both crystals, the

LRR domain of FLRT1 assumes the canonical horseshoe solenoid shape observed in the structures of LRR domains from many other proteins (see Figure 1.8, and Figure 5.4). For example, the structure is very similar to those of both decorin and biglycan (Scott et al., 2006); alpha carbon root mean square deviation (C_α rmsd) values are 2.5 Å between $FLRT1_{LRR}$ (model from crystal 1) and decorin (PDB accession code 1XKU; 286 equivalent C_α positions), and 2.4 Å between $FLRT1_{LRR}$ and biglycan (PDB accession code 2FT3; 286 equivalent C_α positions). The solenoid comprises 10 LRR repeat units, which are each 21-26 residues in length and form a β-strand connected by loops to helical secondary structural elements as described in Section 1.5.3. Together, the β-strands from each repeat make up the parallel β-sheet which forms the concave face of the solenoid, while the helical elements form the convex face and the loop regions form the sides. As for other LRR structures, the LRR repeats are situated between N- and C-terminal cysteine-rich cap structures, each stabilised by two disulphide bonds. Structural superposition of monomers from the two models confirms that differences between the two are relatively subtle; the C_α rmsd over the complete structures (306 equivalent C_α positions) is 0.9 Å.

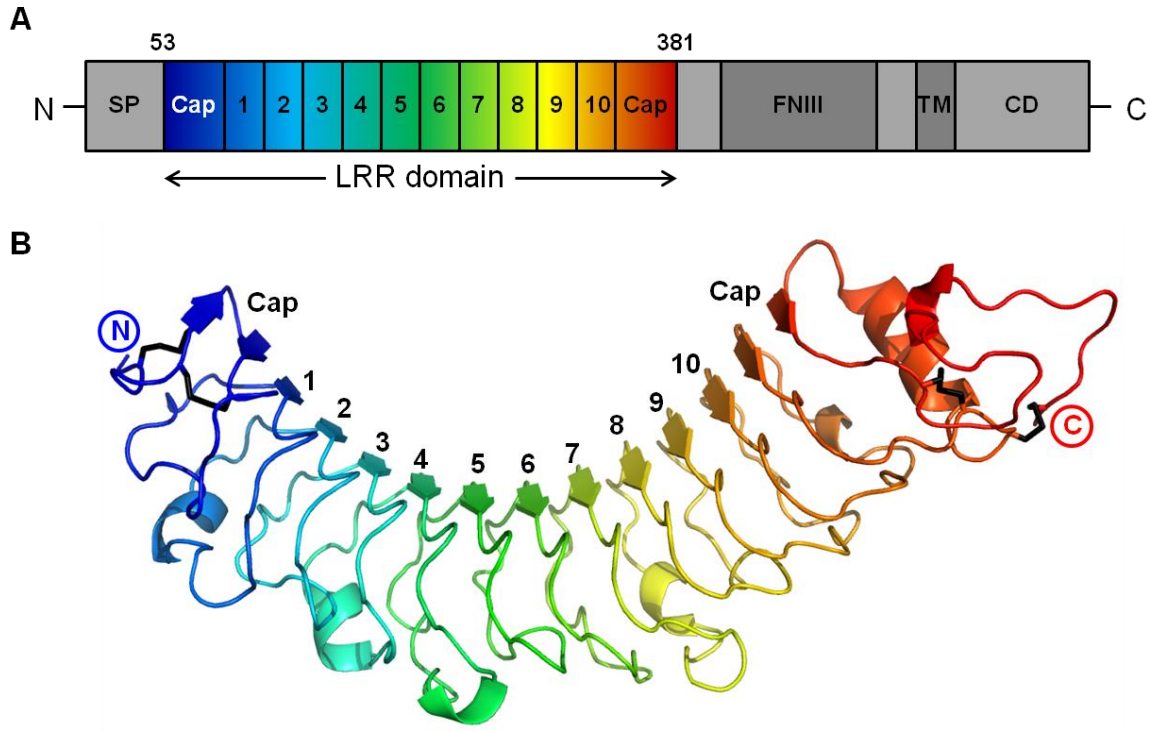


Figure 5.4 Crystal structure of the FLRT1 LRR domain (model from crystal 1). (A) Schematic representation of the structure of the $FLRT1_{LRR}$ construct in the context of the full-length protein. Residue numbers indicate the position of the FLRT1LRR construct in the full mouse FLRT1 sequence (Q14DT7). SP, signal peptide; Cap, N- and C-terminal cap regions; FNIII, fibronectin type III domain; TM, transmembrane region; CD, cytoplasmic domain. (B) Ribbon diagram of the structure of $FLRT1_{LRR}$ (residues T53-F379). LRR units are numbered 1-10; black lines denote disulphide bridges in the N- and C-terminal caps; N- and C-termini are indicated by encircled N and C.

An electrostatic surface representation of the crystal 1 model (Figure 5.5) reveals acidic and basic patches on the concave face and side of the structure; the convex face appears not to have such strongly-charged patches, and is particularly uncharged toward the C-terminus. The concave face also includes a patch of solvent-exposed hydrophobic, aromatic and histidine residues towards its N-terminus, indicated in Figure 5.5; some of these residues (Y87, Y112, H133 and H157) are further illustrated in a close-up view of the electron density map of this region. Because the $FLRT1_{LRR}$ protein was not subjected to deglycosylation prior to crystallisation, both asparagine residues predicted to be sites for *N*-linked

glycosylation (N249 and N305; see Figure 4.4) would be expected to be occupied by glycans. In fact, the first GlcNAc residues which join these glycans to N249 and N305 can be resolved at both sites in the crystal 1 model (see Figure 5.5), for which higher resolution data were collected.

The model from crystal 1 (crystallised under acidic conditions) has a single molecule in the asymmetric unit, though a dimer-type interface utilising the concave face of the LRR domain is generated using two-fold symmetry related molecules (Figure 5.6). As seen in previously-determined structures of LRR domains known to function as homodimers (Scott et al., 2004; Scott et al., 2006; Seiradake et al., 2009; Kajander et al., 2011), this dimer uses the concave face of both molecules, with a relatively large interface surface area of 1250 $Å^2$ (though surprisingly there is a marginally positive free energy change of 0.4 kcal/M upon interface formation, as calculated using the PISA server (Krissinel and Henrick, 2007); the next largest calculated interface is 480 $Å^2$). Additionally, in common with these published structures, the crystal 1 $FLRT1_{LRR}$ dimer has the C-termini of the two $FLRT1_{LRR}$ molecules pointing in opposite directions, so that the arrangement may be termed "antiparallel". The molecules approach each other most closely at their termini, and the N-terminal patch of aromatic residues (Figure 5.5) interfaces with C-terminal residues from the other molecule (Figure 5.6).

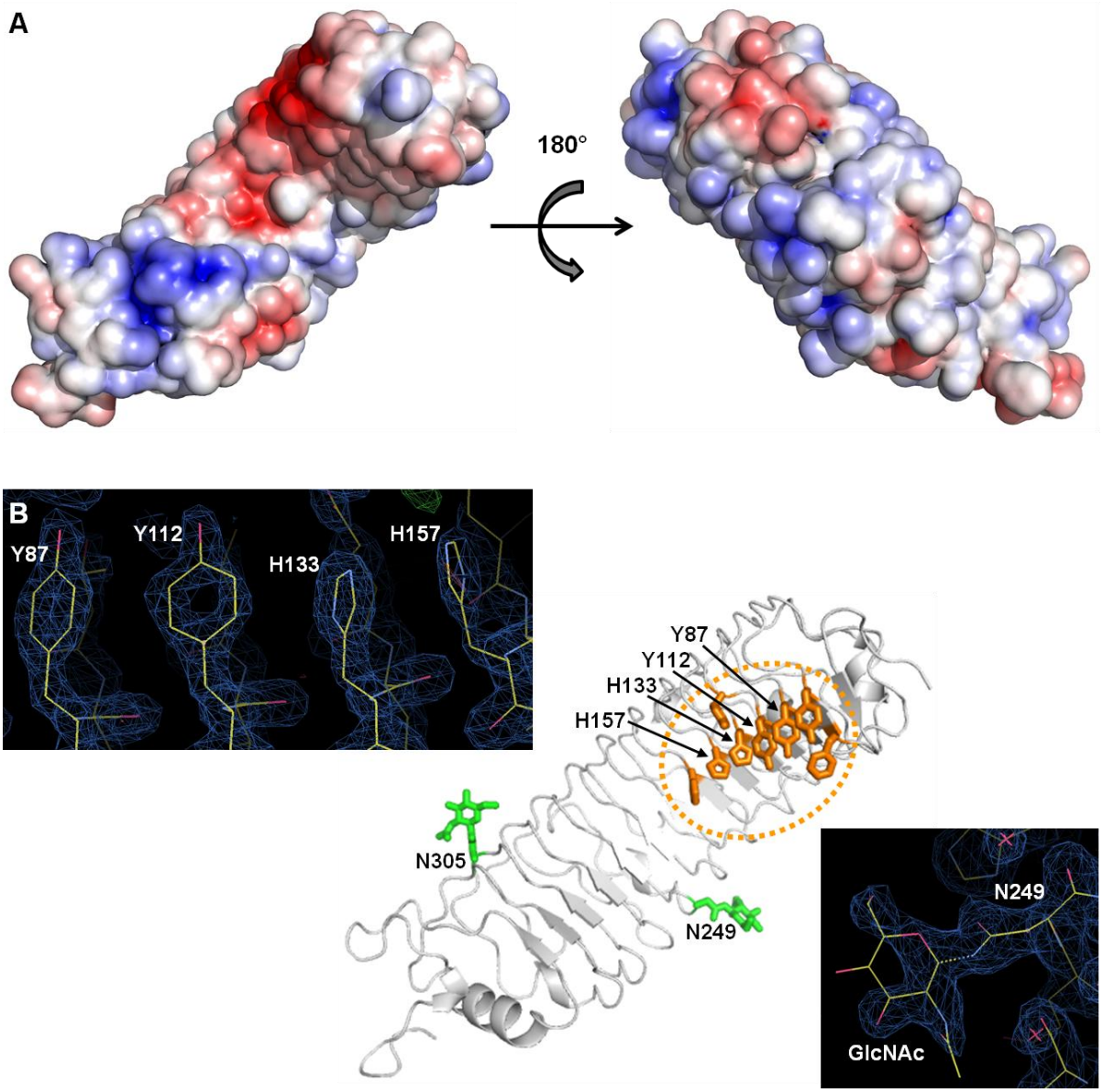


Figure 5.5 Surface properties of the $FLRT1_{LRR}$ monomer (model from crystal 1). (A) Solvent accessible surface representation of the concave face (left panel) and convex face (right panel) of $FLRT1_{LRR}$ coloured by electrostatic potential contoured at ±5 e/kT (red, acidic; blue, basic). (B) View of the concave face highlighting the patch of aromatic residues (in orange, ringed by an orange dashed line); selected residues are also shown in a close-up view of the electron density map (upper-left panel; $2F_o$-F_c map (blue) contoured at 2 σ). Asparagine residues predicted to be sites for *N*-linked glycosylation are highlighted in green, with the first GlcNAc residue shown; electron density for this sugar residue could be observed at both these sites (lower-right panel; $2F_o$-F_c map (blue) contoured at 1 σ; pink crosses indicate the presence of a solvent water molecule).

Unlike that from crystal 1, the model from crystal 2 (crystallised under neutral-like conditions) has two molecules within the asymmetric unit. Like the dimer-type arrangement from crystal 1 described above, these two molecules interface *via* their concave faces, with an interface surface area of 1470 $Å^2$ (Figure 5.6). This is associated with a calculated free energy change of 16.2 kcal/M upon interface formation, indicating that this interface may be less energetically favourable than the crystal 1 interface; the next largest calculated interface for crystal 2 is 750 $Å^2$. However, this arrangement is different to that seen in crystal 1 and elsewhere in that the C-termini of the two molecules point in the same direction; this novel homodimer arrangement is therefore referred to as "parallel". Because the N-termini are situated close together, the aromatic patches of the two molecules are in close proximity, though they interface most closely with another group of residues, as shown in Figure 5.6.

A structure solved from another crystal which formed under the same crystallisation conditions as crystal 2 also showed FLRT1 LRR domains in this parallel arrangement, but due to pathologies within the collected data, this model proved challenging to refine to completion. Following successful refinement of the model from crystal 2 (Table 5.2), this was therefore used to illustrate the parallel mode of $FLRT1_{LRR}$ homodimerisation observed in these crystals.

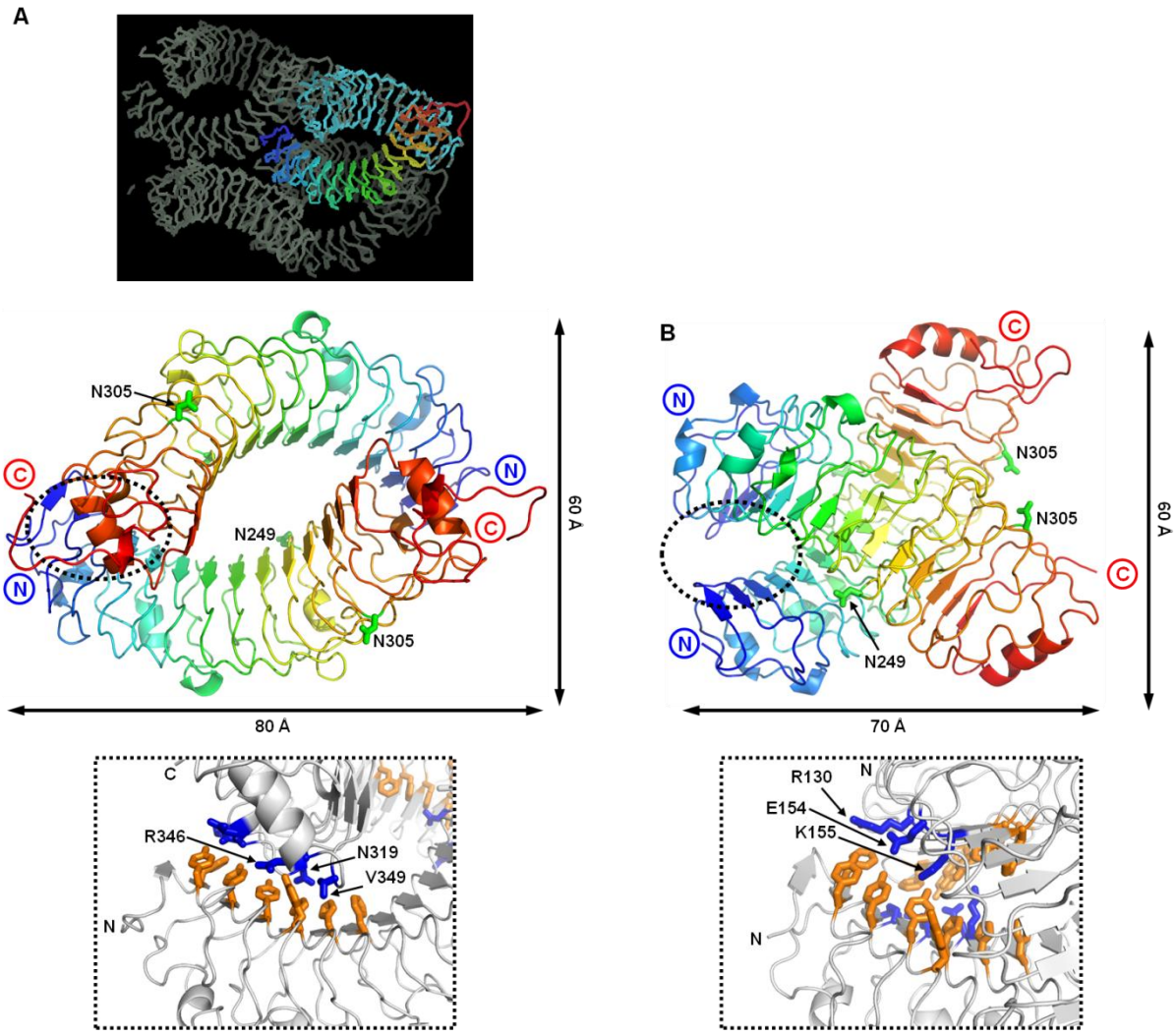


Figure 5.6 Two alternative models for FLRT1$_{LRR}$ homodimerisation, based on crystal structures. (A) The model from crystal 1 only includes a single monomer in the asymmetric unit, but a dimer model is created by superposition of the monomer model on a molecule related by crystallographic symmetry (upper panel). The monomer model is highlighted with colouring as in the upper panel; symmetry related molecules are shown in grey, with one of these representing the asymmetric dimer highlighted in blue. A ribbon diagram of the dimer model is shown in the middle panel, with approximate dimensions. Asparagine residues predicted to be sites for *N*-linked glycosylation are highlighted in green; N- and C-termini are indicated by encircled N and C; regions where the aromatic patch contributes to the dimerisation interface are circled with a dashed line, with a close-up view shown in the lower panel. Aromatic patch residues shown in orange, interfacing residues shown in blue. (B) Ribbon diagram of the dimer model from crystal 2, which includes two molecules in the asymmetric unit (upper panel). A close-up view of the interface region involving the aromatic patch is also shown (lower panel). Colouring and labelling as for (A).

5.3.3 Structure-led mutagenesis

The presence of dimer-type arrangements of $FLRT1_{LRR}$ in crystal structures is in agreement with data from biophysical analyses which suggest that $FLRT1_{LRR}$ homodimerises in solution (see Section 4.4). It is possible that one (or both) of these structures represents the arrangement of the solution dimer form, and/or the arrangement of LRR domains as they mediate FLRT1 homotypic interactions at the cell surface; such interactions may in turn underlie the observed accumulation of FLRT1 at interfaces between transfected cells (see Section 4.2.2). To confirm whether homotypic interactions at the cell surface are indeed dependent on formation of a dimer between LRR domains as shown in these crystal structures, the models were used to guide design of a mutant form of $FLRT1_{LONG}$ in which dimerisation would be disrupted. Such a mutant, which should be designed to disrupt both the antiparallel (crystal 1) and parallel (crystal 2) models of homodimerisation within a single construct, could be studied using FRET in Cos-7 cells as described in Section 5.2.1. As the wild type FLRT1 protein was observed to give rise to a significant FRET signal (Figure 5.1), failure of the mutant protein to interact in the same way could indicate that this behaviour is dependent upon FLRT1 homodimerisation *via* the LRR domain. The mutant construct could also be tested in other experimental systems e.g. to study protein accumulation at cell-cell interfaces in HEK 293 cells (see Section 4.2.2).

Both structures were analysed to find residues which, if mutated, could be expected to abrogate formation of both dimer forms observed in the crystal structures (Figure 5.7). In the antiparallel model from crystal 1, a group of residues from the N-terminal aromatic patch appear to interact with an asparagine residue (Asn319; see also Figure 5.6). This residue is followed by leucine and alanine residues in the

wild type FLRT1 sequence, but if the alanine is mutated to a serine, then an Asn-Leu-Ser motif is formed, which would be predicted to make the asparagine a site for *N*-glycosylation (confirmed by NetNGlyc 1.0 server (Gupta et al., in preparation)). Addition of a bulky glycan adduct to Asn319 would preclude such close association of the C-terminus of the $FLRT1_{LRR}$ molecule with the N-terminal aromatic patch of another molecule, especially His133, which forms a hydrogen bond with Asn319 (PISA server (Krissinel and Henrick, 2007)). Additionally, because of the symmetry of the model, this would occur at both ends of the antiparallel dimer, effectively preventing formation of such an arrangement.

Unlike the antiparallel arrangement, the interface between molecules in the parallel dimer includes residues in the centre of the concave face, such as a pair of arginine residues (Arg225 and Arg226; Figure 5.7), which form salt bridge and hydrogen bond interactions with residues Asp69 and Glu115 towards the N-terminus of the other chain (PISA server (Krissinel and Henrick, 2007)). Reversal of the positive electrostatic charge of these arginine residues (at the crystallisation condition of pH 7.8) by mutation to a pair of negatively-charged glutamate residues would be expected to destroy these salt bridge interactions and create a region of localised electrostatic repulsion, which could prevent formation of the parallel dimer.

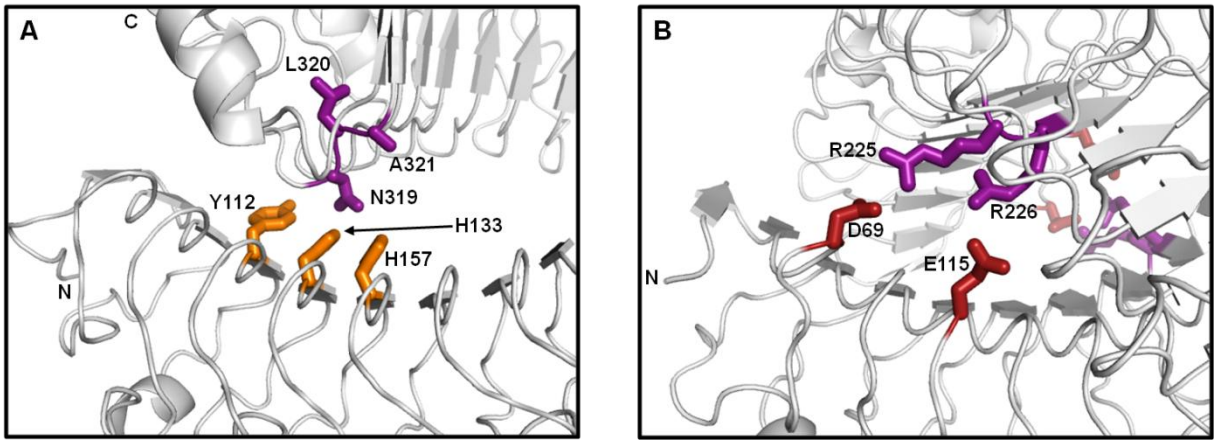


Figure 5.7 Residues selected for site-directed mutagenesis (within a single construct) to prevent dimer formation in the antiparallel (A) and parallel (B) arrangements observed in crystal structures. (A) Asn319 approaches residues in the N-terminal aromatic patch closely, but mutation of Ala321 to serine would create an *N*-glycosylation site. Addition of a glycan here would prevent formation of this arrangement. Residues from the aromatic patch are shown in orange, with residues that would form the novel *N*-glycosylation motif shown in purple; positions of molecule N- and C-termini are indicated. (B) Arg225 and Arg226 form salt bridge interactions with Asp69 and Glu115 respectively. Mutation of the arginine pair to glutamate residues would destroy the salt bridges and lead to electrostatic repulsion between the two molecules in this region. Residues to be mutated shown in purple, with residues interacting *via* salt bridges shown in red.

The Ala321-Ser, Arg225-Glu and Arg226-Glu mutations were all created within single FLRT1$_{LONG}$-mVenus and FLRT1$_{LONG}$-mCerulean constructs (as described in Section 2.1.2; primers shown in Appendix B) in preparation for FRET analysis in combination with wild type FLRT1$_{LONG}$ constructs with the other fluorescent tag. However, inspection of transfected Cos-7 cells (transfected as described for HEK 293 cells in Section 2.2.1) revealed that unlike wild type protein, the mutant protein did not appear to be expressed uniformly across the cell surface (Figure 5.8; see also Section 4.2.1). Instead, fluorescence from the mutant protein appears to be restricted to the endoplasmic reticulum network inside the cell, and does not appear to be present at the cell surface. This precludes FRET analysis, which requires expression of both constructs bearing the FRET pair of fluorophores at the cell

surface to ensure that they are able to interact; it was not therefore possible to test the dependence of FLRT1 FRET on LRR-mediated dimerisation as planned.

However, this result may indicate that FLRT1 dimerisation, mediated by the LRR domain as indicated in the crystal structures, may be essential for correct cell surface expression. It is clear that mutant FLRT1 is expressed, as the fluorescence from the C-terminal tag can be visualised, but it appears that unlike wild type FLRT1, which is expressed uniformly across the cell surface, the mutant form is either unable to reach the cell surface, or its expression there is much reduced. Retention in the endoplasmic reticulum and failure to reach the cell surface may indicate that the mutant protein is not correctly folded, which may in turn be due to the failure of the protein to form homodimers. LRR domain-mediated dimerisation has been shown to be essential for native folding and stability in decorin and biglycan (Scott et al., 2006), and prevention of dimer formation has been shown to preclude correct folding and block cell surface expression in AMIGO-1 (Kajander et al., 2011). These results may indicate that, in common with these other proteins, FLRT1 exists as an obligate homodimer in a cellular context, with dimerisation mediated by the LRR domain.

