Investigating the non-protein binding targets of the Teneurin extracellular domain and their biological significance



Owen Crowther

Exeter College

University of Oxford

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

Hilary Term 2022

Investigating the non-protein binding targets of the Teneurin extracellular domain and their biological significance

Owen Crowther

Exeter College
University of Oxford

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

Hilary Term 2022

Abstract

The adhesion protein teneurin is known to control brain development and neuronal migration. Much of the research centered around teneurin thus far has probed its protein binding partners, of which the aGPCR Lphn, and FLRT protein are most extensively studied. One as yet understudied binding partner of the teneurin extracellular domain are the glycosaminoglycan (GAG) macromolecules abundant in the brain extracellular matrix (ECM). These highly charged glycans are made up of five main families: Heparin/ Heparan sulfate (HS); Chondroitin sulfate (CS); Dermatan sulfate; Keratan sulfate; and Hyaluronic acid. These glycans, specifically HS and CS species, have been shown to be important modulators of neuronal migration, as well as factors dictating CNS plasticity and trauma recovery. The work in this thesis proves that binding of HS and CS ligands occurs on a highly charged patch on the large YD-shell scaffold which forms the majority of the teneurin ecto-domain. Two crystal structures were solved here with glycan ligands, proving the binding site contacts 6 Arginine and Lysine residues on the teneurin surface. Subsequent mutation of this binding site abolished binding, and was confirmed by pulldown assay. In vitro cell based migration assays began to unearth the effects of glycan binding of HEK and N2A cell migration, with a view to the effect of glycan binding in the developing brain. This work illustrates the first time teneurin–glycan complexes have been investigated structurally, and represents an introduction into the study of glycobiology with respect to teneurin and how this may control neuronal development.

Acknowledgements

Throughout my DPhil I have been incredibly fortunate to be supported by a network of colleagues, friends, and family. For this I am, and always will be, forever grateful. I am indebted to the Department of Biochemistry for funding my studies.

First and foremost I would like to thank my supervisor, Prof. Elena Seiradake. She has been a constant source of inspiration and encouragement throughout my work, and has helped me to develop into a better scientist in the most welcoming and supportive laboratory environment I could imagine. I would also like to extend the warmest of thank you's to my co supervisor, Dr Kamel El Omari. His expertise and teaching of crystallography, coupled with kind reassurance and seemingly endless patience has made much of the work in this thesis possible.

Many of the skills I learnt were taught to me by the supportive and patient Maria Carrasquero Ordaz. Her incredible knowledge of all the going ons in the lab is astonishing. I must also give a massive thank you to my DPhil compatriot Amy Chu-Antypas. She has been a constant inspiration to me throughout our time together in the lab, always pushing me, whilst offering the perfect distractions when needed. I wish you, Nick, and baby Mia all the best for the future.
My other close friends and colleagues in the lab: Celine Zheng-Gerard, Miguel Berbeira Santana, and Marilee Kokolaki, without each of you, and the other members of the Seiradake lab, my experience and enjoyment would have suffered greatly. I would also like to thank Dr Metin Aksu and Dr Edward Lowe for their advice whenever called upon.

I would not be in a position to complete this DPhil without the support of my friends. In particular, I would like to thank Henry Sawczyc and Joseph Bluck, and Christopher Kin-Cleaves and Kate Dicker. My time at Oxford would have been immeasurably more miserable without you. I must also thank everyone from Exeter College who I have had the pleasure of meeting over my stint at the university. To my school friends: Josef, Yusef, and the whole of the Fellowship, thank you for being an outlet whenever I needed. I must also thank Sandra Ionescu, for being the perfect lockdown companion, and of course Linguini and Ravioli for bringing happiness when I needed it the most.

Finally and most importantly I would like to thank my family, particularly my parents and brother. Your never ending support and belief has motivated me to achieve more than I could on my own. I owe my joy of learning and desire to succeed to you. I cannot thank you enough for everything you have done for me.
Without the belief and support of my former school teacher Dr Andy Love, I would not have the love for science and discovery that I do today. I am forever grateful for the passion that you instilled in me.

Contents

Abstract **iii**

Abbreviations **xi**

1 Introduction **1**
1.1 Neuronal Guidance 1
1.2 Teneurins 4
1.2.1 The Structure of the Teneurin Extracellular Domain 4
1.2.2 Teneurin Binding Partners 8
1.2.3 Teneurin as a Neuronal Guidance Cue 10
1.3 Glycosaminoglycans (GAGs) 12
1.3.1 Extracellular Glycan Families 14
1.3.2 Glycans in Neuronal Guidance 21
1.4 Project Aims 28

2 Materials and Methods **29**
2.1 Molecular Biology and Cloning 29
2.1.1 Genes and Constructs 29
2.1.2 Cloning Vectors and Digestion 29
2.1.3 Polymerase Chain Reaction (PCR) 30
2.1.4 In-Fusion™ Cloning 30
2.1.5 Transformation 31
2.2 Protein Expression and Purification 31
2.2.1 Cell Lines 31
2.2.2 Mammalian Cell Culture Maintenance 32
2.2.3 Small-Scale Transient Transfection 33
2.2.4 Large-Scale Protein Expression 33
2.2.5 Protein Purification 34
2.2.6 Polyacrylamide Gel Electrophoresis 34
2.2.7 Western Blotting 35
2.3 X-ray Crystallography 35
2.3.1 Principles of X-ray Crystallography 35

2.3.2 Crystallisation and Crystal Optimisation 36
2.3.3 Data Collection 39
2.3.4 Data Processing 41
2.3.5 Phasing 43
2.3.6 Model Building and Refinement 45
2.3.7 Validation 48
2.4 Biophysical Approaches 49
2.4.1 Heparin Column Purification 49
2.4.2 Surface Plasmon Resonance (SPR) 49
2.4.3 Streptavidin Bead Pulldown 54
2.4.4 Heparan Sulfate Multimerisation 54
2.5 Cell Biology 55
2.5.1 Cell Surface Staining 55
2.5.2 Glycan Dot Assay 56

3 The Structures of Teneurin–Glycan Complexes 59
3.1 Summary 59
3.2 The Teneurin–SOS Complex 60
3.2.1 Crystallisation of the Teneurin–SOS Complex 60
3.2.2 Data Collection and Processing 62
3.2.3 The Teneurin–SOS Binding Site 71
3.3 The Teneurin–Glycan Complexes 75
3.3.1 Crystallisation of the Teneurin–Glycan Complexes 75
3.3.2 Heparan Sulfate (HS) Complex Structures 76
3.3.3 Chondroitin Sulfate (CS) Complex Structures 84
3.3.4 The Teneurin–Glycan Binding Site 87

4 Biophysical Characterisation of Teneurin-Glycan Complexes 89
4.1 Summary 89
4.2 Mutational Disruption of the Glycan Binding Site 90
4.3 Surface Plasmon Resonance (SPR) 93
4.3.1 Glycan Binding to *Gg*Ten2$^{\Delta EGF}$ 93
4.3.2 Glycan Dependancy of the Teneurin–Latrophilin Interaction 97
4.4 Streptavidin Bead Pulldowns 98
4.5 Teneurin–Heparan Sulfate Multimerisation 103

5 Cellular Effects of Teneurin-Glycan Binding 107
5.1 Summary 107
5.2 *Gg*Ten2^{Full} Teneurin Cell Surface Expression 108
5.3 Glycan Drop Assays 110
5.3.1 HEK Cell Migration 110
5.3.2 N2A Cell Migration 113

6 Discussion and Future Work **119**
6.1 The Teneurin–Glycan complexes 119
6.2 Glycan Binding Candidate Screening 124
6.3 In Vivo Effects of Glycan Complexes 128
6.4 Future Functional Assays 133
6.5 Summary of Findings in this Thesis 137

Appendices

A Constructs and Primers **143**

B Crystallographic Data **145**

C Conservational Analysis **149**

References **153**

Abbreviations

ABD . . . . . Antibiotic Binding Domain

aGPCR . . . . Adhesion G-Protein Coupled Receptor

ASU . . . . . Asymmetric Unit

BME . . . . . β-Mercaptoethanol

BSA . . . . . Bovine Serum Albumin

ChABC . . . . Chondroitinase ABC

CS . . . . . Chondroitin Sulfate

CSPG . . . . . Chondroitin Sulfate Proteoglycan

DMEM . . . . Dulbecco's Modified Eagle's Medium

DRG . . . . . Dorsal Root Ganglion

DS . . . . . Dermatan Sulfate

ECM . . . . . Extracellular Matrix

EDTA . . . . . Ethylenediamine Tetraacetic Acid

FBS . . . . . Foetal Bovine Serum

FFT . . . . . Fast Fourier Transform

FN . . . . . Fibronectin Type-III

FLRT . . . . . Fibronectin Leucine Rich Transmembrane Protein

GAG . . . . . Glycosaminoglycan

GlcA . . . . . Glucouronic Acid

GalNAc . . . . N-Acetylgalactosamine

GlcNAc . . . . N-Acetylglucosamine

HA . . . . . Hyaluronic Acid

HS . . . . . Heparan Sulfate

HSPG . . . . . Heparan Sulfate Proteoglycan

HRP . . . . . Horse Radish Peroxidase

ICD . . . . . Intracellular Domain
IdoA . . . . . Iduronic Acid
IMAC . . . . . Immobilised Metal Affinity Chromatography
KS . . . . . Keratan Sulfate
KSPG . . . . . Keratan Sulfate Proteoglycan
KO . . . . . Knockout
LB . . . . . Lysogeny Broth
LLGI . . . . . Log-Likelihood Gain on Intensity
LN_2 . . . . . Liquid Nitrogen
Lphn . . . . . Latrophilin
LRR . . . . . Leucine-Rich Repeat
MR . . . . . Molecular Replacement
N2A Neuro-2a
NCS . . . . . Non-Crystallographic Symmetry
NDSB-256 . . Non-Detergent Sulfobetaine 256
PBS . . . . . Phosphate Buffered Saline
PBST . . . . . Phosphate Buffered Saline + 0.1% Tween 20
PCR . . . . . Polymerase Chain Reaction
PEI . . . . . Polyethylenimine
PG . . . . . Proteoglycan
PNN . . . . . Perinueronal Net
RT . . . . . Room Temperature
RU . . . . . Response Units
SAD . . . . . Single Anomalous Dispersion
SDS . . . . . Sodium Dodecyl Sulfate
SEC . . . . . Size Exclusion Chromatography
SOS . . . . . Sucrose Octasulfate
SPR . . . . . Surface Plasmon Resonance
TFZ . . . . . Translational Function Z-Score
TLS . . . . . Translation, Libation, Screw
TTR . . . . . Transthyretin-Related
WT . . . . . Wild Type
YD . . . . . Tyrosine-Aspartate

1 Introduction

1.1 Neuronal Guidance

Brain development and pathfinding is an incredibly complicated process, involving many independent factors combining to produce neuronal migration, synapse formation, and preservation. These factors have to be acutely regulated, both spatially and temporally, to result in a fully developed, functional brain. The neuronal cells themselves must be guided from their origin to their destination within the correct time frame, and produce synaptic connections with the correct target cell. This central dogma surrounding cell migration was hypothesised as early as 1892[1], and renewed in the 1960's, with the 'chemoaffinity hypothesis' that proposed the role of chemicals to specify the properties of neuronal cells[2]. The guidance cues that are responsible for this cellular migration are beginning to be understood, with both long and short distance, attractive and repulsive, signalling molecules now being characterised (Figure 1.1A). These guidance cues have been classed into four main families, termed the 'canonical' guidance cues owing to their early discovery: the semaphorins[3], ephrins[4], netrins[5], and slits[6]. As well as these four main groups, many other morphogens have been reported to induce neuronal guidance[7]. These include Wnt, bone morphogenetic protein (BMP), and sonic hedgehog (Shh). A large number of cell adhesion molecules, such as the aGPCR

latrophilin (Lphn) and fibronectin leucine rich transmembrane proteins (FLRTs)[8] also play a role in guidance.

The neuronal cells themselves have developed specialised architecture at the leading edge or tip of their processes to detect the guidance cues and environment. These so called growth cones have evolved to express large numbers of receptors, acting as a focal point for integrating the external signals, as well as cytoskeletal organising proteins to extend or retract the cell process in response to the same signal (Figure 1.1B). The growth cones are formed of two regions, the peripheral domain, and a stable central domain. The peripheral domain is highly actin dense, and these components form different structures close to the membrane. Filipodia are finger like processes navigating the cell membrane towards an attractive signal, where as lamellipoia remain more closely associated with the stable cell sections. The central domain bridges between the leading edge and the rest of the cell. It contains an abundance of microtubule structures, with the function of vesicle transport to the dynamic peripheral domain, as well as mitochondria, working to synthesise ATP for rapid actin turnover and powering the force generating machinery responsible for growth cone motility and axon guidance[9].

