

UNIVERSITY OF OXFORD

DOCTORATE OF PHILOSOPHY

Mechanistic Studies of the adhesion-GPCR Latrophilin and its Interactions in Neural Guidance

Author

Verity Amanda JACKSON

Supervisors

Prof. Elena SEIRADAKE

Prof. Mark SANSOM



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

in the

Department of Biochemistry

Magdalen College

Michaelmas 2017

UNIVERSITY OF OXFORD

Abstract

Department of Biochemistry

Doctor of Philosophy

Mechanistic Studies of the adhesion-GPCR Latrophilin and its Interactions in Neural Guidance

by Verity Amanda Jackson
Magdalen College
Trinity 2017

The adhesion GPCRs are a poorly understood and evolutionarily ancient family of cell surface receptors, several of which have emerging functions in the development of the nervous system. aGPCRs comprise a large extracellular domain, providing binding sites for a variety of ligands, alongside a seven transmembrane domain characteristic of GPCRs. It has been proposed that aGPCRs may function as “context-recognisers”, using their large ectodomains to bind different combinations of ligands depending on the molecular make-up of the environment. However there is a lack of direct evidence for this at a molecular level. The ectodomain of one subfamily of aGPCRs, the Latrophilins (Lphns) has been shown to directly interact with several ligands with roles in synaptogenesis and neural guidance. The best-validated of these interactions are those with Fibronectin Leucine-Rich Transmembrane (FLRT) proteins and the Teneurins. In addition, FLRT proteins, also interact with Uncoordinated5 (Unc5) proteins, mediating cell repulsion. Here I reveal that the FLRT-binding site of Lphn is bifunctional, mediating both cell adhesion and repulsion, and that Unc5 is capable of influencing the functional outcome of this interaction. Biophysics and structural studies show that fragments of the Lphn, FLRT and Unc5 ectodomains interact in an unusual and homologue-dependent stoichiometry. Despite the fact that Teneurin interacts with Lphn at a distinct site, Teneurin seems incapable of interacting with the Lphn-FLRT-Unc5 complex, but can form a ternary complex with Lphn and FLRT in the absence of Unc5. Alongside this I present a crystal structure of a large portion of a Teneurin ectodomain, revealing the ancient evolutionary origins of this receptor. Together these data provide strong molecular evidence for a role of Lphns as context-recognisers, by their abilities to bind diverse ligands in distinct combinations and variable stoichiometries.

Acknowledgments

I am hugely grateful for the network of colleagues, collaborators, friends and family who have supported me throughout my D.Phil and helped to make the past four years the most enjoyable of my life.

First and foremost I would like to thank my supervisor, Elena Seiradake, for her never-ending support, encouragement and good ideas. I am deeply appreciative of the training and mentorship she has provided over the last four years, as well as for the opportunities to explore projects independently and determine the direction of my own research. I am also immensely grateful for the support of my secondary supervisor, Mark Sansom, for his sage advice, kind reassurance, and seemingly endless patience.

This project would not have been possible without Maria Carrasquero, a wonderful lab technician and friend. Her assistance with cell culture and stripe assay imaging has been invaluable to the project. I would like to thank collaborators, all of which have contributed greatly to the success of this project. Firstly, Rüdiger Klein (Max Planck Institute of Neurobiology) for welcoming me into his lab to perform the stripe assay experiments in this thesis, and his postdoc Dani del Toro for his teaching, guidance and never-ending stream of ideas. Secondly, I am grateful to Carol Robinson (Chemistry Department) for a productive native mass spectrometry collaboration, and to Shahid Mehmood for acquiring and analysing native mass spectrometry data.

Others I would like to thank are Ed Lowe (Department of Biochemistry), for his help with all aspects of crystallography and for helping me to phase Ten2 Δ^{EGF} ; David Staunton (also Department of Biochemistry) for assistance with SEC-MALS; Matthieu Chavent (previously Department of Biochemistry) for teaching me molecular dynamics simulations and for introducing me to Galette des Rois; Paul Emsley (MRC-LMB), for help with glycan building; Jacqueline Ferralli (Friedrich Miescher Institute), for retrieving the chicken teneurin cDNA and beamline scientist Andrew McCarthy (ESRF). I would also like to acknowledge Rob Gilbert, for his support as my Magdalen advisor. I am grateful for generous funding from the Wellcome Trust.

Over the course of my D.Phil, my friends have been of utmost importance and have filled my time in Oxford with happy memories, both in and out of the lab. In particular I would like to thank Jack, Ines, Luke and Fay, and of course my “Wellfies”, Harriet, Helen and Paul. I would also like to thank the Oxford Sirens, for putting a smile on my face no matter what, and for reminding me of the importance of fun when I forget.

Finally I would like to thank my family, particularly my parents, for their endless belief in my ability and support in everything I do (even if they don’t quite know what that might be). I owe my love of learning and desire to succeed to you, and for that I will be forever grateful.

Contents

1	Introduction	1
1.1	Principles of Neural Guidance and Wiring	1
1.1.1	Attraction and Repulsion in Neural Guidance	2
1.1.2	CAMs in Neural Guidance and Migration	4
1.1.3	CAMs in Synaptogenesis	5
1.2	Adhesion-GPCRs	6
1.2.1	Molecular Organisation of aGPCRs	7
1.2.2	aGPCR Activation and Signalling	13
1.3	aGPCRs in Neural Guidance and Synaptogenesis	16
1.3.1	BAIs	17
1.3.2	CELSRs	17
1.3.3	GPR56	18
1.4	Latrophilins	18
1.4.1	Lphns as Receptors for α -Latrotoxin	19
1.4.2	Expression of Lphns	20
1.4.3	Neuronal Functions of Lphns	20
1.4.4	Lphns in Neurological Diseases	22
1.4.5	Non-neuronal Functions of Lphns	23
1.5	FLRTs	23
1.5.1	The Diverse Functions of FLRTs	24
1.5.2	Homophilic FLRT Interactions	25
1.5.3	Heterophilic FLRT Interactions with Unc5s	25
1.5.4	Other Heterophilic FLRT Interactions	30
1.6	Teneurins	30
1.6.1	Domain Architecture of Teneurins	31
1.6.2	Proteolytic Processing of Teneurins	36
1.6.3	Teneurin Expression Patterns	38
1.6.4	Teneurin Homophilic Interactions	38
1.6.5	Teneurins in Axon Guidance	40
1.6.6	Teneurins at the Synapse	40
1.7	Project Aims	42
2	Materials and Methods	43
2.1	Molecular Biology and Cloning	43
2.1.1	Genes and Constructs	43
2.1.2	Cloning Vectors	44
2.1.3	Polymerase Chain Reaction (PCR)	44
2.1.4	DNA Agarose Gel Electrophoresis	45
2.1.5	Restriction Digestion and Ligation	45

2.1.6	In-Fusion TM Cloning	45
2.1.7	Transformation and Clone Verification	45
2.2	Protein Expression and Purification	46
2.2.1	Cell Lines and Control of <i>N</i> -linked Glycosylation	46
2.2.2	Mammalian Cell Culture Maintenance	47
2.2.3	Small-scale Transient Transfection	47
2.2.4	Polyacrylamide Gel Electrophoresis	48
2.2.5	Western Blotting	48
2.2.6	Roller Bottle Preparation	49
2.2.7	Roller Bottle Transfection	49
2.2.8	<i>In Vivo</i> Biotinylation	49
2.2.9	Purification of His-tagged Proteins	50
2.2.10	Size-Exclusion Chromatography	50
2.3	Biophysical Methods	51
2.3.1	Surface Plasmon Resonance (SPR)	51
2.3.2	Native Mass Spectrometry	52
2.3.3	Size-Exclusion Chromatography-Multi-Angle Light Scatter (SEC-MALS)	55
2.4	X-ray Crystallography	57
2.4.1	Crystallisation and Crystal Optimisation	57
2.4.2	Data Collection	59
2.4.3	Data Processing	60
2.4.4	Phasing	63
2.4.5	Model Building and Refinement	71
2.4.6	Validation	74
2.4.7	Crystallisation and Structure Solution of the Lphn3 ^{Lec-Olf} , FLRT2 ^{LRR} , Unc5D ^{IgIgTsp} Complex	74
2.4.8	Crystallisation and Structure Solution of GgTen2 ^{ΔEGF}	76
2.5	Cell Biology	78
2.5.1	Cell-Binding Assay	78
2.5.2	Stripe Assay	79
3	Interaction of Lphn with FLRT and Unc5	83
3.1	Summary	83
3.2	Surface Conservation of Lphn3 ^{Lec-Olf}	84
3.3	Mutational Disruption of the FLRT-Lphn Interaction	85
3.3.1	Mapping the FLRT-binding Site on Lphn	85
3.3.2	Mapping the Lphn-binding Site on FLRT	88
3.4	Cell Guidance Consequences of the Lphn-FLRT Interaction	90
3.4.1	Investigating the Role of Unc5 in Lphn-FLRT-mediated Cell Adhesion	93
3.5	Lphn-FLRT-Unc5 Ternary Complex Formation	94
3.6	Discussion	98
4	Lphn-FLRT-Unc5 Super-Complexes	101
4.1	Summary	101
4.2	Native Mass Spectrometry of Lphn3-FLRT2-Unc5D Complexes	102
4.3	Crystallisation and Data Collection for Lphn3 ^{Lec-Olf} -FLRT2 ^{LRR} -Unc5D ^{IgIgTsp}	105

4.3.1	Crystallisation	105
4.3.2	Data Collection and Processing	107
4.4	Solving the Lphn3 ^{Lec-Olf} -FLRT2 ^{LRR} -Unc5D ^{IgIgTsp} Structure	108
4.5	Analysis of the Lphn3-FLRT2-Unc5D Octameric Complex Structure	110
4.6	Mutational Validation of Novel Super-Complex Interfaces	115
4.6.1	Disrupting Binding of Lphn3B	115
4.6.2	Disrupting the FLRT2 ^{LRR} -Unc5D ^{Tsp1} Interaction	118
4.7	Conservation of Novel Super-Complex Interfaces	118
4.8	Stoichiometry of Other Lphn-FLRT-Unc5 Complexes	121
4.9	Discussion	125
4.9.1	Three Receptors - At Least Five Possible Complexes	125
4.9.2	Potential Roles of Homologue-Dependent Super-Complexes	126
4.9.3	Lphn Dimerisation	127
5	Purification and Biophysical Characterisation of Teneurin and its Interaction with Lphn	129
5.1	Summary	129
5.2	Bioinformatic Analysis of Teneurin Proteins	130
5.3	Expression and Purification of Teneurin Constructs from HEK293 cells	134
5.4	Functional Validation of Recombinant GgTen2 Proteins	138
5.4.1	SEC-MALS Analysis of Teneurin Multimerisation	138
5.4.2	SPR Analysis of the Interaction of GgTen2 with Lphn	140
5.4.3	GgTen2 Interactions with the Lphn-FLRT-Unc5 Super-Complex	143
5.5	Discussion	145
6	Solving the Crystal Structure of Teneurin	148
6.1	Summary	148
6.2	Crystallisation and Data Collection for GgTen2 ^{ΔEGF}	148
6.2.1	Crystallisation	148
6.2.2	Experimental Phasing of GgTen2 ^{ΔEGF}	151
6.2.3	Solving the GgTen2 ^{ΔEGF} Structure	155
6.2.4	Native Data Processing, Molecular Replacement and Refinement	157
6.3	Analysis of the GgTen2 ^{ΔEGF} structure	161
6.3.1	A Tour of the GgTen2 ^{ΔEGF} Structure: from N- to C-terminus	163
6.3.2	Indications of Functionally Important Surfaces on GgTen2	170
6.3.3	GgTen2 ^{ΔEGF} Lattice Contacts	172
6.4	Discussion	174
7	Discussion and Future Work	177
7.1	Lphns as Context-Recognisers	177
7.2	Lphns and its Complexes at Synapses	179
7.3	Lphn Dimerisation	180
7.4	Lphn Complexes During Evolution	181
7.5	Summary of Findings in this Thesis	182
A	Constructs and Primers	223
B	Pseudo-symmetry of GgTen2^{ΔEGF} Crystals	226

List of Figures

1.1	Long- and Short-Range Attraction and Repulsion	3
1.2	CAMs In Synapse Formation	5
1.3	The Diverse Ectodomains of aGPCRs	9
1.4	Hybrid Architecture of aGPCRs	12
1.5	Allosteric Antagonist Model of aGPCR Activation	14
1.6	Domain Organisation of Vertebrate Lphns	19
1.7	Domain Organisation of FLRTs	24
1.8	Domain Organisation of Unc5s	26
1.9	Crystal Structure of the Minimal FLRT-Unc5 Complex	29
1.10	Literature-based Domain Annotation of Teneurins	31
1.11	Electron Micrographs of MmTen1 Ectodomains	32
1.12	The Crystal Structure of YenB-YenC2	35
1.13	Sequence Similarity between TCAP Sequences and CRFs	37
2.1	Control of <i>N</i> -linked Glycosylation for Crystallography	46
2.2	Graphical Representation of SAD Phasing Equations	70
2.3	Stripe Assay Preparation	80
3.1	The Crystal Structure of Lphn3 ^{Lec-Olf}	85
3.2	SPR analysis of the FLRT2-Lphn3 Interaction and Mutants	87
3.3	Wild type and Mutant Lphn3-FLRT2 Cell-binding Assay	88
3.4	Lphn and Unc5 Bind Distinct Surfaces on the FLRT LRR Domain	89
3.5	The FLRT-binding Site on Lphn3 Regulates Cell Adhesion	91
3.6	The FLRT-binding site on Lphn3 regulates repulsion of cortical neurons	92
3.7	Unc5 Controls FLRT-Lphn-mediated Cell Adhesion	94
3.8	Lphn3 Enhances the Affinity of the Unc5D-FLRT2 Interaction	95
3.9	Unc5D Tsp1 Multimerises the Lphn-FLRT-Unc5 Ternary Complex	97
3.10	Model for Cell Guidance Mediated by Lphn, FLRT and Unc5D	100
4.1	Native MS of Lphn3 ^{Lec-Olf} -FLRT2 ^{LRR} -Unc5D ^{Ig}	102
4.2	Native MS of Lphn3 ^{Lec-Olf} -FLRT2 ^{LRR} -Unc5D ^{Ecto}	104
4.3	Crystals of Lphn3 ^{Lec-Olf} -FLRT2 ^{LRR} -Unc5D ^{IgIgTsp}	106
4.4	Electron Density for the Lphn3 ^{Lec-Olf} -FLRT2 ^{LRR} -Unc5D ^{IgIgTsp} Complex	109
4.5	The Structure of the 4:2:2 Lphn3-FLRT2-Unc5D Complex	111
4.6	Unc5D Bridges Two Pseudo-Symmetric 2:1:1 Tetramers	113
4.7	Structural Rearrangements of Lphn3 ^{Lec-Olf} and UncD ^{IgIgTsp}	114
4.8	The Lphn3 DDD-loop Mediates Binding of Lphn3B	117
4.9	Mutations in Unc5D Tsp1 Abolish Octamer Formation	118
4.10	Sequence Conservation Across the Unc5 and FLRT Protein Families	120
4.11	Lphn3, FLRT3 and Unc5B form a 2:1:1 Tetramer	122

4.12	Octamerisation is Restricted to FLRT2	123
4.13	Lphn1 Can Form 2:1:1 Tetrameric and 4:2:2 Octameric Complexes . .	124
4.14	The Diversity of FLRT, Unc5 and Lphn Complexes	125
4.15	Differential Expression of FLRT and Unc5 Homologues in Development	126
4.16	The Lphn2 “DDD loop” Contains a N-linked Glycosylation Site . . .	128
5.1	Literature-based Domain Annotation of the Teneurin Ectodomain . .	130
5.2	Secondary Structure Prediction for GgTen2 Residues 908-1198	131
5.3	Summary of Homology Models Built for the Teneurin Ectodomain . .	132
5.4	Multiple sequence alignment of HsTen1/2/3/4 C-terminal Regions. . .	133
5.5	Expression and Purification of Teneurin Ectodomains	135
5.6	Initial Constructs of the GgTen2 Lphn-binding Region.	136
5.7	Optimised Constructs of GgTen2 Lphn-binding Region	137
5.8	Purification of GgTen2 ^{ΔEGF}	137
5.9	SEC-MALS Analysis of GgTen2 Dimerisation	139
5.10	GgTen2 Binds to the Lphn2 Lec Domain	141
5.11	Binding of GgTen2 ^{Ecto} and GgTen2 ^{ΔEGF} to Lphn	142
5.12	FLRT and Teneurin bind non-competitively to Lphn	143
5.13	Teneurin is Excluded from the Lphn3-FLRT2-Unc5D Super-Complex	144
6.1	Crystals of GgTen2 ^{ΔEGF}	150
6.2	Diffraction of GgTen2 ^{ΔEGF} Crystals	151
6.3	GgTen2 ^{ΔEGF} Derivative Data Collection	153
6.4	Anomalous Signal Against Resolution for Pt-derivative Crystals . . .	154
6.5	Initial Model Building of GgTen2 ^{ΔEGF}	156
6.6	Packing of GgTen2 ^{ΔEGF} in the P2 ₁ Asymmetric Unit	159
6.7	Final Electron Density for GgTen2 ^{ΔEGF}	160
6.8	Overall Architecture of GgTen2 ^{ΔEGF}	162
6.9	Architecture of the FN-plug Domain	164
6.10	The NHL Domain is an Elaborate Six-bladed β -propellor	165
6.11	The YD-barrel of GgTen2 Resembles those of Bacterial Toxins.	166
6.12	The YD-barrel Encloses Part of the GgTen2 C-terminus	168
6.13	Similarity to Bacterial Antibiotic-binding Proteins	169
6.14	Structural Similarity of the Tox-GHH Domain to Colicin DNases . . .	170
6.15	Conserved Surface Patches on GgTen2 ^{ΔEGF} Contain Cancer Mutations	171
6.16	A Conserved Surface on GgTen2 ^{ΔEGF} Mediates Crystal Contacts. . .	172
6.17	An alternatively spliced loop of GgTen2 ^{ΔEGF} mediates crystal contacts	173
6.18	The Hybrid Architecture of the GgTen2 Ectodomain	175
6.19	Bacterial Catalytically-dead Tox-GHH Domains	176
B.1	GgTen2 ^{ΔEGF} Crystal Packing	227
B.2	Alternative Crystal Packing of the GgTen2 ^{ΔEGF} C-terminal Domain .	229

List of Tables

1.1	Extracellular Ligands of aGPCRs	10
2.1	Summary of cDNAs Used for Sub-cloning	43
2.2	Ratios of Reagents Used for Small-Scale Mammalian Cell Culture . .	47
2.3	Native MS Sample Preparation	54
2.4	Molecular Weights of <i>N</i> -linked Glycans Used for $\frac{dn}{dc}$ Adjustment . . .	56
4.1	Lphn3 ^{Lec-Olf} -FLRT2 ^{LRR} -Unc5D ^{IgIgTsp} Crystallisation Conditions. . . .	107
4.2	GgTen2 ^{ΔEGF} Native Data Collection	107
4.3	Data Reduction Statistics for the Lphn3-FLRT2-Unc5D Complex . .	108
4.4	Refinement Statistics for the Lphn3-FLRT2-Unc5D Complex	109
5.1	Domain Boundaries Predicted by BLASTp and Homology Modelling.	132
5.2	cDNAs used for Expression Testing of the Teneurin Ectodomain . . .	134
6.1	GgTen2 ^{ΔEGF} Native Data Collection	151
6.2	Data Reduction Statistics for GgTen2 ^{ΔEGF} Pt-derivative Crystals . .	155
6.3	Data Reduction Statistics for GgTen2 ^{ΔEGF} Native Crystals.	158
6.4	Refinement Statistics for GgTen2 ^{ΔEGF}	160
A.1	Constructs Cloned by Others and Used in this Thesis	223
A.2	Mutant Constructs Cloned as Part of this Thesis	224
A.3	Teneurin and Lphn Constructs Cloned as Part of this Thesis.	225
B.1	Crystallographic Symmetry of GgTen2 ^{ΔEGF} Crystals	226

Abbreviations

7TM	7 Transmembrane region
ABD	Antibiotic-binding domain
aGPCR	Adhesion- G -Protein-Coupled Receptor
BAI	Brain-specific Angiogenesis Inhibitor
BFPP	Bilateral Frontoparietal Polymicrogyria
CAM	Cell Adhesion Molecule
CELSR	Cadherin EGF LAG Seven-pass G-type Receptor
CTF	C -Terminal Fragment
DCC	Deleted in Colorectal Cancer
DMEM	Dulbecco's Modified Eagle's Medium
EGF	Epidermal Growth Factor-like
ES	Electrospray
FBS	Foetal Bovine Serum
FN	Fibronectin
GAIN	GPCR Autoproteolysis Inducing Domain
GPS	GPCR Proteolysis Site
HA	Heavy Atom
Horm	Hormone Receptor Domain
Ig	Immunoglobulin-like
LLG	Log-Likelihood Gain
Lphn	Latrophilin
LRR	Leucine-Rich Repeat
MR	Molecular Replacement
MS	Mass Spectrometry
NCS	Non-Crystallographic Symmetry
NHL	NCL-1 HT2A Lin-41
NTF	N -Terminal Fragment
Rhs	Rearrangement Hotspot
SAD	Single Anomalous Dispersion
SEC-MALS	Size-exclusion Chromatography-Multi-Angle Light Scatter
SPR	Surface Plasmon Resonance
TCAP	Teneurin C -terminal Associated Peptide
Ten	Teneurin
TFZ	Translation Function Z -score
TLS	Translation Libration Screw
tNCS	Translational Non-Crystallographic Symmetry
Tsp	Thrombospondin-like
TTR	Transthyretin
PDZ	PSD95 DLG1 ZO1
PBM	PDZ-Binding Motif
SHANK	SH3 and multiple Ankyrin repeat domains
Unc5	Uncoordinated5
YD	Tyrosine-Aspartate

Chapter 1

Introduction

The brain is the most complicated and poorly understood organ. It must take information from a wide variety of sensory inputs, integrate it with other sensory inputs or stored information, and then elicit the correct physiological or behavioural response. For these functions to occur with high fidelity, brain development requires the correct positioning of millions of neurons alongside the formation of the correct number and type of synaptic connections with appropriate partners.

1.1 Principles of Neural Guidance and Wiring

Initial neural wiring occurs through a combination of guidance events, which are independent of neuronal activity. The idea that axons extend in a directed manner and are actively guided to their targets was first predicted in 1892 by Ramón y Cajal in his *Histology of the Nervous System of Man and Vertebrates* [207], where he writes:

“If one admits that neuroblasts are endowed with chemotactic properties, then one might also imagine that they are capable of ameoboid movements, initiated by factors secreted from epithelial, neural, or mesodermal elements. As a result, their processes may be oriented in the direction of chemical gradients, and thus guided to the secreting cells”

The general applicability of this idea was reaffirmed by Sperry’s 1963 “Chemoaffinity Hypothesis” [244]. Based on experiments in the visual system of amphibians, he hypothesised that unique chemical “labels” specify individual neuronal populations and also provide positional information, potentially in the form of gradients. The

basic notion of the Chemoaffinity Hypothesis has become dogma in developmental neurobiology [168]. Inspired by the ideas of Cajal and Sperry, alongside the advent of modern biochemical, genetic and molecular biological techniques, a large number of these so-called “labels” have now been identified, revealing a diverse selection of neural guidance molecules, comprising both soluble and membrane-bound factors, which can also be components of the extracellular matrix [122]. A handful of these, and the first to be identified, are known as the “Canonical” Neural Guidance Cues: the Slits, Netrins, Ephrins and Semaphorins. In this section I will use well-studied receptor systems as examples to introduce some of the principles of neural guidance and wiring relevant to the data presented in this thesis.

1.1.1 Attraction and Repulsion in Neural Guidance

In order to respond appropriately to this diverse range of factors, developing neurons have specialised, dynamic structures called growth cones at the tips of their axons. These growth cones are capable of sensing a diverse range of cues, integrating positional information and then responding accordingly, allowing them to accurately navigate towards their targets. Neural guidance relies on a complex combination of cell attraction and repulsion to guide neurons toward their correct synaptic partners. Attractive molecules can act in either a long- or short-range manner, steering the growth cone towards its correct target (Figure 1.1). One example of this would be in the embryonic development of the spinal cord, where the long-held view is that Netrin1 is secreted by the ventral midline floor plate promoting the outgrowth of commissural neurons [229, 113]. However, recently this view has been challenged by the finding that a *Netrin1* conditional knockout mouse shows normal development of hindbrain and spinal cord commissural neurons, and it has instead been suggested that Netrin1 primarily acts as a short-range factor, promoting growth cone adhesion [56]. In contrast, repulsive guidance cues act to prevent growth cone advancement, exemplified by the canonical role of Slit in preventing midline crossing [115].

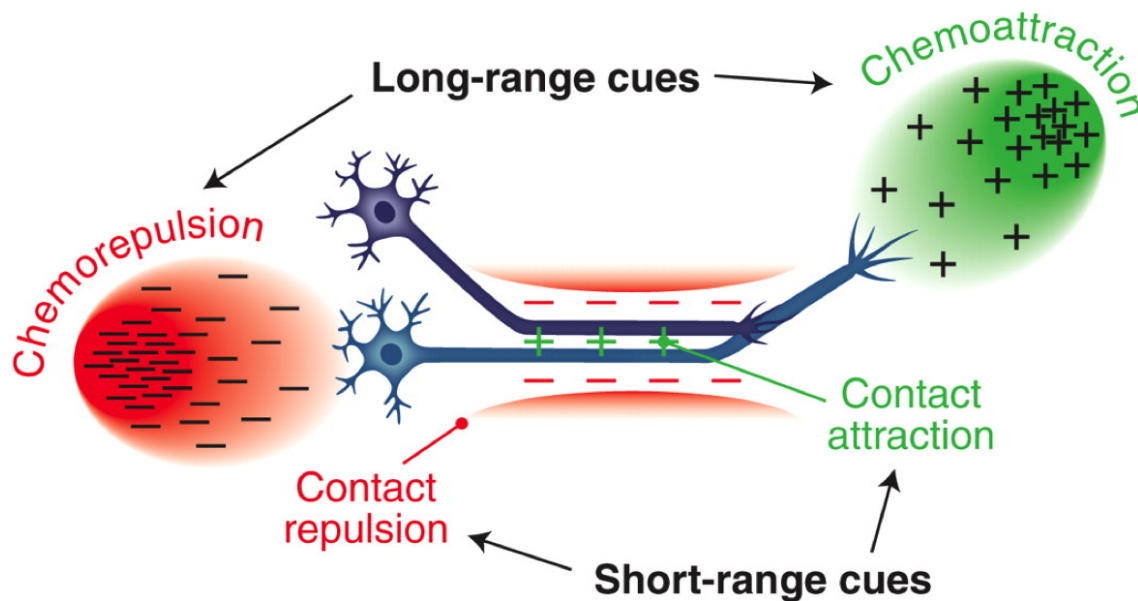


Figure 1.1: **Long- and Short-Range Attraction and Repulsion.** The effects of long- and short-range repulsion in neural guidance. Figure taken from [122].

It was originally thought that individual classes of guidance cues always functioned in either an attractive or repulsive manner. This, along with the idea that different subpopulations of neurons express unique chemical labels, would require an extraordinarily large number of individual guidance cues. It is becoming clear that the limited number of neural guidance molecules function in a highly context-specific fashion, forming heteromeric multi-subunit receptor complexes and modulating their signals according to the combination of extracellular factors and other receptors present [59, 254]. This cooperation between the factors allows the limited number of neural guidance molecules to exert a diverse range of effects, increase fidelity and reduce noise in guidance decisions [59].

As well as functional cooperativity, the involvement of factors not traditionally considered neural guidance molecules also provides a mechanism by which functional complexity can be increased. For example, the Wnt and Hedgehog morphogens (which were initially identified for their roles in tissue patterning during embryonic development) are now known to also function in axon guidance [284]. Another class of molecules which are not considered canonical neural guidance molecules, but play

an undeniably important role in neural guidance, migration and synaptogenesis are the Cell Adhesion Molecules.

1.1.2 CAMs in Neural Guidance and Migration

Cell Adhesion Molecules (CAMs) are transmembrane receptors which interact with components of the extracellular matrix or molecules expressed on the surfaces of other cells. They promote adhesion by providing an indirect physical link between the extracellular and intracellular environments. These interactions provide spatial stability to stationary cells and traction for the migrating growth cone [241]. In neural guidance, CAMs function either homophilically (binding to molecules of the same class, e.g. Cadherins) or heterophilically (binding a ligand on another cell or in the extracellular matrix, e.g. Integrins). One school of thought is that CAMs generally act in a permissive manner by providing a substrate to promote neurite outgrowth, rather than actively inducing growth cone turning [149]. However, CAMs often possess signalling activity alongside physical linkage, and the idea is emerging that they might regulate attraction or outgrowth by functioning as signalling molecules, often through heterophilic interactions [122]. This is exemplified by the growing body of data indicating that the canonical neural guidance factors, the Netrins, Semaphorins, Slits and Ephrins can regulate cell motility by modulating Integrin-mediated adhesion [180]. For example, Sema7a has been shown to promote axon outgrowth by interacting with Integrins and activating the mitogen-activated protein kinase pathway [198].

Although CAMs play important roles in neural guidance, they do not always play adhesive roles. For example, in *Drosophila*, Down Syndrome Cell Adhesion Molecule 1 (DSCAM1) can exist as over 19,000 different isoforms [221]. Each isoform can bind homophilically, but shows little binding to non-self isoforms. Although classically considered a CAM, homophilic DSCAM interactions are actually repulsive and are important in axonal and dendritic self-avoidance [281, 90].

1.1.3 CAMs in Synaptogenesis

After neurons have been guided to their correct positions, their next task is to establish synaptic connections with the proper partners, a process which relies heavily on CAMs interacting *in trans* across the cleft (Figure 1.2). Synapse formation begins with target selection, where initial motile dendritic spines are extended to locate their presynaptic partners. Synapses are then formed by the recruitment of further CAMs and intercellular contacts stabilised by remodelling of the cytoskeleton. All of these steps require cell-cell recognition by extracellular domains of CAMs expressed on presynaptic and postsynaptic membranes [79].

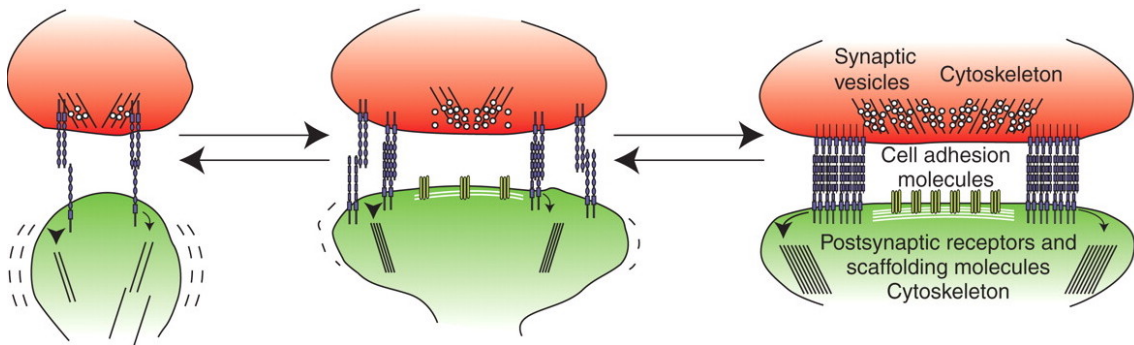


Figure 1.2: **CAMs In Synapse Formation.** From left to right: Target selection occurs when dendrites extend, searching for their presynaptic partners. Initial contact is made by CAMs. As the synapse matures, further CAMs are recruited and the pre- and post-synapse differentiate. All of these steps are reversible - a phenomenon known as synaptic plasticity which is outside the scope of this thesis. Figure adapted from [79].

CAMs play a hugely important role at each of these steps, where they can act in a solely adhesive manner (stabilising the contact between and aligning the two membranes), or they can act as synaptic organisers, promoting pre- and post-synaptic differentiation. Synaptic organiser proteins, such as the well-characterised Neurexins, are often capable of binding to several distinct ligands [234]. Neurexins are best studied for their interactions with Neuroligins, both of which can be considered CAMs. Overexpression of either Neurexins or Neuroligins in heterologous cell lines induces postsynaptic and presynaptic differentiation respectively [220, 82, 181], suggesting that this interaction is sufficient for synapse formation. Neuroligin-mediated postsynaptic differentiation involves clustering of the postsynaptic scaffolding pro-

tein, PSD95 and glutamate receptors, the latter requiring the intracellular PDZ (PSD95-DLG1-ZO1)-binding motif (PBM) of Neuroligin [40]. Although sufficient to induce synapse formation, knockout of either Neuroligin or Neurexin in mice does not affect synapse number or morphology, suggesting that these proteins are not necessary *in vivo* for synapse formation [171, 263]. This observation may be due to redundancy with the vast array of other synaptogenic CAM proteins, or might be indicative that the Neurexin-Neuroligin interaction is necessary at a later developmental stage [79]. These data demonstrate that CAMs provide much more than a physical link across the synapse, and can possess diverse and potent signalling functions in synaptogenesis.

Alongside the CAMs, another class of molecules with well-established cell adhesive properties and roles at all stages of neural development are the adhesion-G-protein-coupled receptors (aGPCRs) [134]. The functions of this important class of molecules during nervous system development was first revealed by the fact that mutations in the aGPCR *GPR56* cause a human brain cortical malformation called bilateral frontoparietal polymicrogyria (BFPP) [201]. Since then, a wealth of information has become available linking the aGPCRs to diverse aspects of neural development. The aGPCRs are discussed in detail in the next section.

1.2 Adhesion-GPCRs

The aGPCRs form one of the five subfamilies of the GPCR superfamily [74]. There are 33 human aGPCRs, divided into nine subfamilies, making it the second largest subfamily of GPCRs (behind the rhodopsin GPCRs) [23]. They are evolutionarily ancient, with multiple orthologues in invertebrates and are thought to be ancestral to the Secretin subfamily of GPCRs [185]. aGPCR nomenclature is diverse for historic reasons, although recently a unified naming system was proposed in which the prefix ADGR (**A**dhesion **G**-protein-Coupled **R**eceptor) is followed by a letter denoting which of the nine subfamilies the aGPCR belongs to, and a number based

on the historic naming system [86]. In this thesis, the historic nomenclature is used, but the new names are shown alongside the historic names for each aGPCR in Figure 1.3.

1.2.1 Molecular Organisation of aGPCRs

aGPCRs are distinct from other GPCRs largely due to their large and highly variable extracellular regions containing arrays of domains typically associated with cell adhesive functions [91]. Like all GPCRs, they also contain seven transmembrane regions (7TMs). Their hybrid architecture, which will be discussed in the following sections, is likely key to their diverse range of functions.

N-terminal Cell-Adhesion Domains

The N-terminal extracellular regions of aGPCRs are highly variable, ranging from non-existent in GPR114 to ~5600 residues in GPR98 (otherwise known as Very Large GPCR1, VLGR1). aGPCRs belonging to the same subfamily tend to have similar types and arrangements of cell adhesion domains in their extracellular regions (Figure 1.3). Note that in Figure 1.3, large parts of some aGPCR ectodomains have no annotated fold. Recently the ectodomain of the aGPCR GPR56 was structurally characterised, revealing a previously unannotated domain of novel fold at its N-terminus. This domain was termed the PLL (Pentraxin/Laminin/Neurexin/sex-hormone-binding-globulin-like) domain [219]. It remains to be seen whether other aGPCRs also contain domains of novel fold in their ectodomains. For other classes of receptors, the types of domains found in aGPCR ectodomains typically provide sites of protein-protein interactions, hence it was reasonable to expect the same for the aGPCRs.

In the mid-1990s, CD55/DAF was identified as the first ligand for an aGPCR, CD97 [88]. Since then, subsequent studies have identified many more factors, including proteins, glycosaminoglycans and lipids that are capable of interacting with

the ectodomains of various aGPCRs (summarised in Table 1.1). Although ligands have now been identified for the majority of aGPCR families, many aGPCRs remain orphan. In line with their diverse architecture, aGPCRs interact with ligands regulating a broad range of functions, from cell adhesion and synaptogenesis, to immunity and apoptosis. The ability of aGPCRs to bind many, often seemingly unrelated, ligands suggests that these receptors may be able to act as so-called “context-recognisers”, tuning their signalling depending upon the precise combination of extracellular factors available [133].

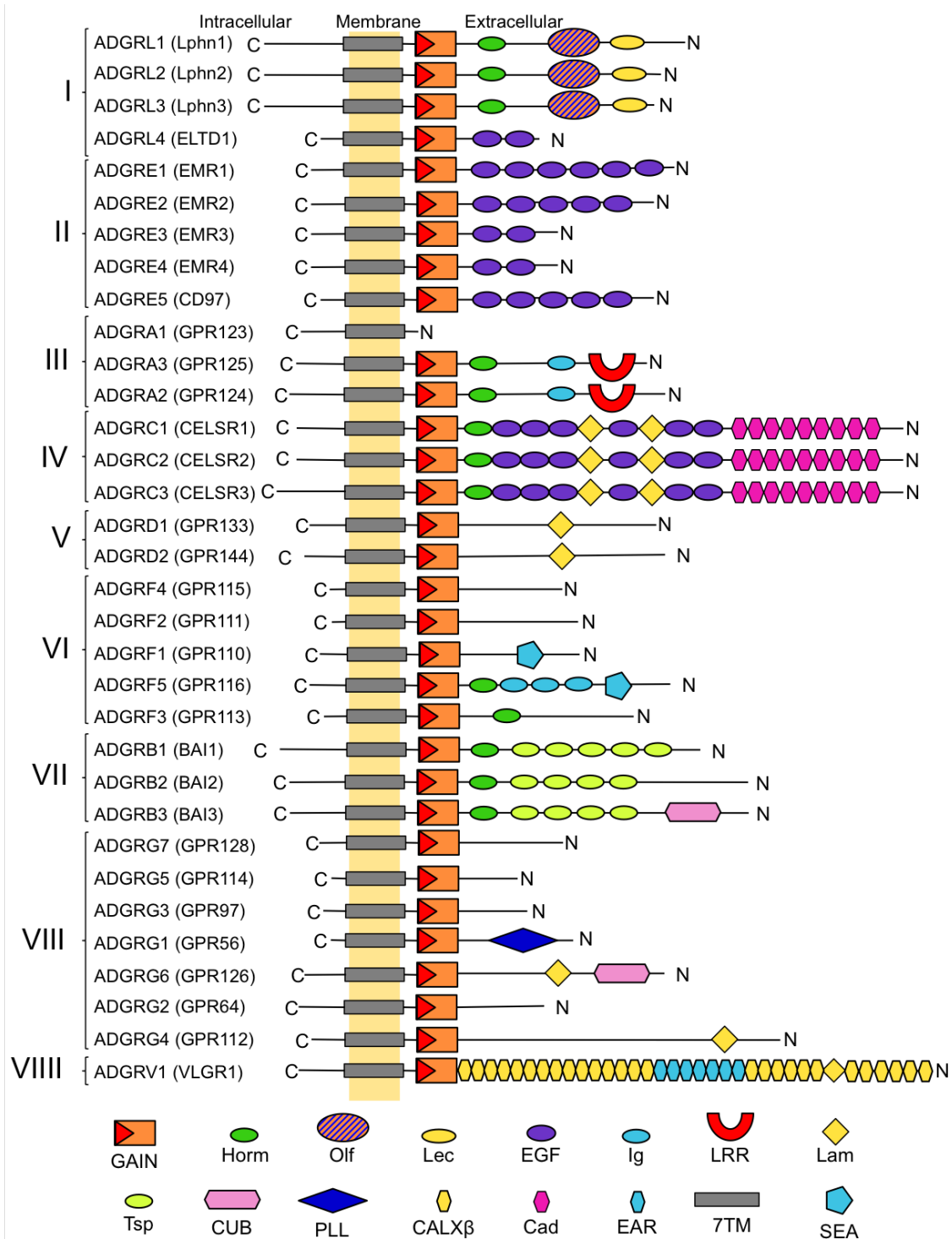


Figure 1.3: The Diverse Ectodomains of aGPCRs. The annotated domains for all aGPCR ectodomains are shown schematically. aGPCRs are ordered by subfamily, and subfamilies numbered using roman numerals. Regions with no predicted fold are shown as black lines. Horm, Hormone receptor motif; Olf, Olfactomedin; Lec, Lectin; EGF, Epidermal Growth Factor-like; Ig, Immunoglobulin-like; LRR, Leucine-Rich Repeat; Lam, Laminin G; Tsp, Thrombospondin; CUB, Cs1 and Csr/Uegf/BMP1; PLL, Pentraxin/Laminin/Neurexin/sex-hormone-binding-globulin-like [219]; CALX β , Calnexin- β ; Cad, Cadherin-like repeat; EAR, Epitempin/Epilepsy-Associated repeat; 7TM, 7 Transmembrane Region; SEA, Sperm protein, Enterokinase, Agrin module. Figure recreated from [86]. aGPCRs are referred to by the new nomenclature proposed in [86], alongside their alternative names.

aGPCR Domains Involved		Proposed Function
Lphn1/2/3		
Contactin6 [294]	Ectodomain	Anti-apoptotic
Teneurins[237]	Lec/Lec-Olf [192, 28]	Transsynaptic receptor pair. Activates Ca^{2+} signalling.
FLRTs[191]	Olf	Transsynaptic receptor pair. Regulates glutamatergic synapse development
Neurexin1s[27]	Lec-Olf-Stalk	Transsynaptic cell adhesion
FMRFamide-like neuropeptides [177]	Ectodomain	Regulation of pharyngeal pumping in nematodes
CD97		
CD55/DAF [88]	EGF1+EGF2 [87]	Cell adhesion
Chondroitin sulphate [245]	EGF4	Interaction with the extracellular matrix
Integrin $\alpha_5\beta_1/\alpha_v\beta_3$ [271]	Ectodomain	Chemoattraction and invasion
CD90/Thy-1 [268]	Ectodomain	Leukocyte recruitment to sites of inflammation
EMR2		
Chondroitin sulphate [245]	EGF4	Cell attachment
GPR124		
Integrin $\alpha_v\beta_3$ [262]	Horm	Promotion of endothelial cell survival
Glycosaminoglycans [262]	Ectodomain	Promotion of endothelial cell survival
GPR116		
Sp-D [75]	Ectodomain	Regulation of lung surfactant secretion and uptake
BAI1		
Phosphatidylserine [197]	Tsp repeats	Recognition and engulfment of apoptotic cells
Lipopolysaccharide [46]	Tsp repeats	Recognition and engulfment of Gram-negative bacteria
CD36 [111]	NTF	Inhibition of endothelial cell migration
Integrin $\alpha_v\beta_3$ [121]	Tsp repeats	Pro-apoptotic
BAI3		
C1qs [24]	Tsp repeats	Regulation of synapse formation and maintenance
GPR56		
Tissue Transglutaminase 2 [282]	Ectodomain	Suppression of tumour growth and metastasis
CollagenIII [160]	Ectodomain	Inhibits neuronal migration
GPR126		
Laminin-211 [200]	Stalk region	Regulation of Schwann cell development
CollagenIV [193]	CUB and Pentraxin domains	Regulation of myelination in ear development
Prion Protein (Prp ^C) [128]	Ectodomain	Promotion of Schwann Cell Myelination

Table 1.1: **Extracellular Ligands of aGPCRs.** A summary of the known extracellular ligands for aGPCRs. The table is based on that presented in [136], but corrected for various omissions and for the inclusion of more recent data. The table is coloured according to the subfamily of each aGPCR: Family I, green; Family II, orange; Family III, pink; Family VI, pale yellow; Family VII, blue; Family VIII, lilac. Where the middle column reads “ectodomain”, no experimental evidence was presented for the aGPCR subdomain involved in the interaction.

The GPCR-Autoproteolysis-Inducing (GAIN) Domain

The diverse range of aGPCRs are united by the almost universal presence of the membrane-proximal GAIN domain (present in all aGPCRs except GPR123) [11] (Figure 1.3). This domain contains the GPCR Proteolysis Site (GPS, ~ 40 amino acids), which is located immediately before the first transmembrane helix and undergoes post-translational, autoproteolytic cleavage in the endoplasmic reticulum [125, 150]. The crystal structure of the GAIN domain (Figure 1.4) revealed that the autoproteolysed scissile bond has a distorted geometry due to its position in a strained loop between the last two β -strands of the GAIN domain. After cleavage, the cleaved β -strand remains tightly associated with the rest of the GAIN domain due to an extensive network of intrastrand hydrogen bonds and side chain interactions [11]. This leads to the cleaved N-terminal fragment, NTF, remaining tightly associated with the rest of the protein (the C-terminal fragment, CTF) and the cell surface (Figure 1.4). Although the GAIN domain is present in almost all aGPCRs, structural variations of it exist, as exemplified by the recently solved crystal structure of the GPR56 ectodomain. The GPR56 GAIN domain lacks the three-most N-terminal helices present in the GAIN domains of Latrophilin (Lphn) 1 and Brain-specific Angiogenesis Inhibitor (BAI) 3 [11, 219].

Alongside the GAIN domain, the hormone receptor domain (Horm) is the second most common domain in the ectodomains of aGPCRs, and is found in 12 out of 33 family members [136] (Figure 1.3). The crystal structures of the Lphn1 and BAI3 Horm domains are highly structurally similar, despite low sequence identity [11]. In the available crystal structures of Horm-GAIN domains, the GAIN domain is blocking the putative hormone-binding site of the Horm domain [11, 136]. The physiological relevance of this domain is poorly understood as there are currently no known hormone binding partners of aGPCRs.

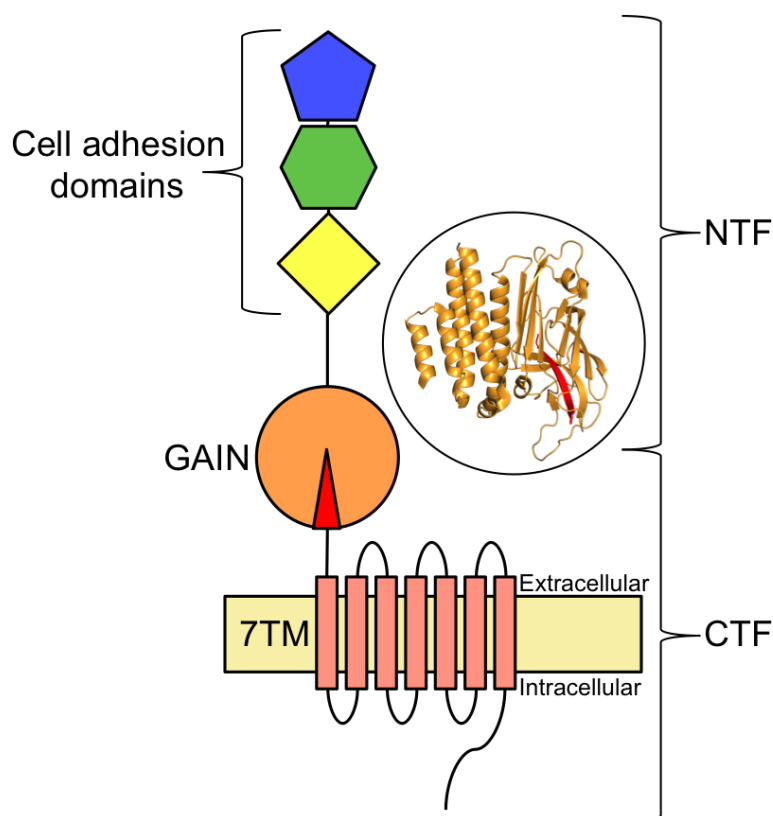


Figure 1.4: **Hybrid Architecture of aGPCRs.** A schematic depicting the generalised hybrid architecture of aGPCRs. The GAIN domain is depicted as an orange circle with the cleaved portion depicted in red. A cartoon view of the crystal structure of the cleaved GAIN domain of Lphn1 is shown alongside. The cleaved β -strand is coloured in red [11].

7TM

Like all GPCRs, aGPCRs have a 7TM region. Overall this region is poorly understood for aGPCRs, but shares the highest overall sequence similarity with the 7TM regions of the secretin GPCR subfamily, sharing almost all highly conserved residues. Homology modelling and sequence alignments have identified that some important features of the 7TM regions of the rhodopsin- and secretin-type GPCRs are approximately conserved in aGPCRs [183]. Interestingly, the aGPCR subfamily classification is based on sequence similarity of the 7TM, but incidentally this results in subfamilies which also share similar organisation of their extracellular domains. This suggests an evolutionarily conserved relationship between the extracellular and 7TM regions. [134] (Figure 1.3).

1.2.2 aGPCR Activation and Signalling

The Tethered Agonist and Allosteric Antagonist Models

The first insights into how the extracellular region of aGPCRs regulates their downstream signalling functions came from the observation that mutant receptors lacking an NTF displayed constitutive signalling activity [189, 194, 286], suggesting that either the extracellular region inhibits the constitutively active CTF, or a remaining part of the extracellular region is agonistic to the 7TM. Using an *in vivo* transgene assay in *C. elegans*, it was suggested that the GPS interacts with the 7TM domain, but that GPS cleavage is not necessary for signalling [205]. Later this idea was substantiated for two aGPCRs, GPR126 and GPR133, by the finding that the extracellular region of aGPCRs harbours a tethered agonist sequence between the natural GPS cleavage site and the N-terminus of the first transmembrane helix (representing the last β -strand of the GAIN domain) [11]. Indeed, synthetic peptides based on this region were able to activate these receptors and the downstream adenylyl cyclase pathway. This sequence has been termed the *Stachel* [148]. *Stachel* sequences have now been identified for several other aGPCRs [246, 178, 51], including GPR114, which is not cleaved at its GPS [277]. Based on these findings, the Tethered Cryptic Agonist Model was proposed, whereby the intact aGPCR is held in an inactive state due to sequestration of the *Stachel* by the GAIN domain. Ligand binding was proposed to cause NTF dissociation, liberating the *Stachel* and allowing it to bind and activate the 7TM.

However, a recent study showed that truncating the entire extracellular region, including the *Stachel* of BAI1, resulted in robust signalling activity in a large range of GPCR signalling assays. For GPR56, the same mutation reduced the activity of the receptor in some signalling assays, but not others [117]. This suggests that aGPCR signalling can proceed via both *Stachel*-dependent and *Stachel*-independent mechanisms, and leads to the current model for aGPCR activation – the Allosteric Antagonism Model. In this model, ligand binding induces a conformational change

in the *Stachel*, inducing *Stachel*-dependent signalling. Meanwhile the NTF directly antagonises the constitutive *Stachel*-independent activity of the 7TM. Only when the NTF becomes dissociated (as discussed in the next section) is this antagonism released, activating both the stalk-dependent and stalk-independent signalling functions of the 7TM (Figure 1.5).

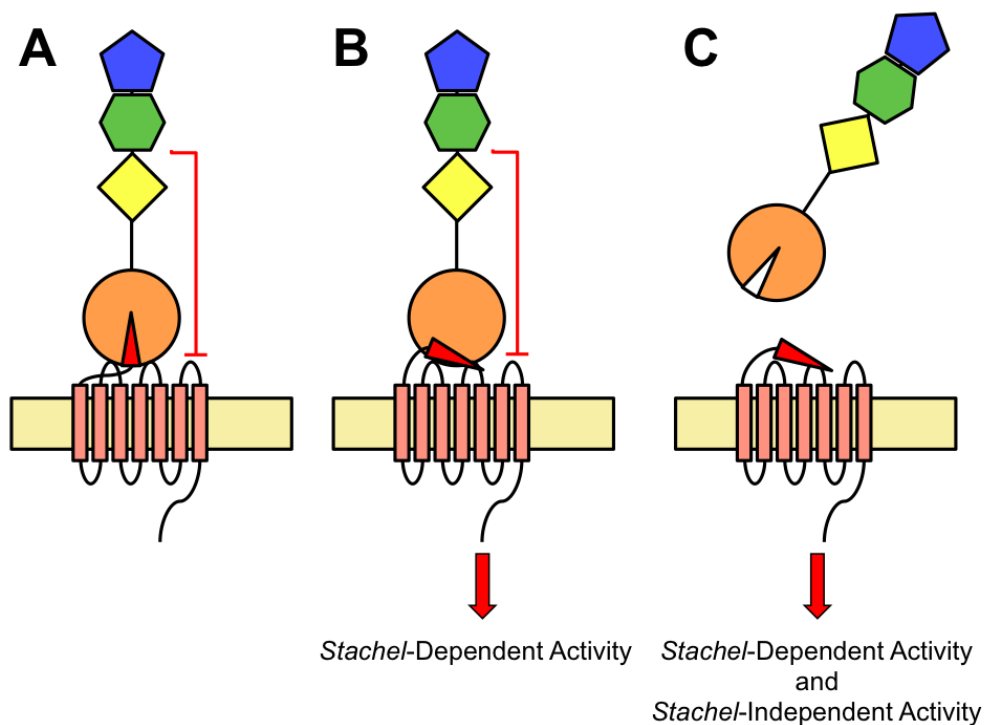


Figure 1.5: **Allosteric Antagonist Model of aGPCR Activation.** In the unliganded receptor the NTF acts to mask the *Stachel* and also directly antagonises the *Stachel*-independent function of the 7TM (A). Upon ligand binding, conformational changes in the NTF enhance *Stachel*-dependent activity (B). Finally, ligand dissociation removes the antagonistic NTF, activating *Stachel*-independent pathway (C).

Proposed *Stachel* Release Mechanisms

How exactly the *Stachel* endogenously activates aGPCRs in the context of the full-length receptor is currently under debate. In the available structures of the GAIN domain the *Stachel* sequence is embedded firmly within the domain, where its movement is presumably restricted by the large number of interactions it makes with the rest of the domain [11]. One of the key questions then is how the *Stachel* could become available to act as an agonist.

Under conditions of sheer stress, engagement of the aGPCR CD97 with its ligand, CD55, results in dissociation of the CD97 NTF from the cell surface [110]. Similarly, polymerised Laminin-211 has been shown to activate GPR126 by direct binding to its stalk region, but only under shaking conditions. It was proposed that mechanical stimulation, alongside ligand binding, may facilitate GPR126 NTF removal and activate signalling [200].

There are multiple lines of evidence indicating that aGPCR signalling is sensitive to mechanical stimulation *in vivo*. The *Drosophila* orthologue of Lphn, CIRL, was shown to be important for determining sensitivity to mechanical stimuli by regulating ionotropic receptor currents. This suggests that CIRL controls the sensitivity of cells towards mechanostimulation, rather than mediating a direct response to mechanical challenge [224, 225]. This response may be regulated by the length and tensile properties of the extracellular region, and requires an intact *Stachel* sequence. Interestingly, GPS cleavage was not required [225]. Mechanosensitive activity has also been proposed for GPR56 in regulating muscle hypertrophy [276]. Further, a missense mutation in the GAIN domain of EMR2 is associated with the disease Vibratory Urticaria, where sufferers develop hives in response to repetitive mechanical stimulation of the skin. This mutation is thought to destabilise the interaction between the NTF and CTF [29]. Together, these data are beginning to suggest that mechanical stimulation may provide a means by which the NTF and CTF dissociate, releasing the *Stachel* peptide.

Downstream Signalling of aGPCRs

Canonical GPCRs mediate many of their effects by coupling to heterotrimeric G-proteins, composed of α (the GTPase), β and γ subunits. There are four classes of $G\alpha$ proteins, each of which mediates distinct intracellular responses upon signalling: $G\alpha_s$, which activates cAMP production via adenylate cyclase; $G\alpha_i$ which inhibits adenylate cyclase; $G\alpha_q$, which activates phospholipase C and $G\alpha_{12/13}$ which acti-

vates small Rho-type GTPases that regulate the cytoskeleton. aGPCRs have been shown to couple to all of these classes [86]. Despite this, very few “activating” ligands have been discovered for aGPCRs, and even fewer of those have been shown to act via G-proteins. For example, type-III Collagen binds GPR56, activating RhoA via $G\alpha_{12/13}$ [160]. GPR126 represents one of the best-studied aGPCRs in terms of its activating ligands, of which there are at least three known examples. Both Collagen IV and the Prion Protein can activate GPR126, causing cAMP elevation. This effect is mediated by a moderately conserved cluster of residues present in both Collagen IV and the Prion Protein [193, 128]. Another GPR126 ligand, Laminin-211, polymerises to promote GPR126 signalling via $G\alpha_s$, elevating intracellular cAMP levels, but can also antagonise cAMP elevation in a $G\alpha_i$ -independent fashion [200].

As well as G-proteins, aGPCRs have been shown to have a variety of C-terminal interactors, and it is evident that their signalling is not always heterotrimeric G-protein-dependent. Both BAI1 and BAI3 bind the intracellular adaptor protein ELMO/Dock180. In the case of BAI1, the interaction with ELMO/Dock180 regulates the actin cytoskeleton via Rac and promotes the engulfment of apoptotic cells recognised by BAI1 binding to phosphatidylserine [197, 137]. aGPCRs have also been shown to interact with intracellular components of morphogen signalling pathways. For example, CELSR1 (cadherin EGF LAG seven-pass G-type receptor 1) interacts with components of the planar cell polarity pathway to promote neural tube closure [184]. Further, almost half of all aGPCRs have a PBM at their C-terminus, allowing the binding of PDZ domain-containing proteins [136], which play important roles as scaffold proteins in the assembly of multi-protein signalling complexes [89].

1.3 aGPCRs in Neural Guidance and Synaptogenesis

aGPCRs play important roles in cell adhesion and migration in various organ systems, including the brain. The identification of GPR56 mutations as drivers in

BFPP was just the beginning of their diverse roles in central nervous system development in both vertebrates and invertebrates [134]. Below, the aGPCRs with the best characterised roles in axon guidance and synaptogenesis are discussed, alongside their involvement in neurological diseases.

1.3.1 BAIs

In the context of nervous system development, BAIs, particularly BAI3, are best studied for their roles in mediating synapse development and maintenance in vertebrates. In the developing cerebellum, BAI3 is expressed on postsynaptic Purkinje cells, where it binds C1q-like proteins (C1ql1-4) expressed by climbing fibres. Initially, many climbing fibres innervate a single Purkinje cell soma, but eventually only a single synapse is formed by the so-called “winning” climbing fibre. C1ql1-BAI3 signalling is a key regulator in this process, and is also involved in the maintenance of the final synapse throughout adulthood [24, 236, 106]. C1ql3 binds the thrombospondin-like (Tsp) repeats of BAI3 [24], but C1ql1 binds to the BAI3 CUB domain [106], suggesting that C1ql family proteins could bind at homologue-specific binding sites on BAI3.

Synaptic dysfunction is implicated in both autism-spectrum disorder and schizophrenia [199, 202]. Consistent with their role in regulating synapse formation and maintenance, mutations in the BAIs have been associated with both of these neurological diseases [52, 147], as well as in other psychiatric disorders such as bipolar disorder and addiction [152, 165].

1.3.2 CELSRs

The CELSRs (or 7TM-cadherins) are best studied for their roles in the planar cell polarity pathway (which regulates organisation of cells in a sheet) but have emerging roles in neuronal migration, axon guidance and dendritic development [255]. Their molecular mechanisms in axon guidance are poorly understood in terms of their

intracellular and extracellular ligands but, along with Frizzled3, CELSR3 controls the development of most of the longitudinal tracts in the central nervous system [36].

CELSR2 uses its large ectodomain to perform a dual role in dendritic maintenance. Deletion of the receptor's extracellular cadherin repeats decreases dendritic complexity, whereas deletion of its EGF-Horm region has the opposite effect [232]. This suggests that the CELSR2 ectodomain may bind multiple ligands at distinct sites, exerting opposing effects on dendritic outgrowth. Further, homophilic interactions between CELSR2 molecules were shown to enhance neurite outgrowth, whereas CELSR3 homophilic interactions have the inverse effect [231]. These results point to a role of CELSRs as context-recognisers in dendritic development. CELSRs have functions early in embryonic development as, in humans, CELSR1 mutations are associated with the neural tube defects craniorachischisis [212] and spina bifida [140].

1.3.3 GPR56

As discussed briefly earlier, mutations in the aGPCR GPR56 are associated with the disease BFPP [201]. BFPP results in a pathologically smooth brain (lissencephaly), characterised by neuronal over-migration and breachment of the pial membrane surrounding the brain [146]. At the pial basement membrane, GPR56 normally encounters its ligand, type-III collagen, inhibiting neuronal migration [160]. This points to a role for GPR56 in regulating cortical lamination and pial membrane integrity.

1.4 Latrophilins

The Latrophilins (Lphn1-3 in most vertebrates) are amongst the best-studied aGPCRs. They have long been linked to the nervous system, as they were first identified as receptors for the neurotoxic protein component of black widow spider venom, α -Latrotoxin [49], which causes massive neurotransmitter release from sensory and

motor neurons. The Lphns represent an evolutionarily ancient class of receptor, playing important functions in both neuronal and non-neuronal cells of vertebrates and invertebrates. The ectodomain of vertebrate Lphn is ~ 100 kDa and comprises an N-terminal Lectin domain (Lec), also termed rhamnose-binding lectin-like domain, an Olfactomedin-like domain (Olf), a glycosylated ~ 100 -residue linker (Stalk-region), and a Horm/GAIN domain. The domain organisation of vertebrate Lphns is shown below (Figure 1.6A). The structure of the murine Lphn1 Lec domain has been solved by NMR, revealing a β -sandwich structure formed by two antiparallel β -sheets decorated by a 10-residue α -helix and two extended loops. One of these loops harbours a low-affinity rhamnose-binding site [260]. Immediately prior to the start of the work described in this thesis, the crystal structure of the Lphn3 N-terminal region (comprising the Lec and Olf domains) was solved by Dr. Elena Seiradake. This revealed that the Olf domain forms a five-bladed β -propellor with two metal ions bound in the central channel of the β -propellor (Figure 1.6B).

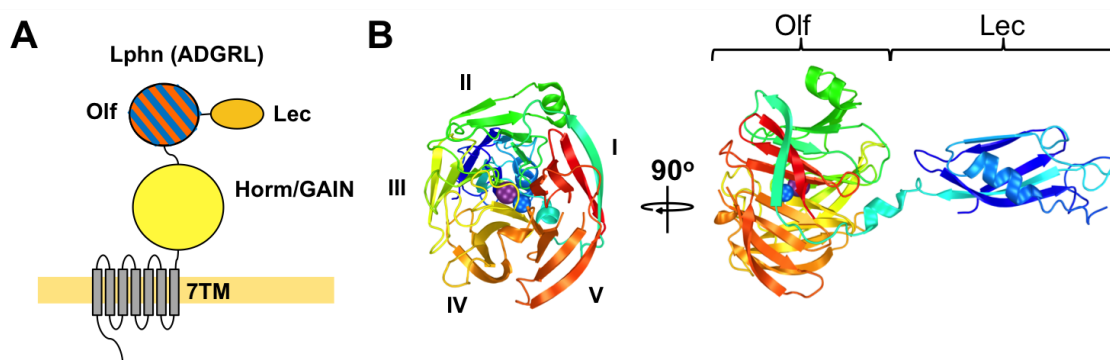


Figure 1.6: **Domain Organisation of Vertebrate Lphns.** The domain structure of vertebrate Lphns is shown as a schematic. The Lec domain is shown as a dark yellow oval, the Olf domain as a blue and orange striped circle and the Horm/GAIN domain as a bright yellow circle. Linkers of unknown fold are shown as black lines (A). The crystal structure of an N-terminal fragment comprising the Lec and Olf domains of the murine Lphn3 is shown in cartoon representation and coloured according to the rainbow (blue = N-terminus, red = C-terminus) (B).

1.4.1 Lphns as Receptors for α -Latrotoxin

Much of the early functional information on Lphns came from studies on the effects of α -Latrotoxin. Lphn1 displays highest affinity binding to α -Latrotoxin, with Lphn2 displaying weaker affinity and Lphn3 displaying no measurable affinity [98].

Stimulation of Lphn1 with α -Latrotoxin resulted in elevation of intracellular cAMP and IP₃ levels, suggesting that α -Latrotoxin stimulates G-protein-dependent signalling from Lphn [142]. However, intracellular signal transduction by Lphn is not required for α -Latrotoxin toxicity [248]. Two mechanisms are currently proposed for α -Latrotoxin toxicity. Firstly, the receptor binds α -Latrotoxin, activating Phospholipase C and Ca²⁺ signalling. In the second mechanism, the signalling activity of Lphn is not required and receptor binding purely serves to recruit α -Latrotoxin to the cell membrane where it can tetramerise and insert into the membrane, forming Ca²⁺ permeable pores. The resulting increase in intracellular Ca²⁺ causes mass exocytosis of neurotransmitter-containing vesicles [159].

1.4.2 Expression of Lphns

Multiple lines of evidence from various vertebrate species suggest that Lphn1 and 3 are primarily expressed in the brain, whereas Lphn2 is ubiquitously expressed [142, 248, 164, 98, 28]. As the mouse brain matures postnatally, the levels of Lphn1 and 3 rise, peaking around postnatal day 15 (P15) [28].

1.4.3 Neuronal Functions of Lphns

Roles in Mechanosensation

The functions of Lphns have been most extensively studied in neurons. In flies, the Lphn homologue, CIRL, is strongly expressed in the neurons of the chordotonal organs – sensory structures that govern the perception of a wide range of mechanical inputs. CIRL mutant neurons displayed decreased action current frequencies in response to vibration stimuli, and the larvae have locomotion defects and reduced sensitivity to gentle touch [224]. Interestingly, it has recently been reported that mechanical stimulation of CIRL reduces cAMP levels in these sensory neurons and that this activity may directly modulate ion flux through TRP channels (well known mechanosensitive ion channels), although the nature of this interaction remains unclear [225].

Extracellular Neuronal Ligands of Lphns

The endogenous roles of vertebrate Lphns are best understood in the context of their interaction partners. Lphns use their N-terminal regions to bind the ectodomains of at least four different ligands: Neurexins, Contactin6, Teneurins and Fibronectin Leucine-Rich Repeat (LRR) Transmembrane proteins (FLRTs) [27, 294, 237, 191], all of which have established roles in axon guidance and synaptogenesis. The interactions with the FLRTs and Teneurins are the most extensively studied so far.

Neurexins are single-pass transmembrane proteins which are best-characterised for mediating heterotypic transsynaptic interactions with Neuroligins [247]. Lphn competes with Neuroligin for binding to Neurexin, although the *in vivo* relevance of this remains unclear [27]. Contactin6 is a newly identified neuronal ligand of Lphn1. This interaction is mediated via the Lphn1 ectodomain *in cis* and is thought to have an anti-apoptotic role [294].

The interaction of Lphn1 with Teneurin2 was first discovered using pull-down experiments from rat brain lysate [237]. This interaction was then later confirmed to occur between Lphn1 and all four Teneurin homologues (Ten1-4) [28]. The Lphn-Teneurin receptor pair is capable of inducing cell aggregation [28] and artificial synapses between hippocampal neurons and transfected HEK293 cells [237], suggesting that this interaction occurs transsynaptically. Teneurins have been shown to bind to the Lec-Olf fragment of Lphns [192], although another report suggests that the Lphn Lec domain is sufficient for Teneurin binding. Interestingly, the affinity of the interaction depends on a developmentally regulated alternative splicing event affecting the length of the linker between the Lphn Lec and Olf domains [28].

All three Lphn homologues also engage in nanomolar affinity interactions with FLRT (FLRT1-3) proteins. Again, this interaction has been shown to occur transsynaptically with Lphn localised to the pre-synapse and FLRT to the post-synapse [191].

Functionally, the interaction regulates the development of glutamatergic synapses [191] and is important in ensuring the correct development of cortical microcircuits in the mouse brain [192]. The Olf domain of Lphns is sufficient for their interaction with FLRTs [192]. Both the Teneurins and the FLRTs will be discussed extensively in later sections.

Intracellular Ligands of Lphns

Lphns have been shown to couple to heterotrimeric G-proteins, including $G\alpha_o$ (of the $G\alpha_i$ -type) [142], $G\alpha_s$ and $G\alpha_q$ [178]. A recent report suggests that G-protein coupling may be specific to the Lphn homologue, cell type or stimulus, as mechanical stimulation of *Drosophila* Lphn resulted in a drop in intracellular cAMP concentration [225].

Like around half of all aGPCRs, Lphns contain a C-terminal PBM. The Lphn PBM has been shown to bind the PDZ domain-containing SH3 and multiple ankyrin repeat domains (SHANK) family proteins [126, 256]. SHANKs are typically considered “master” scaffolding proteins, tethering and organising other scaffolding proteins at synaptic membranes and forming a dense network between clustered cell surface receptors and the actin cytoskeleton [172].