Prior to the solution of crystal structures of $FLRT2_{LRR}$ and $FLRT3_{LRR}$ (see Section 5.4), mutant forms of these proteins were also designed based on the $FLRT1_{LRR}$ mutant described above, to test the effect of these mutations in the other FLRT proteins. Creation of an *N*-glycosylation site to disrupt the antiparallel $FLRT1_{LRR}$ dimer was replicated in the FLRT2 construct by mutation of Lys321 (numbering corresponding to FLRT1 sequence) to serine, to create an NLS motif. In FLRT3, N319 is already a predicted glycosylation site in an Asn-Ile-Thr motif in the wild type sequence (see Figure 4.4; this is N296 in the FLRT3 sequence), so that mutation to create a glycosylation site here was unnecessary. The salt bridge

stabilising the $FLRT1_{LRR}$ parallel dimer arrangement was destroyed in FLRT2 by mutating Arg226 to glutamate to create a pair of glutamate residues; in FLRT3, the same was accomplished by mutating both Lys225 and Arg226 to glutamate. Upon examination of Cos-7 cells transfected with constructs of $FLRT2_{LONG}$-mVenus and $FLRT3_{LONG}$-mVenus containing these mutations, the same pattern was seen as for the FLRT1 mutant (Figure 5.8). Once again, the proteins appear to be expressed, though cell surface expression is blocked and the mutant proteins are retained in the endoplasmic reticulum. This was a preliminary indication, in the absence of structural data, that dimerisation contacts are conserved between the three FLRTs, and that all three are obligate dimers in a cellular context.

As constructs containing mutations designed to disrupt *both* antiparallel and parallel models of $FLRT1_{LRR}$ homodimerisation within a single construct appear to prevent correct protein trafficking to the cell surface, further mutations were designed to disrupt *either* the antiparallel or the antiparallel model. Assuming that formation of at least one of these dimers will permit cell surface expression, it was hoped that this approach would allow investigation of which dimer arrangement, if either, might represent a physiological interaction in cells that could underlie behaviour such as accumulation of FLRT protein at interfaces between HEK 293 cells (Figure 4.2). For example, if constructs containing mutations designed to disrupt only the antiparallel dimerisation model accumulate at cell-cell interface like wild type proteins, while constructs with mutations designed to disrupt the parallel remained uniformly distributed over the cell surface, this could indicate that the parallel model (from crystal 2) is likely to be relevant in the cellular context.

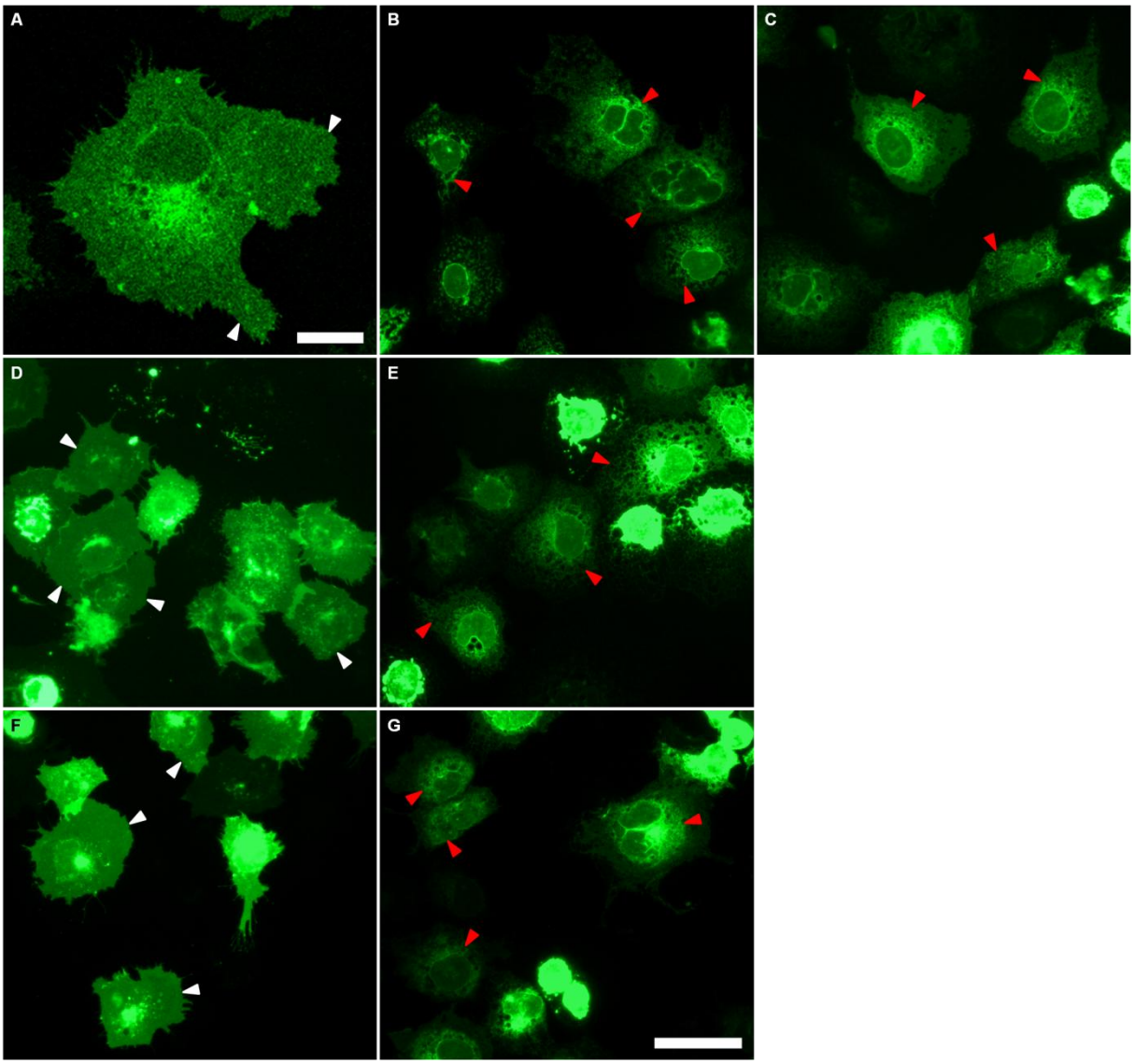


Figure 5.8 Expression patterns of wild type and mutant FLRTs in Cos-7 cells. Cells express wild type $FLRT1_{LONG}$-mVenus (A), mutant $FLRT1_{LONG}$-mVenus (B), $FLRT1_{LONG}$-mVenus containing only the R225E and R226E mutations (C), wild type $FLRT2_{LONG}$-mVenus (D), mutant $FLRT2_{LONG}$-mVenus (based on mutant FLRT1) (E), wild type $FLRT3_{LONG}$-mVenus (F) and mutant $FLRT3_{LONG}$-mVenus (based on mutant FLRT1) (G). Expression of wild type FLRT proteins (A, D, F) is distributed uniformly across the cell surface (indicated by white arrowheads; see also Figure 4.1), while mutant proteins (B, C, E, G) appear to be retained in the endoplasmic reticulum, as shown by a restricted, web-like pattern of fluorescence within the cell (red arrowheads). Scale bar represents 20 µm (A) or 50 µm (B-G).

The PISA server (Krissinel and Henrick, 2007) was used to identify candidate residues for mutation, which would affect one model of dimerisation but not the other (Figure 5.9). Both mutations from the overall dimerisation knockout mutant described above (Ala321-Ser and Arg225E, Arg226E; see also Figure 5.7) were candidates, and were cloned into wild type $FLRT1_{LONG}$-mVenus separately to create two novel constructs. The asparagine residue that is predicted to be glycosylated in the Ala321-Ser mutant (Asn319) was also mutated to a leucine to destroy hydrogen bond interactions with residues on the other chain in the antiparallel (crystal 1) model. In another construct, two residues (Arg346 and Val349) that form salt bridge and Van der Waals' interactions with the other molecule were mutated to reverse the electrostatic charge and substitute a bulkier side chain respectively (Arg346-Glu, Val349-Leu), that should not be accommodated in the antiparallel dimer model. Finally, Thr84 was mutated to a much bulkier tryptophan, which should prevent formation of the parallel (crystal 2) dimer.

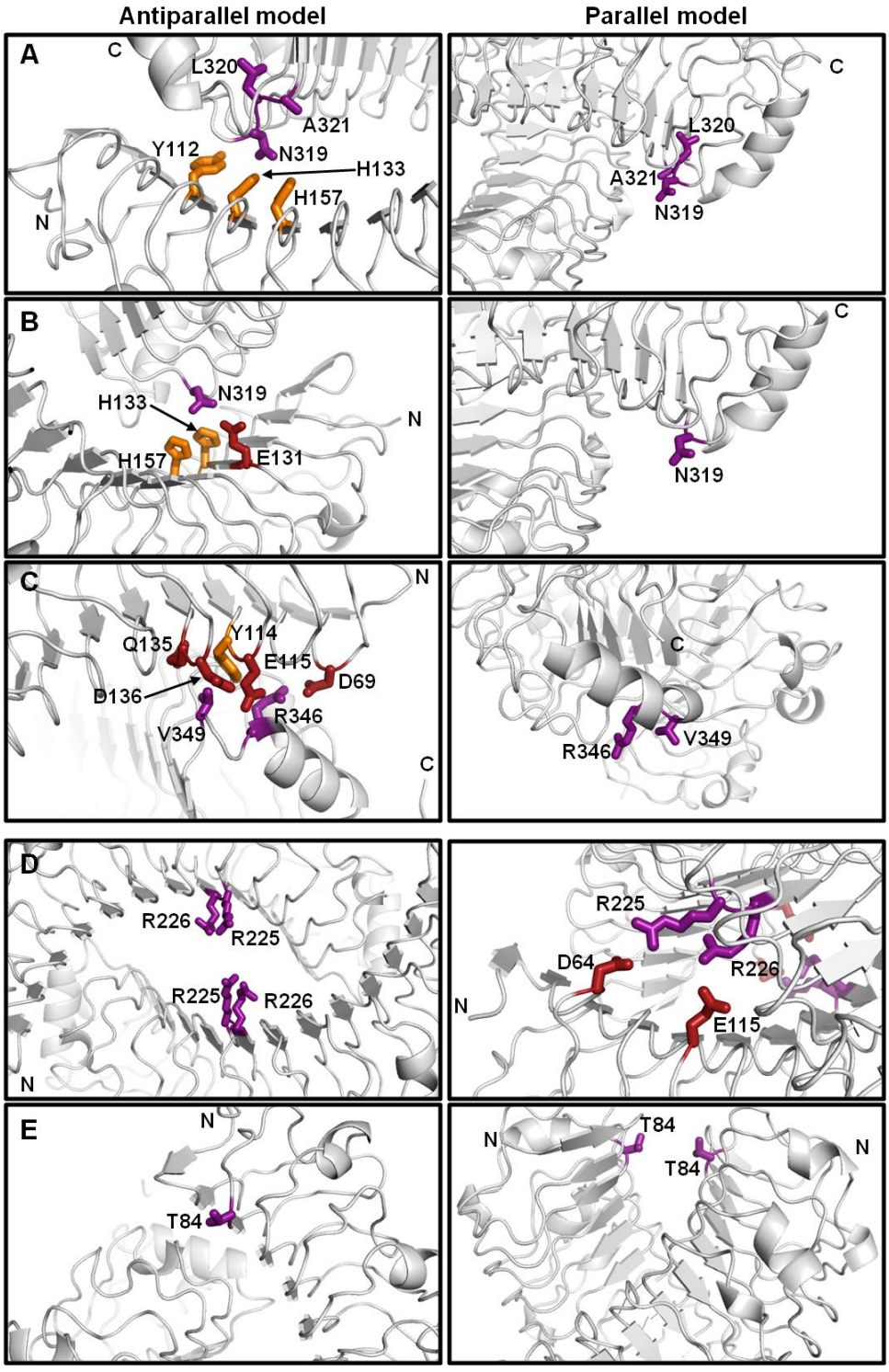


Figure 5.9 Further FLRT1 mutations designed to selectively disrupt either the antiparallel (crystal 1; A, B and C) or parallel (crystal 2; D and E) dimer models. A321S (A), N319L (B) and R346E/V349L (C) were designed to disrupt the antiparallel model (left-hand panels) while not affecting the parallel model (right-hand panels). R225E/R226E (D) and T84W (E) were designed to preclude dimer formation according to the parallel model (right-hand panels) while not affecting the antiparallel model (left-hand panels).

These mutations were cloned into the $FLRT1_{LONG}$-mVenus construct as described in Section 2.1.2 in preparation for testing their behaviour in HEK 293 cells relative to the wild type protein, which accumulates at cell-cell interfaces (Section 4.2.2). However, expression in Cos-7 cells revealed that, like the double dimerisation knockout mutants described above, selective mutants were retained within the endoplasmic reticulum, and were not expressed uniformly at the cell surface; this is illustrated by the Arg225-Glu/Arg226-Glu mutant in Figure 5.8 (C). Unfortunately, the failure of the mutant constructs to be expressed at the cell surface means that they are not suitable for assays which rely on cell surface expression, such as studying their behaviour at cell-cell interfaces in HEK 293 cells, or using FRET to assess cell surface molecular interactions in Cos-7 cells. Therefore, it was not possible to use these mutants to draw any further conclusions regarding which model of dimerisation observed in the crystal structures is most likely to correspond to the arrangement found in cells.

5.3.4 Implications for the association state of FLRT1 in the secretory pathway and at the cell surface

The two crystal structures obtained for $FLRT1_{LRR}$ appear to confirm results from AUC and MALS (see Section 4.4) that suggest the soluble LRR domain is capable of homodimerisation, and may mediate homodimerisation of FLRT1. However, it remains unclear which of these models underlies physiological FLRT homotypic associations, due to at least partial failure of mutants, designed to selectively probe the interfaces of the parallel and antiparallel models, to express at the cell surface. It is also not possible to discern which model most closely corresponds to the $FLRT1_{LRR}$ dimer observed using biophysical methods due to the similarity in dimensions of the dimer models, despite the differences in the arrangement of the monomers (see Figure 5.6).

It is possible that both dimeric models might be relevant, and that different arrangements are adopted at different stages of the secretory pathway, according to the different conditions the protein encounters (Figure 5.10). Results from AUC indicate that pH affects homotypic interactions between soluble $FLRT1_{LRR}$ molecules (Figure 4.10), and pH-dependence of dimer formation has also been demonstrated for other cell surface adhesion molecules (e.g. receptor protein tyrosine phosphatase μ (Aricescu et al., 2007)), so it is possible that the variation in the pH between different cellular compartments may affect the FLRT1 association state. This could be dependent upon the group of histidine residues towards the N-terminus of the concave face; histidine side chains have a pKa value of approximately 6.0 (although this varies depending on the microenvironment of the residue within the protein), and so will be protonated at the lower pH of crystal 1, but deprotonated at the higher pH of crystal 2. The local electrostatic environment of this patch of residues will therefore vary considerably between the two conditions, and it may be that this affects self-association under different pH conditions. It appears that mutants designed to preclude formation of either dimer are retained in the endoplasmic reticulum, which may mean that FLRT1 is an obligate homodimer, and cannot be properly folded and expressed at the cell surface in monomeric form. The low pH conditions (pH 4.6) under which the antiparallel dimer form of $FLRT1_{LRR}$ was crystallised approximate those found within the secretory pathway, so it is possible that prior to reaching the cell surface, FLRT1 associates *via* an antiparallel interaction between the LRR domains, using protonated histidine residues from the N-terminal patch (Figure 5.7).

However, FRET data indicates that at the cell surface, the C-termini of a pair of FLRT1 molecules (which in the FRET assay are attached to the FRET pair of

fluorophores) are situated in close proximity. Due to the flexibility of linker regions between the LRR and FNIII domains, and between the FNIII domain and the transmembrane region (confirmed using the RONN server (Yang et al., 2005)), this could indicate the presence of parallel and/or antiparallel dimers at the cell surface (Figure 5.10). The parallel model was observed in a crystal formed under pH conditions similar to those found at the cell surface (pH 7.8). It is possible, therefore, that as it reaches the cell surface and the pH increases, FLRT1 might switch between dimer arrangements, to form a parallel *cis* dimer; however, the presence of a cell surface antiparallel *cis* dimer cannot be discounted. The possibility of such switching behaviour is supported by the relatively low affinity (Kd ≈ 100 μM; Figure 4.9) calculated using MALS data, as well as the flexibility of interdomain regions of the protein.

In addition to possible pH-dependent switching between dimerisation arrangements as the protein approaches the cell surface, it is also conceivable that homotypic interactions between FLRT1 molecules differ between protein found on the surface of isolated cells, and protein found at cell-cell interfaces, where FLRT1 is known to accumulate (Figure 4.2). This behaviour, as well as the observation that FLRT1-transfected HEK 293 cells sort out from non-transfected cells (Karaulanov et al., 2006), indicates the presence of FLRT1-mediated *trans* interactions between cells. One possibility is that the antiparallel dimer form represents such a *trans* dimer, with the two FLRT molecules originating from the opposing cells at a cell-cell interface (Figure 5.10). This would require dissociation of the cell surface *cis* dimer (either parallel or antiparallel arrangement) upon the approach of another cell expressing FLRT1, followed by formation of an antiparallel *trans* dimer at the cell-cell interface; this has been suggested for AMIGO-1, which is also expressed as an

obligate homodimer, and dimerises *via* its LRR domain (Figure 1.7 (Kajander et al., 2011)). Another possibility is that the *trans* interaction observed between cells occurs between two *cis* dimers presented by the opposing cells. However, analysis of the crystal structures indicates that this type of interaction would be mediated by the convex or side faces of the $FLRT1_{LRR}$ molecules, with a smaller interaction surface area, so would be likely to be significantly weaker than that mediated by a single antiparallel *trans* dimer, which utilises a close interaction between the concave faces of two LRR domains.

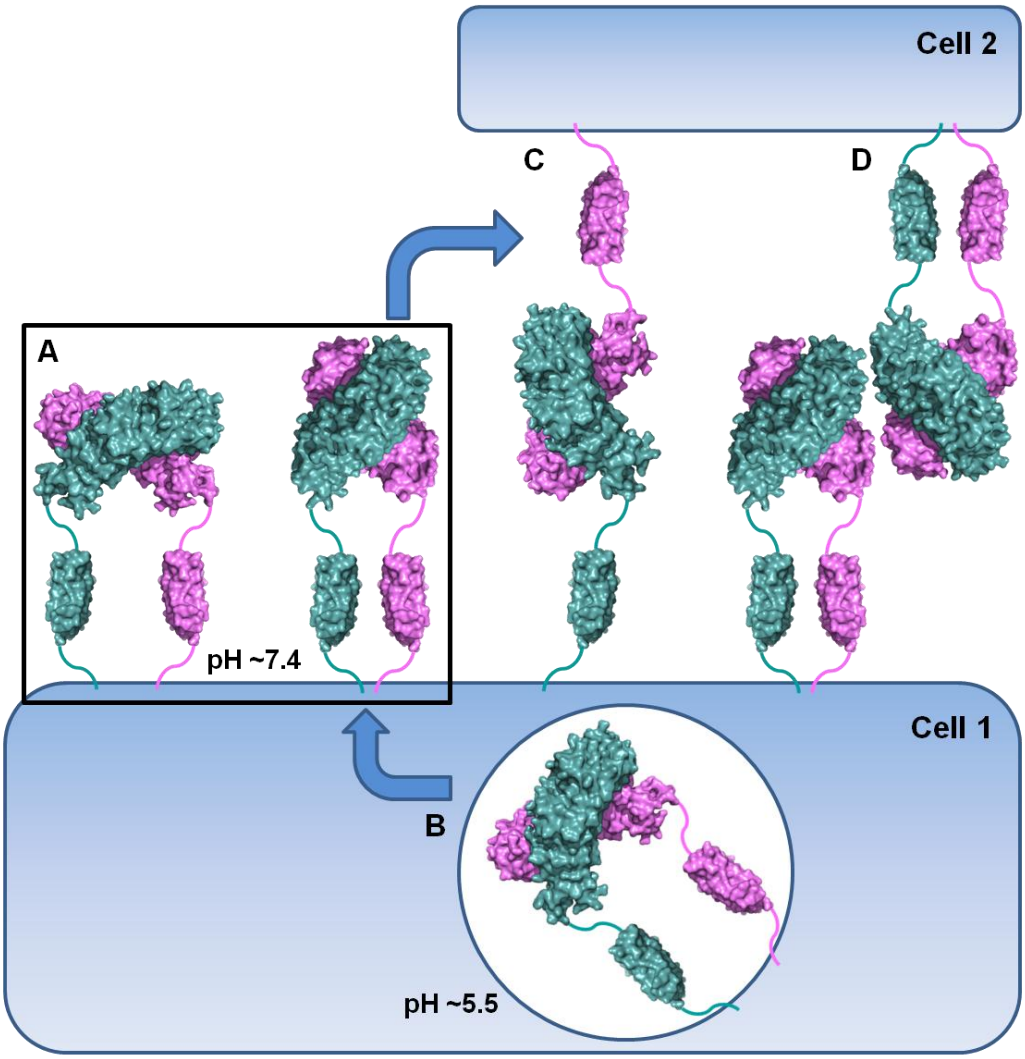


Figure 5.10 A model representing ways in which FLRT1 might self-associate within the secretory pathway (B), and at isolated (A) and opposing (C, D) cell surfaces. (A) Two possible configurations of a FLRT1 *cis* dimer at the cell surface, based on antiparallel (crystal 1, left) and parallel (crystal 2, right) models of LRR dimerisation; both possibilities are consistent with FRET results, which were measured on non-interfacing cell surfaces. The two arrangements, observed in crystals formed under distinct pH conditions, may also represent the association states of FLRT1 at different stages of the secretory pathway (B); at the more acidic pH found in intracellular secretory compartments, the antiparallel form (crystallised at pH 4.6) might predominate, with the parallel form (crystallised at pH 7.8) found at the cell surface. A further possibility is that the approach of another FLRT1-expressing cell might induce switching between *cis* (A) and antiparallel *trans* (C) cell surface dimer arrangements. Alternatively the *trans* interaction between cells could be mediated by interactions between *cis* dimer units (D). The FNIII domain is modelled by 3TES.pdb, with flexible interdomain regions modelled by curved lines.

5.4 Crystallisation and structural analysis of the LRR domains of FLRT2 and FLRT3

5.4.1 Crystallisation and structure determination of $FLRT2_{LRR}$ and $FLRT3_{LRR}$

$FLRT2_{LRR}$ and $FLRT3_{LRR}$ proteins were produced in HEK 293S cells and purified by IMAC followed by size exclusion chromatography (see Section 2.2.5). Pooled fractions were supplemented with non-detergent sulphobetaine 256 to prevent protein loss during concentration, and concentrated to 6 mg/ml ($FLRT2_{LRR}$) and 2 mg/ml ($FLRT3_{LRR}$). At this stage, $FLRT2_{LRR}$ was combined with purified Unc5D protein in a 1:1 molar ratio, in an attempt to form a complex which could be captured by crystallisation (Yamagishi et al., 2011). Prior to crystallisation, recombinant endoglycosidase F1 (Grueninger-Leitch et al., 1996; Chang et al., 2007) was added, to remove any *N*-glycans, at a concentration of 10 µg/ml.

Both proteins crystallised at 20.5 °C (Figure 5.11), in conditions listed in Table 5.3; although $FLRT2_{LRR}$ had been mixed with Unc5D prior to crystallisation, only $FLRT2_{LRR}$ was present in the needle cluster-type crystals. Crystals were flash frozen using 20 % ($FLRT2_{LRR}$) or 25 % ($FLRT3_{LRR}$) glycerol as a cryoprotectant, in 75-80 % reservoir solution. X-ray diffraction data were collected at Diamond Light Source beamlines I24 ($FLRT2_{LRR}$) and I03 ($FLRT3_{LRR}$), and processed using XDS (Kabsch, 2010), xia2 (Winter, 2010) and programmes from the CCP4 suite (Collaborative Computational Project 4, 1994).

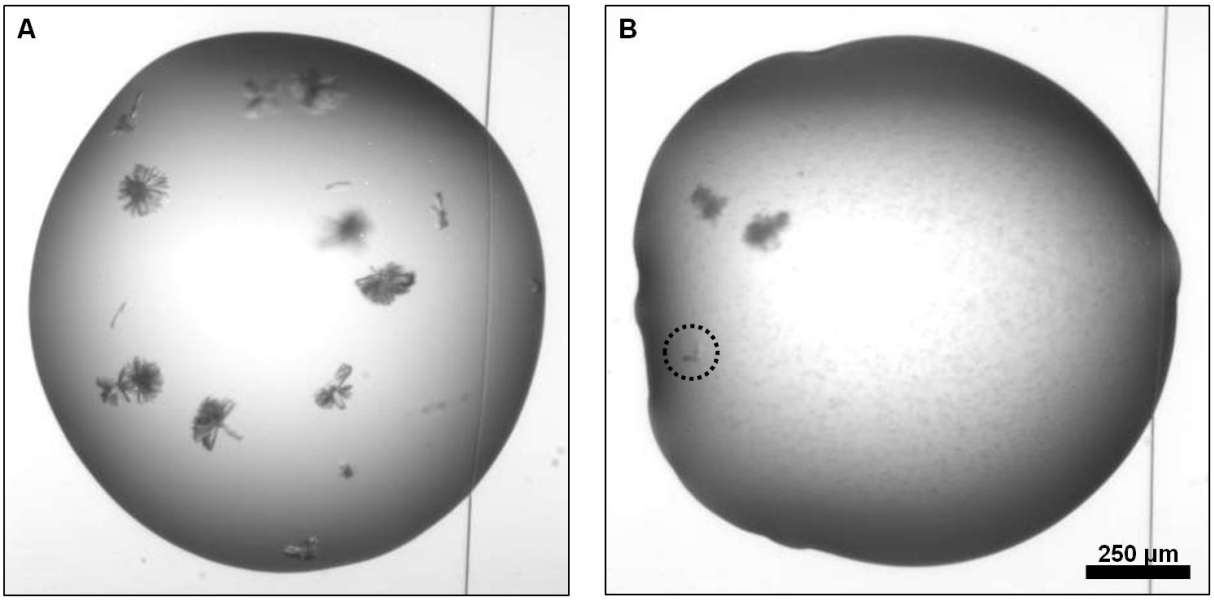


Figure 5.11 Crystals of (A) $FLRT2_{LRR}$ (crystallised by Elena Seiradake (Strubi) in the presence of Unc5D) and (B) $FLRT3_{LRR}$ (highlighted by the dashed black circle).