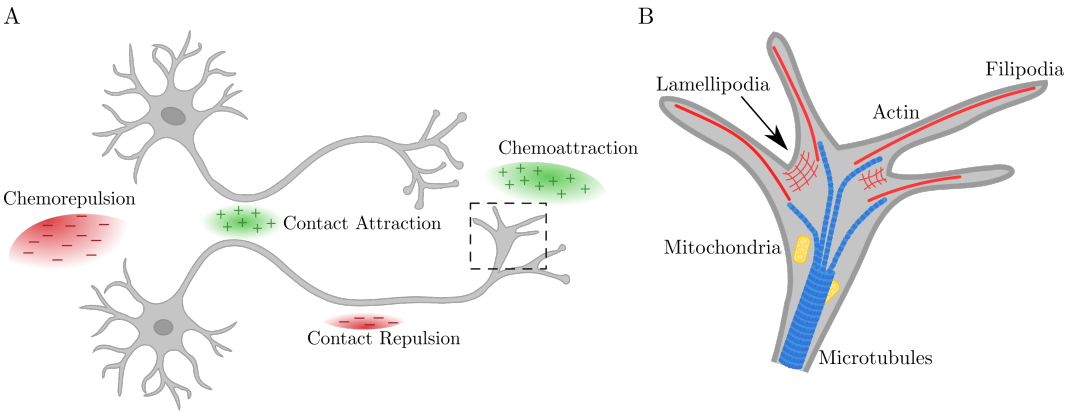


Figure 1.1: Guidance cues in neuronal migration. **A.** Long range guidance cues (chemoattraction and chemorepulsion) direct the neuron to its destination. Short range cues (contact attraction and contact repulsion) regulate cell-cell contact at a closer scale. **B.** A close up representation of the growth cone from A. The actin cytoskeleton (*red*) drives the formation of cellular protrusions (lamellipodia and filipodia) and can remodel in response to guidance cues with the help of abundant mitochondria (*yellow*).

The growth cone has also been shown to remodel its surrounding extracellular matrix (ECM) by secretion of matrix metalloproteinases (MMPs) and disintegrin and metalloproteases (ADAMs) proteins during axonal guidance[10, 11]. For example, the ADAM proteases have been implicated in terminating the high affinity interaction between ephrins and their Eph receptors[12]. It was originally thought that neurons passed through pre-formed channels in the brain tissue, which allowed for paths of least resistance to their target local[13]. This however has since been disproven, and instead it was found that the growth cone can degrade its surrounding ECM to allow more efficient extension into the remaining space[14, 15]. Resent research both in vitro and in vivo has shown that these degradation pathways are very target specific, and can in fact act on their own growth cone to interrupt a continued migratory signal by removing protein receptors on the leading edge membrane[16, 17]. One target for the expressed MMPs are proteoglycans expressing specific sugar species along their structure. The saccharide macromolecules attached to the protein core have been generally accepted to pose as both stimulatory and inhibitory in a species, spacial, and temporal manner, and are described in considerable detail in Section 1.3. The most notoriously inhibitory glycan, chondroitin sulfate (CS) was shown to be inactivated by expressed MMP-2, which recovered the neurite promoting activity of laminin and other ECM components[14]. The interaction between MM-7 and the stimulatory glycan heparan sulfate (HS) has also been described, however the effects of this interaction were not investigated with respect to neuronal migration and development[18].

As well as these proteases, another class of enzymes known as the sulfatases play roles in removing specific sulfur groups from the glycan chains present at the cell surface, in their native environment. This results in modulation of the glycan function over a short time span, where the signalling requirements of the cell can be more rapidly modulated compared to relying on the expression and presentation of the glycans[19]. These enzymes are elaborated upon in Section 1.3.2.

1.2 Teneurins

Outside of the classical neuronal guidance proteins, many other macromolecules play a role in migration, pathfinding, and synaptogenesis. One such neuronal protein are the teneurins. Teneurins are a family of type II cell adhesion proteins expressed in the neurons of many eukaryotic organisms (Ten-1–4 in vertebrates) (Figure 1.2). They mediate cell–cell adhesion at the synapse, by binding to a number of other proteins, and regulate aspects of the nervous system including organisation and development. The evolution of teneurins is still not completely understood, with many theories postulated to explain their unique evolution from ancient bacteria to the simple, single cellular, eukaryote *Monosiga brevicollis* and on to more evolutionarily complex eukaryotes[20, 21].

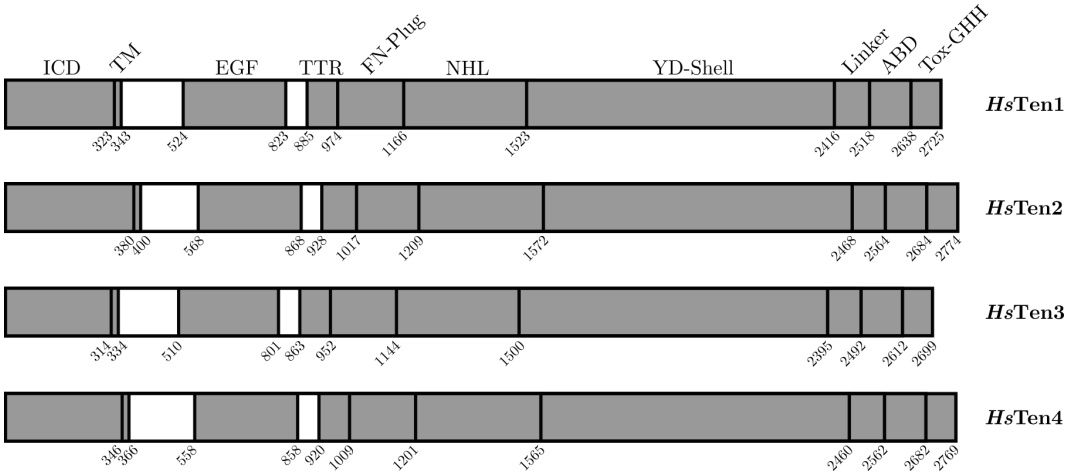


Figure 1.2: Domain architecture of the human Teneurin isoforms. Homo sapien (*HS*) teneurin isoforms are shown, with amino acid boundries for each domain quoted below. *ICD*, intracellular domain; *TM*, transmembrane; *EGF*, epidermal growth factor-like repeats; *TTR*, transthyretin-related; *FN-plug*, fibronectin-plug; *NHL*, NCL, HT2A and Lin-41 like domain; *YD-shell*, tyrosine-aspartate repeat shell; *ABD*, antibiotic-binding domain-like; *Tox-GHH*, toxin like domain.

1.2.1 The Structure of the Teneurin Extracellular Domain

The first crystal structure of the large, 207 kDa, *Gallus gallus* teneurin-2 extracellular domain was solved in 2018[22] allowing the function and interactions to be probed in

more detail (Figure 1.3). The elucidation of the crystal structure revealed a surprisingly compact structure (approximately 7.2 nm x 12.6 nm) accounting for around 210 kDa. The N-terminus of the protein is formed from a transthyretin-related (TTR) domain, which are common among all domains of life. In other examples, TTR-like domains are associated with lipid recognition[23] and aggregation[24]. Following this domain, a fibronectin type-III (FN) domain extends to form a plug for the large centralised β-barrel shell cavity. This domain makes close contacts to P1680, a mutation of which in Ten-1 is associated with congential general anomsima (P1610L), and has been displayed in mice to cause olfactory deficit[25]. Proceeding this plug domain, the NHL domain wedges against the main scaffold, held in place by the N-terminal domains. This NHL is formed of a 6-bladed propeller domain. The exposed loops in this region are thought to be important for trans interaction between opposing teneurin molecules, as this region includes a splice site (R1303–H1309) which has been implicated in the ability for teneurin isoforms to create this architecture. The largest feature of the teneurin extracellular domain is composed of a large 78-stranded YD-shell, forming a single ~800 residue β-sheet. This large cage like domain is made up of repeating Tyrosine and Aspartate residues, and has recently been discovered to contain the binding site for the Lphn and FLRT protein binding partners described below. It also bears structural similarities to the bacterial TcB-TcC toxin[26]. Extending from the large shell, an internal linker composing 96 amino acids seals exposed faces of the YD-shell at both the top and alongside the FN-plug, filling most of the open ~130 nm^3 cavity, leaving only 44 nm^3 solvent accessible. This linker extends out of the YD-shell, through a small gap allowing the C-terminus of the protein to reside on the outside of the scaffold. The antibiotic binding domain (ABD) shows similarity to bacterial small molecule antibiotic binding domains. Finally, the C-terminal Tox-GHH domain has structural similarity to that of bacteriocins, most closely related to Colicin E9 (r.m.s.d. of 2.65 Å), a family distinct from the TcB-TcC toxins.

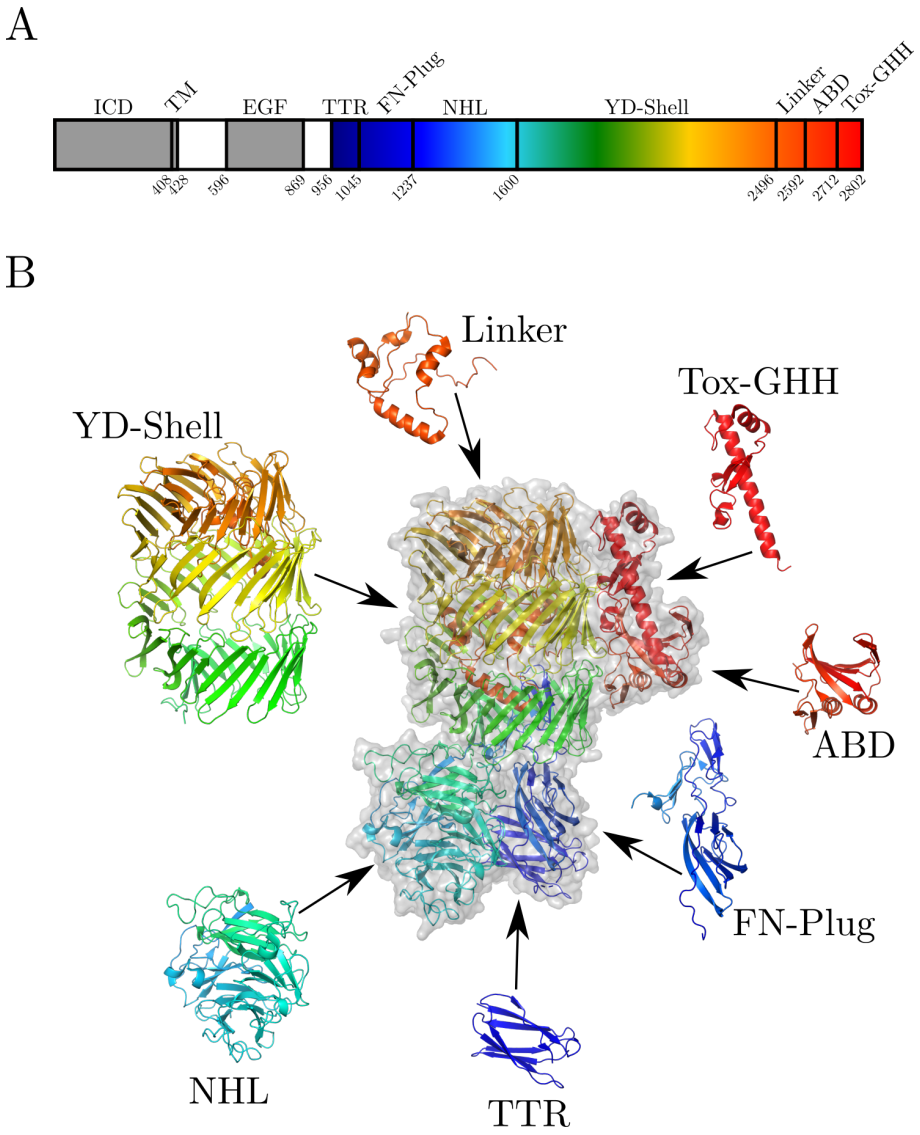


Figure 1.3: Domain architecture of the Teneurin Ecto-domain. **A.** Domain organisation of the whole *Gallus gallus* teneurin 2 protein. *ICD*, intracellular domain; *TM*, transmembrane; *EGF*, epidermal growth factor-like repeats; *TTR*, transthyretin-related; *FN-plug*, fibronectin-plug; *NHL*, NCL, HT2A and Lin-41 like domain; *YD-shell*, tyrosine-aspartate repeat shell; *ABD*, antibiotic-binding domain-like; *Tox-GHH*, toxin like domain. **B.** The crystal structure of the teneurin extracellular region[22], represented as a cartoon and coloured from *blue* to *red* from the N- to C-termiuns. Each domain is shown individually with respect to the whole structure, PDB ID: 6FB3.

Teneurin is known to be able to form both homophilic cis and trans dimers. It has been well documented that two teneurin proteins expressed on the same cell surface are able to dimerise via free cysteine residues within their EGF repeats. These residues are found on EGF repeat 2 and 5, and facilitate the formation of the native

teneurin dimer which is found on the cell surface. Trans interactions between two separate cells have also been described, although less extensively. Teneurin has been shown to be able to form these interactions in a splice variant dependant manner. Each of the four teneurin isoforms in humans utilise two extracellular splice sites (one located in the EGF repeats and the other in the NHL domain). These splice sites have been shown to control the ability of each teneurin isoform to interact with both its own isoform and the others. This gives rise to a complex network of potential trans interactions in the brain, and could be partly responsible for the intricacies needed for the correct function of teneurin during development.