1.4.4 Lphns in Neurological Diseases

Consistent with their extensive role in the brain, mutations in Lphns have been associated with a variety of neurological problems such as mental retardation and psychomotor and language delay [25]. Linkage studies strongly associate mutations in Lphns with the development of Attention Deficit Hyperactivity Disorder (ADHD) [13, 55]. This link has now been more directly validated in *lphn3.1* knockdown studies in zebrafish, showing aberrant dopaminergic neuron development alongside a hyperactive motor phenotype [132].

1.4.5 Non-neuronal Functions of Lphns

Besides their neuronal functions, Lphns have emerging roles in regulating diverse biological processes. The *C. elegans* homologue of vertebrate Lphns, Lat-1, plays an essential role in establishing anterior-posterior polarity in early embryonic development by helping to specify division plane alignment in epidermal seam cells [135]. Lat-1 is also important in the production of functional sperm [205]. Establishment of anterior-posterior polarity requires *Stachel*-mediated cAMP elevation via $G\alpha_s$, although the extracellular cue for this is still unknown [178].

Studies in vertebrates have also indicated non-neuronal functions for the Lphns. In chicken embryos, Lphn2 has been implicated in the epithelial-mesenchymal transition and in early heart development [57]. Further, a recent study implicates Lphns in the development of asthma in humans, as expression of Lphn1 and 3 was enhanced in the airway smooth muscle cells (ASMCs) of asthmatic patients. Interestingly, application of FLRT3 induced both ASMC proliferation and contraction, suggesting that Lphn3 may partly be responsible for increased airway contractility in asthmatics [67].

Many of the studies discussed above implicate the interactions of Lphns with their extracellular ligands as crucial in driving a diverse range of biological processes, particularly in the brain. In the following sections, two of the best-characterised Lphn ligands, the FLRTs and the Teneurins are discussed in detail.

1.5 FLRTs

FLRTs (FLRT1-3) are broadly expressed [131], vertebrate-specific proteins which are emerging as potent regulators of early embryonic, neural and vascular development. FLRTs are single-pass transmembrane proteins with ~ 500 residue extracellular domains, and smaller intracellular domains of ~ 100 amino acids. The extracellular domain comprises an N-terminal leucine-rich repeat (LRR) domain followed by a

~60 residue glycosylated linker and a fibronectin (FN)-type III domain. The crystal structure of the FLRT2/3 LRR domain is published [226]. All FLRTs contain a metalloprotease cleavage site in their extracellular juxtamembrane region [285] (Figure 1.7). As well as functioning as transsynaptic ligand for Lphns, FLRTs have a variety of *in vivo* functions which are discussed below.

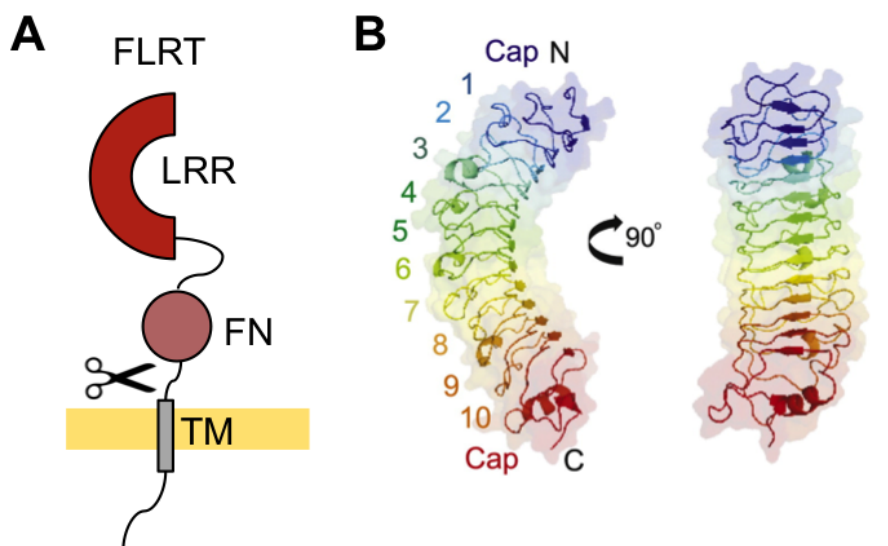


Figure 1.7: **Domain Organisation of FLRTs.** The domain structure of FLRTs is shown as a schematic. The position of the metalloprotease cleavage site is shown by a scissor symbol (A). The crystal structure of the FLRT3 LRR domain is shown in cartoon and transparent surface representation and coloured according to the rainbow. The 10 LRRs are numbered (B). Panel B was taken from [226].

1.5.1 The Diverse Functions of FLRTs

FLRT proteins have been implicated in the development of a wide variety of tissues and organs, and play powerful roles in early embryonic development. In mouse embryos, FLRT3 is required to maintain tissue integrity and organise the underlying basement membrane in the anterior visceral endoderm signalling centre (governing establishment of the anterior-posterior body axis) [61, 161]. FLRT2 is thought to play a similar role during heart development, suggesting that the FLRTs play conserved roles in maintaining cell-cell contacts in diverse body systems [179]. FLRT3 has also been shown to promote neurite outgrowth from both cerebellar and sensory spinal neurons (dorsal root ganglia) [257, 213]. Recent work has proposed that the adhesive properties of FLRTs regulate neuronal migration and guidance in the

developing mouse cortex [285, 226] and that expression levels of FLRT1/3 may act as a key determinant of cortical folding [50].

Like many CAMs, FLRTs have functions other than their role in promoting cell adhesion. For example, FLRT3 has also been shown to use its intracellular domain to recruit the small GTPase, Rnd1, promoting deadhesion by Cadherin endocytosis [188]. A similar function has been proposed for FLRT2 during chondrogenesis [283]. Recently it has been proposed that FLRTs have the unique ability to act as so-called Repulsive Cell Adhesion Molecules (ReCAMs), combining finely tuneable cell adhesion with repulsive functions by the recruitment of a heterophilic receptor [226].

1.5.2 Homophilic FLRT Interactions

One of the first proposed functions of FLRT proteins was in homotypic cell adhesion, mediated by the concave surface of the FLRT LRR domain [108, 226]. This interaction is low affinity (one report suggests a K_D of $\sim 40 \mu\text{M}$ for FLRT3) [208], but is presumably enhanced by avidity effects since FLRT-transfected cells cluster in a cell aggregation assay [108, 226]. Many of the cell adhesive properties of FLRTs have been attributed to their ability to form these homophilic interactions. For example, FLRT homophilic interactions are proposed to be important in mediating the tangential spread of pyramidal neurons during the development of the cerebral cortex [226].

1.5.3 Heterophilic FLRT Interactions with Unc5s

How FLRT functions from a molecular perspective is best understood in terms of its interactions with the Uncoordinated5 family of receptors (Unc5A-D in mammals). Unc5s are single-pass transmembrane receptors which have a largely conserved architecture across species. The ectodomain consists of two N-terminal immunoglobulin-like (Ig1 and Ig2) domains and two thrombospondin-like (Tsp) domains. The single-pass transmembrane helix leads into a large intracellular region including ZU5 (as

initially found in ZO-1 and Unc5) and UPA (Unc5, PIDD and ankyrin) domains, and a Death Domain (DD) (Figure 1.8A). Crystal structures of the two Ig and Tsp1 domains of Unc5A (Figure 1.8B) [226] and most of the intracellular region of Unc5B [270] are available.

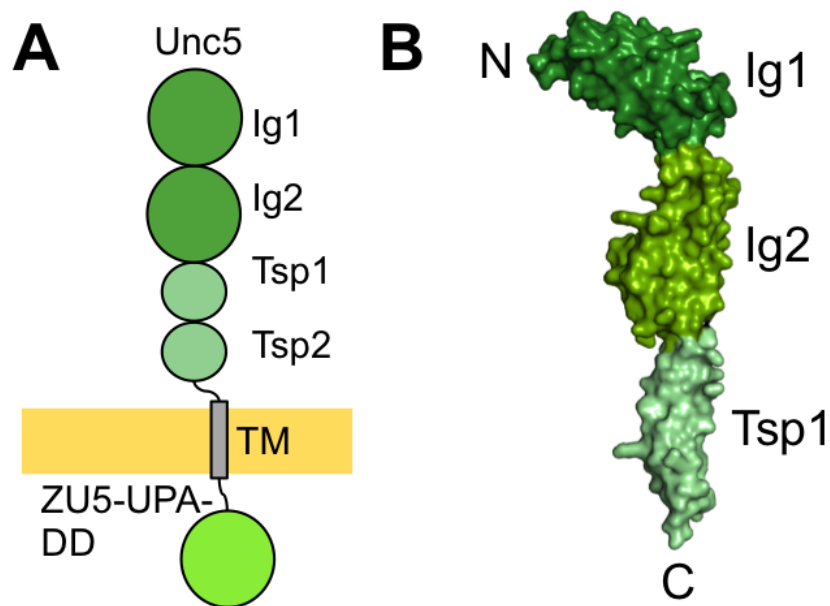


Figure 1.8: **Domain Organisation of Unc5s.** The domain structure of Unc5 receptors is shown as a schematic (A). The crystal structure of a fragment of the human Unc5A ectodomain (PDB accession: 4V2A) is shown in surface representation [226].

Unc5s as Netrin Receptors

Unc5s have mainly been studied for their roles as repulsive receptors for secreted Netrins in a variety of biological scenarios, including neural guidance and angiogenesis. Netrins have two receptors, Deleted in Colorectal Cancer (DCC) and Unc5 [112, 143] and the cellular effects of Netrin largely depend on which receptor it binds to. In all species, the attractive effects of Netrins are mediated by the DCC receptor family, whereas the repulsive functions require Unc5 family receptors [122]. The long-held view is that Netrin1, secreted by the ventral midline floor plate of the embryonic spinal cord, diffuses to bind DCC receptors expressed on the surface of commissural neurons, promoting their outgrowth [229, 113]. However, when Unc5

is present *in cis*, it forms a ternary complex with Netrin-DCC, converting attraction into repulsion [96]. This has recently been contested by reports suggesting that Netrin1 may function in a short-range adhesive manner in this scenario [56]. Netrin1 has been shown to bind the Ig1 and Ig2 domains of Unc5 family members [83].

As well as regulating neural guidance in response to Netrins, Unc5 (and DCC) receptors also have important roles in regulating cell survival by acting as so-called “dependence receptors” for secreted Netrins [154]. When bound to Netrin1, these receptors transduce pro-survival signals such as mitogen-activated protein kinase activation [73]. However, in the absence of Netrin1, Unc5 receptors actively trigger apoptosis [153]. Crystallographic evidence reveals that the intracellular domain of Unc5B adopts an auto-inhibited closed conformation in which the ZU5 and Death Domains bind one another, preventing apoptosis. The apoptotic activity of mutant Unc5B with a constitutively open conformation of the intracellular domain is not inhibited by Netrin1 binding, suggesting that Netrin binding normally stabilises the closed conformation of the receptor [270]. Consistent with a role in regulating cell survival, Unc5 receptors are implicated in various diseases. Unc5s act as potent tumour suppressor genes and are downregulated in many cancers [253]. Mutations in Unc5 family members are most strongly implicated in colorectal cancers [233, 21, 42, 130]. Further, a mutation in the Unc5C intracellular domain has been linked to late-onset Alzheimer’s disease and increased neuronal cell death [274].

The FLRT-Unc5 Interaction

Unc5B and Unc5D were initially identified as high-affinity FLRT3 interactors in a cell-binding expression screen [107, 242]. Since then, all four mammalian Unc5 homologues have been shown to interact with FLRT proteins [191]. FLRT3 and Unc5B act synergistically in *Xenopus* embryos to induce deadhesion, and both proteins can bind the small GTPase Rnd1 via their intracellular domains, which is thought to be important in transmitting the downstream signal [107].

The FLRT-Unc5 interaction is best understood in the context of cortical development, where it acts in a repulsive manner. In the mouse cortex, FLRT2 expression is restricted to the cortical plate, but metalloprotease cleavage in the juxtamembrane region of FLRT2 releases the ectodomain, allowing it to act as a long-range diffusible guidance cue. The FLRT2 ectodomain can then diffuse to the subventricular zone where it acts as a “molecular brake” for young Unc5D+ neurons, delaying their migration until the correct developmental time point [285]. The shed ectodomain of FLRT2 has been shown to interact with Fibronectin, a component of the extracellular matrix, which could help establish a gradient of these soluble molecules [70].

The crystal structure of the minimal FLRT2-Unc5D complex (FLRT2^{LRR} bound to Unc5D^{Ig1}) reveals that the Unc5D^{Ig1} domain binds to the lateral side of the FLRT2^{LRR}, a site distinct from the FLRT-FLRT homophilic interaction site (Figure 1.9). Elegant work using mutations based on this complex crystal structure revealed that the repulsive activity mediated by the FLRT2-Unc5D interaction controls the radial migration of cortical pyramidal neurons in the developing mouse cortex, whereas the FLRT concave surface (which mediates FLRT homophilic interactions) regulates pyramidal neuron migration in the tangential axis [226].

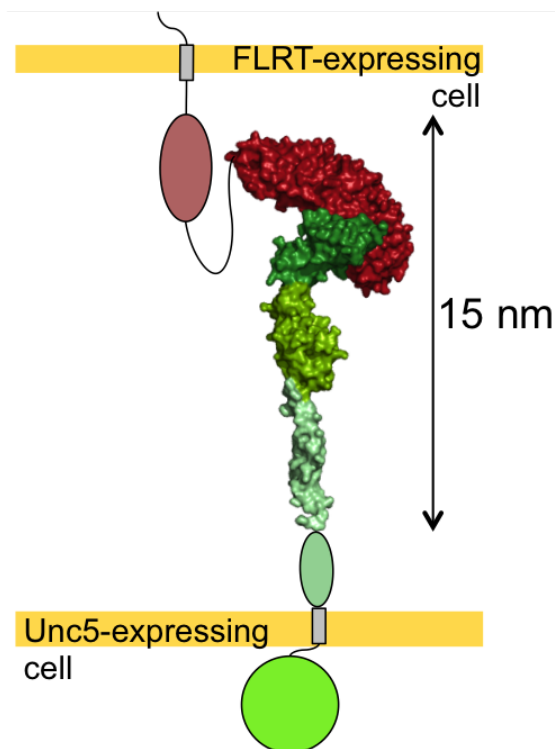


Figure 1.9: **The Crystal Structure of the Minimal FLRT-Unc5 Complex.** The crystal structure of the minimal FLRT2^{LRR}-Unc5D^{Ig1} complex (PDB accession: 4V2C) was used to superpose the crystal structure of Unc5A Ig1-Ig2-Tsp (PDB accession: 4V2A). This provides a model FLRT-Unc5 complex formed between two opposing cells [226]. All domains are coloured as in previous figures.

The FLRT-Unc5 repulsive interaction has also been shown to have important biological functions outside of the brain. For example, FLRT2-Unc5C-mediated repulsion is important in controlling neuronal laminar targeting in the inner plexiform layer of the mouse retina [265]. FLRT-Unc5 interactions are also important in vascular development. Fundamentally the mechanisms of FLRT-mediated adhesion (via homophilic interactions) and repulsion (via Unc5) are conserved between both neural and vascular development [226], placing FLRTs amongst the ever-growing number of neural guidance molecules involved in regulating the development of these distinct systems [5].

1.5.4 Other Heterophilic FLRT Interactions

Role in Fibroblast Growth Factor (FGF) Signalling

FLRTs have been shown to be necessary for FGF signalling and form *in cis* interactions with FGF receptors (FGFRs), via the FLRT FN-III domain [26]. The intracellular domain of FLRT1 is phosphorylated in an FGFR1-dependent fashion, a modification that is thought to suppress FLRT-mediated potentiation of FGF signalling and regulate neurite outgrowth [275]. Interestingly, the various FLRT homologues display differential affinities for the different FGFRs, suggesting that different FLRTs could influence different FGF signalling pathways [85].

Interactions with the Canonical Neural Guidance Cues

FLRTs also regulate the activity of other receptors *in cis*. FLRT3 can bind *in cis* to the guidance receptor Robo1. Upon binding Slit1, the FLRT3-Robo1 complex activates protein kinase A and upregulates DCC cell surface expression. This allows rostral thalamocortical axons to be attracted to sources of Netrin1. In contrast, intermediate thalamocortical axons, which do not express FLRT3, display no response when Netrin1 and Slit1 are present in combination [145].

1.6 Teneurins

The other Lphn ligands that will be discussed extensively in this thesis are the Teneurins. Teneurins are evolutionarily ancient cell surface receptors present in both vertebrates and invertebrates and are implicated in axon guidance, neuronal migration and development. They are believed to have originated in choanoflagellates, unicellular eukaryotes thought to be the closest living relatives of the animals [259]. The Teneurin gene family was first described in *Drosophila*, with two homologues having been identified, *ten-a* [16] and *ten-m* [17, 144]. Later, a murine orthologue of *ten-m*, DOC4 (later Ten4), was discovered in a screen for proteins induced during the endoplasmic reticulum unfolded protein response [272]. Ten4

was later revealed to be one of four murine Teneurin homologues (Ten1-4) [190].

1.6.1 Domain Architecture of Teneurins

The Teneurins are type-II single-pass transmembrane proteins, meaning they have an extracellular C-terminus [190] (Figure 1.10). The vertebrate Teneurin intracellular domain contains two EF-hand like sequences, two polyproline sequences and two tyrosine residues which are predicted to be phosphorylated. Although the intracellular domains of the vertebrate Teneurins align well, none of the intracellular domains of the invertebrate Teneurins align with those of vertebrates, or with each other [258].

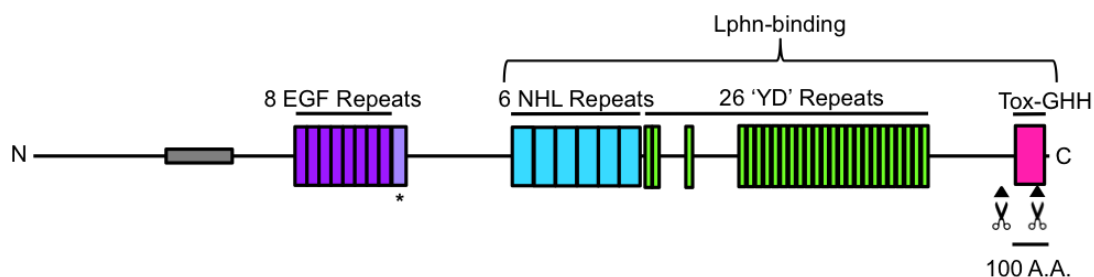


Figure 1.10: **Literature-based Domain Annotation of Teneurins.** A schematic of the Teneurin ectodomain annotated according to domains identified in the literature. Note that much of the ectodomain does not contain any predicted domains. The two putative C-terminal protease cleavage sites are denoted by scissor symbols. The cysteine-rich motif is marked with an asterisk. The portion of the ectodomain that binds to Lphn is shown [237].

The large extracellular domains of all Teneurins (>2000 amino acids in both vertebrates and invertebrates) show a much higher degree of conservation than the intracellular domains. Approximately 200 amino acids separate the transmembrane helix and eight phylogenetically conserved Epidermal Growth Factor-like (EGF) repeats, which bear sequence similarity to those of the extracellular matrix glycoprotein, Tenascin [16, 17, 144]. One EGF repeat normally contains six cysteine residues, forming a stable structure with three disulphide bonds. Uniquely, one cysteine residue in both EGF repeats 2 and 5 is replaced by a bulky hydrophobic residue in all vertebrate Teneurins. This allows Teneurins to form disulphide-bonded homo- and heterodimers [190, 69]. Early electron microscopy studies revealed the overall

architecture of these dimers, showing pairs of roughly spherical domains, connected by thin, elongated rods, which were presumed to be the EGF repeats (Figure 1.11) [190].

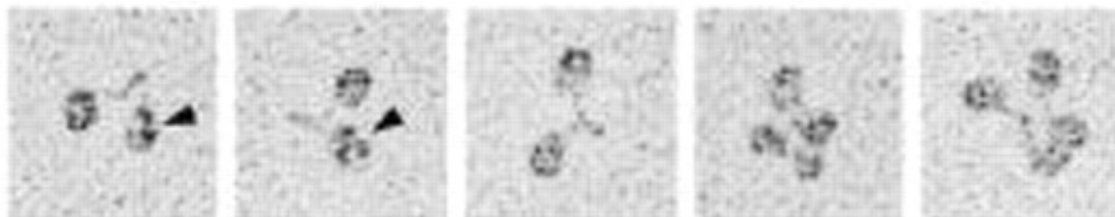


Figure 1.11: **Electron Micrographs of MmTen1 Ectodomains.** Electron micrographs after glycerol spraying/rotary shadowing of recombinant MmTen1 extracellular domains. The MmTen1 ectodomain exhibited a variety of conformations, including forming dimers-of-dimers on the electron microscopy grid (last two panels). Figure adapted from [190].

Immediately following the EGF repeats is a Cys-rich motif containing six cysteine residues. This represents one of the most highly conserved regions of the protein, and aligns well with predicted metal-binding motifs of unrelated proteins [259]. The next ~ 290 amino acids of vertebrate Teneurins are completely unannotated in the literature. This unannotated region is followed by an approximately 400-amino acid NCL-1, HT2A and Lin-41 (NHL) domain, which is present in a variety of growth regulator proteins and has been shown to form a β -propellor structure in the *Drosophila* Brain Tumour transcriptional regulator protein [60]. The Lphn-binding region of Teneurin has been mapped to a region spanning from the beginning of the NHL domain to the protein's C-terminus (Figure 1.10) [237].

Downstream of the NHL domain, Teneurins harbour 26 tyrosine-aspartate (YD) or rearrangement hotspot (Rhs) repeats. This region of the protein has been shown to bind heparin [170]. Each YD-repeat motif (consensus sequence: GXXTXYXY-DXXGRLXXXTXXX) corresponds to a single strand-turn-strand motif, many repeats of which form large shell-like structures, which are plugged at their C-terminal end by a so-called Rhs-associated core motif (also conserved in Teneurins, Figure 1.12) [34]. The final ~ 300 amino acids of the Teneurin ectodomain are poorly annotated, but the final ~ 75 amino acids share sequence homology ($\sim 40\%$) with an

inactive form of a Tox-GHH domain, a type of bacterial nuclease [290]. Teneurins are also thought to be extensively proteolytically processed in the region C-terminal to their YD-repeats (Figure 1.10). Almost all vertebrate Teneurins have a consensus Furin cleavage sequence (RXRR) ~100 amino acids from their C-terminus – proteolytic processing at this point would release the intact Tox-GHH domain [290]. Additionally, a second cleavage site has been predicted ~50 residues from the C-terminus, which would liberate a short peptide with similarity to bioactive peptides. This peptide has been termed the Teneurin C-terminal-associated peptide (TCAP), and is discussed in detail in Section 1.6.2 [157].

Interestingly, the arrangement of N-terminal YD-repeats, followed by a C-terminal toxin-like domain is reminiscent of a class of bacterial polymorphic toxins called the Rhs toxins. These toxins typically exhibit an N-terminal core region containing Rhs/YD-repeats, but have highly variable C-terminal “tip” regions, encoding their cytotoxic parts [94, 101]. These C-terminal tip regions harbour a variety of different toxic activities and can target both prokaryotic [124] and eukaryotic [129] cells. Of these Rhs-containing toxins, the ABC toxins are the best structurally characterised. The structures of these toxins have been used to make predictions about the Teneurin C-terminal region and are briefly introduced below.

Bacterial ABC Toxins as Example Rhs Toxins

The ABC toxins comprise three protein components, A, B and C, which must assemble for full toxicity. The first structure of an Rhs/YD-repeat containing protein was of the BC complex (YenB-YenC2) from an ABC toxin from *Yersinia entomophaga* [34]. This was followed shortly by an almost identical structure of a B-C fusion protein from an ABC toxin of *Photobacterium luminescens* (TcdB2-TccC3) [167] (Figure 1.12). The C protein is similar to the bacterial “polymorphic toxins”, having a conserved N-terminal region harbouring YD-repeats and a highly variable C-terminal region which harbours the main cytotoxic activity of the protein [290]. When assem-

bled with the B protein (which also contains YD-repeats), the C protein undergoes an autoproteolytic cleavage event between its conserved N-terminus and variable C-terminus yielding two fragments, the C^{NTR} and C^{CTR} . The crystal structures of B- C^{NTR} complexes of the two ABC toxins (PDB accessions 4IGL and 4O9X respectively) reveal that the B and C proteins assemble to form a large hollow barrel-like structure, with the C-terminus of C^{NTR} enclosed within the barrel (Figure 1.12A). The barrel-like structure is composed of 90 β -strands, which are contributed by both the B and the C proteins. The shell is closed at its B-end by a β -propellor domain and at the C-end by a short β -sheet which spirals inwards to form a plug (this domain was termed the Rhs-associated core domain and is highly conserved in other bacterial YD toxins and in Teneurins) (Figure 1.12A)

Data from Small-Angle X-ray Scattering experiments and Cryo Electron Microscopy indicate that the cleaved cytotoxic C^{CTR} of both YenC2 and TccC3 resides inside the barrel structure formed by the YD repeats of the B protein and C^{NTR} [34, 167]. Cryo-Electron Microscopy data indicate that the β -propellor of the B protein docks onto the pentameric pore formed by the A protein, opening the barrel and forming a continuous channel, allowing the cytotoxic domain to be translocated into the receiving cell [167].

All of these findings point to a model whereby the cytotoxic domains of these proteins are encapsulated within a shell formed by N-terminal YD-repeats, meaning that the toxin-producing cell is never directly exposed to the cytotoxic domain [34, 167]. Even during toxin delivery, the cytotoxic domain is translocated directly from the inside of the barrel through the A-protein pore and into the target cell. This has been suggested to be a self-protection mechanism for the producing cell. Given that the C-terminal domain of Teneurin harbours an (albeit inactive) Tox-GHH domain, it has been suggested that the C-terminal regions of Teneurins would be enclosed within similar barrels [34].

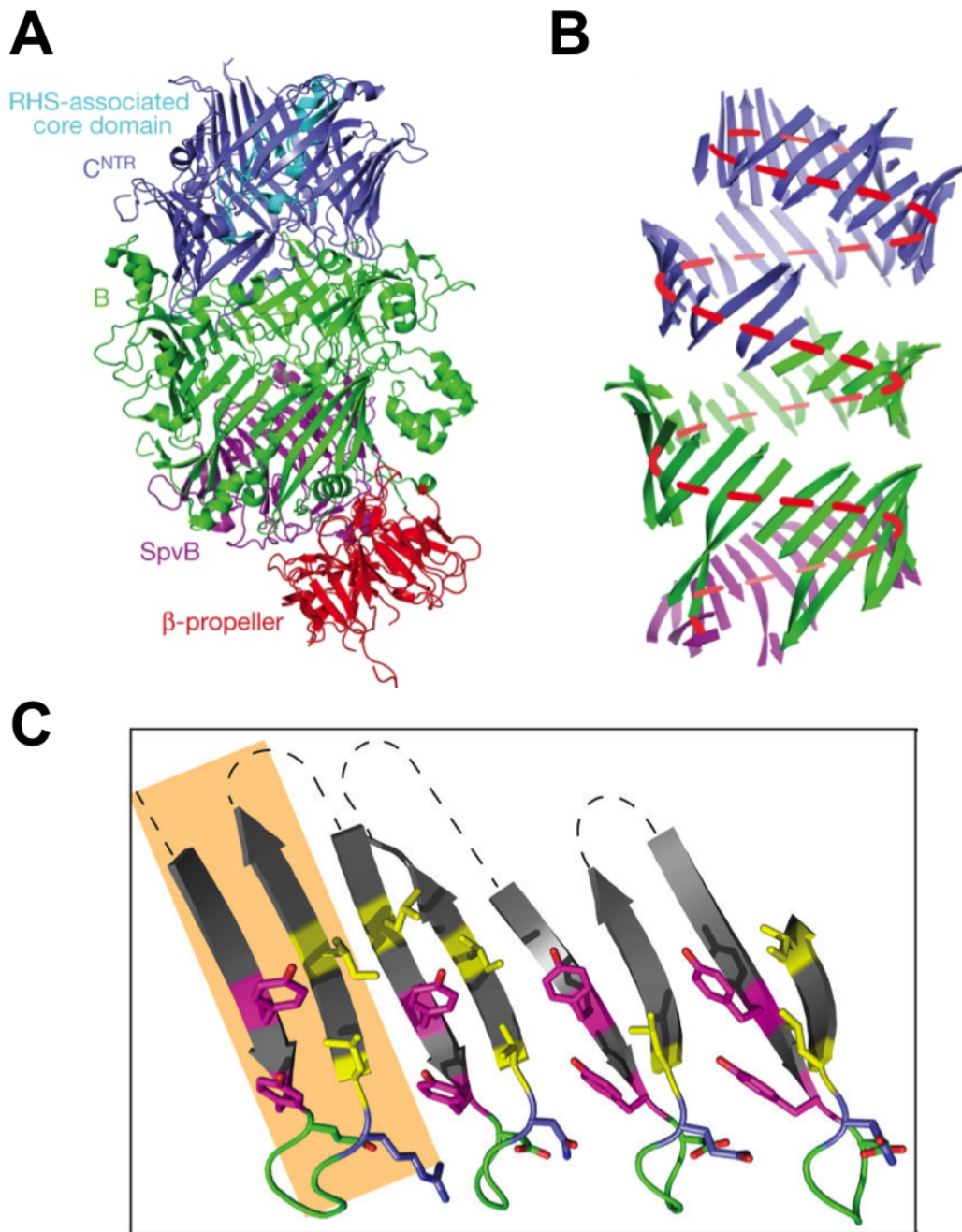


Figure 1.12: **The Crystal Structure of YenB-YenC2.** Side view of the crystal structure of the B- C^{NTR} complex of YenB2-YenC2 from *Y.entomophaga*. The β -propeller domain of B is coloured red, the rest of B is coloured in purple and green. C^{NTR} is coloured in blue and the Rhs-associated core domain is coloured in cyan (A). A continuous β -sheet wraps around the central cavity to form the barrel (B). A section of the YD-repeats from C^{NTR} are shown in cartoon representation (viewed from inside the barrel). The conserved residues are coloured: tyrosine, magenta; other hydrophobic residues, yellow; hairpin turn made by the DXXGR motif, green and blue. One YD-repeat sequence motif is highlighted in orange and forms a single strand-turn-strand motif (C). Figure adapted from [34].

1.6.2 Proteolytic Processing of Teneurins

Interestingly, Teneurins are thought to be post-translationally proteolytically cleaved in both a homologue-dependent and homologue-independent manner. This is discussed below.

Proteolytic Processing of the Intracellular Domain

Early work on Ten2 revealed that the intracellular domain is proteolytically cleaved, allowing it to translocate to the nucleus and act as a transcriptional repressor for certain genes [15]. Recently, additional mechanistic insight was gained when it was discovered that the Ten1 intracellular domain binds a transcriptional repressor, upregulating transcription of genes under the control of Microphthalmia-associated Transcription Factor, MITF [223]. The intracellular domains of many, but not all, Teneurin proteins contain nuclear localisation signals, suggesting that this could be a conserved mechanism of Teneurin signalling [114, 259]. The extracellular cue that regulates this cleavage event is currently poorly understood, but homophilic or heterophilic interactions of the Teneurin extracellular domain is one possible mechanism [15].

Proteolytic Processing Near the Transmembrane Domain

Many of the vertebrate Ten2/3 proteins, but not Ten1/4, contain predicted Furin cleavage sites between their single-span transmembrane helices and their EGF domains [259]. Recently it has been shown that the ectodomain of Ten2 constitutively undergoes Furin cleavage at this site, however the cleaved ectodomain remains associated with the cell surface by association with its transmembrane helix. A second cleavage event, which occurs within the transmembrane helix of Ten2, releases the Ten2 ectodomain into the medium [267]. In neurons, this second cleavage event is highly inefficient (only ~8% of Ten2 dimers are released from the cell surface) [267]. The *in vivo* role of the soluble Ten2 ectodomain is currently unknown.

TCAPs

The TCAP sequence was first discovered using a low-stringency screen for homologues of the stress-response peptide hormone, Corticotrophin-Releasing Factor (CRF) (Figure 1.13). These peptides have many of the features typical of bioactive peptides, including a pro-hormone cleavage motif in their N-terminal region and an amidation motif at their C-terminus [206, 157]. It is not entirely clear how TCAPs may be generated *in vivo*. The pro-hormone cleavage motif lies in the middle of the inactive Tox-GHH domain, and cleavage at this site would result in the liberation of the ~50 most C-terminal amino acids of the receptor [206]. Alternatively, it has been shown that a TCAP-1-containing transcript can be transcribed independently of the rest of the receptor, and it is thought that the protein product of this transcript is further processed to yield a 10-15 kDa fragment [37].

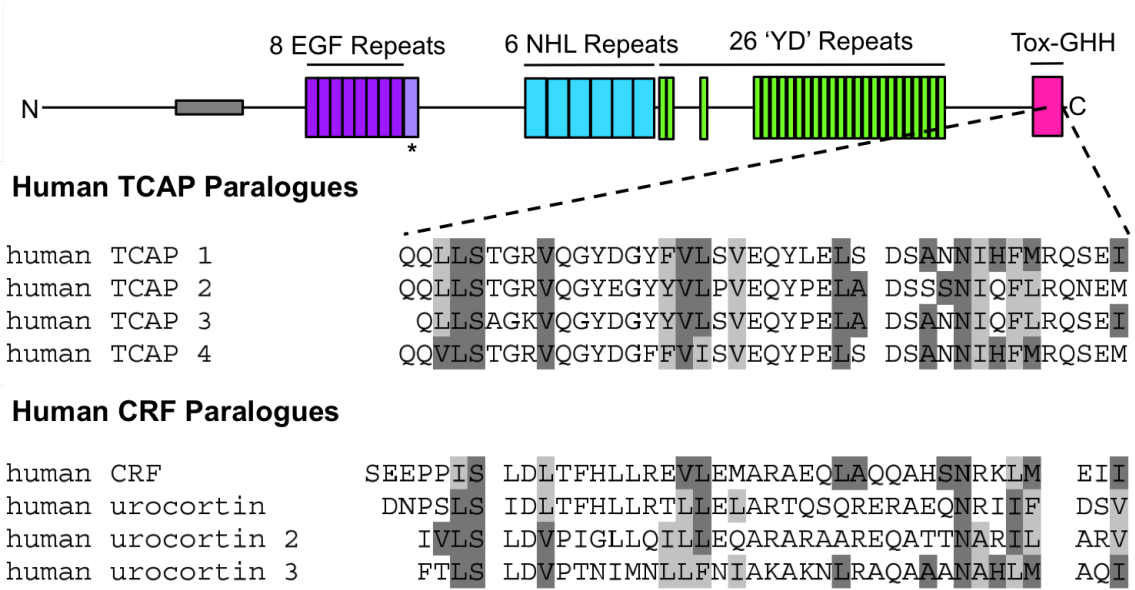


Figure 1.13: **Sequence Similarity between TCAP Sequences and CRFs.** Alignment of the TCAPs with members of the CRF family. Note that the TCAP amidation motif is not shown in this alignment. Dark grey boxes indicate regions of amino acid identity. Light grey regions indicate amino acids replaced with one with similar physicochemical properties. The location of the TCAP peptide in the full-length Teneurin receptor is indicated. Alignment taken from [157].

Since the discovery of TCAPs, synthetic peptides based on the sequences of various TCAPs have been shown to have diverse biological activities. Application of synthetic TCAPs to immortalised mouse neurons results in intracellular cAMP el-

evation and increased cell proliferation [206, 269, 8]. Synthetic TCAPs can also regulate neuronal ultrastructure by increasing neurite outgrowth [7] and by controlling dendritic morphology [251]. They also have behavioural effects, reducing anxiety behaviours in rats [252, 251]. The identity of the TCAP receptor is still a matter of debate, although a C-terminal (TCAP-containing) fragment of Ten2 has been shown to stimulate Ca^{2+} signalling in Lphn1-transfected cell lines and hippocampal neurons expressing Lphn1 [237].

1.6.3 Teneurin Expression Patterns

Expression studies in many species have revealed most prominent expression of Teneurins in the nervous system, with some family members also expressed in non-neuronal tissues, in particular at sites of pattern formation [19]. In chicken embryos, Ten1 and Ten2 are expressed in non-overlapping populations of neurons in interconnected regions of the brain, particularly in regions associated with the visual and olfactory systems. This suggested a role for Teneurins in regulating neuronal connectivity [114]. In the murine brain, expression of Teneurins is highly regulated during embryonic development. Ten3 and Ten4 are expressed early in development (first expressed in the neural plate at E7.5), whereas Ten2 and Ten1 expression begins later (at E10.5 and E15.5 respectively). Analysis of Teneurin expression in the adult mouse brain revealed that all four Teneurin homologues are expressed in the cortex, thalamus, hippocampus and cerebellum, but that different homologues are expressed in distinct populations of neurons in a partially overlapping manner [292].

1.6.4 Teneurin Homophilic Interactions

The vast majority of Teneurin functions have been attributed to homophilic interactions *in trans*. Here the use of the term “homophilic” will refer to interactions between the same Teneurin (e.g. Ten2-Ten2 interactions). Many *in vivo* studies in which Teneurins are over-expressed on the surface of cells reveal that they promote cell aggregation [97, 215]. This effect is observed for both vertebrate and inverte-

brate Teneurins, and is dependent on the region of the ectodomain C-terminal of the EGF repeats [215]. However, in another study, cell aggregation experiments using agitated cultures failed to show any aggregation of Teneurin-transfected cells, but could detect binding of soluble Ten2 ectodomain to the surface of cells transfected with full-length Ten2 [28]. These discrepancies remain to be resolved, but are consistent with the idea that Teneurins can indeed form homophilic complexes, but that the formation of these complexes *in trans* may not be sufficient to mediate intercellular adhesion under all conditions.

Atomic Force Microscopy has shown that homophilic Teneurin interactions provide substantially stronger adhesion than heterophilic interactions between different Teneurin homologues. This specificity is provided by the protein's NHL domain [18]. Interestingly, these adhesive forces strengthen over time. Truncation of the intracellular domain or drug-mediated disruption of the cytoskeleton perturbs this gradual strengthening [18]. Functionally, establishment and strengthening of Teneurin homophilic interactions halts neurite outgrowth, presumably leading to synapse formation [18, 139].

The observation that cytoskeletal disruption perturbs the strengthening of Teneurin homophilic interactions is consistent with their emerging role as cytoskeletal organisers. The intracellular domain of Ten1 interacts with the SH3 domain-containing cytoskeletal adaptor protein, CAP/Ponsin [186]. Indeed several cytoskeletal regulators such as Focal Adhesion Kinase, Cdc42 and Rac1 have been shown to be downstream of, and activated by Ten4, promoting neurite outgrowth [249, 250]. The interactions of Teneurins with the cytoskeleton are proposed to be important in their role as synaptic organisers, which is discussed below [176, 175].

1.6.5 Teneurins in Axon Guidance

One of the earliest proposed functions for the Teneurins is in axon guidance. Early experiments showed that both Ten1 and Ten2 can promote neurite outgrowth *in vivo* [170, 216]. This role in axon guidance is thought to be evolutionarily conserved as RNAi knockdown of the *C. elegans* Teneurin, *ten-1*, results in disturbed axonal pathfinding for some neurons [58].

The roles of Teneurins in neural guidance have been most extensively investigated in developing visual and olfactory systems. In particular, Teneurins are thought to be important in the formation of functional binocular visual circuits [139]. Ten3 is expressed in topographically relevant gradients in the developing visual system of the mouse [138], wallaby [35] and zebrafish [10], suggesting that Ten3 may play a role as a chemoaffinity molecule in specifying axonal targeting in the vertebrate visual pathway. Indeed, Ten3 knockout mice display defects in the ipsilateral wiring of their visual circuitry, whereas contralateral wiring is unaffected [138]. Similarly, overexpression of Ten3 in the wallaby retina causes misrouting of ipsilateral projections, but not contralateral projections [35]. Interestingly, studies using Ten2 knockout mice reveal that Ten2 is also critical for the development of the ipsilateral retinal pathway, but in a manner distinct from Ten3 [287]. These phenotypes can be partly attributed to the loss of Teneurin homophilic interactions, although other axon guidance molecules may also be involved [139]. Together these data suggest complementary roles for different Teneurin homologues in the formation of binocular circuits.

1.6.6 Teneurins at the Synapse

Two pioneering studies in *Drosophila* have revealed roles for Teneurins in synaptic partner matching and synaptic organisation at the neuromuscular junction and at synapses in the olfactory system [97, 176]. These studies have led to an overall model whereby Teneurins function homophilically during partner matching, whilst synap-

tic organisation occurs heterophilically [175]. During partner matching Teneurins are highly expressed at select pairs of pre- and post-synaptic partners which express high levels of the same Teneurin [176, 97]. In contrast, a lower, basal expression of Teneurins persists at all neuromuscular junction and olfactory synapses, where (presumably weaker) heterophilic interactions between presynaptic Ten-a and post-synaptic Ten-m play a role in synaptic organisation [176]. This role in synaptic organisation relies on the ability of Teneurins to interact with and order the cytoskeleton, indeed post-synaptic Ten-m can interact with α -spectrin [176]. Consistent with a role in synaptic organisation, the Teneurins were shown to regulate correct synaptic structure and density in olfactory receptor neurons [174]. Together these findings point to a conserved function of Teneurins in synapse regulation in both the peripheral and central nervous systems.

In vertebrates, the roles of Teneurins at the synapse have been most extensively studied in terms of their interactions with the Lphns (as discussed above). It is likely that Teneurins also play important roles in wiring the olfactory system in vertebrates as a missense mutation in Ten1, P1610L, is associated with congenital general anosmia (the inability to smell), and mice carrying this mutation display an olfactory deficit [9].

1.7 Project Aims

Several aGPCRs are emerging as crucial regulators in neural guidance, although mechanistic and molecular understanding of their interactions is lacking. It has been hypothesised that aGPCRs could play a role as context recognisers, using their large and modular extracellular domains to bind multiple ligands at distinct sites, although direct molecular evidence for this is sparse [133]. Given the relatively low numbers of receptors implicated in axon guidance, how growing axons sense and respond to combinations of signals is thought to be key to diversifying cellular responses. Given the abilities of many aGPCRs to bind multiple and diverse ligands, and to elicit downstream signals, this class of molecules provides a relatively unexplored molecular basis by which such signal integration could occur.

The aim of my D.Phil. project was to provide an integrated molecular understanding of how the ectodomain of the aGPCR Lphn interacts with neural guidance receptors, namely FLRT, Unc5 and Teneurin. This work would provide some of the first mechanistic indications of how aGPCR ectodomains physically integrate multiple distinct ligand-binding events. To achieve this my objectives were four-fold:

1. Use structure-based mutations to map the FLRT-Lphn binding surfaces and provide functional insight into this interaction using cell guidance assays.
2. Biophysically and functionally assess how Unc5 interacts with the FLRT-Lphn complex and solve the crystal structure of a ternary complex comprising the ectodomains of all three proteins.
3. Biophysically characterise the interaction of Teneurin with Lphn and assess how Teneurin binding affects the interaction of Lphn with FLRT and Unc5.
4. Solve a crystal structure of the Lphn-binding region of a Teneurin protein to reveal its overall fold and identify functionally relevant surfaces that may be important in complex formation with Lphn and its other ligands.

Chapter 2

Materials and Methods

2.1 Molecular Biology and Cloning

2.1.1 Genes and Constructs

Details of the cDNAs for Lphn, FLRT, Unc5 and Ten homologues used in this work are detailed in Table 2.1. The gene encoding *MmTen1* was synthesised as a codon-optimised gene by GeneArt™. All other cDNAs were obtained from the mammalian gene collection or were gifts from collaborators. Many of the constructs used throughout this thesis existed prior to the start of this work and are summarised in Appendix A.

Protein	Species	Accession	Protein	Species	Accession
Lphn1	<i>Mus musculus</i>	XP_006531185.1	Ten1	<i>Homo sapiens</i>	NP_001156750.1
Lphn2	<i>Mus musculus</i>	Q8JZZ7.3	Ten1	<i>Mus musculus</i>	NP_035985.2
Lphn3	<i>Mus musculus</i>	XP_011247768.1	Ten1	<i>Gallus gallus</i>	XP_015133677.1
FLRT1	<i>Mus musculus</i>	NP_958813.1	Ten2	<i>Homo sapiens</i>	NP_001073897.2
FLRT2	<i>Mus musculus</i>	NP_958926.1	Ten2	<i>Gallus gallus</i>	XP_015149056.1
FLRT3	<i>Mus musculus</i>	NP_848469.1	Ten4	<i>Homo sapiens</i>	NP_001092286.2
Unc5A	<i>Homo sapiens</i>	BC167811.1			
Unc5B	<i>Homo sapiens</i>	XR_945617.1			
Unc5D	<i>Rattus norvegicus</i>	NM_001107319.2			

Table 2.1: **Summary of cDNAs used for sub-cloning.** The accession number refers to the protein-product of the cDNA.

2.1.2 Cloning Vectors

The main vector for expression in mammalian cells was pHLSec and variants thereof [14]. pHLSec encodes an N-terminal signal peptide derived from receptor-type tyrosine-protein phosphatase S and a C-terminal hexahistidine tag. A variant in which an “Avi” tag is inserted before the C-terminal hexahistidine tag is also available, allowing constructs to be biotinylated. The N-terminal secretion signal allows secretion of recombinant proteins into the extracellular medium. Constructs cloned via the *AgeI/KpnI* cloning sites are fused to this signal sequence. Alternatively, use of the *EcoRI/KpnI* cloning sites allows the use of a protein’s native signal peptide. For some of the described stripe assays, constructs cloned into the pCAGIG vector were used, allowing expression of the gene of interest on a bicistronic mRNA also encoding an internal ribosome entry site (IRES) and GFP [163].

2.1.3 Polymerase Chain Reaction (PCR)

PCR primers used to generate the various FLRT, Unc5, Lphn and Teneurin constructs are summarised in Appendix A. PCRs employed either Pyrobest DNA Polymerase (Takara-Bio, Clontech) or CloneAmp HiFi PCR Premix (Takara-Bio, Clontech) according to the manufacturer’s instructions. For Pyrobest PCRs, a typical reaction contained 500 ng plasmid DNA, 100 pmol of the forward and reverse primers, 0.25 mM dNTPs and 5 U of Pyrobest DNA Polymerase in Pyrobest Buffer II (Takara-Bio, Clontech). For CloneAmp PCRs, a typical reaction contained 100 ng plasmid DNA and 7.5 pmol of the forward and reverse primers, along with CloneAmp HiFi PCR Premix. The PCRs were carried out in an MJ Mini Personal Thermal Cycler (Bio-Rad). Pyrobest PCRs were initially heated to 95 °C for 5 min, followed by 30 thermal cycles (40 s at 95 °C, 40 s at 55 °C and then 1-10 min at 72 °C depending on the length of the desired product), and ending with 5 min at 72 °C. CloneAmp PCRs were performed over 35 thermal cycles (10 s at 98 °C, 15 s at 55 °C, 15 s at 72 °C). PCR products were then separated by agarose gel electrophoresis.

2.1.4 DNA Agarose Gel Electrophoresis

Agarose was dissolved at a concentration of 1.0% in TAE buffer (40 mM Tris Base, 20 mM acetic acid, 1 mM EDTA). The mixture was then melted and SYBRTM Safe DNA Gel Stain (Invitrogen) added (1:10000 dilution), and the gel allowed to set. The DNA fragments were electrophoresed at 150 V for 35 min. Bands were visualised using a Dark Reader[®] transilluminator.

2.1.5 Restriction Digestion and Ligation

Restriction digest reactions typically contained 3-20 μ g DNA and 5 U of each of the relevant restriction enzymes (NEB) in the appropriate NEB buffer. Reactions were incubated for 2 hours (for PCR products) or 6 hours (for plasmids). Plasmid digestions also included 10 U calf intestinal alkaline phosphatase (NEB). Digested DNA was purified (Qiagen Gel Extraction kit) either from agarose gel or directly. To ligate digested PCR products and vectors, 1 μ l Quick Ligase was mixed with approximately 50 ng insert and 50 ng vector in 1x Quick Ligation Reaction Buffer (NEB). The ratio of insert to vector was adjusted according to their relative molecular weights. Ligation reactions were then incubated for 10 min at room temperature.

2.1.6 In-FusionTM Cloning

1 μ l of In-FusionTM reaction premix was mixed with 50-100 ng insert and 50 ng vector. The final mix was adjusted to a volume of 5 μ l with milliQ water and incubated at 55 °C for 15 min.

2.1.7 Transformation and Clone Verification

Sub-cloning-efficiency chemically-competent DH5 α cells (Life Technologies) were transformed with 3 μ l of the ligation or In-FusionTM reaction. Cells were then incubated on ice for 30 min, heat shocked for 45 s at 42 °C, and then returned to ice for 2 min before plating on agar plates containing the required antibiotic

(100 ug/ml ampicillin). Plates were then incubated at 37°C overnight in a static incubator. Single colonies were selected and used to inoculate 10 ml of ampicillin-containing Lysogeny broth (LB). Cultures were incubated at 37°C overnight with shaking, before DNA extraction (Qiagen Miniprep kit). Clones were verified by Sanger Sequencing (Source Biosciences, Nottingham, UK).

2.2 Protein Expression and Purification

2.2.1 Cell Lines and Control of *N*-linked Glycosylation

HEK293 cells were used for secretory recombinant expression of receptor ectodomains. HEK293T cells, which produce complex-type *N*-linked glycans, were most extensively used throughout this project. Complex-type *N*-linked glycosylation is highly heterogenous and is, in general, a barrier to crystallisation. Treatment of HEK293T cells with an α -mannosidase I inhibitor called kifunensine restricts cells to forming only oligomannose-type *N*-glycans [38]. Another cell line, HEK293S, lacks the gene for the enzyme *N*-acetylglucosaminyltransferase I (GnT1^{-/-}) [210], resulting in predominantly Man₅GlcNAc₂ *N*-linked glycans (Figure 2.1). Glycans produced by these cell lines are substrates for the enzyme endoglycosidase F1, which reduces *N*-linked glycans to single GlcNAc moieties, further reducing heterogeneity [38].

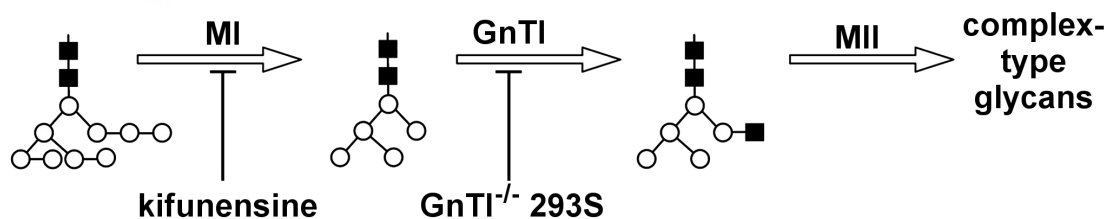


Figure 2.1: **Control of *N*-linked glycosylation for crystallography.** Adapted from [38]. Monosaccharide units are depicted as follows: GlcNAc - closed squares; Man - open circles. MI - α -mannosidase I, GnT1^{-/-} - *N*-acetylglucosaminyltransferase I, MII - α -mannosidase II.

2.2.2 Mammalian Cell Culture Maintenance

Mammalian cells (HEK293T, HEK293S and HeLa) were maintained in T75 flasks in Dulbecco's Modified Eagle's Medium (DMEM; Sigma) supplemented with 10% foetal bovine serum (FBS; Gibco), 1% non-essential amino acids (Gibco) and 1% L-Glutamine (Gibco). Cells were grown at 37 °C in a 5% CO₂ atmosphere. The cells were split at 80-90% confluence. For splitting, the medium was removed and the cells washed with 10 ml PBS (Lonza). Cells were detached by treatment with 2 ml 0.05% Trypsin-EDTA (Gibco) at 37 °C for 4 min and then resuspended in fresh medium to the desired density.

2.2.3 Small-scale Transient Transfection

Twenty-four hours prior to transfection, cells were seeded into either 24-well or 6-well plates at the densities described below (Table 2.2). Plasmid DNA was diluted in the appropriate volume of OptiMEM (Gibco) and Fugene6 (Promega, 1 $\mu\text{g}/\mu\text{l}$) transfection reagent added at a ratio of 1:3 (μg of DNA: μg of Fugene, Table 2.2). Transfection cocktails were mixed by vortexing and incubated for 10 min at room temperature to allow DNA:Fugene complexation. The transfection cocktail was added drop-wise to the cells in 10% FBS DMEM. Cells were then grown for the time period required for the experiment (for expression trials this was typically 2-3 days, for stripe assays just 1 day). Similar protocols were used for transient transfection with polyethyleneimine (PEI, Sigma).

Plate type	Surface Area (cm ²)	Seeding density	Growth medium (ml)	Amount of DNA for transfection (μg)	Amount of Fugene6 for transfection (μg)
6-well	9	0.3×10^6	2	3	9
24-well	2	0.05×10^6	0.5	0.75	2.25

Table 2.2: Ratios of reagents used for small-scale mammalian cell culture and transfection

2.2.4 Polyacrylamide Gel Electrophoresis

Protein samples were mixed with 6x loading dye (375 mM Tris-HCl pH 6.8, 6% SDS, 48% glycerol, 9% β -mercaptoethanol, 0.03% bromophenol blue) and boiled for 5 min at 95 °C. Samples were loaded onto NuPAGE® 4-12% Bis-Tris Gels and run using NuPAGE® MOPS SDS Running Buffer (both Invitrogen). Gels were run in an XCell SureLock™ Mini-Cell at 200 V for 45 min, and then stained with SimplyBlue™ SafeStain (Life Technologies).

2.2.5 Western Blotting

To detect epitope-tagged proteins, proteins from an unstained gel were transferred onto a nitrocellulose membrane (GE Healthcare) using the NuPAGE® XCellIII™ Blot module and NuPAGE® Transfer Buffer. The membrane was soaked for 10 min in transfer buffer before assembling the transfer cassette. The membrane and the gel were placed between Whatman blotting paper and six sponges, all of which were soaked in transfer buffer before use. A potential difference of 30 V was applied across the transfer cassette for 70 min. The membrane was blocked in PBS + 5% milk + 0.1 % Tween-20 (blocking buffer) for 1 hour at room temperature. The blot was incubated with either 1:500 anti-pentaHis (mouse, Invitrogen) or 1:1000 anti-FLAG (rabbit, Sigma) as primary antibody, depending on the epitope being detected, in blocking buffer for 1 hour at room temperature or overnight at 4 °C. The membrane was washed three times with PBS + 0.1% Tween-20 (PBST) and then incubated with either goat anti-mouse (Sigma) or goat anti-rabbit (Abcam) antibody conjugated to horse radish peroxidase (1:10000) in blocking buffer for 1 hour. Finally, the blot was washed three times with PBST and developed using the Amersham™ ECL™ Western Blotting Detection Kit (GE Healthcare) and bands were visualised using Amersham Hyperfilm™ ECL.

2.2.6 Roller Bottle Preparation

HEK293T or HEK293S cells were seeded into T175 flasks (5×10^6 cells) and grown to 90-100% confluency. The cells from one T175 flask were used to seed one roller bottle containing 250 ml DMEM supplemented with 10% FBS. Roller bottles were incubated at 37°C in a 5% CO₂ atmosphere in a Wheaton roller bottle incubator for 3 days prior to transfection. Roller bottles were typically prepared in batches of either 6 or 12 by Mrs. Maria Carrasquero.

2.2.7 Roller Bottle Transfection

Plasmid DNA was prepared for transfection by vortexing the DNA stock with 80 μ l chloroform (Sigma) per 1 ml DNA and centrifuging at 20,000 x g for 10 min. The transfection cocktail for one roller bottle was prepared by mixing 0.5 mg DNA with 1 mg PEI in 50 ml serum-free DMEM. The transfection cocktail was incubated for 10 min at room temperature. Medium from the roller bottles was discarded and replaced with 200 ml DMEM supplemented with 2% FBS plus 50 ml of the transfection cocktail. Roller bottles containing HEK293T cells were also supplemented with 0.38 mg kifunensine per roller bottle to prevent the formation of branched glycans [38]. Cells were grown for 4 days (HEK 293T) or 7 days (HEK 293S) post-transfection, allowing for the production of secreted recombinant protein. Roller bottles were typically transfected by Mrs. Maria Carrasquero.

2.2.8 *In Vivo* Biotinylation

For *in vivo* biotinylation, constructs were cloned into the pHLSec vector and fused to a C-terminal Avi- and hexahistidine tag via a GGS linker region [14]. The AviTag comprises a biotin ligase (BirA) recognition site (sequence GLNDIFEAQKIEWHE). HEK293T cells seeded into 6-well plates were transfected with 2.25 μ g plasmid DNA encoding the Avi-tagged ectodomain plus 0.75 μ g vector encoding BirA fused to an endoplasmic reticulum retention signal. Cell culture medium was supplemented with 100 mM biotin. Cells were grown for 2-3 days post-transfection, after which the

medium was harvested and clarified by centrifugation. Excess biotin was removed by overnight dialysis against PBS at 4 °C.

2.2.9 Purification of His-tagged Proteins

Medium from roller bottles was harvested and clarified by centrifugation at 6000 x g for 20 min, followed by filtration through 0.22 μ m filters (Millipore, Merck). The cell culture medium was exchanged into diafiltration buffer (20 mM Tris (pH 8.0) 500 mM NaCl) with a tangential flow filtration device (Pall) and concentrated to $\sim$ 500 ml. The concentrated cell culture medium was supplemented with 20 mM imidazole (pH 8.0) and applied to a 5 ml HisTrap FF column (GE Healthcare). The column was washed with diafiltration buffer and then with 100 ml wash buffer (20 mM Tris pH 8.0, 300 mM NaCl, 40 mM imidazole). Bound protein was either eluted using 100% elution buffer (20 mM Tris pH 8.0, 300 mM NaCl, 500 mM imidazole) in 1.5 ml fractions, or eluted using a 0-100% gradient of elution buffer against 20 mM Tris pH 8.0, 300 mM NaCl.

2.2.10 Size-Exclusion Chromatography

Size-exclusion chromatography (SEC) is the final step of protein purification, and is used to separate aggregates from monodisperse protein. Typically a Superdex 200 16/60 HiLoad column (GE Healthcare) was equilibrated with 10 mM Tris (pH 7.5), 150 mM NaCl and the protein loaded onto the column. Fractions of 1-2 ml were collected and analysed using SDS-PAGE. Fractions containing sufficiently pure protein were pooled and concentrated using an Amicon Ultra centrifugal filter device. Protein was either aliquoted and stored at -80°C or used immediately.

2.3 Biophysical Methods

2.3.1 Surface Plasmon Resonance (SPR)

SPR is a biophysical method for measuring molecular interactions between unlabelled molecules [65]. In an SPR experiment one of the molecules (the ligand) is immobilised on a chip surface and analyte (often protein, but this can also be a small molecule) is flowed over.

An SPR chip comprises a glass support coated with a thin gold film and a flow cell containing a layer of dextran used for immobilising the ligand. The glass support is illuminated with a wedge of plane-polarised light under conditions of total internal reflection, resulting in the reflection of the polarised light. An evanescent wave is also generated, and the parallel-polarised component of this evanescent wave enters the gold layer. At a certain angle of incidence, the evanescent wave excites delocalised surface electrons, or plasmons, in the gold layer, greatly magnifying the evanescent wave. This results in the reduction in the observed intensity of the reflected light at this resonance angle. This resonance angle depends upon the refractive index of the flow cell at the sensor surface, which is directly influenced by the mass of protein bound to the surface. As the analyte is flowed over and binds the ligand, the mass of protein bound to the surface increases, increasing the refractive index in the flow cell, resulting in a change in the SPR angle. The change in SPR angle is measured over time, allowing complex association and dissociation to be followed. This signal is measured in arbitrary resonance units (RUs).

SPR equilibrium experiments were performed using a Biacore T200 instrument (GE Healthcare). To test binding of FLRT2 to mutant Lphn3^{Lec-Of} constructs, biotinylated wild-type and mutant Lphn3^{Lec-Of} constructs were immobilised on a CAP sensor chip using the Biotin CAPture kit (GE Healthcare) at densities of 450-600 RU. A concentration series of FLRT2^{Ecto} was injected over the chip for 500 s at 5

$\mu\text{l}/\text{min}$, followed by a 100 s dissociation time. The chip surface was regenerated with CAPture regeneration solution for 30 s at $5 \mu\text{l}/\text{min}$.

For all other experiments, CM5 chips were used (GE Healthcare). CM5 chips were coated with streptavidin via amine coupling (GE Healthcare Kit), allowing biotinylated proteins to be attached to the chip. Ligands were immobilised at densities of 500-1000 RU. A concentration series of the analyte was prepared, depending on the expected K_D . Each concentration of the analyte was injected for 300-500 s contact time, followed by a 100 s dissociation time. The chip was regenerated using 2 M MgCl_2 for 30 s. All steps were carried out at a flow rate of $5 \mu\text{l}/\text{min}$. All experiments were performed at 25°C at pH 7.5 (PBS + 0.005% Tween-20).

To obtain values for the specific binding, the response from a control flow cell coated only with either CAPture reagent (GE Healthcare) or streptavidin (depending on the experimental set-up) was subtracted from the response of each experimental flow cell. K_D and maximum analyte binding (B_{max}) values were obtained by nonlinear curve fitting of a 1:1 Langmuir interaction model:

$$\text{bound} = \frac{B_{\text{max}}C}{K_D + C}$$

where C is the concentration of the analyte as a monomer. Data were analysed using BIAevaluation software 1.0 (GE Healthcare) and Prism 7 (GraphPad).

2.3.2 Native Mass Spectrometry

Native mass spectrometry (MS) relies on the ability of non-covalent interactions in protein complexes to be maintained in the gas phase. Protein complexes are ionised using a technique called electrospray (ES) ionisation and maintained intact in a mass spectrometer [214]. ES is a soft ionisation technique, allowing noncovalent interactions to be preserved, meaning that macromolecular complexes can be studied. In ES, a high, positive electrostatic charge is applied to the tip of the sample capillary,

leading to production of droplets of sample and solvent. For native MS, nanoflow ES is the preferred ionisation technique. This allows for a reduced flow rate compared to conventional ES techniques, resulting in smaller droplets, improved sensitivity and a greater tolerance of buffer salts [93].

ES ionisation requires volatile buffer components, and so many common buffer components (salts, detergents, glycerol etc.) are incompatible. The most commonly used buffer is aqueous ammonium acetate as both ammonia and acetic acid evaporate readily during ES [264]. Protein samples for MS were concentrated to 10 to 15 μ M and buffer exchanged using dialysis into 200 mM or 500 mM ammonium acetate (pH 7.5) at room temperature overnight or using micro Bio-Spin Columns (BioRad). The samples in 500 mM ammonium acetate were diluted with milliQ water to a final ammonium acetate concentration of 200 mM immediately prior to MS analysis. Sample preparation for each MS experiment described in this thesis is summarised in Table 2.3.

Sample	Final Protein Concentration	Buffer Exchange Method	Final ammonium acetate concentration
Lphn3 ^{Lec-Olf} , FLRT2 ^{LRR} , Unc5D ^{Ig}	0.5 mg/ml	Bio-Spin into 500 mM ammonium acetate (pH 7.5)	200 mM
Lphn3 ^{Lec-Olf} , FLRT2 ^{LRR} , Unc5D ^{Ecto}	1.5 mg/ml	Dialysis into 500 mM ammonium acetate (pH 7.5)	200 mM
Lphn3 ^{Lec-Olf} D397R D398R D399R, FLRT2 ^{LRR} , Unc5D ^{Ig}	1.5 mg/ml	Dialysis into 500 mM ammonium acetate (pH 7.5)	200 mM
Lphn3 ^{Lec-Olf} D397N D399T, FLRT2 ^{LRR} , Unc5D ^{Ig}	1.5 mg/ml	Dialysis into 500 mM ammonium acetate (pH 7.5)	200 mM
Lphn3 ^{Lec-Olf} , FLRT3 ^{LRR} , Unc5B ^{Ecto}	1.6 mg/ml	Dialysis into 500 mM ammonium acetate (pH 7.5)	200 mM
Lphn3 ^{Lec-Olf} D397N D399T, FLRT3 ^{LRR} , Unc5B ^{Ig}	1.6 mg/ml	Dialysis into 500 mM ammonium acetate (pH 7.5)	200 mM
Lphn1 ^{Lec-Olf} , FLRT2 ^{LRR} , Unc5D ^{Ig}	1.6 mg/ml	Dialysis into 500 mM ammonium acetate (pH 7.5)	200 mM

Table 2.3: **Native MS Sample Preparation.**

Intact macromolecular complexes are analysed using hybrid mass spectrometers, containing both quadrupole (Q) mass filters and a time-of-flight (ToF) analyser [239]. In a mass spectrometer, ions are separated by their mass-to-charge ratio (m/z value). Ions with different m/z values will have different velocities, causing them to reach the detector at different times. Sample solutions are introduced into the mass spectrometer using gold-coated capillary needles, which are prepared in-house using a Polaron range sputter coater and then trimmed to size under a microscope [93]. Partially desolvated ions are accelerated through a Hexapole ion guide and then sequentially enter the Quadrupole analyser and Hexapole collision cell, passing through pressure and potential differences. During this time, ions continue to desolvate until multiply-charged protein ions remain. Ions then enter the ToF analyser where they are deflected by the reflectron, this reduces the spread of flight times of ions with the same m/z and improves the mass resolution, before they hit the microchannel-plate detector. For tandem MS experiments (MS/MS), ions with a particular m/z are selected in the Quadrupole. The voltage in the collision cell is

raised so as to accelerate the precursor ions and facilitate their collision with argon atoms, leading to their dissociation.

Native MS experiments were performed on a modified Q-ToF tandem mass spectrometer [239], and on a Synapt G1 mass spectrometer (both Waters, Manchester, UK). Experiments were carried out under conditions optimised for efficient desolvation and transmission of high molecular weight protein complexes. Typical instrument settings were: capillary voltage 1.5 kV, cone voltage 100-200 V and collision cell energy 20-50 V. For tandem MS experiments the collision energy was raised to 150 V. All mass spectra were calibrated off-line using a 10 mg/ml solution of cesium iodide. All native mass spectrometry data acquisition and analysis was performed in collaboration with Dr. Shahid Mehmood (Robinson lab, University of Oxford).

2.3.3 Size-Exclusion Chromatography-Multi-Angle Light Scattering (SEC-MALS)

SEC-MALS is a technique which uses light scattering to measure the molecular weight of macromolecules in solution. In a SEC-MALS experiment, a sample of macromolecules is injected onto a size-exclusion column, allowing the components to be separated by size. A concentration detector (differential refractometer, dRI) and a set of light scattering (MALS) detectors at different angles are connected in series after the column and acquire data at each elution volume. As the sample leaves the column it is irradiated by a laser beam. This induces an oscillating dipole in the molecules which re-radiates (scatters) light. When a protein is present as a monomer, individual molecules undergo random Brownian motion and, as such, the phase of the scattered light from two individual monomers is incoherent. In a dimer, the scattered light from one monomer is coherent with the scattered light from the other monomer, meaning that the light scattered from a dimer is twice as intense as that scattered by two independent monomers. The intensity of the scattered light hitting a detector at angle, θ , is proportional to the molecular weight

of the macromolecule, M , its concentration, c and its refractive index, $\frac{dn}{dc}$.

$$I(\theta)_{scattered} \propto Mc \left(\frac{dn}{dc} \right)^2$$

Light scattered from different parts of the macromolecule reaches the detectors with different phases, leading to destructive or constructive interference, which varies depending on the measurement angle. This phase shift is highly dependent on the shape of the protein, which is taken into account in the calculation of the molecular weight.

The refractive index $\frac{dn}{dc}$ of the macromolecule is largely independent of the amino acid composition, but does vary according to the carbohydrate contribution to the molecular weight [12]. For all calculations for data from glycoproteins, the $\frac{dn}{dc}$ value was adjusted accordingly (Table 2.4) .

Cell Line	Molecular weight of glycan (kDa)	Reference
HEK293T	1.26	Radu Aricescu, Unpublished
HEK293T + Kif	2.06	Chang <i>et al</i> (2007)[38]
HEK293S	1.4	Crispin <i>et al</i> (2006)[45]

Table 2.4: **Molecular weights of *N*-linked glycans used for $\frac{dn}{dc}$ adjustment**

In the experiments described in this thesis, all samples were loaded onto a Superose6 10/30 column (GE Healthcare) equilibrated in 10 mM Tris (pH 7.5) and 150 mM NaCl. The eluate was analysed using laser light scattering detected at 662 nm wavelength at scattering angles between 20.6° and 149.1° using a Heleos 8 instrument (Wyatt Technology). The data were analysed using the ASTRA 6.1 software (Wyatt Technology).