Statistics are listed in Table 5.3. These show that data for both $FLRT2_{LRR}$ and $FLRT3_{LRR}$ structures is closer in quality to crystal 2 of $FLRT1_{LRR}$ than to the high resolution crystal 1. The resolution is 3.0 Å for both structures, meaning higher resolution information such as side chain conformations cannot be discerned, and the relatively low number of reflections recorded resulted in low completeness values (completeness is only 88.6 % for $FLRT3_{LRR}$). The $FLRT3_{LRR}$ R_{merge} (and R_{meas}, another multiplicity-weighted version of R_{merge}) values are strikingly high (R_{merge} is 46.6 % across all data) indicating poor reliability of the data, although the $I/\sigma I$ value is more acceptable at 2.64 for the highest resolution shell.

The $FLRT1_{LRR}$ model from crystal 1 (Section 5.3.2) was used as the initial search model for molecular replacement to phase the $FLRT3_{LRR}$ dataset, using Phaser (McCoy et al., 2007). The Rosetta implementation in Phenix (DiMaio et al., 2009; DiMaio et al., 2011) and manual adjustment in Coot (Emsley and Cowtan, 2004) were used to complete the building of the two copies of the LRR domain within the

asymmetric unit, with iterative rounds of refinement using autoBuster (Blanc et al., 2004); refinement statistics are included in Table 5.3 and show that the $FLRT2_{LRR}$ and $FLRT3_{LRR}$ models both have a lower-than-ideal degree of agreement with the collected data (R_{work} values are 26.1 and 28.7 % respectively), perhaps due in part to the paucity of data originally collected. The final model was then used for molecular replacement of the $FLRT2_{LRR}$ dataset in Phaser, with details of building and refinement as for $FLRT3_{LRR}$. MolProbity (Davis et al., 2007) was used to assess the quality of both models.

The final model of $FLRT2_{LRR}$ contains only one molecule within the asymmetric unit, and includes residues corresponding to 54-378 from the FLRT1 sequence (36-360 of the mouse FLRT2 sequence (GenBank accession number NP_958926)). However, DNA sequencing revealed that a base substitution had caused a point mutation (Glu370-Ala); as it is situated in the disulphide bond-stabilised C-terminal cap region, and is not involved in the largest intermolecular interface (see below), it is not thought that this would have a significant effect on $FLRT2_{LRR}$ structure or homotypic association. The final refined model of $FLRT3_{LRR}$ contains residues corresponding to 53-373 (molecule A) and 54-373 (molecule B) from the FLRT1 sequence (30-350 and 31-350 of the mouse FLRT3 sequence (GenBank accession number NP_001165631), respectively).

	$FLRT2_{LRR}$	$FLRT3_{LRR}$
Crystal data		
Crystallisation conditions	19% Peg 4000 19% 2-propanol 5% glycerol 0.095 M tri-sodium citrate pH 5.6 20 .5 °C	20% Peg 3350 0.2 M Ammonium Nitrate Final pH 7.4 20.5 °C
Space group	$P2_1$	$P2_1$
Cell dimensions (Å)	a = 45.07, b = 41.41, c = 87.27 α = γ = 90°, β = 102.89°	a = 59.81, b = 61.97, c = 92.97 α = γ = 90°, β = 106.38°
Monomers per asymmetric unit	1	2
Solvent content (%)	43.2	45.9
Data collection		
X-ray source	I24 (Diamond Light Source)	I03 (Diamond Light Source)
Image plate system	Pilatus 600	Pilatus 6M-F
All data		
Resolution (Å)*	43.9 – 3.0 (3.15 – 3.0)	29.9 - 3.0 (3.3 - 3.0)
No. reflections*	6374 (1678)	11741 (582)
R_{meas} (%)*	17.1 (70.3)	37.8 (71.4)
R_{merge} (%)*	23.1 (90.8)	46.6 (82.7)
$I/\sigma I$*	6.93 (1.98)	5.26 (2.64)
Completeness (%)*	98.2 (96.1)	88.6 (74.7)
Redundancy*	3.3 (2.8)	3.1 (2.8)

	$FLRT2_{LRR}$	$FLRT3_{LRR}$
Refinement		
R_{work}/R_{free} (%)	26.1/28.0	28.7/31.6
No. atoms protein	2581	5112
No. atoms glycan	0	0
No. atoms water	0	0
Average B factor ($Å^2$)	55.06	44.39
r.m.s.d. bond lengths (Å)	0.008	0.008
r.m.s.d. bond angles (°)	0.96	0.98
Model quality (Ramachandran plot)		
Most favoured residues (%)	302 (93.5)	598 (93.9)
Additional allowed residues (%)	21 (6.5)	38 (6.0)
Disallowed residues (%)	0 (0.0)	1 (0.2)

Table 5.3 Crystallisation, data collection and refinement statistics for $FLRT2_{LRR}$ and $FLRT3_{LRR}$ crystals; data collection and refinement were performed by Elena Seiradake (Strubi). *Values in parentheses correspond to the highest resolution shell. Pertinent formulae can be

5.4.2 Key features of the $FLRT2_{LRR}$ and $FLRT3_{LRR}$ crystal structures

As expected, given the high degree of sequence conservation between the $FLRT_{LRR}$ constructs (57 % between $FLRT1_{LRR}$ and $FLRT2_{LRR}$; 68 % between $FLRT1_{LRR}$ and $FLRT3_{LRR}$; 58 % between $FLRT2_{LRR}$ and $FLRT3_{LRR}$), the structures obtained for $FLRT2_{LRR}$ and $FLRT3_{LRR}$ are very similar to those obtained for $FLRT1_{LRR}$ (Figure 5.12; see also Section 5.3.2). Over the complete structures, comparing each model to that from $FLRT1_{LRR}$ crystal 1, the C_α rmsd values are 1.8 Å ($FLRT2_{LRR}$; 319 equivalent C_α positions) and 1.5/1.6 Å ($FLRT3_{LRR}$, molecules A (319 equivalent C_α positions) and B (316 equivalent C_α positions) respectively). Comparison with the Protein Data Bank using the Dali server (Holm and Rosenström, 2010) reveals that, like $FLRT1_{LRR}$, the $FLRT2_{LRR}$ and $FLRT3_{LRR}$ structures are highly similar to those of decorin and biglycan (Scott et al., 2006), as well as other proteins containing LRR domains.

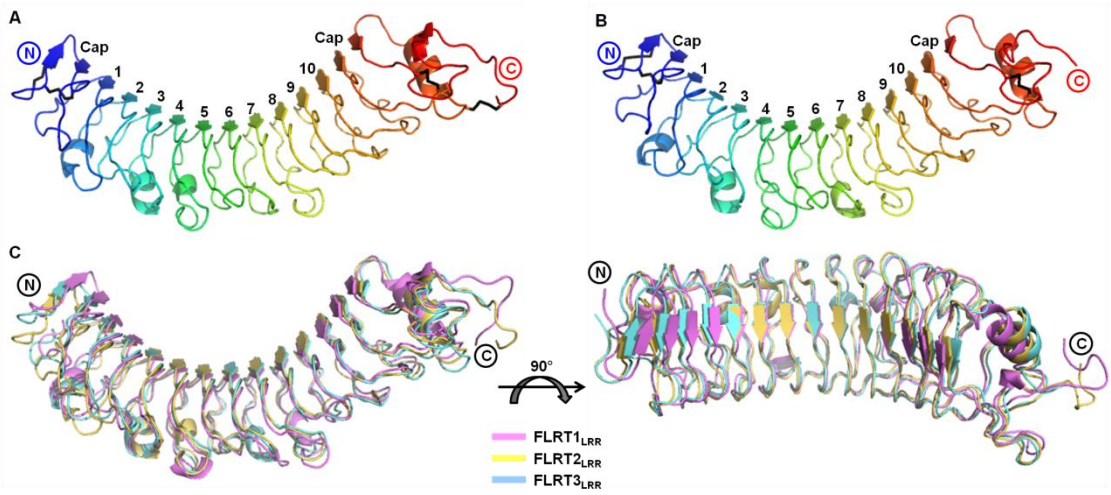


Figure 5.12 Crystal structures of the FLRT2 and FLRT3 LRR domains (latter using molecule A). Ribbon diagrams of the structures of (A) $FLRT2_{LRR}$ (residues C54-P378 (residue numbering from FLRT1 sequence)) and (B) $FLRT3_{LRR}$ (residues S53-A373) indicate the structural similarity of the three FLRT LRR domains. LRR units are numbered 1-10; black lines denote disulphide bridges in the N- and C-terminal caps; N- and C-termini are indicated by encircled N and C. (C) Structural alignment of the three structures, confirming the similarity between them.

Analysis of the surface properties of the structures shows that many features are conserved from the $FLRT1_{LRR}$ structure (Figure 5.13). The acidic and basic patches at the N- and C-termini, respectively, of the concave face are found in all three structures, though the acidic patch appears to be less intensely charged than in $FLRT1_{LRR}$. On the convex face, patterns of electrostatic charge appear to be broadly similar; however, an acidic patch seen only in $FLRT2_{LRR}$ is substituted by a basic patch in $FLRT3_{LRR}$. The aromatic patch at the N-terminal end of the concave face is entirely conserved. Although these proteins were deglycosylated prior to crystallisation, consideration of the positions of predicted *N*-glycosylation sites may be informative when considering possible association states (see below). Two such sites are conserved from FLRT1 to FLRT3, with the addition of one further site, while $FLRT2_{LRR}$ contains a single predicted site not found in the other two constructs; all sites are positioned on the side faces of the solenoid.

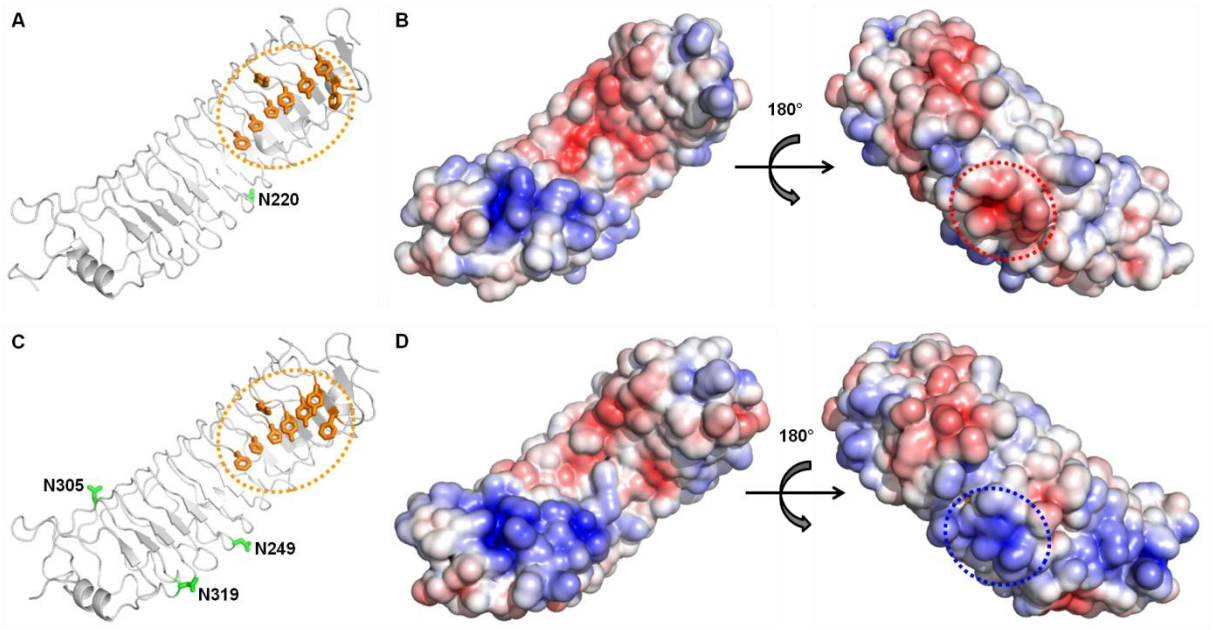


Figure 5.13 Surface characteristics of the $FLRT2_{LRR}$ (A, B) and $FLRT3_{LRR}$ (C, D; model from molecule A) crystal structures. Views of the concave face of the LRR solenoid (A, C) show that the N-terminal aromatic patch (in orange, ringed by an orange dashed line) is entirely conserved. Asparagine residues predicted to be sites for *N*-linked glycosylation are highlighted in green, with all positioned on the two side faces. Solvent accessible surface representations of the concave face (left panel) and convex face (right panel) of $FLRT2_{LRR}$ (B) and $FLRT3_{LRR}$ (D) coloured by electrostatic potential are contoured at ±5 e/kT (red, acidic; blue, basic). A patch of residues on the concave face that is acidic in $FLRT2_{LRR}$ but basic in $FLRT3_{LRR}$ is highlighted by dashed circles.

The model of $FLRT2_{LRR}$ contains a single molecule in the asymmetric unit, though as for the crystal 1 model of $FLRT1_{LRR}$, dimer-type interfaces can be generated using symmetry-related molecules (Figure 5.14). However, in this case, there were two possible interfaces with similar surface areas (calculated using the PISA server (Krissinel and Henrick, 2007)), both distinct from those discussed above for $FLRT1_{LRR}$. The first (interface surface area 820 $Å^2$, with calculated free energy change of 3.3 kcal/M upon interface formation) uses the concave face of only one molecule, interacting with the C-terminus of the other, while the second (770 $Å^2$; with a favourable calculated free energy change of -9.7 kcal/M upon interface formation) utilises the convex faces of one molecule and the C-terminus of the other; the next

largest interface after these has a calculated surface area of 210 $Å^2$. Although these arrangements do not correspond with those of $FLRT1_{LRR}$, similar arrangements to the first dimer have been observed in crystallographic dimers of other LRR domain-containing proteins, including the Nogo-66 receptor (PDB accession code 1P8T (Barton et al., 2003)). The conserved patch of aromatic residues is implicated in the interface of the first arrangement (Figure 5.14) and forms hydrogen bond and Van der Waals' interactions with the C-terminus of the other molecule.

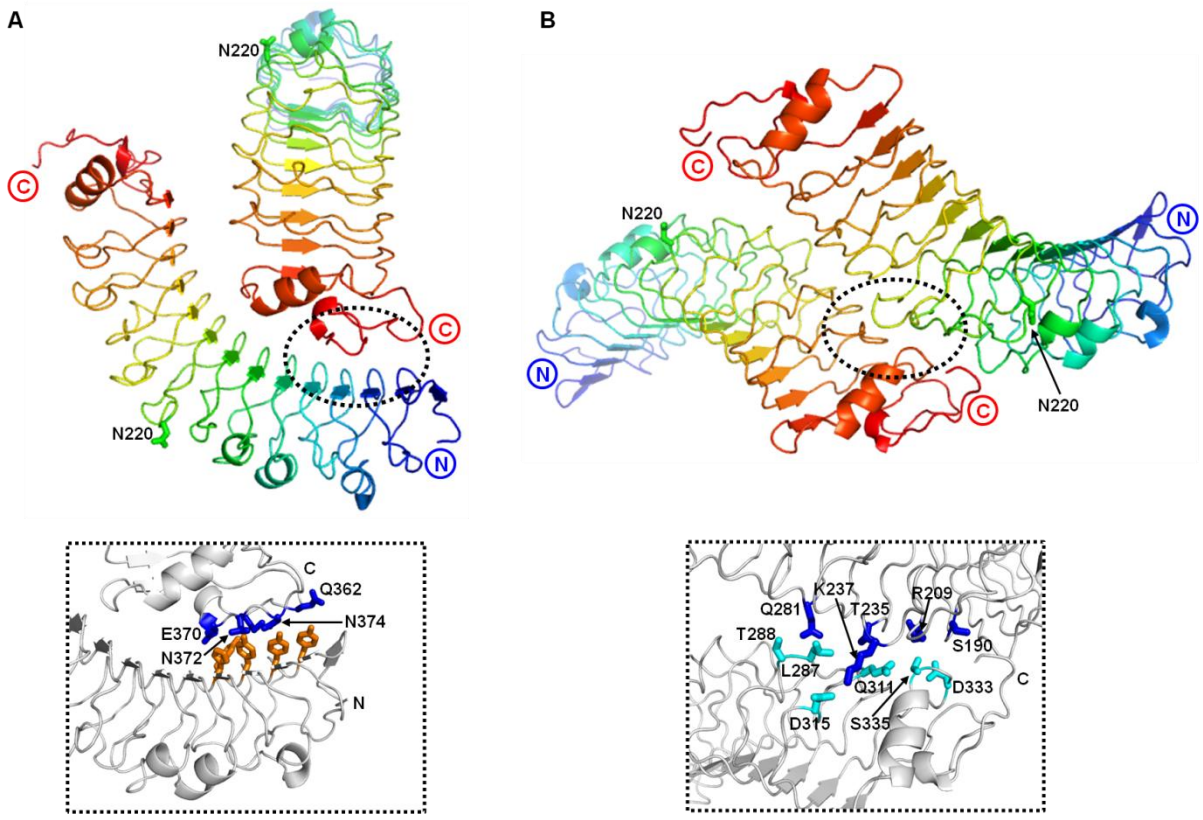


Figure 5.14 Crystallographic dimers of $FLRT2_{LRR}$. (A) An arrangement employing the concave face of one molecule (including the conserved aromatic patch) and C-terminus of the other (top panel), with a close-up view of the interface region (bottom panel) highlighted by a black dashed circle. The predicted *N*-glycosylation site is labelled and highlighted in green, with N- and C-termini also labelled. The close up view includes a selection of closely interfacing residues, with residues from the aromatic patch shown in orange, and other residues shown in blue. (B) A second arrangement where the molecules interface using the convex face and C-terminus. In the close-up view, interfacing residues from one molecule are coloured in dark blue, with residues from the other molecule that interact with these *via* hydrogen bonds and salt bridges shown in light blue. Colouring and labelling otherwise as for (A).

The model of $FLRT3_{LRR}$ includes two molecules within the asymmetric unit, though unlike the crystal 2 model of $FLRT1_{LRR}$, these do not associate *via* the concave faces of both, but rather utilise the concave face of one molecule and the convex face of the other (Figure 5.15) in an antiparallel arrangement, with an interface surface area of 1190 $Å^2$ (the calculated free energy change upon interface formation is surprisingly unfavourable given the interface size, at 9.8 kcal/M). Analysis using the PISA server

(Krissinel and Henrick, 2007) also revealed three other plausible interfaces using either the side faces (interface surface areas of 1380 Å^2 and 710 Å^2, with associated free energy changes of 7.4 and 5.9 kcal/M respectively) or an alternative combination of the concave and convex faces (690 Å^2, with an associated free energy change of 4.8 kcal/M upon interface formation; the next largest interface after this is 410 Å^2). All of these are illustrated in Figure 5.15, but it appears from the positions of the three predicted *N*-glycosylation sites that, in the glycosylated protein, bulky glycan adducts would prevent the formation of those interfaces using the side faces of the LRR domains, so that these are less likely to represent the arrangement at the cell surface.

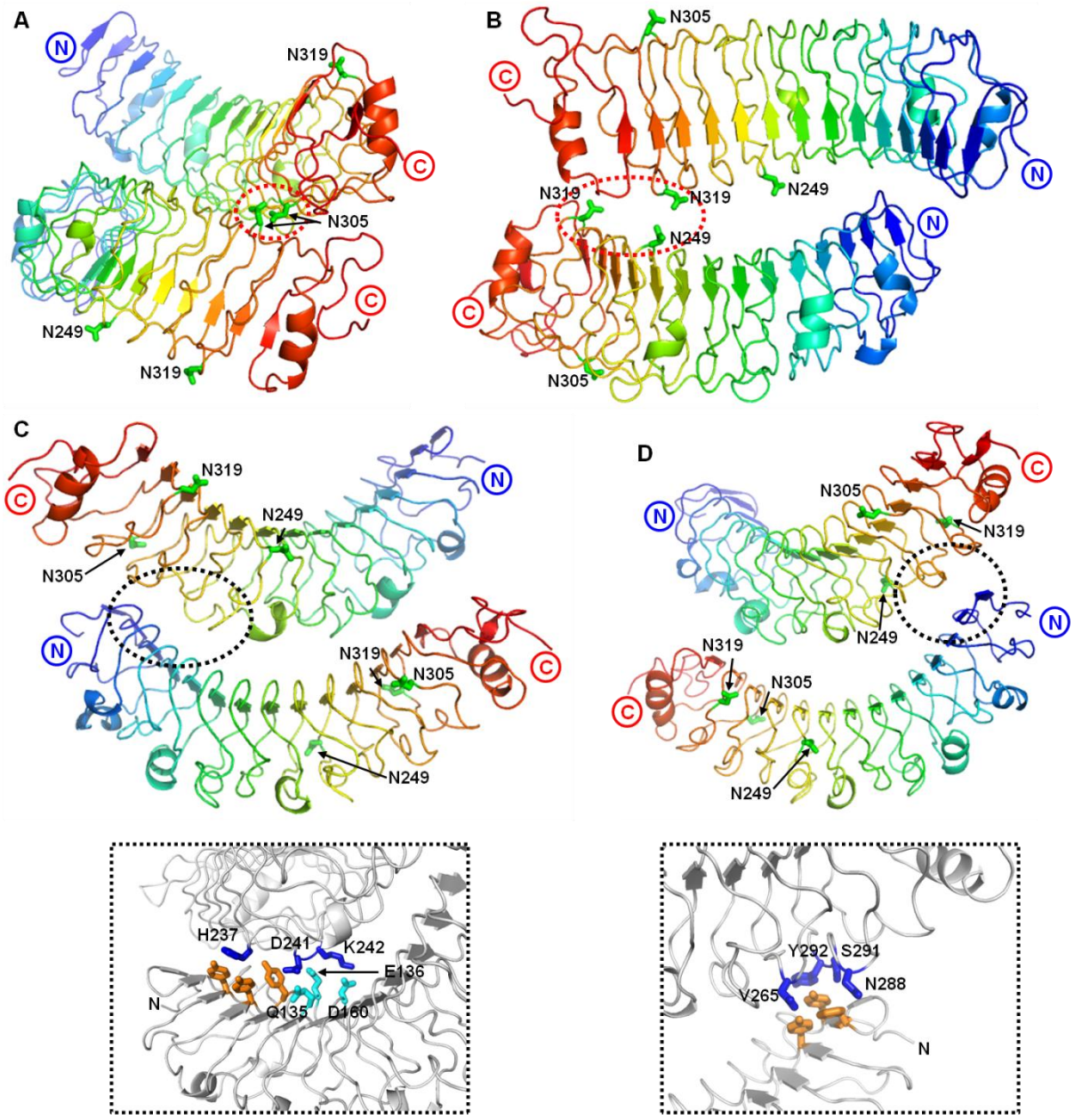


Figure 5.15 Possible dimeric arrangements for FLRT3$_{LRR}$. Two arrangements (A, B) utilise the side faces of both molecules, but because of the positions of the predicted *N*-glycosylation sites (highlighted by red dashed circles), are unlikely to represent physiological associations. Two others (C, D) use the concave face of one molecule, and the convex face of the other, and the predicted glycosylation sites would not interfere with these. Close-up views of selected areas where the molecules interface (highlighted by a black dashed circle) are shown in the lower panel, with residues from the aromatic patch shown in orange, other residues from the same molecule shown in light blue, and interfacing residues shown in dark blue.