When comparing the eukaryotic teneurin structure to that of the ancient bacterial TcB-TcC toxin from which teneurin is postulated to evolve and shows structural homology, the cage domain is very highly conserved (Figure 1.4). The main structural difference between the two proteins occurs at the N-terminus of the scaffold, involving the plug region which seals the YD shell. Teneurin is plugged by a combination of the FN-plug and the β-propellers contributed by the NHL domain. In the TcB toxin, this plug is formed by a SpvB domain which is wedged in place, in much the same way as teneurin, with a single β-propeller region. The more simplified plug architecture found in the TcB toxin still maintains the homophilic interaction described above in teneurin. In the case of the bacterial toxin this allows the TcB-TcC toxin complex to attach to the pentameric TcA toxin to form the holo-toxin required for delivery[27]. Interestingly, in the bacterial protein, the cage encapsulates a toxin domain, which the bacteria present and inject into a target cell. This Tox-GHH domain is still present in teneurin, although appearing on the outside of the YD shell rather than insulated inside. In eukaryotes this domain has been shown, somewhat controversially, to be autoproteolytically cleavable, like the bacterial toxin domain[28]. It has been implicated in vitro to control dendritic morphology[29, 30], and in vivo it has been shown to control anxiety behaviour in rats[30, 31]. This Tox-GHH domain has been shown to bind and cleave a specific DNA sequence originating from mitochondria, even though the DNase catalytic

triad (HNH) is mutated and was presumed non-functional[32].

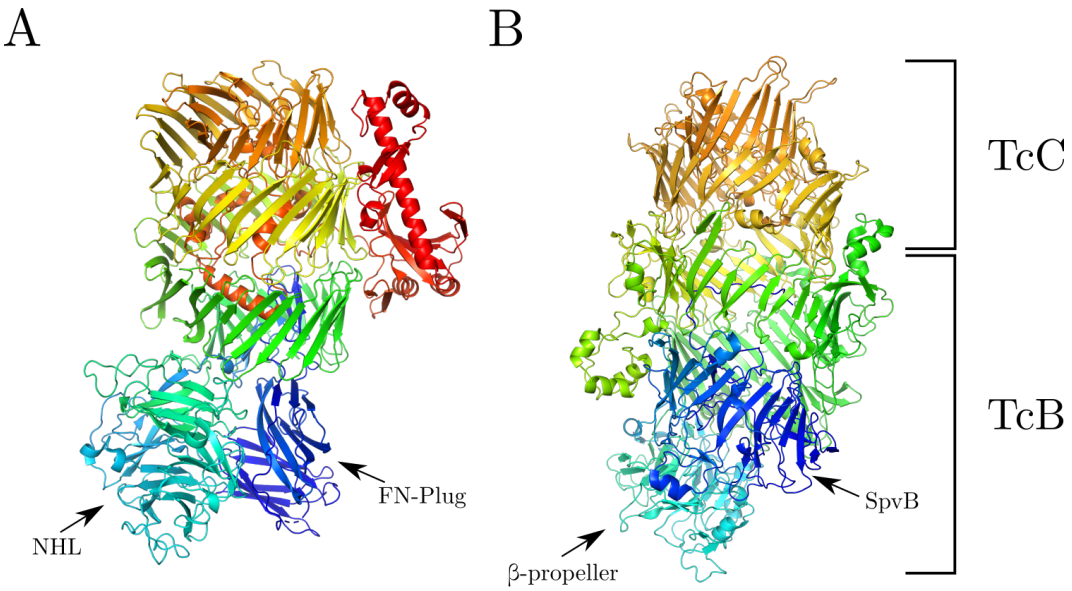


Figure 1.4: Comparison between the Teneurin and TcB-TcC toxin structures. **A.** The structure of the teneurin extracellular domain, highlighting the NHL and FN-plug which seal the large YD shell. PDB ID: 6FB3. **B.** The structure of a TcB-TcC toxin complex, not including the hyper-variable toxin domain. The similar YD shell is plugged by a SpvB and β-propeller domain. PDB ID: 6H6G.

1.2.2 Teneurin Binding Partners

Much of the existing research surrounding teneurin focusses on the protein binding partners involved in teneurin signalling, mainly within its role in neuronal pathways. Classically, teneurin has been described as a binding partner for the G-protein coupled receptor (GPCR) Latrophilin (Lphn). Together these two proteins are co-expressed in the developing brain and have been shown to have overlapping expression patterns[33, 34]. Historically, the binding of these two proteins was proven via non-structural cell binding methods[35], and there have been some differing hypotheses for the exact binding site for Lphn on the teneurin extracellular domain. However, more recently crystal structures elucidated the binding site directly, in more detail[36, 37]. The binding site has been shown to cover much of the surface of the YD-shell of teneurin. Two crystal structures were solved, initially showing that

the minimal binding domain of Lphn, the Lec domain, covers approximately 1500 Å^2 of the teneurin surface. A second crystal structure, containing the proceeding Lphn domain, the Olf domain, reveals a larger total binding site of around 2500 Å^2 (Figure 1.5). In contrast to previous indirect methods, the location of the binding site was shown to be on the YD-shell surface away from the Tox-GHH C-terminus, whereas previously published work had postulated that the Tox-GHH domain is involved in Lphn binding[38].

Another protein binding partner for teneurin is the protein FLRT (Fibronectin leucine-rich repeat transmembrane protein), when already complexed with Lphn. Alone, teneurin is unable to bind to the Leucine Rich Repeat (LRR) domain of FLRT2. The Latrophilin Olf domain is the primary binder to FLRT. However, when teneurin and Lphn are complexed together, the affinity for FLRT increases compared to Lphn alone[36, 39]. It has therefore been postulated that teneurin can act to form a super-complex by increasing the affinity of Lphn and FLRT. This has led to the belief that a super-complex of teneurin-Lphn-FLRT is possible to form in vivo, and work has been ongoing to further characterise it.

As described above, the vast majority of research involving teneurin and its binding partners has centred around protein–protein interactions. This has left the interactions with other abundant macromolecules less understood. Indeed, the interaction between teneurin and extracellular glycans, as well as potential interactions with lipids, have been studied sparingly, if at all. The field of glycobiology is still emerging, and is beginning to be appreciated as important as protein binding with respect to modulating correct function in the organism. An association between heparin and teneurin has been known for some time[40]. At the time, teneurin was first being described in *Drosophila*, found using different approaches for homologues of known targets[41–43]. This paper was the first to describe a vertebrate homologue of the *Drosophila* Ten-m; discovering teneurin-1, named for its expression in the brain. Recombinant teneurin-YD repeats were retained on a heparin column, only eluting with an increased salt concentration to 350 mM

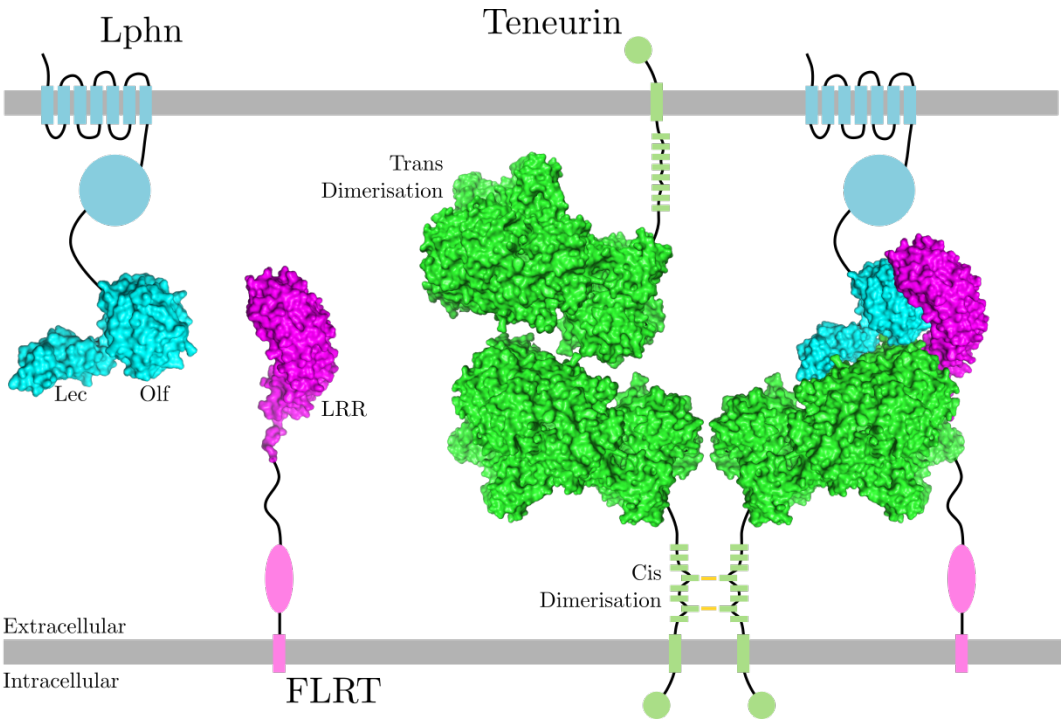


Figure 1.5: Teneurin and its protein binding partners. Known protein binding partners for teneurin are shown in their proposed organisation. The aGPCR Lphn (*cyan*) is known to bind to teneurin (*green*), and is thought to be expressed by the opposing cell to the teneurin (*right*)[36]. PDB ID: 6SKA. The protein FLRT (*pink*) is also a known binder, and is thought to contribute from the same cell as teneurin, this can be superimposed onto the Teneurin-Lphn structure. Teneurin itself is able to form cis-homodimers through open disulfide bonds on EGF repeats 2 and 5, as well as a trans-dimerisation, although work is still ongoing to show this structurally.

(above physiological conditions). They identified aggregate species of YD-shell that could be released after incubation with heparinase, and more effectively with chondroitinase, revealing a binding of teneurin to multiple glycan species.

1.2.3 Teneurin as a Neuronal Guidance Cue

Vertebrate teneurin has classically been identified as a synaptic protein, expressed at the synapse in the developing brain to form connections across the synaptic cleft. In *Drosophila*, teneurin has been known to mediate protein-protein interactions between the presynaptic motor neurons and postsynaptic muscles via a homophilic trans interaction[44, 45] as well heterophilically[46].

In the vertebrate systems studied, the role of teneurin is closely related to that of its latrophilin binding partner. The interaction between the Olf and Lec domains of Lphn and teneurin plays a role in normal brain development, and disruption of this interaction impairs synaptic density[47], however the role of the Lphn-FLRT interaction which binds to the same Lphn domain also disrupts the normal wiring[48]. In the model organism Zebrafish, knockout of *teneurin-3* results in impaired retinal ganglion cells, resulting in incorrect orientation selectivity[49].

Alongside the extracellular interactions of teneurin at the synapse, there are also relevant intracellular roles for teneurin. The intracellular domain (ICD) has been shown to be cleavable and act as a transcription factor on genes involved in cytoskeletal organisation. These genes, specifically involved in the remodelling of actin inside the cell[45, 50], are believed to be the pathway by which teneurin can modulate cell migration and pathfinding in the developing brain. The exact roles of multiple neuronal proteins on the cytoskeleton is still not fully understood. Many proteins with vast effects on the cytoskeleton, for example Nectins[51] and Wnt[52–54], can be mutated to cause catastrophic effects. Other proteins involved, Neurexin and Neruoligin, play a role in cytoskeletal organisation which is still retained after protein mutation[55, 56], however teneurin disruption has more of a catastrophic effect on the cytoskeleton[45]. This redundancy in cytoskeletal organisation hints at its importance for cellular function and specifically synapse formation. The roles of Teneurin and Neuroligin specifically have been compared and shown that Teneurin is primarily involved in mediating cytoskeletal order, but have minor roles in synapse formation and retention; Neuroligin was shown to have these roles reversed[45].

More recently, spearheaded by research in the Seiradake lab, the existence of teneurin at an earlier stage of development has been investigated. Teneurin was known to be expressed during embryonic stages for the development of synapses (~E17), as described above, but has been shown to be expressed earlier, at a stage of neurodevelopment preceding synapse formation (E13.5)[36]. This earlier co-expression with the protein binding partners described above has been shown to

affect cell migration in the developing mouse brain. It has been shown that teneurins and FLRTs are mainly expressed on the migrating neuron, and latrophilins on the neuron and radial glial cells, which act as a scaffold to direct correct migration and pathfinding. The interactions between the three proteins was shown to have a repulsive effect on migration, which is surprising considering the synaptic role of this protein complex results in the attractive formation at the synaptic cleft. This reduced migration with teneurin binding may be due to the proteins inherent repulsive effects, where expression, or over-expression in the study, results in neuronal cells adhering to the extracellular matrix far stronger than with natural teneurin levels. This is not simply an effect of spatial expression, as teneurin and FLRT are uniformly expressed along the neuronal surface, in both the cell body and axonal terminals during migration. Recent work has also described a reciprocal interaction between teneurin and latrophilin[57]. It was shown that Ten3–Ten3 interactions promote migration, and spatial expression of Lphn in neighbouring brain regions repulses Ten neurons to guide them to their correct destination where they are attracted by homophilic binding. This double headed interaction relationship has been described before, as with Erythropoietin producing human hepatocellular (Eph) receptor and ephrin signalling[58], as well as in the reverse manner with Semaphorin 3A[59]. Semaphorin 2A and Semaphorin 2B have also been identified to act reciprocally with the PlexB receptor, mediating repulsion and attraction, respectively[60].