2.4 X-ray Crystallography

X-ray crystallography involves the elastic scattering of an X-ray beam incident on a crystal. X-rays are diffracted by the electron cloud surrounding the atomic nucleus of the atoms in the crystal. During a diffraction experiment, the crystal is rotated by sequential small increments on a goniometer and a series of diffraction patterns are collected by a detector. The organisation of protein molecules into arrays, as seen in a crystalline lattice, means that X-rays diffracted by the crystal form distinct beams, each of which produces a distinct spot (reflection) on the detector. During the structure determination process, information about the positions and intensities of the spots is combined with phase information, allowing a three-dimensional electron density map of the crystal to be calculated using Fourier methods. The next section discusses aspects of the structure-determination pipeline and the principles behind the methods used in this thesis.

2.4.1 Crystallisation and Crystal Optimisation

Structure solution by X-ray crystallography requires that well-ordered crystals of the protein being studied are available. Protein crystals are produced when protein molecules are driven out of solution by the presence of a precipitant. A solution containing a precipitant is mixed with the highly concentrated protein solution (typically 5-25 mg/ml), generating a supersaturated solution. Both protein and precipitant concentration are crucial factors in determining whether or not crystals will form. When the concentration of both the protein and precipitant are high, the protein is likely to precipitate and form amorphous material. At lower concentrations, crystals may form if the concentration is still sufficiently high to allow crystal nuclei to appear. At still lower concentrations, proteins are likely to be soluble.

Crystallisation trials can be performed in various possible formats, the most common being the vapour diffusion method. Vapour diffusion employs a drop containing protein mixed with the reservoir solution. The reservoir solution typically contains

different mixtures of precipitant, buffers, additives and detergents. Vapour diffusion relies on the gradual diffusion of water from the protein-containing drop (where the concentration of precipitant is lower), to the reservoir, concentrating the protein and reducing its solubility.

Protein crystallisation is empirical in nature and typically involves screening a large number of crystallisation conditions, including systematically screening both crystallisation temperature and protein concentration. The most common types of screens employed during initial crystallisation experiments are sparse matrix screens and grid screens. Sparse matrix screens include conditions that have previously been successful in crystallising other proteins [102], and include commercially-available screens such as JCSG+ and Proplex (both Molecular Dimensions). Traditional grid screens, such as PACT premier (Molecular Dimensions), involve varying two factors systematically in a two-dimensional matrix and can also prove successful in identifying initial crystal hits. The MORPHEUS screen has a three-dimensional grid formulation, where eight additive mixes are combined with four precipitant mixes and three buffer systems [81].

Once initial crystals have been obtained, optimisation of the crystallisation condition is often required to yield diffraction-quality crystals. This is usually done by generating grid screens around the initial crystallisation condition and altering the protein:reservoir ratio in the crystallisation drop. Additive screens, such as Silver Bullets and Additive Screen HT (both Hampton Research) can be added to either the reservoir or the crystallisation drop. Microseeding is also commonly used during optimisation to promote crystal nucleation [20], whereby small crystals are harvested and vortexed with a “Seed-Bead” to generate sub-microscopic crystal microseeds. Microseeds are then added to new crystallisation drops where they can act as nucleation sites for further crystals. Because it is difficult to control the number of micro-seeds that are transferred, they are usually used as a serial dilution across

the plate.

Protein crystals are exquisitely sensitive to ionising X-ray radiation. During a diffraction experiment a large proportion of the X-ray beam is absorbed by the crystal and this energy generates free hydrogen and hydroxyl radicals from components of the crystal (either solvent or protein). These radicals then propagate through the crystal causing secondary damage to the protein and reducing the diffraction potential of the crystal. In order to slow the propagation of these radicals, data collection is routinely carried out at 100 K. Before X-ray exposure the crystal is “fished” from the crystallisation drop in a cryoloop and flash-cooled in liquid nitrogen. This type of “flash-freezing” will generate ice in the crystal and destroy its diffraction potential unless the crystal is cryoprotected first. This involves brief transfer of the crystal to a solution containing the mother liquor supplemented with a sufficiently high concentration of a cryoprotectant, such as glycerol, ethylene glycol or 2-methyl-2,4-pentanediol. These cryoprotectants prevent water in the mother liquor from freezing, allowing it to vitrify instead. The best cryoprotectant for particular crystals often must be determined empirically.

2.4.2 Data Collection

For most crystal structures solved today, data are collected at synchrotrons. For native data collection, X-rays with wavelengths of 0.9-1.0 Å are used. Initially the crystal is centred, and three to four “test” shots are taken at intervals of 45°. If sufficient strong reflections are present, the crystal can be indexed and a data collection strategy devised to minimise the X-ray dose given to the crystal. These test shots are also used to inform decisions on experiment parameters such as the detector distance, beam size, transmission, exposure time and oscillation angle.

If the crystal exhibits high-resolution diffraction and low mosaicity, a full dataset is recorded. This involves exposing the crystal to the X-ray beam whilst rotating

it around a single axis in small increments (typically 0.1° to 1.5° , depending on the mosaicity and size of the unit cell). The diffraction image during each of these small oscillations is called a frame. The total oscillation required to collect a complete dataset depends on the symmetry of the crystal. Increasing the amount of data recorded increases the multiplicity (i.e. increases the ratio of measured reflections to unique reflections), and improves accuracy if radiation damage is insignificant.

Single Anomalous Dispersion (SAD) Data Collection

Anomalous data collection is typically carried out at longer wavelengths and lower energies than for native data. For a SAD experiment, data is collected above the X-ray absorption edge in order to amplify the anomalous signal from heavy atoms (HAs). This is an element-specific energy, above which the X-ray absorption by the HA rapidly increases, forming an edge in the absorption spectrum. A fluorescence scan of the crystal can be used to indicate the presence or absence of anomalous scatterers in the crystal and provide an experimental determination of the edge energy. The crystal will absorb the most energy at the absorption edge, and so measurements are normally made slightly to the right of the peak.

2.4.3 Data Processing

Data processing begins with indexing, cell refinement and integration which, in this thesis, was performed using the program XDS [105]. The reciprocal lattice spacings, and thus the unit cell parameters, of the crystal determine the spacing of the reflections observed on the detector. When a diffraction pattern is indexed, the spacing between the reflections on the detector is used to calculate a consistent set of vectors spanning the reciprocal lattice represented by the diffraction spots. Modern indexing algorithms use a Fast Fourier Transform (FFT) algorithm to find a consistent set of lattice directions.

Once the crystal and instrument parameters have been optimised by refinement,

integration is performed. Integration of the diffraction spots allows determination of reflection intensities (I) and their errors (σI). Modern integration programs typically use 3D spot profile fitting, whereby the intensity of an individual spot is calculated over a series of images and the spot profile can vary for the same reflection recorded on different images. This allows very accurate background subtraction for low-intensity reflections. In XDS, all pixels belonging to a single reflection are ranked according to increasing intensity. Pixels with the highest intensity are sequentially removed until the distribution of intensities matches that expected for background values. Spot profiles can vary in different parts of the image and so spot profiles are scaled across images and parts of images, such that the sum of its signal pixels normalises to one [104]. At this stage, XDS merges partially recorded reflections and calculates intensities.

The integrated intensities of reflections must be corrected for various physical factors arising from the experiment. These include slow variation in beam intensity, absorption of X-rays by the crystal, sensitivity variations in the detector plane and radiation damage [66]. This is achieved using the program XSCALE [105]. XSCALE minimises the differences between the individual observations of reflection I and the weighted mean of all its symmetry-related equivalents.

Once the observed intensities are on a common scale, reflections related by rotations characteristic of the Bravais lattice symmetry are merged into unique observations. For anomalous diffraction experiments, Friedel pairs should not be merged. The final step in data processing is truncation. Truncation is the process by which intensities are converted to structure factor amplitudes, which are part of the requirement for calculation of an electron density map.

Data Quality Indicators

Data reduction statistics are indicators of data quality. These are calculated for the overall dataset and individual resolution shells (wedges of data containing equal numbers of reflections). This is informative in deciding where to cut the resolution of the dataset. Because diffraction strength decreases with resolution, it is important to apply a high resolution cut-off in order to discard data that are considered so noisy that their inclusion could hinder resulting model quality. Traditional metrics used for this purpose are $I/\sigma I$ and the “merging” R -values. $I/\sigma I$ represents the signal-to-noise ratio, which drops at lower resolutions. R_{merge} measures the spread of n redundant measurements of a reflection’s intensity, $I_i(hkl)$, around their average, $\bar{I}(hkl)$.

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)}$$

However, $I_i(hkl)$ and $\bar{I}(hkl)$ are not independent. The result of this is that, the higher the multiplicity, the higher the value of R_{merge} , making it a poor data quality indicator. Later, R_{meas} a multiplicity-independent merging R-factor was defined [53]. R_{meas} requires that the R_{merge} definition is adjusted by a factor of $\sqrt{n/(n-1)}$, and reliably reports on the consistency of the individual measurements of each reflection, independent of the multiplicity.

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{i=1}^n |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^n I_i(hkl)}$$

More recently, $I/\sigma I$ and merging R -values have been revealed as unreliable indicators for selecting resolution cut-offs, with traditional cut-off values demonstrated to discard data which can improve the quality of refined models [109]. The Pearson correlation coefficient (CC) of two “half” datasets ($CC_{1/2}$) is calculated by dividing the data into two random “half-sets” and calculating the Pearson CC between the average intensities of each subset. $CC_{1/2}$ is the one data-quality indicator which accurately reflects how different data-handling strategies affect the quality of the

refined model [54, 109]. For test cases, inclusion of data to the resolution at which $CC_{1/2}$ is in the range 0.1 -0.2, leads to improvement of the model, although this may vary case-by-case and is ideally determined by a paired refinement [54].

Many statistics are also reported with respect to their image number. This can be useful in determining the extent of radiation damage the crystal has suffered over the course of the experiment. For example, R_{meas} should remain relatively constant across all images, but will increase if radiation damage is occurring due to the data becoming less internally consistent [78]. This allows the user to exclude images for which radiation damage is significant, at the expense of reducing the multiplicity.

2.4.4 Phasing

During an X-ray diffraction experiment, intensities of scattered X-rays are recorded by a detector, yet their phase information is lost - this is known as “the phase problem”. In a unit cell of volume V , the electron density ρ at a point (x, y, z) for each reflection is calculated by a fourier transform shown below. This requires both the structure-factor amplitudes $|F_{hkl}|$ (these are proportional to the square root of the recorded intensities of each spot on the diffraction pattern), and the phases α_{hkl} .

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i((hx+ky+lz)-\alpha_{hkl})}$$

This means that an electron density map cannot be calculated without the phase information of each reflection. The phases must either be recovered using molecular replacement or experimentally. These approaches traditionally make use of the Patterson function, which defines the set of vectors between the atoms of a structure. The Patterson function is essentially the Fourier transform of the intensities (rather than the structure factors) and thus can be used to reconstruct a map in the absence of phases. This is known as a Patterson Map. The Patterson function may be calculated as:

$$P(u, v, w) = \sum_{hkl} |F_{hkl}|^2 e^{-2\pi i(hu + kv + lw)}$$

where (u, v, w) represents an interatomic vector. This three dimensional map contains peaks at all interatomic distances observed in the structure (including self-vectors), and the peak heights are proportional to the product of the number of electrons in the atoms concerned. The Patterson map is always centrosymmetric, irregardless of the structure symmetry, and is independent of the position of the structure in the unit cell. The Patterson map has a number of practical applications in crystallographic structure determination, which will be discussed below.

Phase determination by Molecular Replacement

The most common method to solve the phase problem in macromolecular crystallography is molecular replacement (MR). MR calculates initial estimates of phases by rotating and translating the known structure of a related protein in the unit cell of the crystal until the solution with the best fit to the observed data is obtained. MR is most likely to be successful in cases for which there a large proportion of the asymmetric unit has a high structural and sequence similarity (>30%) to a protein of known structure.

The first step in MR is the generation of good search models, i.e. making the search model maximally similar to the structure of interest. In the absence of an obvious structural homologue, more distant homologues are identified using bioinformatics programs, for example the HHpred server [240]. Sequence alignments can also be used to predict gaps and insertions, and the search model modified accordingly. For distant homologues, polyalanine truncation of the model (either full or partial) can improve results. In addition, an important pre-requisite for a successful MR search is knowledge of how many copies of the protein are present in the asymmetric unit. This can be calculated using the Matthew's Coefficient. The Matthew's coefficient is the ratio of the asymmetric unit volume to the protein molecular weight. A range of

solvent contents are assumed and the number of molecules in the asymmetric unit is calculated for each solvent content value. The most likely solution(s) are then decided based on the probable solvent content of the crystal and/or knowledge of the oligomeric state of the protein in solution.

MR solutions are found by sequential, yet independent, rotation and translation of the search model. Traditional MR approaches are based on Patterson methods. The Patterson map for a given molecule is independent of its position in the unit cell, but rotating the structure in the unit cell rotates the Patterson map about its origin. The rotational search involves superimposing the origins of two three-dimensional Patterson maps and then rotating the Patterson map of the phasing model through three different angles (ϕ, ψ, χ) . The best orientation of the phasing model is found when the correlation between the intermolecular vectors of the two Patterson maps is maximal, as assessed by a rotation function. A simple rotation function, R , can be expressed:

$$R(\phi, \psi, \chi) = \int_{u,v,w} P^{target}(u, v, w) P^{model}[(u, v, w)(\phi, \psi, \chi)] du dv dw$$

At rotation angles (ϕ, ψ, χ) , the rotation function is the integral of the product of two Patterson functions; that of the target molecule $P^{target}(u, v, w)$ and that of the search model $P^{model}(u, v, w)$ rotated by the rotation matrix (ϕ, ψ, χ) . Around the maximum of this function, the rotation search is refined. The top solutions from refinement are used for Patterson translational searches.

For the translation function, only intermolecular Patterson vectors, which measure distances between atoms of neighbouring molecules in the unit cell, are important. Intramolecular Patterson vectors are subtracted from the Patterson map. The fast translation function is used to match intermolecular Patterson vectors. Translation solutions then undergo a “packing test”, whereby overlaps between the components in the unit cell are penalised, and the solutions re-ranked accordingly. Rigid body re-

finement is used to improve the positioning of the molecule(s) and involves minimising the differences between the observed ($|F_{\text{obs}}|$) and calculated ($|F_{\text{calc}}|$) structure factor amplitudes. A value below 50% is indicative of a correct solution.

$$R = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

In this work Phaser was used for MR [166]. Phaser employs maximum likelihood approaches; the idea that the best MR solution is the one which has the highest probability of reproducing the experimental data. Maximum likelihood approaches are able to account for errors in the search model. Another function, the Log-Likelihood Gain on Intensity (LLGI) target adds the ability to account for experimental error. Critically, maximum likelihood MR is optimised when both the estimates of the structure factors and the estimates of their errors are closest to their real values used to generate the data. In addition, maximum-likelihood approaches can account for intensity modulations caused by crystal pathologies such as anisotropic diffraction and translational NCS (tNCS). The combination of these factors increases the chance of finding the correct solution. Phaser also performs more sophisticated packing tests, and does not use the packing test to re-order the translation peaks.

Phaser reports a variety of statistics for each solution it calculates, these can be important indicators of a solution's accuracy. The LLGI directly measures the probability of the solution being correct, and thus its value should be positive and as high as possible. It should also increase as more components of the solution are added. The translation function Z-score (TFZ) indicates the number of standard deviations above the mean value, and the value for the correct solution is generally well separated from the rest of the solutions. A TFZ of >8 is generally indicative of a valid solution. Visual inspection of the electron density maps is also an important part of validating an MR solution.

Density modification, for example solvent flattening, using programs such as PAR-

ROT [44] can be used to drastically improve initial phase estimates. This entails placing a three-dimensional grid over the initial protein map. For each grid point the average electron density within a sphere of a given radius is computed. Density values above a certain cut-off are deemed protein and set to a certain average value. For values below the cut-off the region is said to be solvent and “flattened”. This map undergoes a reverse Fourier Transform to compute structure factors with improved phase estimates. These structure factors are further iteratively modified until convergence is reached.

Experimental Phasing

For crystals of proteins for which a large proportion of the asymmetric unit has no known structural homologues, the initial phases have to be determined experimentally. This is typically achieved by incorporating HAs into the asymmetric unit in an ordered fashion. Due to their high number of electrons, HAs scatter more X-rays than those atoms typically found in amino acids. Unless the protein contains endogenous heavy metal ions, crystals must be derivatised in order to incorporate them.

Crystal Derivatisation

A soaking experiment is the classic method for crystal derivatisation, whereby a heavy metal-containing compound is added to pre-formed native crystals in the crystallisation drop. HAs must be incorporated with high occupancy, meaning they must bind specifically to defined regions of the protein. However, it is crucial that binding of HAs does not disrupt crystal contacts or destroy the integrity of the crystal in any way as this will severely diminish its diffraction potential. The most common ways to produce a HA-derivatised crystal are either by a soaking experiment (where HAs are incorporated into pre-formed crystals) or by selenomethionine incorporation.

In a soaking experiment, many factors can influence if and how a HA binds to a protein. For example the buffer and pH of the crystallisation drop can affect the reactivity and solubility of HA-compounds by influencing the protonation state of the active groups on the protein and by chelating the HAs. The Heavy Atom Data-bank contains information on successful cases of protein derivatisation, including the buffer conditions, pH range and the details of the types of interactions formed with the protein. For example, this information was used to assess which HAs had the highest chance of derivatisation at the pH in the crystallisation drop of the GgTen2^{ΔEGF} crystals (pH 8.5) described in Chapter 6 [99]. The other key variable in a soaking experiment is the time for which the crystal is incubated with the heavy metal. This is highly crystal-dependent and thus must be determined empirically. The final step in a soaking experiment is often a back-soak, where the crystal is transferred out of the HA-containing solution into a cryoprotectant lacking HAs. This allows non-specifically bound HAs to diffuse away and aims to reduce background signal and radiation damage [77].

Assessing Anomalous Signal

Before attempting phase calculation it is sensible to assess if the data collected from the crystal contains significant anomalous signal, and if so, to what resolution it is present. Anomalous signal can be assessed using the anomalous correlation and SigAno statistics generated by XSCALE. Like the other data reduction statistics, these statistics are generated for each resolution shell as well as for the overall dataset. The anomalous correlation is the correlation coefficient between random half sets of anomalous intensity differences. Values down to about 25-30% are still thought to contain significant anomalous signal [222]. SigAno reflects the ratio of the mean anomalous difference to its standard deviation, thus giving a measure of the anomalous signal relative to its error. Values down around 1.0 are considered to have significant anomalous signal.

Phase Calculation from Experimental Data

The first macromolecular structures were solved using a phasing technique known as isomorphous replacement. As Patterson map peak heights are proportional to the product of the number of electrons in the relevant atoms, HA positions can be identified by comparison of the Patterson maps for isomorphous native and derivative crystals, allowing the HA substructure to be solved. Once the marker atom positions are known, phasing equations for the protein reflections can be solved.

More recently anomalous dispersion techniques have supplanted isomorphous replacement. Single/Multiple Anomalous Dispersion (SAD/MAD) experiments exploit the anomalous signal from one (SAD) or more (MAD) datasets from the same crystal. In the absence of HAs, the measured intensities of reflections known as Friedel pairs are identical, this is known as Friedel's law. Above certain wavelengths, HAs scatter X-rays anomalously, leading to a pronounced anomalous scattering component of the structure factor and causing Friedel's law to break down. This results in intensity differences between the two reflections of a Bijvoet pair (acentric Friedel pairs, centric Friedel pairs remain unchanged).

The first step in a SAD structure solution is the location of the HA substructure, this is accomplished using Patterson methods in programs such as SHELXD [230]. A Patterson map is calculated from the anomalous intensity differences, this generates a map that only relates to vectors between anomalously scattering atoms. The search begins with atoms consistent with the anomalous Patterson map, this is known as Patterson seeding. Dual-space direct methods are then used to modify the calculated substructures by real space fit to the density and reciprocal space refinement of the phases. The Correlation Coefficient between the observed and calculated normalised structure factors are used to identify the correct solution.

Once the HA substructure is known, the phase information for the rest of the struc-

ture can be reconstructed. This can be done using either SHELXE [230] or Phaser [166]. To understand the theory behind phase calculation using SAD, it is convenient to describe a structure factor in terms of the components which contribute to it:

$$\mathbf{F}_{PA} = \mathbf{F}_P + \mathbf{F}_A + i \cdot \mathbf{F}_A''$$

Where $\mathbf{F}_{PA}$ is a vector representing the total scattering by both the protein and anomalous scatterers, $\mathbf{F}_P$ is the contribution of protein (not observed in a SAD experiment) and $\mathbf{F}_A$ and $\mathbf{F}_A''$ are the real and imaginary anomalous components, respectively. $\mathbf{F}_{PA}$ is experimentally determined, and $\mathbf{F}_A$ is known from the substructure determination. $\mathbf{F}_A''$ is calculated using knowledge of the HA present and the wavelength used and is always at 90° to $\mathbf{F}_A$ (Figure 2.2A).

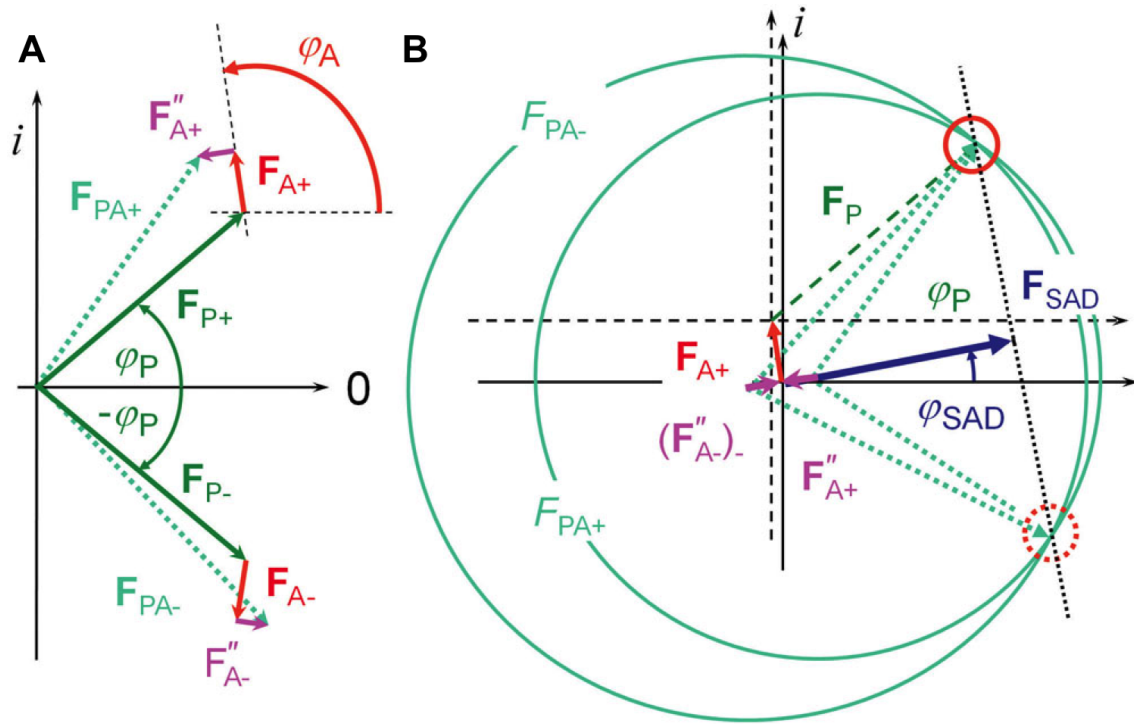


Figure 2.2: **Graphical Representation of SAD Phasing Equations.** Figure taken from [218].

The real anomalous components ($\mathbf{F}_{A+}$ and $\mathbf{F}_{A-}$) are used to offset the imaginary anomalous components ($\mathbf{F}_{A+}''$ and $\mathbf{F}_{A-}''$) from the origin. Circles of radius $\mathbf{F}_{PA+}$ and $\mathbf{F}_{PA-}$ are constructed around the new origins of $\mathbf{F}_{A+}''$ and $\mathbf{F}_{A-}''$ respectively and

the points of intersection of these circles can be used to calculate two solutions for the phase. Density modification is used to improve the phase estimation. Visual inspection of the density-modified map calculated for each phase solution is usually sufficient to determine which enantiomorph is the correct solution.

Experimental phasing protocols have now been combined with a variety of other structure solution softwares in order to extract initial phases from data with weaker anomalous signal or low resolution. For example, Crank2 [196, 238] combines experimental phasing, density modification, model building and refinement leading to an improvement in automated structure solution in some cases. Phaser MR-SAD [209] can combine both MR and SAD approaches and use a poor MR model to assist in determining the substructure. Manual improvement of the model can then be used to iteratively improve the SAD phases.

2.4.5 Model Building and Refinement

After successful initial phasing (either by MR or experimental phasing) and density modification, the electron density map can be interpreted by placing atomic coordinates in it. The process of model building entails adjusting the atomic coordinates in order to improve the agreement between the amplitudes calculated from the atomic model and the measured amplitudes from the diffraction experiment. As this agreement improves, so does the quality of the electron density maps.

If the electron density is of sufficient quality, autobuilding is attempted, using programs such as Buccaneer [43]. Initially likely C_α positions are found and then grown and joined into chains. If the density allows it, sequence will also be docked. This process is usually most successful for defined secondary structure elements. This model is then manually edited in Coot in order to connect pieces of secondary structure and improve the fit to the density [63]. At this early stage in model building, manually-edited autobuilt models are typically used to iteratively improve the

phases of the electron density map. Care must be taken to prevent biasing the map towards poor phases calculated from an inaccurate model.

Once an initial model has been built, a restrained reciprocal space refinement is performed. This involves iterative adjustment of the model parameters in order to improve the agreement of the model with the data. These parameters include atomic parameters (the coordinates x, y, z , B-factors and occupancy) and global parameters (overall B-factor and bulk solvent scattering). This can be carried out using programs such as BUSTER [30], which uses Bayesian statistics for reciprocal space refinement. The B-factor is a measure of the atomic displacement of an atom from its mean position, resulting from thermal motion, lattice disorder or several partially occupied conformations. B-factors can be refined either isotropically (where the thermal motion available to an atom is distributed as a sphere) or anisotropically (using translation, libration, screw motion, TLS). In TLS refinement, TLS groups are defined by identifying groups of atoms whose motion can be parametrised as a combination of translation, libration and screw motion. By default one individual chain is treated as a TLS group, although a more sophisticated analysis is possible based on B-factors using the TLSMD web server [195].

In macromolecular crystallography the experimental data-to-parameter ratio is typically too low for unrestrained refinement. Stereochemical restraints allow the model stereochemistry to be kept within a physically reasonable range and impose limits on features such as bond lengths, angles and torsion angles etc. For small molecule ligands, including monosaccharides of *N*-linked glycans, additional stereochemical restraints must be generated using programs such as Acedrg [156]. Additional restraints can be generated from non-crystallographic symmetry if there is more than one copy of a protein in the asymmetric unit. This causes the NCS copies to be treated as identical and restrains them to remain similar to each other. For refinement of crystal structures at low resolution, a higher resolution structure of a highly

similar protein can also be used to generate restraints for the refinement.

Each refinement cycle should be allowed to run until convergence, at which point a new electron density map with improved phase estimates will be calculated. To minimise the effects of model bias, alternative Fourier transforms are used to calculate the electron density. The most commonly used of these is a $2\mathbf{F}_{\text{obs}} - \mathbf{F}_{\text{calc}}$ map, which reduces the contribution of the structure factors calculated from the model. This is calculated as follows:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (2|F_{\text{obs}}| - |F_{\text{calc}}|) e^{-2\pi i((hx+ky+lz)-\alpha_{\text{calc}})}$$

A “difference map” ($\mathbf{F}_{\text{obs}} - \mathbf{F}_{\text{calc}}$) is calculated, and displays positive density in unmodelled regions and negative density where atoms have been placed erroneously. This map is then used for real-space refinement to place missing components or fix problem areas. This process is repeated iteratively over the course of refining a structure.

Statistically, refinement aims to minimise the sum of residuals squared between the experimentally-determined structure factors ($|F_{\text{obs}}|$) and those calculated from the model ($|F_{\text{calc}}|$). The success of the process is assessed using the R -factor.

$$R = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$$

Before any refinement of the model, reflections are divided into randomly chosen “working” and “test” sets. R -factors are calculated independently for each set during refinement. The working set of reflections is used in refinement, whereas the reflections from the test ($\sim 5\%$ of the total reflections) set never enter the refinement procedure and are only used to calculate the cross-validation R -value (R_{free}) [31]. R_{free} is related to the mean phase error [32], making it an ideal metric to measure model quality. As more parameters are introduced and the model becomes more

complete, the R -factors should fall. After a point the refinement will be optimal, and introduction of further parameters (such as unjustified solvent molecules) will not improve the model, causing the R_{free} to inflate, whilst the R_{work} continues to fall. This is indicative of an over-fitted model.

2.4.6 Validation

The final step in a structure solution is model validation, which ensures that the model is geometrically plausible. MolProbity [48] reports incidences of clashes, Ramachandran outliers, C_{β} deviations, poor rotamers and cis peptides, all of which should be minimised or, ideally, absent in a well-refined structure (unless there is good evidence otherwise). Privateer can be used to validate glycan geometry [6].

2.4.7 Crystallisation and Structure Solution of the Lphn3^{Lec-Olf}, FLRT2^{LRR}, Unc5D^{IgIgTsp} Complex

Crystallisation

Lphn3^{Lec-Olf}, FLRT2^{LRR} and Unc5D^{IgIgTsp} produced in HEK293T cells treated with kifunensine were purified as described in Section 2.2. After elution from the Ni column, all proteins were deglycosylated at 4 °C overnight and then subjected to size-exclusion chromatography. Purified proteins were mixed in a 2:1:1 molar ratio and concentrated to 8 mg/ml in 10 mM Tris-HCl (pH 7.5), 150 mM NaCl and 100 mM NDSB-256. Crystals were grown by the vapour diffusion method at 18 °C by mixing protein and crystallisation solution in a 1:1 (v/v) ratio for four separate crystallisation solutions (crystallisation solutions 1-4). Crystallisation solution 1: 0.1 M MES (pH 6.0), 20% (v/v) 2-methyl-2,4-pentanediol (MPD), 2 mM HEPES (pH 6.8), 0.025% (w/v) benzidine, 0.025% (w/v) nicotinamide, 0.025% (w/v) pyromellitic acid and 0.025% (w/v) sulphaguanidine. Crystallisation solution 2: 0.1 M MES (pH 6.0), 20% MPD, 2 mM HEPES (pH 6.8), 0.016% (w/v) 3-Indolebutyric acid, 0.016% (w/v) hexadecanedioic acid, 0.016% (w/v) oxamic acid, 0.016% (w/v) pyrometal-

lic acid, 0.016% (w/v) sebacic acid and 0.016% (w/v) suberic acid. Crystallisation solution 3: 0.1 M MES (pH 6.0), 20% (v/v) MPD, 2 mM HEPES (pH 6.8), 0.02% (w/v) 2,5-pyridinedicarboxylic acid, 0.02% (w/v) pyromellitic acid, 0.02% (w/v) salicylic acid, 0.02% (w/v) trans-1,2-cyclohexanedicarboxylic acid and 0.02% (w/v) trans-Cinnamic acid. Crystallisation solution 4: 0.1 M ammonium sulphate, 0.3 M sodium formate, 3% (w/v) poly-gamma-glutamic acid LM, 10% (w/v) polyethylene glycol monomethyl ether 2000 and 0.1 M sodium cacodylate (pH 6.5). Crystals from crystallisation conditions 1-3 were cryo-protected using 95% crystallisation solution plus 5% MPD. Crystals from crystallisation solution 4 were cryo-protected with 80% crystallisation solution plus 20% glycerol.

Structure Solution

Diffraction data from four crystals were collected up to 3.4 Å resolution at the Diamond Light Source (beamline I04-1, $\lambda=0.9281$ Å) and the European Synchrotron Radiation Facility (ESRF; beamline ID29, $\lambda=0.97625$ Å) at 100 K. Data were integrated using XDS (via XIA2) [105, 279]. Integrated intensities were merged and scaled using Aimless and truncated using ctruncate [278]. The high resolution cut-off was chosen by selecting shells which still fulfil $CC_{1/2}>25\%$ [109] and $I/\sigma I>0.5$. The structure was solved by molecular replacement in PHASER [166], by sequentially placing the previously solved FLRT2^{LRR}-UncD^{Ig1} complex [226] and the high resolution structures of the unliganded Lphn3^{Lec-Olf} and Ig2 domain of Unc5A^{Ecto}. After MR, the Unc5A Ig2 domain was replaced with a homology model of Unc5D Ig2 domain, made using SWISS-MODEL [22, 226]. A homology model of Unc5D Tsp1 was placed into the density clearly visible after initial MR and refinement. The model was manually adjusted in Coot [63]. The model was all-atom refined in autoBUSTER [30] with the command line options -r 0.008 to restrain the geometry, -w 5 and AdjustXrayWeightAutomatically=no to fix the X-ray weight and -autoncs to use non-crystallographic symmetry restraints. Parameterisation of thermal motion was based on results from the TLSMD web server [195] and each chain of Unc5D

and Lphn3A was divided into two TLS bodies. The final quality of the model was assessed using MolProbity [48].

2.4.8 Crystallisation and Structure Solution of GgTen2^{ΔEGF}

Crystallisation

GgTen2^{ΔEGF} was produced in HEK293T cells treated with kifunensine and purified as described in Section 2.2. To produce the initial poorly diffracting crystals, purified GgTen2^{ΔEGF} was mixed with endoglycosidase F1 in a 1:1000 (endoglycosidase F1: GgTen2^{ΔEGF}) and concentrated to 4.7 mg/ml in a buffer containing 20 mM Tris (pH 7.5), 200 mM NaCl, 200 mM non-detergent sulfobetaine-256. Initial crystals were obtained after 14 days at 18 °C in a 200 nl drop containing a 1:1 mixture of protein solution and reservoir solution. The reservoir solution was 0.1 M Tris (base)/Bicine (pH 8.5), 20% v/v glycerol, 10% w/v PEG4K, 30 mM di-ethyleneglycol, 30 mM tri-ethyleneglycol, 30 mM tetraethyleneglycol, and 30 mM pentaethyleneglycol.

The optimised GgTen2^{ΔEGF} crystals used for structure solution were grown in the absence of endoglycosidase F1. The original reservoir condition was diluted ~5-15% with water. Microseeding was successful in increasing the final size of the crystals. To prepare microseeds, a crystallisation drop containing small crystals was resuspended in its reservoir condition (typically 50 μ l) and vortexed with a “Seed-Bead”. Serial dilutions of this microseed-containing solution were prepared by further diluting it with the reservoir condition. The best diffracting GgTen2^{ΔEGF} crystal (~2.5 Å) was obtained by mixing 250 nl protein solution (5 mg/ml) with 83 nl seed stock (diluted 10⁶-fold) and 166 nl reservoir solution. For crystallisation conditions where the concentration of glycerol was diluted significantly below 20%, additional glycerol was added as cryoprotectant.

To produce heavy atom-derivatised GgTen2^{ΔEGF} crystals, a small amount of solid heavy metal compound was collected on a pipet tip and then thoroughly resus-

pended in the reservoir solution. 2 μ l of this heavy metal containing solution was then added directly to the droplet containing pre-formed crystals. After 5, 7 and 21 hours, crystals were either harvested directly or briefly back-soaked in a drop containing reservoir solution, but no HAs. The back-soaked crystals were then cryo-cooled in liquid nitrogen. The Pt-derivative crystals used to collect the anomalous diffraction data for experimental phasing were produced using K_2PtCl_4 as the heavy metal compound, with an incubation time of 21 hours. These crystals were back-soaked in crystallisation solution prior to flash-cooling.

Structure Solution

Native data from GgTen2 $^{\Delta\text{EGF}}$ crystals were collected up to 2.4 Å resolution at Diamond Light Source at 100 K (beamline I03, $\lambda = 0.97626$ Å). GgTen2 $^{\Delta\text{EGF}}$ crystals were radiation-sensitive and so native data were collected in 90° wedges from three isomorphous crystals. Wedges of native data having suffered the least radiation damage were selected for merging to produce the final native dataset (Table 6.3). Data were indexed and integrated using XDS, and scaled and merged using XSCALE (via Xia2) [105, 279]. For anomalous data collection, datasets from two isomorphous crystals were collected at the European Synchrotron Radiation Facility (ESRF; beamline ID29) at 100 K. To maximise anomalous signal, data were collected at the L-III Pt edge ($\lambda=1.072$ Å). Data were processed as above, but Bijvoet pairs were unmerged (Table 6.2). Crank2 [196, 238] was used to calculate phases and build an initial model. The model was manually improved using Coot [63] and the NHL domain was placed by MR in Phaser using a modified polyalanine model of the β -propeller domain of a distant structural homologue (PDB accession 3HRP). The resulting model was subsequently used for iterative phase improvement using the Phaser SAD pipeline [209] and manual re-building in Coot [63]. Partial sequence assignment was facilitated by homology models of the Ten2 Rhs-associated core domain and YD-shell region based on the structure of YenC2 (PDB 4IGL) [34]

and the knowledge that Pt atoms typically bind to either methionine or histidine residues. The resulting structure was used to phase the higher-resolution native data by MR with Phaser [166] to place four copies of the model in the $P2_1$ asymmetric unit. The resulting map was density-modified using Parrot [44] and then supplied to Buccaneer [43] for autobuilding, allowing reliable assignment of much of the sequence. The remainder of the sequence was assigned manually based on the density. Glycans were built using the automatic glycan building implementation in Coot [63]. The model was all-atom refined in autoBUSTER using NCS restraints and -r 0.009 to restrain the geometry [30]. Thermal motion was parameterised by treating each macromolecular chain as a single TLS group. The quality of the final model was assessed using MolProbity [48].

2.5 Cell Biology

2.5.1 Cell-Binding Assay

HeLa cells seeded onto coverslips in six-well plates were transfected with 2 μ g FLRT2-ires-GFP plasmid using FuGENE 6 transfection reagent (Promega) as described in Section 2.2.3. Thirty-six hours after transfection, cells were incubated at 4 °C for 20 min in Hank's balanced salt solution containing 0.5 μ g Lphn3^{Lec-Olf} or Lphn3^{Lec-Olf} LF mutant protein pre-clustered with 1.6 μ g His-probe rabbit polyclonal immunoglobulinG (IgG) (Santa Cruz Biotechnology). Cells were washed twice with PBS and fixed with 4% paraformaldehyde at 4 °C for 20 min. Cells were blocked for 20 min with 1% BSA in PBS and bound Lphn3^{Lec-Olf} proteins were visualised using Alexa561-conjugated anti-rabbit IgG antibody (Molecular Probes, 1:400). Before mounting with Immu-MountTM (Invitrogen), nuclei were counter-stained with DAPI (1:10000).

2.5.2 Stripe Assay

A stripe assay is a cell-migration assay in which cells are confronted with a “carpet” of alternating stripes of different proteins. The stripe assay is used to study the responses of migrating cells to guidance molecules by quantifying the distribution of the cells across the stripes. The cells used can either be from cell culture (transfected or non-transfected), or can be neurons from a primary culture.

In all stripe assay experiments described in this thesis, the red stripe contained the protein-of-interest (protein solution 1), and the black stripe contained a control protein, which is neutral with respect to its cell guidance properties (protein solution 2). Protein solution 1 contained the protein-of-interest and a Cy3-conjugated anti-Fc antibody, meaning that the stripes containing the protein-of-interest fluoresce red. Protein solution 2 contains Fc as a control protein and anti-Fc as a control antibody.

Stripe Carpet Preparation

Silicone matrices placed on 60 mm tissue culture dishes were used to produce the stripe pattern [120]. 50 $\mu\text{g}/\text{ml}$ of the recombinant protein-of-interest was mixed with 120 $\mu\text{g}/\text{ml}$ Cy3-conjugated anti-Fc (Life Technologies) in PBS (protein solution 1 in Figure 2.3B). This protein mixture was injected into the stripe matrix via the inlet slot using a Hamilton syringe (Figure 2.3A). The stripes were incubated for 30 min at 37°C to allow protein solution 1 to adhere to the tissue culture dish (Figure 2.3B). Stripes were washed three times with PBS and the matrices removed. Dishes were then coated with 50 μl 50 $\mu\text{g}/\text{ml}$ Fc protein mixed with 120 $\mu\text{g}/\text{ml}$ anti-hFc (Jackson ImmunoResearch) (protein solution 2, control) for 30 min at 37°C (Figure 2.3C.) Stripes were then washed three times with PBS. The end result is alternating stripes of protein solutions 1 and 2 (Figure 2.3D).

For experiments with neurons, stripe carpets were overlaid with 100 μl laminin (20

$\mu\text{g}/\text{ml}$ in PBS) and incubated for at least one hour and then washed three times with PBS before use.

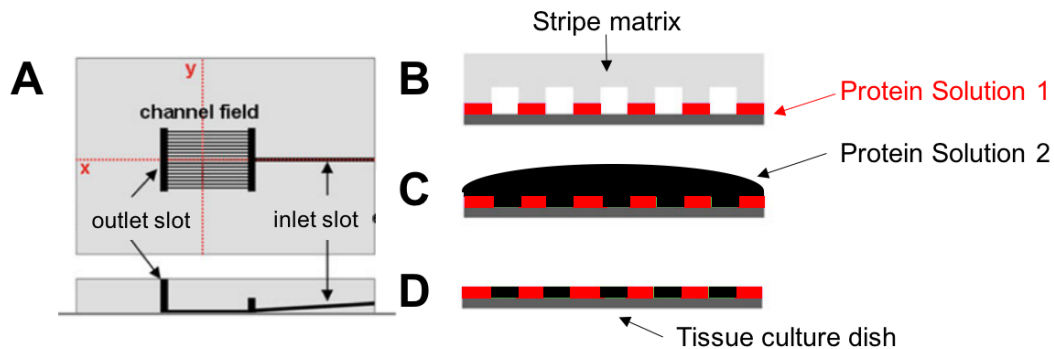


Figure 2.3: **Stripe Assay Preparation**

Stripe Assays with FLRT-expressing cells on $\text{Lphn3}^{\text{Lec-Olf}}$ (Wild Type vs. LF) Stripes

Stripe carpets were produced as described above. Protein solution 1 contained 50 $\mu\text{g}/\text{ml}$ of $\text{Lphn3}^{\text{Lec-Olf}}$ (WT or mutant, LF) and 120 $\mu\text{g}/\text{ml}$ Cy3-conjugated anti-Fc (Life Technologies). Protein solution 2 was 50 μl 50 $\mu\text{g}/\text{ml}$ Fc protein mixed with 120 $\mu\text{g}/\text{ml}$ anti-hFc (Jackson ImmunoResearch).

For the experiments with HeLa cell described in Figure 3.5, HeLa cells were grown in 6-well plates and transfected with 2 μg FLRT2-ires-GFP (WT or mutant, FF, see Table A.1). Twenty-four hours after transfection, cells were washed with PBS and dissociated using 1 ml 5 mM EDTA (pH 8.0) for 5 min at 37°C. Cells were pelleted (5 min, 200 xg) and resuspended in the desired volume of DMEM (2% FBS) to achieve a cell density of $2.0 \times 10^5/\text{ml}$. 150 μl of cell suspension was added to each stripe, giving a total of 30,000 cells per experiment. HeLa cells were cultured for 3 hours at 37°C and then fixed with 2% sucrose/4% PFA in PBS for 10-20 min at room temperature. Cells were washed with PBS and the nuclei counterstained with DAPI before mounting with Immu-Mount™ (Thermo Fisher). The number of GFP+ pixels on red stripes quantified with ImageJ.

For the experiments with cortical neurons described in Figure 3.6, dissociated cortical neuronal cultures (E15.5) were prepared by Dr. Daniel del Toro (Klein lab, Max Planck Institute of Neurobiology) and applied to stripes at a concentration of 50,000 cells per experiment. The neurons were cultured on the stripes for 16 hours and then fixed with 2% sucrose/4% PFA in PBS for 10-20 min at room temperature. Cortical neurons were visualised by staining for β III-tubulin as follows. Neurons were washed with PBS, permeabilised (1% BSA, 0.1% Triton X-100/PBS) for 10 min and then stained using rabbit monoclonal anti- β -III tubulin antibody (1:1000; Sigma) as a primary antibody and Cy2 or Cy3 anti-rabbit IgG (1:400; Jackson) as a secondary antibody. After mounting with Immu-MountTM (Thermo Fisher) the number of GFP+ pixels on red stripes quantified with ImageJ.

Stripe Assay with FLRT2-Unc5D-expressing cells on Lphn3^{Lec-Olf} stripes

For the experiments described in Figure 3.7, stripe carpets were prepared as in Section 2.5.2. Protein solution 1 contained 50 μ g/ml of wild-type Lphn3^{Lec-Olf} and 120 μ g/ml Cy3-conjugated anti-Fc (Life Technologies). Protein solution 2 was 50 μ l 50 μ g/ml Fc protein mixed with 120 μ g/ml anti-hFc (Jackson ImmunoResearch).

HeLa cells were grown in 6-well plates and transfected with 2 μ g Unc5D-ires-GFP and 1 μ g FLAG-FLRT2-Avi (WT or mutant, UF). For the FLRT2 alone sample the FLRT2-ires-GFP construct was used. See Table A.1 for details of these constructs. Twenty-four hours after transfection, cells were washed with PBS and dissociated using 1 ml 5 mM EDTA (pH 8.0) for 5 min at 37°C. Cells were pelleted (5 min, 200 xg) and resuspended in the desired volume of DMEM (2% FBS) to achieve a cell density of 2.0×10^5 /ml. 150 μ l of cell suspension was added to each stripe, giving a total of 30,000 cells per experiment. HeLa cells were cultured for 7 hours at 37°C and then fixed with 2% sucrose/4% PFA in PBS for 10-20 min at room temperature. To verify expression of FLAG-tagged FLRT constructs, cells were washed with PBS, permeabilised (1% BSA, 0.1% Triton X-100/PBS) for 10 min at room temperature

and then incubated with rabbit anti-FLAG (1:400, Sigma) in PBS, 1% BSA for 1 hour at room temperature. Cells were washed with PBS and then incubated with Alexa-Fluor 647 Donkey anti-rabbit IgG (1:200) in PBS, 1% BSA for 1 hour at room temperature to visualise the cells. Cells were washed with PBS again and nuclei counterstained with DAPI before mounting with Immu-MountTM (Thermo Fisher). The number of GFP+ pixels on red stripes quantified with ImageJ.

Chapter 3

Interaction of Lphn with FLRT and Unc5

3.1 Summary

Upon first starting this project, a significant amount of information was available regarding the molecular basis for the interaction between FLRT and Unc5 [226]. However, the basis for the interaction between FLRT and Lphn was poorly understood. Using a newly-solved crystal structure of a fragment of Lphn3 comprising its Lec and Olf domains, and guided by surface sequence conservation, I designed mutations in the protein with the aim of identifying the FLRT binding site. I found that FLRT binds a highly evolutionarily conserved patch on the Lphn Olf domain, and that a mutant, Lphn3^{LF}, abolishes this interaction completely in SPR and cell-binding experiments. Using stripe assays with transfected HeLa cells I showed that, consistent with the literature, the Lphn-FLRT interaction is strongly adhesive, an effect that is ablated for the Lphn3^{LF} mutant. In contrast, stripe assays using cortical neuronal cultures showed that the Lphn FLRT-binding site is repulsive. Further stripe assays, along with biophysical studies demonstrated that Lphn3, FLRT2 and Unc5D can form a multimeric ternary complex, with a function distinct from binary interactions between its subcomponents.

3.2 Surface Conservation of Lphn3^{Lec-Olf}

To better understand the molecular basis of the Lphn-FLRT interaction I aimed to map the FLRT-binding site on Lphn. This interaction is mediated by the Lphn Olf domain, is observed for multiple homologues of both FLRT and Lphn and is highly important in synaptic development and brain wiring [191, 192]. For these reasons, I hypothesised that the FLRT-Lphn binding surfaces would be conserved across species. Surface sequence conservation, calculated by the server ConSurf, can be used to reveal patches of highly conserved residues that are often important to biological function [80]. In the absence of a complex crystal structure this information can be useful in predicting conserved ligand-binding sites [226], these predictions can then be validated experimentally.

I calculated surface sequence conservation scores for newly-solved structure of Lphn3^{Lec-Olf} (solved by Dr Elena Seiradake, prior to the start of this project) using an alignment of sequences from murine Lphn1+3, chicken Lphn2, Lphn3 from *Danio rerio* and Lphn3 from *Xenopus laevis* and used these scores to predict the likely FLRT interaction site on the Lphn3^{Lec-Olf}. This analysis revealed a highly-conserved patch of residues spanning blades II and III of the Lphn3 Olf domain (Figure 3.1).

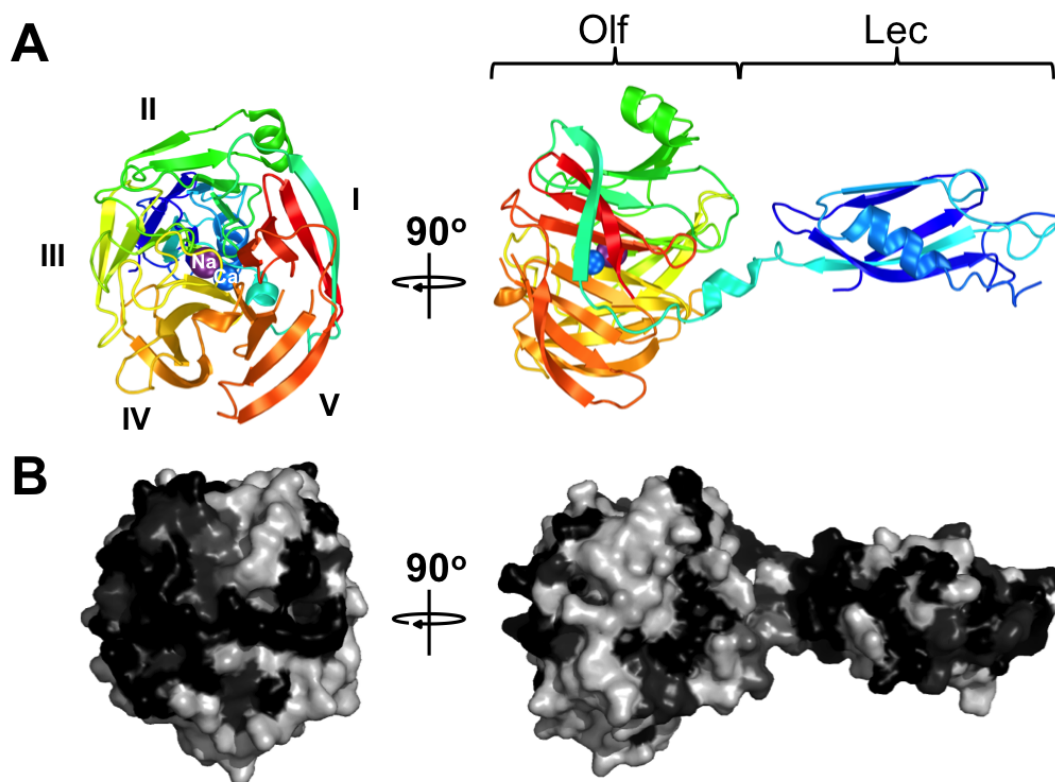


Figure 3.1: **The crystal structure of Lphn3^{Lec-Olf}**. The crystal structure of Lphn3^{Lec-Olf} shown as cartoon and coloured according to the rainbow. The bound sodium and calcium ions are shown as purple and blue spheres respectively (A). The same structure is shown as surface representation and coloured according to sequence conservation (B). Sequence conservation scores were calculated using the ConSurf server [80] from a sequence alignment of murine Lphn1/3, chicken Lphn2 and Lphn3 from *Danio rerio* and *Xenopus tropicalis*. Black indicates the highest conservation scores and white the lowest.

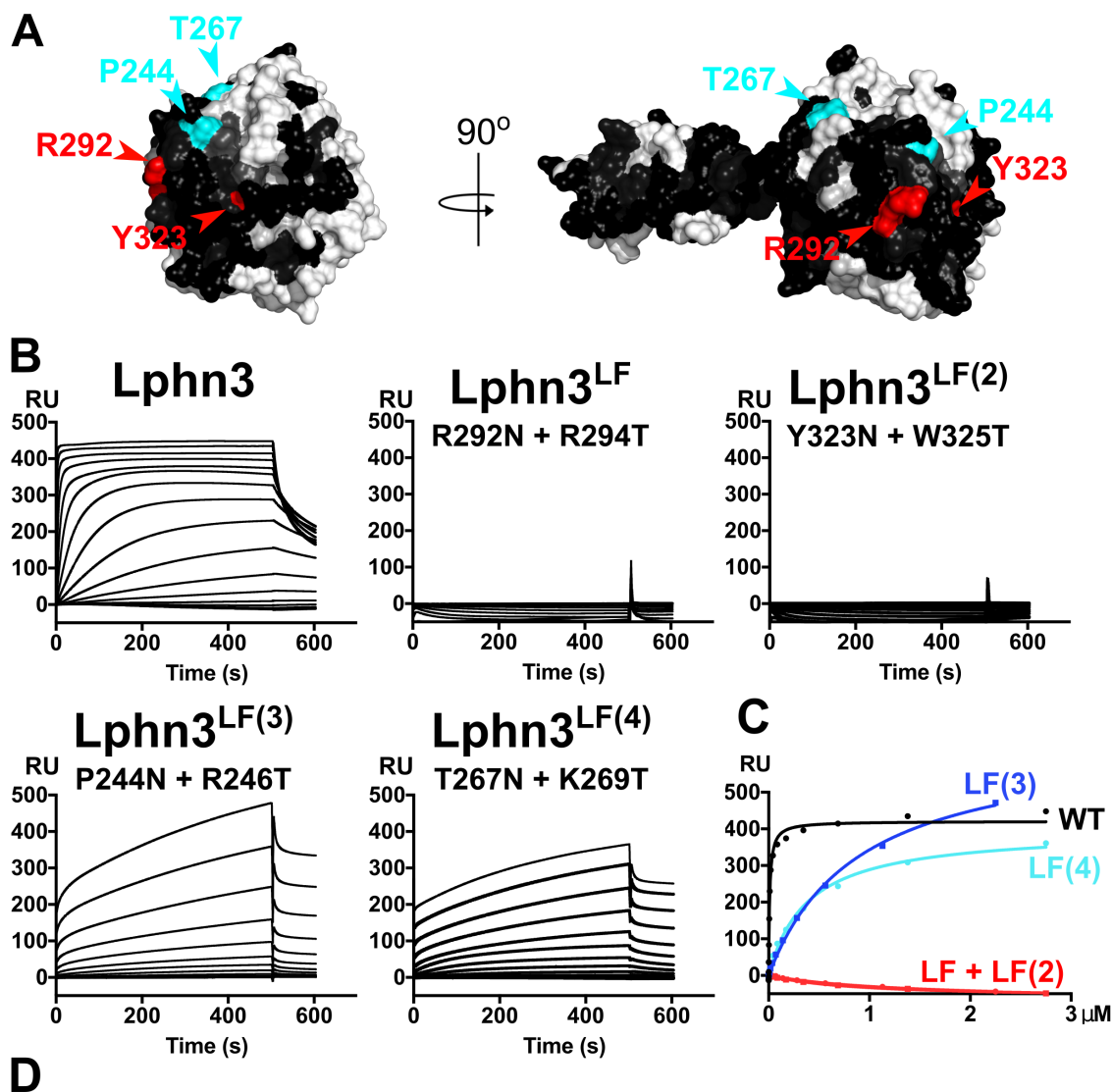
3.3 Mutational Disruption of the FLRT-Lphn Interaction

3.3.1 Mapping the FLRT-binding Site on Lphn

To probe this conserved surface region on the Lphn3 Olf, I generated a series of mutant proteins in which consensus N-linked glycosylation motifs were introduced at various positions in the conserved patch (Figure 3.2A). This is an established way of disrupting protein-protein interactions [226, 288]. The mutations made were R292N R294T (Lphn3^{LF}), Y323N W325T (Lphn3^{LF(2)}), T267N K269T (Lphn3^{LF(3)})

and P244N R246T (Lphn3^{LF(4)}) (Figure 3.2A).

Surface plasmon resonance (SPR) experiments show high affinity binding between the FLRT2^{Ecto} and wild-type Lphn3^{Lec-Olf} (Figure 3.2). However, two of the mutants, LF and LF(2) (with the mutant glycosylations located in the centre of the conserved patch) showed no measurable binding to FLRT2^{Ecto}. LF(3) and LF(4), which have their mutations at the edge of the conserved Olf patch, still interacted with FLRT2^{Ecto}, but with a 35- to 80-fold reduced affinity (Figure 3.2B, C, D). For subsequent cell-based assays, the non-FLRT binding Lphn3 mutant, R292N R294T, was used and is herein referred to as the Lphn3 LF mutant.



Ligand	Analyte	K_D (nM)	B_{max} (RU)	$B_{surface}$ (RU)
MmLphn3 ^{Lec-Olf} WT	FLRT2 ^{Ecto}	11	421	420
MmLphn3 ^{Lec-Olf} LF	FLRT2 ^{Ecto}	N/A	N/A	610
MmLphn3 ^{Lec-Olf} LF(2)	FLRT2 ^{Ecto}	N/A	N/A	460
MmLphn3 ^{Lec-Olf} LF(3)	FLRT2 ^{Ecto}	890	649	54
MmLphn3 ^{Lec-Olf} LF(4)	FLRT2 ^{Ecto}	380	398	340

Figure 3.2: Mutations in the conserved patch on the Lphn3 Olf domain impair binding to FLRT2^{Ecto}. The crystal structure of Lphn3^{Lec-Olf} is shown in surface representation and coloured according to sequence conservation scores shown in Figure 3.1B. Mutated residues are coloured according to their effects on FLRT2 binding (A). SPR sensorgrams for the interaction between FLRT2^{Ecto} and Lphn3^{Lec-Olf} and its mutants. Biotinylated wild-type and mutant Lphn3^{Lec-Olf} proteins were attached to streptavidin-coupled CAPture chips and a 1:1 dilution series of FLRT2^{Ecto} used as analyte. Results are from two consistent experiments with the top concentration of FLRT2^{Ecto} being 2-3 μ M (B). K_D values were calculated using a 1:1 binding model. No meaningful K_D value could be calculated for the LF(3) and LF(4) mutants (C,D).

I also verified my SPR results in a cell-binding assay. Wild-type and LF-mutant Lphn3^{Lec-Olf} proteins (which are His-tagged) were applied to cells transfected with a FLRT2-ires-GFP vector [226]. The His-tag on the Lphn3^{Lec-Olf} constructs allowed detection of protein bound on the cell surface. Consistent with the SPR results, wild-type Lphn3^{Lec-Olf}, but not the LF mutant, bound to the surface of the FLRT2-transfected cells (Figure 3.3).

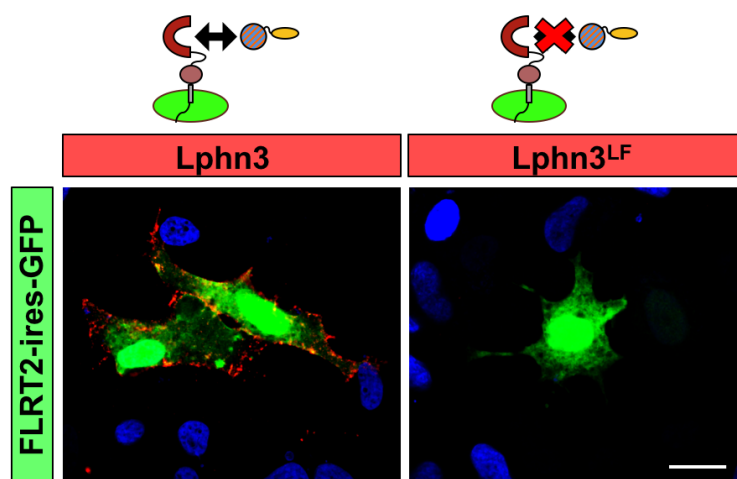
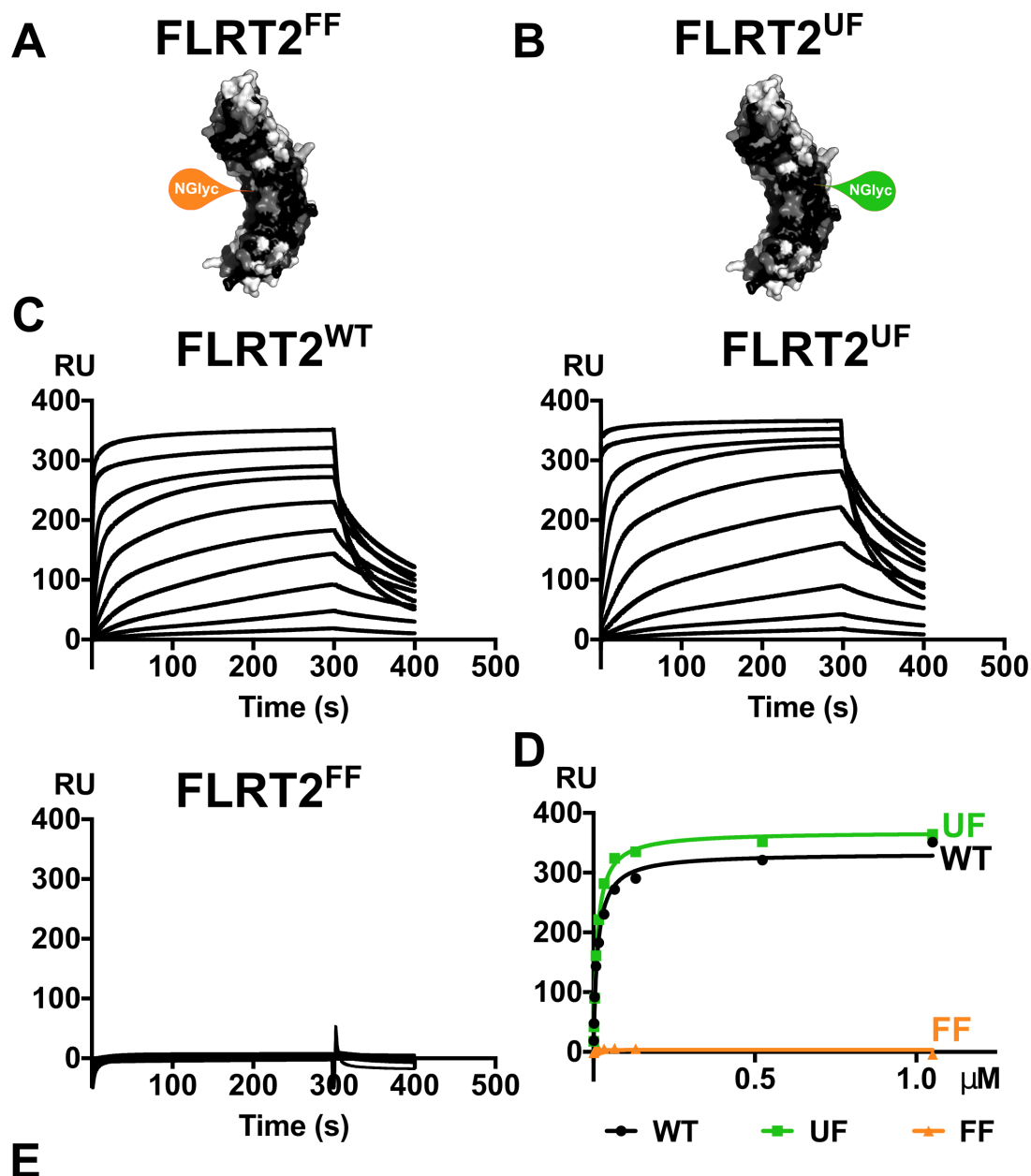


Figure 3.3: **Wild type and mutant Lphn3-FLRT2 cell-binding assay.** Binding of wild-type or mutant (LF) Lphn3^{Lec-Olf} proteins to HeLa cells transfected with FLRT2-ires-GFP (green). Bound Lphn3^{Lec-Olf} proteins were visualised using an antibody against the hexahistidine tag (red). Cells were counter-stained with DAPI before mounting (blue). Scale bar, 20 μ m. This experiment was supervised by Dr Daniel del Toro, Max Planck Institute of Neurobiology, Munich.

3.3.2 Mapping the Lphn-binding Site on FLRT

Having mapped the FLRT-binding site on Lphn3, I decided to map the equivalent site on FLRT. To do this, I used previously described mutants, FLRT2^{FF} and FLRT2^{UF}, which knock out FLRT-FLRT homophilic interactions and the FLRT-Unc5 interaction respectively (Figure 3.4A, B) [226]. SPR experiments with these mutants revealed that, whilst FLRT2^{UF} binds the Lphn3^{Olf} with wild-type affinity, FLRT2^{FF} showed no measurable binding to Lphn3^{Olf} (Figure 3.4C, D). These data suggest that Unc5 and Lphn bind distinct surfaces on the FLRT2 LRR domain and that the FLRT-Lphn interaction would compete with low-affinity FLRT-FLRT homophilic interactions.



Ligand	Analyte	K_D (nM)	B_{max} (RU)	$B_{surface}$ (RU)
FLRT2 ^{Ecto-WT}	Lphn3 ^{OLF}	13	332	2280
FLRT2 ^{Ecto-UF}	Lphn3 ^{OLF}	11	368	2580
FLRT2 ^{Ecto-FF}	Lphn3 ^{OLF}	N/A	N/A	3000

Figure 3.4: **Lphn and Unc5 bind distinct surfaces on the FLRT LRR domain.** The structure of the FLRT2 LRR domain coloured according to its sequence conservation [226]. The positions of mutant N-linked glycosylation sites in these mutants are indicated for FLRT2^{FF} (A) and FLRT2^{UF} (B). Full ectodomain proteins of wild-type FLRT2, the non-Unc5-binding mutant FLRT2^{UF} and the non-multimerising FLRT2^{FF} mutant were immobilised on a CM5 Biacore chip by direct amine-coupling. A 1:1 dilution series of FLRT2^{OLF} was used as analyte, with a top concentration of 1.05 μ M (C). K_D values were calculated from the equilibrium SPR data assuming a 1:1 binding model (D,E).

3.4 Cell Guidance Consequences of the Lphn-FLRT Interaction

The FLRT-Lphn interaction has been postulated to mediate transsynaptic adhesion [191]. To assess these adhesive properties I performed stripe assays, which are used to assess the “guidance properties” of biological molecules. Lphn3^{Lec-Olf} proteins were immobilised in an alternating stripe pattern with control F_c protein, which is neutral with respect to its cell guidance properties. HeLa cells transfected with FLRT2-ires-GFP were then seeded onto the stripe “carpet” and allowed to choose between growth on either Lphn3^{Lec-Olf} or F_c control protein (Figure 3.5A). Consistent with the literature, the FLRT2-transfected cells adhered strongly to the Lphn3^{Lec-Olf} stripes. However, exchanging wild-type Lphn3^{Lec-Olf} with the non-FLRT-binding mutant, LF, significantly reduced this adhesive effect. Similarly, cells transfected with the non-Lphn-binding FLRT2^{FF}-ires-GFP did not show this adhesive effect (Figure 3.5). These data from transfected HeLa cells indicate that the interaction between the FLRTs and Lphn3 Olf is strongly adhesive, consistent with the proposed role of this interaction in regulating synapse formation [191]. Note that the small amount of adhesion seen for the FLRT2^{FF}-expressing cells is likely due to endogenous expression of FLRT proteins in these cells (Figure 3.5C).