The putative dimer structures were also analysed to discern the locations of mutations made based on the structure of $FLRT1_{LRR}$ (Section 5.3.3); these mutations were made to abrogate the $FLRT1_{LRR}$ dimer interactions, but resulted in retention of all three $FLRT_{LONG}$-mVenus proteins in the endoplasmic reticulum and a failure in surface expression, indicating that they might be expressed at the cell surface as obligate homodimers. To investigate what effect these mutations may have had on the various possible dimer arrangements of $FLRT2_{LRR}$ and $FLRT3_{LRR}$ described above, the corresponding residues were identified for each arrangement (Figure 5.16). The mutations would be predicted to have a significant effect on the $FLRT2_{LRR}$ dimer arrangement using the concave face of one molecule and the C-terminus of another, with the asparagine residue that is used to create a novel *N*-glycosylation site forming a hydrogen bond with Arg326, and the Arg226 that is mutated to glutamate forming a hydrogen bond with Tyr345. The other dimer, which does not use the concave face of either molecule, does not include the mutated residues in its interface, and therefore might be expected not to be affected by the mutations, unless they have a gross effect on the native fold of the protein. This could indicate that the first arrangement is more likely to be physiological, given the effect of the mutations.

However, examination of the $FLRT3_{LRR}$ structures (Figure 5.16) does not yield an obvious candidate for the physiological arrangement in the same way. Three of the structures, including the two which utilise different interfaces between the concave and convex faces of the two molecules, do not include the mutated region, or existing glycosylation site, at points of contact between the monomers, and therefore might not be expected to be affected by the mutations. The fourth arrangement, utilising the side faces, has the mutated residues approaching each other quite closely,

though because Asn319 is already predicted to be a site of *N*-glycosylation in this construct, the molecules would be unlikely to be able to form this arrangement if the protein was glycosylated. It is not, therefore, clear from this analysis why the mutations affect the cell surface expression of $FLRT3_{LONG}$-mVenus as well as the other two FLRTs.

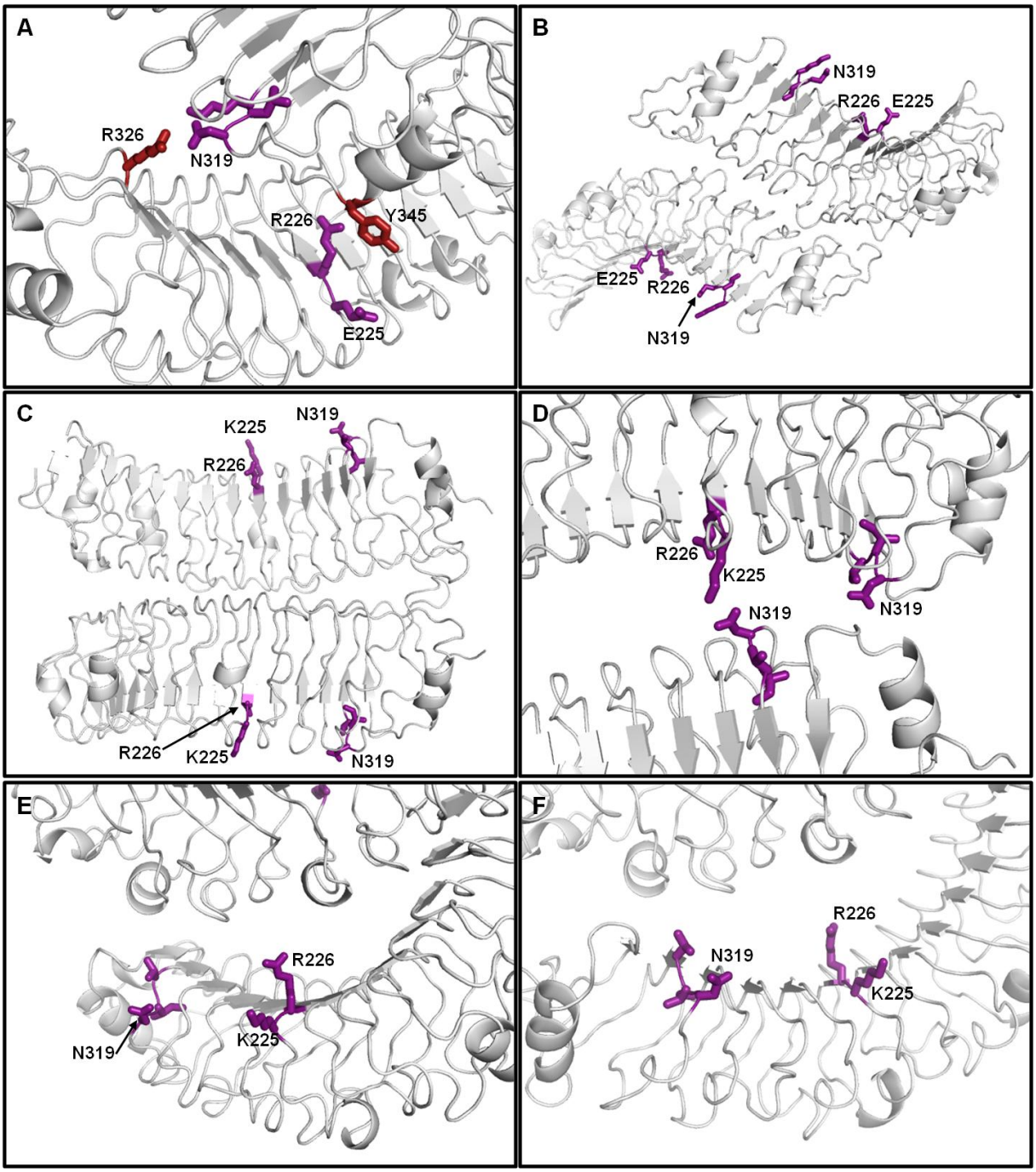


Figure 5.16 Examination of the positions of residues mutated based on the $FLRT1_{LRR}$ structure, which prevent proper cell surface expression of all three FLRT proteins. The arrangement of $FLRT2_{LRR}$ which includes the concave face in its interface (A) reveals that these mutations would be likely to affect this interface and may disrupt the dimer; however, the other interface (B) is distant from the mutations and is less likely to be affected. The same is true for one side-to-side (C) and both concave-convex arrangements (E, F) of $FLRT3_{LRR}$; the other side-to-side arrangement (D) includes the mutated residues in its interface, though because of the pre-existing *N*-glycosylation site, would be unlikely to form.

5.4.3 Possible implications for the association states of FLRT2 and FLRT3 at the cell surface

The involvement of the LRR domain in *trans* interactions between FLRT-transfected cells was first observed for FLRT3 (Karaulanov et al., 2006), and the crystal structures (as well as other data from biophysical analyses (Section 4.4)) appear to confirm that this domain is capable of forming significant intermolecular interactions in both FLRT3 and FLRT2. Both of these proteins have also been shown to induce clumping of transfected Sf9 cells (Figure 4.3), further confirming their role in forming *trans* interactions between cells. It is, however, somewhat surprising that in models of possible dimers derived from the crystal structures, neither of the LRR domains from these proteins associate in a similar way to that of FLRT1 and many other LRR domain-containing proteins by using the concave faces of both molecules. Given that there are multiple models for each protein, it is challenging to conclusively identify which, if any, most closely resembles the arrangements adopted at the cell surface - indeed, it may be that the biological arrangements are not represented in these crystal structures, especially if different arrangements are formed under different conditions, as I have suggested for FLRT1 (Figure 5.10).

However, it is possible that results presented here may give some indications about the likelihood of each arrangement. As for $FLRT1_{LRR}$, it appears from mutants designed to prevent dimerisation using the concave faces of both molecules (according to models of $FLRT1_{LRR}$ association, prior to the solution of $FLRT2_{LRR}$ and $FLRT3_{LRR}$ structures) that destruction of the dimerisation interface prevents correct cell surface expression (Figure 5.8), indicating that FLRTs form homodimers in the secretory pathway. Structural analysis of the location of these mutations in the FLRT2 models (Figure 5.16) indicates that an association using the concave face and C-terminus of the two molecules would be more likely to be disrupted by these

mutations than an alternative model using the convex face and C-terminus, and may therefore represent an arrangement adopted by the FLRT2 LRR domains in the secretory pathway (FLRT2 was crystallised at 5.6). Analysis of the mutated residues in $FLRT3_{LRR}$ does not identify a clear candidate arrangement that would be maximally disrupted in the mutant protein (Figure 5.16); as $FLRT3_{LRR}$ was crystallised at pH 7.4, the crystallographic dimers may not include arrangements that would only occur in the acidic conditions of the secretory system. However, two side-to-side arrangements can be discounted on the basis that, in the cell, *N*-glycosylation at three sites, all situated on the side faces of the molecule, would preclude formation of these arrangements as the molecules would be unable to approach each other closely enough (Figure 5.15).

As for $FLRT1_{LRR}$, FRET analysis indicates that FLRT2 and FLRT3 form homotypic associations in *cis* at the cell surface; for a significant FRET signal, the C-terminal fluorophore tags must be within 1-10 nm. However, because of the flexibility of interdomain linkers (between the LRR and FNIII domains, and between the FNIII domain and the transmembrane region), this does not exclude any of the association models.

This flexibility also means that it is possible for the two putative $FLRT2_{LRR}$ arrangements to represent either *cis* or *trans* dimers that might occur at the surfaces of isolated and interfacing cells, respectively (Figure 5.17). Switching between *cis* and *trans* arrangements upon the approach of another cell expressing FLRT molecules on its surface, as suggested for FLRT1, may therefore also occur for FLRT2 and FLRT3. An additional possibility for FLRT3, not observed for the other two FLRTs, is that the two concave-convex dimer arrangements could be combined at a cell-cell interface to form a theoretically infinite two-dimensional "zipper" array of

FLRT3 molecules between the two cells (Figure 5.17). However, at this stage, it is not clear which arrangements of FLRT2 and FLRT3 LRR domains are found at the cell surface; any suggestions are highly speculative and necessarily based solely upon the crystal structures available.

5.5 Analysis of the LRR domain dimerisation interface across the FLRT family

The high level of sequence identity between the mouse FLRT LRR domain sequences (57 % between $FLRT1_{LRR}$ and $FLRT2_{LRR}$; 68 % between $FLRT1_{LRR}$ and $FLRT3_{LRR}$; 58 % between $FLRT2_{LRR}$ and $FLRT3_{LRR}$) is illustrative of the high level of sequence conservation across members of the FLRT family from multiple higher vertebrate species (Figure 5.18). Levels of conservation are slightly lower than this across the entire FLRT sequence (43 % between mouse $FLRT1_{LRR}$ and mouse $FLRT2_{LRR}$; 57 % between $FLRT1_{LRR}$ and $FLRT3_{LRR}$; 45 % between $FLRT2_{LRR}$ and $FLRT3_{LRR}$), as might be expected given the inherent conservation of the LRR repeat, though are still relatively high across a range of vertebrate species; intriguingly, however, no avian FLRT1 homologue has yet been identified.

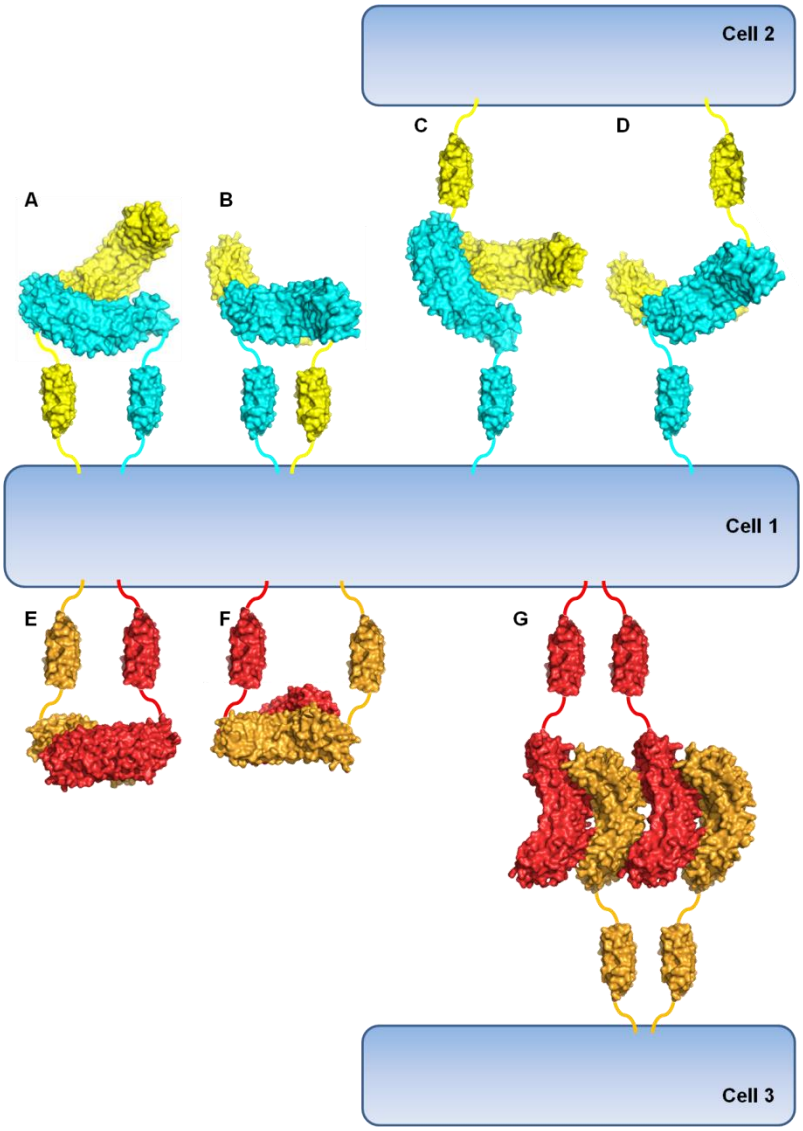
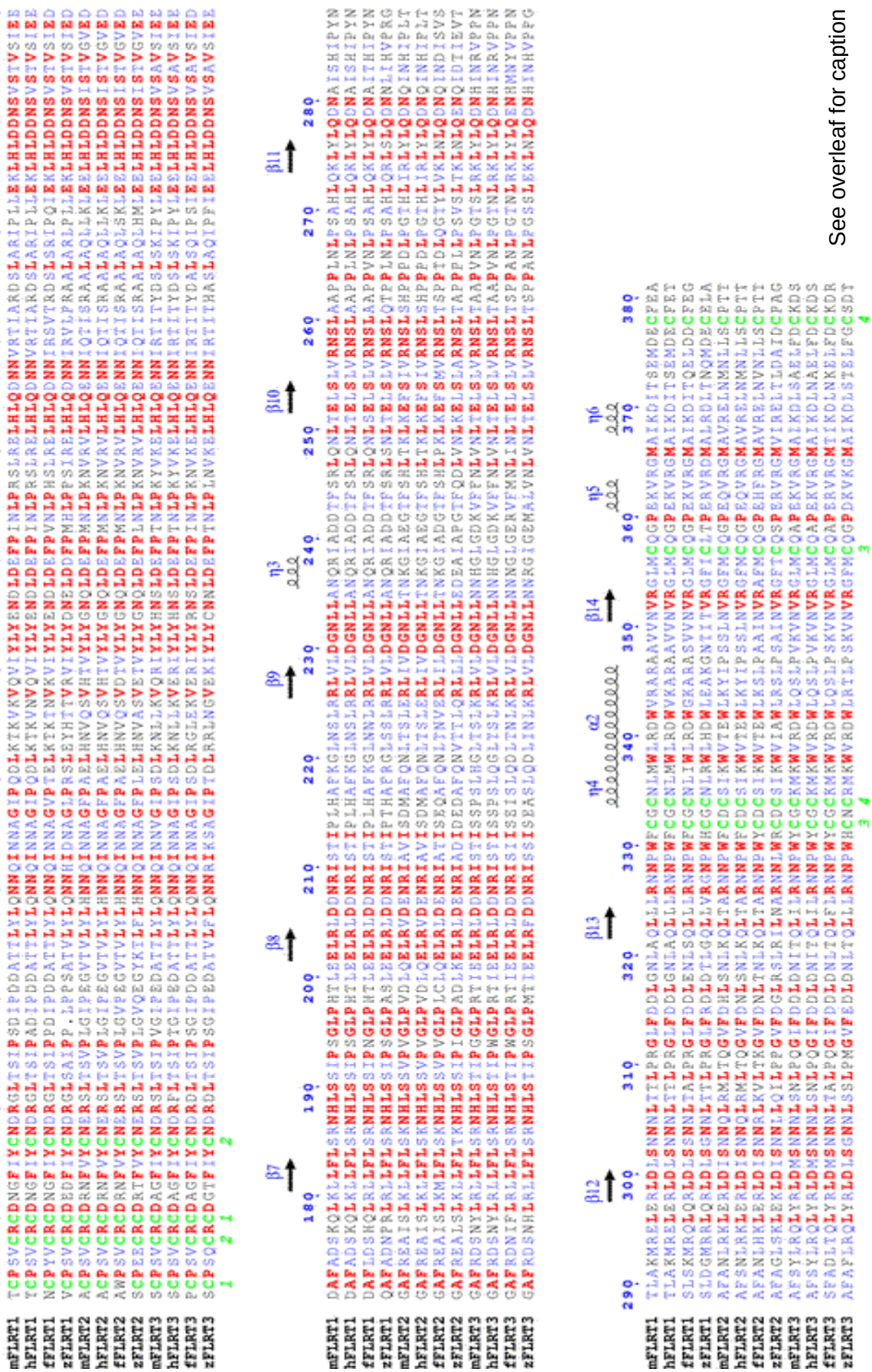


Figure 5.17 Representations of how FLRT2 (top) and FLRT3 (bottom) might self-associate at the surface of isolated (A, B, E, F) and opposing (C, D, G) cells. (A, B) Two possible configurations of a FLRT2 *cis* dimer at the cell surface, based on two crystallographic dimers (from a single crystal) involving the C-terminus and (A) concave face, or (B) convex face; both are consistent with FRET results, which were measured on non-interfacing cell surfaces. These arrangements may also represent *trans* cell surface dimers (C, D) between opposing cells; *trans* interactions between cells could also be mediated by interactions between *cis* dimer units (not shown; it is not clear from crystal structures how this might occur). (E, F) Two possible arrangements of a FLRT3 *cis* dimer, both involving the concave and convex faces of two molecules; further side-to-side crystallographic dimers have been discounted due to the positioning of *N*-glycosylation sites. One possibility for a *trans* interaction between cells could be a two-dimensional "zipper" array (G) formed by a combination of these two interfaces. The FNIII domain is modelled by 3TES.pdb, with flexible interdomain regions modelled by curved lines.



See overleaf for caption

Figure 5.18 Sequence alignment of FLRT1, 2 and 3 LRR domains across species. Amino acid sequences were taken from the following sources (accession numbers are for GenBank; m=mouse, h=human, f=frog, z=zebrafish): mFLRT1: NP_958813; hFLRT1: NP_037412; fFLRT1: NP_001086411; zFLRT1: XP_003198770; mFLRT2: NP_958926; hFLRT2: AAI43937; fFLRT2: CAE54087; zFLRT2: XP_003200389; mFLRT3: NP_001165631; hFLRT3: NP_037413; fFLRT3: AAI70046; zFLRT3: NP_001243575. Numbering above the alignment corresponds to the mouse FLRT1 sequence, where residue 1 is the initial methionine. Secondary structure elements derived from the structure of $mFLRT1_{LRR}$ (assigned using DSSP (Kabsch and Sander, 1983)) are also indicated above the alignment. Identical residues are coloured red, residues which are similar are coloured blue. Cysteine pairs forming disulphide bridges are coloured green and paired by numbers below the alignment.

An overlay of the three mouse $FLRT_{LRR}$ structures (Figure 5.12 (C)) shows that the sequence similarity is translated into high levels of conservation of the domain architecture, which is also very similar to LRR domain structures from other proteins. As for the majority of these other proteins, several of the intermolecular interfaces suggested by crystal structures utilise the concave face of the domain, and closer analysis of the surface residues which contribute to these interactions (using the PISA server (Krissinel and Henrick, 2007)) reveals that many are common across different arrangements, and across the different FLRTs (Figure 5.19). Specifically, the portion of the concave face towards the N-terminus, including residues from the patch of aromatic residues, is the site of most of the residues common to several interfaces, which indicates the importance of this region of the protein in forming homodimer, and possibly other intermolecular, interactions.

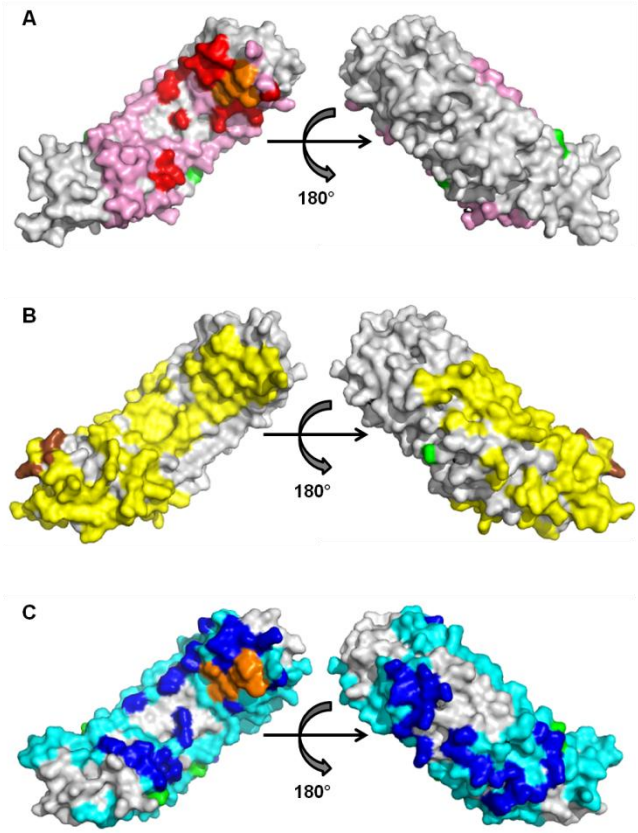


Figure 5.19 Surface residues of FLRT1 (A), 2 (B) and 3 (C) LRR domains which are implicated in significant dimerisation interfaces suggested by the PISA server. Residues involved in each interface discussed above were identified and plotted on the surface of the FLRT LRR structures (pink (A): $FLRT1_{LRR}$ crystal 1 and crystal 2 interfaces; yellow (B): two crystallographic $FLRT2_{LRR}$ interfaces; light blue (C): four crystallographic $FLRT3_{LRR}$ interfaces). Where residues were found in more than one arrangement, different colouring is used (red (A): common $FLRT1_{LRR}$ interface residues; brown (B; 2 residues only): common $FLRT2_{LRR}$ interface residues; dark blue (C): common $FLRT3_{LRR}$ interface residues). Where residues from the aromatic patch are included in the set of common interface residues, they are coloured in orange. Asparagine residues corresponding to predicted *N*-glycosylation sites are coloured in bright green.

To determine whether residues identified as contributing to multiple crystallographic interfaces might be relevant to the biological function of the FLRT proteins, the degree of conservation of multiple FLRT amino acid sequences was plotted onto the $FLRT1_{LRR}$ structure (using the ConSurf server (Ashkenazy et al., 2010)). To ensure the results were representative across FLRTs and different species, a wide range of

sequences was analysed, including two rodent (mouse, rat), two primate (human, Rhesus monkey), two amphibian (*Xenopus laevis*, *Xenopus tropicalis*), two fish (zebrafish, tilapia) and two avian (zebra finch, turkey) sequences for each of the three FLRTs; as there is no avian FLRT1 homologue in the database, avian sequences were only for FLRT2 and FLRT3. Residues were coloured according to the degree to which they were conserved throughout the sequences, allowing identification of regions which were highly conserved and therefore more likely to be relevant to the biological function of the proteins, and regions which were less well conserved, where a higher level of variability can be tolerated structurally and functionally.

This analysis highlighted areas of high surface conservation on the concave face and on one side face of the LRR structure; the convex face and other side face were generally less well conserved. The region of high conservation on the concave face includes all residues from the aromatic patch, confirming the importance of these residues. This analysis indicates that crystallographic interfaces involving the concave face of the LRR domain (such as those used by the parallel and antiparallel dimers suggested from crystal 1 and 2 structures of $FLRT1_{LRR}$) are more likely to be physiologically relevant, and to represent the homotypic interactions between FLRT molecules that occur at the surfaces of cells. The same analysis was also performed on separate groups of FLRT1, FLRT2 and FLRT3 homologues, and produced strikingly similar results to the analysis across all three FLRTs.