1.3 Glycosaminoglycans (GAGs)

Glycosaminoglycans (GAGs) are polysaccharides made up of repeating disaccharide units that are expressed in every mammalian tissue. They are almost exclusively found attached to a core protein to form a proteoglycan (PG), where specific protein motifs allow for glycan addition in a tissue and time dependent manner. The superstructure of these glycans, that is the modifications that can be present on their chain, are modulated by enzymes expressed in the ER and Golgi. These are primarily

responsible for addition of sulfur groups onto the sugar, ultimately this sulfation is controlled by the expression patterns of the facilitating sulfotransferases[61, 62].
Specific GAGs are implicated in a wide variety of biological functions, most notoriously with anticoagulation[63–65], and cartilage compression[66]. The inherent heterogeneity of the glycan structures and sulfation patterns, as well as their size, flexibility, and difficulty to isolate, has hindered glycan research. This is no more apparent than with structural studies, particularly X-ray crystallography, where a very limited number of structures have been published with bound glycan ligands, owing to difficulty in purifying large enough amounts of homogenous, pure glycan, and isolating a single conformation in the crystal lattice. The structural studies that have been performed discovered that the glycan chains actually adopt a roughly helical structure in solution studies[67, 68], and utilise their inherent flexibility to conform to the binding site when complexed with protein binding partners[69–71].

One of the common starting points for GAG binding protein identification is therefore using glycan mimics, commonly heparin (for binding and in vitro studies), or small disaccharide sugars such as Sucrose octasulfate (SOS) for structural investigation. Heparin is comonly used as it differs only from the more physiologically relevant heparan sulfate (HS) by the density of sulfation. Infact, heparin is identical to the most highly sulfated patches of HS, and as such this mimics the highest density charge based interaction between a protein and an identical HS patch possible. This initial screening can however falsely open avenues of research for proteins that bind to the structurally distinct heparin but not HS, or vice versa exclude non-heparin binding proteins that could in fact bind to specific glycan families or sub-species of glycans within families.
Glycans are isolated and purified by exploiting their naturally long linear chains, and high density of negative charges caused by varying sulfation levels. The glycan chains are isolated from the core protein to which they are bound by degradation using proteolysis and cleavage of the polypeptide, where the free glycan can subsequently be purified by strong anion-exchange (SAX) chromatography[72].

Further disaccharide analysis is one of the most important tools for downstream GAG production. The fact that the small disaccharide repeats from a large sugar chain are fundamental in their function has necessitated the ability to isolate and study specific small glycans. This product is achieved by sequential degradation of longer 'parental' glycans by glycosidic enzymes. These enzymes act on specifically sulfated patches of the sugar chain, and thus allow isolation of different sized, homogeneously sulfated, glycans from a larger mix. These small fragments are then often identified using further techniques such as HPLC[73, 74] or ESI-MS[75] when compared to a reference standard.

1.3.1 Extracellular Glycan Families

The GAGs can be categorised into 5 families, shown in Figure 1.6, based upon their repeating disaccharides: Heparin/ Heparan Sulfate (HS); Chondroitin Sulfate (CS); Dermatan Sulfate (DS); Keratan sulfate (KS); and Hyaluronic acid (HA).

Heparin and Heparan sulfate (HS) are the most abundant extracellular glycans expressed in the body. Heparin has classically been characterised as an anticoagulant as it binds to and activates the enzyme inhibitor antithrombin III, which itself then inactivates various clotting factors[76]. The physiological role of heparin is not well understood, but it is thought that its anticoagulation role is a by-product of an anti-bacterial role, this is supported by the fact that heparin is secreted exclusively by mast cells, and is only released at the site of vascular injury. Due to its highly sulfated disaccharide repeats (Figure 1.7), it has the highest negative charge density of any biological molecule (~3.3 negative charges per disaccharide)[77]. This lends itself to biophysical applications investigating charge based interactions and a result of it this, it is commonly used as an indirect method to detect DNA binding proteins by pulldowns[78].

Heparan sulfate (HS) is a more ubiquitously expressed glycan, with regions of high and low sulfation along its chain. The regions of highest sulfation are

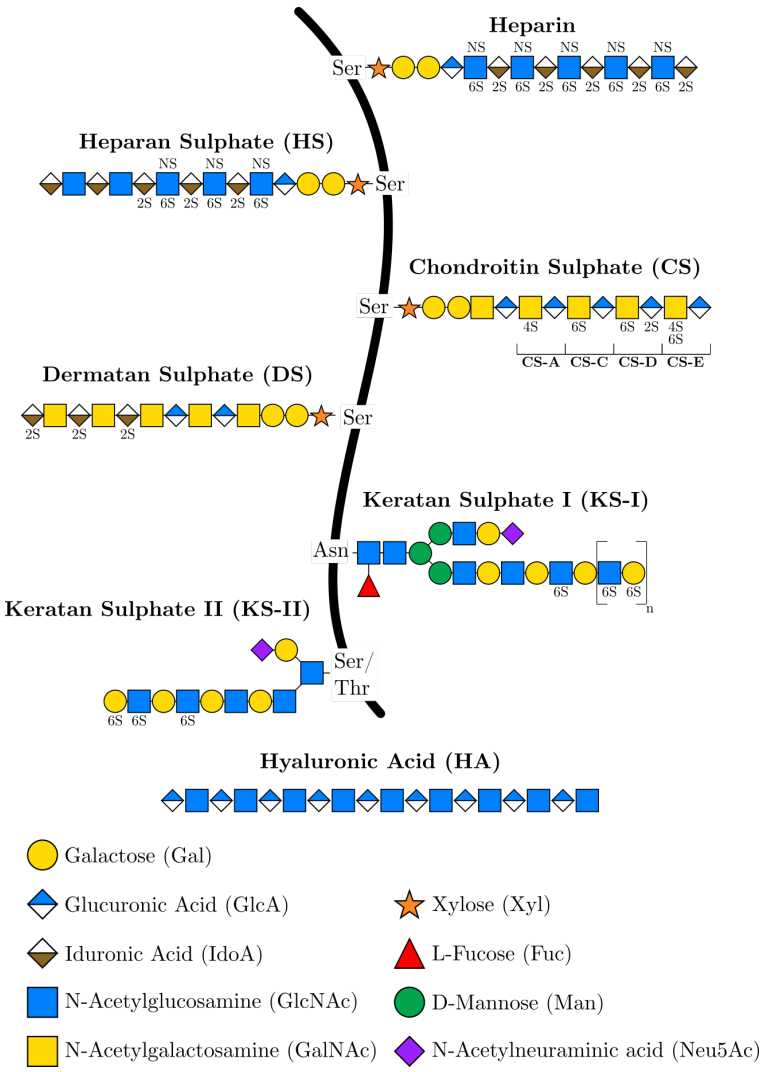


Figure 1.6: Structures of Glycosaminoglycans (GAGs). Glycans are represented by their standardised symbols. Heparin and Heparan sulfate (HS) are differentiated by their degree of sulfation. Chondroitin sulfate (CS) has sub-families made from differentially sulfated repeats. Dermatan sulfate (DS) is closely realted to CS. Keratan sulfates are classified depending on their protein binding residue. Hyaluronic Acid (HA) is the only GAG not bound to a core protein.

indistinguishable from pure heparin, and as such the long (parental) HS chain can be thought of as regions of heparin interspersed with regions of non sulfated sugar backbone. It is produced by every cell type[79], finally being identified in erythrocytes after a dedicated search effort[80]. HS is presented by the cell, as with all of the GAGs except Hyaluronic Acid, attached to a core protein, forming a Heparan sulfate proteoglycan (HSPG). The HSPGs only constitute a very small family,

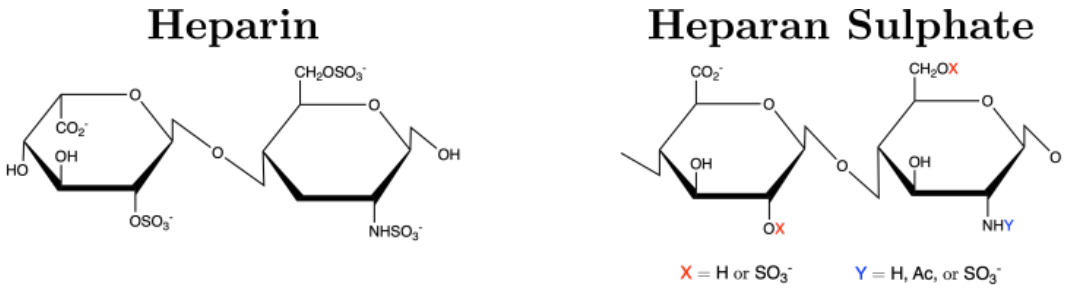


Figure 1.7: Detailed structures of the Heparin and Heparan sulfate disaccharides. Both Heparin and Heparan sulfate are made up of repeating Iduronic acid (IdoA), and N-Acetyleglucosamine (GlcNAc) monomers. The difference between the two species is the degree of sulfation along the sugar chain, with heparin more densely studded with sulfur groups.

made up of 16 members classified into three groups depending on their sub-cellular location: integral membrane HSPGs, secreted ECM HSPGs, and secretory vesicle proteoglycans. The biggest class of HSPGs, the membrane bound proteins, consist of the Syndecans (SDC-1–4), Glycpicans (GPC-1–6), Betaglycan, and CD44v3. The ECM proteins include Perlecan, Agrin, and Collagen XVIII. The only vesicle bound HSPG is the protein Serglycin which has the highest number of HS chains attached to its core (Figure 1.8).

HSPGs have been described as regulators of Wnt ligand dispersal in *Drosophila*, by enhancing signalling by Wnt ligands or by delivering the ligands to neighbouring cells[81, 82]. With respect to brain development, the most important class of HSPG are the integral membrane proteins, specifically GPC-1, GPC-2, and SDC-3. Specifically, glypican proteins have been shown to play important roles for signalling at the synapse, where they can interact with multiple protein binding partners across the synaptic cleft. With dis-regulation of their HS chains, the synapse is destabilised, which leads to malfunction and is associated with autism spectrum disorder[83]. Glypican 1 has been shown to control mammalian brain size, with a reduction in overall cell numbers due to a specific relationship between progenitor cell proliferation or differentiation through its signalling partners in early neurogenesis[84]. This is a phenomenon that has been measured with GPC-2, as it has been reported to indicate neurogenesis in adult brain[85] as well as promoting cell differentiation when

applied exogenously[86]. Both of the effects of Glypican expression have been shown to be modulated through signalling via FGF receptors, highlighting the effect of multiple signalling molecules in close proximity.

Syndecans have also been studied for their roles in neuronal development, with research discovering a function more aligned with neuronal migration than cell proliferation. Studies have shown that syndecan 3 knockout mouse brains exhibit structural changes in the adult cerebral cortex, as a result of delayed cell migration[87]. Syndecan 1 is highly expressed at the growth cone where it stabilises this structure in the migrating cell[88]. After injury it is further up-regulated, along with the HS synthesising enzyme EXT2, suggesting a role in neuronal regeneration after injury[89]. As syndecans are the only membrane spanning HSPG with a large intracellular domain, they are beginning to be investigated as signal transducers. This would allow signals from binding to either the HS chain or core protein to be acted on inside the cell. Alongside these HS-presenting proteins, the expression pattern of HS-modifying enzymes are also important for neuronal patterning and guidance, dictating the amount and place of sulfation on the HS sugar chains, which modulates the interacting partners.

Chondroitin sulfate (CS) is formed of a different repeating disaccharide to heparin and HS. This family of glycans is formed of repeating N-Acteylgalactosamine (GalNAc) and Glucuronic Acid (GlcA). They have a generally shorter chain length, averaging between 2 and 50 kDa. The most interesting feature of the CS chain are the distinct sub-populations of CS formed by differential sulfation locations within the disaccharide. The detailed disaccharide structures of each isoform are highlighted in Figure 1.9. These sub-populations are often found in distinct regions along a larger parental CS chain which can contain any number of sub-classes of CS along its structure. These differing regions are known to modulate different functions, we believe by promoting different binding partners at different distances from the core protein. This is especially well known with respect to neuronal guidance and development, which is described in more detail in Section 1.3.2. CS,