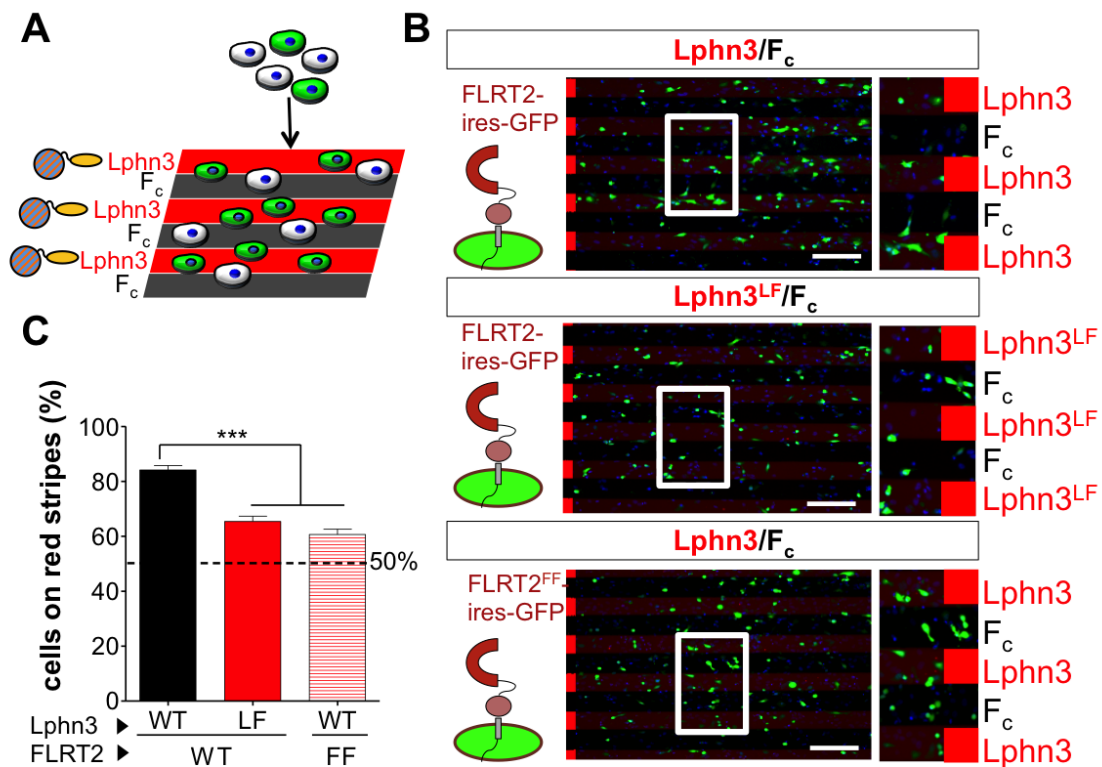


Figure 3.5: The FLRT-binding site on Lphn3 regulates cell adhesion. Schematic demonstrating the experimental set-up for stripe assays with transfected HeLa cells (A). Composite images showing the distribution of HeLa cells transfected with either FLRT2-ires-GFP or FLRT2^{FF}-ires-GFP on alternating stripes of wild-type or mutant (LF) Lphn3^{Lec-Olf} (red) and F_c protein (black). Mutational knock-out of the interaction between the Lphn3 Olf and FLRT2 LRR domains abolishes adhesion to the Lphn3^{Lec-Olf} stripes. Higher magnification images of the boxed regions are shown as insets to clearly show the distribution of cells on the stripes. The positions of the stripes are indicated by red bars. Cells were counter-stained with DAPI before mounting (blue). Scale bar, 200 μ m (B). Quantification was performed by measuring the number of green pixels present on the red (wild-type or LF Lphn3^{Lec-Olf}) stripes. Statistical analysis was performed using a one-way ANOVA test with Tukey's post-hoc analysis ($p < 0.001$, ***). Error bars represent the standard error of the mean (C). This experiment was supervised by Dr Daniel del Toro, Max Planck Institute of Neurobiology, Munich.

Although a useful tool, transfected HeLa cells do not accurately reflect the complexity of the neuronal cell surface and their downstream signalling. To assess the guidance role of the FLRT-Lphn interaction in a more native environment I performed stripe assays using primary neuronal cultures. Cortical neurons were chosen as they express high levels of both FLRT2 and FLRT3 [226]. Dissociated primary neuronal cultures were produced from E15.5 mouse embryos by Dr Daniel del Toro, Max Planck Institute of Neurobiology, Munich. Interestingly, these cortex-derived cultures behaved in an opposite fashion to the HeLa cells, with the cortical

neurons being repelled from the wild-type $Lphn3^{Lec-Olf}$ stripes. The repulsive effect was eliminated when cortical neurons were grown on non-FLRT-binding $Lphn3^{Lec-Olf}$ LF stripes (Figure 3.6).

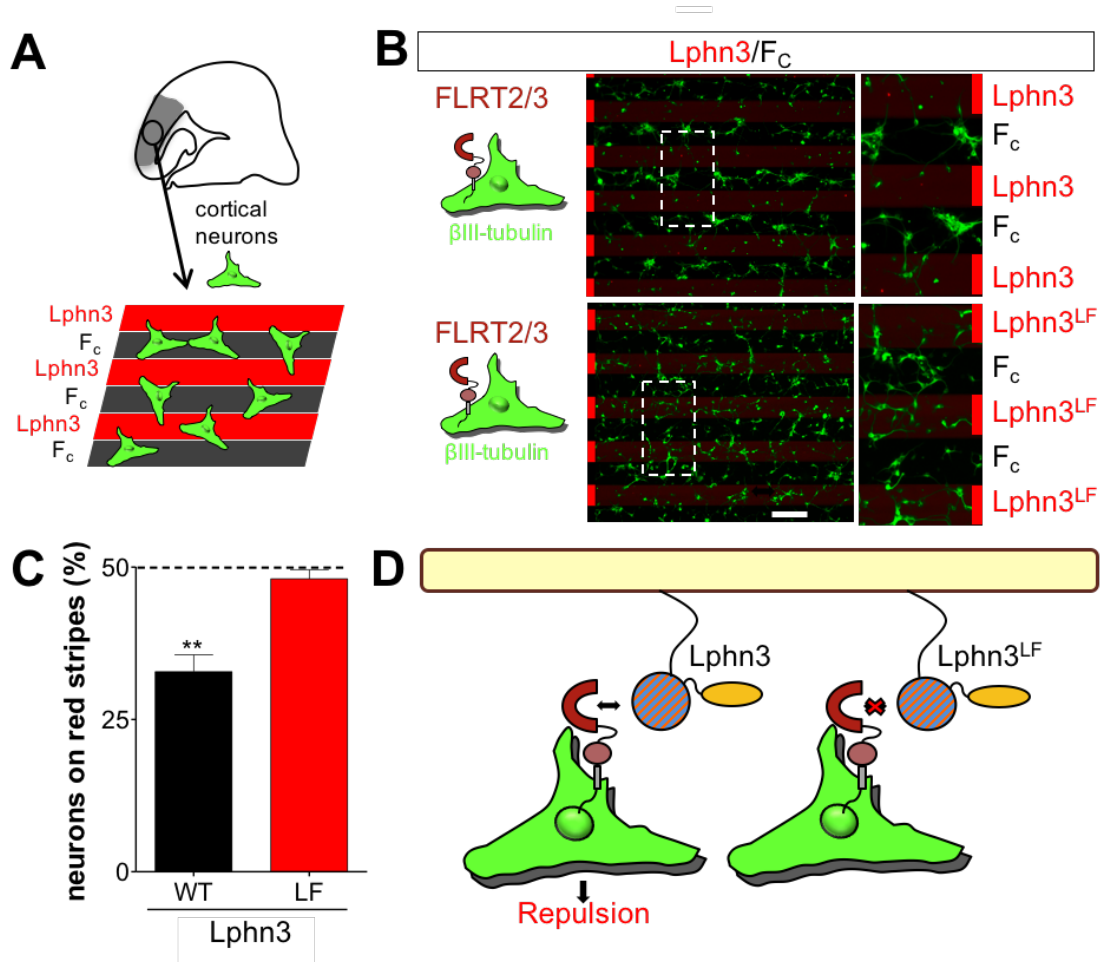


Figure 3.6: The FLRT-binding site on $Lphn3$ regulates repulsion of cortical neurons. Schematic of the experimental set-up for cortical neurons grown on alternating stripes of $Lphn3^{Lec-Olf}$ and F_c (A). Composite images showing the distribution of dissociated cortical neuronal cultures (E15.5, expressing high levels of endogenous FLRTs [226]) on alternating stripes of $Lphn3^{Lec-Olf}$ and F_c protein. Cortical neurons were visualised by staining for βIII -tubulin, a neuronal-specific tubulin (green). The neurons are repelled from the wild-type $Lphn3^{Lec-Olf}$ protein, but not the LF mutant, stripes. Insets are higher magnification images of the regions enclosed in the dashed white boxes. Red bars indicate the positions of the red ($Lphn3^{Lec-Olf}$ -containing) stripes. Scale bar, 200 μm . Dissociated neuronal cultures were prepared by Dr Daniel del Toro, Max Planck Institute of Neurobiology, Munich (B). Data of the type shown in (B) were quantified by calculating the percentage of green pixels on the red stripes using ImageJ. Statistical significance was calculated using a two-tailed Student t-test ($p < 0.01$, **). Error bars represent the standard error of the mean (C). The findings of this experiment are summarised in a schematic (D).

3.4.1 Investigating the Role of Unc5 in Lphn-FLRT-mediated Cell Adhesion

This surprising result from the primary cortical neurons suggests that the FLRT-binding site on Lphn3 may play a context-dependent bifunctional role, being adhesive under some conditions, and repulsive under others. One mechanism by which this could be mediated is by recruitment of repulsive guidance receptors *in cis* by FLRT. Examples of such receptors could be Robo1 [145] or Unc5 [226], both of which are known FLRT ligands. Given that my SPR results had indicated that Lphn3 and Unc5 bind the FLRT2 LRR domain at distinct sites, it seemed feasible that FLRT could engage both receptors at the same time. To assess whether the presence of Unc5 *in cis* was capable of influencing the guidance effect induced by the interaction between FLRT and Lphn, I performed further Lphn3^{Lec-Olf} stripe assays using HeLa cells co-transfected with both Unc5 (Unc5D-ires-GFP) and FLRT2 (FLAG-FLRT2-Avi) as described in Section 2.5.2. I used the equivalent construct for the non-Unc5-binding FLRT mutant (FLRT2^{UF}) [226] as a control in these experiments.

Consistent with previous experiments, HeLa cells transfected with full-length FLRT2 alone strongly adhered (>80%) to the Lphn3^{Lec-Olf} stripes. Cells co-transfected with FLRT2 and Unc5D adhere significantly less (~70% of cells) to Lphn3^{Lec-Olf}, similar to control (GFP-transfected) cells or cells transfected with Unc5D alone (Figure 3.7C). In contrast, co-expression of Unc5D with the non-Unc5-binding mutant, FLRT2^{UF}, restores the response to (>80%) adhesion (Figure 3.7B, C). This suggests that *in cis* recruitment of Unc5D by FLRT2 can act as a switch to attenuate adhesion mediated by the FLRT-Lphn interaction (Figure 3.7D).

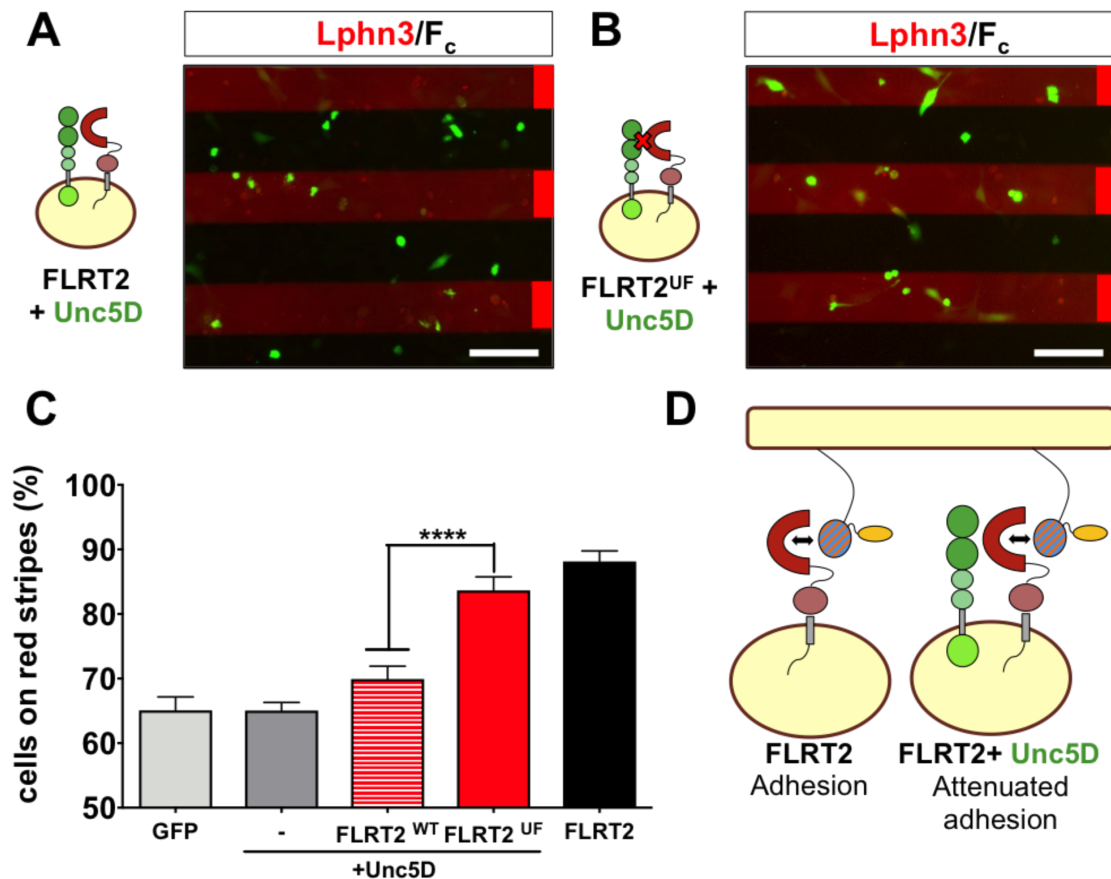


Figure 3.7: **Unc5 controls FLRT-Lphn-mediated cell adhesion.** Composite images showing the distribution of cells co-transfected with Unc5D and FLRT2 (WT or mutant, UF) on alternating stripes of Lphn3^{Lec-Olf} (red bars) and F_c control protein (A, B). Quantification of experiments shown in (A) and (B), along with relevant controls. Cells expressing FLRT2 alone (FLRT2-ires-GFP) adhere strongly to the Lphn3^{Lec-Olf} (red) stripes. Co-expression of Unc5D significantly reduces this adhesion. Co-transfection of Unc5D with a non-Unc5-binding mutant of FLRT2 (FLRT2^{UF}) restores this strong adhesion. Statistical significance was determined using an unpaired, two-tailed t-test ($p < 0.0001$, ****). The error bars represent the standard error of the mean. Imaging and quantification were assisted by Mrs Maria Carrasquero (C). The main finding of this experiment is summarised as a schematic (D).

3.5 Lphn-FLRT-Unc5 Ternary Complex Formation

Given that Unc5 and Lphn bind to distinct surfaces on the FLRT2 LRR, and that the presence of all three receptors mediates a cell guidance effect distinct from individual pairs of receptors, it seemed highly likely that FLRT, Lphn and Unc5 would be capable of forming a ternary complex containing all three proteins. I used SPR experiments to biophysically validate whether FLRT2 was capable of binding both Lphn3^{Lec-Olf} and Unc5D at the same time, and thus form a ternary complex. I immobilised a construct comprising the whole Unc5D ectodomain (Unc5D^{Ecto}) on a

streptavidin-coated CM5 chip and then used a dilution series of either FLRT2^{LRR} alone or a mixture of FLRT2^{LRR} plus excess Lphn3^{Lec-Olf} as analyte (Figure 3.8).

This experiment revealed an increased B_{max} value for FLRT2^{LRR}+Lphn3^{Lec-Olf} binding, as opposed to FLRT2^{LRR} binding alone. This indicates that, in the presence of Lphn3^{Lec-Olf}, total protein binding to Unc5D^{Ecto} is higher. Given that there is no detectable binding between Lphn3 alone and Unc5D (data not shown), these data suggest that FLRT2 is capable of mediating binding to both Lphn3 and Unc5D at the same time (Figure 3.8A,B). This observation was in line with cell-based assays performed by the Klein lab, Max Planck Institute of Neurobiology. Interestingly, my SPR experiment revealed that the presence of Lphn3^{Lec-Olf} in the analyte enhanced the affinity of the interaction between FLRT2 and Unc5D by about 5-fold (Figure 3.8C, D).

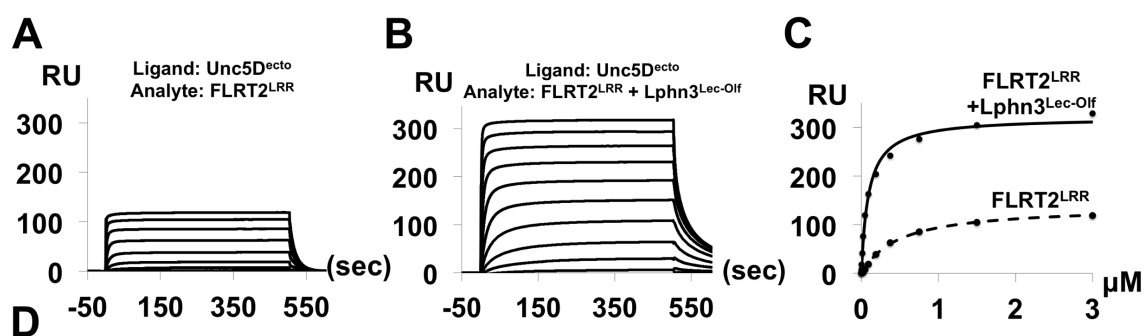


Figure 3.8: The presence of Lphn3 enhances the affinity of the Unc5D-FLRT2 interaction. Unc5D^{Ecto} was coupled to a streptavidin-coated CM5 Biacore chip. Analytes were either FLRT2^{LRR} (A) alone or a mixture of FLRT2^{LRR} and an excess of Lphn3^{Lec-Olf} (B). A 1:1 dilution series of these analytes was prepared for the run, where the top concentration in each case was 3 μ M. To ensure that results were comparable with one another the dilution series was prepared such that the top concentrations contained an equal amount of FLRT2^{LRR}. The presence of Lphn3^{Lec-Olf} in the analyte enhances the affinity of the interaction between FLRT2 and Unc5D by about 5-fold. A control experiment revealed no measurable affinity between Lphn3 and Unc5D. K_D values were calculated according to a simple 1:1 binding model using the equilibrium binding data (C, D).

To further biophysically characterise the ternary complex formed by Lphn3, FLRT2 and Unc5D, I performed SEC-MALS experiments. Samples were prepared by mixing Lphn3^{Lec-Olf} with FLRT2^{LRR} and different length constructs of the Unc5D ectodomain (Ig, comprising the first Ig domain only, IgIg, containing both Ig domains, and IgIgTsp, having both Ig domains, plus the first Tsp domain). Complexes formed using Unc5D constructs Unc5D^{Ig} and Unc5D^{IgIg} gave masses of ~ 100 kDa, which would approximately equate to 1:1:1 complexes. I was intrigued to find that complexes formed using the Unc5D^{IgIgTsp} construct gave masses which were significantly larger than those expected for simple 1:1:1 complexes. Note that that Tsp1 domain has a mass of only ~ 6 kDa and so is not, in itself, responsible for this large increase in the apparent molecular weight of the complex, and is largely acting to mediate some type of complex multimerisation. Unfortunately, SEC-MALS was not an appropriate method to calculate accurate masses for these complexes, as their formation was concentration-dependent and so dilution on the gel filtration column caused a fraction of the complex molecules separate into their subcomponents.

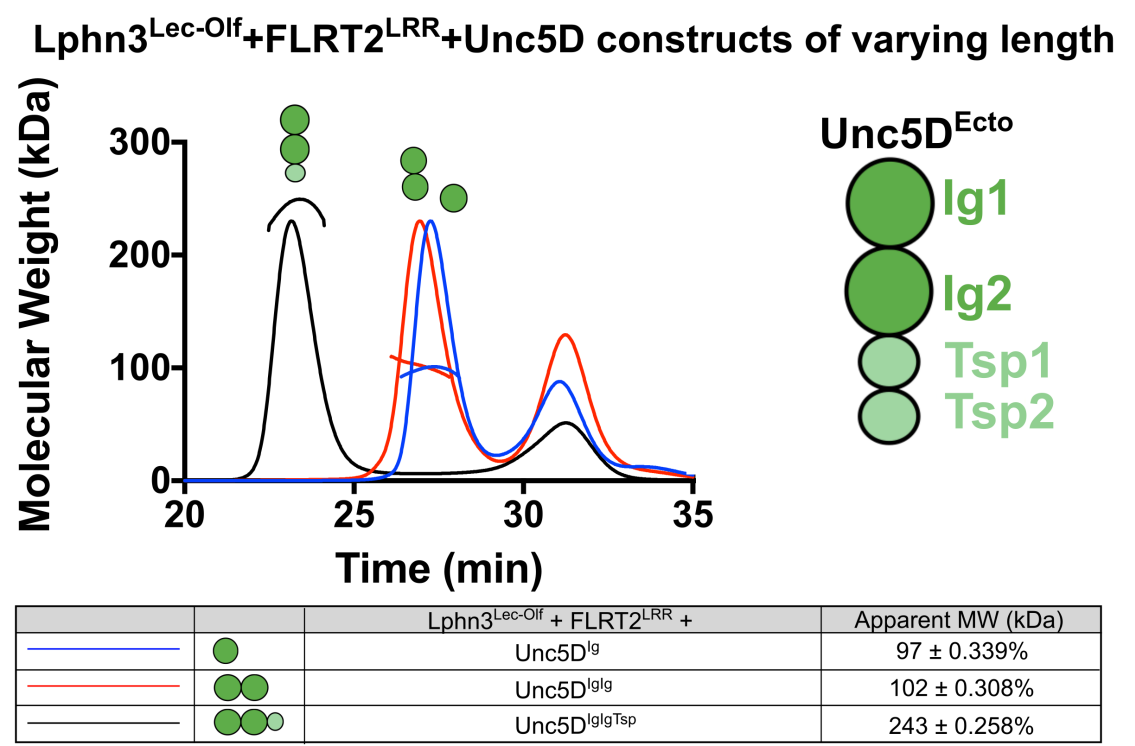


Figure 3.9: **Unc5D Tsp1 mediates multimerisation of the Lphn3-FLRT2-Unc5D ternary complex.** SEC-MALS indicates the formation of a large multimer ($\gg 200$ kDa) when Lphn3^{Lec-Olf} and FLRT2^{LRR} are mixed with Unc5D^{IgIgTsp}, but not when mixed with shorter Unc5D constructs without the Tsp1 domain. For these experiments, concentrations injected were 7 - 8 mg/ml (peak concentrations are at least 10-fold diluted by the SEC method). The Unc5D construct used in each experiment is indicated by the schematic above the relevant peak. The apparent molecular weight (MW) stated here reflects the average values within a given peak and so underestimates the true molecular weight of the fully-formed complexes.

3.6 Discussion

The data presented in this chapter reveal a bifunctional role for a highly conserved patch of residues on the surface of the Lphn3 Olf domain. This patch of residues mediates the interaction with the FLRT LRR domain and promotes cell adhesion in a simple stripe assay using transfected HeLa cells. However, the same patch mediates repulsion of cortical neurons.

In addition to mapping the FLRT-binding surface on Lphn3, I also mapped the Lphn-binding surface on FLRT. These experiments indicate that FLRT binds to Lphn via the concave surface of its LRR domain. This suggests that the FLRT-Lphn interaction would compete with low affinity FLRT dimerisation, which has been shown to occur via the concave surface on FLRT, mediating cell adhesion [226]. Given the relative affinities of these two interactions ($\sim 40 \mu\text{M}$ for FLRT3-FLRT3 [208] v.s. $\sim 10 \text{ nM}$ for Lphn3-FLRT2), Lphn-binding would easily out-compete FLRT homophilic interactions (assuming the absence of significant avidity effects). Although the cell guidance effect of the FLRT-FLRT and FLRT-Lphn interactions can be similar (both lead to adhesion, albeit different strengths), engagement of these two different receptors is likely to elicit different intracellular signals, likely making these two adhesive events distinct from one another.

Before the Lphn-binding site on FLRT was known, assays employing the FLRT^{FF} mutant indicated that adhesive FLRT-FLRT interactions regulate neuronal migration in the tangential axis of the cortex, and also guide the migration of vascular cells [226]. As well as preventing FLRT-FLRT homophilic interactions, the data presented in this chapter suggest that the FLRT^{FF} mutant is also incapable of interacting with Lphns. Lphns are broadly expressed in the mouse cortex [191, 192], meaning it is possible that differences between wild-type FLRT and the FLRT^{FF} mutant could be due to its lack of Lphn binding, rather than the abolishment of FLRT-FLRT interaction. Similarly, Lphn2 is highly expressed in vasculature-derived

cell lines [243], and thus the Lphn-FLRT interaction may also be responsible for the apparent adhesive effect of FLRT3 on these cells (in the absence of interactions with Unc5 receptors) [226]. Similar experiments using the Lphn LF mutant may go some way to answering these questions.

The mechanism by which the Lphn-FLRT interaction site is adhesive in a simple stripe assay using transfected HeLa cells, but repulsive to primary cortical neurons in the same assay is currently unknown. One potential mechanism is that Lphn3 binding to FLRTs may activate FLRT reverse repulsive signalling via its cytoplasmic tail, analogous to ephrin reverse signalling [119], although there are not yet any published examples of this phenomenon. Alternatively, when Lphn3 engages FLRT, FLRT may recruit other guidance receptors *in cis* and exploit their repulsive activity. In this chapter I show that the presence of Lphn3 enhances the affinity of FLRT2 for Unc5D about five-fold, suggesting that *in cis* recruitment of Unc5 by FLRT may be enhanced when it is bound by Lphn3. In my stripe assay where Unc5D is presented *in cis* with FLRT2 in HeLa cells, Unc5D recruitment by FLRT2 attenuates the adhesive effect of the FLRT-Lphn interaction, presumably using the repulsive activity of the Unc5D intracellular domain [96]. This leads to a model whereby the highly conserved patch on the Lphn3 Olf domain is bifunctional, and recruitment of Unc5 acts as a molecular switch to attenuate FLRT2-mediated Lphn3 adhesion (Figure 3.10). Indeed, recruitment of Unc5 *in cis* has been shown to modulate the response of other receptors to ligand-binding. For example Netrin bound to its receptor, DCC, elicits an attractive response [68], which is switched to repulsion by Unc5 binding [96]. This is analogous to the apparent role of Unc5 in the Lphn-FLRT-Unc5 system.

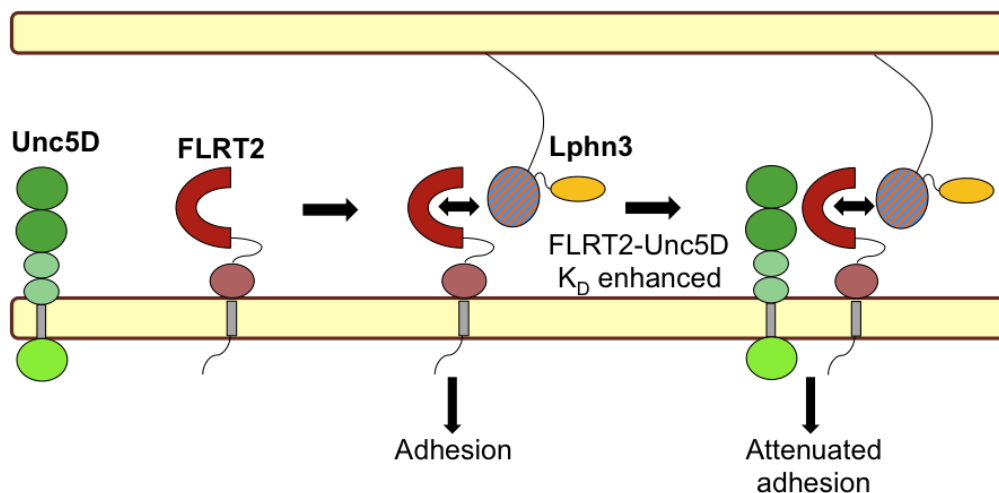


Figure 3.10: **Potential model for cell guidance mediated by Lphn, FLRT and Unc5D.** Simple model for cell guidance mediated by Lphn3, FLRT2 and Unc5D in HeLa cells. The arrangement of receptors is based on the stripe assay set-up in the chapter. *In trans* engagement of FLRT2 and Lphn3 enhances the affinity of the FLRT2-Unc5D interaction, leading to recruitment of Unc5D *in cis*. The formation of this ternary complex attenuates the adhesion caused by the binary FLRT-Lphn interaction.

A significant finding from this chapter is that the Lphn-FLRT-Unc5 ternary complex undergoes higher-order oligomerisation mediated by the Tsp1 domain of Unc5D. Complex multimerisation has been shown to be functionally relevant in a variety of systems, and understanding the molecular details of these complexes has yielded functional insights [227, 103, 228, 92]. From the SEC-MALS data presented in this chapter, complex stoichiometry was not clear and, although the individual ligand-binding sites on each protein have been mapped, the structural basis for this complex multimerisation remained unknown. In the next chapter I aim to address these questions using biophysical and structural methods.

Chapter 4

Lphn-FLRT-Unc5 Super-Complexes

4.1 Summary

Following on from the finding in the previous chapter that FLRT can bind both Lphn and Unc5 at the same time, and that this complex has an unusual stoichiometry, I set out to biophysically characterise the complex stoichiometry and gain structural insights on complex formation. I used native mass spectrometry and X-ray crystallography to determine that Lphn3, FLRT2 and Unc5D ectodomains bind to each other with a 4:2:2 stoichiometry. In the crystal structure each FLRT2 dimerises Lphn3 and then the whole complex is dimerised again by Unc5D. I also mutationally validated the crystal structure, selectively breaking these super-complexes down into smaller sub-complexes. These mutants will provide useful tools for future functional analysis of these complexes. Surface sequence conservation analysis along with SEC-MALS and native mass spectrometry reveal that formation of 4:2:2 complexes is restricted to FLRT2 and Unc5D homologues, and that other FLRT and Unc5 homologues form 2:1:1 tetrameric complexes, half the size of the octamers formed by Lphn3, FLRT2 and Unc5D.

4.2 Native Mass Spectrometry of Lphn3-FLRT2-Unc5D Complexes

Due to the technical difficulties in determining Lphn-FLRT2-Unc5D complex stoichiometries using SEC-MALS, Native Mass Spectrometry (MS) was employed to determine accurate masses of the intact complexes in the gas phase. In MS, molecular masses of complexes are measured without dilution, overcoming difficulties caused by complex dissociation at low concentration. In order to characterise the simplest possible system, I performed an experiment on the “minimal” Lphn3-FLRT2-Unc5D complex by mixing, FLRT2^{LRR}, Unc5D^{Ig} and Lphn3^{Lec-Olf} (Figure 4.1).

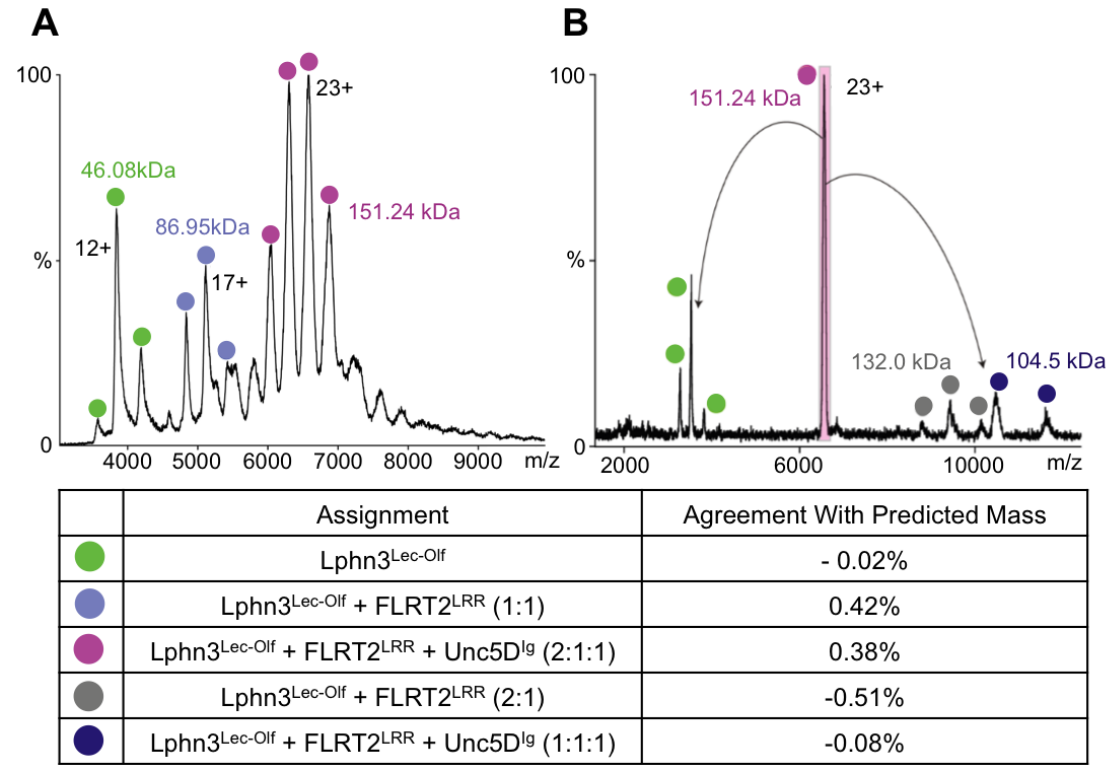


Figure 4.1: Native MS of Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{Ig}. Native MS was used to determine the molecular weights of complexes formed by Lphn3^{Lec-Olf}, FLRT2^{LRR} and Unc5D^{Ig}. The mass spectrum reveals masses corresponding to a 2:1:1 Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{Ig} tetramer, as well as 1:1 Lphn3^{Lec-Olf}-FLRT2^{LRR} dimers and Lphn3^{Lec-Olf} alone (A). MS/MS was used to validate the composition of the 151.24 kDa peak (B). The sample (~0.5 mg/ml) was transferred into 200 mM acetate (pH 7.5) immediately before collecting data. Native MS data collection and analysis was performed by Dr Shahid Mehmood.

This experiment revealed an unusual stoichiometry, not predicted by the SEC-MALS, but confirming low resolution crystallography results previously obtained

in the lab. Whereas SEC-MALS predicted a molecular weight of ~ 100 kDa for this complex (see Section 3.5), the highest molecular weight complex in the native mass spectrum was 151.24 kDa (Figure 4.1A), which is most consistent with a 2:1:1 (tetrameric) complex of $\text{Lphn3}^{\text{Lec-Olf}}\text{-FLRT2}^{\text{LRR}}\text{-Unc5D}^{\text{Ig}}$. Dissociation of the 151.24 kDa peak by tandem MS (MS/MS) resulted in peaks of 132.0 kDa (consistent with a 2:1 $\text{Lphn3}^{\text{Lec-Olf}}\text{-FLRT2}^{\text{LRR}}$ complex) and 104.5 kDa (consistent with a 1:1:1 complex of $\text{Lphn3}^{\text{Lec-Olf}}\text{-FLRT2}^{\text{LRR}}\text{-Unc5D}^{\text{Ig}}$), suggesting that one of the two $\text{Lphn3}^{\text{Lec-Olf}}$ molecules is more weakly bound than the other (Figure 4.1B).

Next I explored the higher-order multimerisation of Lphn3-FLRT2-Unc5D complexes observed when constructs of Unc5D including the Tsp1 domain were used in SEC-MALS experiments (Section 3.5). To do this, I performed a native MS experiment in which I mixed Lphn3^{Lec-Olf}, FLRT2^{LRR} and Unc5D^{Ecto} in a 2:1:1 ratio (Figure 4.2). The resulting mass spectrum revealed a peak corresponding to a mass of 367.39 kDa, which would be consistent with a 4:2:2 complex of Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{Ecto} (a “dimer of tetramers”). Peaks corresponding FLRT2^{LRR}-Unc5D^{Ecto} dimers and Lphn3^{Lec-Olf} alone were also observed (Figure 4.2).

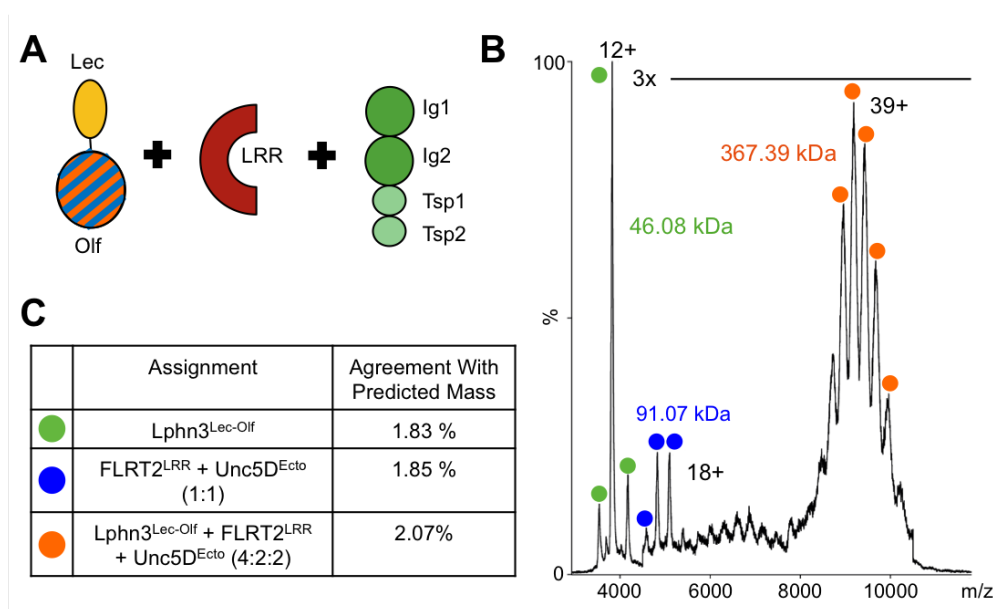


Figure 4.2: Native MS of Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{Ecto}. Native MS reveals that Lphn3^{Lec-Olf}, FLRT2^{LRR} and Unc5D^{Ecto} form a large oligomer, twice the size of a putative 2:1:1 complex. The complex was dialysed into 500 mM ammonium acetate (pH 7.5) and then diluted to a final concentration of 200 mM ammonium acetate. The sample concentration was ~ 1.5 mg/ml. Native MS data collection and analysis was performed by Dr Shahid Mehmood.

4.3 Crystallisation and Data Collection for

Lphn3^{Lec-Olf}-FLRT2^{LR}-Unc5D^{IgIgTsp}

In order to further understand the molecular basis for the formation of the 4:2:2 Lphn3-FLRT2-Unc5D complex I performed crystallisation trials using Lphn3^{Lec-Olf}, FLRT2^{LR} and Unc5D^{IgIgTsp}.

4.3.1 Crystallisation

All proteins for crystallisation were produced in kifunensine-treated HEK293T cells and purified as described in Section 2.2. Prior to preparative gel filtration, all three proteins were individually incubated with endoglycosidase F1 at 4 °C overnight to reduce the heterogeneity of N-linked glycans, which can impede crystallisation. Given that SEC-MALS data suggested that gel filtration resulted in complex dissociation, samples for crystallisation trials were prepared by mixing the purified components in the expected stoichiometric ratio (2:1:1). The high concentrations of protein typically used during crystallisation trials would be expected to enhance protein complex formation. The sample was concentrated to 8 mg/ml in 10 mM Tris-Base (pH 7.5), 150 mM NaCl, 100 mM non-detergent sulfobetaine-256. Crystals were obtained by mixing protein and reservoir solutions 1:1 in the sitting-drop vapour diffusion method. Initial trials were performed using multiple commercially available crystallisation screens, and small crystal hits were obtained from the JCSG I, III and Morpheus screens after one to four days at 18 °C. These initial hits were screened on beamline I02 at Diamond Light Source, with crystals from one condition from JCSGI (0.1 M MES pH 6.0, 20% 2-methyl-2,4-pentanediol, MPD) (Figure 4.3A) diffracting to ~ 8 Å. This was the condition I carried forward for optimisation.

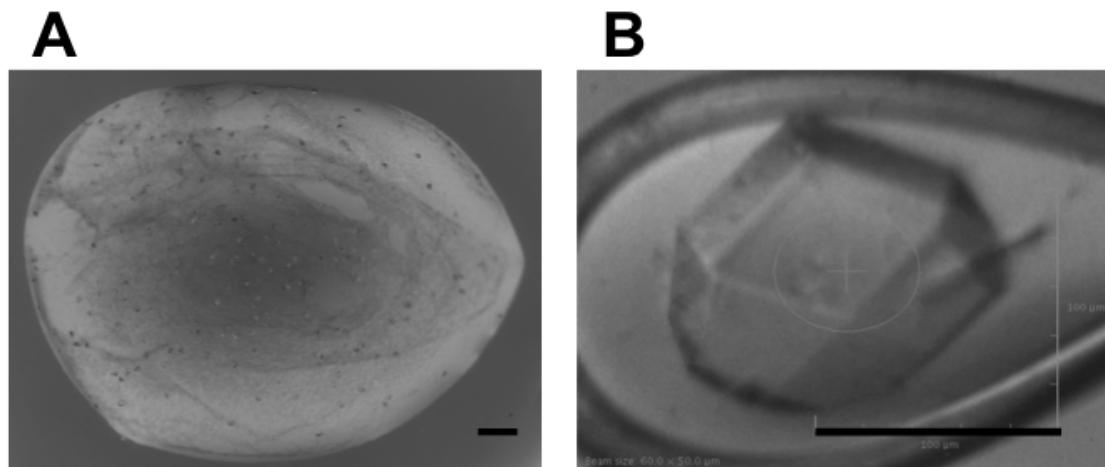


Figure 4.3: **Crystals of Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp}** Initial hits (A) and one of the optimised crystals from the Silver Bullets screen (~ 3.6 Å diffraction, B). Both scale bars = 100 μm .

In order to improve crystal diffraction, I employed a variety of standard techniques to optimise the crystal reservoir solution. The most successful of these was the use of additive screens. Various additives from the Silver Bullet screen (Hampton) (which contains a diverse range of small molecules) produced three crystals which diffracted to between 3.5 and 4.0 Å (crystals 1-3, Table 4.1). Figure 4.3B shows an example of one of these crystals. In addition, a slow-growing crystal from one of the initial Morpheus plates (crystal 4, Table 4.1) diffracted to ~ 3.5 Å. A summary of the reservoir conditions for the crystals used for structure solution is shown in Table 4.1.

	Crystallisation Solution	Cryoprotectant
Crystal 1	0.1 M MES (pH 6.0), 20% (v/v) MPD, 2 mM HEPES (pH 6.8), 0.025% (w/v) benzidine, 0.025% (w/v) nicotinamide, 0.025% (w/v) pyromellitic acid, 0.025% (w/v) sulphaguanidine.	95% crystallisation solution plus 5% MPD
Crystal 2	0.1 M MES (pH 6.0), 20% (v/v) MPD, 2 mM HEPES (pH 6.8), 0.016% (w/v) 3-Indolebutyric acid, 0.016% (w/v) hexadecanedioic acid, 0.016% (w/v) oxamic acid, 0.016% (w/v) pyrometallic acid, 0.016% (w/v) sebacic acid, 0.016% (w/v) suberic acid.	95% crystallisation solution plus 5% MPD
Crystal 3	0.1 M MES (pH 6.0), 20% (v/v) MPD, 2 mM HEPES (pH 6.8), 0.02% (w/v) 2,5-pyridinedi- carboxylic acid, 0.02% (w/v) pyromellitic acid, 0.02% (w/v) trans-Cinnamic acid, 0.02% (w/v) salicylic acid, 0.02% (w/v) trans-1,2-cyclohexanedicarboxylic acid.	95% crystallisation solution plus 5% MPD
Crystal 4	0.1 M ammonium sulphate, 0.3 M sodium formate, 3% (w/v) poly-gamma-glutamic acid LM, 10% (w/v) polyethylene glycol monomethyl ether 2000, 0.1 M sodium cacodylate (pH 6.5).	80% crystallisation solution plus 20% glycerol

Table 4.1: **Reservoir conditions used to grow optimised Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} crystals.**

4.3.2 Data Collection and Processing

For structure solution, data were collected on Diamond Light Source beamline i04-1 and European Synchrotron Radiation Facility beamline ID29 from four crystals. Experimental parameters for data collection are shown in Table 4.2.

Beamline	i04-1	Beamline	ID29
Wavelength (Å)	0.9282	Wavelength (Å)	0.97625
Beam size (μm)	60x50	Beam size (μm)	50x30
Oscillation (°)	0.2-0.3	Oscillation (°)	0.05
Exposure (s)	0.8	Exposure (s)	0.02
Transmission (%)	100	Transmission (%)	8.44
Number of Images	900	Number of Images	3600

Table 4.2: **Experiment parameters for native data collection from Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} crystals at Diamond Light Source (left) and European Synchrotron Radiation Facility (right).**

Data from all four crystals were integrated using XDS (via Xia2), revealing that all datasets could be considered isomorphous [71]. Integrated intensities from all four crystals were merged using Aimless. XDS identified the space group as C2. All

datasets were carefully assessed for the presence of radiation damage (indicated by a drop in the $CC_{1/2}$ in the high resolution shells and increased R_{meas} in all resolution shells), and images exhibiting significant radiation damage were excluded. Data reduction statistics for the final merged data are shown in Table 4.3.

Space group	C2
Cell dimensions	
a, b, c (Å)	231.96, 141.49, 151.49
α, β, γ (°)	90.00, 117.94, 90.00
Resolution (Å)	133.83 - 3.38 (3.47 - 3.38)
R_{merge} (%)	14.3 (136.3)
R_{meas} (%)	15.1 (171.6)
$CC_{1/2}$ (%)	99.1(39.2)
$I/\sigma I$	7.5 (0.5)
Highest resolution shell with $I/\sigma I > 2$	3.95-3.81
Completeness (%)	98.4 (92.4)
Multiplicity	8.0 (2.4)

Table 4.3: Data reduction statistics for the **Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} Complex**.

4.4 Solving the Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} Structure

Given that there are known structures for each of the proteins in these crystals [226], phasing could be performed using MR. Matthew’s coefficient calculations revealed a solvent content of 62% with one 4:2:2 Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} octamer in the asymmetric unit. MR was performed using Phaser, first searching for two copies of the FLRT2^{LRR}-Unc5D^{Ig1} complex [226] (TFZ=21.7, LLG=535), then two copies of Lphn3 Olf (TFZ=49.7, LLG=2953). Phaser located one copy of Lphn3 Lec and the other was placed by exploiting the 2-fold NCS (TFZ=45.1, LLG=2599). Unc5A Ig2 [226] was placed similarly (TFZ=52.3, LLG=2908). Finally, two additional copies of Lphn3 Olf were placed (TFZ=69.9, LLG=5143). Rigid-body refinement using BUSTER resulted in a model with $R_{\text{work}}/R_{\text{free}}=0.315/0.326$. The Unc5A Ig2 domain was replaced with a homology model of the Unc5D Ig2 domain and a homology model of Unc5D Tsp1 was placed into the density clearly visible after ini-

tial MR and refinement. Both homology models were made using SWISS-MODEL [22]. The model then underwent iterative cycles of manual adjustment in COOT [63] and all-atom refinement with BUSTER [30]. During refinement NCS, X-ray weight restraints and bond-length restraints were used. The final model was of good geometric quality (Table 4.4) and the electron density maps of highly quality, showing some side chain densities (Figure 4.4).

Resolution	133.83-3.4 (3.49 - 3.40)
No. reflections	58,669 (2,832)
R_{work}/R_{free}	0.225/0.245 (0.261/0.253)
No. atoms	19,616
Average B-factors ($\text{\AA}^2$)	185
R.m.s. deviations	
Bond lengths ($\text{\AA}$)	0.007
Bond angles ($^\circ$)	0.92
Ramachandran Statistics	
Most favoured regions (%)	91.1
Outliers (%)	0.29

Table 4.4: **Refinement statistics for Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} Complex.** Values for the highest resolution shell are shown in parentheses.

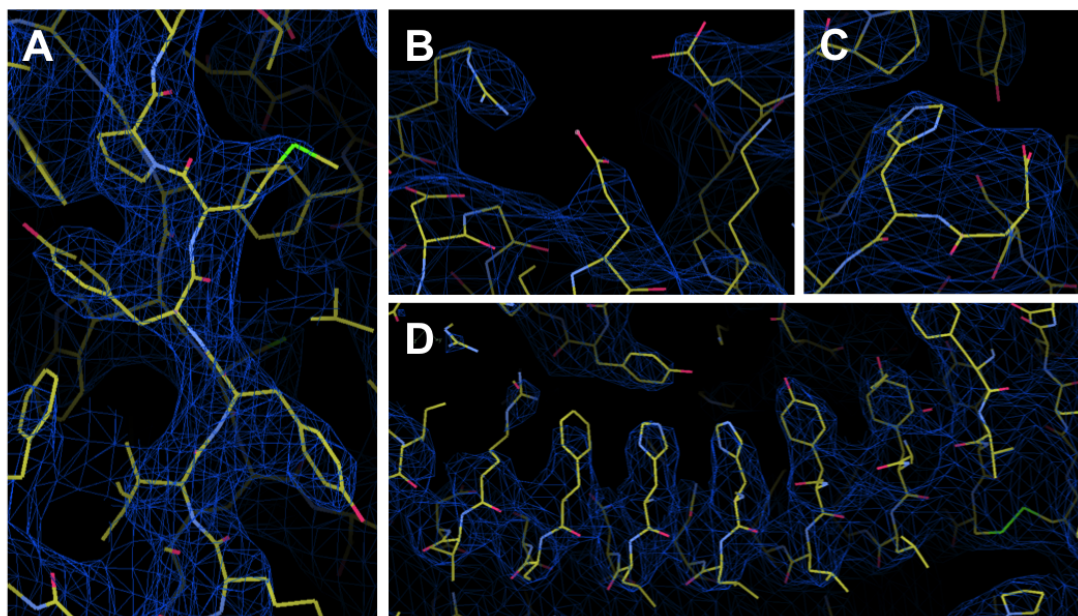


Figure 4.4: **Final electron density for the Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} complex.** Electron density maps (1.0σ) showing portions of Lphn3 (A, C) and FLRT2 (B, D).

4.5 Analysis of the Lphn3-FLRT2-Unc5D Octameric Complex Structure

The crystal structure of Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} revealed an intricate octameric arrangement of the three proteins with a total buried surface area of $\sim 19,000 \text{ \AA}^2$ (Figure 4.5). Two chains of Unc5D^{IgIgTsp} pack in an antiparallel arrangement, bridging the two pseudo-symmetric halves of the complex, each of which comprises a 2:1:1 Lphn3-FLRT2-Unc5D tetramer (Figure 4.6). The Unc5D Ig1 domain binds the lateral side of the FLRT2^{LRR} in the same manner as in the 1:1 FLRT2^{LRR}-Unc5D^{Ig1} complex solved previously (r.m.s.d. $0.7 \text{ \AA}$ for 423 aligned C α atoms) [226]. The Unc5D Tsp1 domain also contacts the lateral side of the FLRT2^{LRR}, adjacent to the binding site for Unc5D Ig1 (Figure 4.6A). Compared with the Unc5A^{Ecto} structure (PDB accession 4V2A), Unc5D^{IgIgTSP} bends at two hinge positions (Ig1-Ig2, Ig2-TSP1; Figure 4.7A,B).

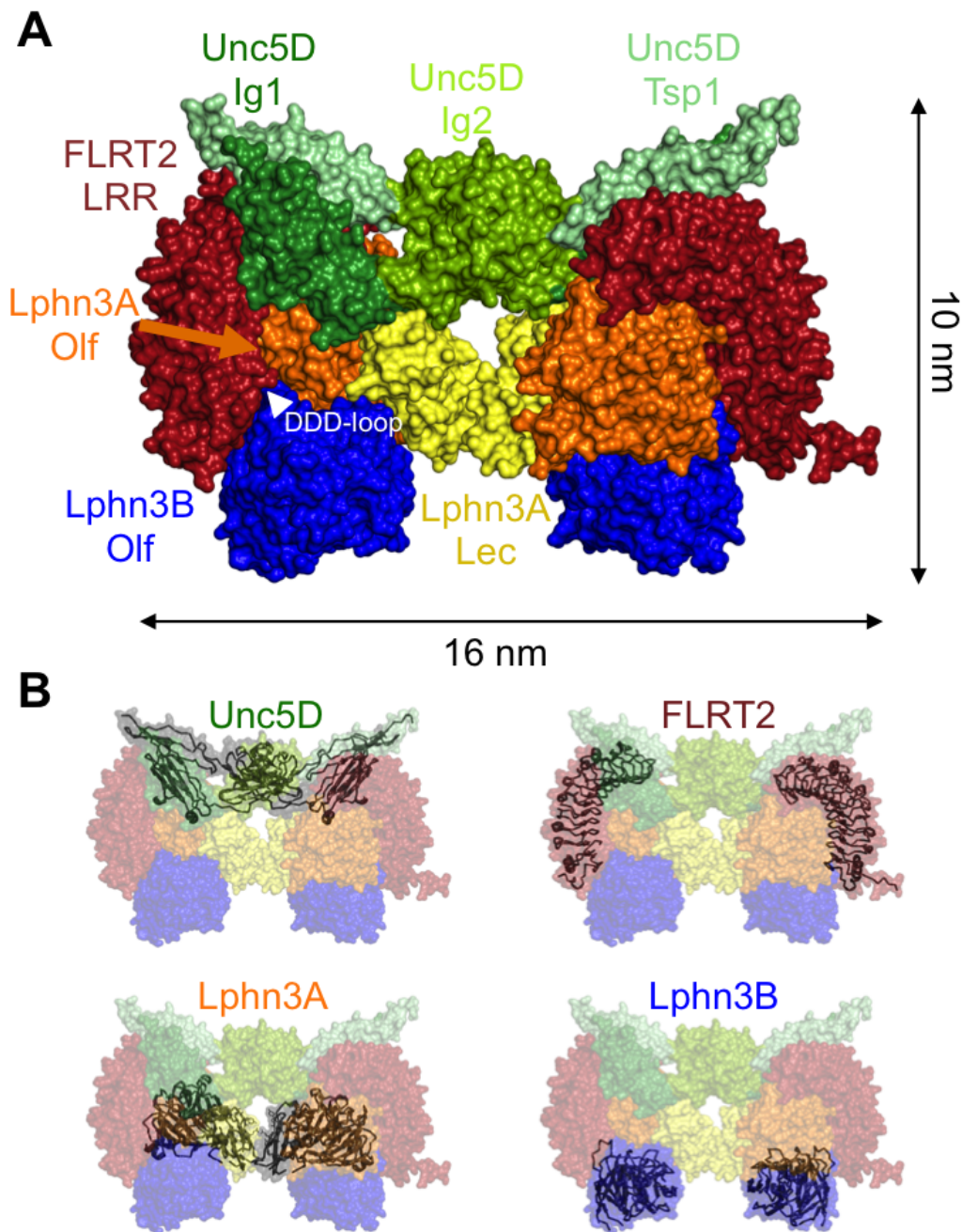


Figure 4.5: **The structure of the 4:2:2 Lphn3-FLRT2-Unc5D Complex.** The protein chains in the structure are coloured: FLRT2^{LRR}, red; Unc5D^{Ig1}, Ig1 - dark green, Ig2 - medium green, Tsp1 - pale green; Lphn3A^{Lec-Olf}, Olf - orange, Lec - yellow; Lphn3B^{Lec-Olf}, blue. The Lec domain of Lphn3B is not resolved in the crystal structure and is presumably disordered in the crystal (A). Each chain is coloured as a black cartoon within a surface model of the complex (B). The figure was prepared using PyMol.

Each copy of FLRT2^{LRR} uses its concave surface to bind two copies of Lphn3^{Lec-Olf} (Lphn3A^{Lec-Olf} and Lphn3B^{Lec-Olf}) (Figure 4.5B). The Olf domain of Lphn3A^{Lec-Olf} occupies a binding site on the concave dimerisation surface of FLRT2^{LRR} [226].

This arrangement strongly agrees with my mutagenesis data presented in Chapter 3. Comparison of Lphn3A^{Lec-Olf} with the unliganded structure of the Lphn3^{Lec-Olf} reveals a re-orientation of the lectin domain to make contacts with the Lphn3A Olf domain and the Unc5D Ig1-Ig2 linker (Figure 4.7C, D). The Olf domain of the second copy of Lphn3^{Lec-Olf} (Lphn3B^{Lec-Olf}) makes extensive contacts with the Lphn3A Lec and Olf domains and inserts a negatively charged loop (S393-N405) into a cleft formed by FLRT2^{LRR} and Lphn3A^{Lec-Olf}. This loop is termed the “DDD loop” as it contains three consecutive aspartate residues (D397, D398 and D399, Figure 4.5A). The DDD-loop was unresolved in the crystal structure of the unliganded Lphn3^{Lec-Olf}, suggesting that it only becomes ordered when bound to FLRT^{LRR}/Lphn3A^{Lec-Olf}. There is no visible electron density for Lphn3B^{Lec-Olf} Lec domain, and it makes no contacts with the complex.

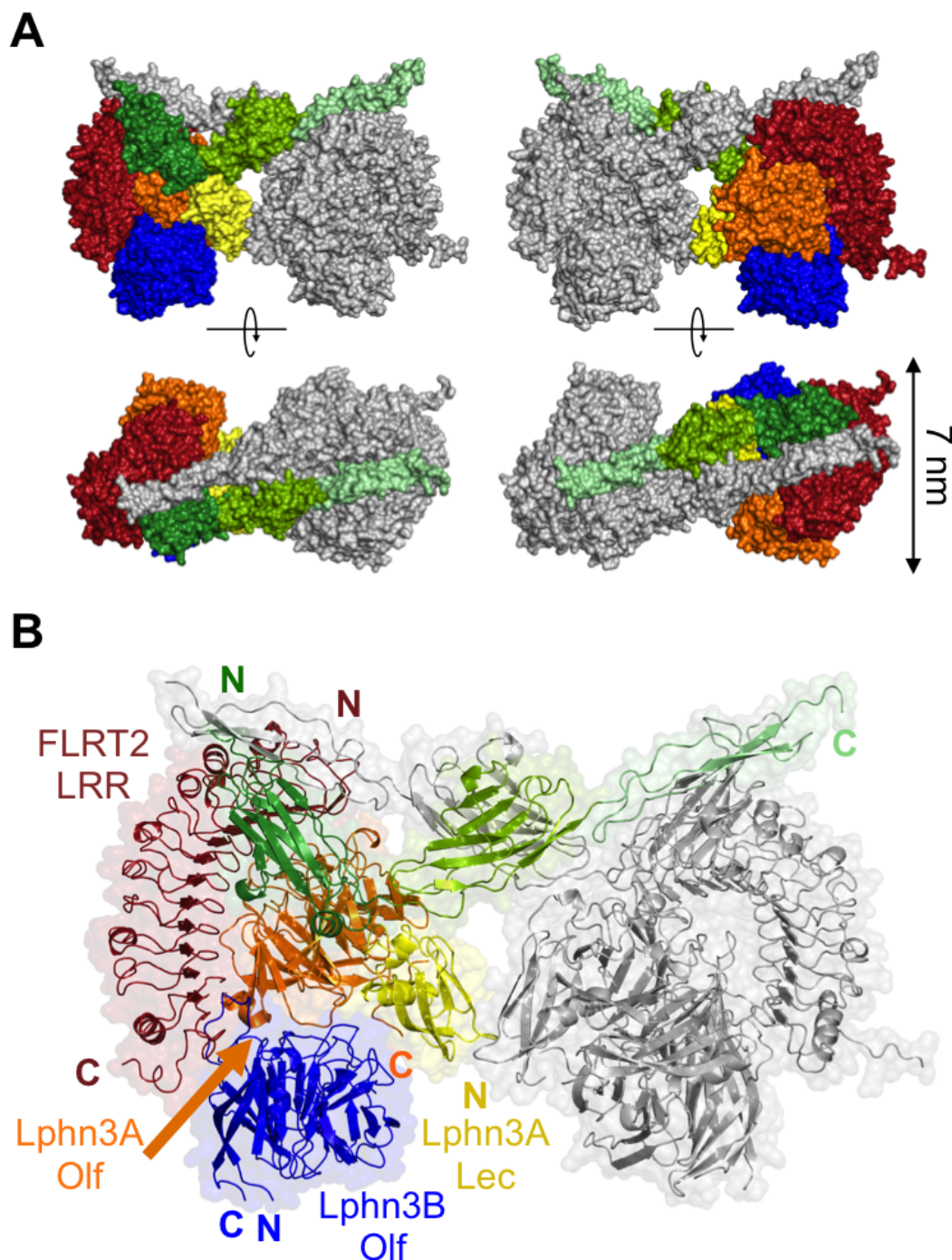


Figure 4.6: **Unc5D bridges two pseudo-symmetric 2:1:1 tetramers.** Two 2:1:1 Lphn3^{Lec}-Olf-FLRT2^{LRR}-Unc5D^{Ig2}Tsp1 tetramers are bridged by antiparallel packing of the Unc5D Ig2 and Tsp1 domains. The pseudo-symmetric octamer is shown in surface (A) and cartoon (B) representation with one half coloured in grey. All other colours are as in Figure 4.5. The figure was prepared using PyMol.

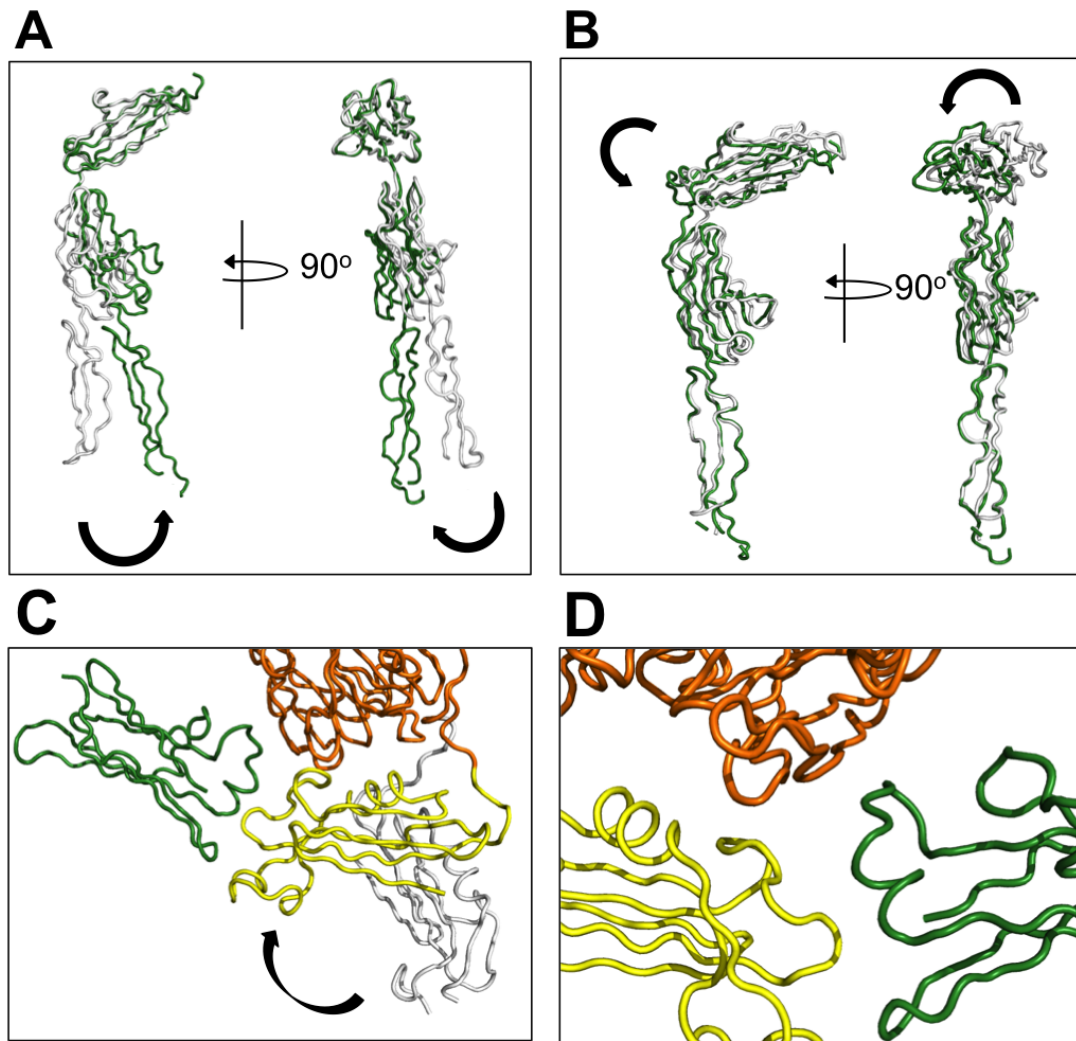


Figure 4.7: **Structural rearrangements of Lphn3^{Lec-Olf} and UncD^{IgIgTsp}.** Comparison of Unc5A^{IgIgTsp} [226] and Unc5D^{IgIgTsp} crystal structures. The structure of Unc5D^{IgIgTsp} (green) from the 4:2:2 Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} crystal structure was aligned with that of the unliganded Unc5A ectodomain (white, PDB accession 4V2A) via either the Ig1 domain (A) or the Tsp1 domain (B), showing major conformational differences at a hinge point in the Ig1-Ig2 linker, with a smaller reorientation around the Ig2-Tsp1 linker. The Lphn3^{Lec-Olf} linker region acts as a flexible hinge to reorientate the Lec domain. In the structure of the unliganded Lphn3^{Lec-Olf} the Lec and Olf domains adopt an extended conformation (white). The Lec domain of Lphn3A^{Lec-Olf} in the Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} structure (yellow) undergoes a rotation and translation (C). This movement brings the Lec domain into contact with the Unc5D Ig-Ig2 linker and the Lphn3A Olf domain (D). The figure was prepared using PyMol.

4.6 Mutational Validation of Novel Super-Complex Interfaces

4.6.1 Disrupting Binding of Lphn3B

In order to validate the arrangement of proteins observed in the crystal structure I made mutations in the novel super-complex interfaces. The first interface to investigate was that formed by the negatively-charged “DDD-loop” of Lphn3B^{Lec-Olf}, which intercalates into a cleft formed by the Lphn3A Olf domain and FLRT2^{LRR}. Four residues in the DDD-loop, D397, D398, D399 and N400 are potentially involved in forming hydrogen bonds with residues on FLRT2^{LRR} and Lphn3A Olf (Figure 4.8A). I designed a charge-reversal mutant Lphn3^{Lec-Olf} in which all three of these aspartate residues were mutated to arginine (D397R D398R D399R). This mutant would presumably prevent binding of Lphn3B^{Lec-Olf} to the complex, but would not affect binding between Lphn3A^{Lec-Olf} and FLRT2^{LRR}. I performed native MS with Unc5D^{Ig} and FLRT2^{LRR} to assess the complexation ability of this Lphn3^{Lec-Olf} mutant. In contrast to wild-type Lphn3^{Lec-Olf} (which forms a 2:1:1 tetrameric complex with FLRT2^{LRR} and Unc5D^{Ig}, Figure 4.8B), using the charge-reversal mutant resulted in peaks of ~110 kDa appearing, corresponding to masses of a 1:1:1 trimeric complex. The amount of species assigned to the 2:1:1 tetramer was relatively reduced (Figure 4.8C). A second mutation, which introduced an artificial N-linked glycan into the DDD-loop (D397N D399T) more completely disrupted the formation of the tetrameric complex, presumably because the bulky glycan is sterically incompatible with Lphn3B DDD-loop binding into the cleft formed by FLRT2^{LRR} and Lphn3 Olf (Figure 4.8D). From herein this mutant is referred to as the “DDD mutant”.

Notably, both mutations still result in masses consistent with 1:1:1 trimeric complexes, indicating that the DDD-loop is important for binding of Lphn3B^{Lec-Olf}, but not Lphn3A^{Lec-Olf}. Consistent with this, SPR experiments using immobilised

Unc5D^{Ecto} and FLRT2^{LRR}+Lphn3^{Lec-Olf} DDD mutant as analyte showed an intermediate level of binding compared to experiments using wild-type Lphn3^{Lec-Olf}+FLRT2^{LRR} or FLRT2^{LRR} alone (Figure 4.8E). The calculated K_D s were similar for Unc5D^{Ecto} binding to wild-type Lphn3^{Lec-Olf}+FLRT2^{LRR} or when Lphn3^{Lec-Olf} DDD mutant was used instead (apparent K_D 94 nM in both cases, compared to 460 nM for FLRT2^{LRR} alone, see Section 3.5). This suggests that the enhanced affinity of FLRT2^{LRR} for Unc5D^{Ecto} in the presence of Lphn3^{Lec-Olf} depends mainly on Lphn3A, and is consistent with the crystal structure as Lphn3B makes no contacts with either of the Unc5D molecules.