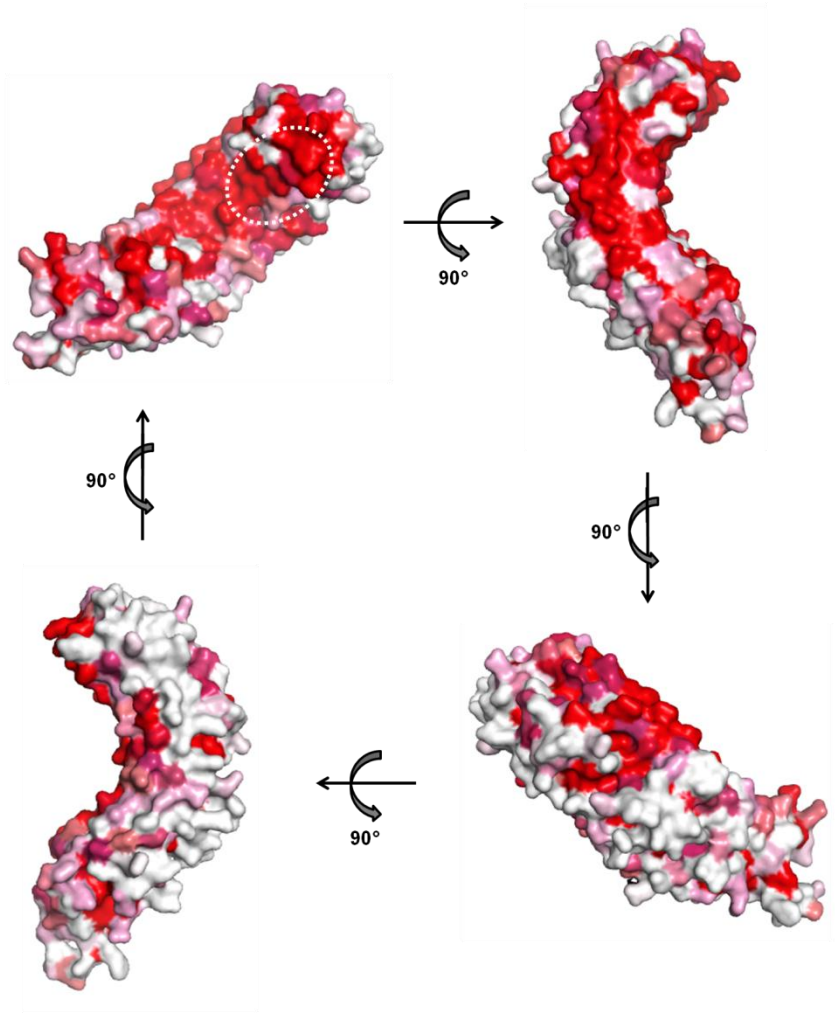


Figure 5.20 Degree of FLRT amino acid sequence conservation mapped onto the $FLRT1_{LRR}$ crystal structure, to indicate surface regions of high and low sequence conservation (scale from white (not conserved) to red (identical)). The position of the N-terminal aromatic patch is highlighted on the concave face (top left panel) by a dashed white circle. Sequences were sourced as follows (unless stated otherwise, accession numbers are for GenBank): mouse FLRT1, NP_958813; rat FLRT1, NP_001102630; human FLRT1, NP_037412; Rhesus macaque FLRT1, AFE65164; *X. laevis* FLRT1, NP_001086411; *X. tropicalis* FLRT1, UniProt F6RB26; zebrafish FLRT1, XP_003198770; tilapia FLRT1, UniProt I3KYE0; mouse FLRT2, NP_958926; rat FLRT2, NP_001100220; human FLRT2, AAI43937; Rhesus macaque FLRT2, NP_001253402; *X. laevis* FLRT2, CAE54088; *X. tropicalis* FLRT2, XP_002933238; zebrafish FLRT2, XP_003200389; tilapia FLRT2, UniProt I3KYM4; zebra finch FLRT2, UniProt H0ZQ68; turkey FLRT2, UniProt G1NR04; mouse FLRT3, NP_001165631; rat FLRT3, NP_001119763; human FLRT3, NP_037413; Rhesus macaque FLRT3, NP_001252801; *X. laevis* FLRT3, AAI70048; *X. tropicalis* FLRT3, AAI60452; zebrafish FLRT3, NP_001243575; tilapia FLRT3, UniProt I3KZM3; zebra finch FLRT3, UniProt H0YW34; turkey FLRT3, XP_003203695.

5.6 Conclusions and future perspectives

Results from biophysical analyses discussed in the previous chapter indicate that weak, transient homotypic interactions exist between FLRT LRR domain constructs in solution, and that these may form the basis of interaction between full-length FLRT molecules at the cell surface. It may be that restriction of spatial movement by localisation in the plane of the cell membrane (and/or the involvement of other binding partners) is responsible for stabilisation of such interactions at the cell surface, allowing accumulation of FLRTs at cell interfaces, and adhesion between cells.

Interactions between FLRT-transfected cells, in *trans*, were investigated in Chapter 4, using cellular assays. The existence of *cis* homotypic interactions between FLRTs at the cell surface is supported by results, discussed in this Chapter, from FRET and, in the case of FLRT1 only, cell surface co-patching studies. The former shows that the fluorophore-tagged C-termini are close enough (1-10 nm) for a significant FRET signal to occur between them, and the latter indicates that FLRT molecules with different N-terminal tags associate with each other strongly enough to pull each other into the same regions of the cell membrane upon patching using antibodies. It would be informative to repeat this type of analysis for FLRT2 and FLRT3 as well.

Crystal structures of the LRR domains of all three mouse FLRTs were solved, representing a significant advance in the characterisation of these proteins. They were found to be structurally highly similar, both to each other and to LRR domain structures from other proteins. Dimeric interfaces represented in the crystals may indicate arrangements of the LRR domains in *cis* and/or *trans* FLRT interactions at the cell surface, though these are only speculative at this stage. Mutations were designed to probe which arrangements were likely to be physiological, but were

retained in the endoplasmic reticulum and not expressed at the cell surface; however, this may indicate that FLRTs are produced and expressed at the cell surface as dimers. The two models of $FLRT1_{LRR}$ interaction, originating from different crystals and with distinct chain orientations, both use the concave faces of the two LRR domains, as has also been observed for other LRR-containing proteins; suggested arrangements for the FLRT2 and FLRT3 LRR domains include other parts of the domain surface. One possible arrangement of FLRT3 molecules between opposing cells, in a zipper type array made up of interactions between the concave and convex faces of multiple LRR domains, rather than isolated dimers as seen for FLRT1, may explain why FLRT3-transfected cells appear to form stronger adhesive *trans* interactions than cells expressing FLRT1 (e.g. in Sf9 cells, Figure 4.3). The adhesive strength of such an interaction may also explain why the phenotype of FLRT3 knockout mice is quite severe, with apparent defects in tissue adhesion (Maretto et al., 2008), while no phenotype has yet been observed for a FLRT1 knockout mouse (personal communication, Liz Robertson and Liz Bikoff (both Sir William Dunn School of Pathology, University of Oxford)). However, no such array-type arrangement is suggested for FLRT2 which, when knocked out, also exhibits a severe developmental phenotype (Müller et al., 2011).

Finally, analysis of the residues involved in the various crystallographic associations indicates that regions implicated in some of these arrangements (particularly those using the concave face) are coincident with regions of high sequence conservation. The N-terminal patch of aromatic residues, for example, is very well conserved across all three FLRTs and a range of species, and is included in the interface footprint in several suggested association models. This may indicate the importance of this region, and others which also exhibit a high level of conservation, in the

biological function of FLRT proteins. More specifically, given evidence that the LRR domain of FLRTs is required for *trans* interactions between cells (Karaulanov et al., 2006), this analysis provides a map of regions on the domain surface that may be implicated in this aspect of FLRT function.

The results discussed here are in agreement with those discussed in Chapter 4, which related to associations of FLRT constructs in solution. It appears that such homotypic associations may also be found at the cell surface, and may underlie aspects of FLRT-dependent behaviour including cell-cell adhesion. These findings represent a more complete structural characterisation of the functionally essential LRR domain, and its involvement in homotypic FLRT associations at the surfaces of isolated and opposing cells, than has previously been available. An intriguing extension to this work might be to investigate the structural basis and functional relevance of heterotypic interactions between the different FLRTs, and this is discussed in further detail in Chapter 6.

Chapter 6 Concluding Remarks and Future Perspectives

Chapter 6 Concluding Remarks and Future Perspectives

6.1 Concluding remarks

Communication between cells *via* their cell-surface receptors is essential to complex processes including development and immunity, and therefore to the overall survival of multi-cellular organisms. The work on structural and functional characterisation presented in this thesis focuses jointly on two classes of cell-surface receptors, namely cross-reactive TCRs from MS patients and FLRTs. Although much remains to be investigated, the results of this work represent an advance in our understanding of both receptor classes, and also indicate the directions that future research might take.

6.1.1 Recombinant production of MS patient-derived TCRs and complex formation with MHC class II molecules

MS is an autoimmune pathology resulting from loss of immunological self tolerance, specifically towards myelin proteins in the brain and CNS. A level of TCR cross-reactivity is essential to ensure that all possible foreign pathogens may be recognised within the size limits of the T cell repertoire (Mason, 1998), though (should peripheral tolerance mechanisms fail) may also allow erroneous recognition of self antigens (e.g. MBP) that are structurally similar to foreign ones, leading to a pathogenic autoimmune response (Oldstone, 1987). Crystal structures of TCR-pMHC complexes have so far proved invaluable for gaining an understanding of the structural basis of TCR cross-reactivity and its relevance to autoimmunity (Jones et al., 2006); the aim of the work described in Chapter 3 was to expand the database of such structures by studying two cross-reactive TCRs from MS patients.

The highly cross-reactive Hy (Hy.2E11) T cell clone was isolated from an MS patient with a relapsing-remitting disease course; Hy TCR recognises EBV and influenza virus peptides as well as the immunodominant MBP_{85-99} peptide in the context of several MS-associated MHC class II molecules ((Wucherpfennig and Strominger, 1995; Lang et al., 2002) and personal communication, Lars Fugger (HIU, WIMM, Oxford)). Hy TCR, and its complexes with these pMHC combinations, therefore represent intriguing targets for structural studies which may permit assessment of the contribution of molecular mimicry to cross-reactivity, though to date attempts to form Hy-pMHC complexes have been confounded by poor yields of the recombinant TCR protein. To address this, a number of constructs were designed for transient expression in mammalian cells (Section 3.2.1), though unfortunately this approach did not yield soluble heterodimeric protein due to the failure of all β chain constructs to be secreted (Section 3.2.2). Although some α chain constructs were secreted when expressed alone, the yields were still very low, making production of sufficient quantities of protein for crystallisation from this system excessively challenging (Section 3.2.3). It was therefore decided prudent to attempt to produce Hy protein in a standard bacterial expression system, though previous attempts had proved unsuccessful (Section 3.3). Small-scale trials of protein production and refolding from inclusion bodies indicate that Hy TCR can be successfully produced in this way, and further α and β chain constructs were designed to allow optimisation of the purification process and to include the antigenic peptides recognised by Hy, in order to facilitate formation of Hy TCR-pMHC complexes as part of future work (see Section 6.2.1).

The MS49 (MS49 D3) T cell clone was isolated from an MS patient with a secondary progressive disease course (Mazza et al., 2002); the MS49 TCR recognises the

MBP_{30-44} peptide in the context of DQ6. Unlike Hy TCR, protocols for production and purification of refolded MS49 protein have already been established (Sections 3.4.1 and 3.4.2), and suitable conditions for complex formation with MBP_{30-44}-DQ6 have been identified (Section 3.4.3), representing a greater degree of progress towards crystallisation and structure solution of the complex (McMahon, 2009). However, the low yields of the component proteins from recombinant production systems, high levels of protein loss and the poor efficiency of complex formation present significant barriers to crystallisation. Several attempts at large-scale complex formation were made, each requiring multiple batches of each protein, though invariably yields were so small that only a limited number of crystallisation conditions could be screened (Section 3.4.4). Despite this, crystals (which appear to be composed of protein rather than salt) were grown in one crystallisation condition; although they diffracted poorly and no structural information could be obtained, this provides a starting point for future optimisation screens so that the limited amounts of protein complex might have a better chance of forming well-diffracting crystals from which structural information can be obtained (see Section 6.2.1).

6.1.2 Structural and functional characterisation of FLRTs

Another family of cell surface receptors, the FLRT proteins, have been shown to be involved in a number of roles including cell sorting and adhesion in the early embryonic development of higher vertebrates. The LRR domain, which is frequently found to mediate protein-protein interactions in a diverse range of processes and systems (Kobe and Kajava, 2001), has been shown to be essential for FLRT-dependent cell sorting behaviour (Karaulanov et al., 2006). The aim of the work described in Chapters 4 and 5 of this thesis was to further our understanding of FLRT function in homotypic cell-cell adhesion by investigating the nature of

homotypic interactions between the different FLRT molecules, both in solution and at the cell surface. The molecular basis of these interactions was also investigated by studying the involvement of the LRR domain using biophysical characterisation, and by solving and characterising crystal structures of the FLRT LRR domains.

Results discussed in this thesis indicate that FLRTs, which were shown to be expressed at the cell surface (Section 4.2.1), affect cell-cell interactions in both adherent and suspension cell lines. Specifically, all three FLRTs were shown to accumulate at sites of cell-cell adhesion between HEK 293 cells, and FLRT2 and FLRT3 were also shown to promote clustering of Sf9 cells, indicating that they are capable of recognition and interaction between cells (Section 4.2.2). It was suggested that a direct homotypic *trans* interaction between FLRT molecules might underlie this behaviour at the cellular level, and soluble constructs encoding the FLRT ectodomains and FLRT LRR domains were produced (Section 4.3) and characterised in terms of their ability to self-associate. AUC showed that at least two oligomeric forms of each FLRT ectodomain construct exist in solution (Section 4.4.2), and MALS indicated the existence of a low-affinity monomer-dimer equilibrium (Section 4.4.1). In the case of FLRT1 in particular, MALS indicates that the interaction between ectodomains is mediated predominantly by the LRR domain, which forms a concentration-dependent monomer-dimer equilibrium in solution characterised by a similar dissociation constant value to that seen for the ectodomain. This self-association behaviour was then shown to be relevant at the cell surface, where FRET and co-patching studies indicate that specific and direct homotypic interactions exist in *cis* between all three FLRTs (Sections 5.2.1 and 5.2.2).

The structural basis of these intermolecular interactions observed in solution, in *cis* at the surface of non-interfacing cells and in *trans* between opposing cells was also investigated by crystallising and solving the structures of the FLRT LRR domains (Sections 5.3.1 and 5.4.1). As well as providing the first structural information on this family of proteins, the resulting models also indicate how self-association of these domains might underlie FLRT dimerisation (Sections 5.3.2 and 5.4.2). Site-directed mutagenesis to probe the physiological relevance of two different putative interfaces between FLRT1 LRR domains was not able to distinguish between them due to a failure of the mutant constructs to express at the cell surface, though this could indicate that homotypic interactions between FLRTs are essential for cell surface expression (Section 5.3.3). The crystal structures were also used, in conjunction with results from other experiments, to suggest possible arrangements that FLRTs might adopt at the surface of isolated and interfacing cells (Sections 5.3.4 and 5.4.3). Finally, interfaces between the FLRT LRR domains, inferred from the crystal structures, were analysed - some of these (including both from the FLRT1 LRR crystal structures) make use of the concave surface of the LRR solenoid structure, which has also been found to contribute to LRR-mediated dimerisation interfaces in a range of other proteins (Bella et al., 2008). A patch of aromatic and histidine residues towards the N-terminus of the concave face is frequently implicated in these FLRT interfaces; this patch, as well as other regions of the concave and side faces of the LRR domain, was found to be highly conserved between members of the FLRT family and across a range of species (Section 5.5). This indicates that interfaces including these regions are likely to be physiologically relevant, and may form the basis for a direct LRR-mediated interaction between FLRT molecules, which in turn may underlie the cell-cell adhesion phenomena described here and elsewhere.

6.2 Future perspectives

6.2.1 Production and characterisation of cross-reactive TCRs

The extreme cross-reactivity of Hy TCR means it represents an intriguing system in which to study the contribution of molecular mimicry (e.g. between DR2a-EBV$_{627-641}$ and DR2b-MBP$_{85-99}$ (Smith et al., 1998; Lang et al., 2002)) to TCR cross-reactivity, and its relevance to the pathogenesis of autoimmune diseases including MS. However, this relies on the production of sufficient quantities of soluble protein for biophysical and structural studies. Despite the considerable effort expended on attempting to produce Hy αβ protein using a mammalian cell system, it appears that it may in fact be easier to produce *via* the conventional route of *in vitro* refolding from bacterial inclusion bodies. It is clear from results presented in this thesis that the refolding and purification protocols require further optimisation to increase the yield of pure αβ protein, and to this end another α chain construct including a C-terminal FLAG tag has already been cloned. If combined with a β chain construct with a C-terminal hexahistidine tag, this should allow αβ heterodimers to be resolved from αα and ββ homodimers also produced in the refolding process using affinity chromatography and IMAC steps. Additional β chain constructs have also been cloned, including three different peptides which have been shown to be recognised by Hy in the contexts of different MS-associated MHC class II molecules. These peptides are linked to the N-terminus of the β chain *via* a flexible glycine-rich linker, which has been shown to drive formation of TCR-pMHC complexes in the cases of other cross-reactive TCRs (e.g. MS49 and Ob (McMahon, 2009)). The use of these novel constructs for optimisation of the protocols for production and purification of Hy offers a potential next step in this project, and if sufficient quantities of protein can be

produced, structural and biophysical studies of the interactions of Hy with its cognate pMHC complexes might be possible.

In this case, formation, crystallisation and structural characterisation of tertiary complexes between Hy TCR and DR2a-EBV$_{627-641}$, DR2b-MBP$_{85-99}$ and DR4-EBV$_{627-641}$ might be expected to make a significant contribution to our understanding of the involvement of structural similarity in TCR cross-reactivity. Structures of DR2a-EBV$_{627-641}$ and DR2b-MBP$_{85-99}$, without TCR bound, suggest conservation of TCR contacts (Smith et al., 1998; Lang et al., 2002); the DR4-EBV$_{627-641}$ structure is yet to be solved, and it would be interesting to compare this to DR2a-EBV$_{627-641}$ and DR2b-MBP$_{85-99}$, both without and ideally with Hy TCR bound. It would also be informative to characterise the TCR-pMHC complexes biophysically, for example by investigation of the TCR binding affinity using surface plasmon resonance; other autoreactive TCRs have been shown to have a lower than usual affinity for the self pMHC complexes they bind (Li et al., 2005), and it would be interesting to see whether the same is true for Hy TCR.

The DR2 haplotype, consisting of the DR2a, DR2b and DQ6 alleles, accounts for just under 50 % of the genetic basis for MS susceptibility. Of these, DR2b appears to confer the highest level of risk of developing the disease, while DR2a plays a modifier role (Gregersen et al., 2006); however, the data on the contribution of DQ6 to MS risk remain inconclusive. If solved, the structure of the MS49 TCR in complex with DQ6-MBP$_{30-44}$ would represent the first model of a DQ6-restricted TCR complex and would hopefully provide structural insight into the role of DQ6 in autoreactive TCR recognition in MS. Although this project is at a more advanced stage than that involving the Hy TCR so that MS49 and other components required for complex formation can already be successfully produced and purified, the low yield of

complex formation has thus far formed a barrier to crystallisation, as has also been observed for the complex between the autoreactive Ob TCR and DR2b-MBP_{85-99} (Harkiolaki et al., 2009; McMahon, 2009). However, as described in this thesis, sufficient protein was produced for crystallisation trials and yielded some crystals; although these did not diffract sufficiently to provide structural information, the conditions under which they formed may provide the basis for future optimisation screens, and the crystals themselves could be used for seeding experiments. This could reduce the crystallisation space to be searched, with a higher probability of finding conditions which produce well-diffracting crystals than with using sparse matrix screens. This in turn could lead to solution of the complex crystal structure which, in combination with biophysical characterisation as suggested for Hy complexes, would extend our knowledge of the molecular basis of TCR autoreactivity further, and specifically indicate the role of DQ6 in the pathogenesis of MS.

6.2.2 Heterotypic interactions between FLRTs

The various roles of FLRTs in neural growth and embryonic development continue to be investigated using a variety of techniques, though until now very little information on the structural and biophysical nature of the interactions underlying these behaviours has been reported. The work described in this thesis has focussed on homotypic interactions between FLRT molecules, which appear (in concurrence with previously published data (Karaulanov et al., 2006)) to be dependent on the LRR domain. However, evidence of heterotypic physical interactions between FLRTs has also been reported (Karaulanov et al., 2006); FLRT3 was shown to co-immunoprecipitate FLRT1 and FLRT2 as well as itself, and FLRT2 was shown to interact strongly with FLRT1 as well as itself using a bioluminescence resonance energy transfer assay. Although the three FLRTs have been shown to have differing

patterns of expression in terms of both tissue restriction and developmental stage (Lacy et al., 1999), there is overlap in these patterns which could allow the different FLRTs to interact heterotypically. For example, all three are expressed to some extent in adult brain, as well as overlaps in expression in the developing embryo (Maretto et al., 2008). It could therefore be very interesting to use the tools described in this thesis, and others, to investigate heterotypic interactions between FLRTs, to add to our existing knowledge of homotypic ones, and further our understanding of the roles of these interactions in early development and neurite growth.

Experiments to examine cell-cell adhesion in Sf9 cells (Section 4.2.3) could be repeated using mixed populations of cells expressing FLRT molecules with different fluorescent tags (e.g. FLRT2 tagged with monoVenus, FLRT3 tagged with monoCherry), to determine whether cells can form heterotypic clumps in additional to the homotypic clustering behaviour observed for FLRT2 and FLRT3. Attempts to replicate this approach in HEK 293 cells were made in the course of the work described in this thesis, but proved to be unsuccessful due to the adherent nature of the cell line. When separately transfected cell populations were mixed, the resuspended cells failed to re-adhere to the bases of wells, and did not form interfaces of the type observed in adherent cells in Section 4.2.2. However, this problem could perhaps be overcome using the Sf9 suspension cell line.

The biophysical basis of any heterotypic interactions observed in the cellular assay could be investigated using the MALS and AUC techniques discussed in Sections 4.4.1 and 4.4.2 of this thesis, respectively. As in the investigation of homotypic association, this would allow the involvement of the LRR domain to be tested. However, the testing of two different constructs would allow the use of surface plasmon resonance to supplement results from MALS and AUC, which could

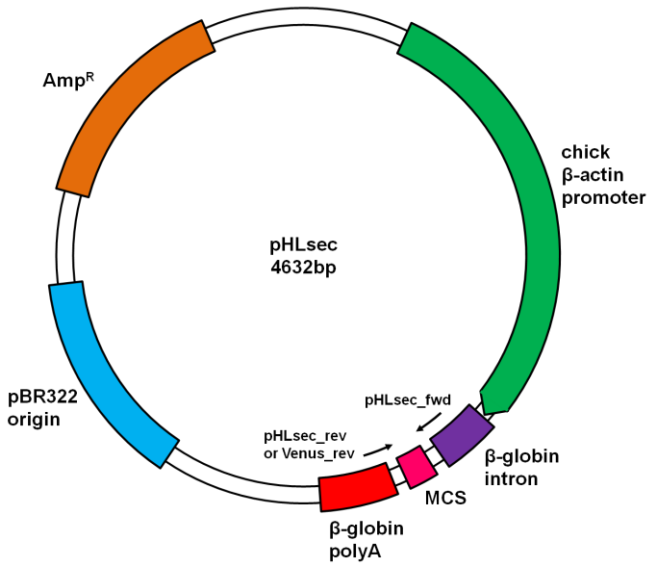
provide quantitative information on the affinity and association/dissociation characteristics of any interactions. These could be compared to the approximate K_d values obtained from analysis of MALS data for the FLRT1 homotypic interaction (Section 4.4.1) to determine whether all FLRT-FLRT interactions are as weak and transient as these values suggest. This information may in turn provide an interesting biophysical insight into results from cell surface FRET and co-patching assays investigating whether heterotypic associations occur at the cell surface.

Finally, as production and purification protocols have been optimised for all three FLRT LRR domains (Section 4.3), these could also be mixed prior to crystallisation to provide information on the molecular basis of any interactions observed, should the structures of any complexes be solved. It is clear from the wealth of biological data that has been accumulated that FLRTs exhibit a degree of specificity of binding, though it appears from data presented in this thesis that the gross structure of the LRR domain is not the basis of this specificity, given the high degree of similarity between the structures (Section 5.4.2). It seems likely, therefore, that any specificity is dependent upon the surface properties of the LRR domains. As discussed in Section 5.5, certain regions of the LRR domain surface are highly conserved, indicating their importance in intermolecular interactions, though for there to be specificity, there may be subtle variation between FLRTs that underlies differences in their affinities for each other. Structural studies of complexes of the different FLRT LRR domains would provide an exciting opportunity to assess the contribution of such variation to the specificity of interactions between the FLRTs, as well as information on the structural basis of interactions observed using other techniques.