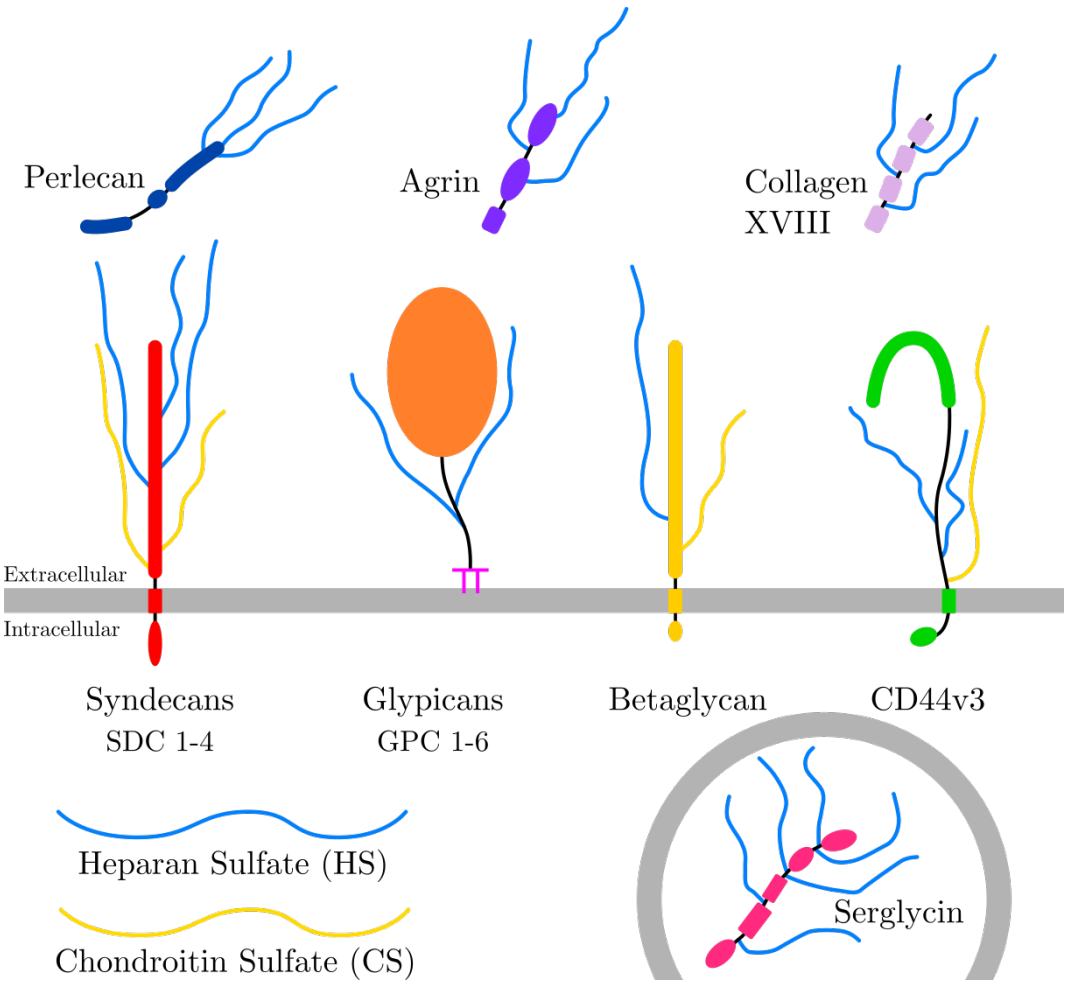


Figure 1.8: Schematics of the Heparan Sulfate Proteoglycans (HSPGs). Representations of the HSPG families are shown. Many contain not only HS (*blue*), but also CS (*yellow*) chains presented on their surface. All HSPGs are found at the membrane or in the ECM, except for syndecan which is vesicle bound.

and their associated proteoglycans, are the most common extracellular glycans in the brain ECM, and have been implicated in various conditions, particularly with respect to axon regeneration and repair after trauma. CS has been shown to be generally inhibitory to migration and re-growth, with the sub-classes allowing for higher fidelity control.[90, 91]. The CS glycans also play roles in inflammation and immune modulation[92], the maintenance of cartilage structure and function, and the inherent shock absorbing properties of cartilage tissue[66].

Dermatan sulfate (DS) was classified as a distinct family of GAG after initially

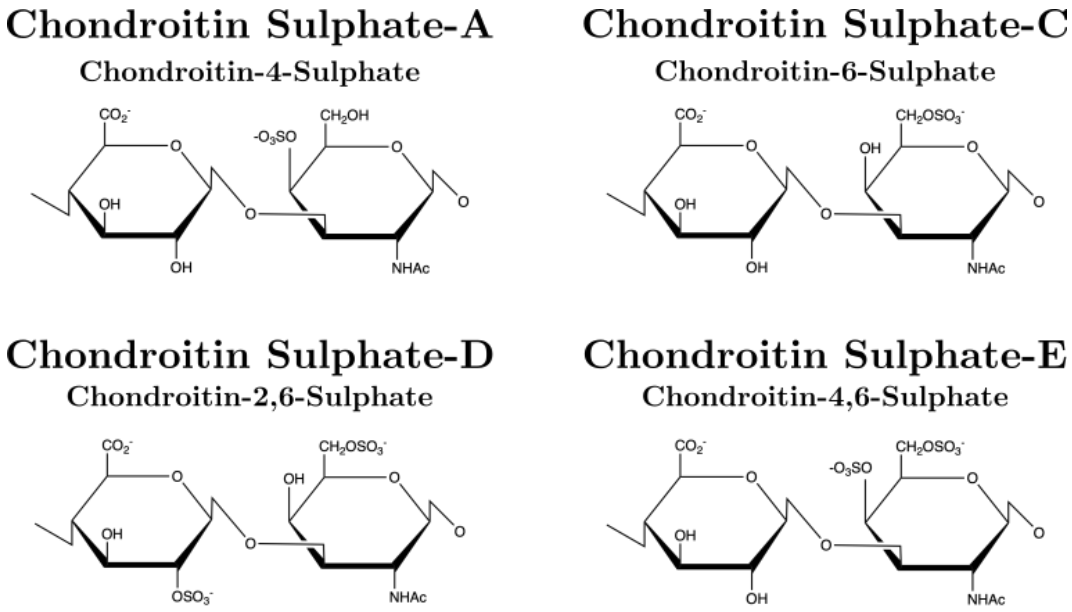


Figure 1.9: Detailed structures of the Chondroitin sulfate disaccharides. Each sub-class of CS is classified based on the sulfation amount and location on the disaccharides. Physiologically, whole CS strands are formed of patches of different sub-classes, each of which can modulate potential function, and are not exclusively made of a single form.

being classed as CS-B due to its close structural similarities to the other chondroitin sulfates, and its inclusion of GalNAc (Figure 1.10). DS plays roles in a number of cellular and tissue specific functions, including stabilisation of cell–cell and cell–ECM interactions, as well as stabilisation of the basement membrane. The best studied proteoglycans containing Dermatan sulfate are Versican and Biglycan, which have been shown to be expressed in bone, cartilage, and skin[92–94]. The protein Biglycan also contains CS glycans attached to its core, and as such the role of DS when attached to this protein closely mirrors chondroitin sulfate.

Keratan sulfate (KS) is evolutionarily the youngest GAG, and is the least well understood. It has a diverse structure, being the only glycan to include a branched chain, and a repeating unit missing the characteristic Uronic acid monomer found elsewhere (Figure 1.11). KS has two forms originally classified by their expression pattern and protein attachment point. KS-I was originally discovered as the main glycan in the cornea, and attaches to its core protein via an Asn residue. KS-II

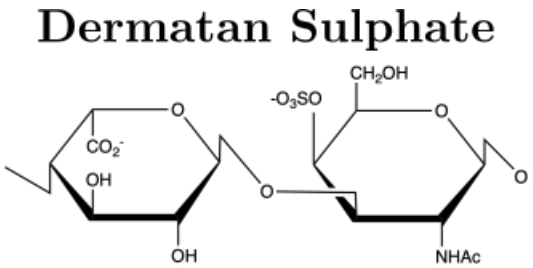


Figure 1.10: Detailed structures of the Dermatan sulfate disaccharides. DS was originally classed as CS-B, due to its inclusion of GalNAc and similarity in abundance to the other chondroitin sulfates.

has been found more widely expressed, and attaches to the core protein via Serine or Threonine residues utilising a GalNAc sugar[95]. It has since been found that KS-I is expressed more widely than its original source, and as such KS classification depends only on the core protein attachment point. Keratan sulfate proteoglycans (KSPGs) have been identified in numerous epithelial and neural tissues in which KS expression responds to embryonic development, physiological variations, and wound healing[96, 97]. Evidence has also been presented supporting functional roles of KS in cellular recognition of protein ligands, axonal guidance, cell motility, and in embryo implantation[98].

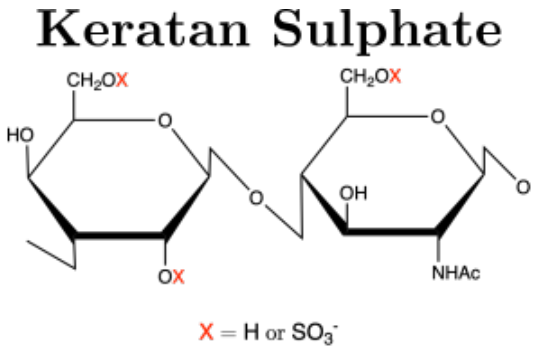


Figure 1.11: Detailed structure of the Keratan sulfate repeating disaccharide. Keratan sulfate is the only branched GAG, and can attach to the core proteoglycan on differing side chains, giving rise to the two KS isoforms.

Hyaluronic Acid (HA) is the only glycosaminoglycan not found bound to a core protein. It is instead secreted by the cell directly into the extracellular space. HA

is synthesised by the integral membrane protein family hyaluronan syntheases by sequentially adding GlcA and GlcNAc to form the full chain[99], which is then extruded through the cell membrane via an ABC transporter[100, 101]. Hyaluronic Acid is also the only completely un-sulfated glycan. It is found as very long linear chains, up to 25,000 repeating disaccharides (up to 20 million Da), making it the longest GAG in the body. Depending on the tissue and required function the glycan is produced to different lengths. For example, in synovial fluid, the average mass of HA is 7,000 kDa (7 million Da), compared to in the umbilical cord where a smaller form (3,000 kDa) is primarily found. The roles of this glycan are mainly involved in soft connective tissue, cartilage elasticity and its shock absorbing properties. Because of its roles in cartilage maintenance and cushioning, HA injections are a common treatment for osteoarthritis to improve the condition of the soft tissue[102]. The glycan can be degraded by a class of enzymes called hyaluronidases, of which there are 7 isoforms in humans. These enzymes are involved in tumour suppression[103], and as such the role of HA in tumour survival and angiogenesis is of clinical relevance[104].

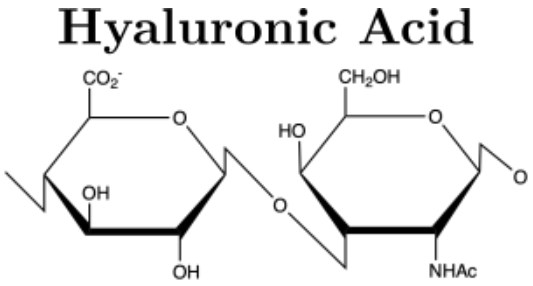


Figure 1.12: Detailed structure of the Hyaluronic Acid repeating disaccharide. HA is the only GAG to not be attached to a core protein, instead being secreted directly into the ECM. It is also the only entirely unsulfated glycan species.

1.3.2 Glycans in Neuronal Guidance

As touched upon above, glycosaminoglycans play a role in a variety of functions, at both a system and cellular level. Of specific interest to my work, and this thesis, is the role these GAGs play during brain development, how they effect axon guidance

and neurite growth towards their target, and then how these GAGs play a role in synaptogenesis. An extension of this function may also have relevance with respect to recovery after trauma to the CNS, with spinal cord injury and regeneration a target for medical research and treatment.

Heparan Sulfate in Neuro–development

Vertebrate studies to identify the function of HS relied on conditional knockout (KO) methods, or using rescue experiments. KO of the mouse HS polymerase (EXT1) significantly impairs neuronal migration, resulting in an abnormal morphology[105]. Further work focusing on removal of the *ext1* gene from the mouse spinal cord inhibited the netrin dependent guidance of axons in the spinal cord, revealing a link between the HS glycan and Netrin signalling[106], thought to be as a result of the glycan's role in the Slit-Robo axon guidance signalling cascade[107]. In Zebrafish, a similar function for both EXT2 and EXTL3 has been shown, with both genes important for retinal axon patterning. EXT2 functions to correctly extend sensory axons to the skin[108], and EXTL3 controls axon guidance of dopaminergic neurons[109]. In *C. elegans*, Heparan sulfate regulates a wide range of neuronal cell migration[110].

The role of the Heparan sulfate sugar is often closely associated with that of the core proteoglycan (HSPG) to which it is attached. The expression of these proteins dictate the pattern and prevalence of the sugars within regions of the organism, as well as the expression of the glycan synthesising and modifying enzymes within the cell. Work in this avenue has exploited the fact that *C. elegans* have single genes for all HS modifying enzymes (with the exception of 3-*O*-sulfotransferase which has two genes)[111]. This has allowed for investigation into the role of specific HS sulfation, and the effects that the inclusion or exclusion of certain sulfate groups has on signalling. It has been discovered that the axonal growth and patterning required different HS modifications. As such, mutation of a certain HS modifying enzyme

is not catastrophic to the organism, as only single pathways are disrupted[112]. Further work has also identified that the same neuronal populations require different modifying enzymes in different contexts. PVQ neurons in *C. elegans* require 6-*O*-sulfotransferase and 2-*O*-sulfotransferase enzymes for nerve projection in some cases but not others. It has also been discovered that the modifying enzymes, specifically 6-*O*-sulfotransferase, can act as signalling initiators in incorrect locales. The enzyme was misexpressed in the skin, where it is not normally found, and caused motor neuron patterning issues throughout the organism. This could be partially rescued by removal of HSPG core proteins, demonstrating a potential function of HS sulfation in the control of signalling during neuronal development[113]. KO mutants of some HS modifying enzymes in mouse models has allowed for their investigation in a more complex organism[114]. Of particular interest and importance is the study of retinal axons during development. Both 6-*O*-sulfotransferase and 2-*O*-sulfotransferase modifying enzymes are necessary for proper pathfinding and axon guidance to their final positions in the optic chiasm[115]. HS modifying enzymes have also been shown to effect migration rates in the formation of mouse corpus callosum, where loss of function of 6-*O*-sulfotransferase and 2-*O*-sulfotransferase leads to over migration of the glial cells past their target location[116].