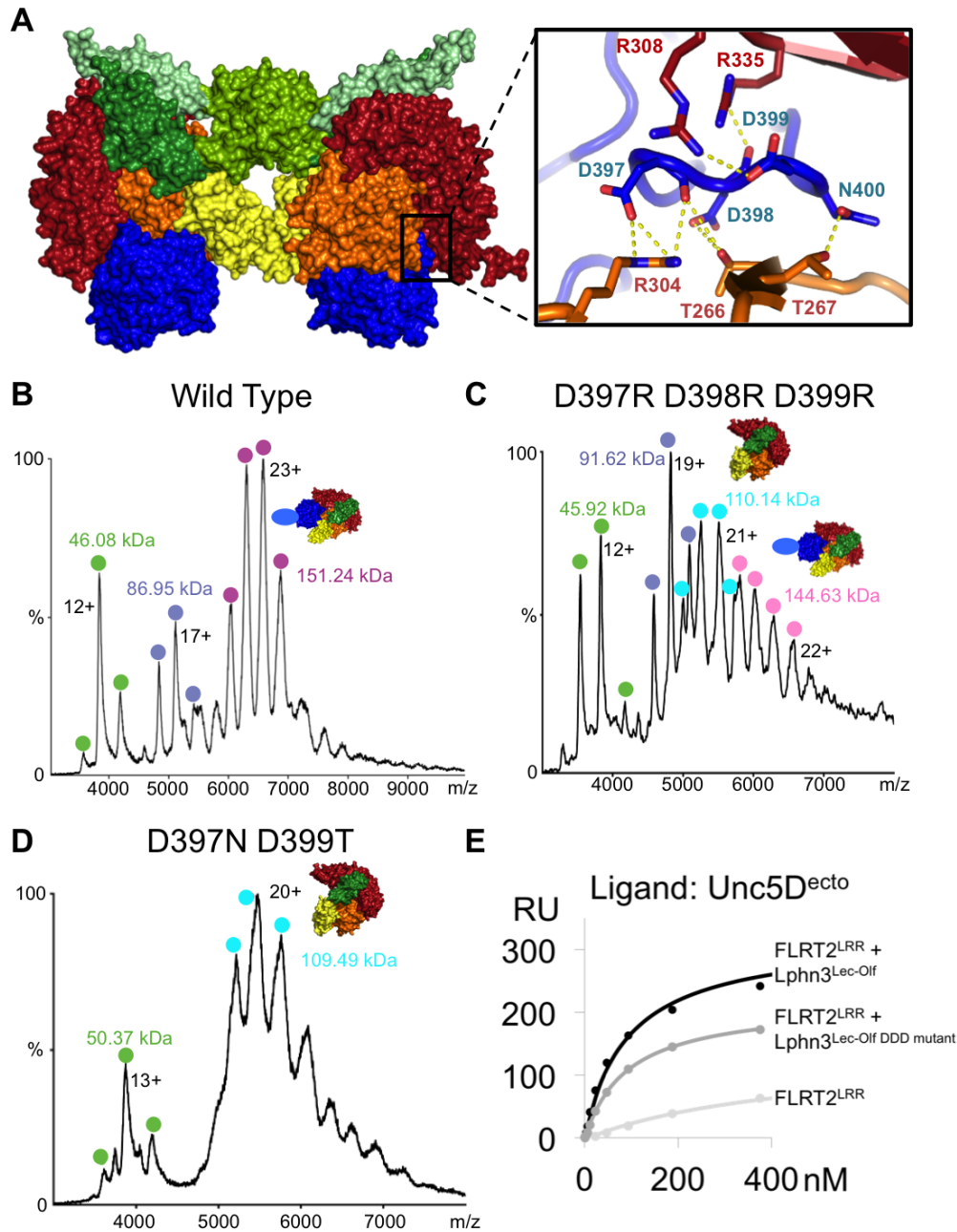


Figure 4.8: The Lphn3 DDD-loop mediates binding of Lphn3B. Charge-reversal and N-linked glycosylation mutations were made in the Lphn3 DDD-loop (A). Charge-reversal reduces the abundance of tetrameric (2:1:1) Lphn3^{Lec-Of}-FLRT2^{LRR}-Unc5D^{Ig} complexes and increases the abundance of 1:1:1 trimeric complexes (B,C). Introduction of an N-linked glycan into the DDD-loop fully disrupts the ability of Lphn3B to bind, resulting in only trimeric 1:1:1 complexes (B,D). For all MS experiments, the most consistent peak assignments are depicted as surface models above their corresponding peaks. Colours are as in Figure 4.5. The Lphn3B Lec domain, which is not resolved in the crystal structure, is shown as a blue oval. All native MS samples were measured at concentrations of ~0.5-2.5 mg/ml in 200 mM ammonium acetate (pH 7.5). Native MS data collection and analysis was performed by Dr Shahid Mehmood. SPR experiments with Unc5D^{Ecto} immobilised on a streptavidin-coated CM5 chip show that the Lphn3^{Lec-Of} DDD mutant (mixed 2:1 with FLRT2^{LRR}) displays an intermediate level of binding ($B_{\max}$) when compared to wild-type Lphn3^{Lec-Of}+FLRT2^{LRR} or FLRT2^{LRR} alone.

4.6.2 Disrupting the FLRT2^{LRR}-Unc5D^{Tsp1} Interaction

The other super-complex interface that was interesting to disrupt was that formed between the lateral side of FLRT2^{LRR} and Unc5D Tsp1. This interaction is clearly important for complex multimerisation as complexes formed by Unc5D constructs lacking the Tsp1 domain were not capable of multimerising (see Section 3.5). Presumably mutations in this interface would disrupt the formation of the octameric complex, but not affect formation of a tetrameric (2:1:1) complex as Unc5D Tsp1 is not necessary to form tetramers. To disrupt this interface, I engineered an artificial N-linked glycosylation site in Unc5D Tsp1 at its interface with FLRT2 LRR (M292N V294T, Figure 4.9A). SEC-MALS experiments show that this mutation reduces the octameric complex to a mass corresponding to a tetramer (Figure 4.9B).

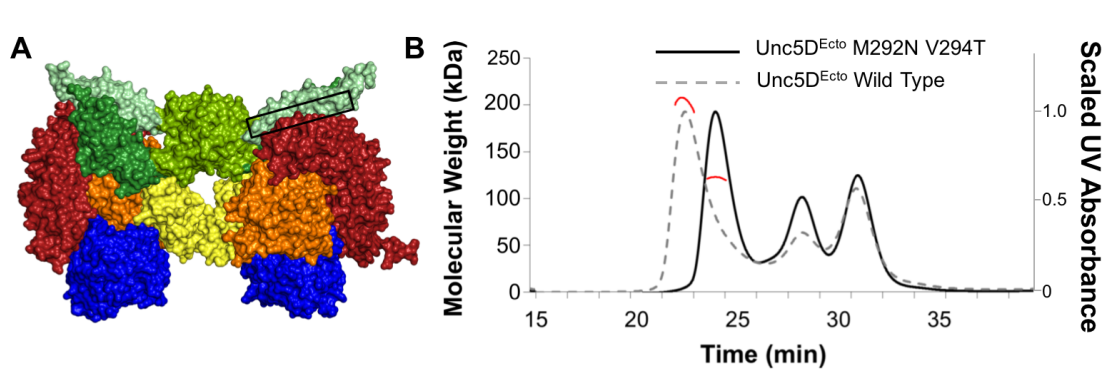


Figure 4.9: Mutations in Unc5D Tsp1 abolish octamer formation. An artificial N-linked glycosylation site was introduced in Unc5D Tsp1 at the LRR-Tsp1 interface (boxed) (A). SEC-MALS experiments using Lphn3^{Lec-Olf}+FLRT2^{LRR} mixed with either wild-type or M292N V294T Unc5D^{Ecto} revealed that the M292N V294T mutation reduces the observed molecular mass of the complex from >200 kDa to a mass corresponding to a tetramer. Both protein samples were injected at ~3.5 mg/ml. The molecular weight of each complex is shown by the red line above each peak compared to the y-axis on the left. The UV absorbance of the SEC profile is also shown for each sample (grey dashed line, wild type; solid black line, mutant). The UV absorbance traces were scaled according to their maximum peak height and are shown compared to y-axis on the right. As discussed in Section 3.5, the apparent molecular weights reflect the average values within a given peak and so can be used only for relative comparisons.

4.7 Conservation of Novel Super-Complex Interfaces

As discussed in the previous chapter, protein-protein interaction sites typically exhibit high surface sequence conservation. To visualise this I used the Consurf server to calculate conservation scores for FLRT and Unc5 and map them onto the struc-

tures of these proteins [80]. Using alignments of all Unc5 (Unc5A-D) homologues from zebrafish, rat, chicken, shark and frog I saw that the novel binding site for the Lphn3A Lec domain on Unc5D Ig1-Ig2 is highly conserved (Figure 4.10A), this is also the case when an alignment of just Unc5D sequences from the same species was used (Figure 4.10C). However, the FLRT^{LRR}-binding surface on Unc5D Tsp1, although well conserved between Unc5D orthologues (Figure 4.10D), was relatively less well conserved across all Unc5 homologues in several species (Figure 4.10B). A similar picture emerged for the FLRTs; alignments of all FLRT homologues from mouse, rat, chicken, frog, zebrafish and puffer fish revealed that the Unc5 Ig1 binding surface on the lateral side of the FLRT2 LRR is very highly conserved (as in [226]), but the adjacent Tsp1-binding surface is relatively less well conserved (Figure 4.10E). However, when an alignment containing just FLRT2 homologues was used, the Tsp1-binding surface is highly conserved (Figure 4.10F). These results suggest that FLRT2 and Unc5D may be the only FLRT and Unc5 homologues forming these interfaces.

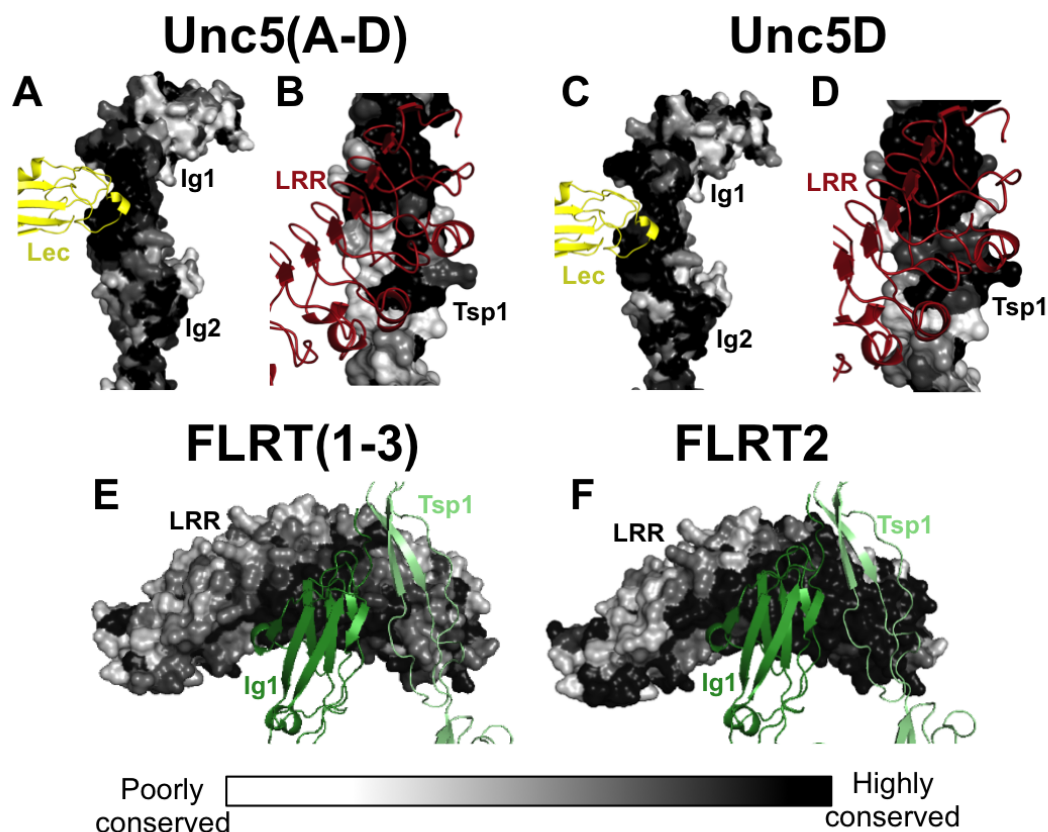


Figure 4.10: **Sequence conservation across the Unc5 and FLRT protein families.** Conservation scores from sequence alignments (rat, chicken, frog, zebrafish and shark) of all Unc5(A-D) (A,B) or only Unc5D homologues (C,D) were mapped onto the surface of Unc5D. The interface formed between Lphn3A Lec (yellow cartoon) and Unc5D Ig1-Ig2 is highly conserved across all Unc5 homologues (A,C). The interface between FLRT2 LRR (dark red cartoon) and Unc5D Tsp1 is less conserved across Unc5A-D (B) compared to Unc5D orthologues only (D). Similarly for the FLRT family, conservation scores computed from alignments from mouse, rat, chicken, frog and zebrafish FLRT sequences reveal that the Tsp1 binding site on the FLRT LRR is highly conserved for FLRT2 orthologues (F), but less well conserved across the whole FLRT family (E). Black = highly conserved, white = poorly conserved. The figure was made using PyMol.

The interaction between Unc5D Ig1-Ig2 and Lphn3A Lec domain reorientates the Lec domain (Figure 4.7) so as to provide stabilising interactions with Lphn3B. Therefore, this interaction is presumably important in forming 2:1:1 Lphn-FLRT-Unc5 tetramers. The fact that the binding site for Lphn3 Lec on Unc5D Ig1-Ig2 is conserved across all Unc5 homologues suggests that all Unc5 proteins may be able to form 2:1:1 tetramers with Lphn and FLRT.

In contrast to this, the key interface in forming the octameric complex is that formed between the FLRT2 LRR and Unc5D Tsp1. The poor conservation of these surfaces

across the Unc5 and FLRT families, but high conservation amongst FLRT2 and Unc5D orthologues suggests that formation of octameric (4:2:2) complexes may be specific to FLRT2 and Unc5D homologues.

4.8 Stoichiometry of Other Lphn-FLRT-Unc5 Complexes

In order to experimentally validate the predictions from the surface sequence conservation analysis I performed a variety of native MS and SEC-MALs experiments using multiple homologues of all three proteins. These experiments gave a consistent set of results suggesting that only a combination of FLRT2 and Unc5D were able to form a 4:2:2 complex with Lphn proteins and that other FLRT and Unc5 combinations formed 2:1:1 complexes with Lphn. To avoid large amounts of repetition, here I present a subset of the experiments that support these conclusions.

Lphn3 also binds FLRT3 [191], the cognate interaction partner of Unc5B. Native MS experiments using Lphn3^{Lec-Olf}, FLRT3^{LRR} and Unc5B^{Ecto} revealed that these proteins form 2:1:1 tetramers, but not 4:2:2 octamers (Figure 4.11). In addition, replacing wild-type Lphn3^{Lec-Olf} with the DDD mutant, reduced these 2:1:1 tetramers to 1:1:1 trimers, suggesting that Lphn3B binds in the same conformation to this complex as it does to the Lphn3^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{IgIgTsp} complex (Figure 4.11).

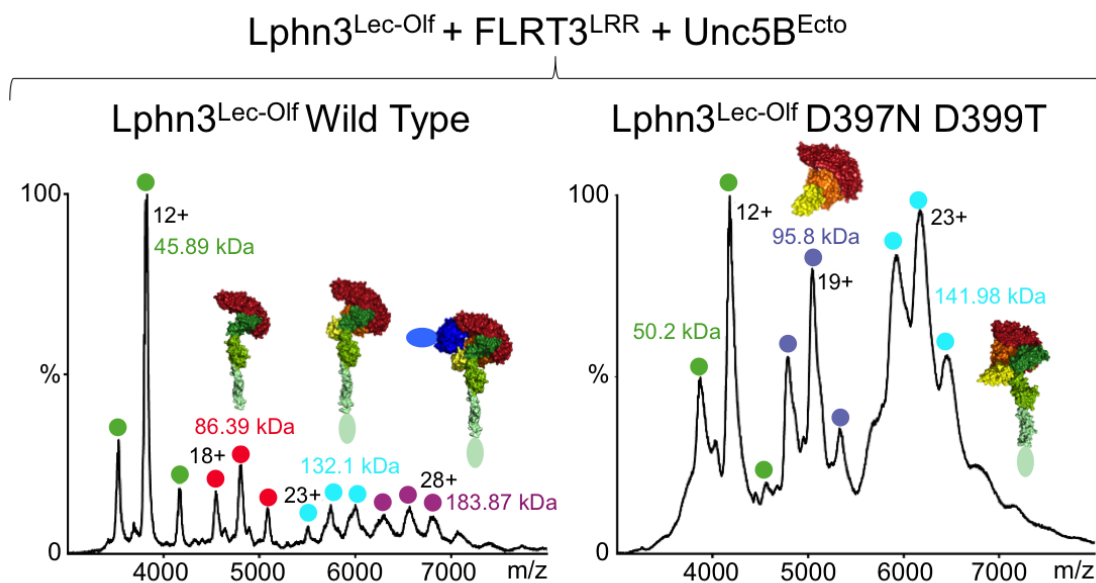


Figure 4.11: **Lphn3, FLRT3 and Unc5B form a 2:1:1 Tetramer** Native MS spectra of complexes formed when mixing either wild-type (left) or DDD mutant (right) $\text{Lphn3}^{\text{Lec-Olf}}$, $\text{FLRT3}^{\text{LRR}}$ and $\text{Unc5B}^{\text{Ecto}}$ in a 2:1:1 ratio. Assigned complexes for each peak are shown as comparable species from the crystal structure of $\text{Lphn3}^{\text{Lec-Olf}}\text{-FLRT2}^{\text{LRR}}\text{-Unc5D}^{\text{IgTsp}}$. The Lphn3B Lec domain, which is unresolved in the crystal structure, is shown as a blue oval. The Unc5 Tsp2 domain, which is not present in the $\text{Unc5D}^{\text{IgTsp}}$ crystallisation construct is depicted as a pale green oval. Samples were measured at ~ 1.6 mg/ml in 200 mM ammonium acetate pH 7.5. Native MS data collection and analysis was performed by Dr Shahid Mehmood.

Given that $\text{Lphn3}^{\text{Lec-Olf}}$, $\text{FLRT3}^{\text{LRR}}$ and $\text{Unc5B}^{\text{Ecto}}$ failed to form an octameric complex, it was interesting to test whether simply switching the FLRT homologue used might also prevent the formation of octamers. SEC-MALS experiments in which $\text{Lphn3}^{\text{Lec-Olf}}$ and $\text{Unc5D}^{\text{Ecto}}$ were mixed with $\text{FLRT1/2/3}^{\text{LRR}}$ revealed that molecular weights corresponding to octamers (>200 kDa) were only observed when $\text{FLRT2}^{\text{LRR}}$ was used (Figure 4.12). Consistent with the surface sequence conservation results, FLRT1 and FLRT3 do not seem capable of mediating octamerisation.

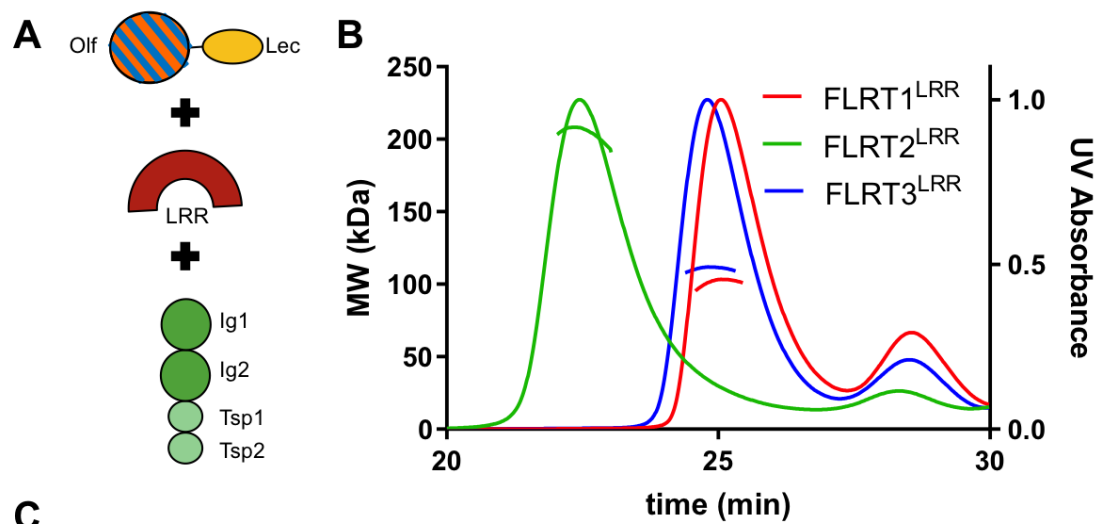


Figure 4.12: **Octamerisation is restricted to FLRT2.** Lphn3^{Lec-Olf}, Unc5D^{Ecto} and FLRT^{LRR} were mixed for each FLRT homologue (FLRT1-3) (A) and complexes formed analysed by SEC-MALS (B). All samples were injected onto the column at a concentration of ~ 3 mg/ml. Observed molecular weights for each complex are given in (C). As discussed in Section 3.5, the apparent molecular weights reflect the average values within a given peak and so can be used only for relative comparisons. MW: Molecular weight.

Finally I decided to assess whether changing the Lphn homologue used would affect the propensity of the complex to form either 2:1:1 tetramers or 4:2:2 octamers. To do this I mixed Lphn1^{Lec-Olf} and FLRT2^{LRR} with either Unc5D^{Ig} or Unc5D^{Ecto} in a 2:1:1 ratio and performed native MS and SEC-MALS experiments (Figure 4.13). Lphn1^{Lec-Olf} behaved similarly to Lphn3^{Lec-Olf} in these experiments, supporting the formation of both 2:1:1 tetramers with FLRT2^{LRR} and Unc5D^{Ig} and 4:2:2 octamers with FLRT2^{LRR} and Unc5D^{Ecto}.

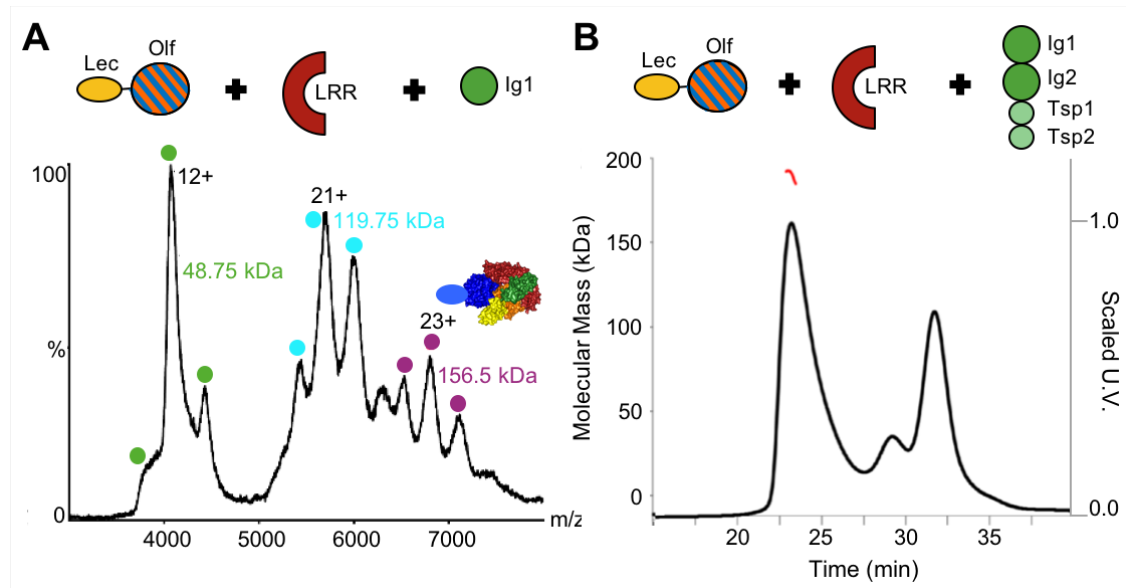


Figure 4.13: Lphn1 can form both 2:1:1 Tetramers and 4:2:2 Octamers with FLRT2 and Unc5D. Lphn1^{Lec-Olf} and FLRT2^{LRR} was mixed with either Unc5D^{Ig} (A) or Unc5D^{Ecto} (B) in a 2:1:1 ratio and used for native MS (~ 1.6 mg/ml, A) or SEC-MALS experiments (~ 3.7 mg/ml, B). The native MS spectrum reveals masses consistent with 2:1:1 Lphn1^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{Ig} tetramers (156.5 kDa) (A) and the SEC-MALS profile of Lphn1^{Lec-Olf}-FLRT2^{LRR}-Unc5D^{Ecto} is consistent with that of an octameric complex. The molecular mass of the main peak in this profile is represented by the red line and shown according the y-axis on the left (B). As discussed in Section 3.5, the apparent molecular weights from SEC-MALS experiments with this complex reflect the average values within a given peak and so can be used only for relative comparisons

4.9 Discussion

4.9.1 Three Receptors - At Least Five Possible Complexes

The data presented in this chapter reveal unexpected structural versatility for the interactions between three different types of cell surface receptor: the adhesion GPCR Lphn; the homophilic adhesion molecule FLRT and the repulsive guidance receptor Unc5. Whilst performing the work detailed in this chapter, two 1:1 crystal structures of Lphn3^{Olf} bound to FLRT3^{LRR} were published [208, 158]. This, along with other published data [226] and data presented in this chapter demonstrates that these three receptors can form at least five cell-surface complexes: FLRT-FLRT dimers [226], binary complexes of 1:1 FLRT-Unc5 [226] and FLRT-Lphn [208, 158] and homologue-specific tetrameric or octameric super-complexes containing multiple copies of each protein (Figure 4.14). The number of possible complexes vastly increases when one considers that each component may be expressed on individual cells in different combinations (i.e. with components present either *in cis* or *in trans*). These results show that the aGPCR Lphn mediates formation of diverse complexes, and are consistent with Lphn acting as a so-called “context-recogniser”, forming distinct complexes depending on the molecular composition of its local environment.

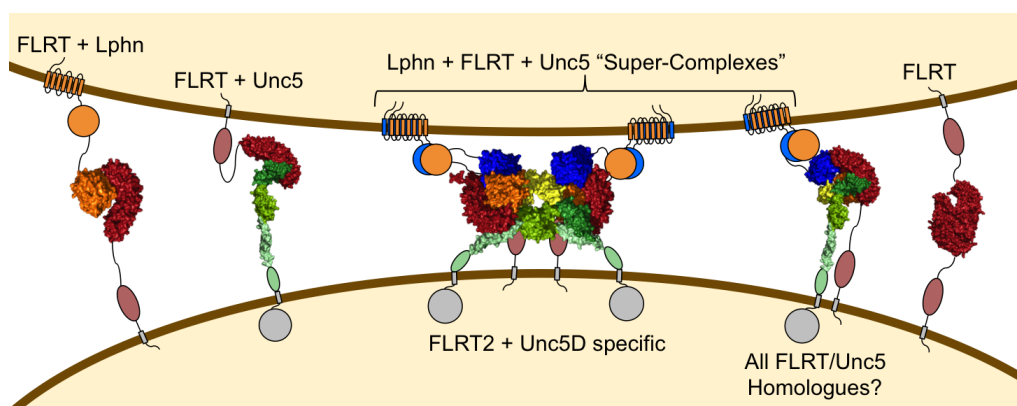


Figure 4.14: **The diversity of FLRT, Unc5 and Lphn Complexes.** Schematic showing the potential arrangement of receptor complexes at the cell surface. Protein domains are shown in surface representation and are coloured as they are throughout this chapter. Intracellular domains and domains for which there are no known structures are depicted as circles.

4.9.2 Potential Roles of Homologue-Dependent Super-Complexes

The data presented in this chapter lead to a model in which binding of Lphn to FLRT2 and Unc5D homologues promotes formation of 4:2:2 octameric complexes, whereas engagement of Lphn with other FLRT and Unc5 homologues promotes formation of 2:1:1 tetramers. FLRT2-Unc5D functions are best understood in the context of cortical development in the mouse, where different FLRT and Unc5 homologues display distinct expression patterns [226]. Unc5D is expressed in non-migrating neurons in the subventricular zone. FLRT2 is strongly expressed in the cortical plate, from where its ectodomain is shed and diffuses to the subventricular zone to delay the migration of Unc5D+ neurons towards the cortical plate. The FLRT3/Unc5B expression pattern is significantly different, with FLRT3 expressed by neurons in the intermediate zone as they migrate toward the cortical plate, where Unc5B is expressed (Figure 4.15). [226].

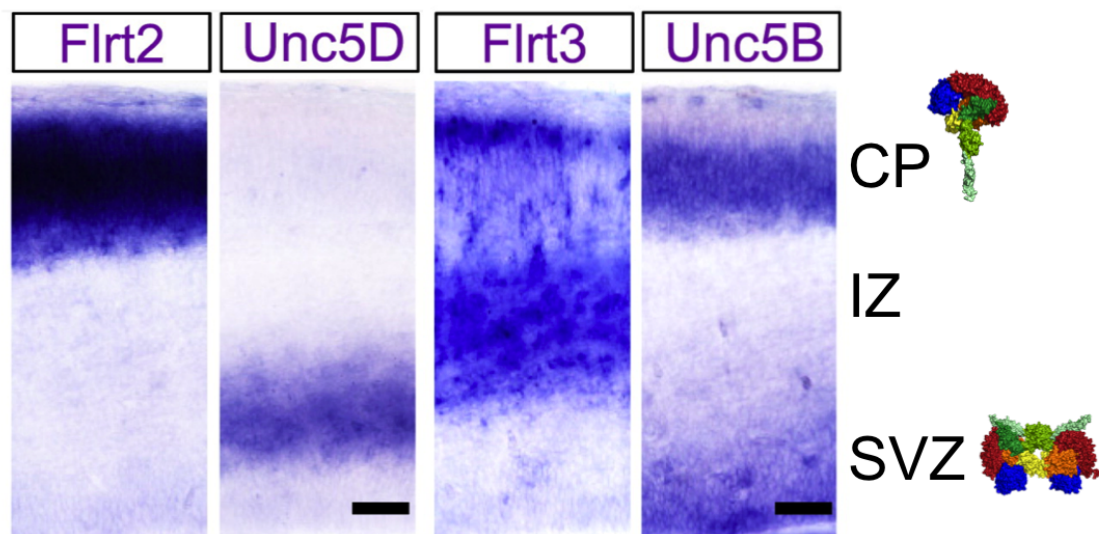


Figure 4.15: **FLRT and Unc5 homologues are differentially expressed in the developing mouse cortex.** *In situ* hybridisation showing FLRT2/3 and Unc5B/D expression in coronal sections of the mouse E15.5 cortex. The proposed architectures of Lphn-FLRT-Unc5 complexes formed in the cortical plate and subventricular zone are shown alongside the *in situ*. The figure was adapted from [226]. CP, cortical plate; IZ, intermediate zone; SVZ, subventricular zone.

Lphns are broadly expressed in the mouse cortex [191, 192]. Therefore it is conceivable that different structural assemblies of Lphn, FLRT and Unc5 homologues could form on migrating and non-migrating neurons, depending on which FLRT and Unc5

they express (or bind to).

4.9.3 Lphn Dimerisation

The complexes discussed in this chapter have unequal stoichiometries, containing two copies of Lphn for each copy of FLRT and Unc5. These results suggest that Lphn may act as a constitutive dimer or that it is dimerised upon engagement with FLRT/Unc5. Previous studies have shown that the extracellular domain of the *C. elegans* Lphn homologue, Lat-1, exists as a mixture of monomers and non-covalently-linked dimers. The authors suggested a mechanism whereby ligand-binding to the Lphn Lec domain promotes dimerisation of the ectodomain, leading to cross-activation of the 7TM by the partner molecule [205]. Further, stimulation of Lphn-expressing cells with α -latrotoxin results in oligomerisation of the Lphn CTF [266], suggesting that dimerisation of the receptor may play a role in its signal transduction mechanism.

On the whole, GPCR dimerisation is a highly controversial area of research. Many GPCRs dimerise, but the functional relevance of such a self-association is unclear in most cases [84]. This is even more poorly understood for members of the aGPCR family. GPR116, ELTD1 and EMR2 have also been shown to exist as dimers or oligomers [3, 182, 47], although for EMR2 this dimerisation is independent of the receptor's ectodomain [47].

Results presented in this chapter indicate that FLRT/Unc5-mediated dimerisation of the Lphn ectodomain is crucially dependent on the Lphn DDD-loop. Alignment of Lphn1, 2 and 3 from mouse reveals that the three consecutive aspartate residues of the DDD-loop are also conserved in Lphn1, consistent with its ability to form 2:1:1 tetrameric complexes. However, in Lphn2, these consecutive aspartate residues are replaced by a consensus N-linked glycosylation site, which would mimic the DDD mutant discussed in this chapter, likely preventing FLRT/Unc5-induced dimerisa-

tion of Lphn2 (at least in the same manner observed for Lphn3). Interestingly, Lphn1 and 3 are predominantly expressed in the brain, whereas Lphn2 exhibits a broader expression pattern [142, 248, 164, 98, 28]. This suggests that FLRT/Unc5-induced Lphn dimerisation may be brain-specific.

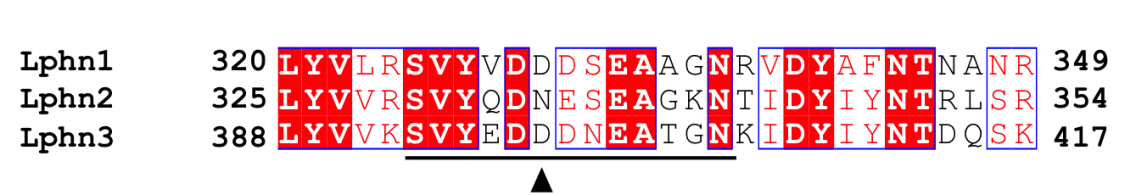


Figure 4.16: **The Lphn2 “DDD loop” contains a N-linked glycosylation site.** Sequences of Lphn 1, 2 and 3 from mouse were aligned using Clustal Omega. The sequence comprising the DDD-loop of Lphn3 is indicated by a black line below the alignment and the predicted glycosylated residue in Lphn2 indicated by a black arrowhead. The alignment was performed using Clustal Omega [235] and the figure was prepared using ESPript [211].

Although the data presented in this chapter give first indications of the molecular basis for potential functional cooperativity between Lphn, FLRT and Unc5, FLRTs are just one of the many neuronal ligands of Lphn. To follow on from the work presented here I decided to study the molecular basis for the interaction between Lphn and another of its binding partners, Teneurin. My work to achieve this is presented in the next chapter.

Chapter 5

Purification and Biophysical Characterisation of Teneurin and its Interaction with Lphn

5.1 Summary

In order to biophysically assess how additional Lphn binding partners, namely Teneurins, interact with Lphn (and the complexes it forms), it was necessary to first generate pure and well-behaved recombinant Teneurin ectodomain constructs. In this chapter I describe how I used bioinformatics to predict individual domains of Teneurin proteins, and then used these domain predictions to inform design of constructs for biophysical and, later, structural studies (see Chapter 6). After extensively screening many constructs of different Teneurin homologues and orthologues I was able to produce several well-behaved constructs of Ten2 from chicken. I then used this protein to confirm binding to Lphn and study the oligomerisation state of the protein. Finally, using SPR, I showed that Lphn can bind FLRT and Ten2 non-competitively, forming a triple complex, but that this complex was not capable of binding to Unc5D. This suggests that Ten2 may inhibit Lphn3-FLRT2-Unc5D super-complex formation.

5.2 Bioinformatic Analysis of Teneurin Proteins

The first step in producing recombinant protein for biophysical and structural studies is to understand the domain composition of the target protein. Constructs are then designed based on this knowledge. The ectodomain of Teneurin proteins are known to contain EGF repeats, a ~ 40 amino-acid Cys-rich motif [259], an NHL domain consisting of several NHL repeats, 26 YD (or Rhs) repeats, which are thought to form an enclosed barrel structure [169, 34], and a C-terminal Tox-GHH domain [290] (Figure 5.1). However, there are still large amounts of sequence for which no fold is yet predicted in the literature.

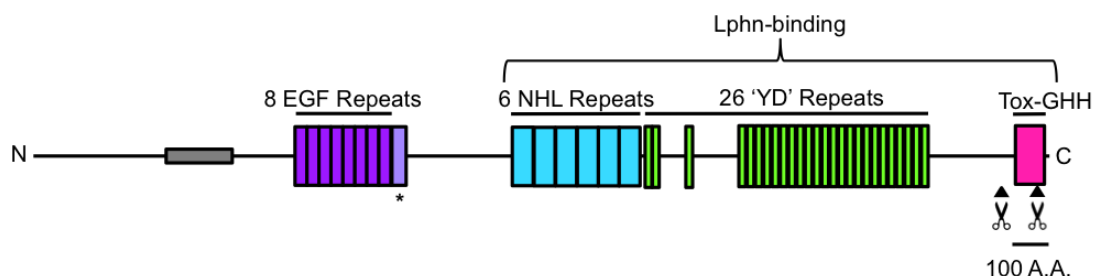


Figure 5.1: **Literature-based domain annotation of the Teneurin ectodomain.** Domains are coloured as follows: EGF repeats, purple; Cys-rich motif (starred), lilac; NHL domain, cyan; YD repeats, green; Tox-GHH, pink. The two C-terminal proteolytic cleavage sites are shown by scissor symbols.

To inform the design of Teneurin constructs for biophysics and crystallography, the primary amino acid sequences of all four Teneurin homologues from multiple species were subjected to thorough bioinformatic analysis. Due to the lack of homology between Teneurins and other vertebrate proteins, the most successful bioinformatic strategy for generating domain boundaries was homology modelling, using the online server SWISS-MODEL [22] to generate homology models based on distantly related proteins.

The most easily identifiable domains from the primary sequence are the EGF, NHL and YD domains. SWISS-MODEL readily identified the NHL domain as a six-bladed β -propellor (consistent with reports for other NHL domains [60]). The YD repeat domain was modelled based on the structures of two dimeric Rhs bacterial

toxins, YenB-YenC2 and TcdB2-TccC3 (discussed in Section 1.6.1). No convincing structural models were generated for the Tox-GHH domain, or indeed the entire region C-terminal of the YD repeat domain (Figure 5.3). The region between the Cys-rich motif and the NHL domain (~ 290 amino acids, indicated by a blue double-ended arrow in Figure 5.2) is completely unannotated in the literature. Secondary structure analysis using PSIPRED [33] predicted a strong β sheet propensity and a ~ 18 residue disordered region (Figure 5.2). Careful dissection of this region and homology modelling using SWISS-MODEL revealed homology to two distinct β sandwich-type domains (Figure 5.3 and Table 5.1).

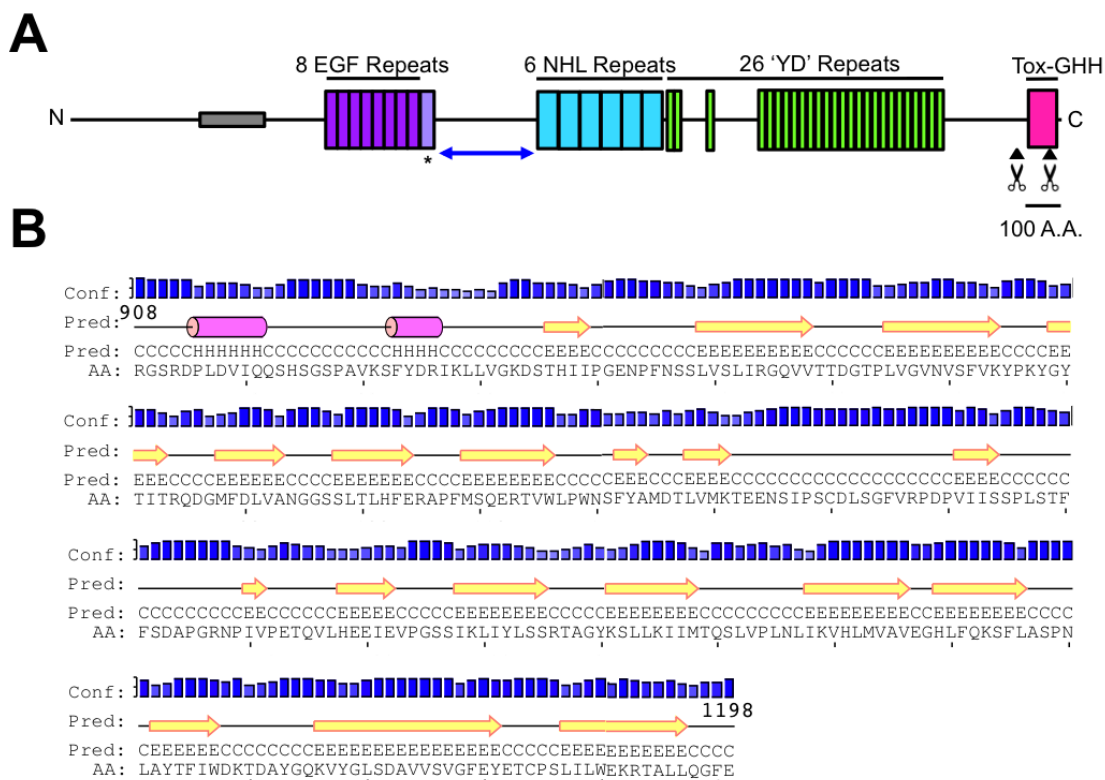


Figure 5.2: Secondary structure prediction for the Ten2 Cys-rich motif-NHL inter-domain region (residues 908-1198 in chicken Ten2 (GgTen2)). This region is marked by a blue double-ended arrow in (A), and was used for secondary structure prediction in PSIPRED [33]. This revealed a strong β -strand propensity with a central coil region of about 18 amino acids (B). The sequence displayed in this case is from GgTen2, but equivalent results were obtained for all other Teneurins tested. Yellow arrows indicate a predicted β -strand, pink cylinders indicate a predicted α -helix. Areas marked by a black line are predicted as coil. The blue bars indicate the confidence level of the prediction.

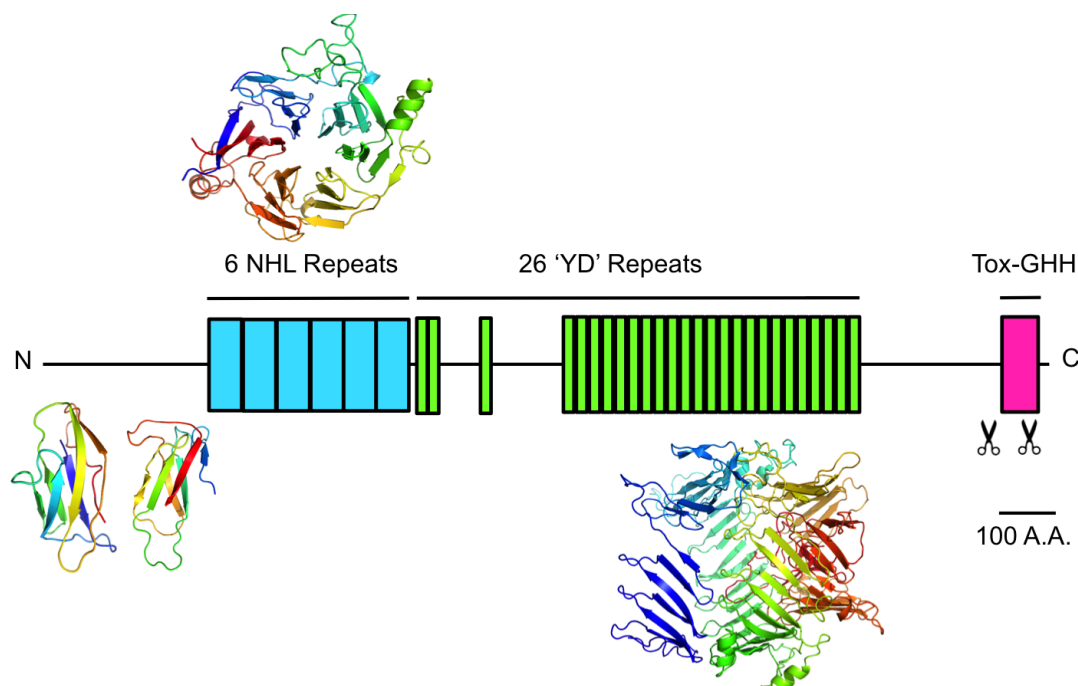


Figure 5.3: **Summary of homology models built for the Teneurin ectodomain.** The EGF domain is omitted as it is facile to predict its domain boundaries from the primary sequence. The model built for each region of sequence is displayed alongside the relevant part of the schematic. Models were generated using the SWISS-MODEL server and are coloured according to the rainbow with the N-terminus of the model coloured in blue and the C-terminus in red.

Domain	BLASTp Prediction	Model built	Template PDB Accession	Template PDB Name	Sequence Identity
β -Sandwich1	Not predicted	932 - 1039	3MN8	Carboxypeptidase D Transthyretin-like domain	15.5%
β -Sandwich2	Not predicted	1084 - 1173	3OSV	FlgD from <i>Xanthomonas campestris</i>	20.0%
NHL	1268 - 1626	1234 - 1602	3KYA	Putative phosphatase	15.4%
YD repeat	1710 - 2482	1629 - 2500	4O9X	TcdB2-TccC3 toxin	
ToxGHH	2720 - 2795	Not predicted	N/A	N/A	N/A

Table 5.1: **Domain boundaries predicted by BLASTp and homology modelling.** Residue numbers refer to GgTen2.

Homology modelling allowed assignment of structural domains to the majority of the ectodomain, although the fold of the most C-terminal ~ 300 amino acids remained mysterious. Alignment of these residues from HsTen1, 2, 3 and 4 indicated that a region about forty amino acids C-terminal of the predicted YD-repeat barrel-like fold is poorly conserved between these different homologues (Figure 5.4). The rest of the C-terminal region has very high levels of conservation between the homologues, indicating that it is likely to be functionally important and may be structured. This was confirmed by secondary structure prediction. To complicate matters, the

Teneurin C-terminal region contains two putative proteolytic cleavage sites; a furin site, ~ 110 residues from the C-terminus [259], and a pro-hormone cleavage motif ~ 50 residues from the C-terminus [206, 157].

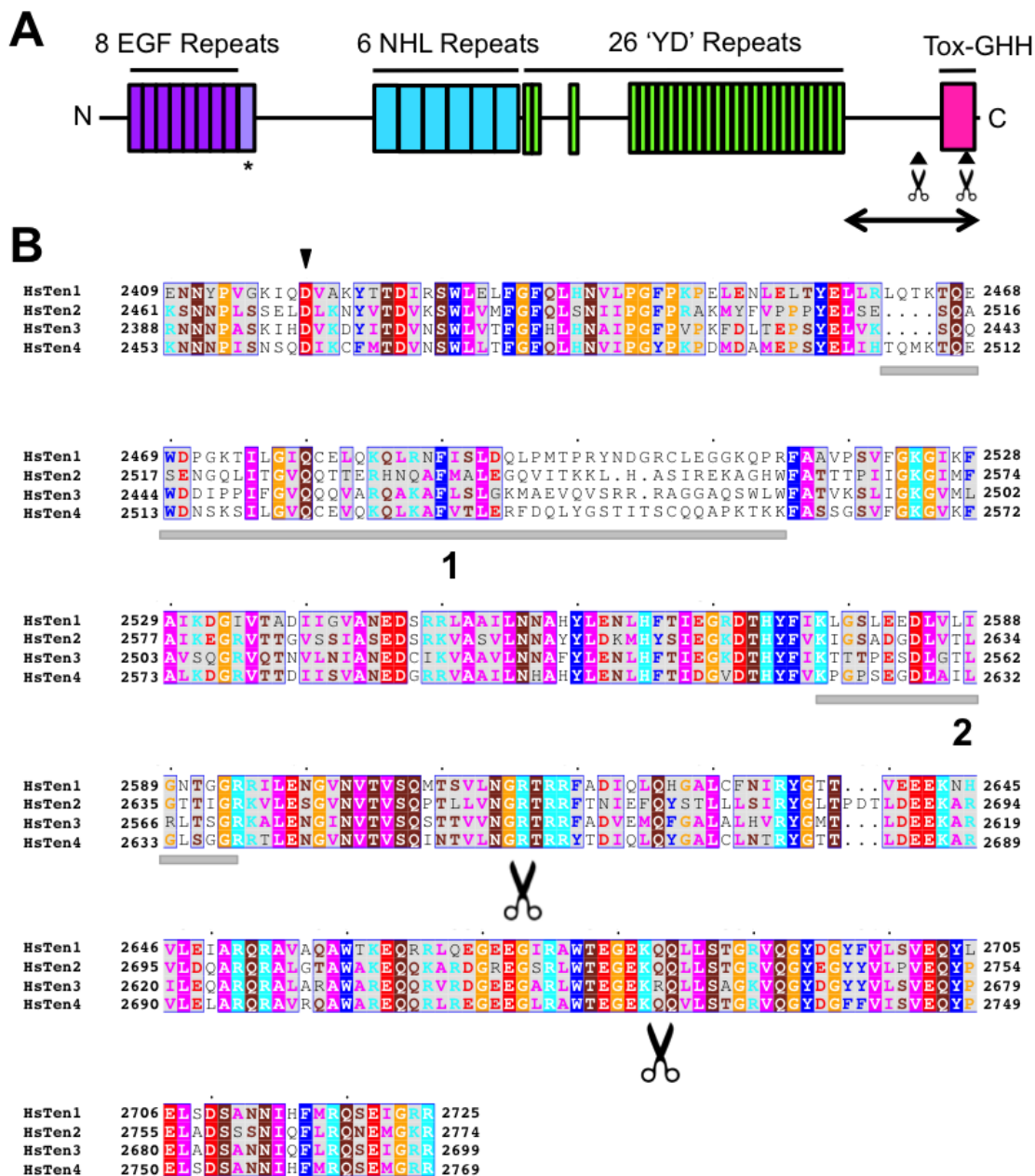


Figure 5.4: **Multiple sequence alignment of HsTen1, HsTen2, HsTen3, HsTen4 C-terminal region.** The residues referred to as the C-terminal region of Teneurin are indicated by the black double-ended arrow in (A). Colours and symbols are the same as for previous figures. These residues from all four human Teneurin homologues were aligned (B). Residues are coloured according to their physicochemical properties: HKR, cyan; DE, red; SNTQ, maroon; AVLIM, magenta; FYW, blue; PGC, orange. Stretches of residues with a global similarity score >0.7 are surrounded by blue frames. The black arrowhead indicates the end of the region modelled as part of the YD-repeat barrel fold by SWISS-MODEL. The grey bars indicate the unconserved regions following the barrel fold, with region 1 being significantly more poorly conserved than region 2. The two predicted protease recognition sites are indicated by scissor symbols. The alignment was performed using Clustal Omega [235] and the figure was prepared using ESPrpt [211].

5.3 Expression and Purification of Teneurin Constructs from HEK293 cells

Expression of large and post-translationally modified mammalian proteins such as Teneurin remains a considerable challenge in structural biology. Expression of secreted ectodomains in HEK293 cells has enabled the production and structural characterisation of some very large ectodomains, such as those of PlexinAs (~ 1200 amino acids) [123]. However, it is often still necessary to screen many different homologues and orthologues, as well as different constructs in order to produce recombinant protein of sufficient quality and quantity for structural studies. In the case of Teneurin, cDNA for an array of Teneurin homologues was obtained and screened for expression (Table 5.2).

Teneurin	Residues included	Source
MmTen1	804 - 2731	GeneArt TM Gene Synthesis
HsTen1	1 - 2725	I.M.A.G.E. Consortium
HsTen2	567 - 2774	I.M.A.G.E. Consortium
HsTen4	663 - 2769	I.M.A.G.E. Consortium
GgTen1	1 - 2705	Chiquet-Ehrismann Lab, Friedrich Miescher Institute, Basel [18]
GgTen2	1 - 2802	Chiquet-Ehrismann Lab, Friedrich Miescher Institute, Basel [18]

Table 5.2: **cDNAs used for expression testing of the Teneurin ectodomain.** Mm, *Mus musculus*; Hs, *Homo sapiens*; Gg, *Gallus gallus*.

Given the large size of the Teneurin cDNA, some clones obtained encoded incomplete ectodomains. For these homologues only constructs encoding fragments of the ectodomain were possible, most of which showed no secretory expression on a western blot. Changing the signal sequence, tag or tag position had no effect on secretion. For this reason I abandoned work on these fragment-type constructs and pursued the cDNAs which were sufficiently complete to create full ectodomain constructs (MmTen1, GgTen1, GgTen2). For this I designed three constructs (“Ecto” constructs) and assessed their secretory expression from HEK293T cells by western blot. These constructs included a FLAG tag at the N-terminus and a hexahistidine tag at the C-terminus, allowing me to detect the N and C-termini independently

in the event of any internal cleavage. I found that the GgTen2 homologue gave by far the best secretory expression, with both Ten1 orthologues producing only a weak band visible only after long exposure of the X-ray film (Figure 5.5B). Large-scale preparation by Ni-affinity chromatography and size-exclusion chromatography resulted in high yields of pure GgTen2^{Ecto} protein (Figure 5.5C, D).

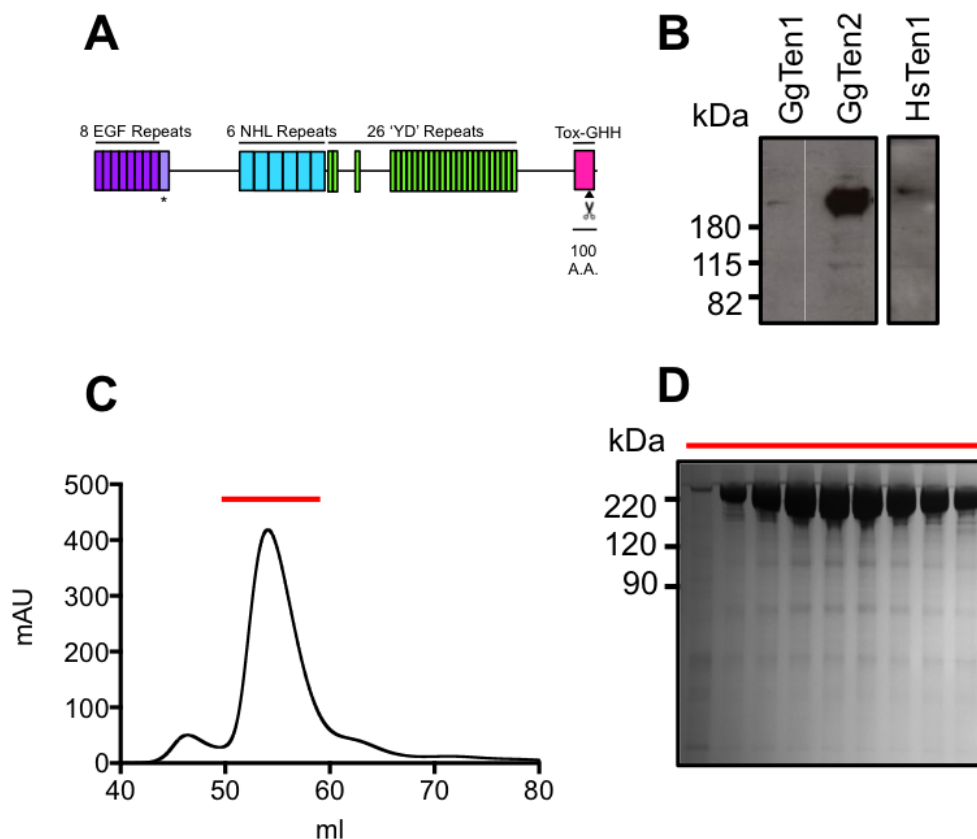


Figure 5.5: **Expression trials of Teneurin ectodomains and purification of GgTen2^{Ecto}.** Schematic showing the domains included in the three constructs. Colours are as in Figure 5.1 (A). Conditioned medium from HEK293T cells transiently transfected with each construct was loaded onto a western blot and probed with an anti-FLAG antibody. The GgTen2 ectodomain shows clear enhanced expression over the other two constructs (B). Elution profile of GgTen2^{Ecto} from an S200 16/60 gel filtration column (C). SDS-PAGE of the fractions from the gel filtration shown in C (D).

Given that the GgTen2^{Ecto} showed vastly superior expression to that of GgTen1 or HsTen1, I focussed my efforts on using this homologue to produce a construct of the Lphn-binding region of Teneurin that was suitable for crystallography (i.e. the minimal construct that would express and fold correctly and still bind to Lphn). The Lphn-binding region of Teneurin has been mapped to a region spanning from

the NHL domain to the end of the protein [237] and so this was the region targeted for crystallisation. I produced a variety of constructs aimed at expressing this region (Figure 5.6).

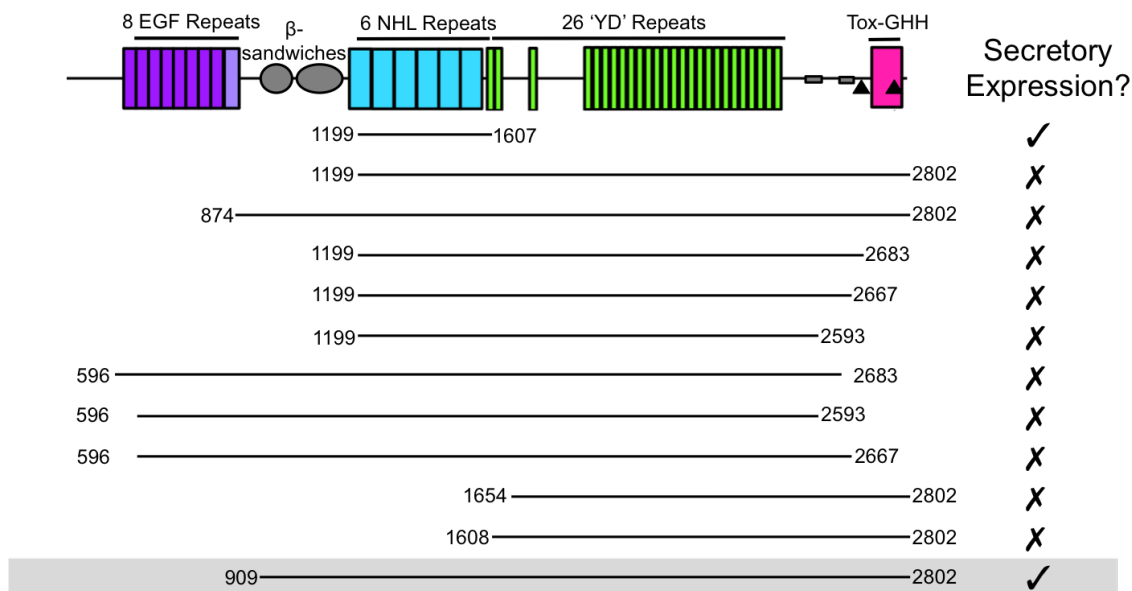


Figure 5.6: **Schematic depicting the initial constructs of the GgTen2 Lphn-binding region.** All domains are coloured as in previous figures and the two bioinformatically annotated β -sandwich domains are indicted by grey ovals. The two unconserved regions in the C-terminus are indicated by grey bars. The black arrowheads represent the two putative proteolytic sites. Domains included in each construct are indicated by black lines. The first and last residue of each construct is indicated by the number on either end of the black line. The presence of a correctly-sized protein band on a western blot of the cell culture medium is indicated by a tick mark. The construct highlighted in grey was carried forward for further optimisation.

Intriguingly, from the constructs generated, it seemed impossible to express the region C-terminal of the NHL domain without the β -sandwich region, suggesting that the β -sandwich region either helps the rest of the protein to fold, or stabilises it in some way. Truncation of the C-terminus was also problematic, with truncation at the three sites chosen (the two unconserved regions (Figure 5.4), and the predicted furin cleavage site) resulting in no secretory expression of the protein. The construct beginning with the first predicted β -sandwich domain and ending at the final residue of the protein (highlighted in grey) was carried forward for further optimisation for crystallisation.

To optimise this construct further, gradual truncations were performed from the N-

terminus. SWISS-MODEL models β -Sandwich1 as starting at residue 959, but the aforementioned construct begins with residue 909. Three truncations of increasing length were made from the N-terminus (Figure 5.7).

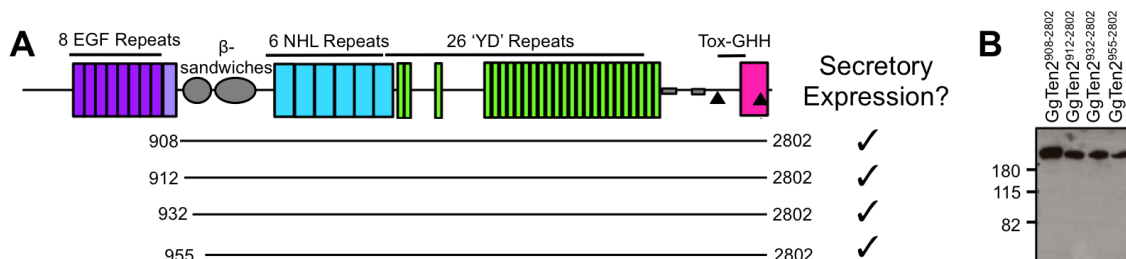


Figure 5.7: **Optimised constructs of GgTen2 Lphn-binding region.** Schematic indicating the domain composition and terminal residues of each N-terminally truncated construct. Symbols are as in Figure 5.6 (A). Anti-His western blot showing secreted protein of the correct size in the cell culture medium of cells transfected with each truncated construct (B).

Given that all three N-terminal truncations still resulted in the production of secreted protein, the shortest of the constructs (GgTen2⁹⁵⁵⁻²⁸⁰²) was carried forward for large-scale purification and biophysical and structural characterisation as it is likely to have the least flexibility. From herein this construct will be called GgTen2 ^{Δ EGF}. GgTen2 ^{Δ EGF} was purified by Ni-affinity chromatography followed by size-affinity chromatography and resulted in high yields of monodisperse and pure protein (Figure 5.8).

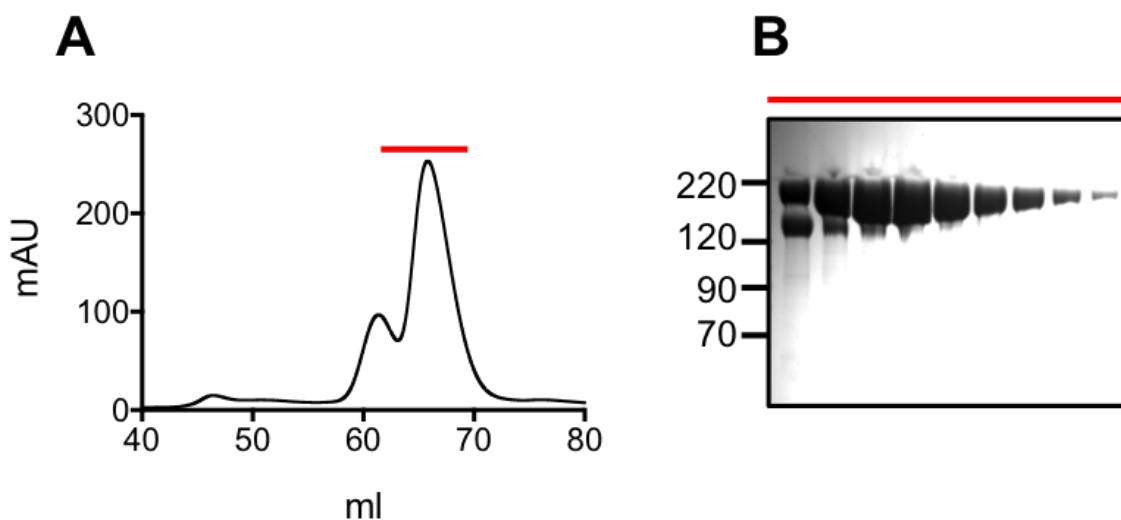


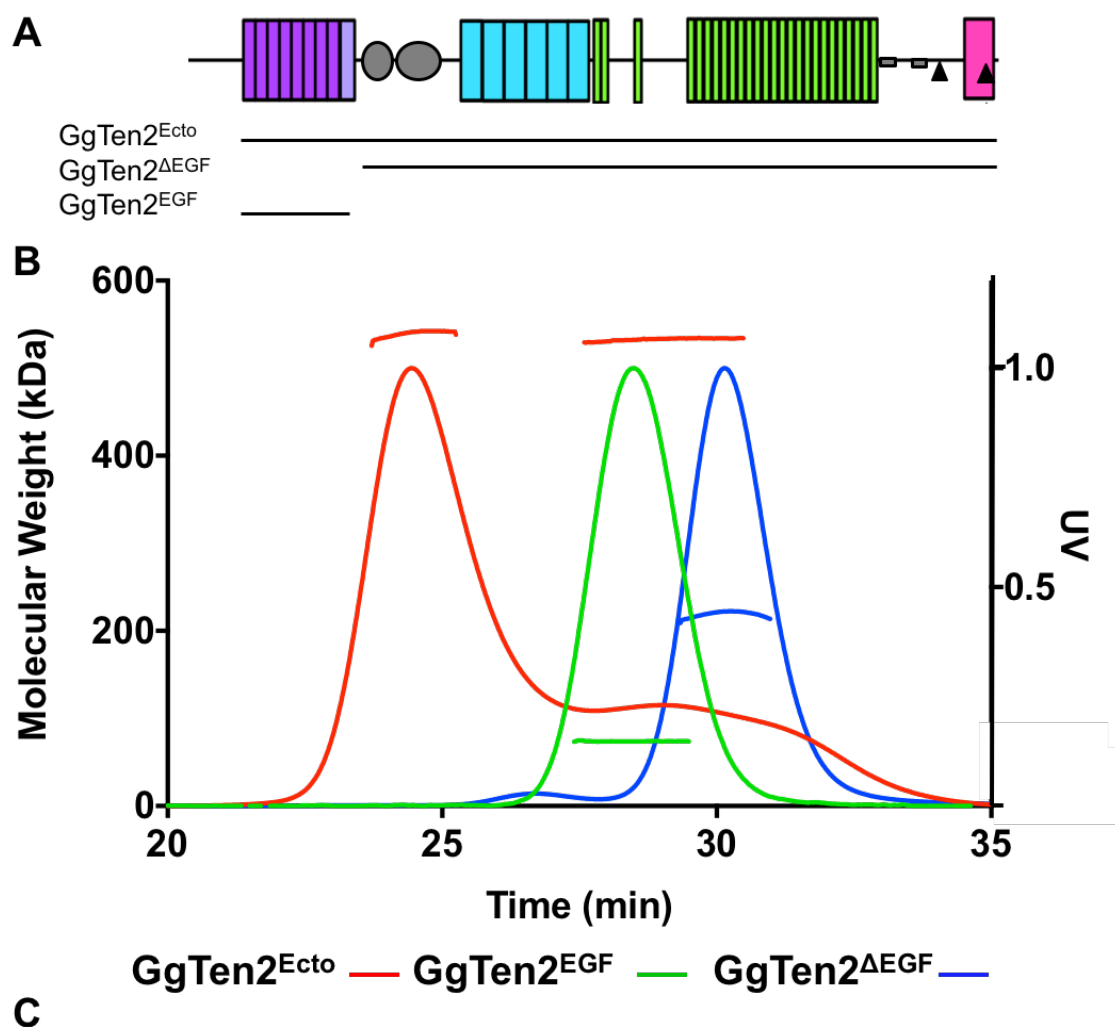
Figure 5.8: **Purification of GgTen2 ^{Δ EGF}.** Size-exclusion chromatography elution profile of GgTen2 ^{Δ EGF} on a Superdex 200 16/60 column (A). Indicated fractions from the gel filtration visualised by SDS-PAGE (B).

5.4 Functional Validation of Recombinant GgTen2 Proteins

Before using any protein for crystallisation experiments it is important to verify that it is functional (for example, it binds to one of its ligands or multimerises appropriately). The multimeric state of each construct was of interest, as it is well established in the literature that Teneurin proteins exist as disulphide-bonded dimers which form due to the uneven number of cysteine residues in EGF repeats 2 and 5 [190, 69].

5.4.1 SEC-MALS Analysis of Teneurin Multimerisation

Given the propensity for Teneurins to form disulphide-bonded dimers via their EGF domains [190, 69], I hypothesised that removal of the EGF domain in GgTen2^{ΔEGF} may monomerise the GgTen2 ectodomain. To validate this I performed a SEC-MALS experiment to compare the multimeric state of GgTen2^{Ecto} with that of GgTen2^{ΔEGF} and also a construct comprising the GgTen2 EGF domain and Cys-rich motif, GgTen2^{EGF} (Figure 5.9).



Construct	Theoretical Monomeric Molecular Weight (kDa)	Observed Molecular Weight (kDa)	Multimeric State
GgTen2 ^{Ecto}	276.2	539.5 ± 0.626 % 532.9 ± 0.575%	Dimer
GgTen2 ^{EGF}	36.7	73.9 ± 1.078 %	Dimer
GgTen2 ^{ΔEGF}	235.3	219.4 ± 0.662%	Monomer

Figure 5.9: **SEC-MALS analysis of GgTen2 dimerisation.** GgTen2^{Ecto} and GgTen2^{EGF} have molecular weights corresponding to dimers, whereas GgTen2^{ΔEGF} is monomeric (A, B). The samples were run a Superose6 10/300 analytical gel filtration column in 10 mM Tris (pH 7.5), 150 mM NaCl. Expected monomeric molecular weights were calculated including the expected glycosylation molecular weight for proteins produced in HEK293 cells treated with kifunensine (2.06 kDa) [38]. N-linked glycosylation sites were predicted using the NetNGlyc server. Calculations of theoretical monomeric molecular weights assumed 100% occupancy of predicted N-linked glycosylation sites (C).