Appendices

Appendix A Expression Vectors

A



B

```
gaattcaagcttgccaccatgggatccttcccagccctgggatgcctgcgctgctctcc
 EcoRI      Kozak  M  G  I  L  P  S  P  G  M  P  A  L  L  S

ctcgtgagccttctctccgtgctgctgatgggttgcgtagctgaaaccggt...insert...
 L  V  S  L  L  S  V  L  L  M  G  C  V  A     AgeI

ggtaccaagcaccaccatcaccatcactaatgatcactcgag
 KpnI  K  H  H  H  H  H  H  -  -     XhoI
```

C

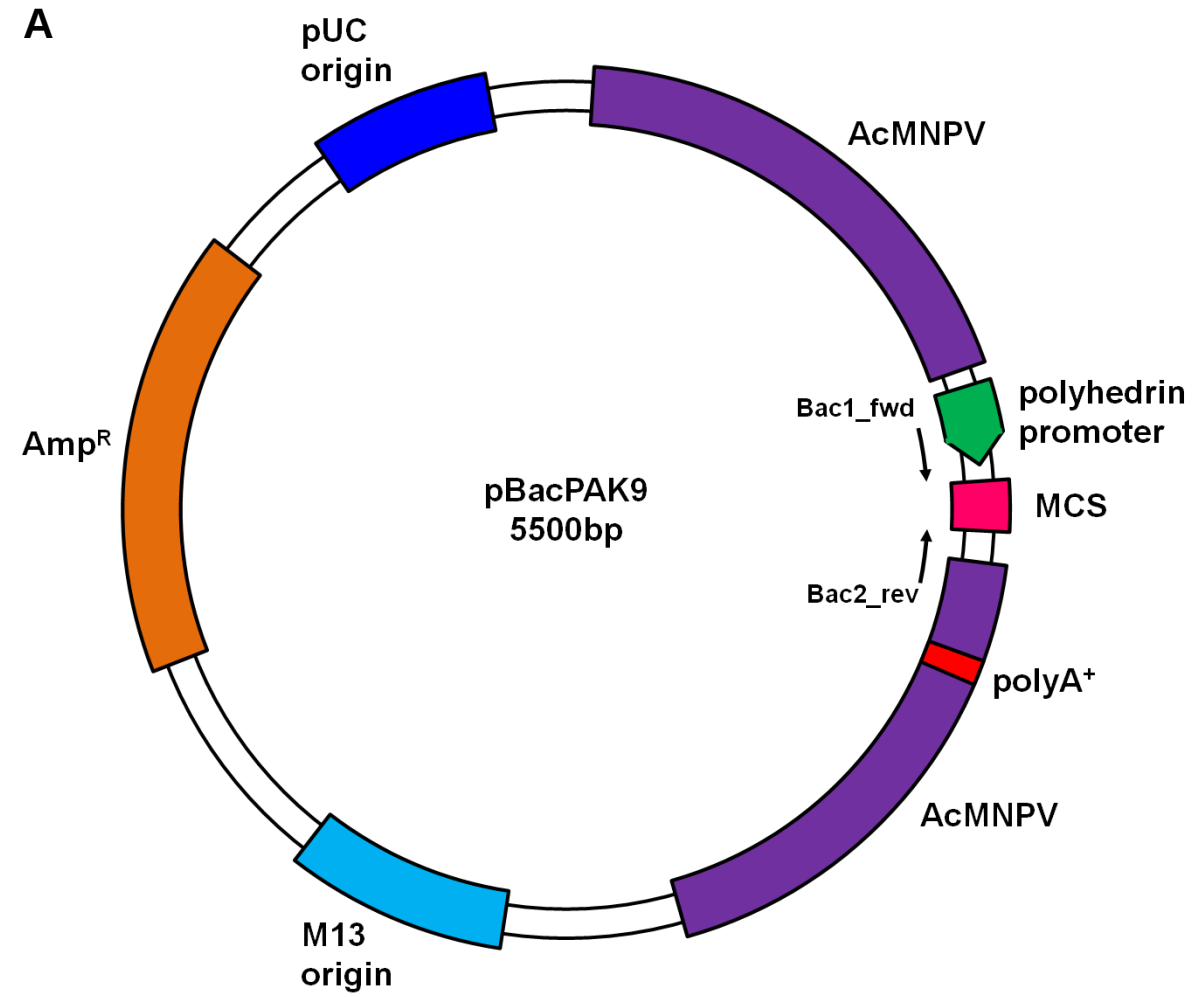
```
gaattcaagcttgccaccatgggatccttcccagccctgggatgcctgcgctgctctcc
 EcoRI      Kozak  M  G  I  L  P  S  P  G  M  P  A  L  L  S

ctcgtgagccttctctccgtgctgctgatgggttgcgtagctgaaaccggt...insert...
 L  V  S  L  L  S  V  L  L  M  G  C  V  A     AgeI

ggtaccctggaggtgctgttccagggccccggaatg...mVenus/mCerulean...aag
 KpnI  L  E  V  L  F  Q  G  P     M                        K

caccaccatcaccatcaccatcattaatgatcactcgag
 H  H  H  H  H  H  H  H  -  -     XhoI
```

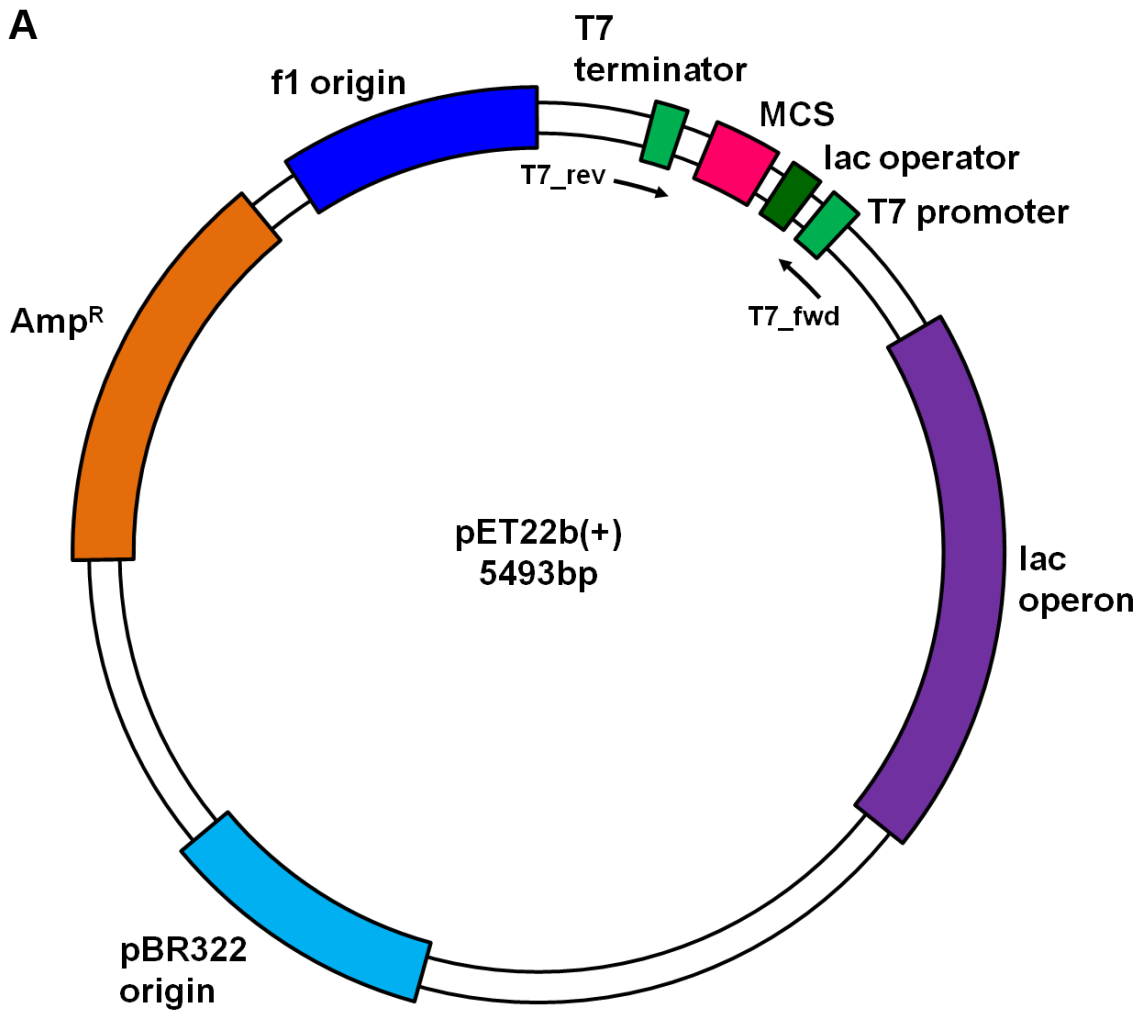
Key features of the pHLsec, pHL-mVenus and pHL-mCerulean vectors (A) Vector map of pHLsec, including sites for pHLsec_fwd and pHLsec_rev/Venus_rev sequencing primers. (B) Multiple cloning site (MCS) of pHLsec. The chick RPTPσ secretion signal amino acid sequence is shown in green, the C-terminal hexahistidine tag in blue and stop codons in red, with relevant restriction enzyme sites also shown. The *EcoRI* site is compatible with *MfeI*, the *AgeI* site is compatible with *NgoMIV*, and the *Acc65I* isoschizomer of the *KpnI* site is compatible with *BsiWI*. (C) MCS of pHL-mVenus/pHL-mCerulean. Colouring is as for (B), with the addition of a 3C protease cleavage site shown in pink, the C-terminal mVenus or mCerulean tag (start indicated by the ATG codon highlighted in green), and an additional two histidine residues in the histidine tag. Figure adapted from (Aricescu et al., 2006).



B

ggatccggaattc...insert...gagctcggtaccagatctctagattcgaactcgag
BamHI EcoRI SacI KpnI XhoI

Key features of the pBacPAK9 vector (Clontech) (A) Vector map of pBacPAK9, including sites for Bac1_fwd and Bac2_rev sequencing primers. AcMNPV, *Autographa californica* nucleopolyhedrovirus. (B) Part of the MCS of pBacPAK9, highlighting the restriction enzyme sites used in this thesis.



B

aaggagatatacatatg...insert...ctcgagcaccaccaccaccaccactga
rbs NdeI XhoI H H H H H H –

Key features of the pET22b(+) vector (Invitrogen) (A) Vector map of pET22b(+), including sites for T7_fwd and T7_rev sequencing primers. (B) Part of the MCS of pET22b(+), highlighting the ribosome binding site (rbs) in purple, the C-terminal hexahistidine tag in blue and stop codon in red, with restriction enzyme sites relevant to this thesis also shown. The second codon of the *NdeI* site also encodes the initial methionine residue of the inserted construct.

Appendix B Thesis Constructs

Protein	Construct	Amino acids	Forward primer (5' to 3')	Reverse primer (5' to 3')	Restriction enzymes	Vector
Human Hy.2E11 TCR α chain (TRAV17*01, TRAJ6*01, TRAC)	α1 (Clone provided by Dr. M. Harkiolaki)	1-222 (TRAV17*01: 1-112; TRAJ6*01 (beginning at residue 2): 113-131; tyrosine: 132; TRAC: 133-222) plus C-terminal ENDGGGCKLE extension	cgaggaattcgccaccATGGAAACT CTCCTGGGAGTGTCTTTGGT	cgcggtaccCTCTAGCTTGCATC CGCCTCCGTCGTTCTCTGGGC TGGGGAAGAAGGTGTCTTCTG GAATAATG*	EcoRI and KpnI	pHLsec
	α2	1-222 (as for α1, except Tyr132 is returned to native Asn)	GGAACCAGCCTTATTGTTCAT CCAAATATCCAGAACCCTGAC CCTGCC	GGCAGGGTCAGGGTTCTGGA TATTTGGATGAACAATAAGGC TGGTTCC	(EcoRI and KpnI)	pHLsec
	α3	1-222 (as for α1, except with C-terminal ESCDSVK extension)	cgaggaattcgccaccATGGAAACT CTCCTGGGAGTGTCTTTGGT	cgcggtaccTTTCACGCTGTCGC AAGACTCAGGGCTGGGGAAG AA*	EcoRI and KpnI	pHLsec
	α4	1-132 (TRAV17*01: 1-112; TRAJ6*01 (beginning at residue 2): 113-131; tyrosine: 132) plus C-terminal glycine	cgaggaattcgccaccATGGAAACT CTCCTGGGAGTGTCTTTGGT	cgcggtaccTCCGTATGGATGAA CAATAAGGCTGGTTCC*	EcoRI and KpnI	pHLsec
	α5	1-131 (as for α4, except Tyr132 is returned to native Asn)	cgaggaattcgccaccATGGAAACT CTCCTGGGAGTGTCTTTGGT	cgcggtaccTCCATTTGGATGAA CAATAAGGCTGGTTCC*	EcoRI and KpnI	pHLsec
	α6	22-222 (as for α1, except using the vector-derived signal sequence)	cgagaccggtCAACAGGGAGAAG AGGATCCTCAGGCC	cgcggtaccCTCTAGCTTGCATC CGCCTCCGTCGTTCTCTGGGC TGGGGAAGAAGGTGTCTTCTG GAATAATG*	AgeI and KpnI	pHLsec
	α7	24-222 (as for α6, except using a shortened N-terminus)	cgagaccggtGGAGAAGAGGATC CTCAGGCCTTGAGC	cgcggtaccCTCTAGCTTGCATC CGCCTCCGTCGTTCTCTGGGC TGGGGAAGAAGGTGTCTTCTG GAATAATG*	AgeI and KpnI	pHLsec
	α8	28-222 (as for α7, except using a shortened N-terminus)	cgagaccggtCCTCAGGCCTTGA GCATCCAGGAGGGTG	cgcggtaccCTCTAGCTTGCATC CGCCTCCGTCGTTCTCTGGGC TGGGGAAGAAGGTGTCTTCTG GAATAATG*	AgeI and KpnI	pHLsec

Protein	Construct	Amino acids	Forward primer (5' to 3')	Reverse primer (5' to 3')	Restriction enzymes	Vector
Human Hy.2E11 TCR α chain (TRAV17*01, TRAJ6*01, TRAC)	$\alpha_{Bact}1$	22-222 (as for α6, except without a signal sequence)	gctcacgccatATGCAACAGGGAG AAGAGGATCCTCAGGCC	cgcctcgagCTCTAGCTTGCATC CGCCTCCGTCGTTCTCTGGGC TGGGGAAGAAGGTGTCTTCTG GAATAATG*	NdeI and XhoI	pET22b(+)
	$\alpha_{Bact}2$	22-222 (as for $\alpha_{Bact}1$, except with an additional DYKDDDDK C-terminal extension)	gctcacgccatATGCAACAGGGAG AAGAGGATCCTCAGGCC	cgcctcgagTCATTTGTCGTCGT CGTCTTTGTAGTCCTCTAGCT TGCATCCGCCTCCGTCG	NdeI and XhoI	pET22b(+)
Human Hy.2E11 TCR β chain (TRBV29-1*01, TRBD1*01, TRBJ1-2*01, TRBC2)	β1	1-261 (TRBV29-1*01: 1-109; TRBD1*01 (reading frame 2): 110-117; TRBJ1-2*01 (beginning at residue 2): 118-130; threonine-glutamate: 131-132; TRBC2: 133-261 (serine: 206)) plus C-terminal QDRGGGCD extension)	cgaggaattcgccaccATGGTGAGT CTTCTGCTCCTTCTCCTG	cgcggtaccGTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*	EcoRI and KpnI	pHLsec
	β2	1-261 (as for β1, except with C-terminal RGGGCD extension)	cgaggaattcgccaccATGGTGAGT CTTCTGCTCCTTCTCCTG	cgcggtaccGTCACATCCACCTC CCCTGTCTGCTCGGCCCCAG GCCTC*	EcoRI and KpnI	pHLsec
	β3	1-261 (as for β1, except with C-terminal CGEQKLIEEDLDG extension)	cgaggaattcgccaccATGGTGAGT CTTCTGCTCCTTCTCCTG	cgcggtaccCCCGTCTAGGTCCT CCTCAATTAGTTTCTGCTCCC CGCAGTCTGCTCGGCCCCAG GCCTCGGCGCTGAC*	EcoRI and KpnI	pHLsec
	β4	1-272 (as for β1, except with native CGFTSESYQQG C-terminus)	cgaggaattcgccaccATGGTGAGT CTTCTGCTCCTTCTCCTG	cgcggtaccCCCCTGCTGGTAGC TCTCGCTGGTGAACCCGCAG TC*	Eco RI and KpnI	pHLsec
	β5	1-272 (as for β4, except with modified CGFTSNSYQQG C-terminus)	cgaggaattcgccaccATGGTGAGT CTTCTGCTCCTTCTCCTG	cgcggtaccCCCCTGCTGGTAGC TATTGCTGGTGAACCCGCAGT C*	EcoRI and KpnI	pHLsec
	β6	1-261 (as for β1, except Ser206 is returned to the native Cys)	GCCCTCAATGACTCCAGATAC TGTCTGAGCAGCCGCCTGAG GGTC	GACCCTCAGGCGGCTGCTCA GACAGTATCTGGAGTCATTGA GGGC	(EcoRI and KpnI)	pHLsec

Protein	Construct	Amino acids	Forward primer (5' to 3')	Reverse primer (5' to 3')	Restriction enzymes	Vector
Human Hy.2E11 TCR β chain (TRBV29-1*01, TRBD1*01, TRBJ1-2*01, TRBC2)	β7	1-261 (as for β3, except Ser206 is returned to the native Cys)	GCCCTCAATGACTCCAGATAC **TGT**CTGAGCAGCCGCCTGAG GGTC	GACCCTCAGGCGGCTGCTCA G**ACA**GTATCTGGAGTCATTGA GGGC	(EcoRI and KpnI)	pHLsec
	β8	1-272 (as for β4, except Ser206 is returned to the native Cys)	GCCCTCAATGACTCCAGATAC **TGT**CTGAGCAGCCGCCTGAG GGTC	GACCCTCAGGCGGCTGCTCA G**ACA**GTATCTGGAGTCATTGA GGGC	(EcoRI and KpnI)	pHLsec
	β9	1-133 (TRBV29-1*01: 1-109; TRBD1*01 (reading frame 2): 110-117; TRBJ1-2*01 (beginning at residue 2): 118-130; threonine: 131) plus native C-terminal Glu-Asp	cgag**gaattc**gccacc**ATG**GTGAGT CTTCTGCTCCTTCTCCTG	cgc**ggtacc**GTCCTCTGTGACCG TTAACCTGGTCCC*	EcoRI and KpnI	pHLsec
	β10	17-261 (as for β1, except using the vector-derived signal sequence)	cgag**accggt**GCTGTCATCTCTC AAAAGCCAAGCAGG	cgc**ggtacc**GTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*	AgeI and KpnI	pHLsec
	β11	21-261 (as for β10, except using a shortened N-terminus)	cgag**accggt**CAAAAGCCAAGCA GGGATATCTGTCAACGTGGAA CC	cgc**ggtacc**GTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*	AgeI and KpnI	pHLsec
	β12	2-261 (as for β10, except for the inclusion of the vector-derived signal sequence)	cgag**accggt**GTGAGTCTTCTGC TCCTTCTCCTG	cgc**ggtacc**GTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*	AgeI and KpnI	pHLsec
	$\beta_{Bact}1$	17-261 (as for β10, except without a signal sequence)	gctcacgc**catATG**GCTGTCATCT CTCAAAAGCCAAGCAGG	cgc**ctcgag**GTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*	NdeI and XhoI	pET22b(+)
	$\beta_{Bact}2$	21-261 (as for β11, except without a signal sequence)	gctcacgc**catATG**CAAAAGCCAA GCAGGGATATCTGTCAACGTG	cgc**ctcgag**GTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*	NdeI and XhoI	pET22b(+)

Protein	Construct	Amino acids	Forward primer (5' to 3')	Reverse primer (5' to 3')	Restriction enzymes	Vector
Human Hy.2E11 TCR β chain (TRBV29-1*01, TRBD1*01, TRBJ1-2*01, TRBC2)	$\beta_{Bact}3$	17-261 (as for $\beta_{Bact}1$, except with an N-terminal extension consisting of MBP_{85-99} followed by a GGSGGGGG linker; Cys28 is substituted by Ser)	gctcacgccatATGGAGAACCCC GTCGTCCACTTCTTCAAGAAC ATCGTCACCCCCAGGGGAGG AAGCGGAGGAGGAGGAGGA GCTGTCATCTCTCAAAAGCC; GCCAAGCAGGGATATCTCTCA ACGTGGAACCTCC (Cys28Ser)	cgcctcgagGTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*; GGAGGTTCCACGTTGAGAGAT ATCCCTGCTTGGC (Cys28Ser)	NdeI and XhoI	pET22b(+)
	$\beta_{Bact}4$	17-261 (as for $\beta_{Bact}3$, except N-terminal extension includes $EBV_{627-641}$)	gctcacgccatATGACCGGAGGA GTCTATCACTTCGTCAAGAAG CACGTCCACGAGAGCGGAGG AAGCGGAGGAGGAGGAGGAG CTGTCATCTCTCAAAAGCC	cgcctcgagGTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*	NdeI and XhoI	pET22b(+)
	$\beta_{Bact}5$	17-261 (as for $\beta_{Bact}3$, except N-terminal extension includes the influenza A peptide YRNLVWFIKKNTRYP)	gctcacgccatATGTATAGGAACC TCGTCTGGTTCATCAAGAAGA ACACCAGGTATCCCGGAGGA AGCGGAGGAGGAGGAGGAGC TGTCATCTCTCAAAAGCC	cgcctcgagGTCACATCCACCTC CCCTATCCTGGTCTGCTCGGC CCCAGGCCTC*	NdeI and XhoI	pET22b(+)
Human MS49 D3 TCR α chain (TRAV39*01, TRAJ34*01, TRAC)	MS49 α (Clone produced by Dr. R. McMahon)	TRAV39*01; TRAJ34*01; TRAC (ending at Pro91) plus C-terminal ENGGGCK extension	ccttggcatATGGAGCTGAAAGTG GAACAAAACCCGCTGTTCCTG AGCATGCAGGAGGGAAAA	cccgctcgagTTTACAACCACCA CCGTTTTCTGGGCTGGGGAAG AAGGTGTCTTCTGG	NdeI and XhoI	pET22b(+)
Human MS49 D3 TCR β chain (TRBV2*01, TRBJ2-1*01, TRBD2*02, TRBC2)	MS49 β (Clone produced by Dr. R. McMahon)	N-terminal extension consisting of MBP_{30-44} followed by a GGSGGGGG linker; TRBV2*01; TRBD2*02; TRBJ2-1*01; TRBC2 (ending at Asp128) plus C-terminal QDRGGGGCD extension; unpaired Cys substituted by Ser	gttccatATGGAACCGGAAGTC ACCCAGACTCCGAGCCATCAG GTCACAC (without N-terminal extension); CTCAATGACTCCAGATACAGC CTGAGCAGCCGCCTGA (Cys-Ser)	ccgggatcctcgagCTAGTCACAA CCACCACCACCCCTGTCCTG GTCTGCTCTACCCCAGGCCTC GGCGCTGACG; TCAGGCGGCTGCTCAGGCTG TATCTGGAGTCATTGAG (Cys-Ser)	NdeI and XhoI	pET22b(+)
Mouse FLRT1 (NCBI NP_958813.1)	HA-$FLRT1_{LONG}$	53-597 plus N-terminal YPYDVPDYA extension	cgcaccggtTACCCTTACGACGT GCCTGACTACGCCACCTGCC CATCAGT	cgcggtaccCTTCCTCCTGCTGC CCCTGTTGT	AgeI and KpnI	pHLsec

Protein	Construct	Amino acids	Forward primer (5' to 3')	Reverse primer (5' to 3')	Restriction enzymes	Vector
Mouse FLRT1 (NCBI NP_958813.1)	FLAG-FLRT1$_{LONG}$	53-597 plus N-terminal DYKDDDDK extension	cgcaccggtGACTACAAGGACGACGACGACAAGACCTGCCCATCAGTGTG	cgcggtaccCTTCCTCCTGCTGCCCCTGTTGT	AgeI and KpnI	pHLsec
	FLRT1$_{LONG}$-mVenus / mCerulean	1-597	cgaggaattcgccaccATGGTGGTGGCACACTCTGCTGC	cgcggtaccCTTCCTCCTGCTGCCCCTGTTGT	EcoRI and KpnI; EcoRI and XhoI †	pHL-mVenus / pHL-mCerulean; pBacPAK9
	Dimer KO FLRT1$_{LONG}$-mVenus / mCerulean	1-597 with Arg225Glu, Arg226Glu and Ala321Ser substitutions	TTCAAGGGTCTCAACAGCCTGGAAGAACTGGTTCTGGATGGCAACCTG; CCTGGGGAACCTGTCACAGCTGCTTCTCAGG	CAGGTTGCCATCCAGAACCAGTTCTTCCAGGCTGTTGAGACCCTTGAA; CCTGAGAAGCAGCTGTGACAGGTTCCCCAGG	(EcoRI and KpnI)	pHL-mVenus / pHL-mCerulean
	RR225-6EE FLRT1$_{LONG}$-mVenus	1-597 with Arg225Glu and Arg226Glu substitutions	TTCAAGGGTCTCAACAGCCTGGAAGAACTGGTTCTGGATGGCAACCTG	CAGGTTGCCATCCAGAACCAGTTCTTCCAGGCTGTTGAGACCCTTGAA	(EcoRI and KpnI)	pHL-mVenus
	A321S FLRT1$_{LONG}$-mVenus	1-597 with Ala321Ser substitution	CCTGGGGAACCTGTCACAGCTGCTTCTCAGG	CCTGAGAAGCAGCTGTGACAGGTTCCCCAGG	(EcoRI and KpnI)	pHL-mVenus
	N319L FLRT1$_{LONG}$-mVenus	1-597 with Asn319Leu substitution	CCTGTTTGATGACCTGGGGCTCCTGGCACAGC	GCTGTGCCAGGAGCCCCAGGTCATCAAACAGG	(EcoRI and KpnI)	pHL-mVenus
	R346E, V349L FLRT1$_{LONG}$-mVenus	1-597 with Arg346Glu and Val349Leu substitutions	GGGTGAGGGCAGAGGCTGCACTGGTCAATGTGC	GCACATTGACCAGTGCAGCCTCTGCCCTCACCC	(EcoRI and KpnI)	pHL-mVenus
	T84W FLRT1$_{LONG}$-mVenus	1-597 with Thr84Trp substitution	CCCCGACGATGCCTGGACCCTCTACC	GGTAGAGGGTCCAGGCATCGTCGGGG	(EcoRI and KpnI)	pHL-mVenus