Alongside these modifying enzymes are the incredibly important extracellular Endo-sulfatases. These degredaing enzymes are expressed in the ECM, where they act to cleave specific classes of glycans in their native, signal tranducing, state. These enzymes therefore unsuprisingly play roles in limiting signal transduction at the leading edge of the cell. Specific enzymes, for example Sulf1 and Sulf2, catalyse the de-sulfation of heparan sulfate glycans present in their vicinity, thus reduced the signalling capabilities requiring these sugar moieteies. Research has shown that this results in reduced signalling through the intreaction of HS with FGF and VEGF partners[117, 118], wherease with regard to Wnt signalling modulation as was decribed above, Wnt is released when the HS chains are desulfated by Sulfatase as there are less intreactions between the glycan and ligand[119–121]. The

Sulf2 enzyme itself even presents a CS/DS chain on its structure, which plays an important role in modulating its own activity. When the glycan is present, Sulf2 activity is reduced due to partial blocking of the enzyme active site by the long sugar chain interacting with the binding site. Upon action of hyaluronidase to cleave the CS/DS chain, the Sulf enzyme activity is promoted, realising its effects as a pro-oncogenic marker[122]. These sulfatase enzymes are incredibly important with regards to signal modulation during neuro–development, by fine tuning the signalling requirements at the membrane, offering a fast response compared to the expression of new or modified glycans from within the cell.

Chondroitin Sulfates in Neuro–development

Chondroitin Sulphate chains are the most abundant GAG in the neuronal ECM, and are best known for their ability to inhibit regeneration after CNS trauma. However, CS is also known to pay important roles in development. The discovery of these functions were hindered by the fact that, unlike HS, Chondroitin Sulphate is only sparsely sulfated in invertebrates[123, 124]. Therefore, much of the research has been performed in more complex vertebrate model organisms. Like above, migration of the optic neurons also requires CS, although for different aspects of retinal axon guidance to the optic chiasm[125, 126]. The growth cone of retinal axons was negatively affected by enzymatic degradation and removal of CS chains, showing differential morphology in vitro[127]. Again, as with HS, the importance of the CS producing and modifying enzymes has been studied quite considerably. The variation in length of the GAG, as well as the sulfation pattern, are major modulators of overall function. Through KO models, it has been shown that uronyl 2-*O*-sulfotransferase and N-acetylgalactosamine 4,6-*O*-sulfotransferase are necessary for cortical neuron migration in vivo[128]. Interestingly it was also shown that KO of dermatan 4-*O*-sulfotransferase-1 had no phenotypical effect on migration. CS has also been shown to be required for Semaphorin 3A repulsion of cortical neurons. Expression of specific CSPGs (mainly versican and neurocan) have been shown to inhibit axon guidance[129], neurite outgrowth[130], and dorsal root ganglion

growth[86] in vitro. This inhibitory affect was overturned after treatment with the CS degrading enzyme chondrotinase ABC[86, 131].

It has long been thought that the human CNS was unable to regenerate after injury. This is largely the case, although there have been reports of supraspinal neurons mounting a short regenerative response to trauma, however this required a permissive environment[132, 133]. So far this short, unsuccessful, response has not been resolved, although the post injury environment has been identified as incredibly inhibitory. The expression of many proteins and constituents of this CNS environment after trauma are therefore of particular interest with respect to solving the CNS regeneration problem. Chondroitin Sulphate proteoglycans (CSPGs) are of particular importance in this process, as they are widely present, and often up-regulated, in these inhibitory environments. The core protein certainly plays a role in the function of these CS sugar chains (the same sugar bound to an aggerecan has different function compared to it bound to a versican protein), however the biological role is also influenced by the sugar sulfation pattern. As described in the previous section, Chondroitin Sulphate can be found with a number of different combinations of sulfation pattern (Figure 1.9). The most common forms of CS are CS–A (4S) and CS–C (6S). These are the most widely studied thus far, and have been implicated in neuronal development. CS–A (4S) has been shown to be highly inhibitory[90, 134]. Some reports state that the 6S form of CS has the potential to actually promote axonal growth, although this is still a controversial finding[91], and contradicts the long standing consensus from other labs[135]. Interestingly the disulphated CS forms, CS–D and CS–E, are both accepted as positive regulators of neuronal growth during development[136, 137]. In the developing brain, the ratio of these CS forms is thought to play a role, with more inhibitory glycans being expressed further along the development process, after much of the migration has occurred. Measurement of specific CS forms in the mouse brain shows an increase in the inhibitory CS–A (4S) glycan from 58% (P7), through 76% (P14), 81% (P21), and 87% (adult). It is hypothesised that this increase in the abundance of the inhibitory CS–A (4S) glycan could coincide with the final stages of migration, where

the cell slows to find its final target and form a synapse, where migration is less abundant. The ratio of CS–C (6S) was reduced over the same time period (28%, 14%, 9%, and 3%), and so was the ratio of disulfated CS–E (4,6S)[138]. This work is corroborated by other studies which have shown the same pattern for CS–A and CS–C during development, although with slightly different values[139].

Chondroitin Sulfates in Trauma Recovery

With respect to neuronal injury, CSPGs are up-regulated at the site of injury, in an architecture known as the glial scar. After trauma, the immune system secretes chemokines that act on a number of the surrounding cells including astrocytes at the border of the injury. The astrocytes go through a process known as reactive astrogliosis, become hypertrophic, and secrete vast numbers of CSPGs[140, 141]. This leads to the formation of the glial scar, a chemophysical barrier blocking axonal regeneration. Further invading macrophages can themselves also secrete CSPGs (containing mainly inhibitory CS–A), and are thought to be a source of the proteoglycans after injury[142, 143]. This response at the injury site imposes an inhibitory environment which blocks the regenerative migration of neuronal cells for recovery[144, 145]. In adult mouse dorsal root ganglion (DRG) axons, this inhibitory effect is mediated to some extent by binding of the CSPG to RPTPσ[146]. Interestingly, in the developing chick environment, DRG axons are promoted by interaction with RPTPσ when interacting with HSPGs. This further illustrates the opposing effects of HS and CS in vertebrate models. After injury, structures known as the perinueronal net (PNN) are formed to protect the surrounding cells from damage, and to encourage inflammation and immune response to treat the injury. These nets are highly abundant in CSPGs, which form the inhibitory environment blocking neuronal migration. As described above with the dynamic CS content of the developing brain, the PNNs present during injury are also dynamic. They go through the same shift as described above, initially containing more CS–C (6S) glycan and switching to the CS–A (4S) inhibitory form over time[147]. This may be

acting as the mechanism by which these nets inhibit recovery and cause plasticity in the CNS.

As a result of the complex inhibitory effect of CS regarding neuronal migration after injury, treatments have been targeted towards removing this inhibitory sugar from the ECM and environment to allow migration to then occur. One of the main treatments for this is the application of chondroitinase enzymes to degrade the sugar chains. Chondroitinase enzymes are classed by their target sequence, with different forms cleaving at specific points with respect to sulfated disaccharides. This can result in specifically sulfated products, and is one of the main methods for isolation of the sugar for biochemical research. In a therapeutic setting, ChABC, an enzyme capable of impartially cleaving the sugar at all sites, is being researched as a potential treatment to degrade the inhibitory environment. This has been used in spinal cord injury and has led to improved repair in both partial[148] and complete[149–153] spinal cord transection. The enzyme has been shown to cleave the PNN formed post trauma, and allows for the beginnings of axonal growth and cell migration to recover after trauma[154]. Unfortunately, there is evidence that CSPGs are turned over after ChABC treatment, and new fully glycated proteins are expressed after around 14 days post enzyme injection[155]. It would therefore require careful repeated ChABC treatment throughout the timescale of neuronal regeneration to allow for recovery. This would have to be balanced with the physiological role of the glial scar and PNN to protect the surrounding tissue, as repeated degradation of this protective site results in more severe damage[156].

1.4 Project Aims

The first indication of teneurin interacting with glycosmainoglycans was proposed in 1999[40]. This paper gave the first characterisation of the interaction between recombinant teneurin YD-shell fragments and heparin. They identified that treatment with heparinase and chondroitinase released the bound protein, and that neurite outgrowth on YD-repeat surfaces is inhibited by the introduction of a glycan. Unfortunately, no structural insights were available at the time, and as such an accurate binding site could not be identified.

This work will aim to expand on the little knowledge that currently exists for this interaction, by utilising structural, biophysical, and cell based assays to probe the interaction of teneurin and glycans.
The aims of the work presented in this thesis can be broadly covered by three main targets.

- Structural investigation of Teneurin–Glycan complexes.
- Biophysical analysis of glycan binding.
- Studying the in vitro effects of glycan binding

2
Materials and Methods

2.1 Molecular Biology and Cloning

2.1.1 Genes and Constructs

The Teneurin construct used throughout this report (Accession number XP_015149056.1) was subcloned from cDNA from the mammalian gene collection[157] and was previously cloned by Dr Verity Jackson[158] (Appendix A Table A.1). For recombinant protein expression, this protein construct was cloned into a pHLSec mammalian expression vector [*Addgene plasmid #99845*][159]. This vector contains an N-terminal signal peptide originating from receptor tyrosine protein phosphatase S, and a C-terminal hexahistidine tag. The teneurin construct was cloned into this plasmid utilising flanking *AgeI*/*KpnI* restriction sites to include the above features.

2.1.2 Cloning Vectors and Digestion

Primers for cloning were designed to allow for adequate template binding and insertion overlap. A template binding region of at least 20 bp was included in each primer sequence, as well as an overlap of 15 bp to the opposite primer in the centre of the inserted sequence. The primers used for cloning are summarised in Appendix A, Table A.2. The vector was removed from the whole plasmid by restriction using AgeI and KpnI [*New England BioLabs*]. 6 μl vector plasmid and

2 units of each restriction enzyme were combined in NEBuffer 1.1 [*New England BioLabs*] and incubated at 37℃ for 3 hr.

2.1.3 Polymerase Chain Reaction (PCR)

Mutant teneurin constructs were generated by polymerase chain reaction (PCR) using the CloneAmp HiFi PCR Premix [*Takara-Bio, Clonetech*] according to manufacturer's instructions. The PCR program used is detailed in Table 2.1.

Table 2.1: PCR program for amplification of the teneurin fragments.

Cycles	Temperature (℃)	Time
10x	98	10 s
	55	15 s
	72	5 min
25x	98	10 s
	55	15 s
	72	3 min
1x	4	∞

2.1.4 In-Fusion™ Cloning

Prior to DNA ligation, both the linearised vector and mutant PCR product were mixed with 6x DNA loading dye [*Thermo Fisher*] and electrophoresed on a 1% agarose-TAE gel for 60 min. DNA bands were visualised using SYBR™ Safe (1 : 10,000 dilution) and a G:BOX Chemi XRQ chemiluminescence gel imager [*SYNGENE*]. Desired DNA bands were extracted using a NucleoSpin Gel and PCR Clean-up kit [*Fisher Scientific*], as per the manufacturers protocol.

An In-Fusion™ HD cloning plus kit [*Takara Bio*] was used for construct–plasmid ligation. 50 ng of both the linearised vector and PCR product were combined with In-Fusion™ premix and adjusted to 5 μl MilliQ H_2O. The infusion reactions were incubated at 55℃ for 15 min.

2.1.5 Transformation

Stellar chemically competent cells (*E. coli* HST08) [*Takara*] were transformed with 3 μl of the In-Fusion™ reaction. Cells were incubated on ice for 30 min, before heat shocking for 45 s at 42℃ and returning to ice for 5 min. Transformed cells were plated on lysogeny broth (LB) agar plates containing 100 μg/ml ampicillin and incubated overnight at 37℃ in a static incubator. Plasmid DNA extraction was performed with the QIAprep Spin Miniprep Kit [*Qiagen*], according to the supplied protocol. To confirm cloning, samples were sequenced by Sanger sequencing [*Source Biosciences*].

2.2 Protein Expression and Purification

2.2.1 Cell Lines

Adherent HEK293 cells were used for all secretory recombinant protein expression described in this thesis. A hallmark of eukaryotic expression is the addition of *N*-linked glycosylation to specific asparagine residues within the protein sequence. Once a mature glycan is formed at the endoplasmic reticulum (ER), this is attached to the protein by recognition of an Asn–*X*–Ser/Thr motif, where *X* can be any amino acid, except proline. This addition of varying oligosaccharides assists with the proper folding of complex protein structures, but is inherently incompatible with downstream applications such as X-ray crystallography, where homogeneous samples are required[160]. To reduce the heterogeneity of glycosylation, HEK293T cells used for protein expression were treated with kifunensine. This compound is an α-mannosidase I inhibitor, which restricts the cells to produce only oligomannose-type *N*–glycans[161]. Without kifunensine treatment, complex glycans are present that cannot be later cleaved with the endoglycosidase F1 (Endo F1) or Peptide-*N*-Glycosidase (PNGase) enzymes due to the presence of a shielding fucose sugar moiety (Figure 2.1).

For later cell migration assays, Neuro-2a (N2A) cells were kindly donated by Prof Esther Becker. This cell line is derived from mouse neuroblastoma, and

has been extensively used to study neuronal migration, differentiation, signalling, and axonal growth.