This experiment suggests that the EGF domain of GgTen2 does indeed act as a dimerisation domain and that, in its absence, the protein is a monomer. Given these observations, and the overall aim to structurally characterise the Lphn-binding region of Teneurin, it was interesting to compare the binding of GgTen2^{Ecto} and GgTen2^{ΔEGF} to Lphn. This would also be important to demonstrate functionality of the constructs, and identify the Teneurin-binding domain of Lphn.

5.4.2 SPR Analysis of the Interaction of GgTen2 with Lphn

SPR experiments using Lphn constructs as ligand and the GgTen2 constructs as analyte were used to validate the functionality of the recombinant GgTen2 proteins and also to identify which domain of Lphn is primarily responsible for binding. Published evidence suggests that Teneurins bind primarily to the Lec-Olf region of Lphn [192]. One study indicates that the Lphn Lec domain is sufficient for Teneurin binding, but that the affinity is enhanced by the presence of the rest of the Lphn ectodomain [28]. SPR experiments using either Lphn2^{Lec} or Lphn2^{Lec-Olf} as ligand and GgTen2^{Ecto} as analyte revealed that these two Lphn constructs bind with very similar affinities, indicating that the Lec domain is the primary domain of Lphn responsible for Teneurin binding (Figure 5.10).

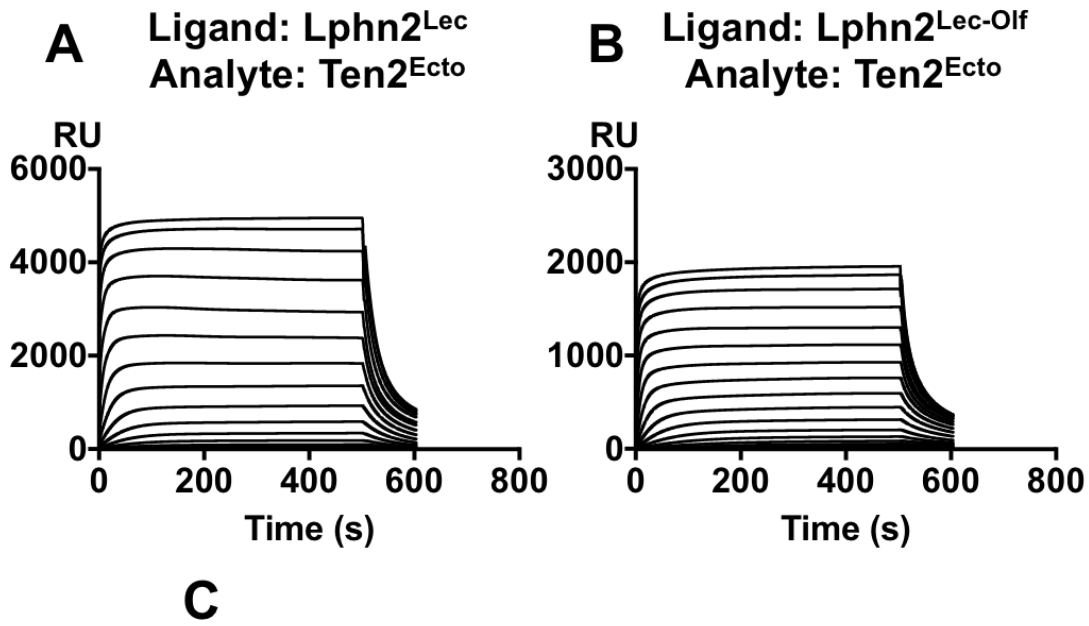
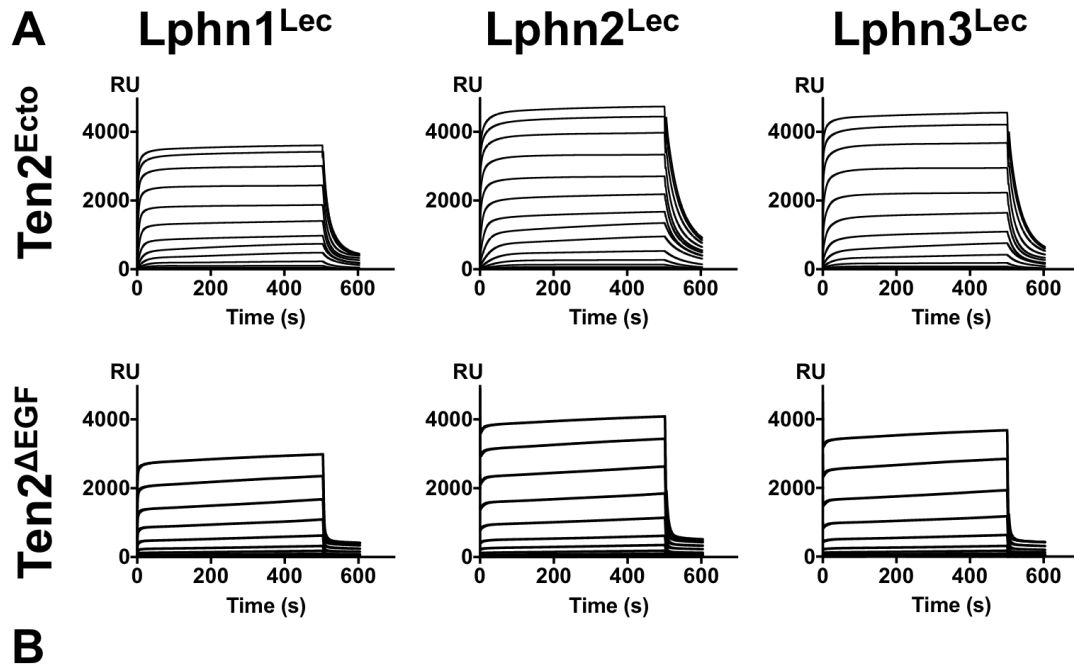


Figure 5.10: **GgTen2^{Ecto} binds to the Lphn2 Lec domain with nanomolar affinity.** Equilibrium surface plasmon resonance analysis of the binding of GgTen2^{Ecto} to Lphn2^{Lec} (A) and Lphn2^{Lec-Olf} (B). A 2-fold dilution series of the analyte was used for the experiment, with the top concentration being 11.2 μ M. The contact time was 500 s, followed by a 100 s dissociation time. K_D values were calculated assuming a 1:1 binding model. Given that GgTen^{Ecto} is dimeric, these K_D values may not reflect absolute values and are for internal comparison only (C).

Having established that GgTen2^{Ecto} interacts with the Lphn2^{Lec} with nanomolar affinity, it was interesting to see how the other Lphn homologues compared, and also how the deletion of the EGF domain affected Lphn binding. A similar experiment to the one above was performed with the Lec domains of mouse Lphn1/2/3 as ligand and either dimeric GgTen2^{Ecto} or monomeric GgTen2 ^{Δ EGF} as analyte (Figure 5.11).



Ligand	Analyte	K_D (μ M)	B_{max} (RU)	$B_{surface}$ (RU)
MmLphn1 ^{Lec}	GgTen2 ^{Ecto}	0.54	3682	822
MmLphn1 ^{Lec}	GgTen2 ^{ΔEGF}	5.4	3983	822
MmLphn2 ^{Lec}	GgTen2 ^{Ecto}	0.31	4511	1115
MmLphn2 ^{Lec}	GgTen2 ^{ΔEGF}	3.4	4819	1115
MmLphn3 ^{Lec}	GgTen2 ^{Ecto}	0.66	4633	1022
MmLphn3 ^{Lec}	GgTen2 ^{ΔEGF}	7.1	5092	1022

Figure 5.11: **Binding of GgTen2^{Ecto} and GgTen2^{ΔEGF} to Lphn.** SPR experiments showing that GgTen2^{Ecto} binds to the Lec domains of all three murine Lphn homologues with high nanomolar affinity. Experiment parameters were the same as those if Figure 5.10 (A). Lphn2^{Lec} binds most strongly, followed by Lphn1^{Lec} and then Lphn3^{Lec}. Removal of the EGF domain, monomerising GgTen2 (see Figure 5.9) results in 10-fold weaker binding to each of the Lec domains. K_D values were calculated assuming a 1:1 binding model (B).

This experiment revealed that all three Lphn Lec domains bind GgTen2^{Ecto} with a similar nanomolar affinity, although GgTen2 binds most tightly to the Lec domain of Lphn2. Interestingly the removal of the EGF domain weakens the interaction by about 10-fold compared to the interaction with the whole ectodomain. A test injection of GgTen2^{EGF} showed that it does not directly bind to Lphn (data not shown). The tighter binding of GgTen2^{Ecto} compared to GgTen2^{ΔEGF} could reflect an avidity effect caused by dimerisation of the ectodomain (see Figure 5.9).

5.4.3 GgTen2 Interactions with the Lphn-FLRT-Unc5 Super-Complex

Given the nanomolar-affinity interaction between Teneurin and the Lphn Lec domain, it was interesting to assess if Teneurin might form larger complexes also involving FLRT and Unc5. Given that Teneurin interacts with the Lphn Lec domain, and FLRT with the Lphn Olf domain, it seemed likely that these proteins might be able bind Lphn at the same time. To test if this was indeed the case I performed an SPR experiment where FLRT2^{Ecto} was used as ligand and an equimolar mixture of Lphn1^{Lec-Olf}+GgTen2^{Ecto} used as analyte. Both proteins were also used individually as control analytes.

The results of this experiment showed that FLRT2^{Ecto} and GgTen2^{Ecto} bind non-competitively to Lphn1^{Lec-Olf} as an equimolar mixture of Lphn1^{Lec-Olf} and GgTen2^{Ecto} displays about twice response of Lphn1^{Lec-Olf} alone at the top concentration used. GgTen2^{Ecto} showed no observable binding to FLRT2^{Ecto} (Figure 5.12). Equivalent results were also obtained for FLRT3^{Ecto}.

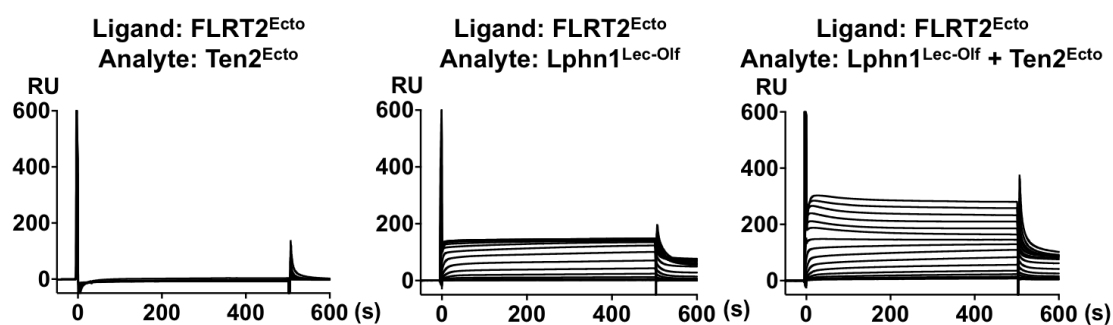


Figure 5.12: **FLRT and Teneurin bind non-competitively to Lphn.** Biotinylated FLRT2^{Ecto} (590 RU) was immobilised on a streptavidin-coated CM5 Biacore chip. The analytes used were GgTen2^{Ecto} alone, MmLphn1^{Lec-Olf} alone and a 1:1 molar mixture of MmLphn1^{Lec-Olf}+GgTen2^{Ecto} (all 10 μ M top concentration). The complex was prepared such that the individual components were both present at a concentration of 10 μ M and so results are comparable between all three sensorgrams.

In the knowledge that Lphn, FLRT and Ten2 ectodomain fragments can form a triple complex, I decided to see how the Teneurin-Lphn interaction would affect formation of the Lphn3-FLRT2-Unc5D super-complex described in Chapter 4. To

test this, I performed an SPR experiment using a similar set-up to that discussed in Figure 3.8, where $\text{Unc5D}^{\text{Ecto}}$ was used as ligand and different complexes and individual control proteins used as analyte (Figure 5.13).

This experiment revealed no detectable interaction between $\text{Unc5D}^{\text{Ecto}}$ and $\text{GgTen2}^{\text{Ecto}}$. $\text{FLRT2}^{\text{Ecto}}$ alone displayed ~ 320 RU binding to $\text{Unc5D}^{\text{Ecto}}$. Similarly to the experiment described in Figure 3.8, a 2:1 mixture of $\text{FLRT2}^{\text{Ecto}} + \text{Lphn3}^{\text{Lec-Olf}}$ showed an approximately two to three times greater signal than $\text{FLRT2}^{\text{Ecto}}$ alone. When a mixture of $\text{GgTen2}^{\text{Ecto}} + \text{FLRT2}^{\text{Ecto}} + \text{Lphn3}^{\text{Lec-Olf}}$ was used as analyte, the response was similar to when $\text{GgTen2}^{\text{Ecto}}$ was omitted. This experiment indicates that formation of the Lphn3-FLRT2-Unc5D complex may preclude $\text{GgTen2}^{\text{Ecto}}$ binding.

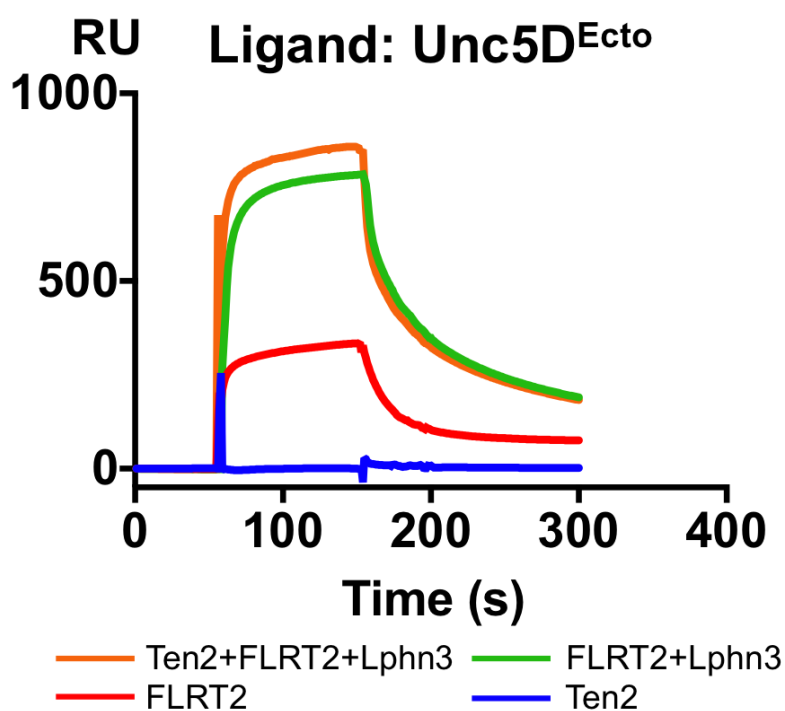


Figure 5.13: **Formation of the Lphn3-FLRT2-Unc5D super-complex precludes Teneurin binding.** Biotinylated $\text{Unc5D}^{\text{Ecto}}$ (630 RU) was immobilised on a streptavidin-coated CM5 Biacore chip. Analytes used were: $\text{FLRT2}^{\text{Ecto}}$ ($5 \mu\text{M}$); $\text{GgTen}^{\text{Ecto}}$ ($5 \mu\text{M}$); $\text{FLRT2}^{\text{Ecto}} + \text{Lphn3}^{\text{Lec-Olf}}$ ($5 \mu\text{M} + 10 \mu\text{M}$); $\text{FLRT2}^{\text{Ecto}} + \text{Lphn3}^{\text{Lec-Olf}} + \text{GgTen2}^{\text{Ecto}}$ ($5 \mu\text{M} + 10 \mu\text{M} + 5 \mu\text{M}$). The experiment was carried out in single-injection mode and so no K_D values were calculated. Each analyte was injected for 100 s, followed by a 100 s wait period and a 30 s regeneration with 2 M MgCl_2 .

5.5 Discussion

Thorough bioinformatic analyses identified the likely domain composition for the majority of the Teneurin ectodomain. Careful understanding of the domain architecture (in particular the region between the cysteine-rich motif and the NHL domain containing β -sandwich domains 1 and 2) was key to producing the GgTen2 ^{Δ EGF} construct. It was only possible to identify these two domains based on homology modelling, presumably due to their poor sequence identity with other structurally-related domains. Intriguingly it seemed impossible to express the region C-terminal of the NHL domain (amino acids 1607-2802 in GgTen2), without the β -sandwich region, suggesting that these domains could be important in allowing the rest of the protein to fold correctly.

When attempting to recombinantly express the Teneurin ectodomain, a variety of orthologues were assessed for expression by Western blot. The only homologue tested with expression levels suitable for biophysical and structural analysis was GgTen2 (from chicken). The ectodomain of Teneurin2 has been shown to exist in a soluble form [267], which might facilitate its expression as a soluble recombinant ectodomain.

A particularly interesting finding from this chapter is how the multimerisation state of GgTen2 affects its binding to Lphn. The data presented here and elsewhere suggest that Teneurins form disulphide bonded dimers *in cis* using unpaired cysteines in EGF repeats 2 and 5 [190, 69]. Receptor dimerisation, which is often ligand-induced, has emerged as a signalling mechanism for several receptor classes [261]. However, there are several other receptors which exist as constitutive covalently-linked dimers, similar to Teneurin, whose modes of activation are significantly less-well studied. One example of particular interest is the Discoidin Domain Receptor (DDR) family, a family of receptor tyrosine kinases which use their N-terminal discoidin domain to bind triple-helix collagen, leading to receptor activation. Enzyme-linked immunosorbent assays have demonstrated that the DDR discoidin domain

recognises collagen as a dimer, but not a monomer, and that the DDR1 discoidin domain alone has around a 10-fold lower affinity for collagen than a construct of the whole ectodomain [141, 2]. This was later revealed to be due the disulphide bond-mediated dimerisation of the ectodomain by its membrane-proximal “stalk” region [1]. This is analogous to the data presented in this chapter for Teneurin and Lphn, whereby the Teneurin ectodomain comprises separate Lphn-binding and dimerisation domains and, although the dimerisation domain does not directly bind the ligand, its presence enhances the affinity of the interaction, presumably due to avidity effects. This may allow the receptor to be in a constitutively “primed” state, facilitating ligand recognition, or may be important for a signalling whereby ligand binding causes a shift of the constrained subunits relative to one another.

The final aspect of this chapter was the interaction of Teneurin with the other complexes discussed in this thesis. Consistent with earlier reports [28], I found that the Lphn Lec domain was sufficient for GgTen2^{Ecto} binding. Given that the Lphn Olf domain is sufficient for FLRT-binding, it seemed likely that the two might bind non-competitively to Lphn, which was indeed the case. However, in the SPR experiment shown in Figure 5.13, GgTen2^{Ecto} seemed incapable of interacting with Lphn3-FLRT2-Unc5D super-complex discussed in the previous chapter. This could be because the Lec domain (at least of Lphn3A) is enclosed within the centre of the super-complex. Perhaps super-complex formation sterically prevents Teneurin binding, or perhaps the SPR set-up used is not compatible with Teneurin binding. Further experiments, for example native mass spectrometry, will be necessary to conclusively determine whether Teneurin can indeed bind to the super-complex.

The data presented in this chapter improve our understanding of Teneurin proteins and their interactions with Lphn on a biophysical level, however these mysterious proteins are entirely uncharacterised on a structural level. The GgTen2^{ΔEGF} construct generated in this chapter is a good candidate for structure determination by

X-ray crystallography due to its monomeric nature, monodispersity, good yield and ability to bind to Lphn. These structure-determination studies form the basis of the next chapter.

Chapter 6

Solving the Crystal Structure of Teneurin

6.1 Summary

There is no known available structural information for the Teneurin ectodomain, impeding our understanding of this complex and multifunctional receptor. Having produced several well-behaved constructs of GgTen2, I decided to perform crystallisation trials, with the ultimate aim of solving the first crystal structure of the Teneurin ectodomain. I obtained crystals of the GgTen2^{ΔEGF} construct described in Chapter 5, which I optimised extensively to produce a 2.5 Å native dataset and a 3.3 Å Pt-derivative dataset. Using the phases calculated from the Pt-derivative dataset, and the native dataset for all-atom refinement I solved the structure of GgTen2^{ΔEGF}. The structure reveals the complex architecture of the Teneurin ectodomain, which is a chimera of domains typically found in vertebrate cell-surface receptors and domains found in bacterial proteins. Here I discuss the fold of each domain in detail and also identify potential functional surfaces on GgTen2^{ΔEGF}.

6.2 Crystallisation and Data Collection for GgTen2^{ΔEGF}

6.2.1 Crystallisation

For crystallisation GgTen2^{ΔEGF} was produced in kifunensine-treated HEK293T cells. Protein was purified as described previously (Section 2.2) and the purest fractions

from gel filtration pooled and mixed with endoglycosidase F1 in a 1:1000 (endoglycosidase F1:GgTen2^{ΔEGF}) ratio. The protein was concentrated to 4.7 mg/ml in a buffer containing 20 mM Tris (pH 7.5), 200 mM NaCl, 200 mM non-detergent sulfobetaine-256. Higher concentrations of protein resulted in precipitation in over 60% of crystallisation drops. A number of broad screens were set up, including JCSGI-IV (QIAGEN), Morpheus, PACT, PGA, and Index (all Molecular Dimensions) at both 18 °C and 4 °C. All screens were performed using 1:1 protein:reservoir drop ratios. Initial crystals were obtained in the Morpheus screen after 14 days at 18 °C in a condition containing 0.1 M Tris (base)/Bicine (pH 8.5), 20% v/v glycerol, 10% w/v PEG4K, 30 mM di-ethyleneglycol, 30 mM tri-ethyleneglycol, 30 mM tetraethyleneglycol, and 30 mM pentaethyleneglycol (Figure 6.1A).

Initial crystals were highly multiple and diffracted only very weakly. The process of crystal optimisation was lengthy. Additive screens such as Silver Bullets and Hampton Additives (both Hampton Research) resulted in only small increases in the diffraction potential of the crystals. The most productive optimisations were the result of diluting the original reservoir solution with water and micro-seeding with varying concentrations of seeds. However, the most significant improvement to the diffraction potential of the crystals came from omitting the endoglycosidase F1 prior to the concentration step. The combination of these procedures resulted in crystals that were much larger, more single and diffracted to around 2.5 Å resolution (Figure 6.1B).

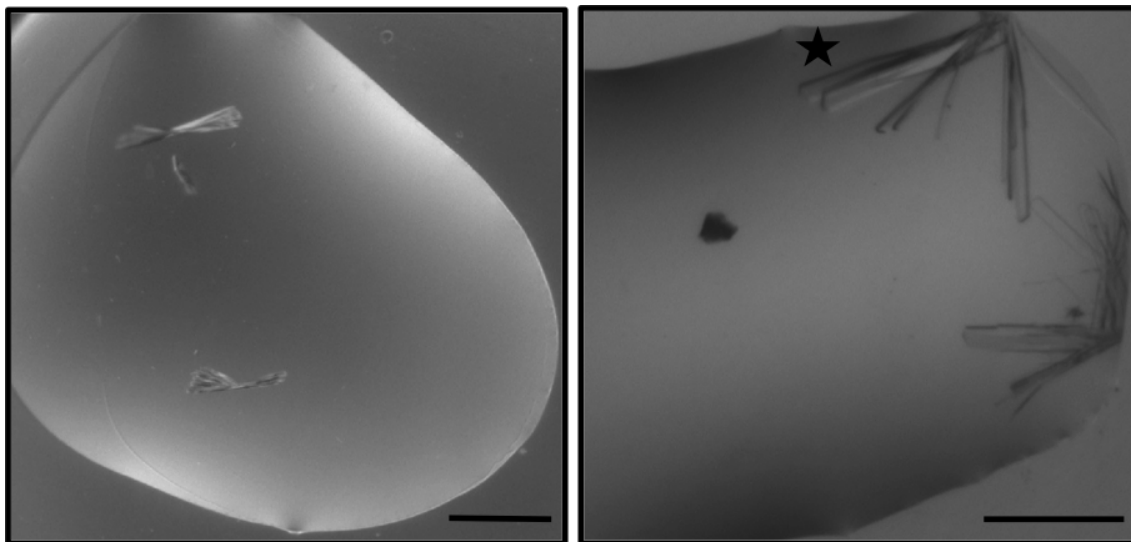


Figure 6.1: **Crystals of GgTen2 Δ EGF**. Initial (left) and optimised (right) crystals of GgTen2 Δ EGF. These optimised crystals diffracted to ~ 2.5 Å. Scale bar=200 μ m.

Despite the optimisation procedure, the optimised crystals still grew in clusters at the edges of the drops. This necessitated that, during harvesting, crystal clusters were carefully dissected in order to liberate the largest and most single crystals. To facilitate this the drop was expanded by adding ~ 2 μ l reservoir solution. In cases where the glycerol content of the reservoir was significantly below 20%, additional glycerol was added as cryoprotectant. Crystals were then harvested directly from this expanded drop and cryo-cooled in liquid nitrogen prior to their trip to the synchrotron.

Native Data Collection

During the crystal optimisation process, various crystals were screened on beamlines at Diamond Light Source and at the ESRF. The highest quality native data used for refinement were collected on beamline I03 at Diamond Light Source. The best crystals diffracted to around 2.5 Å resolution (Figure 6.2A). From the diffraction images it was evident that the reciprocal lattice spacing of the crystals was very small (indicating a very large unit cell). For this reason, the data was collected with the detector as close as possible to the crystal (to capture the high resolution reflections), whilst still allowing adequate separation of the reciprocal lattice points

(RLPs) in the direction of the long cell axis (Figure 6.2B). The crystals were also radiation sensitive, and thus a strategy was devised whereby successive 90° wedges of data were collected from different areas on three of the best diffracting crystals (Table 6.1). These data were merged together to yield a complete dataset with minimal radiation damage. Native data processing is discussed in Section 6.2.4.

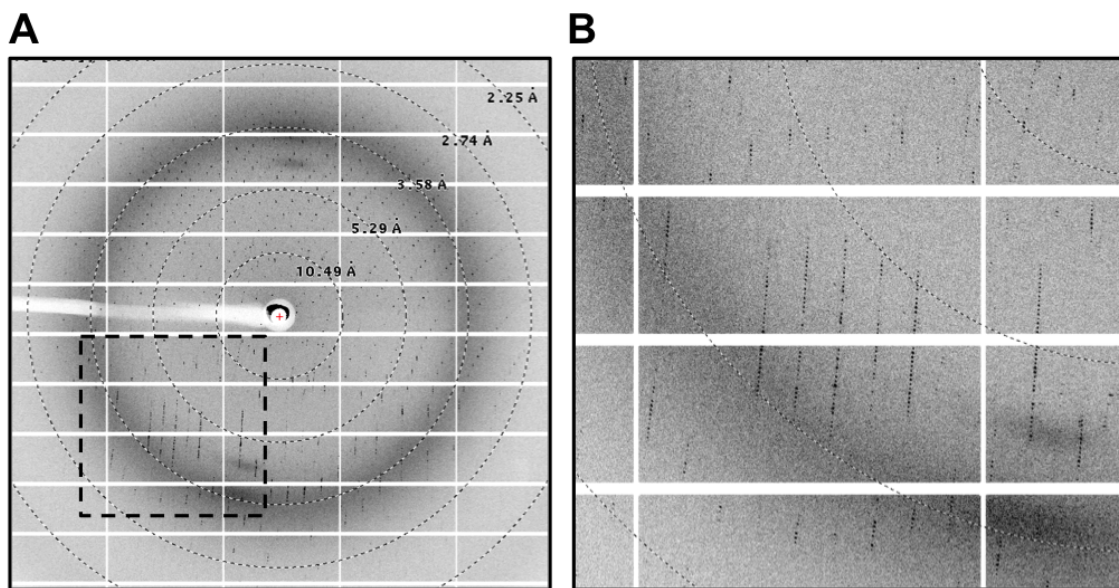


Figure 6.2: **Diffraction of GgTen2 Δ EGF Crystals.** A GgTen2 Δ EGF diffraction image from beamline I03 at Diamond Light Source with resolution rings displayed (A). A magnified view of the region of the diffraction image indicated by the dashed box is shown, demonstrating the small separation of the RLPs in the direction of the long unit cell axis (B).

Beamline	i03
Wavelength (Å)	0.98
Beam size (μ m)	20x20
Oscillation (°)	0.2
Exposure (s)	0.3
Transmission (%)	100
Number of Images	450

Table 6.1: **Experiment parameters for native data collection from GgTen2 Δ EGF crystals.**

6.2.2 Experimental Phasing of GgTen2 Δ EGF

Molecular replacement requires a model of a closely-related structural homologue of the protein in the crystal. There are structures available of proteins distantly related to Teneurins (for example the NHL domain of the *Drosophila* RNA bind-

ing protein Brat (PDB accession: 4ZLR) which is 19% identical [155]), but these models have low sequence identity (<25%) and represent only a low proportion of the total scattering mass in the crystal, making them unlikely candidates for successful molecular replacement. For this reason, experimental phasing attempts were initiated. A Single Anomalous Dispersion (SAD) experiment was the first choice as this allows phase information to be calculated from a single dataset obtained from a single crystal. A SAD experiment requires a crystal derivatised with heavy atoms. At certain wavelengths heavy atoms absorb X-rays, resulting in the break down of Friedel's law (F_{hkl} and $F_{-h -k -l}$ are no longer equal in intensity). The differences in intensity of Friedel pairs (the anomalous signal) can be used to calculate initial phases.

Derivatisation of GgTen2^{ΔEGF} crystals

Production of a good derivative crystal is key for a successful SAD experiment. Soaking experiments were carried out using multiple compounds of four different heavy metals: platinum (potassium tetrachloroplatinate (II), potassium tetracyanoplatinate (II) hydrate, potassium hexachloroplatinate (IV)); gold (gold (I) potassium cyanide, potassium tetrachloroaurate (II) hydrate, sodium tetrachloroaurate (III) dihydrate); tantalum (hexatantalum tetradecabromide) and mercury (mercury (II) acetate).

Derivative Data Collection

In total, around fifty crystals of GgTen2^{ΔEGF} were soaked with heavy metal compounds and harvested for anomalous data collection. Anomalous data collection was performed on the tuneable beamline, ID29, at the European Synchrotron Radiation Facility (ESRF). To maximise anomalous signal, SAD data collection was carried out above an element-specific energy known as the the X-ray absorption edge. A fluorescence scan was used to locate the absorption edge for each crystal. Once the edge was located, the beam was tuned such that the beam energy was slightly above

the edge energy. Most of the derivatised crystals diffracted poorly or not at all. However, two crystals soaked with potassium tetrachloroplatinate (II) hydrate for 21 hours (and then briefly back-soaked) exhibited a sharp edge in the fluorescence scan (indicating that Pt was bound, Figure 6.3) and diffracted to ~ 3.3 Å resolution. 360° of data was collected from each of these crystals (Figure 6.3).

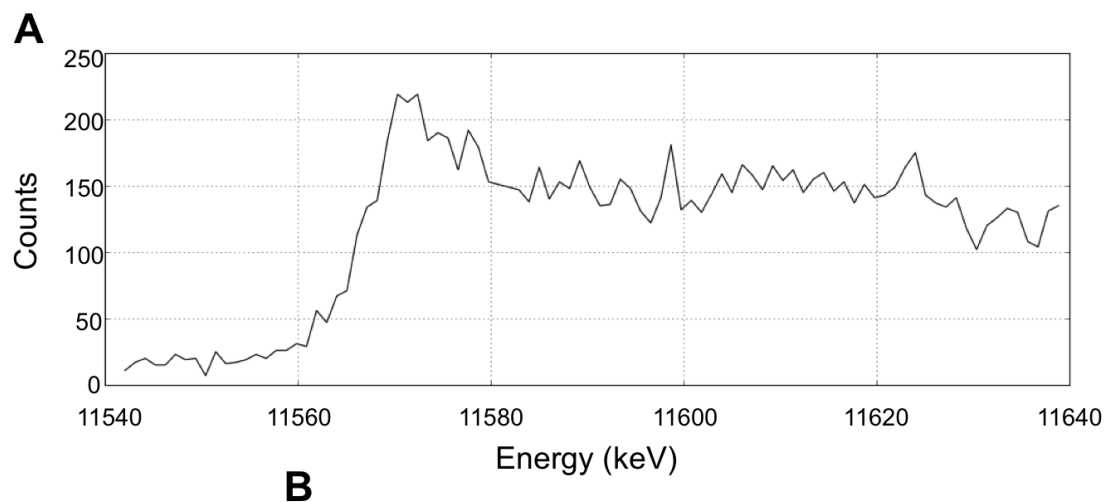


Figure 6.3: **GgTen2^{ΔEGF} derivative data collection.** Fluorescence scan of a GgTen2^{ΔEGF} crystal soaked with potassium tetrachloroplatinate (II) (A). Experiment parameters for data collection from the two Pt-derivative crystals (B).

Derivative Data Processing

The Xia2 pipeline was used to process the derivative data. Data from both derivative crystals were indexed and integrated using XDS to produce intensities. XDS identified the space group as C222₁, with refined cell parameters $a=83.32$, $b=156.13$, $c=452.95$, $\alpha=\beta=\gamma=90^\circ$. Data from the two crystals were isomorphous and so were merged, multiple observations of individual reflections were also merged. Bi-

jovet pairs remained unmerged for the detection of anomalous differences for phase calculations. Scaling was performed using XSCALE and intensities converted to structure factors using the program Truncate. The presence of an anomalous signal was assessed using the anomalous correlation (Anomal Corr) and SigAno statistics generated by XSCALE. Both of these statistics can be seen plotted by resolution in Figure 6.4, and indicate that a significant anomalous signal is present to about 5 Å resolution.

The overall resolution limit of the dataset was set to 3.26 Å where $CC_{1/2}=25.4\%$ in the highest resolution shell. The data reduction statistics are summarised in Table 6.2. The Matthew's coefficient indicated a solvent content of 65.4% with just one copy of GgTen2^{ΔEGF} in the asymmetric unit.

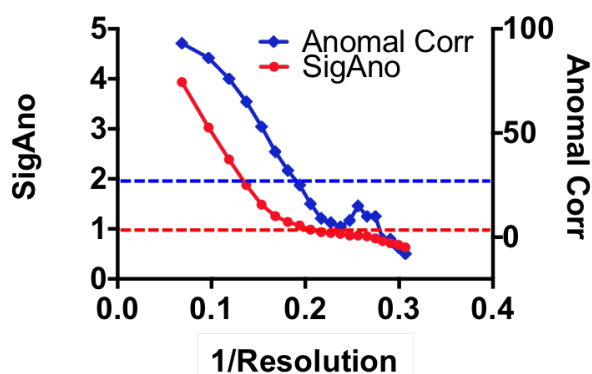


Figure 6.4: Anomalous signal plotted against resolution for GgTen2^{ΔEGF} crystals. SigAno and the anomalous correlation are plotted against the reciprocal of the resolution. The dashed lines represent the typical cut-off values below which no significant anomalous signal is considered to be present.

Space group	C222 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	83.32, 156.13, 452.95
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	113.24-3.26 (3.34-3.26)
R_{meas} (%)	34.3 (156.6)
CC_{1/2} (%)	99.3 (25.4)
<i>I</i>/σ<i>I</i>	7.17 (0.81)
Completeness (%)	99.9 (99.3)
Multiplicity	12.2 (3.6)

Table 6.2: **Data reduction statistics for GgTen2^{ΔEGF} Pt-derivative crystals.** Values for the highest resolution shell are given in parentheses.

6.2.3 Solving the GgTen2^{ΔEGF} Structure

Initially the Crank2 pipeline was used to perform substructure detection, obtain initial phases, carry out density modification and autobuild into the modified density. This allowed the placement of the best-ordered secondary structure elements (Figure 6.5A). The structure built by Crank2 was then used in Phaser to re-calculate the phases using a combined molecular-replacement-SAD (MR-SAD) approach. The resultant maps were solvent-flattened using PARROT and fed to Buccaneer for autobuilding. Buccaneer built a large, non-continuous, barrel-like structure which resembled the B-C component of bacterial ABC toxin complexes [34, 167]. This model was manually corrected in Coot, reversing the strand direction of incorrect β -strands and connecting the N- and C-termini of continuous regions. This model was then used to re-calculate the phases using Phaser MR-SAD to further improve the maps.

Building the NHL domain was more tricky than building the barrel. The NHL domain has distant structurally characterised homologues and thus is a candidate for MR, albeit a difficult one. HHPred was used to identify structurally characterised homologues. Superposition of a selection of these homologues identified a core region of the β -propellor that is highly structurally conserved. Based on this superposition, the search models were trimmed to include just this core region. The

side chains were also truncated to alanine residues. For MR, the barrel structure was first placed, followed by the search for the NHL domain. One of the homologues, a protein of unknown function from the bacterium *Bacteroides thetaiotaomicron* (PDB accession: 3HRP), was placed with convincing TFZ scores (TFZ=14.0) and its N- and C- placed close to the termini of the barrel (Figure 6.5B).

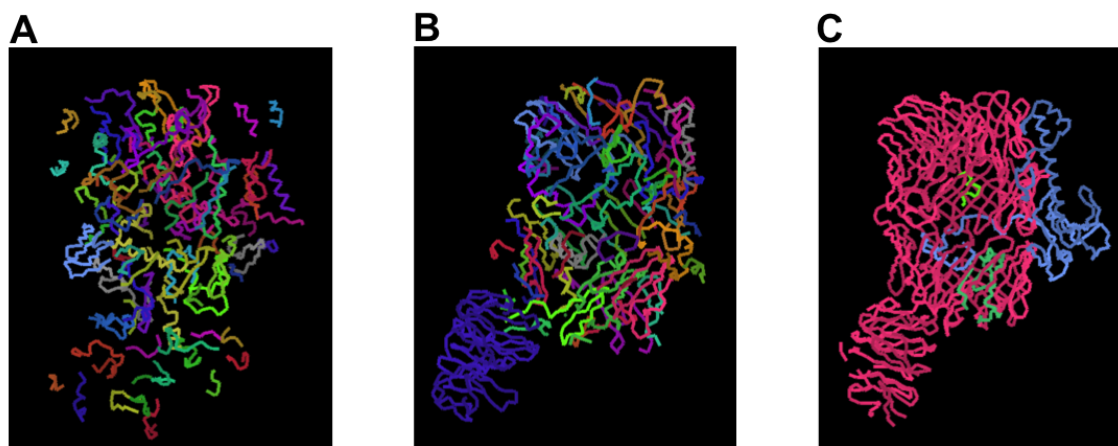


Figure 6.5: **Initial model building of GgTen2 Δ EGF**. Some secondary structure elements were initially built by the Crank2 pipeline (A). Manual re-building was used to improve the model for the barrel region and the NHL domain was located using a trimmed polyalanine model of a distantly-related homologue (B). Iterative rounds of model improvement and phase recalculation using Phaser MR-SAD were used to build a continuous chain for the NHL and barrel domains (C). The figure was prepared using Coot [63].

After manual editing of the improved model, recalculation of the phases by MR-SAD and solvent-flattening using Parrot, the maps were of sufficient quality to begin assigning sequence to some parts of the model. Initial sequence-assignment was informed using the homology models shown in Figure 5.3 and validated by the high frequency of aromatic residues in the YD repeats, alongside the knowledge that the platinum atoms used for derivatisation typically bind to either histidine or methionine residues. The presence of densities for N-linked glycosylations were also used to infer the positions of asparagine residues. Further rounds of phase-recalculation by MR-SAD followed by density modification allowed the NHL and barrel domains to be joined as a single chain (Figure 6.5C). Following this I was able to manually build β -sandwich domain 2. At this stage, there was only very poor density visible for β -sandwich domain 1. The directionality and register of the

region C-terminal of the barrel remained unclear, although it was possible to model several secondary structure components.

6.2.4 Native Data Processing, Molecular Replacement and Refinement

Once the model built based on the experimental phases was deemed sufficiently accurate (R_{work} and R_{free} were both below 40%), it became possible to use the phases calculated from this model to phase the native data using MR. As the native crystals diffracted to significantly higher resolution than the derivative crystals (2.5 Å vs. 3.3 Å), there was significantly more high resolution information present in the native dataset which would be highly important for final sequence assignment and refinement of the structure.

Native data processing was performed similarly to the derivative data. Six 90° datasets collected from three separate native crystals were identified as being isomorphous with each other. These six datasets were carefully examined for evidence of radiation damage (indicated by a drop in the $CC_{1/2}$ in the high resolution shells and increased R_{meas} in all resolution shells). Images were excluded from many of the datasets on this basis. Images without significant radiation damage were merged together to produce a single complete dataset. In contrast to the derivative data (space group = C222₁), XDS identified the space group P2₁ for the final merged native data, and for all six of the individual datasets. Unit cell lengths were similar to those of the derivative data, except the lengths of the b and c axes were exchanged. Matthews coefficient analysis indicated a solvent content of 65% with 4 molecules in the asymmetric unit. At this stage the reason for the differences between the native and derivative data was unclear, but is discussed in detail in Appendix B. A summary of the data reduction statistics for the native data is shown below.

Space group	P2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	88.56, 452.56, 146.36
α, β, γ (°)	90.00, 95.12, 90.00
Resolution (Å)	90.51-2.38 (2.44-2.38)
R_{meas}	25.0 (147.6)
CC_{1/2}	99.0 (31.5)
<i>I</i>/σ<i>I</i>	4.69 (0.39)
Completeness (%)	96.3 (86.8)
Multiplicity	5.6 (1.8)

Table 6.3: **Data reduction statistics for GgTen2^{ΔEGF} native crystals.** Values for the high resolution shell are shown in parentheses.

The native dataset was phased using the model built from the derivative dataset as an MR model. The four molecules in the asymmetric unit displayed alternate tail-to-tail and head-to-head packing (Figure 6.6). The map output by Phaser was solvent flattened and input to Buccaneer for another round of model building. For one of the four chains in the asymmetric unit a model was built as a continuous chain for the region spanning from the beginning of the barrel domain until just a few amino acids from the C-terminal end. Much of this model was convincingly sequenced. Buccaneer failed to re-build the N-terminal region of the protein (the β -sandwich regions and the NHL domain), presumably because the density for this part was poorer. For this reason the continuous model of the barrel and C-terminal region built by Buccaneer was transplanted onto my existing model for the β -sandwich domain 2 and the NHL domain. Four copies of this chimeric model were superposed on those placed by Phaser. At this point, the model was sufficiently complete to commence all-atom refinement with NCS restraints in BUSTER ($R_{\text{work}}/R_{\text{free}}=0.326/0.337$).

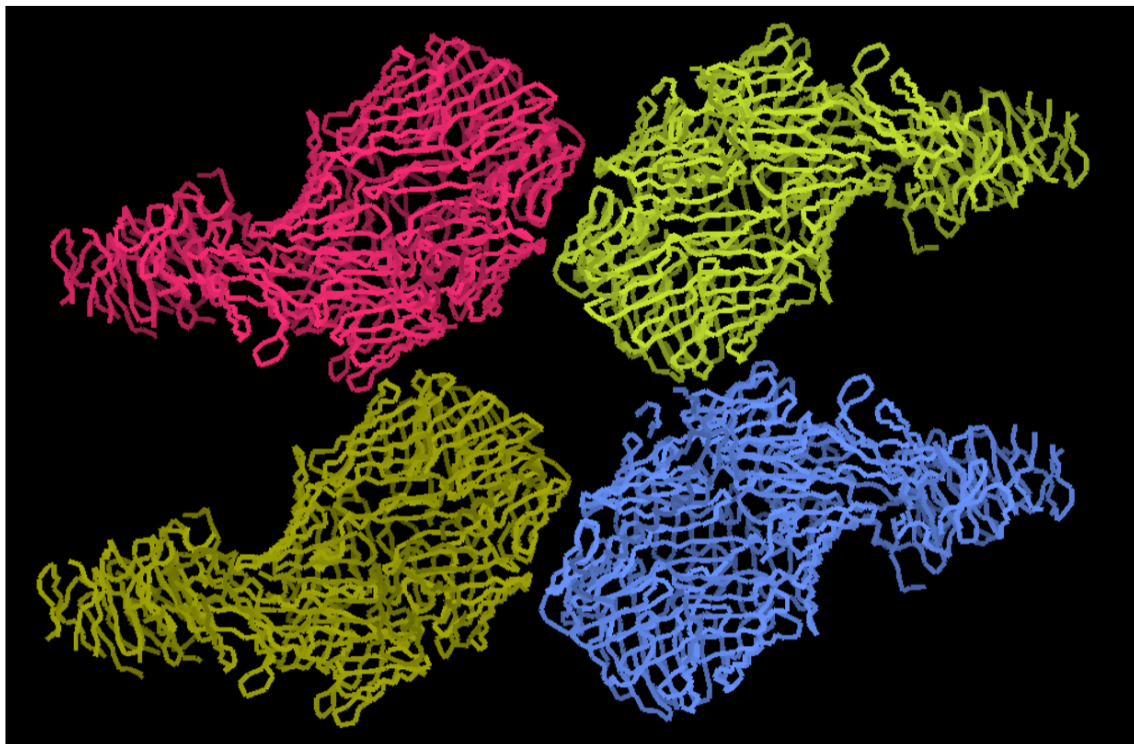


Figure 6.6: **Packing of GgTen2^{ΔEGF} in the P₂₁ asymmetric unit.** Packing of four molecules of GgTen2^{ΔEGF} in the P₂₁ asymmetric unit showing alternate head-to-head and tail-to-tail packing. GgTen2^{ΔEGF} is shown as a Ca trace. There are eight molecules in total in the unit cell. The figure was prepared using Coot [63]

After this point, iterative rounds of manual rebuilding in Coot and all-atom refinement in BUSTER were sufficient to complete the rest of the model. This involved assigning sequence to the NHL domain and C-terminal region and building β -sandwich domain 1 based on the homology model shown in Figure 5.3. N-linked glycans were built automatically using the carbohydrate module in Coot (P. Emsley, unpublished). The final electron density is shown in (Figure 6.7), and the final refinement statistics in Table 6.4. The geometry statistics (bond length, bond angles and Ramachandran plot) all indicate that the model was of good geometric quality.

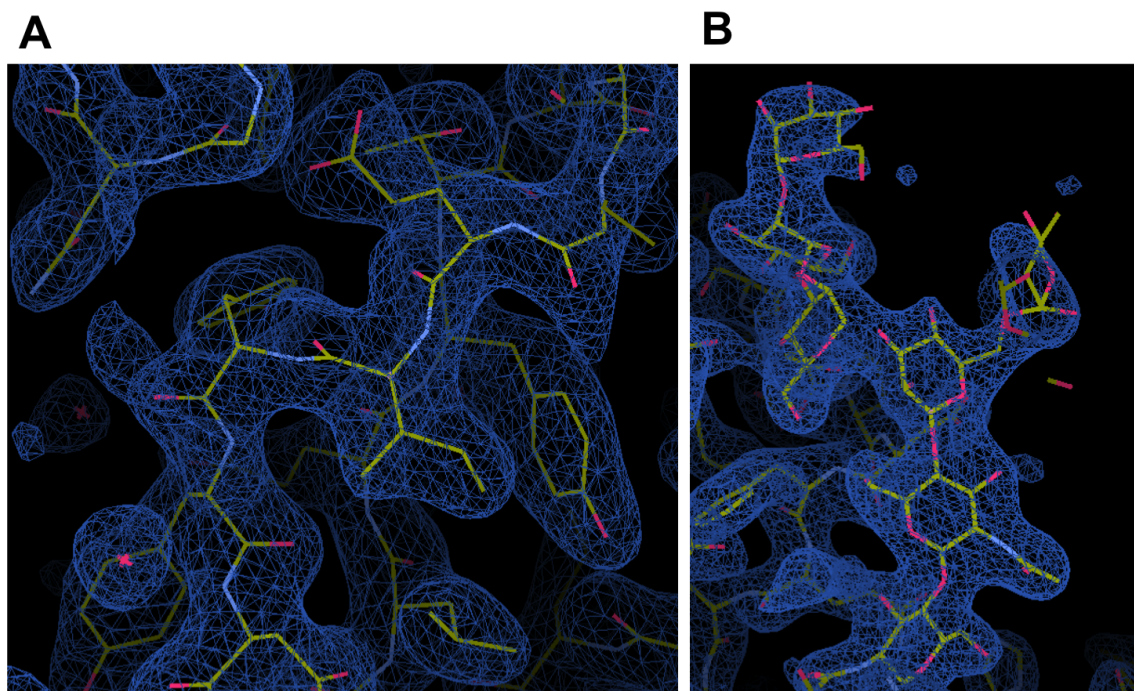


Figure 6.7: **Final electron density for GgTen2 Δ EGF**. Final electron density (2Fo-Fc, contour 1.5 σ) for protein (A) and carbohydrate (B) components of GgTen2 Δ EGF. Atoms are coloured yellow (carbon), blue (nitrogen) and pink (oxygen). The figure was prepared using Coot [63].

Resolution ($\text{\AA}$)	89.38-2.38 (2.44-2.38)
No. reflections (R_{free})	21455
$R_{\text{work}}/R_{\text{free}}$ (%)	22.7/24.4
No. atoms	
Protein	57940
Ligand	1160
Water	153
Average B-factors ($\text{\AA}^2$)	76.1
R.m.s. deviations	
Bond lengths ($\text{\AA}$)	0.007
Bond angles ($^\circ$)	1.01
Ramachandran statistics	
Most favoured regions (%)	96.2
Outliers (%)	0.32

Table 6.4: **Refinement statistics for GgTen2 Δ EGF**. Values for the high resolution shell are shown in parentheses.

6.3 Analysis of the GgTen2^{ΔEGF} structure

The structure of GgTen2^{ΔEGF} reveals a surprisingly compact and highly intricate domain arrangement, centred around the barrel-like structure predicted by homology modelling. At its longest dimensions it measures ~ 12.6 nm in length and ~ 7.2 nm in width. The structure is highly modular and is formed of seven distinguishable domains (Figure 6.8), however their complex arrangement means that these “domains” are unlikely to be stable in isolation. This probably underlies the difficulty in generating truncation constructs. In the next sections, each domain and its similarity to existing structures is discussed in detail.

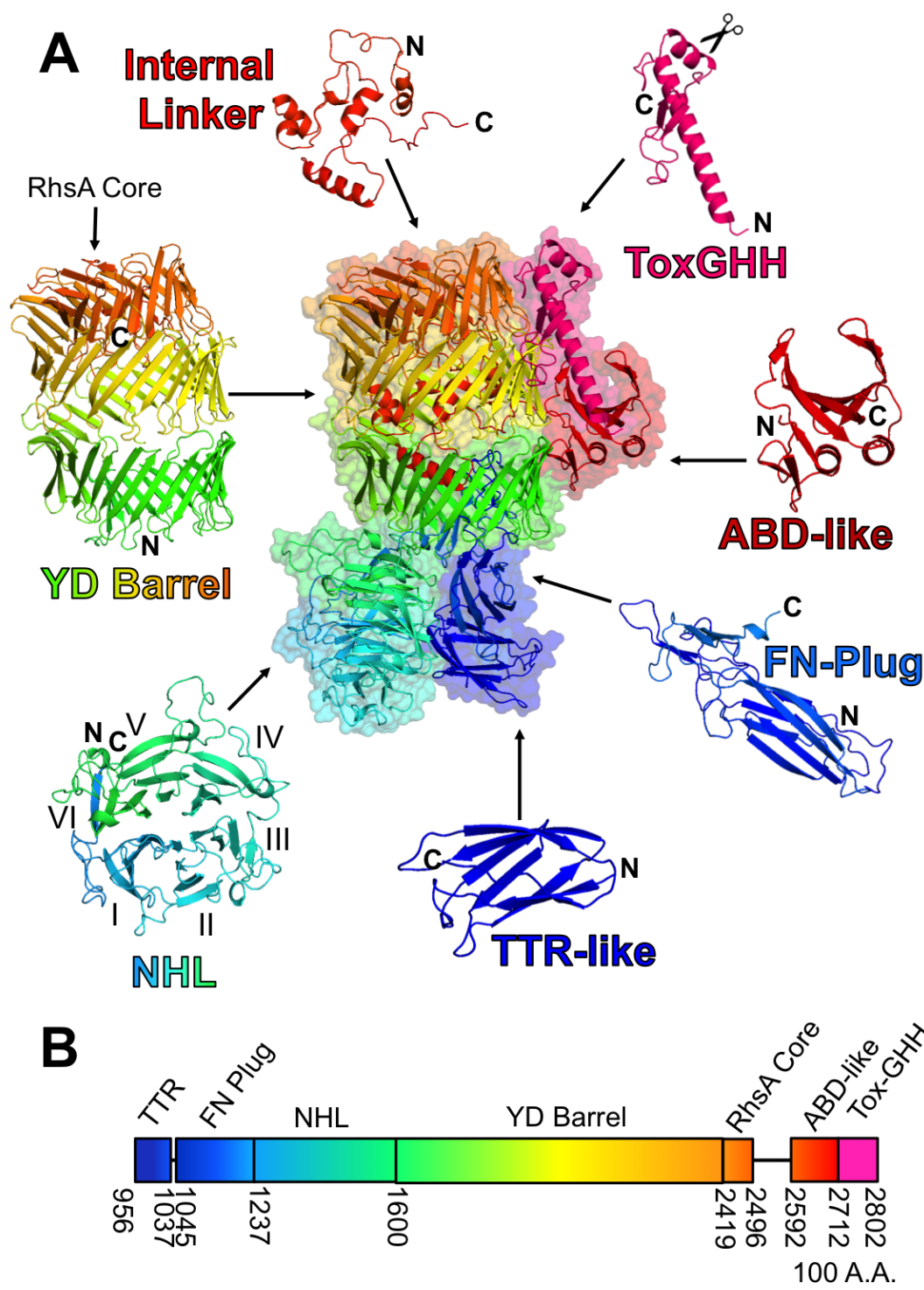


Figure 6.8: **Overall architecture of GgTen2 Δ EGF**. GgTen2 Δ EGF is shown in both cartoon and surface representation and is coloured according to the rainbow: blue, N-terminus; red, C-terminus. Its constituent domains are shown in cartoon representation and their positions in the intact structure indicated by arrows. The position of the predicted TCAP cleavage site is indicated by a scissor symbol. The figure was prepared with PyMol. Abbreviations are as follows: FN, fibronectin; TTR, transthyretin; RhsA Core, Rhs-associated core domain; ABD-like, antibiotic-binding-like domain (A). Annotation of the primary amino acid sequence according to the domain arrangement revealed by the crystal structure. The positions of each domain shown in (A) are indicated. Colours are also as above (B).

6.3.1 A Tour of the GgTen2^{ΔEGF} Structure: from N- to C-terminus

β-sandwich domain 1 becomes a TTR domain

An 81-amino acid *β*-sandwich domain comprises the very N-terminus of GgTen2^{ΔEGF}. This domain is formed of seven *β*-strands arranged in two *β*-sheets, one antiparallel four-stranded and the other mixed three-stranded, both sandwiching a hydrophobic core (Figure 6.8A). This domain strongly resembles the homology model generated by SWISS-MODEL to aid construct design. Comparison of this *β*-sandwich domain with the Dali database [95] reveals that it is most similar with the transthyretin (TTR)-like domain of human carboxypeptidase N (r.m.s.d. 1.85 Å for 72 aligned C α , PDB accession: 5AQ0). TTR-like domains are widely distributed in species of bacteria, plants and animals, and gain their names from their structural similarity to transthyretin, a transport protein for two thyroid hormones and retinol. They are thought to play a role in protein multimerisation/aggregation [273, 76].

β-sandwich domain 2 is a novel type of FN domain

A short linker ($\sim$ 8 amino acids, which are unresolved in my crystal structure) separates the TTR-like domain from the next domain, comprising residues 1045-1237 (Figure 6.8). This domain can be split into two subdomains. Like the TTR domain, subdomain I has seven *β*-strands split across two *β*-sheets. Both *β*-sheets are antiparallel, one four-stranded and one three-stranded. The Dali server [95] identified that subdomain I is highly structurally similar to FN type III domain 8 of fibronectin (r.m.s.d. 2.14 Å for 82 aligned C α , PDB accession: 2GEE) (Figure 6.9). These FN-III domains are commonly found in metazoan extracellular proteins, including FLRT proteins, and can function as protein-protein interaction modules [26].

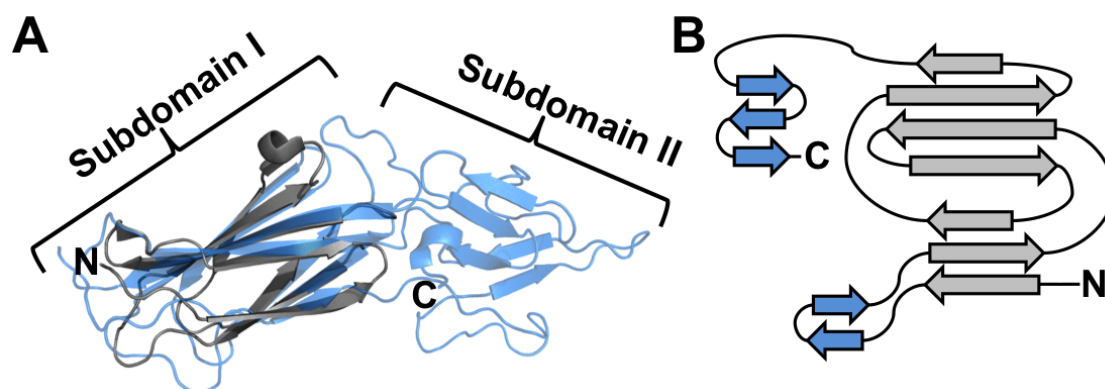


Figure 6.9: **Architecture of the FN-plug domain.** Structural alignment of subdomain I of the FN-plug domain with the crystal structure an FN-III domain from fibronectin (PDB accession: 2GEE). Subdomain II is formed by an insertion between the first and second β -strands of the FN-III domain, along with a C-terminal extension. The FN-plug domain is shown in transparent blue cartoon representation. The FN-III domain of 2GEE is shown as a grey cartoon. The figure was prepared using PyMol (A). The secondary structure arrangement of the FN-plug domain is shown in 2D representation. Arrows indicate individual β -strands. Arrows coloured in grey represent β -strands typical of a standard FN-III domain, the additional extensions present in the GgTen2 FN-plug (subdomain II) are shown as blue arrows (B).

Subdomain II is formed by a 44 amino acid insertion between the first and second β -strands of the FN-III domain and a ~ 38 amino acid extension of its C-terminus. The insertion between β -strands one and two largely lacks defined secondary structure, but does contain a short β -hairpin. The C-terminal extension forms three short antiparallel β -strands. The Dali server identified no convincing structural homologues for subdomain II. Interestingly, subdomain II resides inside the barrel cavity, acting to “seal” the barrel from its N-terminal side. Subdomain I is situated outside of the barrel cavity and packs against the TTR-like domain (Figure 6.8).

The NHL domain

As predicted, the NHL domain forms a six-bladed β -propellor, with Dali [95] identifying a protein of unknown function from *Bacteroides thetaiotaomicron* as its closest structural homologue (r.m.s.d. 2.31 Å for 251 aligned C α atoms, PDB accession: 3HRP). Incidentally this is the same structure which was truncated to provide the MR model to initially locate the NHL domain. Although the core β -propellor region of the domain is highly conserved, the loop regions between individual β -strands are considerably longer in the GgTen2 NHL domain than in the other six-bladed β -

propellers identified by Dali (Figure 6.10A). These extensions frequently contain intra-loop disulphides, presumably rigidifying them. All ten of the cysteines involved in forming the five intra-loop disulphide bonds are highly conserved across the vertebrate Teneurin NHL domains. It is worth noting that the TcdB2-TccC3 and YenB-YenC2 toxin structures discussed in the next section also have a six-bladed β -propellor domain at the same relative position (immediately N-terminal of their barrel-like domains). This occludes the N-terminal end of the barrel. The propellor domain of YenB2 is shown in Figure 6.10B.

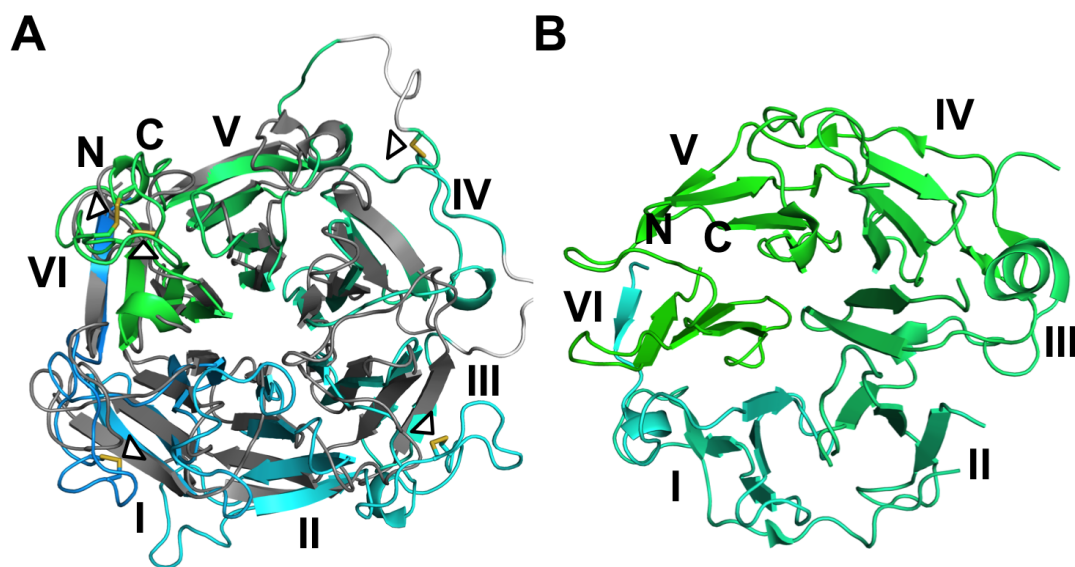


Figure 6.10: **The NHL domain is an elaborate six-bladed β -propellor.** The GgTen2 NHL domain is a highly elaborate six-bladed β -propellor with many long and structured loops separating the individual β -strands of the propellor. Five disulphide bonds, all located in external loop regions, are shown as yellow sticks and are marked by white arrowheads. The GgTen2 Δ^{EGF} NHL domain is coloured from N- to C-terminus according to a blue-to-green gradient. Two regions which are unresolved in the crystal structure (residues 1419-1426 and 1485-1491) were modelled in geometrically sensible positions and are coloured in white. The six-bladed β -propellor of the closest structural homologue identified by the Dali server is shown as a grey cartoon (A). The six-bladed β -propellor domain of YenB2 is shown in cartoon representation and coloured from cyan to green (N- to C-terminus) (B).

The YD barrel domain

Like the structures of TcdB2-TccC3 and YenB-YenC2, the YD repeat-containing region of GgTen2 forms a large shell [34, 167]. The shell is formed by a single, long β -sheet consisting of 74 β -strands. The GgTen2 barrel measures ~ 78 Å in length and is

less extended than that of the YenB-YenC2 heterodimer (87 Å length). The barrels are similar widths (~ 48 Å for GgTen2 as compared to 42 Å for YenB-YenC2). The YenB-YenC2 barrel encloses a considerably larger volume than GgTen2 (~ 76000 Å³ as opposed to ~ 44000 Å³) [39]. The GgTen2 YD barrel is formed solely of β -sheet, unlike the barrel of TcdB2-TccC3 and YenB-YenC2, which is externally decorated with many short α -helical extensions between the β -strands (Figure 6.11A).

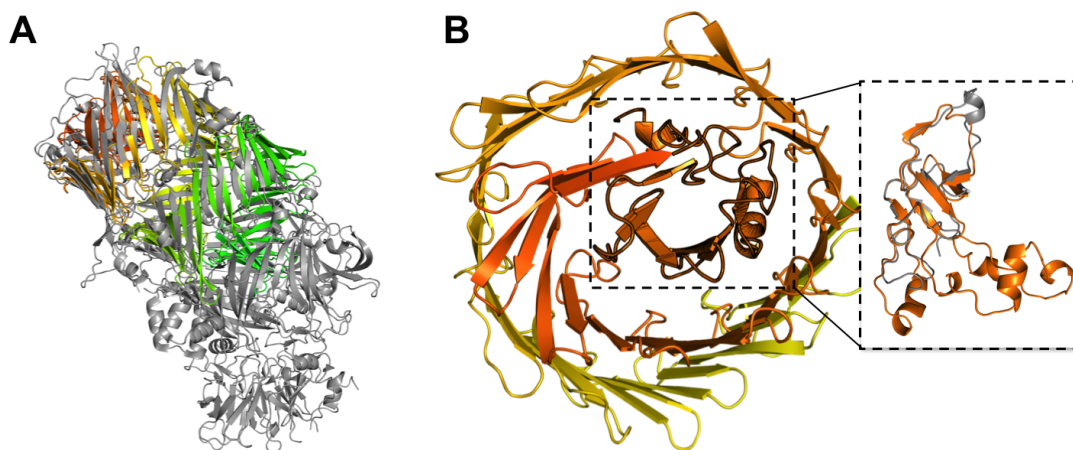


Figure 6.11: The YD barrel of GgTen2 resembles those of bacterial toxins. Alignment of the GgTen2 barrel structure (rainbow) with that of YenB-YenC2 (grey) reveals that the overall architecture of the barrel is conserved between the two proteins, but the barrel of YenB-YenC2 is substantially larger than that of GgTen2. The structures were aligned via their Rhs-associated core region (A), which is highly structurally conserved between the two proteins and forms the plug at the C-terminal end of the barrel (boxed) (B). This domain shows remarkable structural homology tho the Rhs-associated core region of YenB-YenC2 (which is shown overlaid in grey cartoon representation).

Like the structures of the BC toxins, the end of the barrel is “plugged” by the Rhs-associated core domain (residues 2419-2496), which is largely structurally conserved between GgTen2 and the BC toxins (r.m.s.d. 1.48 Å for 66 aligned C α atoms for the Rhs-associated core domain of YenB-YenC2) (Figure 6.11B). The Rhs-associated core domain of YenB-YenC2 has been shown to contain pH-dependent autoproteolytic activity, with residues R650, D663 and D686 being essential for catalysis [34]. Although two of these catalytic residues are conserved in GgTen2, the third (D686), which is essential for catalysis, is not present, and the backbone where it should be located is substantially shifted. This makes it unlikely that GgTen2 harbours such autocatalytic activity. In agreement with this there is no break in the density for

the backbone of GgTen2^{ΔEGF} at this position.

The Internal Linker domain

One of the most curious features of the GgTen2^{ΔEGF} structure is the region C-terminal of the Rhs-associated core region. Similar to the structures of the TcdB2-TccC3 [167] and YenB-YenC2 [34] complexes, the C-terminal end of the YD-barrel of GgTen2 is located within the barrel itself. Residues 2467-2592 of GgTen2 are located within the interior of the barrel and have some helical secondary structure. The Dali server identifies no structural homologues for these residues. The most C-terminal forty amino acids of this region represent the residues unconserved between the four human Teneurins shown in Figure 5.4. The helices largely pack against the interior surface of the barrel, leaving an empty cavity in its centre (Figure 6.12).

One of the most striking features of the structure is that, around 120 amino acids after the Rhs-associated core region, the C-terminus leaves the interior of the barrel via a gap between two β -strands of the shell (Figure 6.12B). This is in contrast to TcdB2-TccC3 and YenB-YenC2, where the entire C-terminus of the protein is thought to reside within the barrel [34, 167].