Protein	Construct	Amino acids	Forward primer (5' to 3')	Reverse primer (5' to 3')	Restriction enzymes	Vector
Mouse FLRT1 (NCBI NP_958813.1)	$FLRT1_{ECD}$	1-550	Clone provided by Dr. M. Harkiolaki		EcoRI and KpnI	pHLsec
	$FLRT1_{LRR}$	53-381	Clone provided by Dr. M. Harkiolaki		AgeI and KpnI	pHLsec
Mouse FLRT2 (NCBI NP_958926.1)	HA-$FLRT2_{LONG}$	35-584 plus N-terminal YPYDVPDYA extension	cgcaccggtTACCCTTACGACGT GCCTGACTACGCCGCCTGCCC TAGTGT	cggcgtacgCCGTCGGCCCCG GTTGTATTTC	AgeI and BsiWI	pHLsec
	FLAG-$FLRT2_{LONG}$	35-584 plus N-terminal DYKDDDDK extension	cgcaccggtGACTACAAGGACGA CGACGACAAGGCCTGCCCTAG TGTATG	cggcgtacgCCGTCGGCCCCG GTTGTATTTC	AgeI and BsiWI	pHLsec
	$FLRT2_{LONG}$-mVenus / mCerulean	1-584	cgaggaattcgccaccATGGGCCTAC AGACTACAAAGTGGC	cggcgtacgCCGTCGGCCCCG GTTGTATTTC	EcoRI and BsiWI; EcoRI and XhoI †	pHL-mVenus / pHL-mCerulean; pBacPAK9
	Dimer KO $FLRT2_{LONG}$-mVenus / mCerulean	1-584 with Arg208Glu and Lys303Ser substitutions	CCTCACAAGCTTGGAGGAGCT GATCGTGGATGG; CATCTCTCCAACCTGTCACAAC TCACTGCG	CCATCCACGATCAGCTCCTC CAAGCTTGTGAGG; CGCAGTGAGTTGTGACAGGT TGGAGAGATG	(EcoRI and BsiWI)	pHL-mVenus / pHL-mCerulean
	$FLRT2_{ECD}$	1-538	Clone provided by Dr. M. Harkiolaki		EcoRI and BsiWI	pHLsec
	$FLRT2_{LRR}$	36-361 with Glu352Ala mutation	Clone provided by Dr. M. Harkiolaki		AgeI and BsiWI	pHLsec
Mouse FLRT3 (NCBI NP_001165631.1)	HA-$FLRT3_{LONG}$	30-572 plus N-terminal YPYDVPDYA extension	cgcgccggcTACCCTTACGACGT GCCTGACTACGCCTCCTGTCC ATCTGT	cggcgtacgCTTTCTCCTCCGC CCTTTGCTGTAC	NgoMIV and BsiWI	pHLsec

Protein	Construct	Amino acids	Forward primer (5' to 3')	Reverse primer (5' to 3')	Restriction enzymes	Vector
Mouse FLRT3 (NCBI NP_001165631.1)	FLAG-FLRT3$_{LONG}$	30-572 plus N-terminal DYKDDDDK extension	cgc**gccggc**GACTACAAGGACG ACGACGACAAGTCCTGTCCAT CTGTATG	cgg**cgtacg**CTTTCTCCTCCGC CCTTTGCTGTAC	NgoMIV and BsiWI	pHLsec
	FLRT3$_{LONG}$-mVenus / mCerulean	1-572	catcag**caattg**gccaccATGATCAGC CCAGCCTGGAGCCT catcag**ggatcc**gccaccATGATCAG CCCAGCCTGGAGCC	cgg**cgtacg**CTTTCTCCTCCGC CCTTTGCTGTAC cgg**gagctc**CTTTCTCCTCCGC CCTTTGCTGTACGC	MfeI and BsiWI; BamHI and SacI	pHL-mVenus / pHL-mCerulean; pBacPAK9
	Dimer KO FLRT3$_{LONG}$-mVenus / mCerulean	1-572 With Lys202Glu, Arg203Glu	GGTCTCACAAGCCTG**GAAGAG** CTGGTTTTAGATGG	CCATCTAAAACCAG**CTCTTC** CAGGCTTGTGAGACC	(MfeI and BsiWI)	pHLsec
	FLRT3$_{ECD}$	1-525	Clone provided by Dr. M. Harkiolaki		MfeI and BsiWI	pHLsec
	FLRT3$_{LRR}$	1-359	Clone provided by Dr. M. Harkiolaki		MfeI and BsiWI	pHLsec
Human CD45 (NCBI NP_563578.2)	CD45-mVenus	26-1145	Clone provided by Dr. R. Aricescu		AgeI and KpnI	pHL-mVenus
Rat CD2 (NCBI NP_036962.1)	CD2	1-239	Clone provided by Dr. R. A. J. McIlhinney			
	CD2-YFP	1-239 plus C-terminal YFP tag	Clone provided by Dr. R. A. J. McIlhinney			
Mouse TrkC (NCBI NM_182809.2)	TrkC-mVenus	31-464	Clone provided by Dr. C. Coles		AgeI and KpnI	pHL-mVenus

***** Reverse primers shown lack a stop codon and will include the vector-derived hexahistidine tag. Reverse primers were also produced including a terminal stop codon so that constructs lacking the hexahistidine tag can also be produced.

† Constructs were cut from pHL-mVenus using the enzymes indicated, then ligated into the pBacPAK9 vector. Cloning of FLRT3$_{LONG}$ into pBacPAK9, with separate insertion of mVenus, is described in Section 2.1.1.

Appendix C List of Formulae

Formulae relating to MALS

Formula 1 If $\mathrm{M_w}(\mathrm{MALS})$, $\mathrm{M_w}(M)$ and $\mathrm{M_w}(D)$ are the molecular weight values determined by a MALS experiment and of monomer and dimer forms of the protein of interest respectively; $[T]$, $[M]$ and $[D]$ are the total protein concentration and molar concentrations of monomer and dimer respectively; and S_T, S_M and S_D are the total scattering, scattering from the monomer and scattering from the dimer respectively:

$$\mathrm{M_w}(\mathrm{MALS}) \times [T] \times S_T = (\mathrm{M_w}(M) \times [M] \times S_M) + (\mathrm{M_w}(D) \times [D] \times S_D)$$

as $[T] = [M] + [D]$ and $S_T = S_M + S_D$.

Formula 2 Because $\frac{[M]}{[T]} + \frac{[D]}{[T]} = 1$, $\frac{S_M}{S_T} + \frac{S_D}{S_T} = 1$, and $\mathrm{M_w}(D) = 2\mathrm{M_w}(M)$, Formula 1 becomes:

$$\mathrm{M_w}(\mathrm{MALS}) \times \frac{[T]}{[T]} \times \frac{S_T}{S_T} = \left(\mathrm{M_w}(M) \times \frac{[M]}{[T]} \times \frac{S_M}{S_T}\right) + \left(2\mathrm{M_w}(M) \times \left(1 - \frac{[M]}{[T]} \times \frac{S_M}{S_T}\right)\right)$$

Formula 3 Reducing Formula 2 gives

$$2 - \frac{\mathrm{M_w}(\mathrm{MALS})}{\mathrm{M_w}(M)} = \frac{[M]}{[T]} \times \frac{S_M}{S_T}$$

which should allow the value of $\left(\frac{[M]}{[T]} \times \frac{S_M}{S_T}\right)$ to be determined from experimental data. The value of $\left(\frac{[D]}{[T]} \times \frac{S_D}{S_T}\right)$ can be determined from this using $\left(\frac{[D]}{[T]} \times \frac{S_D}{S_T}\right) = 1 - \left(\frac{[M]}{[T]} \times \frac{S_M}{S_T}\right)$.

Formula 4 Rearranging Formula 3 gives

$$\frac{[M]}{[T]} = \left.2 - \frac{\mathrm{M_w(MALS)}}{\mathrm{M_w}(M)}\right/ \frac{S_M}{S_T} \text{ and } \frac{[D]}{[T]} = \left.1 - \left(2 - \frac{\mathrm{M_w(MALS)}}{\mathrm{M_w}(M)}\right)\right/ \frac{S_D}{S_T}$$

and because $\frac{[M]}{[T]} + \frac{[D]}{[T]} = 1$ and $S_D = 2S_M$,

$$2 - \frac{\mathrm{M_w(MALS)}}{\mathrm{M_w}(M)} + \frac{1-\left(2-\frac{\mathrm{M_w(MALS)}}{\mathrm{M_w}(M)}\right)}{2} = \frac{S_M}{S_T}$$

Formula 5 Simplifying Formula 4 gives

$$\frac{3-\frac{\mathrm{M_w(MALS)}}{\mathrm{M_w}(M)}}{2} = \frac{S_M}{S_T}$$

$\frac{[M]}{[T]}$ and $\frac{[D]}{[T]}$ can then be calculated using Formula 3, and as $[T]$ is known, $[M]$ and $[D]$ can also be calculated.

Formula 6 $\mathrm{K_d} = \frac{[M]^2}{[D]}$

where $\mathrm{K_d}$ is the equilibrium dissociation constant for the dimerisation reaction $M + M \rightleftharpoons D$. In this way (using Formulae 1 to 6), $\mathrm{K_d}$ may be calculated using data from a MALS experiment.

Formulae relating to X-ray crystallography

Formula 7 $F_{hkl} = |F_{hkl}| \exp[i\alpha(hkl)]$

The structure factor F_{hkl} in terms of its amplitude, $|F_{hkl}|$, and its phase, $\exp[i\alpha(hkl)]$.

Formula 8 $I_{hkl} \propto |F_{hkl}|^2$

where I_{hkl} is the intensity of a reflection hkl.

Formula 9 $R_{merge} = \frac{\sum_{hkl}\sum_i I(hkl;i) - \langle I(hkl)\rangle}{\sum_{hkl}\sum_i I(hkl;i)}$

where $I(hkl;i)$ is the intensity of a single measurement, i, of the reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity, from multiple measurements, of the reflection hkl.

Formula 10 $R_{pim} = \frac{\sum_{hkl}\sqrt{\frac{1}{n-1}}\sum_i I(hkl;i) - \langle I(hkl)\rangle}{\sum_{hkl}\sum_i I(hkl;i)}$

As for R_{merge} (Formula 9) with an additional $\sqrt{\frac{1}{n-1}}$ term, where n is the redundancy of the data (see Formula 12), to correct for oversampling.

Formula 11 $\text{Completeness} = \frac{\text{No. unique reflections observed}}{\text{No. unique reflections expected}} \times 100$

Formula 12 $\text{Redundancy} = \frac{\text{Total no. reflections observed}}{\text{No. unique reflections observed}}$

Formula 13 $R_{work} = \frac{\sum_{hkl}||F_{obs}| - |F_{calc}||}{\sum_{hkl}|F_{obs}|}$

where $|F_{obs}|$ are the experimentally observed structure factor amplitudes and $|F_{calc}|$ are the structure factor amplitudes calculated from the model.

Formula 14 $R_{free} = \frac{\sum_{hkl\ (\text{test set})}||F_{obs}|-|F_{calc}||}{\sum_{hkl\ (\text{test set})}|F_{obs}|}$

Calculated as for R_{work} (Formula 13), with (typically) a randomly selected 5 % of data that has been excluded from the refinement process, to provide an unbiased evaluation of the model.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D *66*, 213-221.

Anderton, S.M., and Wraith, D.C. (2002). Selection and fine-tuning of the autoimmune T-cell repertoire. Nat Rev Immunol *2*, 487-498.

Andreae, L.C., Peukert, D., Lumsden, A., and Gilthorpe, J.D. (2007). Analysis of Lrrn1 expression and its relationship to neuromeric boundaries during chick neural development. Neural Dev *2*, 22.

Andrews, R.K., Gardiner, E.E., Shen, Y., and Berndt, M.C. (2004). Platelet interactions in thrombosis. IUBMB Life *56*, 13-18.

Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. Acta crystallographica *62*, 1243-1250.

Aricescu, A.R., Siebold, C., Choudhuri, K., Chang, V.T., Lu, W., Davis, S.J., van der Merwe, P.A., and Jones, E.Y. (2007). Structure of a tyrosine phosphatase adhesive interaction reveals a spacer-clamp mechanism. Science *317*, 1217-1220.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Research *38*, W529-W533.

Attri, A.K., and Minton, A.P. (2005). New methods for measuring macromolecular interactions in solution via static light scattering: basic methodology and application to nonassociating and self-associating proteins. Analytical Biochemistry *337*, 103-110.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. Proceedings of the National Academy of Sciences *98*, 10037-10041.

Barton, W.A., Liu, B.P., Tzvetkova, D., Jeffrey, P.D., Fournier, A.E., Sah, D., Cate, R., Strittmatter, S.M., and Nikolov, D.B. (2003). Structure and axon outgrowth inhibitor binding of the Nogo-66 receptor and related proteins. EMBO J *22*, 3291-3302.

Bella, J., Hindle, K., McEwan, P., and Lovell, S. (2008). The leucine-rich repeat structure. Cellular and Molecular Life Sciences *65*, 2307-2333.

Benoist, C., and Mathis, D. (2001). Autoimmunity provoked by infection: how good is the case for T cell epitope mimicry? Nat Immunol *2*, 797-801.

Bixby, J.L., Lilien, J., and Reichardt, L.F. (1988). Identification of the major proteins that promote neuronal process outgrowth on Schwann cells in vitro. The Journal of Cell Biology *107*, 353-361.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallographica Section D *60*, 2210-2221.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol *294*, 1351-1362.

Bottcher, R.T., Pollet, N., Delius, H., and Niehrs, C. (2004). The transmembrane protein XFLRT3 forms a complex with FGF receptors and promotes FGF signalling. Nat Cell Biol *6*, 38-44.

Bowden, T.A., Crispin, M., Harvey, D.J., Aricescu, A.R., Grimes, J.M., Jones, E.Y., and Stuart, D.I. (2008). Crystal structure and carbohydrate analysis of Nipah virus attachment glycoprotein: a template for antiviral and vaccine design. Journal of Virology *82*, 11628-11636.

Brochet, X., Lefranc, M.-P., and Giudicelli, V. (2008). IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Research *36*, W503-W508.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., and Wiley, D.C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature *364*, 33-39.

Busch, R., Reich, Z., Zaller, D.M., Sloan, V., and Mellins, E.D. (1998). Secondary structure composition and pH-dependent conformational changes of soluble recombinant HLA-DM. Journal of Biological Chemistry *273*, 27557-27564.

Cavallaro, U., and Christofori, G. (2004). Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. Nat Rev Cancer *4*, 118-132.

Cavallaro, U., Niedermeyer, J., Fuxa, M., and Christofori, G. (2001). N-CAM modulates tumour-cell adhesion to matrix by inducing FGF-receptor signalling. Nat Cell Biol *3*, 650-657.

Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I.*, et al.* (2007). Glycoprotein structural genomics: solving the glycosylation problem. Structure (London, England : 1993) *15*, 267-273.

Chen, X., Koh, E., Yoder, M., and Gumbiner, B.M. (2009). A protocadherin-cadherin-FLRT3 complex controls cell adhesion and morphogenesis. PLoS ONE *4*, e8411.

Chen, Y., Aulia, S., Li, L., and Tang, B.L. (2006). AMIGO and friends: An emerging family of brain-enriched, neuronal growth modulating, type I transmembrane proteins with leucine-rich repeats (LRR) and cell adhesion molecule motifs. Brain Research Reviews *51*, 265-274.

Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A.A., and Strominger, J.L. (1992). Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. Nature *358*, 764-768.

Collaborative Computational Project 4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallographica Section D *50*, 760-763.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S.*, et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res *35*, W375-383.

DeLano, W.L. (2008). The PyMOL molecular graphics system (DeLano Scientific LLC, Palo Alto, CA, USA).

DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., and André, I. (2011). Modeling symmetric macromolecular structures in Rosetta3. PLoS ONE *6*, e20450.

DiMaio, F., Tyka, M.D., Baker, M.L., Chiu, W., and Baker, D. (2009). Refinement of protein structures into low-resolution density maps using Rosetta. Journal of Molecular Biology *392*, 181-190.

Dolan, J., Walshe, K., Alsbury, S., Hokamp, K., O'Keeffe, S., Okafuji, T., Miller, S.F., Tear, G., and Mitchell, K.J. (2007). The extracellular leucine-rich repeat superfamily; a comparative survey and analysis of evolutionary relationships and expression patterns. BMC genomics *8*, 320.

Ebers, G.C. (2008). Environmental factors and multiple sclerosis. The Lancet Neurology *7*, 268-277.

Egea, J., Erlacher, C., Montanez, E., Burtscher, I., Yamagishi, S., Heß, M., Hampel, F., Sanchez, R., Rodriguez-Manzaneque, M.T., Bösl, M.R.*, et al.* (2008). Genetic ablation of FLRT3 reveals a novel morphogenetic function for the anterior visceral endoderm in suppressing mesoderm differentiation. Genes & Development *22*, 3349-3362.

Elbein, A.D., Tropea, J.E., Mitchell, M., and Kaushal, G.P. (1990). Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. Journal of Biological Chemistry *265*, 15599-15605.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta crystallographica *60*, 2126-2132.

Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T., and Matsushima, N. (2004). Structural principles of leucine-rich repeat (LRR) proteins. Proteins: Structure, Function, and Bioinformatics *54*, 394-403.

Eswarakumar, V.P., Lax, I., and Schlessinger, J. (2005). Cellular signaling by fibroblast growth factor receptors. Cytokine & Growth Factor Reviews *16*, 139-149.

Ettinger, R.A., and Kwok, W.W. (1998). A peptide binding motif for HLA-DQA1*0102/DQB1*0602, the class II MHC molecule associated with dominant protection in insulin-dependent Diabetes Mellitus. The Journal of Immunology *160*, 2365-2373.

Evdokimov, A.G., Anderson, D.E., Routzahn, K.M., and Waugh, D.S. (2001). Unusual molecular architecture of the Yersinia pestis cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. Journal of Molecular Biology *312*, 807-821.

Fawcett, J.W., and Keynes, R.J. (1990). Peripheral nerve regeneration. Annual Review of Neuroscience *13*, 43-60.

Feng, D., Bond, C.J., Ely, L.K., Maynard, J., and Garcia, K.C. (2007). Structural evidence for a germline-encoded T cell receptor-major histocompatibility complex interaction 'codon'. Nat Immunol *8*, 975-983.

Filbin, M.T. (2003). Myelin-associated inhibitors of axonal regeneration in the adult mammalian CNS. Nat Rev Neurosci *4*, 703-713.

Fournier, A.E., and Strittmatter, S.M. (2001). Repulsive factors and axon regeneration in the CNS. Current Opinion in Neurobiology *11*, 89-94.

Frayser, M., Sato, A.K., Xu, L., and Stern, L.J. (1999). Empty and peptide-loaded class II major histocompatibility complex proteins produced by expression in *Escherichia coli* and folding *in vitro*. Protein Expression and Purification *15*, 105-114.

Giudicelli, V., Chaume, D., and Lefranc, M.-P. (2004). IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V–J and V–D–J rearrangement analysis. Nucleic Acids Research *32*, W435-W440.

Gong, S.G., Mai, S., Chung, K., and Wei, K. (2009). Flrt2 and Flrt3 have overlapping and non-overlapping expression during craniofacial development. Gene Expression Patterns *9*, 497-502.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305-308.

Goverman, J. (2009). Autoimmune T cell responses in the central nervous system. Nat Rev Immunol *9*, 393-407.

Gregersen, J.W., Kranc, K.R., Ke, X., Svendsen, P., Madsen, L.S., Thomsen, A.R., Cardon, L.R., Bell, J.I., and Fugger, L. (2006). Functional epistasis on a common MHC haplotype associated with multiple sclerosis. Nature *443*, 574-577.

Grothe, C., Meisinger, C., and Claus, P. (2001). In vivo expression and localization of the fibroblast growth factor system in the intact and lesioned rat peripheral nerve and spinal ganglia. Journal of comparative neurology *434*, 342-357.

Grueninger-Leitch, F., D'Arcy, A., D'Arcy, B., and Chène, C. (1996). Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. Protein Science *5*, 2617-2622.

Gupta, R., Jung, E., and Brunak, S. (in preparation). Prediction of N-glycosylation sites in human proteins.

Hahn, M., Nicholson, M.J., Pyrdol, J., and Wucherpfennig, K.W. (2005). Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. Nat Immunol *6*, 490-496.

Haines, B.P., Wheldon, L.M., Summerbell, D., Heath, J.K., and Rigby, P.W.J. (2006). Regulated expression of FLRT genes implies a functional role in the regulation of FGF signalling during mouse development. Developmental Biology *297*, 14-25.

Haines, J.L., Terwedow, H.A., Burgess, K., Pericak-Vance, M.A., Rimmler, J.B., Martin, E.R., Oksenberg, J.R., Lincoln, R., Zhang, D.Y., Banatao, D.R.*, et al.* (1998). Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. Human Molecular Genetics *7*, 1229-1234.

Hammerschmidt, M., and Wedlich, D. (2008). Regulated adhesion as a driving force of gastrulation movements. Development *135*, 3625-3641.

Han, B.W., Herrin, B.R., Cooper, M.D., and Wilson, I.A. (2008). Antigen recognition by variable lymphocyte receptors. Science *321*, 1834-1837.

Hansen, S.M., Berezin, V., and Bock, E. (2008). Signaling mechanisms of neurite outgrowth induced by the cell adhesion molecules NCAM and N-cadherin. Cellular and Molecular Life Sciences *65*, 3809-3821.

Harkiolaki, M., Holmes, S.L., Svendsen, P., Gregersen, J.W., Jensen, L.T., McMahon, R., Friese, M.A., van Boxel, G., Etzensperger, R., Tzartos, J.S.*, et al.* (2009). T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. Immunity *30*, 348-357.

He, X.-l., Radu, C., Sidney, J., Sette, A., Ward, E.S., and Garcia, K.C. (2002). Structural snapshot of aberrant antigen presentation linked to autoimmunity: the immunodominant epitope of MBP complexed with I-Au. Immunity *17*, 83-94.

Helft, L., Reddy, V., Chen, X., Koller, T., Federici, L., Fernández-Recio, J., Gupta, R., and Bent, A. (2011). LRR conservation mapping to predict functional sites within protein leucine-rich repeat domains. PLoS ONE *6*, e21614.

Hennecke, J., Carfi, A., and Wiley, D.C. (2000). Structure of a covalently stabilized complex of a human αβ T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. EMBO J *19*, 5611-5624.

Hennecke, J., and Wiley, D.C. (2002). Structure of a complex of the human α/β T cell receptor (TCR) HA1.7, Influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401). The Journal of Experimental Medicine *195*, 571-581.

Herrin, B.R., and Cooper, M.D. (2010). Alternative adaptive immunity in jawless vertebrates. The Journal of Immunology *185*, 1367-1374.

Hillert, J., and Olerup, O. (1993). Multiple sclerosis is associated with genes within or close to the HLA-DR-DQ subregion on a normal DR15,DQ6,Dw2 haplotype. Neurology *43*, 163-168.

Hogquist, K.A., Baldwin, T.A., and Jameson, S.C. (2005). Central tolerance: learning self-control in the thymus. Nat Rev Immunol *5*, 772-782.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Research *38*, W545-W549.

Howitt, J.A., Clout, N.J., and Hohenester, E. (2004). Binding site for Robo receptors revealed by dissection of the leucine-rich repeat region of Slit. EMBO J *23*, 4406-4412.

Hunt, D., Michel, H., Dickinson, T., Shabanowitz, J., Cox, A., Sakaguchi, K., Appella, E., Grey, H., and Sette, A. (1992). Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-Ad. Science *256*, 1817-1820.

Ide, C. (1996). Peripheral nerve regeneration. Neuroscience Research *25*, 101-121.