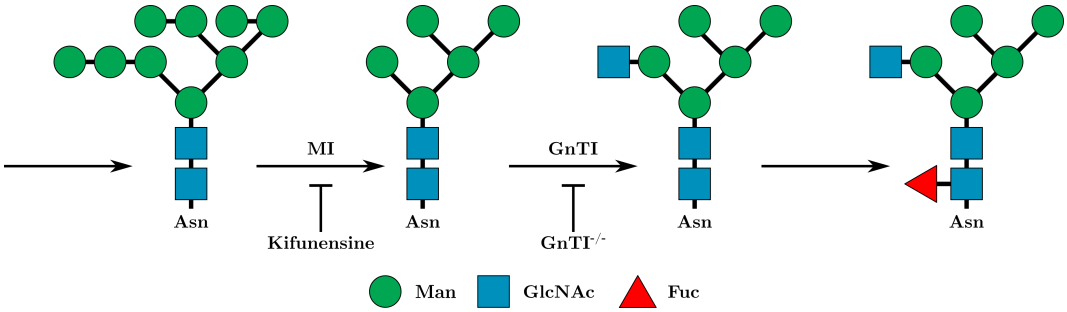


Figure 2.1: Inhibiting N-glycosylation maturation. N-Acetylglucosamine (*GlcNAc*) depicted as blue squares, Mannose (*Man*) shown as green circles, and Fucose (*Fuc*) shown as a red triangle. Inhibition of α-mannosidase I (*MI*) by kifunensine stops the desired glycan structure from being further processed into more diverse species. Cell lines can be generated that lack the N-acetylglucosaminyltransferase I (*GnTI*) enzyme and are thus only able to produce smaller glycans than the kifunensine treated populations (*GnTI*). Without regualtion of the N-linked glycan species present in the cell, complex glycans which are not transferable to structural studies such as X-ray crystallograhy are present (*right*).

2.2.2 Mammalian Cell Culture Maintenance

HEK293T cells were maintained in T75 flasks [*Griener-Bio*] 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*]. Cell lines were maintained by splitting once 90-100% confluency was reached. Cell media was removed by aspiration and cells were carefully washed with 10 ml PBS [*Lonza*]. PBS was removed before addition of 2 ml 0.05% trypsin-EDTA [*Gibco*], the flask was then incubated for 4 min (37℃, 5% CO_2). After incubation the trypsin-cell solution was diluted with 8 ml DMEM (10% FBS) and seeded in a new flask, containing pre-warmed fresh media, at the required density. Fresh cell flasks were stored at 37℃ in 5% CO_2.

N2A Cells were maintained in the same manner, however these cells were grown in 10% FBS media containing 1% L-Glutamine and 1% Penicillin-Streptomycin solution. The old media was aspirated, and the cells washed with 2 ml 0.25%

trypsin-EDTA [Gibco]. Attached cells were detached with another treatment of 2 ml 0.25% trypsin-EDTA [*Gibco*] which was incubated for 5 min (37℃, 5% CO_2). Detached cells were then seeded into fresh media in a 1 : 5 dilution.

2.2.3 Small-Scale Transient Transfection

24 hr prior to transfection, 100% confluent mammalian cells were seeded, at a density of 3.0 x 10^5, into 6-well plates and grown at 37℃, 5% CO_2. For transfection, 100 μl Opti-MEM [*Gibco*] was added to a sterile 1.5 ml eppendorf tube. 3 μg DNA was added to the tube with either polyethylenimine (PEI) [*Sigma*], or Fugene® [*Promega*] at a ratio of 3:1 (PEI/Fugene:DNA). This transfection mixture was briefly vortexed and incubated for 10 min. The incubated transfection cocktail was added to the side of each well gently to avoid disturbing the cells, before the plate was carefully mixed and incubated at 37℃, 5% CO_2 for 3 days. Secreted protein was harvested by centrifugation in a sterile eppendorf in a bench top centrifuge (max speed, 4 min).

2.2.4 Large-Scale Protein Expression

To allow for sufficient levels of adherent cells to be grown, the cell stocks were scaled up to T175 flasks [*Greiner-Bio*] and allowed to reach confluency before being transferred into 2125 cm^2 sterile grooved roller bottles [*CELLMASTER*; *Greiner-Bio*]. 250 ml of DMEM, supplemented with 10% FBS, was added to each bottle along with 10 ml of suspended cells from the large 175 cm^2 cell culture flasks. Roller bottles were incubated for 2 days at 37℃ in a roller bottle incubator [*Wheaton*].
The DNA used for transient transfection was prepared from Gigapreps [*Thermo Fisher*], as per the provided protocol. Purified DNA was sterilised by adding 100 μl chloroform per 1 mg DNA aliquot and briefly vortexed. The DNA was centrifuged at 20,000 x g for 10 min. The transfection cocktail was prepared by mixing 0.5 mg DNA with 1 mg PEI in 50 ml serum-free DMEM, per bottle, which was then incubated at room temperature (RT) for 10 min. 0.38 mg of kifunensine was added to each transfection cocktail post incubation. The growth medium in the roller bottles was discarded and replaced with 200 ml DMEM supplemented with 2%

Foetal Bovine Serum (FBS), 1% non essential amino acids, and 1% L-Glutamine, before the transfection cocktail was added. After transfection, the roller bottles were returned to the incubator for 4 days, before harvesting by centrifugation (7,000 rpm, 4℃) and sterile filtering through a 0.22 μm polyethersulfone (PES) filter [*Millipore*]. Filtered supernatant was stored at 4℃ until purification could be performed.

2.2.5 Protein Purification

The filtered cell culture medium was concentrated and the medium was exchanged into diafiltration buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 137 mM phosphate buffered saline (PBS)), using a tangential flow filtration device [*Pall*]. The medium was concentrated to approximately 200 ml and was loaded onto a 5 ml HisTrap FF column [*GE Healthcare*] at a rate of 5 ml/min. The column was washed with 150 ml wash buffer (20 mM Tris, pH 7.5, 300 mM NaCl, 40 mM imidazole). The resulting bound protein was eluted into 1.5 ml fractions in elution buffer (20 mM Tris, pH 7.5, 300 mM NaCl, 500 mM imidazole).

A Bradford assay was carried out on the eluted fractions to qualitatively determine the highest protein sample. 90 μl Bradford's Reagent [*Sigma*] was mixed with 10 μl sample and sealed overnight to assess protein content.

Size Exclusion Chromatography (SEC) was carried out with a superdex 200 16/60 HiLoad column [*GE Healthcare*] equilibrated with gel filtration buffer (10 mM Tris, pH 7.5, 150 mM NaCl). HisTrap elution samples were loaded onto the column and the resulting 1.5 ml fractions collected in a sterile 96 deep-well plate [*Greiner*]. Fractions with highest $UV_{280\,nm}$ trace from the elution profile were analysed by denaturing polyacrylamide gel electrophoresis (SDS-PAGE) to visually confirm protein expression.

2.2.6 Polyacrylamide Gel Electrophoresis

Samples were mixed with 6x SDS-loading dye (375 mM Tris-HCl, pH 6.8, 6% SDS, 48% glycerol, 0.03% bromophenol blue and 5% β-mercaptoethanol (BME)). NuPAGE™ 4-12% Bis-Tris gels were use with NuPAGE™ MOPS running buffer

[*Thermo Fisher*] to visualise the high molecular weight protein. The gels were run in an XCell surelock™ Mini-Cell [*Thermo Fisher*] at 180 V until completion. Protein bands were visualised using InstantBlue™ Coomasie stain [*Expedeon*]. Gels were removed from their casing and stained for at least 15 min at RT, or until the bands became visible. For increased sensitivity, the gels were left staining overnight.

2.2.7 Western Blotting

Western blotting was used to detect protein expression of epitope-tagged samples. An unstained SDS-PAGE gel was transferred onto nitrocellulose membrane [*GE Healthcare*] using a sandwich of 5 sponges and 2 blotting papers [*Whattman*] soaked in transfer buffer (1x NuPAGE™ Transfer Buffer, 20% ethanol) and run in a NuPAGE™ XCell II™ blot module [*Thermo Fisher*] (30 V, 90 min). Blocking of the membrane was achieved by addition of 3% bovine serum albumin (BSA) dissolved in 1x PBS with an additional 0.1% Tween 20 [*Fisher*] (PBST) for 30 min. Mouse anti-pentaHis [*Invitrogen*] primary antibody was diluted 1 : 1,000 in 3% BSA for 60 min at RT. The membrane was then rinsed with PBST three times before addition of the goat anti-mouse [*Sigma*] secondary antibody conjugated to horse radish peroxidase (HRP) diluted 1 : 10,000 in BSA for 60 min. The secondary antibody was removed and the membrane was once again washed with PBST for 3 cycles and then left in PBST for 30 min. Development was performed using the Amersham™ ECL™ western blotting detection kit [*GE Healthcare*] and bands were visualised using Amersham Hyperfilm™ ECL™ [*GE Healthcare*].

2.3 X-ray Crystallography

2.3.1 Principles of X-ray Crystallography

X-ray diffraction experiments employ the ability of an electron environment within the target molecule, most commonly a macromolecule such as a protein, to diffract the incident radiation beam and generate diffraction 'spots' on a detector; which can be recorded and back transformed to generate a molecular model of the structure. X-rays have a wavelength in the same order of magnitude (0.7 - 1.54 Å) to that

of the bond lengths in a protein (C-C bonds are 1.54 Å), so in ideal cases the resolution allows individual atomic details to be resolved. Unlike in other structural biology approaches, crystallography requires the molecule of interest to be uniformly packed in a crystal lattice, giving a time and space average of multiple molecular conformations in one final model. To generate a diffraction spot, incoming X-ray radiation must elastically scatter off the electron environment, preserving the wavelength and energy of the incident beam. This elastic scattering contributes only a small percentage of the total effect of the X-ray beam. The vast majority of the total X-ray beam is not scattered, instead being stopped by a physical beamstop which prevents damage to the detector. Still, the majority of the remaining radiation is absorbed by the electron causing radiation damage in the crystal. Computationally, the diffraction pattern from a crystal can be determined using the known intensity and relative position of each spot or reflection, and the calculated phases. A model can then be generated to best fit the data collected, and the structure probed to understand details about the target.

This section discusses the aspects of single-crystal X-ray crystallography used throughout this thesis, including the pipeline used to determine the structure of the complexes solved in this work.

2.3.2 Crystallisation and Crystal Optimisation

In order to perform X-ray diffraction experiments, the protein of interest must be able to pack to form well ordered crystal lattices from an initially homogenous sample. These protein crystals are obtained when the sufficiently highly concentrated, pure, protein is mixed with a precipitant in the correct conditions (pH, salt concentration, temp, etc.) to favourably drive the protein out of solution. When the protein and precipitant concentrations are too high, the constituents are likely to form amorphous material. Too low of a concentration of one or both of the protein and precipitant will allow the protein to stay in solution. At optimal concentrations, crystal nuclei are able to appear, which leads to crystal growth via incorporation of

more soluble protein into the lower energy crystalline state.

The most common approach to generate crystals within macromolecular crystallography is the vapour diffusion method. This employs a drop containing a mixture of protein and reservoir solution, typically containing different mixtures of precipitant, buffer and various additives. Vapour diffusion relies on the gradual diffusion of water from the protein containing sample drop to the reservoir, which contains a higher concentration of the constituent mixture. This results in a gradual increase in both protein and precipitant concentration, which, if successful, reaches the nucleation zone allowing crystal nuclei to form. Once a crystal nuclei has been generated, the incorporation of more protein during growth sees a sharper decrease in the protein concentration. Sitting drop vapour diffusion, an approach utilising this principle, was employed throughout this thesis. A phase diagram of vapour diffusion (Figure 2.2) illustrates the formation and growth of crystals using this approach.

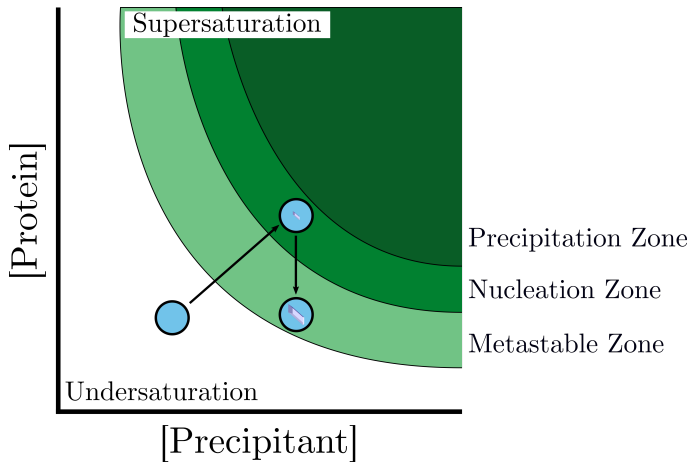


Figure 2.2: Phase Diagram of Vapour Diffusion Crystallography. Initially a drop is set up with a known concentration of precipitant and protein. By gradual diffusion the concentration of all components in the drop increase until the crystal enters the nucleation zone where a nucleating crystal can from. As the crystal grows, the component concentrations in the drop solution decrease until it reaches the undersaturated zone.

The protein concentration required for successful crystallisation is routinely calculated empirically. Because the protein construct used for crystallisation in this thesis

has been successfully crystallised previously by Dr Verity Jackson[158], the starting concentration was matched to the previous result. Non-detergent sulfobetaine 256 (NDSB-256) was added to the concentrated protein to prevent aggregation in the Eppendorf. To empirically assess the crystallisation conditions required in each experiment, multiple sparse matrix and grid screens are commercially available, making use of conditions and buffer combinations that have commonly been successful in crystallising previous proteins[162]. Different crystallisation conditions were trialled from the Morpheus I[163] and II[164], JCSGCore I-IV and JCSG+[165, 166], Proplex[167], Midas[168], PGA[169] [*all Molecular Dimensions*], and Index [*Hampton Research*] screens. The use of these screens depended on the protein yield from each purification, and successful previous attempts. Invariably Morpheus I and II were used in every case. The Morpheus screen is a three-dimensional grid which tests four precipitant mixtures, three buffer systems, and eight classes of additive. Morpheus II builds on the original Morpheus screen, testing less commonly observed reagents. The JCSG I-IV sparse matrix screens were individually tested if the protein yield allowed. If not, JCSG+, a data mined consolidation of the four core screens would be used. Following these initial screens, if protein yield permitted, further screens would be tested.