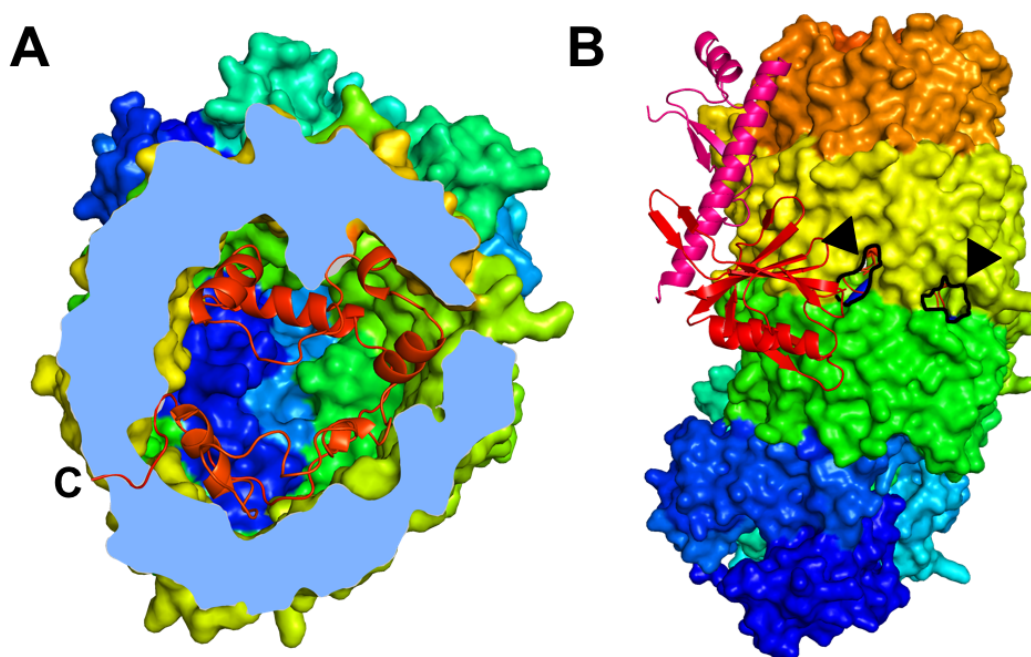


Figure 6.12: **The YD-barrel encloses part of the GgTen2 C-terminus.** Clipped view through the top of the YD barrel. The YD barrel encloses ~ 120 residues immediately C-terminal of the Rhs-associated core region (the “internal linker”, shown in red cartoon representation). The rest of the protein is shown in surface representation and is coloured according to the rainbow (A). Side view of GgTen2 $^{\Delta\text{EGF}}$ shown in surface representation. The C-terminal ~ 200 amino acids are shown as cartoon. The hole in the barrel through which the C-terminus protrudes is outlined in black. Another hole in the barrel is also indicated by a black outline. Both of these holes are indicated by black arrowheads (B).

The antibiotic-binding (ABD)-like domain

Upon leaving the barrel, the ~ 200 most C-terminal residues of GgTen2 pack against the exterior of the barrel (Figure 6.12B). The first ~ 120 of these residues fold to form an intricate domain that has high β -strand composition. The main structural feature is a mixed five-stranded β -sheet, with a short α -helix separating the first and second strands of the sheet. This sheet bears structural homology to a class of proteins involved in binding the small molecule antibiotics bleomycin and zorbamycin (r.m.s.d. $2.38 \text{ \AA}$ for 52 aligned $C\alpha$ atoms, PDB accession: 1EWJ, bleomycin-binding protein [162]) (Figure 6.13). These bacterial proteins confer resistance to the antibiotics by sequestering them, preventing their activation by oxygen [217]. The bleomycin-binding surface has been identified crystallographically [162]. The equivalent surface in GgTen2 $^{\Delta\text{EGF}}$ is solvent-exposed.

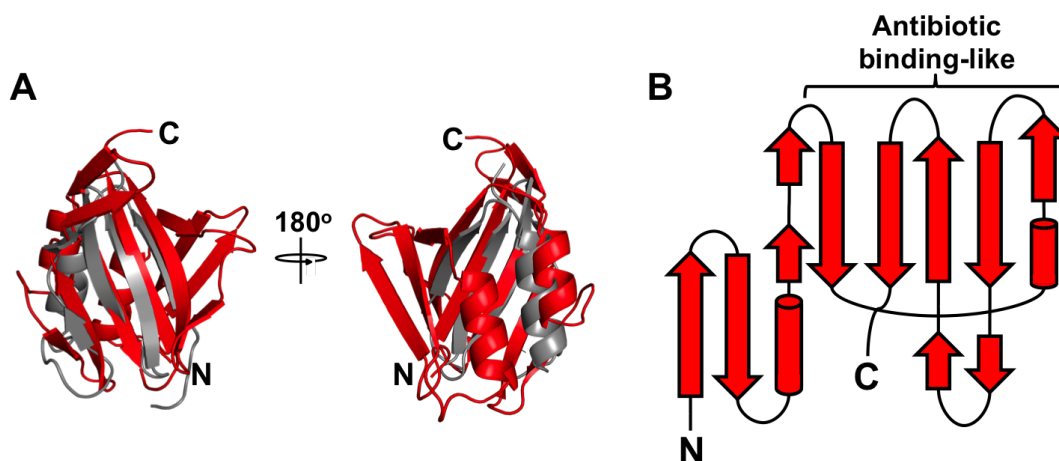


Figure 6.13: **Similarity to bacterial antibiotic-binding proteins.** Structural alignment of GgTen2 residues 2594-2713 with a bleomycin-binding protein (PDB accession: 1EWJ). GgTen2 is coloured in red and 1EWJ is coloured in grey (A). Topology diagram of GgTen2 residues 2594-2713. β -strands are represented by arrows and α -helices by cylinders (B).

The Tox-GHH domain

The very C-terminal region of Teneurins (residues 2720-2795 in GgTen2) has previously been identified to have sequence similarity to a novel family of DNases, the Tox-GHH family [290]. The Teneurin Tox-GHH domain is thought to be catalytically inactive. Structural comparison of the residues annotated as the Tox-GHH domain by BLASTp (residues 2720 - 2795) with the Dali database [95] reveals highest structural homology to the catalytic nuclease domains of a class of bacteriocins known as A-type colicins (the top hit, Colicin E9, has r.m.s.d. 2.65 Å for 60 aligned C α atoms, PDB accession: 1BXI) (Figure 6.14A).

The Tox-GHH family are predicted to be HNH nucleases, meaning they use a catalytic triad of two histidine residues and one asparagine residue to cleave nucleic acids. The first histidine residue of the motif is thought to act as a general base, activating water for nucleophilic attack on the sugar-phosphate backbone of nucleic acids [203]. The asparagine residue plays a structural role, forming a stabilising hydrogen bond with the backbone of an adjacent β -strand. The third histidine residue is involved in coordination of a catalytically important metal ion [118]. Structural and sequence alignment of GgTen2, plus all four human Teneurin homologues, with

various nucleic acid-cleaving colicins reveals that, in Teneurins, the two catalytic histidine residues of the HNH motif are mutated to tyrosine and glutamine residues respectively. Conversely, the asparagine residue of the Teneurin “HNH” motif (N2790) is conserved and plays a stabilising structural role by hydrogen bonding to the main chain of the adjacent β -strand, similar to its role in the structure of colicin E9 [118] (Figure 6.14B). These mutations would presumably make the domain catalytically inactive. Intriguingly, this vestigial HNH motif is located within the “TCAP” region of the protein, that is predicted to be cleaved *in vivo* and act as a bioactive peptide [157].

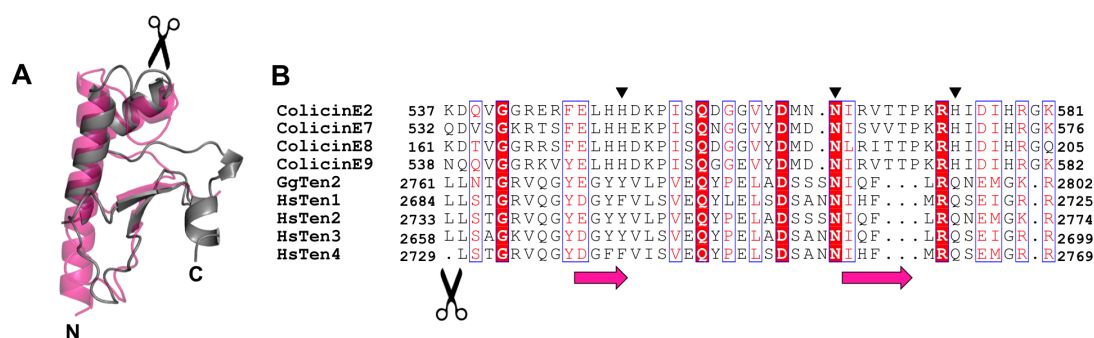


Figure 6.14: **Structural similarity between the Tox-GHH domain and bacterial colicin DNases.** Structural alignment of the GgTen2 Tox-GHH domain (residues 2720 - 2795) with the C-terminal nuclease domain of ColicinE9 (PDB accession: 1BXI). The GgTen2 Tox-GHH domain is shown in hot pink and the ColicinE9 nuclease domain in grey (A). Sequence alignment of the *E. coli* colicin DNase domains and the TCAP regions GgTen2 and HsTen1, 2, 3 and 4. Black arrowheads indicate the HNH motif of the *E. coli* colicins. The pink arrows represent β -strands. The scissor symbol represents the predicted TCAP cleavage site for the Teneurins. Colicins are not thought to be proteolytically cleaved at this position. The alignment was performed using Clustal Omega [235] and the figure was prepared using ESPript [211].

6.3.2 Indications of Functionally Important Surfaces on GgTen2

As shown in earlier chapters, protein interaction sites often reside in conserved surface patches. The surface sequence conservation of GgTen2 ^{Δ EGF} was analysed using the ConSurf server [80] to give indications of functional regions of the protein. Analyses were performed using the Teneurin sequences from a diverse range of vertebrates. Due to the extraordinarily high identity between the vertebrate Teneurin orthologues (for example, the human and chicken Teneurin2 ectodomain sequences are 93% identical), sequences of all four Teneurin homologues were used in the

alignment in order to create sufficient diversity for the calculations to be reliable. To supplement the information from the surface conservation, and due to the link between Teneurins and cancers [293], missense mutations in all four Teneurin homologues from the COSMIC cancer database were scored according to their frequency of occurrence. The highest scoring mutations were mapped onto the surface of the structure [72].

This analysis reveals two highly conserved patches on the surface of the GgTen2 barrel structure (outlined and shown from four different angles in Figure 6.15A). Comparison with the positions of Teneurin mutations from cancer samples reveals that many of the most highly occurring mutations can be found in these conserved patches (Figure 6.15B).

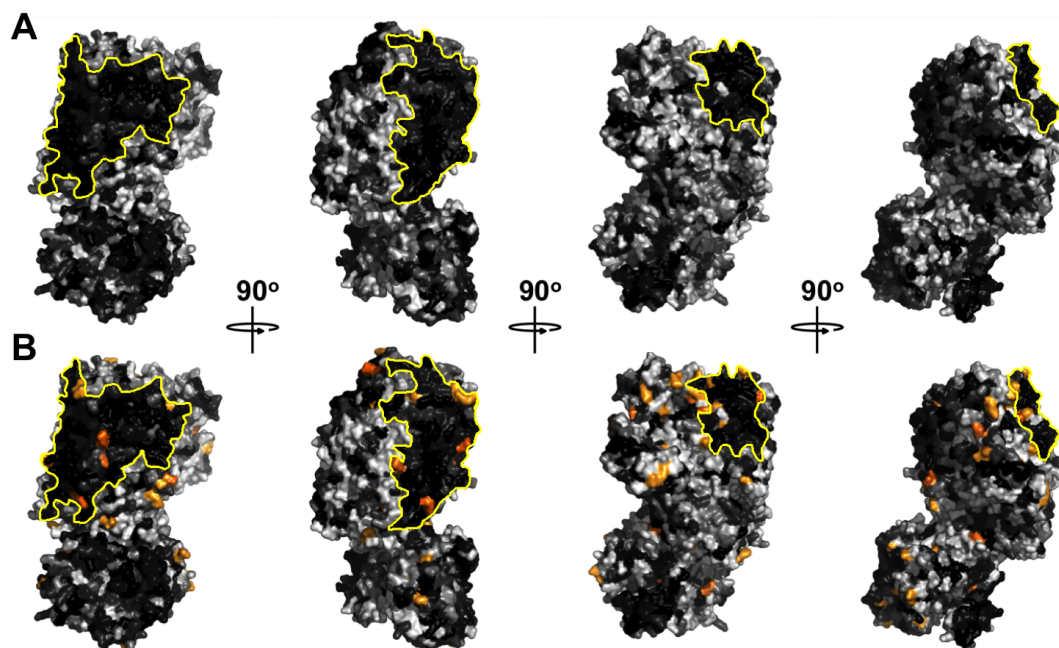


Figure 6.15: Conserved surface patches on GgTen2^{ΔEGF} contain cancer mutations. The structure of GgTen2^{ΔEGF} shown in surface representation. A sequence alignment of all Teneurin homologues from human (*Homo sapiens*), chicken (*Gallus gallus*), fish (*Danio rerio*), shark (*Callorhynchus milii*), frog (*Xenopus laevis*, *Xenopus tropicalis*) and alligator (*Alligator mississippiensis*) was used to produce surface conservation scores with ConSurf [80]. Black indicates highest conservation scores and white the lowest conservation scores (A). Mutations from the COSMIC cancer database are shown in orange. Mutations at positions highlighted in the brightest orange occur 6-9 times in the database. Mutations in the fainter orange colour occur 3 - 5 times. Mutated positions occurring fewer than three times are not shown (B).

6.3.3 GgTen2^{ΔEGF} Lattice Contacts

Crystal lattice contacts can be used to formulate hypotheses about physiological protein-protein interaction surfaces. Inspection of the crystal contacts formed by GgTen2^{ΔEGF} reveals that the smaller of the two conserved patches indicated in Figure 6.15A mediates the head-to-head packing of GgTen2^{ΔEGF} monomers in the asymmetric unit (Figure 6.16). The buried surface area for this interaction is 993 Å² per monomer.

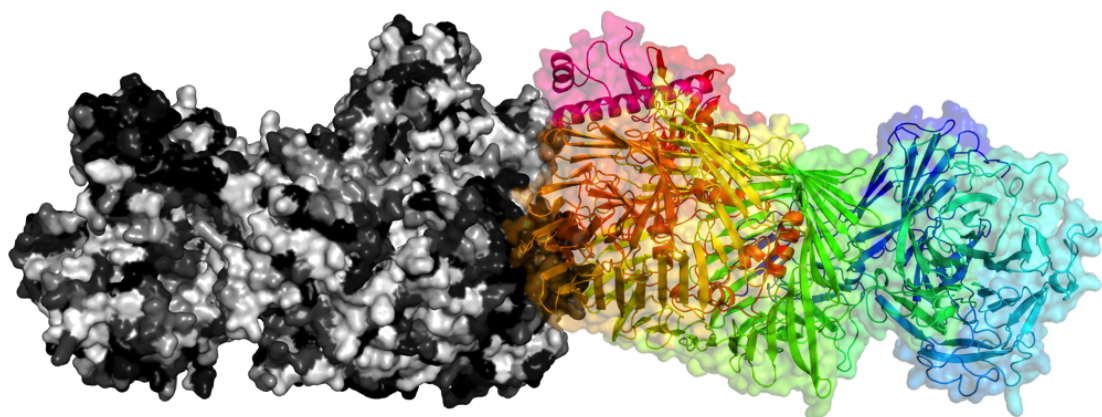


Figure 6.16: **A conserved surface on GgTen2^{ΔEGF} mediates crystal contacts.** One GgTen2^{ΔEGF} monomer is shown as surface and coloured according to its sequence conservation as in Figure 6.15. The other GgTen2^{ΔEGF} monomer is shown in transparent surface and cartoon representation and coloured as in Figure 6.8.

Another interesting crystal contact is formed by the GgTen2^{ΔEGF} NHL domain. The Teneurin NHL domain has previously been implicated in determining the specificity of Teneurin homophilic interactions [18]. In the GgTen2^{ΔEGF} crystal lattice, the NHL domain provides an extended crystal lattice contact between adjacent antiparallel monomers, burying 1100 Å² surface area (Figure 6.17). Key to forming this interface are residues of the alternatively spliced NHL loop (residues 1303-1309) [237].

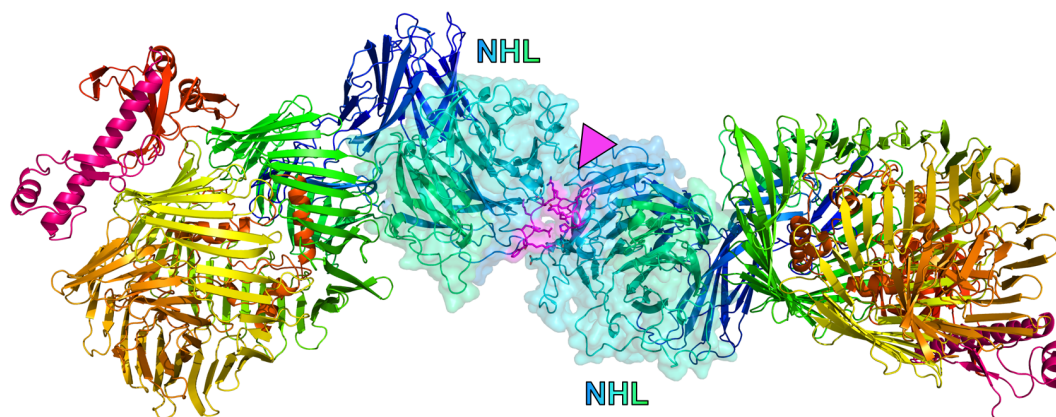


Figure 6.17: **An alternatively spliced loop in the NHL domain of GgTen2 Δ EGF mediates crystal contacts.** The alternatively spliced loop in the NHL domain is indicated by a magenta arrowhead with side chains shown as sticks. The NHL domain is highlighted in transparent surface representation. Each GgTen2 Δ EGF monomer is shown in cartoon representation and coloured as in Figure 6.8.

There are several examples where crystal contacts have represented physiologically relevant interaction surfaces [227, 228, 226]. However, crystallisation is an entirely unphysiological process for most proteins and so conclusions based on crystal lattice contacts must always be thoroughly validated using complementary methods (i.e. biophysical methods in solution and also ideally in cell-based assays). Such experiments will be necessary to determine if either of the crystal lattice contacts discussed here represent a physiologically relevant interaction surface.

6.4 Discussion

The crystal structure of GgTen2^{ΔEGF} provides a rich source of information about the function of this mysterious protein, although also raises very many questions about its function and evolution.

In terms of evolution, Teneurins are thought to have arisen in choanoflagellates, unicellular organisms commonly considered to be the closest living relatives of the metazoans [116]. Bioinformatic studies have implied that Teneurins are the result of a horizontal gene transfer of both a prokaryotic gene (contributing the YD-repeats) and a diatom or fungal gene, resulting in the formation of a complex hybrid fusion protein [259]. This is strikingly corroborated by the domain architecture of the GgTen2^{ΔEGF} crystal structure. The most N-terminal domains of the structure (the TTR-like, FN-plug and NHL domains) bear strong structural homology to domains commonly found in both eukaryotes and prokaryotes [151, 64, 127]. Similar domains persist in the ectodomains of diverse metazoan cell surface receptors and extracellular matrix proteins. However, the remainder of the ectodomain bears closest structural homology to a range of bacterial proteins (Figure 6.18). The most obvious example of this is the Teneurin YD-barrel which shares high structural homology to the central domain of Rhs family bacterial polymorphic toxins [290, 34, 167]. This “central domain” is thought to encapsulate the protein’s highly variable C-terminal toxic domain, protecting the producing cell from its deleterious effects [101, 34]. The very C-terminal ~200 amino acids of GgTen2 could be considered analogous to the C-terminal toxic domain of these bacterial toxins. Indeed, the striking finding that the Tox-GHH domain is structurally similar to the cytotoxic DNase domain of Colicin E9 reaffirms the idea that the Teneurins arose from a single transfer of a complete bacterial Rhs family toxin [290]. The presence of the ABD-like domain in the C-terminus of GgTen2 was entirely unexpected, but its fold is likely a relic of its prokaryotic ancestry. Whether or not the ABD-like domain retains its small-molecule-binding ability remains an open question.

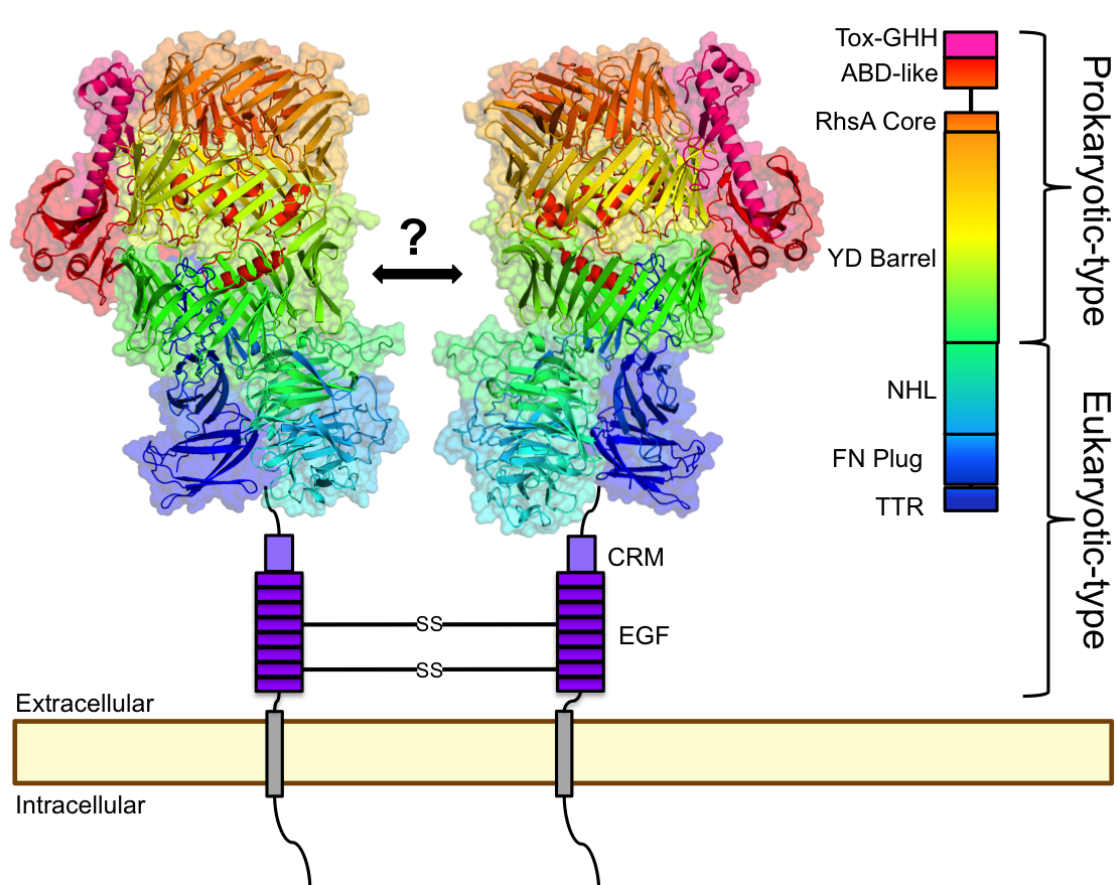


Figure 6.18: **The hybrid architecture of the GgTen2 ectodomain.** Potential *in cis* arrangement of GgTen2 Δ EGF at the cell surface. “Prokaryotic-type” and “Eukaryotic-type” domains of the receptor are indicated demonstrating the hybrid nature of this receptor. The exact molecular mechanism by which Teneurin dimerises *in cis* is not yet understood. The two GgTen2 Δ EGF monomers are shown in transparent surface and cartoon representation and coloured as in Figure 6.8. SS, disulphide bond; CRM, cysteine-rich motif.

Previous speculation had suggested that the C-terminal region of Teneurins may be enclosed within the YD-barrel, similar to the ABC toxins [34]. However, the crystal structure of GgTen2 Δ EGF demonstrates that this is not the case, and that the C-terminal domains (the ABD-like and Tox-GHH domains) are displayed outside of the barrel. This suggests that their exposure to the surroundings is important to their function. Although the Teneurin Tox-GHH domain is unlikely to harbour nuclease activity, it may serve an alternate function, such as mediating interactions with lipids, as observed for the ColicinE9 DNase domain [173]. Interestingly, a BLASTp search of the GgTen2 Tox-GHH domain against all bacterial proteins identified a small number of bacterial Tox-GHH domains which would also be predicted to be

catalytically inactive (Figure 6.19), suggesting that nucleic acid hydrolysis is not the only function of these domains in bacteria.

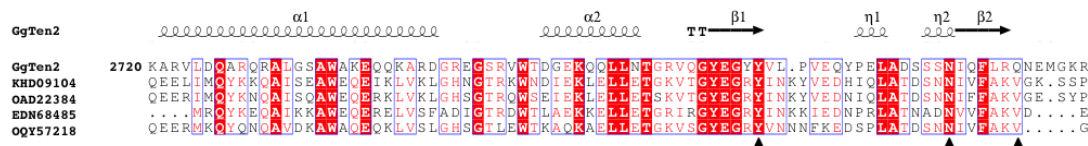


Figure 6.19: **BLAST searches identified catalytically-inactive Tox-GHH domains in bacteria.** Sequences of the bacterial proteins are referred to by their GenBank accession codes. KHD09104 and OAD22384 are hypothetical proteins from *Thiomargarita nelsonii*, whereas EDN68485 and OQY57218 are both from *Beggiatoa sp.* All of these proteins have three predicted transmembrane helices at their C-terminal end, with the Tox-GHH domain located N-terminally. No other predicted domains are present. The secondary structure of the GgTen2 Tox-GHH domain is depicted above the alignment. Invariant residues are shown with red backgrounds. Stretches of residues with >0.7 global similarity score are surrounded by blue frames. The alignment was performed using Clustal Omega [235] and the figure was prepared using ESPript [211].

Several other eukaryotic protein domains are thought to have evolved from horizontal transfer from bacterial toxin systems. The SUKH domain, which in eukaryotes is thought to act as an adaptor regulating protein ubiquitination or polyglutamylation, is thought to have originated from a bacterial toxin immunity protein, exploiting its ability to bind diverse targets [291]. Similarly, multiple RNA modifiers for tRNAs, mRNAs and short non-coding RNAs are thought to have their evolutionary origin in nucleic acid deaminases from bacterial toxin systems [100]. Bacterial toxin systems are under intense selective pressure for constant diversification and thus it has been suggested that these systems provide rich sources of molecular variation for the evolution of new and distinct systems to drive major evolutionary transitions [290].

Chapter 7

Discussion and Future Work

The development of the nervous system requires a complex series of cell guidance events, which are directed by relatively few cell guidance receptors. How such few factors control the development of a complex organ such as the brain is poorly understood and a matter of intense research. Compensating for the relatively small number of receptors is the ability of many guidance factors to form heteromeric multi-subunit receptor complexes, modifying their downstream signals and allowing these receptors to function in a context-specific manner [59, 254]. Central to this idea is the ability of some receptors to act as “context-recognisers”. The aGPCRs present a relatively unexplored class of molecules that could fulfil this role.

7.1 Lphns as Context-Recognisers

The work in this thesis lends strong support to the idea that Lphns act as context-recognisers. The extracellular domain of Lphn directly binds to both FLRTs and Teneurins and also indirectly interacts with Unc5 proteins. Crystal structures of FLRT3^{LRR}-Lphn3^{Olf} and FLRT2^{LRR}-Unc5D^{Ig} reveal 1:1 complexes [208, 158], but in my crystal structure with Unc5 an unusual 4:2:2 Lphn3:FLRT2:Unc5D complex is observed. Although Lphn does not bind directly to Unc5s, by interacting with FLRTs, Lphn is clearly capable of recognising Unc5s. Moreover, I have also shown that the stoichiometry of these complexes depends on the homologues of Unc5 and FLRT present, revealing that Lphns are capable of finely recognising their surround-

ings and tuning the complexes they form according to homologue-specific expression patterns. In Section 3.4.1 I showed that interaction of Unc5D and FLRT2 *in cis* attenuates FLRT2-mediated adhesion to Lphn3^{Lec-Olf} stripes. Although this result establishes that these three receptors functionally interact, mediating an outcome distinct from individual receptor pairs, its physiological significance is unclear. Advanced *in vivo* techniques such as *in utero* electroporation of FLRT and Unc5 non-binding mutants, coupled with whole-mount brain clarification [41] have proved useful in elucidating the functions of FLRTs and Unc5s in the brain [226]. Similar experiments using the new mutants described in this thesis may eventually lead to clues on the functions of distinct Lphn complexes in the brain.

The preliminary finding that Lphn3-FLRT2-Ten2 are capable of forming a triple complex, but that Ten2 is excluded from the Lphn3-FLRT2-Unc5D complex, further supports the conclusion that Lphns act as context-recognisers. A structure of the Lphn3-FLRT2-Ten2 complex would likely provide clues as to the molecular basis for this observation. Given the unusual stoichiometry of Lphn-FLRT-Unc5 complexes, it would be interesting to perform native MS experiments on the Lphn3-FLRT2-Ten2 complex to determine if Lphn3 is also present in an unequal stoichiometry in this complex. Similar to FLRT and Unc5, the Teneurins (Ten1-4) display distinct expression patterns in the developing mouse brain [292]. Given that the stoichiometry of the Lphn-FLRT-Unc5 complex is homologue-dependent, it is tempting to speculate that the complexes Lphn forms with Teneurins may also be homologue-dependent. In performing these experiments one of the main challenges will be producing recombinant protein for the other Teneurin homologues, although now that the GgTen2^{ΔEGF} structure is solved this could be used to create homology models of the other Teneurin homologues to inform rational construct design.

7.2 Lphns and its Complexes at Synapses

Lphns interact with a variety of intracellular and extracellular ligands predominantly found at synapses (SHANKs, FLRTs, Teneurins, Neurexins and Contactin6) [136]. Although both FLRTs and Teneurins are well-characterised at synapses [191, 97, 176], there is little evidence of Unc5 expression at synapses. In fact, in *C. elegans* Unc5 has been shown anti-synaptogenic and to be important in excluding presynaptic components from dendrites [204]. Therefore Lphn:FLRT:Unc5 complexes may not form at the synapse. One important question to answer will be in what context each of the complexes discussed in this thesis forms. This will require a detailed understanding of when and where each of these proteins is expressed. Although traditional expression profiling methods (such as *in situ* hybridisations and immunohistochemistry) are likely to be useful here, the advent of single-cell RNA sequencing technology provides an opportunity to identify cell type-specific molecular signatures with unprecedented resolution and would likely provide useful information with regard to the exact cells these complexes might form on [187].

In addition to understanding the cell types on which these complexes form, it will also be important to determine whether the complex components are all contributed by the same cell or by different cells. For technical reasons the crystal structures presented in this thesis comprise just fragments of receptor extracellular domains. Both Lphn and FLRT contain long and glycosylated linkers in their extracellular domains as well as additional structured domains. This means that, based on the Lphn3-FLRT2-Unc5D crystal structure, it is impossible to predict the relative positions of the transmembrane regions of each component. Structures using full ectodomain constructs of Lphn3, FLRT2 and Unc5D (and also Ten2) would allow a more accurate prediction of the position of each transmembrane region, and thus allow the cellular arrangement of each component in the complex to be inferred. Although such a complex is highly unlikely to be amenable to crystallography, the “resolution revolution” in cryo-electron microscopy [62] presents new possibilities to study large

macromolecular complexes with a degree of structural heterogeneity. In addition, cryo-electron microscopy offers opportunities to structurally study transmembrane regions themselves in more native-like membrane environments than is possible using X-ray crystallography. New technology in which recombinant proteins are inserted into extracellular vesicles allows membrane proteins to be studied with their correct topology in a native lipid bilayer [289]. Structural studies of full-length receptors in these vesicles would give valuable insight into how exactly the complexes discussed in this thesis form at the cell surface.

7.3 Lphn Dimerisation

The crystal structure of Lphn3-FLRT2-Unc5D reveals two copies of Lphn3 bound per copy of FLRT2 and Unc5D. Additionally, Ten2 forms a constitutive dimer, although can still bind to Lphn in a monomeric form, meaning there is a strong possibility that the Ten-Lphn interaction also dimerises Lphn. This raises the possibility that Lphn acts as a constitutive dimer, or may be dimerised upon ligand binding. As discussed in Section 4.9, published evidence suggests that Lphn dimerisation may play a role in its signal transduction mechanism [266, 205]. Fluorescence Resonance Energy Transfer (FRET, or similar) experiments with fluorescently-tagged Lphns expressed on the cell surface would likely be useful in elucidating this mechanism (assuming that both Lphns in the dimer are expressed from the same cell). FRET experiments have already been applied successfully to study the homo- and hetero-oligomerisation properties of another (unliganded) aGPCR, EMR2 [47]. Monitoring the FRET signal in the presence and absence of Lphn ligands would likely reveal insights into ligand-induced Lphn oligomerisation. Many of the mutant proteins discussed in this thesis would provide useful negative controls in these experiments.

The FRET experiments described above would provide a means to validate if Lphns dimerise or oligomerise in cells in the presence of their ligands. To elucidate the functional significance of this it would be important to interpret this information

alongside data from cell signalling assays (such as GPCR signalling assays). As a first step it will be important to characterise whether Lphn elicits downstream signals in response to application of soluble ectodomains of any of its ligands discussed in this thesis. Indeed, the identification of an endogenous Lphn-activating ligand would mark a substantial step forward in our understanding of this receptor. Again, many of the mutants discussed in this thesis will provide important negative controls in these assays. Lphns have been shown to signal by modulating intracellular cAMP and IP₃ levels [142, 178, 225]. However, the recent demonstration that aGPCRs activate distinct downstream pathways depending on their mechanism of activation means that it will be important to assess ligand-induced downstream Lphn signalling using a variety of readouts [117].

7.4 Lphn Complexes During Evolution

The elucidation of the crystal structure of a large part of the Ten2 ectodomain reveals the ancient evolutionary origins of this receptor. The Ten-Lphn interaction is currently the only known example of a transsynaptic adhesive receptor pairing common to both vertebrates and invertebrates [280]. Unc5 proteins are also conserved between vertebrates and invertebrates. In comparison, FLRTs are a relatively new evolutionary invention and are vertebrate-specific. Interestingly, invertebrate Lphn lacks an extracellular Olf domain (the domain responsible for FLRT-binding), whereas the Lec domain (the Teneurin-binding domain) is conserved. FLRT-binding by vertebrate Lphn endows Lphn with the ability to not only recognise FLRTs, but also Unc5s, despite being unable to directly bind to these proteins. The increased complexity of the Lphn ectodomain in vertebrates may be an indication of its role as a context-recogniser, reflecting a requirement to recognise a broader range of molecular contexts in vertebrates.

7.5 Summary of Findings in this Thesis

Lphn is an evolutionarily-ancient aGPCR which interacts with several neural guidance molecules and has important roles in synaptogenesis. In this thesis, I revealed that the vertebrate-specific protein FLRT binds to Lphn via a conserved and bi-functional patch on the Lphn Olf domain. Further I revealed that FLRT2 is able to recruit Unc5D *in cis* and modulate its functional response to Lphn. Indeed, crystal structures and native mass spectrometry demonstrate that Lphn3-FLRT2-Unc5D form a triple complex with an unusual 4:2:2 stoichiometry, but other FLRT and Unc5 homologues form complexes of half the size (2:1:1 stoichiometry). Although another binding partner, Teneurin, binds to Lphn at a distinct site, Teneurin seems unable to bind to the Lphn3-FLRT2-Unc5D ternary complex. The crystal structure of a large fragment of the Ten2 ectodomain reveals the ancient evolutionary origins of this receptor. Together, these results provide strong molecular evidence for a role of Lphn as a context-recogniser by its ability to bind diverse ligands in distinct combinations and variable stoichiometries.

Bibliography

- [1] R. Abdulhussein, D. H. H. Koo, and W. F. Vogel. Identification of disulfide-linked dimers of the receptor tyrosine kinase DDR1. *Journal of Biological Chemistry*, 283(18):12026–33, 2008.
- [2] R. Abdulhussein, C. McFadden, P. Fuentes-Prior, and W. F. Vogel. Exploring the collagen-binding site of the DDR1 tyrosine kinase receptor. *Journal of Biological Chemistry*, 279(30):31462–70, 2004.
- [3] J. Abe, H. Suzuki, M. Notoya, T. Yamamoto, and S. Hirose. Ig-Hepta, a novel member of the G protein-coupled Hepta-helical receptor (GPCR) family that has immunoglobulin-like repeats in a long N-terminal extracellular domain and defines a new subfamily of GPCRs. *Journal of Biological Chemistry*, 274(28):19957–19964, 1999.
- [4] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221, 2010.
- [5] R. H. Adams and A. Eichmann. Axon Guidance Molecules in Vascular Patterning. *Cold Spring Harbor Perspectives in Biology*, 2(5):a001875, 2010.
- [6] J. Agirre, J. Iglesias-Fernández, C. Rovira, G. J. Davies, K. S. Wilson, and K. D. Cowtan. Privateer: software for the conformational validation of carbo-

- hydrate structures. *Nature Structural & Molecular Biology*, 22(11):833–834, 2015.
- [7] A. Al Chawaf, K. St. Amant, D. Belsham, and D. A. Lovejoy. Regulation of neurite growth in immortalized mouse hypothalamic neurons and rat hippocampal primary cultures by teneurin C-terminal-associated peptide-1. *Neuroscience*, 144:1241–1254, 2007.
- [8] A. Al Chawaf, K. Xu, L. Tan, F. J. Vaccarino, D. A. Lovejoy, and S. Rotzinger. Corticotropin-releasing factor (CRF)-induced behaviors are modulated by intravenous administration of teneurin C-terminal associated peptide-1 (TCAP-1). *Peptides*, 28(7):1406–1415, 2007.
- [9] A. Alkelai, T. Olender, R. Haffner-Krausz, M. M. Tsoory, V. Boyko, P. Tatarsky, R. Gross-Isseroff, R. Milgrom, S. Shushan, I. Blau, E. Cohn, R. Beerli, E. Levy-Lahad, E. Pras, and D. Lancet. A role for TENM1 mutations in congenital general anosmia. *Clinical Genetics*, 90(3):211–219, 2016.
- [10] P. Antinucci, N. Nikolaou, M. P. Meyer, and R. Hindges. Teneurin-3 specifies morphological and functional connectivity of retinal ganglion cells in the vertebrate visual system. *Cell Reports*, 5(3):582–92, 2013.
- [11] D. Araç, A. A. Boucard, M. F. Bolliger, J. Nguyen, S. M. Soltis, T. C. Südhof, and A. T. Brunger. A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *The EMBO Journal*, 31(6):1364–78, 2012.
- [12] T. Arakawa and J. Wen. Determination of carbohydrate contents from excess light scattering. *Analytical Biochemistry*, 299(2):158–61, 2001.
- [13] M. Arcos-Burgos, M. Jain, M. T. Acosta, S. Shively, H. Stanescu, D. Wallis, S. Domené, J. I. Vélez, J. D. Karkera, J. Balog, K. Berg, R. Kleta, W. A. Gahl, E. Roessler, R. Long, J. Lie, D. Pineda, A. C. Londoño, J. D. Palacio, A. Arbelaez, F. Lopera, J. Elia, H. Hakonarson, S. Johansson, P. M. Knappskog, J. Haavik, M. Ribases, B. Cormand, M. Bayes, M. Casas, J. A. Ramos-

- Quiroga, A. Hervas, B. S. Maher, S. V. Faraone, C. Seitz, C. M. Freitag, H. Palmason, J. Meyer, M. Romanos, S. Walitza, U. Hemminger, A. Warnke, J. Romanos, T. Renner, C. Jacob, K.-P. Lesch, J. Swanson, A. Vortmeyer, J. E. Bailey-Wilson, F. X. Castellanos, and M. Muenke. A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication. *Molecular Psychiatry*, 15(11):1053–66, 2010.
- [14] A. R. Aricescu, W. Lu, and E. Y. Jones. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica. Section D, Biological Crystallography*, 62:1243–50, 2006.
- [15] C. Bagutti, G. Forro, J. Ferralli, B. Rubin, and R. Chiquet-Ehrismann. The intracellular domain of teneurin-2 has a nuclear function and represses zic-1-mediated transcription. *Journal of Cell Science*, 116(14):2957–2966, 2003.
- [16] S. Baumgartner and R. Chiquet-Ehrismann. Tena, a Drosophila gene related to tenascin, shows selective transcript localization. *Mechanisms of Development*, 40(3):165–176, 1993.
- [17] S. Baumgartner, D. Martin, C. Hagios, and R. Chiquet-Ehrismann. tenm, a Drosophila gene related to tenascin, is a new pair-rule gene. *The EMBO Journal*, 13(16):3728–3740, 1994.
- [18] J. Beckmann, R. Schubert, R. Chiquet-Ehrismann, and D. J. Müller. Deciphering Teneurin Domains That Facilitate Cellular Recognition, Cell-Cell Adhesion, and Neurite Outgrowth Using Atomic Force Microscopy-Based Single-Cell Force Spectroscopy. *Nano Letters*, 13(6):2937–46, 2013.
- [19] T. Ben-zur, E. Feige, B. Motro, and R. Wides. The Mammalian Odz Gene Family : Homologs of a Drosophila Pair-Rule Gene with Expression Implying Distinct yet Overlapping Developmental Roles. *Developmental Biology*, 120:107–120, 2000.

- [20] T. Bergfors. Seeds to crystals. *Journal of Structural Biology*, 142:66–76, 2003.
- [21] A. Bernet, L. Mazelin, M. M. Coissieux, N. Gadot, S. L. Ackerman, J. Y. Scoazec, and P. Mehlen. Inactivation of the UNC5C Netrin-1 Receptor Is Associated With Tumor Progression in Colorectal Malignancies. *Gastroenterology*, 133(6):1840–1848, 2007.
- [22] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli, and T. Schwede. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42:W252–W258, 2014.
- [23] T. K. Bjarnadóttir, R. Fredriksson, P. J. Höglund, D. E. Gloriam, M. C. Lagerström, and H. B. Schiöth. The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. *Genomics*, 84(1):23–33, 2004.
- [24] M. F. Bolliger, D. C. Martinelli, and T. C. Südhof. The cell-adhesion G protein-coupled receptor BAI3 is a high-affinity receptor for C1q-like proteins. *Proceedings of the National Academy of Sciences*, 108(6):2534–9, 2011.
- [25] M. C. Bonaglia, S. Marelli, F. Novara, S. Commodaro, R. Borgatti, G. Minardo, L. Memo, E. Mangold, S. Beri, C. Zucca, D. Brambilla, M. Molteni, R. Giorda, R. G. Weber, and O. Zuffardi. Genotype–phenotype relationship in three cases with overlapping 19p13.12 microdeletions. *European Journal of Human Genetics*, 18115(10):1302–1309, 2010.
- [26] R. T. Böttcher, N. Pollet, H. Delius, and C. Niehrs. The transmembrane protein XFLRT3 forms a complex with FGF receptors and promotes FGF signalling. *Nature Cell Biology*, 6(1):38–44, 2004.
- [27] A. A. Boucard, J. Ko, and T. C. Südhof. High Affinity Neurexin Binding to Cell Adhesion G-protein-coupled Receptor CIRL1/Latrophilin-1 Produces an Intercellular Adhesion Complex. *Journal of Biological Chemistry*, 287(12):9399–9413, 2012.

- [28] A. A. Boucard, S. Maxeiner, and T. C. Südhof. Latrophilins Function as Heterophilic Cell-adhesion Molecules by Binding to Teneurins: REGULATION BY ALTERNATIVE SPLICING. *Journal of Biological Chemistry*, 289(1):387–402, 2014.
- [29] S. Boyden, A. Desai, G. Cruse, M. Young, H. Bolan, L. Scott, A. Eisch, R. Long, C.-C. Lee, C. Satorius, A. Pakstis, A. Olivera, J. Mullikin, E. Chouery, A. Megerbane, M. Medlej-Hashim, K. Kidd, D. Kastner, D. Metcalfe, and H. Komarow. Vibratory Urticaria Associated with a Missense Variant in ADGRE2. *The New England Journal of Medicine*, 7374:656–663, 2016.
- [30] W. T. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart OS, Vonrhein C. BUSTER version 2.10.0, 2011.
- [31] A. T. Brünger. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355(6359):472–475, 1992.
- [32] A. T. Brünger. Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallographica Section D Biological Crystallography*, 49(1):24–36, 1993.
- [33] D. W. A. Buchan, F. Minneci, T. C. O. Nugent, K. Bryson, and D. T. Jones. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, 41(W1):W349–W357, 2013.
- [34] J. N. Busby, S. Panjikar, M. J. Landsberg, M. R. H. Hurst, and J. S. Lott. The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature*, 501(7468):547–50, 2013.
- [35] O. P. Carr, K. A. Glendining, C. A. Leamey, and L. R. Marotte. Overexpression of Ten-m3 in the retina alters ipsilateral retinocollicular projections in the wallaby (*Macropus eugenii*). *International Journal of Developmental Neuroscience*, 31(7):496–504, 2013.

- [36] G. Chai, A. M. Goffinet, and F. Tissir. Celsr3 and Fzd3 in axon guidance. *The International Journal of Biochemistry & Cell Biology*, 64:11–14, 2015.
- [37] D. Chand, C. A. Casatti, L. de Lannoy, L. Song, A. Kollara, D. Barsyte-Lovejoy, T. J. Brown, and D. A. Lovejoy. C-terminal processing of the teneurin proteins: Independent actions of a teneurin C-terminal associated peptide in hippocampal cells. *Molecular and Cellular Neuroscience*, 52:38–50, 2013.
- [38] V. T. Chang, M. Crispin, A. R. Aricescu, D. J. Harvey, J. E. Nettleship, J. A. Fennelly, C. Yu, K. S. Boles, E. J. Evans, D. I. Stuart, R. a. Dwek, E. Y. Jones, R. J. Owens, and S. J. Davis. Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure*, 15(3):267–73, 2007.
- [39] C. R. Chen and G. I. Makhatadze. ProteinVolume: calculating molecular van der Waals and void volumes in proteins. *BMC Bioinformatics*, 16(101), 2015.
- [40] B. Chih, H. Engelman, and P. Scheiffele. Control of Excitatory and Inhibitory Synapse Formation by Neuroligins. *Science*, 307(5713):1324–1328, 2005.
- [41] K. Chung and K. Deisseroth. CLARITY for mapping the nervous system. *Nature Methods*, 10(6):508–513, 2013.
- [42] M. M. Coissieux, J. Tomsic, M. Castets, H. Hampel, S. Tuupanen, N. Andrieu, I. Comeras, Y. Drouet, C. Lasset, S. Liyanarachchi, L. Mazelin, A. Puisieux, J. Saurin, J. Scoazec, Q. Wang, L. Aaltonen, S. M. Tanner, A. De La Chapelle, A. Bernet, and P. Mehlen. Variants in the netrin-1 receptor UNC5C prevent apoptosis and increase risk of familial colorectal cancer. *Gastroenterology*, 141(6):2039–2046, 2011.
- [43] K. Cowtan. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D Biological Crystallography*, 62:1002–1011, 2006.
- [44] K. Cowtan. Recent developments in classical density modification. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):470–478, 2010.

- [45] M. Crispin, D. J. Harvey, V. T. Chang, C. Yu, A. R. Aricescu, E. Y. Jones, S. J. Davis, R. A. Dwek, and P. M. Rudd. Inhibition of hybrid-and complex-type glycosylation reveals the presence of the GlcNAc transferase I-independent fucosylation pathway. *Glycobiology*, 16(8):748–756, 2006.
- [46] S. Das, K. A. Owen, K. T. Ly, D. Park, S. G. Black, J. M. Wilson, C. D. Sifri, K. S. Ravichandran, P. B. Ernst, and J. E. Casanova. Brain angiogenesis inhibitor 1 (BAI1) is a pattern recognition receptor that mediates macrophage binding and engulfment of Gram-negative bacteria. *Proceedings of the National Academy of Sciences*, 108(5):2136–41, 2011.
- [47] J. Q. Davies, G. W. Chang, S. Yona, S. Gordon, M. Stacey, and H. H. Lin. The role of receptor oligomerization in modulating the expression and function of leukocyte adhesion-G protein-coupled receptors. *Journal of Biological Chemistry*, 282(37):27343–27353, 2007.
- [48] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, and D. C. Richardson. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, 35:W375–W383, 2007.
- [49] B. A. Davletov, O. G. Shamotienko, V. G. Lelianova, E. V. Grishin, and Y. A. Ushkaryov. Isolation and Biochemical Characterization of a Ca^{2+} - independent α -Latrotoxin-binding Protein. *Journal of Biological Chemistry*, 271(38):23239–23245, 1996.
- [50] D. del Toro, T. Ruff, E. Cederfjäll, A. Villalba, G. Seyit-Bremer, V. Borrell, and R. Klein. Regulation of Cerebral Cortex Folding by Controlling Neuronal Migration via FLRT Adhesion Molecules. *Cell*, 169(4):621–635.e16, 2017.
- [51] L. M. Demberg, S. Rothmund, T. Schoneberg, and I. Liebscher. Identification of the tethered peptide agonist of the adhesion G protein-coupled receptor

- GPR64/ADGRG2. *Biochemical and Biophysical Research Communications*, 464:743–747, 2015.
- [52] P. Derosse, T. Lencz, K. E. Burdick, S. G. Siris, J. M. Kane, and A. K. Malhotra. The Genetics of Symptom-Based Phenotypes: Toward a Molecular Classification of Schizophrenia. *Schizophrenia Bulletin*, 34(6):1047–1053, 2008.
- [53] K. Diederichs and P. A. Karplus. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Biology*, 4(4):269–75, 1997.
- [54] K. Diederichs and P. A. Karplus. Better models by discarding data? *Acta Crystallographica Section D Biological Crystallography*, 69:1215–1222, 2013.
- [55] S. Domené, H. Stanescu, D. Wallis, B. Tinloy, D. E. Pineda, R. Kleta, M. Arcos-Burgos, E. Roessler, and M. Muenke. Screening of human LPHN3 for variants with a potential impact on ADHD susceptibility. *American Journal of Medical Genetics Part B*, 156(1):11–8, 2011.
- [56] C. Dominici, J. A. Moreno-Bravo, S. R. Puiggros, Q. Rappeneau, N. Rama, P. Vieugue, A. Bernet, P. Mehlen, and A. Chédotal. Floor-plate-derived netrin-1 is dispensable for commissural axon guidance. *Nature*, 545(7654):350–354, 2017.
- [57] S. E. Doyle, M. J. Scholz, K. A. Greer, A. D. Hubbard, D. K. Darnell, P. B. Antin, S. E. Klewer, and R. B. Runyan. Latrophilin-2 is a novel component of the epithelial-mesenchymal transition within the atrioventricular canal of the embryonic chicken heart. *Developmental Dynamics*, 235(12):3213–3221, 2006.
- [58] K. Drabikowski, A. Trzebiatowska, and R. Chiquet-Ehrismann. *ten-1*, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in *Caenorhabditis elegans*. *Developmental Biology*, 282(1):27–38, 2005.

- [59] I. Dudanova and R. Klein. Integration of guidance cues: Parallel signaling and crosstalk. *Trends in Neurosciences*, 36(5):295–304, 2013.
- [60] T. A. Edwards, B. D. Wilkinson, R. P. Wharton, and A. K. Aggarwal. Model of the brain tumor-Pumilio translation repressor complex. *Genes & Development*, 17(20):2508–13, 2003.
- [61] J. Egea, C. Erlacher, E. Montanez, I. Burtscher, S. Yamagishi, M. Hess, F. Hampel, R. Sanchez, M. T. Rodriguez-Manzanique, M. R. Bösl, R. Fässler, H. Lickert, and R. Klein. Genetic ablation of FLRT3 reveals a novel morphogenetic function for the anterior visceral endoderm in suppressing mesoderm differentiation. *Genes & Development*, 22(23):3349–62, 2008.
- [62] E. H. Egelman. The Current Revolution in Cryo-EM. *Biophysical Journal*, 110(5):1008–1012, 2016.
- [63] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*, D60:2126–2132, 2004.
- [64] T. Eneqvist, E. Lundberg, L. Nilsson, R. Abagyan, and A. E. Sauer-Eriksson. The transthyretin-related protein family. *European Journal of Biochemistry*, 270:518–532, 2003.
- [65] P. Englebienne, A. V. Hoonacker, and M. Verhas. Surface plasmon resonance: principles, methods and applications in biomedical sciences. *Spectroscopy*, 17:255–273, 2003.
- [66] P. Evans. Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography*, 62:72–82, 2006.
- [67] A. Faiz, C. Donovan, M. A. Nieuwenhuis, M. van den Berge, D. S. Postma, S. Yao, C. Y. Park, R. Hirsch, J. J. Fredberg, G. Tjin, A. J. Halayko, K. L. Rempel, J. P. T. Ward, T. Lee, Y. Bossé, D. C. Nickle, M. Obeidat, J. M.

- Vonk, J. L. Black, B. G. Oliver, R. Krishnan, B. McParland, J. E. Bourke, and J. K. Burgess. Latrophilin receptors: novel bronchodilator targets in asthma. *Thorax*, 72(1):74–82, 2017.
- [68] A. Fazeli, S. L. Dickinson, M. L. Hermiston, R. V. Tighe, R. G. Steen, C. G. Small, E. T. Stoeckli, K. Keino-Masu, M. Masu, H. Rayburn, J. Simons, R. T. Bronson, J. I. Gordon, M. Tessier-Lavigne, and R. A. Weinberg. Phenotype of mice lacking functional Deleted in colorectal cancer (Dec) gene. *Nature*, 386(6627):796–804, 1997.
- [69] K. Feng, X.-H. Zhou, T. Oohashi, M. Mörgelin, A. Lustig, S. Hirakawa, Y. Nomiya, J. Engel, U. Rauch, and R. Fässler. All four members of the Tenm/Odz family of transmembrane proteins form dimers. *Journal of Biological Chemistry*, 277(29):26128–35, 2002.
- [70] K. Flintoff, Y. Arudchelvan, and S.-G. Gong. FLRT2 Interacts With Fibronectin in the ATDC5 Chondroprogenitor Cells. *Journal of Cellular Physiology*, 229(10):1538–1547, 2014.
- [71] J. Foadi, P. Aller, Y. Alguel, A. Cameron, D. Axford, R. L. Owen, W. Armour, D. G. Waterman, S. Iwata, and G. Evans. Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*, 69:1617–32, 2013.
- [72] S. A. Forbes, N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton, and P. A. Futreal. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Research*, 39:D945–950, 2011.
- [73] C. Forcet, E. Stein, L. Pays, V. Corset, F. Llambi, M. Tessier-Lavigne, and

- P. Mehlen. Netrin-1-mediated axon outgrowth requires deleted in colorectal cancer-dependent MAPK activation. *Nature*, 417(6887):443–447, 2002.
- [74] R. Fredriksson, M. C. Lagerst Om, L.-G. Lundin, and H. B. Sch Oth. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralagon Groups, and Fingerprints. *Molecular Pharmacology*, 63(6):1256–1272, 2003.
- [75] T. Fukuzawa, J. Ishida, A. Kato, T. Ichinose, D. M. Ariestanti, T. Takahashi, K. Ito, J. Abe, T. Suzuki, S. Wakana, A. Fukamizu, N. Nakamura, and S. Hirose. Lung Surfactant Levels are Regulated by Ig-Hepta/GPR116 by Monitoring Surfactant Protein D. *PLoS ONE*, 8(7):e69451, 2013.
- [76] J. Garcia-Pardo, R. Graña-Montes, M. Fernandez-Mendez, A. Ruyra, N. Rother, F. X. Aviles, J. Lorenzo, and S. Ventura. Amyloid formation under native conditions 1 Amyloid Formation by Human Carboxypeptidase D Transthyretin-like Domain Under Physiological Conditions. *Journal of Biological Chemistry*, 289:33783–33796, 2014.
- [77] E. Garman and J. W. Murray. Heavy-atom derivatization. *Acta Crystallographica Section D Biological Crystallography*, D59:1903–1913, 2003.
- [78] E. F. Garman. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica Section D Biological Crystallography*, 66:339–351, 2010.
- [79] N. Giagtzoglou, C. V. Ly, and H. J. Bellen. Cell Adhesion, the Backbone of the Synapse: “Vertebrate” and “Invertebrate” Perspectives. *Cold Spring Harbor Perspectives in Biology*, 1:a003079, 2009.
- [80] F. Glaser, T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, and N. Ben-Tal. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, 19(1):163–4, 2003.

- [81] F. Gorrec. The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography*, 42:1035–1042, 2009.
- [82] E. R. Graf, X. Zhang, S. X. Jin, M. W. Linhoff, and A. M. Craig. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell*, 119(7):1013–1026, 2004.
- [83] M. Grandin, M. Meier, J. G. Delcros, D. Nikodemus, R. Reuten, T. R. Patel, D. Goldschneider, G. Orriss, N. Krahn, A. Boussouar, R. Abes, Y. Dean, D. Neves, A. Bernet, S. Depil, F. Schneiders, K. Poole, R. Dante, M. Koch, P. Mehlen, and J. Stetefeld. Structural Decoding of the Netrin-1-UNC5 Interaction and its Therapeutical Implications in Cancers. *Cancer Cell*, 29:173–185, 2015.
- [84] V. Gurevich and E. Gurevich. How and why do GPCRs dimerize? *Trends in Pharmacological Sciences*, 29(5):234–240, 2008.
- [85] B. P. Haines, L. M. Wheldon, D. Summerbell, J. K. Heath, and P. W. J. Rigby. Regulated expression of FLRT genes implies a functional role in the regulation of FGF signalling during mouse development. *Developmental Biology*, 297:14–25, 2006.
- [86] J. Hamann, G. Aust, D. Araç, F. B. Engel, C. Formstone, R. Fredriksson, R. A. Hall, B. L. Harty, C. Kirchhoff, B. Knapp, A. Krishnan, I. Liebscher, H.-H. Lin, D. C. Martinelli, K. R. Monk, M. C. Peeters, X. Piao, S. Prömel, T. Schöneberg, T. W. Schwartz, K. Singer, M. Stacey, Y. A. Ushkaryov, M. Vallon, U. Wolfrum, M. W. Wright, L. Xu, T. Langenhan, and H. B. Schiöth. International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G Protein-Coupled Receptors. *Pharmacological Reviews*, 67:338–367, 2015.
- [87] J. Hamann, C. Stortelers, E. Kiss-Toth, B. Vogel, W. Eichler, and R. A. van Lier. Characterization of the CD55 (DAF)-binding site on the seven-span

- transmembrane receptor CD97. *European Journal of Immunology*, 28(5):1701–1707, 1998.
- [88] J. Hamann, B. Vogel, G. M. van Schijndel, and R. A. van Lier. The seven-span transmembrane receptor CD97 has a cellular ligand (CD55, DAF). *Journal of Experimental Medicine*, 184(3):1185–1189, 1996.
- [89] B. Z. Harris and W. A. Lim. Mechanism and role of PDZ domains in signaling complex assembly. *Journal of Cell Science*, 114(18):3219–3231, 2001.
- [90] D. Hattori, Y. Chen, B. J. Matthews, L. Salwinski, C. Sabatti, W. B. Grueber, and S. L. Zipursky. Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature*, 461:644–649, 2009.
- [91] J. S. Hayflick. A Family of Heptahelical Receptors With Adhesion- Like Domains: A Marriage Between Two Super Families. *J. OF RECEPTOR & SIGNAL TRANSDUCTION RESEARCH*, 20(3):119–131, 2000.
- [92] E. G. Healey, B. Bishop, J. Elegheert, C. H. Bell, S. Padilla-Parra, and C. Siebold. Repulsive guidance molecule is a structural bridge between neogenin and bone morphogenetic protein. *Nature Structural & Molecular Biology*, 22(6):458–465, 2015.
- [93] H. Hernández and C. V. Robinson. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols*, 2(3):715–726, 2007.
- [94] C. W. Hill, C. H. Sandt, and D. A. Vlazny. Rhs elements of Escherichia coli: A family of genetic composites each encoding a large mosaic protein. *Molecular Microbiology*, 12(6):865–871, 1994.
- [95] L. Holm and P. Rosenström. Dali server: conservation mapping in 3D. *Nucleic Acids Research*, 38:W545–9, 2010.

- [96] K. Hong, L. Hinck, M. Nishiyama, M. M. Poo, M. Tessier-Lavigne, E. Stein, T. Daniel, J. Lecocq, M. Masu, H. Rayburn, and E. Al. A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. *Cell*, 97(7):927–41, 1999.
- [97] W. Hong, T. J. Mosca, and L. Luo. Teneurins instruct synaptic partner matching in an olfactory map. *Nature*, 484(7393):201–207, 2012.
- [98] K. Ichtenko, M. A. Bittner, V. Krasnoperov, A. R. Little, O. Chepurny, R. W. Holz, and A. G. Petrenko. A Novel Ubiquitously Expressed α -Latrotoxin Receptor Is a Member of the CIRL Family of G-protein-coupled Receptors. *Journal of Biological Chemistry*, 274(9):5491–5498, 1999.
- [99] S. A. Islam, D. Carvin, M. J. Sternberg, and T. L. Blundell. HAD, a data bank of heavy-atom binding sites in protein crystals: a resource for use in multiple isomorphous replacement and anomalous scattering. *Acta Crystallographica Section D Biological Crystallography*, 54:1199–206, 1998.
- [100] L. M. Iyer, D. Zhang, I. B. Rogozin, and L. Aravind. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Research*, 39(22):9437–9497, 2011.
- [101] A. P. Jackson, G. H. Thomas, J. Parkhill, and N. R. Thomson. Evolutionary diversification of an ancient gene family (rhs) through C-terminal displacement. *BMC Genomics*, 2009.
- [102] J. Jancarik and S. H. Kim. Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of Applied Crystallography*, 24(4):409–411, 1991.
- [103] B. Janssen, R. Robinson, F. Perez-Branguli, C. Bell, K. Mitchell, C. Siebold,

- and E. Jones. Structural basis of semaphorin-plexin signalling. *Nature*, 467:1118–1124, 2010.
- [104] W. Kabsch. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):133–144, 2010.
- [105] W. Kabsch. XDS. *Acta Crystallographica Section D Biological Crystallography*, 66:125–132, 2010.
- [106] W. Kakegawa, N. Mitakidis, E. Miura, M. Abe, K. Matsuda, Y. H. Takeo, K. Kohda, J. Motohashi, A. Takahashi, S. Nagao, S.-i. Muramatsu, M. Watanabe, K. Sakimura, A. Radu Aricescu, and M. Yuzaki. Anterograde C1ql1 Signaling Is Required in Order to Determine and Maintain a Single-Winner Climbing Fiber in the Mouse Cerebellum. *Neuron*, 85:316–329, 2015.
- [107] E. Karaulanov, R. T. Böttcher, P. Stannek, W. Wu, M. Rau, S. Ogata, K. W. Y. Cho, and C. Niehrs. Unc5B interacts with FLRT3 and Rnd1 to modulate cell adhesion in *Xenopus* embryos. *PloS One*, 4(5):e5742, 2009.
- [108] E. E. Karaulanov, R. T. Böttcher, and C. Niehrs. A role for fibronectin-leucine-rich transmembrane cell-surface proteins in homotypic cell adhesion. *EMBO Reports*, 7(3):283–90, 2006.
- [109] K. Karplus, P. Andrew, Diederichs. Linking Crystallographic Model and Data Quality. *Science*, 336(6084):1030–1033, 2012.
- [110] O. N. Karpus, H. Veninga, R. M. Hoek, D. Flierman, J. D. van Buul, C. C. VandenAkker, E. VanBavel, M. E. Medof, R. A. W. van Lier, K. A. Reedquist, and J. Hamann. Shear Stress-Dependent Downregulation of the Adhesion-G Protein-Coupled Receptor CD97 on Circulating Leukocytes upon Contact with Its Ligand CD55. *The Journal of Immunology*, 199(6), 2013.
- [111] B. Kaur, S. M. Cork, E. M. Sandberg, N. S. Devi, Z. Zhang, P. A. Klenotic, M. Febbraio, H. Shim, H. Mao, C. Tucker-Burden, R. L. Silverstein, D. J.

- Brat, J. J. Olson, and E. G. Van Meir. Vasculostatin Inhibits Intracranial Glioma Growth and Negatively Regulates In vivo Angiogenesis through a CD36-Dependent Mechanism. *Cancer Research*, 69(3), 2009.
- [112] K. Keino-Masu, M. Masu, L. Hinck, E. Leonardo, S. S.-Y. Chan, J. G. Culotti, and M. Tessier-Lavigne. Deleted in Colorectal Cancer (DCC) Encodes a Netrin Receptor. *Cell*, 87(2):175–185, 1996.
- [113] T. E. Kennedy, T. Serafini, J. de la Torre, and M. Tessier-Lavigne. Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell*, 78(3):425–435, 1994.
- [114] D. Kenzelmann, R. Chiquet-Ehrismann, N. T. Leachman, and R. P. Tucker. Teneurin-1 is expressed in interconnected regions of the developing brain and is processed in vivo. *BMC Developmental Biology*, 8(30), 2008.
- [115] T. Kidd, K. S. Bland, C. S. Goodman, M. Tessier-Lavigne, C. Goodman, W. Henzel, M. Tessier-Lavigne, and T. Kidd. Slit Is the Midline Repellent for the Robo Receptor in Drosophila. *Cell*, 96(6):785–94, 1999.
- [116] N. King, M. J. Westbrook, S. L. Young, A. Kuo, M. Abedin, J. Chapman, S. Fairclough, U. Hellsten, Y. Isogai, I. Letunic, M. Marr, D. Pincus, N. Putnam, A. Rokas, K. J. Wright, R. Zuzow, W. Dirks, M. Good, D. Goodstein, D. Lemons, W. Li, J. B. Lyons, A. Morris, S. Nichols, D. J. Richter, A. Salamov, J. Sequencing, P. Bork, W. A. Lim, G. Manning, W. T. Miller, W. McGinnis, H. Shapiro, R. Tjian, I. V. Grigoriev, and D. Rokhsar. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature*, 451:783–788, 2008.
- [117] A. Kishore, R. H. Purcell, Z. Nassiri-Toosi, and R. A. Hall. Stalk-dependent and stalk-independent signaling by the adhesion G protein-coupled receptors GPR56 (ADGRG1) and BAI1 (ADGRB1). *Journal of Biological Chemistry*, 291(7):3385–3394, 2016.

- [118] C. Kleanthous, U. Kuhlmann, A. Pommer, N. Ferguson, S. Radford, G. Moore, R. James, and A. Hemmings. Structural and mechanistic basis of immunity toward endonuclease colicins. *Nature Structural Biology*, 6(3):243–252, 1999.
- [119] R. Klein. Eph/ephrin signalling during development. *Development*, 139(22):4105–4109, 2012.
- [120] B. Knöll, C. Weinl, A. Nordheim, and F. Bonhoeffer. Stripe assay to examine axonal guidance and cell migration. *Nature Protocols*, 2(5):1216–1224, 2007.
- [121] J. T. Koh, H. Kook, H. J. Kee, Y.-W. Seo, B. C. Jeong, J. H. Lee, M.-Y. Kim, K. C. Yoon, S. Jung, and K. K. Kim. Extracellular fragment of brain-specific angiogenesis inhibitor 1 suppresses endothelial cell proliferation by blocking $\alpha\beta 5$ integrin. *Experimental Cell Research*, 294(1):172–184, 2004.
- [122] A. L. Kolodkin and M. Tessier-Lavigne. Mechanisms and Molecules of Neuronal Wiring: A Primer. *Cold Spring Harbor Perspectives in Biology*, 3(6):a001727, 2010.
- [123] Y. Kong, B. J. Janssen, T. Malinauskas, V. R. Vangoor, C. H. Coles, R. Kaufmann, T. Ni, R. J. Gilbert, S. Padilla-Parra, R. J. Pasterkamp, and E. Y. Jones. Structural Basis for Plexin Activation and Regulation. *Neuron*, 91(3):548–560, 2016.
- [124] S. Koskiniemi, J. G. Lamoureux, K. C. Nikolakakis, C. T ’kint De Roodenbeke, M. D. Kaplan, D. A. Low, and C. S. Hayes. Rhs proteins from diverse bacteria mediate intercellular competition. *Proceedings of the National Academy of Sciences*, 110(17):7032–7037, 2013.
- [125] V. Krasnoperov, Y. Lu, L. Buryanovsky, T. A. Neubert, K. Ichchenko, and A. G. Petrenko. Post-translational Proteolytic Processing of the Calcium-independent Receptor of α -Latrotoxin (CIRL), a Natural Chimera of the Cell Adhesion Protein and the G Protein-coupled Receptor. *Journal of Biological Chemistry*, 277(48):46518–46526, 2002.