Jin, M.S., Kim, S.E., Heo, J.Y., Lee, M.E., Kim, H.M., Paik, S.-G., Lee, H., and Lee, J.-O. (2007). Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell *130*, 1071-1082.

Jones, E.Y., Fugger, L., Strominger, J.L., and Siebold, C. (2006). MHC class II proteins and disease: a structural perspective. Nat Rev Immunol *6*, 271-282.

Jukkola, T., Lahti, L., Naserke, T., Wurst, W., and Partanen, J. (2006). FGF regulated gene expression and neuronal differentiation in the developing midbrain–hindbrain region. Developmental Biology *297*, 141-157.

Kabsch, W. (2010). XDS. Acta Crystallographica Section D *66*, 125-132.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577-2637.

Kajander, T., Kuja-Panula, J., Rauvala, H., and Goldman, A. (2011). Crystal structure and role of glycans and dimerization in folding of neuronal leucine-rich repeat protein AMIGO-1. Journal of Molecular Biology *413*, 1001-1015.

Kalandadze, A., Galleno, M., Foncerrada, L., Strominger, J.L., and Wucherpfennig, K.W. (1996). Expression of recombinant HLA-DR2 molecules. Journal of Biological Chemistry *271*, 20156-20162.

Kantardjieff, K.A., and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. Protein Science *12*, 1865-1871.

Karaulanov, E., Böttcher, R.T., Stannek, P., Wu, W., Rau, M., Ogata, S., Cho, K.W.Y., and Niehrs, C. (2009). Unc5B interacts with FLRT3 and Rnd1 to modulate cell adhesion in *Xenopus* embryos. PLoS ONE *4*, e5742.

Karaulanov, E.E., Bottcher, R.T., and Niehrs, C. (2006). A role for fibronectin-leucine-rich transmembrane cell-surface proteins in homotypic cell adhesion. EMBO Reports *7*, 283-290.

Kim, S.-D., Liu, J.L., Roscioli, T., Buckley, M.F., Yagnik, G., Boyadjiev, S.A., and Kim, J. (2012). Leucine-rich repeat, immunoglobulin-like and transmembrane domain 3 (LRIT3) is a modulator of FGFR1. FEBS Letters *586*, 1516-1521.

Kiselyov, V.V., Skladchikova, G., Hinsby, A.M., Jensen, P.H., Kulahin, N., Soroka, V., Pedersen, N., Tsetlin, V., Poulsen, F.M., Berezin, V*., et al.* (2003). Structural basis for a direct interaction between FGFR1 and NCAM and evidence for a regulatory role of ATP. Structure *11*, 691-701.

Kleywegt, G. (2007). Crystallographic refinement of ligand complexes. Acta Crystallographica Section D *63*, 94-100.

Ko, J., and Kim, E. (2007). Leucine-rich repeat proteins of synapses. Journal of Neuroscience Research *85*, 2824-2832.

Kobe, B., and Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature *366*, 751-756.

Kobe, B., and Deisenhofer, J. (1995). A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature *374*, 183-186.

Kobe, B., and Kajava, A.V. (2001). The leucine-rich repeat as a protein recognition motif. Current Opinion in Structural Biology *11*, 725-732.

Kouhara, H., Hadari, Y.R., Spivak-Kroizman, T., Schilling, J., Bar-Sagi, D., Lax, I., and Schlessinger, J. (1997). A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. Cell *89*, 693-702.

Krantz, D.E., and Zipursky, S.L. (1990). Drosophila chaoptin, a member of the leucine-rich repeat family, is a photoreceptor cell-specific adhesion molecule. EMBO journal *9*, 1969-1977.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology *372*, 774-797.

Kuja-Panula, J., Kiiltomäki, M., Yamashiro, T., Rouhiainen, A., and Rauvala, H. (2003). AMIGO, a transmembrane protein implicated in axon tract development, defines a novel protein family with leucine-rich repeats. The Journal of Cell Biology *160*, 963-973.

Lacy, S.E., Bonnemann, C.G., Buzney, E.A., and Kunkel, L.M. (1999). Identification of FLRT1, FLRT2, and FLRT3: a novel family of transmembrane leucine-rich repeat proteins. Genomics *62*, 417-426.

Lang, H.L.E., Jacobsen, H., Ikemizu, S., Andersson, C., Harlos, K., Madsen, L., Hjorth, P., Sondergaard, L., Svejgaard, A., Wucherpfennig, K.*, et al.* (2002). A functional and structural basis for TCR cross-reactivity in multiple sclerosis. Nat Immunol *3*, 940-943.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J.*, et al.* (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168-176.

Lemmon, M.A., and Schlessinger, J. (1994). Regulation of signal transduction and signal diversity by receptor oligomerization. Trends in Biochemical Sciences *19*, 459-463.

Li, Y., Huang, Y., Lue, J., Quandt, J.A., Martin, R., and Mariuzza, R.A. (2005). Structure of a human autoimmune TCR bound to a myelin basic protein self-peptide and a multiple sclerosis-associated MHC class II molecule. EMBO J *24*, 2968-2979.

Li, Y., Li, H., Martin, R., and Mariuzza, R.A. (2000). Structural basis for the binding of an immunodominant peptide from myelin basic protein in different registers by two HLA-DR2 proteins. Journal of Molecular Biology *304*, 177-188.

Lin, J.C., Ho, W.-H., Gurney, A., and Rosenthal, A. (2003). The netrin-G1 ligand NGL-1 promotes the outgrowth of thalamocortical axons. Nat Neurosci *6*, 1270-1276.

Linhoff, M.W., Harton, J.A., Cressman, D.E., Martin, B.K., and Ting, J.P.-Y. (2001). Two distinct domains within CIITA mediate self-association: involvement of the GTP-binding and leucine-rich repeat domains. Molecular and Cellular Biology *21*, 3001-3011.

Liu, J., Ernst, S.A., Gladycheva, S.E., Lee, Y.Y.F., Lentz, S.I., Ho, C.S., Li, Q., and Stuenkel, E.L. (2004). Fluorescence resonance energy transfer reports properties of syntaxin1A interaction with munc18-1 *in vivo*. Journal of Biological Chemistry *279*, 55924-55936.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics *50*, 437-450.

Madsen, L.S., Andersson, E.C., Jansson, L., Krogsgaard, M., Andersen, C.B., Engberg, J., Strominger, J.L., Svejgaard, A., Hjorth, J.P., Holmdahl, R.*, et al.* (1999). A humanized model for multiple sclerosis using HLA-DR2 and a human T-cell receptor. Nat Genet *23*, 343-347.

Maretto, S., Muller, P.S., Aricescu, A.R., Cho, K.W., Bikoff, E.K., and Robertson, E.J. (2008). Ventral closure, headfold fusion and definitive endoderm migration defects in mouse embryos lacking the fibronectin leucine-rich transmembrane protein FLRT3. Developmental Biology *318*, 184-193.

Marrie, R.A. (2004). Environmental risk factors in multiple sclerosis aetiology. The Lancet Neurology *3*, 709-718.

Mason, D. (1998). A very high level of crossreactivity is an essential feature of the T-cell receptor. Immunology Today *19*, 395-404.

Matthews, B.W. (1968). Solvent content of protein crystals. Journal of Molecular Biology *33*, 491-497.

Maulik, A., Sarkar, A.I., Devi, S., and Basu, S. (2012). Polygalacturonase-inhibiting proteins – leucine-rich repeat proteins in plant defence. Plant Biology *14*, 22-30.

Maynard, J., Petersson, K., Wilson, D.H., Adams, E.J., Blondelle, S.E., Boulanger, M.J., Wilson, D.B., and Garcia, K.C. (2005). Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. Immunity *22*, 81-92.

Mayo, C.J., Diprose, J.M., Walter, T.S., Berry, I.M., Wilson, J., Owens, R.J., Jones, E.Y., Harlos, K., Stuart, D.I., and Esnouf, R.M. (2005). Benefits of automated crystallization plate tracking, imaging, and analysis. Structure *13*, 175-182.

Mazza, G., Ponsford, M., Lowrey, P., Campbell, M.J., Zajicek, J., and Wraith, D.C. (2002). Diversity and dynamics of the T-cell response to MBP in DR2+ve individuals. Clinical & Experimental Immunology *128*, 538-547.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of Applied Crystallography *40*, 658-674.

McMahon, R.M. (2009). DPhil thesis: Structures of autoimmune peptide-MHC and TCR peptide-MHC complexes. In Division of Structural Biology and MRC Human Immunology Unit (University of Oxford, Oxford).

Milán, M., Weihe, U., Pérez, L., and Cohen, S.M. (2001). The LRR proteins capricious and tartan mediate cell interactions during DV boundary formation in the *Drosophila* wing. Cell *106*, 785-794.

Morlot, C., Thielens, N.M., Ravelli, R.B.G., Hemrika, W., Romijn, R.A., Gros, P., Cusack, S., and McCarthy, A.A. (2007). Structural insights into the Slit-Robo complex. Proceedings of the National Academy of Sciences *104*, 14923-14928.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.-y. (2004). The ankyrin repeat as molecular architecture for protein recognition. Protein Science *13*, 1435-1448.

Mühlfriedel, S., Kirsch, F., Gruss, P., Chowdhury, K., and Stoykova, A. (2007). Novel genes differentially expressed in cortical regions during late neurogenesis. European Journal of Neuroscience *26*, 33-50.

Müller, P.-S., Schulz, R., Maretto, S., Costello, I., Srinivas, S., Bikoff, E., and Robertson, E. (2011). The fibronectin leucine-rich repeat transmembrane protein Flrt2 is required in the epicardium to promote heart morphogenesis. Development *138*, 1297-1308.

Muraro, P.A., Vergelli, M., Kalbus, M., Banks, D.E., Nagle, J.W., Tranquill, L.R., Nepom, G.T., Biddison, W.E., McFarland, H.F., and Martin, R. (1997). Immunodominance of a low-affinity major histocompatibility complex-binding myelin basic protein epitope (residues 111-129) in HLA-DR4 (B1*0401) subjects is associated with a restricted T cell receptor repertoire. The Journal of Clinical Investigation *100*, 339-349.

Murphy, K., Travers, P., Walport, M., and Janeway, C. (2012). Janeway's Immunobiology, 8th edn (Garland Science, London; Taylor & Francis (distributor)).

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat Biotech *20*, 87-90.

Nakamura, K., Watakabe, A., Hioki, H., Fujiyama, F., Tanaka, Y., Yamamori, T., and Kaneko, T. (2007). Transiently increased colocalization of vesicular glutamate transporters 1 and 2 at single axon terminals during postnatal development of mouse neocortex: a quantitative analysis with correlation coefficient. European Journal of Neuroscience *26*, 3054-3067.

Napier, R. (2004). Plant hormone binding sites. Annals of Botany *93*, 227-233.

Nashmi, R., Dickinson, M.E., McKinney, S., Jareb, M., Labarca, C., Fraser, S.E., and Lester, H.A. (2003). Assembly of α4β2 nicotinic acetylcholine receptors assessed with functional fluorescently labeled subunits: effects of localization, trafficking, and nicotine-induced upregulation in clonal mammalian cells and in cultured midbrain neurons. The Journal of Neuroscience *23*, 11554-11567.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallographica Section A *50*, 157-163.

Ng, A.C.Y., Eisenberg, J.M., Heath, R.J.W., Huett, A., Robinson, C.M., Nau, G.J., and Xavier, R.J. (2011). Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. Proceedings of the National Academy of Sciences *108*, 4631-4638.

Niehrs, C., and Meinhardt, H. (2002). Developmental biology: Modular feedback. Nature *417*, 35-36.

Nobes, C.D., Lauritzen, I., Mattei, M.-G., Paris, S., Hall, A., and Chardin, P. (1998). A new Member of the Rho family, Rnd1, promotes disassembly of actin filament structures and loss of cell adhesion. The Journal of Cell Biology *141*, 187-197.

Nose, A., Mahajan, V.B., and Goodman, C.S. (1992). Connectin: A homophilic cell adhesion molecule expressed on a subset of muscles and the motoneurons that innervate them in Drosophila. Cell *70*, 553-567.

Nose, A., Umeda, T., and Takeichi, M. (1997). Neuromuscular target recognition by a homophilic interaction of connectin cell adhesion molecules in Drosophila. Development *124*, 1433-1441.

Nürnberger, T., Brunner, F., Kemmerling, B., and Piater, L. (2004). Innate immunity in plants and animals: striking similarities and obvious differences. Immunological Reviews *198*, 249-266.

O'Sullivan, Matthew L., de Wit, J., Savas, Jeffrey N., Comoletti, D., Otto-Hitt, S., Yates, John R., and Ghosh, A. (2012). FLRT proteins are endogenous latrophilin ligands and regulate excitatory synapse development. Neuron *73*, 903-910.

Ogata, S., Morokuma, J., Hayata, T., Kolle, G., Niehrs, C., Ueno, N., and Cho, K.W.Y. (2007). TGF-β signaling-mediated morphogenesis: modulation of cell adhesion via cadherin endocytosis. Genes & Development *21*, 1817-1831.

Oldstone, M.B. (1987). Molecular mimicry and autoimmune disease. Cell *50*, 819-820.

Olsson, T., Zhi, W.W., Höjeberg, B., Kostulas, V., Jiang, Y.P., Anderson, G., Ekre, H.P., and Link, H. (1990). Autoreactive T lymphocytes in multiple sclerosis determined by antigen-induced secretion of interferon-gamma. The Journal of Clinical Investigation *86*, 981-985.

Orgel, J.P.R.O., Eid, A., Antipova, O., Bella, J., and Scott, J.E. (2009). Decorin core protein (decoron) shape complements collagen fibril surface structure and mediates its binding. PLoS ONE *4*, e7028.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, Charles W. Carter, Jr., ed. (Academic Press), pp. 307-326.

Papageorgiou, A.C., Shapiro, R., and Acharya, K.R. (1997). Molecular recognition of human angiogenin by placental ribonuclease inhibitor - an X-ray crystallographic study at 2.0 A resolution. EMBO J *16*, 5162-5177.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth *8*, 785-786.

Pettinelli, C.B., and McFarlin, D.E. (1981). Adoptive transfer of experimental allergic encephalomyelitis in SJL/J mice after in vitro activation of lymph node cells by myelin basic protein: requirement for Lyt 1+ 2- T lymphocytes. The Journal of Immunology *127*, 1420-1423.

Price, S.R., Evans, P.R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645-650.

Raguenez, G., Desire, L., Lantrua, V., and Courtois, Y. (1999). BCL-2 is upregulated in human SH-SY5Y neuroblastoma cells differentiated by overexpression of fibroblast growth factor 1. Biochemical and Biophysical Research Communications *258*, 745-751.

Reeves, P.J., Callewaert, N., Contreras, R., and Khorana, H.G. (2002). Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative

HEK293S stable mammalian cell line. Proceedings of the National Academy of Sciences *99*, 13419-13424.

Robinson, M., Parsons Perez, M.C., Tébar, L., Palmer, J., Patel, A., Marks, D., Sheasby, A., De Felipe, C., Coffin, R., Livesey, F.J.*, et al.* (2004). FLRT3 is expressed in sensory neurons after peripheral nerve injury and regulates neurite outgrowth. Molecular and Cellular Neuroscience *27*, 202-214.

Romero, M.I., Rangappa, N., Garry, M.G., and Smith, G.M. (2001). Functional regeneration of chronically injured sensory afferents into adult spinal cord after neurotrophin gene therapy. The Journal of Neuroscience *21*, 8408-8416.

Sawcer, S., Ban, M., Maranian, M., Yeo, T.W., and A., C. (2005). A high-density screen for linkage in multiple sclerosis. The American Journal of Human Genetics *77*, 454-467.

Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends in Plant Science *13*, 663-670.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophysical Journal *78*, 1606-1619.

Scott, P.G., Dodd, C.M., Bergmann, E.M., Sheehan, J.K., and Bishop, P.N. (2006). Crystal structure of the biglycan dimer and evidence that dimerization is essential for folding and stability of class I small leucine-rich repeat proteoglycans. Journal of Biological Chemistry *281*, 13324-13332.

Scott, P.G., Grossmann, J.G., Dodd, C.M., Sheehan, J.K., and Bishop, P.N. (2003). Light and X-ray scattering show decorin to be a dimer in solution. Journal of Biological Chemistry *278*, 18353-18359.

Scott, P.G., McEwan, P.A., Dodd, C.M., Bergmann, E.M., Bishop, P.N., and Bella, J. (2004). Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan. Proceedings of the National Academy of Sciences of the United States of America *101*, 15633-15638.

Seabold, G.K., Wang, P.Y., Chang, K., Wang, C.-Y., Wang, Y.-X., Petralia, R.S., and Wenthold, R.J. (2008). The SALM family of adhesion-like molecules forms heteromeric and homomeric complexes. Journal of Biological Chemistry *283*, 8395-8405.

Seiradake, E., von Philipsborn, A.C., Henry, M., Fritz, M., Lortat-Jacob, H., Jamin, M., Hemrika, W., Bastmeyer, M., Cusack, S., and McCarthy, A.A. (2009). Structure and functional relevance of the Slit2 homodimerization domain. EMBO Rep *10*, 736-741.

Sethi, D.K., Schubert, D.A., Anders, A.-K., Heroux, A., Bonsor, D.A., Thomas, C.P., Sundberg, E.J., Pyrdol, J., and Wucherpfennig, K.W. (2011). A highly tilted binding mode by a self-reactive T cell receptor results in altered engagement of peptide and MHC. The Journal of Experimental Medicine *208*, 91-102.

Sibley, W., Bamford, C., and Clark, K. (1985). Clinical viral infections and multiple sclerosis. The Lancet *325*, 1313-1315.

Siebold, C., Hansen, B.E., Wyer, J.R., Harlos, K., Esnouf, R.E., Svejgaard, A., Bell, J.I., Strominger, J.L., Jones, E.Y., and Fugger, L. (2004). Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy. Proceedings of the National Academy of Sciences of the United States of America *101*, 1999-2004.

Sloan, V.S., Cameron, P., Porter, G., Gammon, M., Amaya, M., Mellins, E., and Zaller, D.M. (1995). Mediation by HLA-DM of dissociation of peptides from HLA-DR. Nature *375*, 802-806.

Smith, K.J., Pyrdol, J., Gauthier, L., Wiley, D.C., and Wucherpfennig, K.W. (1998). Crystal structure of HLA-DR2 (DRA*0101, DRB1*1501) complexed with a peptide from human myelin basic protein. The Journal of Experimental Medicine *188*, 1511-1520.

Smith, T., and Tickle, C. (2006). The expression of Flrt3 during chick limb development. The International journal of developmental biology *50*, 701-704.

Söllner, C., and Wright, G. (2009). A cell surface interaction network of neural leucine-rich repeat receptors. Genome Biology *10*, R99.

Sospedra, M., and Martin, R. (2005). Immunology of multiple sclerosis. Annual Review of Immunology *23*, 683-747.

Stern, J.N.H., Illés, Z., Reddy, J., Keskin, D.B., Fridkis-Hareli, M., Kuchroo, V.K., and Strominger, J.L. (2005). Peptide 15-mers of defined sequence that substitute for random amino acid copolymers in amelioration of experimental autoimmune encephalomyelitis. Proceedings of the National Academy of Sciences of the United States of America *102*, 1620-1625.

Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L., and Wiley, D.C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature *368*, 215-221.

Stern, L.J., and Wiley, D.C. (1992). The human class II MHC protein HLA-DR1 assembles as empty αβ heterodimers in the absence of antigenic peptide. Cell *68*, 465-477.

Stewart-Jones, G.B.E., McMichael, A.J., Bell, J.I., Stuart, D.I., and Jones, E.Y. (2003). A structural basis for immunodominant human T cell receptor recognition. Nat Immunol *4*, 657-663.

Stuart, D.I., Levine, M., Muirhead, H., and Stammers, D.K. (1979). Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. Journal of Molecular Biology *134*, 109-142.

Tanabe, K., Bonilla, I., Winkles, J.A., and Strittmatter, S.M. (2003). Fibroblast growth factor-inducible-14 is induced in axotomized neurons and promotes neurite outgrowth. The Journal of Neuroscience *23*, 9675-9686.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J., and Adams, P.D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallographica Section D *64*, 61-69.

Thisse, B., and Thisse, C. (2005). Functions and regulations of fibroblast growth factor signaling during embryonic development. Developmental Biology *287*, 390-402.

Tomás, A.R., Certal, A.C., and Rodríguez-León, J. (2011). FLRT3 as a key player on chick limb development. Developmental Biology *355*, 324-333.

Tossell, K., Andreae, L.C., Cudmore, C., Lang, E., Muthukrishnan, U., Lumsden, A., Gilthorpe, J.D., and Irving, C. (2011). Lrrn1 is required for formation of the midbrain-hindbrain

boundary and organiser through regulation of affinity differences between midbrain and hindbrain cells in chick. Developmental Biology *352*, 341-352.

Tretter, V., Altmann, F., and MÄRz, L. (1991). Peptide-N4-(N-acetyl-β-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached α1 → 3 to the asparagine-linked N-acetylglucosamine residue. European Journal of Biochemistry *199*, 647-652.

Trombetta, E.S., and Mellman, I. (2005). Cell biology of antigen processing *in vitro* and *in vivo*. Annual Review of Immunology *23*, 975-1028.

Tsang, M., and Dawid, I.B. (2004). Promotion and attenuation of FGF signaling through the Ras-MAPK pathway. Sci STKE *2004*, pe17-.

Tsuji, L., Yamashita, T., Kubo, T., Madura, T., Tanaka, H., Hosokawa, K., and Tohyama, M. (2004). FLRT3, a cell surface molecule containing LRR repeats and a FNIII domain, promotes neurite outgrowth. Biochemical & Biophysical Research Communications *313*, 1086-1091.

Walker, L.S.K., and Abbas, A.K. (2002). The enemy within: keeping self-reactive T cells at bay in the periphery. Nat Rev Immunol *2*, 11-19.

Walter, T.S., Diprose, J.M., Mayo, C.J., Siebold, C., Pickford, M.G., Carter, L., Sutton, G.C., Berrow, N.S., Brown, J., Berry, I.M*., et al.* (2005). A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. Acta crystallographica *61*, 651-657.

Wei, K., Xu, Y., Tse, H., Manolson, M.F., and Gong, S.-G. (2011). Mouse FLRT2 interacts with the extracellular and intracellular regions of FGFR2. Journal of Dental Research *90*, 1234-1239.

Wheldon, L.M., Haines, B.P., Rajappa, R., Mason, I., Rigby, P.W., and Heath, J.K. (2010). Critical role of FLRT1 phosphorylation in the interdependent regulation of FLRT1 function and FGF receptor signalling. PLoS ONE *5*, e10264.

Willer, C.J., Dyment, D.A., Risch, N.J., Sadovnick, A.D., Ebers, G.C., and Group, T.C.C.S. (2003). Twin concordance and sibling recurrence rates in multiple sclerosis. Proceedings of the National Academy of Sciences *100*, 12877-12882.

Williams, E.-J., Williams, G., Howell, F.V., Skaper, S.D., Walsh, F.S., and Doherty, P. (2001). Identification of an N-cadherin motif that can interact with the fibroblast growth factor receptor and is required for axonal growth. Journal of Biological Chemistry *276*, 43879-43886.

Wim Ang, C., Jacobs, B.C., and Laman, J.D. (2004). The Guillain–Barré syndrome: a true case of molecular mimicry. Trends in Immunology *25*, 61-66.

Winter, G. (2010). Xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography *43*, 186-190.

Wucherpfennig, K.W., Call, M.J., Deng, L., and Mariuzza, R. (2009). Structural alterations in peptide–MHC recognition by self-reactive T cell receptors. Current Opinion in Immunology *21*, 590-595.

Wucherpfennig, K.W., and Strominger, J.L. (1995). Molecular mimicry in T cell-mediated autoimmunity: Viral peptides activate human T cell clones specific for myelin basic protein. Cell *80*, 695-705.

Wunnenberg-Stapleton, K., Blitz, I.L., Hashimoto, C., and Cho, K.W. (1999). Involvement of the small GTPases XRhoA and XRnd1 in cell adhesion and head formation in early Xenopus development. Development *126*, 5339-5351.

Xu, Y., Wei, K., Kulyk, W., and Gong, S.G. (2011). FLRT2 promotes cellular proliferation and inhibits cell adhesion during chondrogenesis. Journal of Cellular Biochemistry *112*, 3440-3448.

Yamagishi, S., Hampel, F., Hata, K., del Toro, D., Schwark, M., Kvachnina, E., Bastmeyer, M., Yamashita, T., Tarabykin, V., Klein, R.*, et al.* (2011). FLRT2 and FLRT3 act as repulsive guidance cues for Unc5-positive neurons. EMBO J *30*, 2920-2933.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics *21*, 3369-3376.

Yin, Y., Li, Y., Kerzic, M.C., Martin, R., and Mariuzza, R.A. (2011). Structure of a TCR with high affinity for self-antigen reveals basis for escape from negative selection. EMBO J *30*, 1137-1148.