- [126] H.-J. Kreienkamp, H. Zitzer, E. D. Gundelfinger, D. Richter, and T. M. Böcker. The Calcium-independent Receptor for α -Latrotoxin from Human and Rodent Brains Interacts with Members of the ProSAP/SSTRIP/Shank Family of Multidomain Proteins. *Journal of Biological Chemistry*, 275(42), 2000.
- [127] A. Krupa and N. Srinivasan. Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes. *BMC Genomics*, 6129, 2005.
- [128] A. Küffer, A. K. K. Lakkaraju, A. Mogha, S. C. Petersen, K. Airich, C. Doucerain, R. Marpakwar, P. Bakirci, A. Senatore, A. Monnard, C. Schiavi, M. Nuvolone, B. Grosshans, S. Hornemann, F. Bassilana, K. R. Monk, and A. Aguzzi. The prion protein is an agonistic ligand of the G protein-coupled receptor Adgrg6. *Nature*, 536(7617):1–21, 2016.
- [129] V. L. Kung, S. Khare, C. Stehlik, E. M. Bacon, A. J. Hughes, and A. R. Hauser. An rhs gene of *Pseudomonas aeruginosa* encodes a virulence protein that activates the inflammasome. *Proceedings of the National Academy of Sciences*, 109(4):1275–1280, 2012.
- [130] S. Küry, C. Garrec, F. Airaud, F. Breheret, V. Guibert, C. Frenard, S. Jiao, D. Bonneau, P. Berthet, C. Bossard, O. Ingster, E. Cauchin, and S. Bezieau. Evaluation of the colorectal cancer risk conferred by rare UNC5C alleles. *World Journal of Gastroenterology*, 20(1):204–213, 2014.
- [131] S. E. Lacy, C. G. Bönnemann, E. a. Buzney, and L. M. Kunkel. Identification of FLRT1, FLRT2, and FLRT3: a novel family of transmembrane leucine-rich repeat proteins. *Genomics*, 62(3):417–26, 1999.
- [132] M. Lange, W. Norton, M. Coolen, M. Chaminade, S. Merker, F. Proft, A. Schmitt, P. Vernier, K.-P. Lesch, and L. Bally-Cuif. The ADHD-susceptibility gene lphn3.1 modulates dopaminergic neuron formation and locomotor activity during zebrafish development. *Molecular Psychiatry*, 17(9):946–54, 2012.

- [133] T. Langenhan, G. Aust, and J. Hamann. Sticky signaling–adhesion class G protein-coupled receptors take the stage. *Science Signaling*, 6(276):re3, 2013.
- [134] T. Langenhan, X. Piao, and K. R. Monk. Adhesion G protein-coupled receptors in nervous system development and disease. *Nature Reviews Neuroscience*, 17:550–561, 2016.
- [135] T. Langenhan, S. Prömel, L. Mestek, B. Esmaili, H. Waller-Evans, C. Hennig, Y. Kohara, L. Avery, I. Vakonakis, R. Schnabel, and A. P. Russ. Latrophilin Signaling Links Anterior-Posterior Tissue Polarity and Oriented Cell Divisions in the *C. elegans* Embryo. *Developmental Cell*, 17(4):494–504, 2009.
- [136] T. Langenhan, Tobias, Schoneberg. *Adhesion G Protein-coupled Receptors Molecular, Physiological and Pharmacological Principles in Health and Disease*. Springer, 2016.
- [137] V. Lanoue, A. Usardi, S. Sigoillot, M. Talleur, K. Iyer, J. Mariani, P. Isope, G. Vodjdani, N. Heintz, and F. Selimi. The adhesion-GPCR BAI3, a gene linked to psychiatric disorders, regulates dendrite morphogenesis in neurons. *Molecular Psychiatry*, 18(8):943–950, 2013.
- [138] C. A. Leamey, S. Merlin, P. Lattouf, A. Sawatari, X. Zhou, N. Demel, K. A. Glendining, T. Oohashi, M. Sur, and R. Fässler. Ten_m3 Regulates Eye-Specific Patterning in the Mammalian Visual Pathway and Is Required for Binocular Vision. *PLoS Biology*, 5(9):2077–2092, 2007.
- [139] C. A. Leamey and A. Sawatari. The teneurins: New players in the generation of visual topography. *Seminars in Cell & Developmental Biology*, 35:173–179, 2014.
- [140] Y. Lei, H. Zhu, W. Yang, E. Ross, G. M. Shaw, R. H. Finnell, and O. El-Maarri. Identification of Novel CELSR1 Mutations in Spina Bifida. *PLoS ONE*, 9(3):e922207, 2014.

- [141] B. Leitingner. Molecular analysis of collagen binding by the human discoidin domain receptors, DDR1 and DDR2. Identification of collagen binding sites in DDR2. *Journal of Biological Chemistry*, 278(19):16761–16769, 2003.
- [142] V. G. Lelianaova, B. A. Davletov, A. Sterling, M. A. Rahman, E. V. Grishin, N. F. Totty, and Y. A. Ushkaryov. α -Latrotoxin Receptor, Latrophilin, Is a Novel Member of the Secretin Family of G Protein-coupled Receptors. *Journal of Biological Chemistry*, 272(34):21504–21508, 1997.
- [143] E. Leonardo, L. Hinck, M. Masu, K. Keino-Masu, S. Ackerman, and M. Tessier-Lavigne. Vertebrate homologues of C.elegans UNC-5 are candidate netrin receptors. *Nature*, 386(24):833–838, 1997.
- [144] A. Levine, A. Bashan-Ahrend, O. Budai-Hadrian, D. Gartenberg, S. Menasherow, and R. Wide. odd Oz: A Novel Drosophila Pair Rule Gene. *Cell*, 77:587–598, 1994.
- [145] E. Leyva-Díaz, D. Del Toro, M. J. Menal, S. Cambray, R. Susín, M. Tessier-Lavigne, R. Klein, J. Egea, and G. López-Bendito. FLRT3 Is a Robo1-Interacting Protein that Determines Netrin-1 Attraction in Developing Axons. *Current Biology*, 24(5):494–508, 2014.
- [146] S. Li, Z. Jin, S. Koirala, L. Bu, L. Xu, R. O. Hynes, C. A. Walsh, G. Corfas, and X. Piao. Neurobiology of Disease GPR56 Regulates Pial Basement Membrane Integrity and Cortical Lamination. *The Journal of Neuroscience*, 28(22):5817–5826, 2008.
- [147] H.-M. Liao, Y.-L. Chao, A.-L. Huang, M.-C. Cheng, Y.-J. Chen, K.-F. Lee, J.-S. Fang, C.-H. Hsu, and C.-H. Chen. Identification and characterization of three inherited genomic copy number variations associated with familial schizophrenia. *Schizophrenia Research*, 139:229–236, 2012.
- [148] I. Liebscher, J. Schön, S. C. Petersen, L. Fischer, N. Auerbach, L. M. Demberg, A. Mogha, M. Cöster, K.-U. Simon, S. Rothmund, K. R. Monk, and

- T. Schöneberg. A tethered agonist within the ectodomain activates the adhesion G protein-coupled receptors GPR126 and GPR133. *Cell Reports*, 9(6):2018–26, 2014.
- [149] A. Lilienbaum, A. Reszka, A. Horwitz, and C. Holt. Chimeric Integrins Expressed in Retinal Ganglion Cells Impair Process Outgrowth in Vivo. *Molecular and Cellular Neuroscience*, 6(2):139–152, 1995.
- [150] H.-H. Lin, G.-W. Chang, J. Q. Davies, M. Stacey, J. Harris, and S. Gordon. Autocatalytic Cleavage of the EMR2 Receptor Occurs at a Conserved G Protein-coupled Receptor Proteolytic Site Motif. *Journal of Biological Chemistry*, 279(30):31823–31832, 2004.
- [151] E. Little, P. Bork, and R. F. Doolittle. Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. *Journal of Molecular Evolution*, 39(6):631–643, 1994.
- [152] Q.-R. Liu, T. Drgon, C. Johnson, D. Walther, J. Hess, G. R. Uhl, Q.-R. Liu, T. Drgon, C. Johnson, and W. D. Hess. Addiction molecular genetics: 639,401 SNP whole genome association identifies many “cell adhesion” genes. *American Journal of Medical Genetics Part B*, 141(141), 2006.
- [153] F. Llambi, F. Calheiros-Lourenço, D. Gozuacik, C. Guix, L. Pays, G. D. Rio, A. Kimchi, and P. Mehlen. The dependence receptor UNC5H2 mediates apoptosis through DAP-kinase. *The EMBO Journal*, 24:1192–1201, 2005.
- [154] F. Llambi, F. Causeret, E. Bloch-Gallego, and P. Mehlen. Netrin-1 acts as a survival factor via its receptors UNC5H and DCC. *The EMBO Journal*, 20(11):2715–22, 2001.
- [155] I. Loedige, L. Jakob, T. Treiber, D. Ray, M. Stotz, N. Treiber, J. Hennig, K. B. Cook, Q. Morris, T. R. Hughes, J. C. Engelmann, M. P. Krahn, and G. Meister. The Crystal Structure of the NHL Domain in Complex with

- RNA Reveals the Molecular Basis of Drosophila Brain-Tumor-Mediated Gene Regulation. *Cell Reports*, 13(6):1206–1220, 2015.
- [156] F. Long, R. A. Nicholls, P. Emsley, S. Gražulis, A. Merkys, A. Vaitkus, and G. N. Murshudov. AceDRG: a stereochemical description generator for ligands. *Acta Crystallographica Section D Biological Crystallography*, 73:112–122, 2017.
- [157] D. A. Lovejoy, A. A. Chawaf, and M. A. Cadinouche. Teneurin C-terminal associated peptides: An enigmatic family of neuropeptides with structural similarity to the corticotropin-releasing factor and calcitonin families of peptides. *General and Comparative Endocrinology*, 148(3):299–305, 2006.
- [158] Y. C. Lu, O. V. Nazarko, R. Sando, G. S. Salzman, T. C. Südhof, and D. Araç. Structural Basis of Latrophilin-FLRT-UNC5 Interaction in Cell Adhesion. *Structure*, 23:1–14, 2015.
- [159] A. Luch. Mechanistic insights on spider neurotoxins. In A. Luch, editor, *Molecular, Clinical and Environmental Toxicology. Volume 2: Clinical Toxicology*, pages 293–315. Birkhäuser Basel, 2010.
- [160] R. Luo, S.-J. Jeong, Z. Jin, N. Strokes, S. Li, and X. Piao. G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *Proceedings of the National Academy of Sciences*, 108(31):12925–30, 2011.
- [161] S. Maretto, P.-S. Müller, a. R. Aricescu, K. W. Y. Cho, E. K. Bikoff, and E. J. Robertson. Ventral closure, headfold fusion and definitive endoderm migration defects in mouse embryos lacking the fibronectin leucine-rich transmembrane protein FLRT3. *Developmental Biology*, 318(1):184–93, 2008.
- [162] M. Maruyama, T. Kumagai, Y. Matoba, M. Hayashida, T. Fujii, Y. Hata, and M. Sugiyama. Crystal structures of the transposon Tn5-carried bleomycin resistance determinant uncomplexed and complexed with bleomycin. *Journal of Biological Chemistry*, 276(13):9992–9999, 2001.

- [163] T. Matsuda and C. L. Cepko. Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proceedings of the National Academy of Sciences*, 101(1):16–22, 2003.
- [164] H. Matsushita, V. G. Leliana, and Y. A. Ushkaryov. The latrophilin family: multiply spliced G protein-coupled receptors with differential tissue distribution. *FEBS Letters*, 443(3):348–52, 1999.
- [165] M. J. McCarthy, C. M. Nievergelt, J. R. Kelsoe, and D. K. Welsh. A Survey of Genomic Studies Supports Association of Circadian Clock Genes with Bipolar Disorder Spectrum Illnesses and Lithium Response. *PLoS ONE*, 7(2), 2012.
- [166] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. Phaser crystallographic software. *Journal of Applied Crystallography*, 40:658–674, 2007.
- [167] D. Meusch, C. Gatsogiannis, R. G. Efremov, A. E. Lang, O. Hofnagel, I. R. Vetter, K. Aktories, and S. Raunser. Mechanism of Tc toxin action revealed in molecular detail. *Nature*, 508:61–67, 2014.
- [168] R. L. Meyer. Roger Sperry and his chemoaffinity hypothesis. *Neuropsychologia*, 36(10):957–80, 1998.
- [169] A. D. Minet and R. Chiquet-Ehrismann. Phylogenetic analysis of teneurin genes and comparison to the rearrangement hot spot elements of *E. coli*. *Gene*, 257(1):87–97, 2000.
- [170] A. D. Minet, B. P. Rubin, R. P. Tucker, S. Baumgartner, and R. Chiquet-Ehrismann. Teneurin-1, a vertebrate homologue of the *Drosophila* pair-rule gene *ten-m*, is a neuronal protein with a novel type of heparin-binding domain. *Journal of Cell Science*, 112:2019–32, 1999.
- [171] M. Missler, W. Zhang, A. Rohlmann, G. Kattenstroth, R. E. Hammer, K. Gottmann, and T. C. Südhof. Alpha-neurexins couple Ca^{2+} channels to synaptic vesicle exocytosis. *Nature*, 423:939–948, 2003.

- [172] P. Monteiro and G. Feng. SHANK proteins: roles at the synapse and in autism spectrum disorder. *Nature Reviews Neuroscience*, 18:147–157, 2017.
- [173] K. Mosbahi, D. Walker, E. Lea, G. R. Moore, R. James, and C. Kleanthous. Destabilization of the colicin E9 Endonuclease domain by interaction with negatively charged phospholipids: implications for colicin translocation into bacteria. *Journal of Biological Chemistry*, 279(21):22145–51, 2004.
- [174] T. Mosca and L. Luo. Synaptic organization of the *Drosophila* antennal lobe and its regulation by the Teneurins. *eLife*, 2014.
- [175] T. J. Mosca. On the Teneurin track: a new synaptic organization molecule emerges. *Frontiers in Cellular Neuroscience*, 2015.
- [176] T. J. Mosca, W. Hong, V. S. Dani, V. Favaloro, and L. Luo. Trans-synaptic Teneurin signalling in neuromuscular synapse organization and target choice. *Nature*, 484(7393):237–41, 2012.
- [177] S. Mühlfeld, H.-P. Schmitt-Wrede, A. Harder, and F. Wunderlich. FMRFamide-like neuropeptides as putative ligands of the latrophilin-like HC110-R from *Haemonchus contortus*. *Molecular and Biochemical Parasitology*, 164(2):162–164, 2009.
- [178] A. Müller, J. Winkler, F. Fiedler, T. Sastradihardja, C. Binder, R. Schnabel, J. Kungel, S. Rothmund, C. Hennig, T. Schöneberg, and S. Prömel. Oriented Cell Division in the *C. elegans* Embryo Is Coordinated by G-Protein Signaling Dependent on the Adhesion GPCR LAT-1. *PLoS Genetics*, 11(10):1–20, 2015.
- [179] P.-S. Müller, R. Schulz, S. Maretto, I. Costello, S. Srinivas, E. Bikoff, and E. Robertson. The fibronectin leucine-rich repeat transmembrane protein Flrt2 is required in the epicardium to promote heart morphogenesis. *Development*, 138(7):1297–308, 2011.

- [180] T. Nakamoto, K. H. Kain, and M. H. Ginsberg. Neurobiology: New Connections between Integrins and Axon Guidance. *Current Biology*, 14(3):R121–R123, 2004.
- [181] C. I. Nam and L. Chen. Postsynaptic assembly induced by neurexin-neurologin interaction and neurotransmitter. *Proceedings of the National Academy of Sciences*, 102(17):6137–42, 2005.
- [182] T. Nechiporuk, L. D. Urness, and M. T. Keating. ETL, a novel seven-transmembrane receptor that is developmentally regulated in the heart. ETL is a member of the secretin family and belongs to the epidermal growth factor-seven-transmembrane subfamily. *Journal of Biological Chemistry*, 276(6):4150–4157, 2001.
- [183] C. Nijmeijer, S., Wolf, S., Ernst, O.P., de Graaf. 7TM domain structure of adhesion GPCRs. In T. Langenhan and T. Schoneberg, editors, *Adhesion G protein-coupled receptors: molecular, physiological and pharmacological principles in health and disease*. Springer, Heidelberg, 2016.
- [184] T. Nishimura, H. Honda, and M. Takeichi. Planar Cell Polarity Links Axes of Spatial Dynamics in Neural-Tube Closure. *Cell*, 149(5):1084–1097, 2012.
- [185] K. J. V. Nordstrom, M. C. Lagerstrom, L. M. J. Waller, R. Fredriksson, and H. B. Schioth. The Secretin GPCRs Descended from the Family of Adhesion GPCRs. *Molecular Biology and Evolution*, 26(1):71–84, 2008.
- [186] S. Nunes, J. Ferralli, K. Choi, M. Brownluedi, A. Minet, and R. Chiquet-Ehrismann. The intracellular domain of teneurin-1 interacts with MBD1 and CAP/ponsin resulting in subcellular codistribution and translocation to the nuclear matrix. *Experimental Cell Research*, 305(1):122–132, 2005.
- [187] D. Ofengeim, N. Giagtzoglou, D. Huh, C. Zou, and J. Yuan. Single-Cell RNA Sequencing: Unraveling the Brain One Cell at a Time. *Trends in Molecular Medicine*, 23(6):563–576, 2017.

- [188] S. Ogata, J. Morokuma, T. Hayata, G. Kolle, C. Niehrs, N. Ueno, and K. W. Y. Cho. TGF-beta signaling-mediated morphogenesis: modulation of cell adhesion via cadherin endocytosis. *Genes & Development*, 21(14):1817–31, 2007.
- [189] D. Okajima, G. Kudo, and H. Yokota. Brain-specific angiogenesis inhibitor 2 (BAI2) may be activated by proteolytic processing. *Journal of Receptors and Signal Transduction*, 30(3):143–153, 2010.
- [190] T. Ohashi, X.-H. Zhou, K. Feng, B. Richter, M. Mörgelin, M. T. Perez, W.-D. Su, R. Chiquet-Ehrismann, U. Rauch, and R. Fässler. Mouse Ten-m/Odz Is a New Family of Dimeric Type II Transmembrane Proteins Expressed in Many Tissues. *The Journal of Cell Biology*, 145(3):563–577, 1999.
- [191] M. L. O’Sullivan, J. de Wit, J. N. Savas, D. Comoletti, S. Otto-Hitt, J. R. Yates, and A. Ghosh. FLRT proteins are endogenous latrophilin ligands and regulate excitatory synapse development. *Neuron*, 73(5):903–10, 2012.
- [192] M. L. O’Sullivan, F. Martini, S. von Daake, D. Comoletti, and A. Ghosh. LPHN3, a presynaptic adhesion-GPCR implicated in ADHD, regulates the strength of neocortical layer 2/3 synaptic input to layer 5. *Neural development*, 9(7):1–11, 2014.
- [193] K. J. Paavola, H. Sidik, J. B. Zuchero, M. Eckart, and W. S. Talbot. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Science Signaling*, 7(338), 2014.
- [194] K. J. Paavola, J. R. Stephenson, S. L. Ritter, S. P. Alter, and R. A. Hall. The N terminus of the adhesion G protein-coupled receptor GPR56 controls receptor signaling activity. *The Journal of Biological Chemistry*, 286(33):28914–21, 2011.
- [195] J. Painter and E. A. Merritt. TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography*, 39(1):109–111, 2006.

- [196] N. S. Pannu, W.-J. Waterreus, P. Skubák, I. Sikharulidze, J. P. Abrahams, and R. A. G. De Graaff. Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallographica Section D Biological Crystallography*, 67:331–337, 2011.
- [197] D. Park, A.-C. Tosello-Tramont, M. R. Elliott, M. Lu, L. B. Haney, Z. Ma, A. L. Klibanov, J. W. Mandell, and K. S. Ravichandran. BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. *Nature*, 450(7168):430–434, 2007.
- [198] R. J. Pasterkamp, J. J. Peschon, M. K. Spriggs, and A. L. Kolodkin. Semaphorin 7A promotes axon outgrowth through integrins and MAPKs. *Nature*, 424:398–405, 2003.
- [199] J. Peça and G. Feng. Cellular and synaptic network defects in autism. *Current Opinion in Neurobiology*, 22:866–872, 2012.
- [200] S. C. Petersen, R. Luo, I. Liebscher, S. Giera, S.-J. Jeong, A. Mogha, M. Ghidnelli, M. L. Feltra, T. Schoneberg, X. Piao, and K. R. Monk. The Adhesion GPCR GPR126 Has Distinct, Domain- Dependent Functions in Schwann Cell Development Mediated by Interaction with Laminin-211. *Neuron*, 85:755–769, 2015.
- [201] X. Piao, R. S. Hill, A. Bodell, B. S. Chang, L. Basel-Vanagaite, R. Straussberg, W. B. Dobyns, B. Qasrawi, R. M. Winter, A. M. Innes, T. Voit, E. Ross, J. L. Michaud, J.-C. Déscarie, A. J. Barkovich, and C. A. Walsh. G Protein-Coupled Receptor-Dependent Development of Human Frontal Cortex. *Science*, 303:2033–2036, 2004.
- [202] A. J. Pocklington, M. O’Donovan, and M. J. Owen. The synapse in schizophrenia. *European Journal of Neuroscience*, 39(7):1059–1067, apr 2014.
- [203] A. J. Pommer, S. Cal, A. H. Keeble, D. Walker, S. J. Evans, U. C. Ku, È. Hlmann, A. Cooper, B. A. Connolly, A. M. Hemmings, G. R. Moore,

- R. James, and C. Kleanthous. Mechanism and Cleavage Specificity of the H-N-H Endonuclease Colicin E9. *Journal of Molecular Biology*, 314:735–349, 2001.
- [204] V. Y. Poon, M. P. Klassen, and K. Shen. UNC-6/netrin and its receptor UNC-5 locally exclude presynaptic components from dendrites. *Nature*, 455(7213):669–73, 2008.
- [205] S. Prömel, M. Frickenhaus, S. Hughes, L. Mestek, D. Staunton, A. Woollard, I. Vakonakis, T. Schöneberg, R. Schnabel, A. P. Russ, and T. Langenhan. The GPS Motif Is a Molecular Switch for Bimodal Activities of Adhesion Class G Protein-Coupled Receptors. *Cell Reports*, 2:321–331, 2012.
- [206] X. Qian, D. Barsyte-Lovejoy, L. Wang, B. Chewpoy, N. Gautam, A. Al Chawaf, and D. A. Lovejoy. Cloning and characterization of teneurin C-terminus associated peptide (TCAP)-3 from the hypothalamus of an adult rainbow trout (*Oncorhynchus mykiss*). *General and Comparative Endocrinology*, 137(2):205–16, 2004.
- [207] S. Ramon y Cajal. *Histology of the Nervous System of Man and Vertebrates (translated from French)*. Oxford University Press, 1995.
- [208] F. M. Ranaivoson, Q. Liu, F. Martini, F. Bergami, S. von Daake, S. Li, D. Lee, B. Demeler, W. A. Hendrickson, and D. Comoletti. Structural and Mechanistic Insights into the Latrophilin3-FLRT3 Complex that Mediates Glutamatergic Synapse Development. *Structure*, 23:1–13, 2015.
- [209] R. J. Read and A. J. McCoy. Using SAD data in Phaser. *Acta Crystallographica Section D Biological Crystallography*, 67:338–44, 2011.
- [210] P. J. Reeves, J.-M. Kim, and H. G. Khorana. Structure and function in rhodopsin: a tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proceedings of the National Academy of Sciences*, 99(21):13413–8, 2002.

- [211] X. Robert and P. Gouet. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42:W320–W324, 2014.
- [212] A. Robinson, S. Escuin, K. Doudney, M. Vekemans, R. E. Stevenson, N. D. Greene, A. J. Copp, and P. Stanier. Mutations in the planar cell polarity genes CELSR1 and SCRIB are associated with the severe neural tube defect craniorachischisis. *Human Mutation*, 33(2):440–447, 2012.
- [213] M. Robinson, M. Parsons Perez, L. Tébar, J. Palmer, A. Patel, D. Marks, A. Sheasby, C. De Felipe, R. Coffin, F. Livesey, and S. Hunt. FLRT3 is expressed in sensory neurons after peripheral nerve injury and regulates neurite outgrowth. *Molecular and Cellular Neuroscience*, 27(2):202–214, 2004.
- [214] A. A. Rostom and C. V. Robinson. Detection of the Intact GroEL Chaperonin Assembly by Mass Spectrometry. *Journal of the American Chemical Society*, 121:4718–4719, 1999.
- [215] B. P. Rubin, R. P. Tucker, M. Brown-Luedi, D. Martin, and R. Chiquet-Ehrismann. Teneurin 2 is expressed by the neurons of the thalamofugal visual system in situ and promotes homophilic cell-cell adhesion in vitro. *Development*, 129(20):4697–4705, 2002.
- [216] B. P. Rubin, R. P. Tucker, D. Martin, and R. Chiquet-Ehrismann. Teneurins: a novel family of neuronal cell surface proteins in vertebrates, homologous to the *Drosophila* pair-rule gene product Ten-m. *Developmental Biology*, 216(1):195–209, 1999.
- [217] J. D. Rudolf, L. Bigelow, C. Chang, M. E. Cuff, J. R. Lohman, C.-Y. Chang, M. Ma, D. Yang, S. Clancy, G. Babnigg, A. Joachimiak, G. N. Phillips, and B. Shen. Crystal Structure of the Zorbamycin-Binding Protein ZbmA, the Primary Self-Resistance Element in *Streptomyces flavoviridis* ATCC21892. *Biochemistry*, 54:6842–6851, 2015.
- [218] B. Rupp. *Biomolecular Crystallography*. Garland Science, 2010.

- [219] G. Salzman, S. Ackerman, C. Ding, A. Koide, K. Leon, R. Luo, H. Stoveken, C. Fernandez, G. Tall, X. Piao, K. Monk, S. Koide, and D. Arac. Structural Basis for Regulation of GPR56/ADGRG1 by Its Alternatively Spliced Extracellular Domains. *Neuron*, 91:1292–1304, 2016.
- [220] P. Scheiffele, J. Fan, J. Choih, R. Fetter, and T. Serafini. Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell*, 101(6):657–69, 2000.
- [221] D. Schmucker, J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, and J. E. Dixon. Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell*, 101:671–684, 2000.
- [222] T. R. Schneider and G. M. Sheldrick. Substructure solution with SHELXD. *Acta Crystallographica Section D Biological Crystallography*, 58:1772–1779, 2002.
- [223] J. Schöler, J. Ferralli, S. Thiry, and R. Chiquet-Ehrismann. The Intracellular Domain of Teneurin-1 Induces the Activity of Microphthalmia-associated Transcription Factor (MITF) by Binding to Transcriptional Repressor HINT1. *Journal of Biological Chemistry*, 290(13):8154–8165, 2015.
- [224] N. Scholz, J. Gehring, C. Guan, D. Ljaschenko, R. Fischer, V. Lakshmanan, R. J. Kittel, and T. Langenhan. The Adhesion GPCR Latrophilin/CIRL Shapes Mechanosensation. *Cell Reports*, 11(6):866–874, 2015.
- [225] N. Scholz, C. Guan, M. Nieberler, A. Grotemeyer, I. Maiellaro, S. Gao, S. Beck, M. Pawlak, M. Sauer, E. Asan, S. Rothmund, J. Winkler, S. Promel, G. Nagel, T. Langenhan, and R. Kittel. Mechano-dependent signaling by Latrophilin/CIRL quenches cAMP in proprioceptive neurons. *eLife*, 6:e28360, 2017.
- [226] E. Seiradake, D. del Toro, D. Nagel, F. Cop, R. Härtl, T. Ruff, G. Seyit-Bremer, K. Harlos, E. C. Border, A. Acker-Palmer, E. Y. Jones, and R. Klein.

- FLRT structure: balancing repulsion and cell adhesion in cortical and vascular development. *Neuron*, 84(2):370–85, 2014.
- [227] E. Seiradake, K. Harlos, G. Sutton, a. R. Aricescu, and E. Y. Jones. An extra-cellular steric seeding mechanism for Eph-ephrin signaling platform assembly. *Nature Structural & Molecular Biology*, 17(4):398–402, 2010.
- [228] E. Seiradake, A. Schaupp, D. del Toro Ruiz, R. Kaufmann, N. Mitakidis, K. Harlos, a. R. Aricescu, R. Klein, and E. Y. Jones. Structurally encoded intraclass differences in EphA clusters drive distinct cell responses. *Nature Structural & Molecular Biology*, 20(8):958–64, 2013.
- [229] T. Serafini, T. E. Kennedy, M. J. Gaiko, C. Mirzayan, T. M. Jessell, and M. Tessier-Lavigne. The netrins define a family of axon outgrowth-promoting proteins homologous to *C. elegans* UNC-6. *Cell*, 78(3):409–424, 1994.
- [230] G. M. Sheldrick. A short history of SHELX. *Acta Crystallographica Section D Biological Crystallography*, 64:112–122, 2008.
- [231] Y. Shima, S.-Y. Kawaguchi, K. Kosaka, M. Nakayama, M. Hoshino, Y. Nabeshima, T. Hirano, and T. Uemura. Opposing roles in neurite growth control by two seven-pass transmembrane cadherins. *Nature Neuroscience*, 10(8):963–969, 2007.
- [232] Y. Shima, M. Kengaku, T. Hirano, M. Takeichi, and T. Uemura. Regulation of Dendritic Maintenance and Growth by a Mammalian 7-Pass Transmembrane Cadherin. *Developmental Cell*, 7:205–216, 2004.
- [233] S. K. Shin, T. Nagasaka, B. H. Jung, N. Matsubara, W. H. Kim, J. M. Carethers, C. R. Boland, and A. Goel. Epigenetic and Genetic Alterations in Netrin-1 Receptors UNC5C and DCC in Human Colon Cancer. *Gastroenterology*, 133(6):1849–1857, 2007.
- [234] T. J. Siddiqui and A. M. Craig. Synaptic organizing complexes. *Current Opinion in Neurobiology*, 21(1):132–143, 2011.

- [235] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(539), 2011.
- [236] S. Sigoillot, K. Iyer, F. Binda, I. Gonzalez-Calvo, M. Talleur, G. Vodjdani, P. Isope, and F. Selimi. The Secreted Protein C1QL1 and Its Receptor BAI3 Control the Synaptic Connectivity of Excitatory Inputs Converging on Cerebellar Purkinje Cells. *CellReports*, 10:820–832, 2015.
- [237] J.-P. Silva, V. G. Leliana, Y. S. Ermolyuk, N. Vysokov, P. G. Hitchen, O. Berninghausen, M. A. Rahman, A. Zangrandi, S. Fidalgo, A. G. Tonevitsky, A. Dell, K. E. Volynski, and Y. a. Ushkaryov. Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proceedings of the National Academy of Sciences*, 108(29):12113–8, 2011.
- [238] P. Skubák and N. S. Pannu. Automatic protein structure solution from weak X-ray data. *Nature Communications*, 4:2777, 2013.
- [239] F. Sobott, H. Hernández, M. McCammon, M. Tito, and C. Robinson. A Tandem Mass Spectrometer for Improved Transmission and Analysis of Large Macromolecular Assemblies. *Analytical Chemistry*, 74:1402–1407, 2002.
- [240] J. Söding, A. Biegert, and A. N. Lupas. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, 33:W244–W248, 2005.
- [241] D. J. Solecki. Sticky situations: recent advances in control of cell adhesion during neuronal migration. *Current Opinion in Neurobiology*, 22(5):791–798, 2012.
- [242] C. Söllner and G. J. Wright. A cell surface interaction network of neural leucine-rich repeat receptors. *Genome Biology*, 10(9):R99, 2009.

- [243] A. Sommer, G. Herberth, A. Stein, J. Glienke, S. Taudien, I. Klamann, A. Herr, and K.-H. Thierach. Human Latrophilin-2 is Expressed in the Cytotrophoblast and Syncytiotrophoblast of Placenta and in Endothelial Cells. *American Journal of Biochemistry and Biotechnology*, 1(3):135–144, 2005.
- [244] R. Sperry. CHEMOAFFINITY IN THE ORDERLY GROWTH OF NERVE FIBER PATTERNS AND CONNECTIONS. *Proceedings of the National Academy of Sciences*, 50:703–10, 1963.
- [245] M. Stacey, G.-W. Chang, J. Q. Davies, M. J. Kwakkenbos, R. D. Sanderson, J. Hamann, S. Gordon, and H.-H. Lin. The epidermal growth factor-like domains of the human EMR2 receptor mediate cell attachment through chondroitin sulfate glycosaminoglycans. *Blood*, 102(8), 2003.
- [246] H. M. Stoveken, A. G. Hajduczuk, L. Xu, and G. G. Tall. Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proceedings of the National Academy of Sciences*, 112(19):6194–9, 2015.
- [247] T. C. Südhof. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature*, 455(7215):903–11, 2008.
- [248] S. Sugita, K. Ichtchenko, M. Khvotchev, and T. C. Südhof. α -Latrotoxin Receptor CIRL/Latrophilin 1 (CL1) Defines an Unusual Family of Ubiquitous G-protein-linked Receptors. *Journal of Biological Chemistry*, 273(49):32715–32724, 1998.
- [249] N. Suzuki, M. Fukushi, K. Kosaki, A. D. Doyle, S. de Vega, K. Yoshizaki, C. Akazawa, E. Arikawa-Hirasawa, and Y. Yamada. Teneurin-4 is a novel regulator of oligodendrocyte differentiation and myelination of small-diameter axons in the CNS. *The Journal of Neuroscience*, 32(34):11586–99, 2012.
- [250] N. Suzuki, T. Numakawa, J. Chou, S. De Vega, C. Mizuniwa, K. Sekimoto, N. Adachi, H. Kunugi, E. Arikawa-Hirasawa, Y. Yamada, and C. Akazawa. Teneurin-4 promotes cellular protrusion formation and neurite outgrowth

- through focal adhesion kinase signaling. *FASEB Journal*, 28(3):1386–1397, 2014.
- [251] L. A. Tan, A. Al Chawaf, F. J. Vaccarino, P. C. Boutros, and D. A. Lovejoy. Teneurin C-terminal associated peptide (TCAP)-1 modulates dendritic morphology in hippocampal neurons and decreases anxiety-like behaviors in rats. *Physiology and Behavior*, 104(2):199–204, 2011.
- [252] L. A. Tan, K. Xu, F. J. Vaccarino, D. A. Lovejoy, and S. Rotzinger. Repeated intracerebral teneurin C-terminal associated peptide (TCAP)-1 injections produce enduring changes in behavioral responses to corticotropin-releasing factor (CRF) in rat models of anxiety. *Behavioural Brain Research*, 188(1):195–200, 2008.
- [253] K. Thiebault, L. Mazelin, L. Pays, F. Llambi, M.-O. Joly, J.-Y. Scoazec, J.-C. Saurin, G. Romeo, and P. Mehlen. The netrin-1 receptors UNC5H are putative tumor suppressors controlling cell death commitment. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7):4173–8, 2003.
- [254] N. K. Thiede-Stan and M. E. Schwab. Attractive and repulsive factors act through multi-subunit receptor complexes to regulate nerve fiber growth. *Journal of Cell Science*, 128(14):2403–14, 2015.
- [255] F. Tissir and A. M. Goffinet. Shaping the nervous system: role of the core planar cell polarity genes. *Nature Reviews Neuroscience*, 14:525–535, 2013.
- [256] S. Tobaben, T. C. Südhof, and B. Stahl. The G Protein-coupled Receptor CL1 Interacts Directly with Proteins of the Shank Family. *Journal of Biological Chemistry*, 275(46):36204–36210, 2000.
- [257] L. Tsuji, T. Yamashita, T. Kubo, T. Madura, H. Tanaka, K. Hosokawa, and M. Tohyama. FLRT3, a cell surface molecule containing LRR repeats and

- a FNIII domain, promotes neurite outgrowth. *Biochemical and Biophysical Research Communications*, 313(4):1086–1091, 2004.
- [258] R. Tucker and R. Chiquet-Ehrismann. Teneurins: A conserved family of transmembrane proteins involved in intercellular signaling during development. *Developmental Biology*, 290(2):237–245, 2006.
- [259] R. P. Tucker, J. Beckmann, N. T. Leachman, J. Scho, and R. Chiquet-Ehrismann. Phylogenetic Analysis of the Teneurins : Conserved Features and Premetazoan Ancestry Research article. *Molecular Biology and Evolution*, 29(3):1019–1029, 2012.
- [260] I. Vakonakis, T. Langenhan, S. Prömel, A. Russ, and I. D. Campbell. Solution structure and sugar-binding mechanism of mouse latrophilin-1 RBL: a 7TM receptor-attached lectin-like domain. *Structure*, 16(6):944–53, 2008.
- [261] C. C. Valley, A. K. Lewis, and J. N. Sachs. Piecing it together: Unraveling the elusive structure-function relationship in single-pass membrane receptors. *Biochimica et Biophysica Acta*, 1859(9):1398–1416, 2017.
- [262] M. Vallon and M. Essler. Proteolytically processed soluble tumor endothelial marker (TEM) 5 mediates endothelial cell survival during angiogenesis by linking integrin alpha(v)beta3 to glycosaminoglycans. *Journal of Biological Chemistry*, 281(45):34179–88, 2006.
- [263] F. Varoqueaux, G. Aramuni, R. L. Rawson, R. Mohrmann, M. Missler, K. Gottmann, W. Zhang, T. C. Südhof, and N. Brose. Neuroligins Determine Synapse Maturation and Function. *Neuron*, 51(6):741–754, 2006.
- [264] U. H. Verkerk and P. Kebarle. Ion-Ion and ion-molecule reactions at the surface of proteins produced by nanospray. Information on the number of acidic residues and control of the number of ionized acidic and basic residues. *Journal of the American Society for Mass Spectrometry*, 16(8):1325–1341, 2005.

- [265] J. J. Visser, Y. Cheng, S. C. Perry, A. B. Chastain, B. Parsa, S. S. Masri, T. A. Ray, J. N. Kay, and W. M. Wojtowicz. An extracellular biochemical screen reveals that FLRTs and Unc5s mediate neuronal subtype recognition in the retina. *eLife*, 4:e08149, 2015.
- [266] K. E. Volynski, J.-P. Silva, V. G. Lelianova, M. Atiqur Rahman, C. Hopkins, and Y. A. Ushkaryov. Latrophilin fragments behave as independent proteins that associate and signal on binding of LTX(N4C). *The EMBO Journal*, 23(22):4423–33, 2004.
- [267] N. V. Vysokov, J.-P. Silva, V. G. Lelianova, C. Ho, M. B. Djamgoz, A. G. Tonevitsky, and Y. A. Ushkaryov. The Mechanism of Regulated Release of Lasso/Teneurin-2. *Frontiers in Molecular Neuroscience*, 9:59, 2016.
- [268] E. Wandel, A. Saalbach, D. Sittig, C. Gebhardt, and G. Aust. Thy-1 (CD90) Is an Interacting Partner for CD97 on Activated Endothelial Cells. *The Journal of Immunology*, 188(3):1442–1450, 2012.
- [269] L. Wang, S. Rotzinger, A. Al Chawaf, C. F. Elias, D. Baršytė-Lovejoy, X. Qian, N.-C. Wang, A. De Cristofaro, D. Belsham, J. C. Bittencourt, F. Vaccarino, and D. A. Lovejoy. Teneurin proteins possess a carboxy terminal sequence with neuromodulatory activity. *Molecular Brain Research*, 133(2):253–265, 2005.
- [270] R. Wang, Z. Wei, H. Jin, H. Wu, C. Yu, W. Wen, L. N. Chan, Z. Wen, and M. Zhang. Autoinhibition of UNC5b Revealed by the Cytoplasmic Domain Structure of the Receptor. *Molecular Cell*, 33(6):692–703, 2009.
- [271] T. Wang, Y. Ward, L. Tian, R. Lake, L. Guedez, W. G. Stetler-Stevenson, and K. Kelly. CD97, an adhesion receptor on inflammatory cells, stimulates angiogenesis through binding integrin counterreceptors on endothelial cells. *Blood*, 105(7), 2005.

- [272] X.-Z. Wang, M. Kuroda, J. Sok, N. Batchvarova, R. Kimmel, P. Chung, H. Zinsner, and D. Ron. Identification of novel stress-induced genes downstream of chop. *The EMBO Journal*, 17(13):3619–3630, 1998.
- [273] P. Westermark, K. Sletten, B. Johansson, and G. G. Cornwell. Fibril in senile systemic amyloidosis is derived from normal transthyretin. *Medical Sciences*, 87:2843–2845, 1990.
- [274] M. K. Wetzel-Smith, J. Hunkapiller, T. R. Bhangale, K. Srinivasan, J. A. Maloney, J. K. Atwal, S. M. Sa, M. B. Yaylaoglu, O. Foreman, W. Ortmann, N. Rathore, D. V. Hansen, M. Tessier-Lavigne, Alzheimer’s Disease Genetics Consortium, R. Mayeux, M. Pericak-Vance, J. Haines, L. A. Farrer, G. D. Schellenberg, A. Goate, T. W. Behrens, C. Cruchaga, R. J. Watts, and R. R. Graham. A rare mutation in UNC5C predisposes to late-onset Alzheimer’s disease and increases neuronal cell death. *Nature Medicine*, 20(12):1452–1457, 2014.
- [275] L. M. Wheldon, B. P. Haines, R. Rajappa, I. Mason, P. W. Rigby, and J. K. Heath. Critical Role of FLRT1 Phosphorylation in the Interdependent Regulation of FLRT1 Function and FGF Receptor Signalling. *PLoS ONE*, 5(4), 2010.
- [276] J. P. White, C. D. Wrann, R. R. Rao, S. K. Nair, M. P. Jedrychowski, J.-S. You, V. Martínez-Redondo, S. P. Gygi, J. L. Ruas, T. A. Hornberger, Z. Wu, D. J. Glass, X. Piao, and B. M. Spiegelman. G protein-coupled receptor 56 regulates mechanical overload-induced muscle hypertrophy. *Proceedings of the National Academy of Sciences*, 111(44):15756–15761, 2014.
- [277] C. Wilde, L. Fischer, V. Lede, J. Kirchberger, S. Rothmund, T. Schöneberg, and I. Liebscher. The constitutive activity of the adhesion GPCR GPR114/ADGRG5 is mediated by its tethered agonist. *FASEB Journal*, 30(2):666–73, 2016.

- [278] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):235–242, 2011.
- [279] G. Winter, C. M. C. Lobley, and S. M. Prince. Decision making in xia2. *Acta crystallographica. Section D, Biological crystallography*, 69(Pt 7):1260–73, jul 2013.
- [280] R. Woelfle, A. L. D’Aquila, T. Pavlović, M. Husić, and D. A. Lovejoy. Ancient interaction between the teneurin C-terminal associated peptides (TCAP) and latrophilin ligand-receptor coupling: a role in behavior. *Frontiers in Neuroscience*, 9:146, 2015.
- [281] W. M. Wojtowicz, W. Wu, I. Andre, B. Qian, D. Baker, and S. L. Zipursky. A Vast Repertoire of Dscam Binding Specificities Arises from Modular Interactions of Variable Ig Domains. *Cell*, 130(6):1134–1145, 2007.
- [282] L. Xu, S. Begum, J. D. Hearn, and R. O. Hynes. GPR56, an atypical G protein-coupled receptor, binds tissue transglutaminase, TG2, and inhibits melanoma tumor growth and metastasis. *Proceedings of the National Academy of Sciences*, 103(24):9023–8, 2006.
- [283] Y. Xu, K. Wei, W. Kulyk, and S. G. Gong. FLRT2 promotes cellular proliferation and inhibits cell adhesion during chondrogenesis. *Journal of Cellular Biochemistry*, 112(11):3440–3448, 2011.
- [284] P. T. Yam and F. Charron. Signaling mechanisms of non-conventional axon guidance cues: the Shh, BMP and Wnt morphogens. *Current Opinion in Neurobiology*, 23(6):965–973, 2013.

- [285] S. Yamagishi, F. Hampel, K. Hata, D. Del Toro, M. Schwark, E. Kvachnina, M. Bastmeyer, T. Yamashita, V. Tarabykin, R. Klein, and J. Egea. FLRT2 and FLRT3 act as repulsive guidance cues for Unc5-positive neurons. *The EMBO Journal*, 30(14):2920–33, 2011.
- [286] L. Yang, G. Chen, S. Mohanty, G. Scott, F. Fazal, A. Rahman, S. Begum, R. O. Hynes, and L. Xu. GPR56 Regulates VEGF Production and Angiogenesis during Melanoma Progression. *Cancer Research*, 71(16), 2011.
- [287] T. R. Young, M. Bourke, X. Zhou, T. Oohashi, A. Sawatari, R. Fassler, and C. A. Leamey. Ten-m2 Is Required for the Generation of Binocular Visual Circuits. *Journal of Neuroscience*, 33(30):12490–12509, 2013.
- [288] M. Zebisch, V. A. Jackson, Y. Zhao, and E. Y. Jones. Structure of the Dual-Mode Wnt Regulator Kremen1 and Insight into Ternary Complex Formation with LRP6 and Dickkopf. *Structure*, 24:1599–1605, 2016.
- [289] T. Zeev-Ben-Mordehai, D. Vasishtan, C. A. Siebert, C. Whittle, and K. Grünewald. Extracellular Vesicles: A Platform for the Structure Determination of Membrane Proteins by Cryo-EM. *Structure*, 22(11):1687–1692, 2014.
- [290] D. Zhang, R. F. de Souza, V. Anantharaman, L. M. Iyer, and L. Aravind. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biology Direct*, 7(1):18, 2012.
- [291] D. Zhang, L. M. Iyer, and L. Aravind. A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic Acids Research*, 39(11):4532–52, 2011.
- [292] X.-h. Zhou, O. Brandau, K. Feng, T. Oohashi, and Y. Ninomiya. The murine Ten-m / Odz genes show distinct but overlapping expression patterns during development and in adult brain. *Gene Expression Patterns*, 3:397–405, 2003.

- [293] A. Ziegler, A. Corvalán, I. Roa, J. A. Brañes, and B. Wollscheid. Teneurin protein family: an emerging role in human tumorigenesis and drug resistance. *Cancer Letters*, 326(1):1–7, 2012.
- [294] A. Zuko, A. Oguro-Ando, H. Post, R. L. R. E. Taggenbrock, R. E. Van Dijk, A. F. Maarten Altelaar, A. J. R. Heck, A. G. Petrenko, B. Van Der Zwaag, Y. Shimoda, R. J. Pasterkamp, D. Tropea, M. J. Schmeisser, and J. Burbach. Association of Cell Adhesion Molecules Contactin-6 and Latrophilin-1 Regulates Neuronal Apoptosis. *Frontiers in Molecular Neuroscience*, 9:143, 2016.
- [295] P. H. Zwart, R. W. Grosse-Kunstleve, A. A. Lebedev, G. N. Murshudov, and P. D. Adams. Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallographica Section D: Biological Crystallography*, 64(1):99–107, 2007.

Appendix A

Constructs and Primers

Construct Name	Mutation	Vector	Cloning Sites	Tags	a.a. _N	a.a. _C
Lphn1						
Lphn1 ^{Lec-Olf}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	29	395
Lphn2						
Lphn2 ^{Lec-Olf}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	30	400
Lphn3						
Lphn3 ^{Lec}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	92	197
Lphn3 ^{Lec-Olf}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	92	463
Lphn3 ^{Lec-Olf}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	92	463
FLRT1						
FLRT1 ^{LRR}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	53	381
FLRT2						
FLRT2 ^{LRR}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	35	362
FLRT2 ^{Ecto}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	35	540
FLRT2 ^{Ecto} UF	H170N	pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	35	540
FLRT2 ^{Ecto} FF	R186N D188T	pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	35	540
FLRT2-ires-GFP		pCAGIG	<i>XhoI</i> / <i>NotI</i>	HA(N), PTP μ SS	35	660
FLRT2 ^{FF} -ires-GFP	R186N D188T	pCAGIG	<i>XhoI</i> / <i>NotI</i>	HA(N), PTP μ SS	35	660
FLAG-FLRT2-Avi		pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), Avi(C)	35	660
FLAG-FLRT2 ^{UF} -Avi	H170N	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), Avi(C)	35	660
FLRT3						
FLRT3 ^{LRR}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	29	359
FLRT3 ^{Ecto}		pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	29	526
Unc5A						
Unc5A ^{Ecto}		pHLSec	<i>EcoRI</i> / <i>KpnI</i>	His6(C)	1	300
Unc5B						
Unc5B ^{Ecto}		pHLSec	<i>EcoRI</i> / <i>KpnI</i>	His6(C)	1	355
Unc5D						
Unc5D-ires-GFP		pCAGIG	<i>XhoI</i> / <i>NotI</i>	HA(N), PTP μ SS	46	956
Unc5D ^{Ig}		pHLSec	<i>EcoRI</i> / <i>KpnI</i>	His6(C)	1	161
Unc5D ^{IgIg}		pHLSec	<i>EcoRI</i> / <i>KpnI</i>	His6(C)	1	244
Unc5D ^{IgIgTsp}		pHLSec	<i>EcoRI</i> / <i>KpnI</i>	His6(C)	1	307
Unc5D ^{Ecto}		pHLSec	<i>EcoRI</i> / <i>KpnI</i>	Avi(C)	1	382
Unc5D ^{Ecto}		pHLSec	<i>EcoRI</i> / <i>KpnI</i>	His6(C)	1	382

Table A.1: **Constructs Cloned by Others and Used in this Thesis.** Abbreviations are: a.a._N, N-terminal amino acid; a.a._C, C-terminal amino acid; (N), N-terminal tag; (C), C-terminal tag; His6, hexahistidine; IRES, internal ribosome entry site; PTP μ SS, signal sequence from receptor-type tyrosine-protein phosphatase S.

Construct Name	Mutation	Vector	Tags	a.a.N	a.a.C	Dir.	Primer
Lphn3							AATAATACCGGTCCCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG CCTGTCTTCAACAAGAGAAATACCAACAACATCGTGAAGTTCGACCTGC GCAGGTGCAACTCAGCATGTTTGTGGTATCTCTTTGTGAAGAACAGG
Lphn3 ^{Lec-Olf} LF	R292N R294T	pHLSec	Avi(C)	92	463	F R F _{mut} R _{mut}	AATAATACCGGTCCCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG CCTGTCTTCAACAAGAGAAATACCAACAACATCGTGAAGTTCGACCTGC GCAGGTGCAACTCAGCATGTTTGTGGTATCTCTTTGTGAAGAACAGG
Lphn3 ^{Lec-Olf} LF (2)	Y323N W325T	pHLSec	Avi(C)	92	463	F R F _{mut} R _{mut}	AATAATACCGGTCCCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG TTACCAAGACACCAAGCCCTAACAGAACAGCGGGCAAGCGACATCGACC GGTGGATGTCGCTCTTCCGCGCTGTTCTGTAGGGGCTGTCGTCTGTTAA
Lphn3 ^{Lec-Olf} LF (3)	T267N K269T	pHLSec	Avi(C)	92	463	F R F _{mut} R _{mut}	AATAATACCGGTCCCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG ATAAGATCTACTACATGCGCTGACCACTACACAAACCGGACACCTGACC GGTGAGGGTCTCGGTTGTGTAGTTGGTCCAGGGCATGTAGTAGATCTTAT
Lphn3 ^{Lec-Olf} LF (4)	P244N R246T	pHLSec	Avi(C)	92	463	F R F _{mut} R _{mut}	AATAATACCGGTCCCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG CGCGGGACACCAACCACTACACACTGCCACAGAGTGGAGTGGACGCA TGCCGTCCACTCTGTGGGCGAGTGTGTAATTGGTGGTCTGTCGGCGG
Lphn3 ^{Lec-Olf} Charge Reversal	D397R D398R D399R	pHLSec	Avi(C)	92	463	F R F _{mut} R _{mut}	AATAATACCGGTCCCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG GTGAAGTCCGTGTACGAGCGCAGACGCAACGAGCCACCGGCAAC GTTCCCGGTGGCCTCGTTCGCTCTGCGCTCGTACACGGACTTCAAC
Lphn3 ^{Lec-Olf} DDD mutant	D397N D399T	pHLSec	Avi(C)	92	463	F R F _{mut} R _{mut}	AATAATACCGGTCCCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG GTGAAGTCCGTGTACGAGAACACCAACGAGCCACCGGCAAC GTTCCCGGTGGCCTCGTTCGCTCTGCGCTCGTACACGGACTTCAAC
Unc5D							AATAATGAAATTCGGCACCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG GTGAAGTCCGTGTACGAGAACACCAACGAGCCACCGGCAAC GTTCCCGGTGGCCTCGTTCGCTCTGCGCTCGTACACGGACTTCAAC
Unc5D ^{Ecto} Octamer Mutant	M292N V294T	pHLSec	Avi(C)	1	382	F R F _{mut} R _{mut}	AATAATGAAATTCGGCACCATCCCTATGGCGTCTGTCGGGAGAGAG AATAATGGTACCGTCCAGGGGCGGCAAGTCCAGGGAGTATTTCAAGACGTG GGTGGGCGCTTTGTGAGGGAAATTCACAGGCGAGAAATAACCTGCACCTGCT AGCAGTGCAGGGTATTTTCTGCGTGTGAATTTCCCTCACAAGGGCCCCACC

Table A.2: **Mutant Constructs Cloned as Part of this Thesis.** Abbreviations are: Dir., direction; F, forward; R, reverse; F_{mut}, forward mutagenesis primer; R_{mut}, reverse mutagenesis primer. All other abbreviations are as above.

Construct Name	Vector	Restriction Sites	Tags	a.a.N	a.a.C	Dir.	Primer
HsTen1							
HsTen1 ^{Ecto}	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	362	2725	F R	GACAAGGGGTACCGGTAGCAAAAGGGAACAGGGGGACCGAGTCC GTGGTGCTTGGTACCCCTCTGCTCCATTTTCGGCTCTGTCTC
GgTen1							
GgTen1 ^{Ecto}	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	343	2705	F R	GACAAGGGGTACCGGTAGCAAAAGGGAATAAAGGAGCAGAATCC GTGGTGCTTGGTACCCCTCTGCTCCATTTTCACCTCTGTCTC
GgTen2							
GgTen2 ^{Ecto}	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	596	2802	F	GACAAGGGGTACCGGTGTATTTTGGATTGATTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTCTTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ¹¹⁹⁹⁻¹⁶⁰⁷	pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	1199	1607	F	GTAGCTGAAACCGGTCTAGATCCTTCCATCTAGGAGGATGGTCTTTGG GTGGTGCTTGGTACCCAGATTAAGAATGGCCCTGTTTAACTGAC
GgTen2 ¹¹⁹⁹⁻²⁸⁰²	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	1199	1607	F	GACAAGGGGTACCGGTCTAGATCCTTCCATCTAGGAGGATGG GTGGTGCTTGGTACCCCTCTTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ⁸⁷⁴⁻²⁸⁰²	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	874	2802	F	GACAAGGGGTACCGGTACCTCTGCGGATACAAAGGATAACGAGG GTGGTGCTTGGTACCCCTCTTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ¹¹⁹⁹⁻²⁶⁸³	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	1199	2683	F	GACAAGGGGTACCGGTCTAGATCCTTCCATCTAGGAGGATGG GTGGTGCTTGGTACCCCTGCTGGCTGGGAGCGGTACACGTTTAC
GgTen2 ¹¹⁹⁹⁻²⁶⁶⁷	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	1199	2667	F	GACAAGGGGTACCGGTCTAGATCCTTCCATCTAGGAGGATGG GTGGTGCTTGGTACCCCTGCTGTCATCGGAGGGGTACGAG
GgTen2 ¹¹⁹⁹⁻²⁵⁹³	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	1199	2593	F	GACAAGGGGTACCGGTCTAGATCCTTCCATCTAGGAGGATGG GTGGTGCTTGGTACCCCTGCTTGTGCAACGAGTGCCTGG
GgTen2 ⁵⁹⁶⁻²⁶⁸³	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	596	2683	F	GACAAGGGGTACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCGTTGGCTGGGAGCGGTACGTTTAC
GgTen2 ⁵⁹⁶⁻²⁵⁹³	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	596	2593	F	GACAAGGGGTACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCGTTGGCTGGGAGCGGTACGTTTAC
GgTen2 ⁵⁹⁶⁻²⁶⁶⁷	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	596	2667	F	GACAAGGGGTACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTGCTGTCATCGGAGGTGACGAG
GgTen2 ¹⁶⁵⁴⁻²⁸⁰²	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	1654	2802	F	GACAAGGGGTACCGGTACCGAGGTGATGACAGCAATGGCAACTCC GTGGTGCTTGGTACCCCTTTCCTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ¹⁶⁰⁸⁻²⁸⁰²	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	1608	2802	F	GACAAGGGGTACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTTTCCTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ⁹⁰⁹⁻²⁸⁰²	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	909	2802	F	GACAAGGGGTACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTTTCCTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ⁹⁰⁹⁻²⁸⁰²	pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	909	2802	F	GTAGCTGAAACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTTTCCTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ⁹¹²⁻²⁸⁰²	pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	912	2802	F	GTAGCTGAAACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTTTCCTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ⁹³²⁻²⁸⁰²	pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	932	2802	F	GTAGCTGAAACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTTTCCTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ^{ΔEGF}	pHLSec	<i>AgeI</i> / <i>KpnI</i>	His6(C)	955	2802	F	GTAGCTGAAACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTTTCCTCCATTTTCATTTCTGTCTTAAAACTGG
GgTen2 ^{ΔEGF}	pHLSec	<i>AgeI</i> / <i>KpnI</i>	FLAG(N), His6(C)	596	924	F R	GACAAGGGGTACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GTGGTGCTTGGTACCCCTTTCCTCCATTTTCATTTCTGTCTTAAAACTGG
Lphn1							
Lphn1 ^{Lec}	pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	26	134	F R	GTAGCTGAAACCGGTACCGGGGTGGCTCCCATTTGGGGTTG GGAACCTCCGGTACCCCTCCACTTTGTAAAGGACACAGTCGTA
Lphn2							
Lphn2 ^{Lec}	pHLSec	<i>AgeI</i> / <i>KpnI</i>	Avi(C)	30	137	F R	GTAGCTGAAACCGGTGTATTTTGGATTTCAGTGCAGAACTGTCCACG GGAACCTCCGGTACCCCTCCACTTTGTCTGTAAGGGACACATTCATATTGAACCTTCAAGG

Table A.3: Teneurin and Lphn constructs cloned as part of this thesis. All abbreviations are as above.

Appendix B

Pseudo-symmetry of GgTen2^{ΔEGF} Crystals

GgTen2^{ΔEGF} crystals suffered an interesting pathology, whereby some datasets would index in space group P2₁, whereas others would index in space group C222₁. This was also observed even for different datasets acquired from the same crystal. Further, it was not possible to convincingly re-index C222₁ datasets into P2₁ or vice versa. For the two space groups, the unit cells were a similar size, but the *b* and *c* unit cell lengths were swapped. A derivative dataset (C222₁, ~ 3.3 Å resolution) was used to build an initial model of GgTen2^{ΔEGF}. The native dataset used to refine the structure was indexed as P2₁ (~ 2.4 Å resolution). A summary of the crystallographic symmetry in each dataset is shown below.

	C222 ₁ (Derivative)	P2 ₁ (Native)
<i>a</i> , <i>b</i> , <i>c</i> (Å)	83.32, 156.13, 452.95	88.56, 452.56, 146.36
α , β , γ (°)	90.00, 90.00, 90.00	90.00, 95.12, 90.00
Molecules in ASU	1	4
Molecules in Unit Cell	8	8

Table B.1: **Crystallographic Symmetry of GgTen2^{ΔEGF} Crystals.** ASU, asymmetric unit.

Native Patterson maps and Phenix Xtriage [4] identified a strong off-origin Patterson peak (46% the height of the origin peak) for the P2₁ dataset. As a fraction of the unit cell lengths the length of this vector was 0.50, 0.00, 0.50, with a distance to the origin of 82.1 Å. The presence of this strong peak was indicative of translation non-crystallographic symmetry (tNCS) in the P2₁ native data. In this case the tNCS vector is identical to one of the possible *C*-centring operators, which can lead to a pathology known as pseudo-centering, where an NCS vector closely corresponds

to a possible crystallographic symmetry operation [295]. Given that both the $P2_1$ and $C222_1$ datasets had similar unit cells, it seemed likely that pseudo-centring was responsible for the apparent difference in space group between the two datasets.

The presence of tNCS in the native ($P2_1$) dataset exacerbated difficulties in MR. Fortunately, the $C222_1$ asymmetric unit of the derivative dataset contained only one copy of GgTen $2^{\Delta\text{EGF}}$, meaning no NCS was possible, and the strong anomalous signal facilitated initial structure solution in this space group. This solution was then used as a MR model to solve the $P2_1$ dataset as described in Chapter 6. As expected, this revealed that the two indexing solutions represented almost identical crystal packing (Figure B.1).

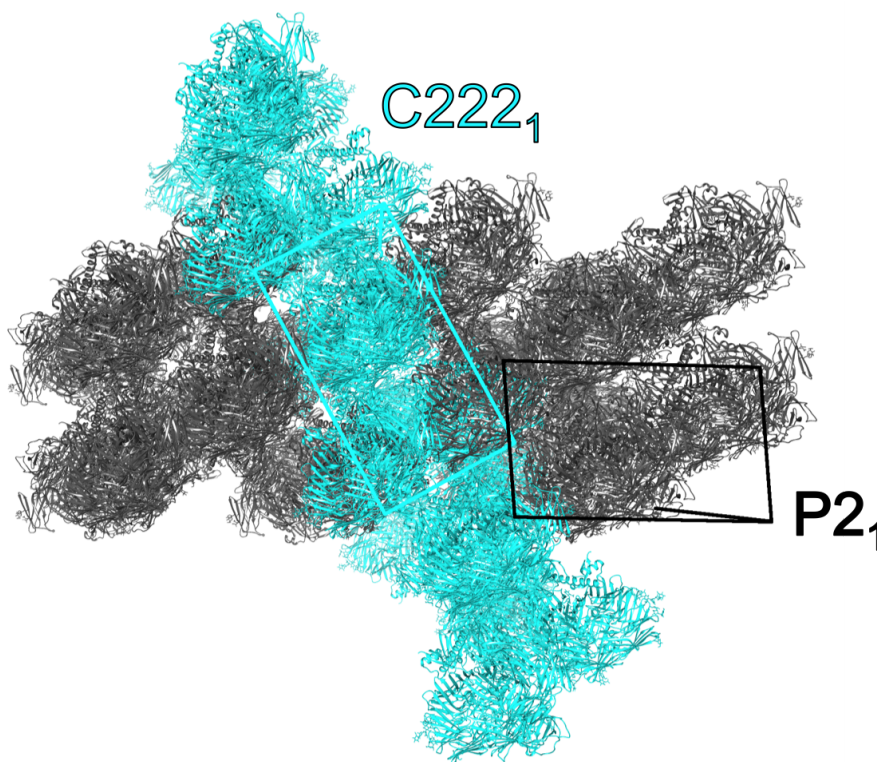


Figure B.1: **GgTen $2^{\Delta\text{EGF}}$ Crystal Packing.** The derivative ($C222_1$, cyan) and native ($P2_1$, black) GgTen $2^{\Delta\text{EGF}}$ datasets share the same crystal packing. The figure was prepared by superposing the GgTen $2^{\Delta\text{EGF}}$ $C222_1$ asymmetric unit onto the $P2_1$ asymmetric unit and then generating symmetry-related molecules for each model. Unit cell outlines are shown in cyan ($C222_1$) and black ($P2_1$).

After refinement of the complete model using the $P2_1$ native dataset, the likely reason for the space group ambiguity became apparent. Of the four copies of

GgTen2^{ΔEGF} present in the P2₁ asymmetric unit, two of the copies (represented by chains A and C) have significantly poorer density for the ABD-like domain. This is particularly obvious for chain C. Note that chains A and C display slightly altered crystal packing compared to chains B and D (Figure B.2).

Removal of the regions which fitted poorly to the density followed by all-atom refinement does not result in the density in this region becoming any clearer, suggesting that this domain is not occupying a discrete alternative conformation in these molecules, and may simply be disordered. Re-processing the native data in the P1 space group also did not result in the density becoming any clearer. Presumably the ABD-like domains of chains B and D are stabilised in one position by their crystal contacts, whereas the slightly altered crystal packing surrounding ABD-like domains of chains A and C allows this domain to be disordered or packed in more than one conformation.

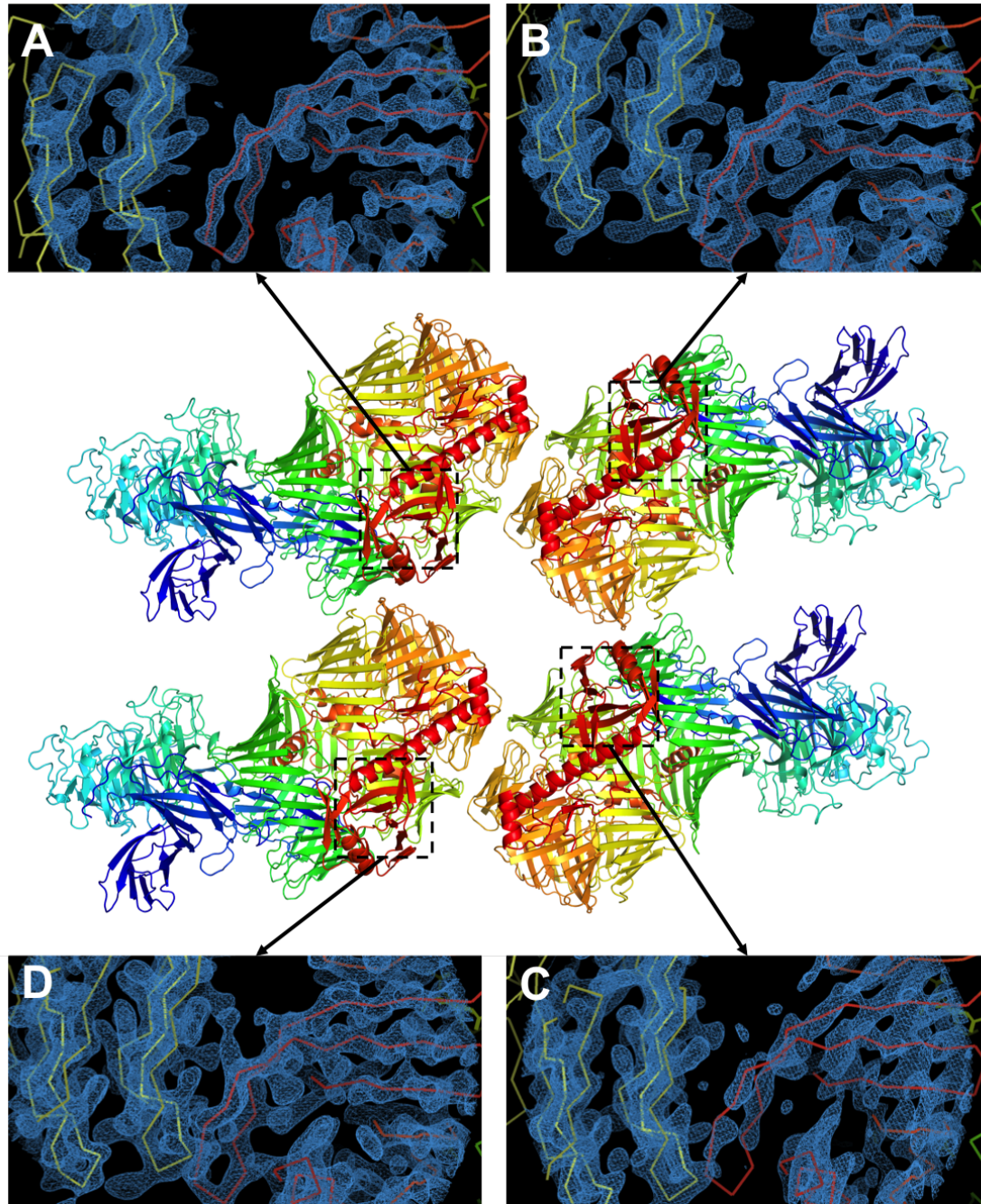


Figure B.2: **Alternative crystal packing of the GgTen2^{ΔEGF} C-terminal domain.** The P2₁ asymmetric unit contains four copies of GgTen2^{ΔEGF} (shown in cartoon representation). Inspection of the electron density (1.5 σ) for each of these four copies reveals striking differences for the GgTen2^{ΔEGF} ABD-like domain only. Chains B and D have clear electron density for the whole of the C-terminal domain, yet the density for the same regions of chains A and C is disjointed. Chains B and D share a similar crystal packing of this domain with a neighbouring molecule, as do chains A and C. Each of the four copies are coloured according to the rainbow (N-terminus = Blue, C-terminus = Red) and symmetry-related molecules are coloured in yellow. Views are centred on residue R2692 of each chain. The chain ID for each chain is displayed alongside the corresponding electron density. The part of the protein affected by this unusual order/disorder phenomenon is surrounded by a dashed box on the cartoon.

Appendix C

Publications Arising From This Thesis

1. **Jackson V.A.**, Mehmood, S., Chavent, M., Roversi, P., Carrasquero, M., del Toro, D., Seyit-Bremer, G., Ranaivoson, F.M., Comoletti, D., Sansom, M.S.P., Robinson, C.V., Klein, R., Seiradake, E. Super-complexes of adhesion GPCRs and neural guidance receptors. *Nat. Commun.* 7, 11184 (2016).
2. **Jackson V.A.**, del Toro, D., Carrasquero, M., Roversi, P., Harlos, K., Klein, R., Seiradake, E. Structural basis of Latrophilin-FLRT interaction. *Structure* **23**, 774-81 (2015).