General crystal plate setups were performed as follows. 96-well 2-drop Swissci plates containing 50 μl commercial screen reservoir solution were prepared by Edward Lowe [*Crystallography facility, Department of Biochemistry, University of Oxford*]. To these plates, a 1 : 1 protein : reservoir drop of 100 nl volume of each solution was added in each first sub-well. If protein yield allowed, a second drop was prepared with a 2 : 1 ratio (133 nl protein to 67 nl reservoir).The Mosquito crystal nanolitre protein crystallisation robot [*TPP Labtech*] was used to stamp out all crystallography plates used in this thesis. Once prepared, the plates were sealed and incubated at 18℃ in a CrystalMation crystal hotel and imager [*Rigaku*]. Both bright-field and UV images were taken at regular intervals and checked for crystal growth by eye. In cases where excess protein was available, crystal plates were also set up in

duplicate and stored at 4℃ for inspection.
Crystals were harvested and flash cooled before delivery to the Diamond Light Source. Crystal hits were harvested using an appropriately sized cryo-loop [*Hampton Research*] and plunge-cooled into liquid nitrogen (LN_2) before mounting in a submerged puck [*MiTeGen*]. For cases where the candidate crystal was not grown in a cryo-protected screen, additional measures had to be taken to ensure the cooling of the crystal without developing ice crystals in and around the sample. For such cases, a cryoprotectant solution was prepared by supplementing a small volume (1 μl) of reservoir solution with glycerol to a final concentration of 25%, and any added ligands to match the original drop concentration. This cryoprotectant solution was added to the protein drop in the sub-well before the crystal was harvested.

If crystal diffraction was not sufficient to obtain a dataset when tested at the synchrotron beamline, the crystallisation condition was often optimised. This involves employing a number of strategies including dilution of the original crystallisation condition from which promising crystals grew, seeding micro-crystals into new drops, or commercial additive screens. The initial reservoir solution was diluted with MilliQ water (100-70%) across a new crystallisation plate and processed as before or with a second dimension in which different dilutions of micro-seeds were added. Micro-seeding was performed by collecting small, poorly diffracting crystals and vortexing in a Seed-Bead eppendorf [*Jena Bioscience*] to generate sub-microscopic micro-seeds. These micro-seeds were serially diluted within the new reservoir diluted plate to generate a unique condition in each subwell. Commercial additive screens such as the Silver Bullet[170] and Additive Screen HT[171, 172] [*Hampton Research*] were used to further investigate potential small molecule additives previously shown to increase stabilisation and solubility to promote crystal growth.

2.3.3 Data Collection

All data collections in this thesis were carried out at the Diamond Light Source synchrotron beamlines, either in person or remotely. For native data collection,

a fixed X-ray wavelength of approximately 0.9 - 1.5 Å (dependent on the exact beamline capabilities) was used. Initially, crystals were moved into the beamline stage and manually centred in the goniometer to ensure complete X-ray beam inclusion during data collection. Three test data collections were performed, with an oscillation of 45°, to inform the user of the probably quality within the crystal. This could be used to detect if strong reflections were present and therefore a full dataset could be collected from the crystal. These test shots often provided information which could influence full data collection, including detector distance, beam size, transmission, exposure time, and oscillation angle. In some cases, a grid scan was performed to identify the region of the crystal which best diffracts. This was performed by separating the crystal into discrete sections, and exposing each region to a single dose of the X-ray beam. Whichever region produced the most intense diffraction pattern was then used as the centre point for which data collection could be carried out.

To record a full data set, the crystal was exposed to the X-ray beam whilst rotating around a single axis in small increments (termed oscillation, typically between 0.1 - 1.5° per frame). Depending on the quality of the initial test images, it may have be possible to estimate a space group for the crystal, which in turn informed the total oscillation required for a full dataset to be collected. Data was collected to obtain at least 180° of frames, for two reasons: firstly to safeguard against inaccurate space group classification from the outset, allowing enough data to be collected to record all possible symmetry operators from the crystal; and secondly to increase the multiplicity of the data recorded. This increases the ratio of measured reflections to unique reflections, increasing the accuracy of the data, so long as there is not too much radiation damage in the process.

Single Anomalous Dispersion (SAD) Data Collection

Single anomalous dispersion (SAD) data was collected at the Diamond Light Source beamline I23. Unlike native data collection, SAD is performed at longer wavelengths (1.1 - 5.9 Å), and lower energy (2.1 - 11 keV). Sulfur SAD phasing is used

to overcome the crystallographic phase problem, using the small anomalous signal from endogenous sulfur in the protein or ligand. The I23 beamline utilises a large semi-cylindrical detector to allow for more data capture at its longer wavelengths, minimising the loss of data as a result of the effect Bragg's law on the crystal. At the longer wavelengths used in SAD experiments, much of the data would be lost with a conventional flat detector. Because of the increased angle of diffraction caused by the longer wavelength ($\lambda = 2d \sin\theta$), a curved detector is used. These data collections are also performed in a vacuum, to minimise the effect of absorption of the long wavelength radiation by the air in the beam path[173].

2.3.4 Data Processing

Data processing in this thesis was carried out using the autoPROC automatic processing program[174], with the added STARANISO package[175] used for analysis of anisotropy. The diffractions patterns collected during an X-ray crystallography experiment inform the user of two sets of information: the intensity and position. The intensity of the spots related to the electron density of the plane from which the X-ray was diffracted, and the position of the spots, hkl, gives information as to the crystal symmetry and lattice dimensions, with the spacing between spots related to the reciprocal vector distance. The first step in data processing is indexing of the spots, this process measures the spacing between the spots to generate a set of vectors in the reciprocal lattice that satisfy the diffraction spacing. Modern indexing programs utilise Fast Fourier Transform (FFT) algorithms to find a constant set of lattice directions. The indexed reflections are iteratively refined to obtain a more accurate lattice description, whilst also correcting for small instrumental errors.

Once indexing has been completed and the crystal and instrumental parameters refined, the spots are integrated. The role of integration is to determine the intensity (I) of each spot, as well as their error (σI). There are two methods for integrating spots, summation integration and profile fit integration, both of which are used to support the other. Spots can be rejected as background noise if they estimations of

their intensities do not match from both approaches. This pipeline is referred to as data reduction, and can be used to cut the resolution of the data if statistical indicators show the data to be too noisy at higher resolution shells. 3D profiling is now commonly used to index spots from modern detectors, using the program XDS[176]. 3D profiling considers each spot over a slice of several images, where the intensity and spot profile can change across the series. This allows for a more accurate consideration of background noise when compared to low-intensity reflections, which are more often caused by higher resolution spots. The background noise is calculated by sequential removal of the highest intensity spots, until those that remain match the expected background values and can then be discarded.

With the data integrated, it can then be scaled and merged. Scaling is a process where the integrated reflections are reviewed as various physical, experimental, or crystal parameters change throughout the data collection. Some of the parameters corrected for are: small variations in beam intensity, anisotropic absorption by the crystal, sensitivity differences throughout the detector, and radiation damage[177]. The observations are scaled relative to one and other to minimise the effects of these parameters on the whole dataset. The program AIMLESS[178] is used to merge the scaled reflections according to their symmetry matches, sorting the reflections into unique observations.

Truncation, also performed by AIMLESS, is the final step of data processing. This is a process by which the spot intensities are converted to structure factor amplitudes as is required for calculation of an electron density map.

There are various data quality indicators generated throughout the data processing pipeline designed to inform the user as to the actual quality of the data collected. These can be used to manually cut-off the highest resolution data that could otherwise could add to the noise of the whole dataset. Traditionally, the $I/\sigma I$ and R_{merge} values were taken as a representation of the signal to noise ratio, and

the redundant measurements of a reflection, respectively. However, R_{merge} values have more recently been found to be misleading in some cases, especially when the data has high multiplicity[179]. As a result R_{meas} was developed as a multiplicity independent value for data evaluation[180]. Now, it is more common to use the Pearson correlation coefficient (CC) of two half datasets ($CC_{1/2}$). The $CC_{1/2}$ measures the correlation between two randomly selected halves of the data set. This is nominally used to cut-off resolutions with a coefficient lower than 0.2–0.5 (20%–50%), as these will not improve the model.

2.3.5 Phasing

The data recorded from an X-ray diffraction experiment contains only the intensities of the scattered X-rays, but lacks the more important phase information that is required to interpret the data into usable models, this phenomena is known as the 'phase problem'. In a unit cell of volume V, the electron density (ρ) at any point (x,y,z) of the model can be calculated from the fourier transformation, utilising the structure factor F_{hkl} where h, k, and l are the indices of reflection hkl. As each structure factor is a wave function, each reflection must have three variables: frequency, amplitude, and phase. The frequency is an inherent property of the X-ray beam and can be calculated from its wavelength; amplitude is proportional to the square root of the measured intensity of the reflection ($I_{hkl}^{\frac{1}{2}}$). Therefore, only the phase of each reflection (α_{hkl}) needs be calculated. The Fourier transform, which connects $\rho(x, y, z)$ to the diffraction pattern is:

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

Molecular Replacement

The most common method of phase determination is molecular replacement (MR), this uses a known homology structure to import phases into the unknown dataset. This approach works best when the reference structure and the experimental data share a high sequence identity (>30%), with a low RMSD between the two molecule's C_α atoms. During MR, the Patterson map of the known structure is rotated and

translated to match that of the experimental data. The Patterson map is generated by a Fourier transform of the intensities of each diffraction spot, without taking into account the phases. An important consideration before MR can be performed is to identify the number of protein copies in the asymmetric unit (ASU), which can be calculated using the Matthew's coefficient. This measures the ratio between ASU volume and protein molecular weight (provided by the amino acid sequence). The number of protein copies is calculated for a range of solvent contents, and the most likely solutions are highlighted based on knowledge of crystal solvent content and oligomeric state of the protein.

For the work in this thesis, Phaser was used for MR[181]. Phaser implements a maximum-likelihood probability to match the best MR solution. This assumes that the best MR solution has the highest probability to reproduce the experimental data, whilst taking into account errors in the search model. Phaser also reports the log-likelihood gain on intensity (LLGI) and translational function Z-score (TFZ) as assessments of its own solution. The LLGI reports the probability of a correct solution, and should increase as the crystal's unit cell is filled with correctly placed protein molecules. The TFZ is a measure of the number of standard deviations above the mean value of a solution, with the expectation that the correct MR solution is separated from the other solutions. A TFZ score greater than 8 represents a valid solution, and is generally used as the criteria for an unambiguously correct solution.

Single anomalous dispersion (SAD) experiments use the anomalous signal from a single dataset (at a single wavelength) to solve the phase problem for unknown crystal structures. The first step in this process is the identification of the heavy atom (HA) substructure, which is accomplished using the program SHELXD[182]. This calculates a Patterson map from the anomalous intensity differences which isolates the anomalously scattering HAs. With the diffraction data, the phases could be solved by MR using the original teneurin structure. This then allowed the position of the intrinsic sulphur atoms within the protein sequence (cysteine and methionine residues) to solve the phases for the S-SAD anomalous Fourier

difference map. As a result of this the Sulphur signal from the ligand could then be investigated at the binding sites.

2.3.6 Model Building and Refinement

Once initial phasing is completed using molecular replacement or experimental phasing, an initial map is generated. This map can be used to place atomic coordinates to represent a model of the data collected. The process of model building and refinement entails improving the positioning of these atoms within the generated model to better fit the data. As each atom has an influence on all other data points within the electron density map, improving the model agreement further improves the maps as a whole. The aim is to improve the agreement between the measured amplitudes in the data and the measured amplitudes in the atomic model.

Once an initial model was built, refinement was performed to improve the atomic positioning with the data. The iterative refinement process involves improving the atomic parameters (x, y, z coordinates, B-factors, and occupancy) and global parameters (global B-factors and bulk solvent scattering) to a point at which they match the observed data. The B-factor of a specific atom is a measure of its displacement from the mean position, which can be used as a measure of the true static or dynamic mobility of the atom. This arises as a result of thermal motion, lattice disorder or partially occupied conformer, or as a result of errors in the model building. These B-factors can be refined either isotropically, where the thermal motion is represented as a sphere of possible motion, or anisotropically, where the representation is ellipsoidal. Usually, refining anisotropic B-factors would require more observations to satisfy the increased parameters[183]. However using translation, libation, and screw motion (TLS) refinement bypasses this requirement. TLS defines groups of atoms with related motion, describing it as translation, libation and screw motion. By default, each chain is treated as a TLS group, which was the approach used in this thesis. Defining more complex TLS groups can be used if B-factors are important in the specific refinement being performed.

In this thesis, refinement was carried out using the program Buster[184] which uses Bayesian statistics for reciprocal space refinement, with Coot[185] used for manual intervention. Initial refinement used both reciprocal space refinement and isotropic atomic displacement factors with secondary structure restraints. In macromolecular crystallography, the data : parameter ratio is poor. This influences the approaches used throughout refinement, as more restraints were used to limit the degrees of freedom. With moderate to low resolution datasets this is especially problematic as there are even less reflections. In the worst cases this can result in overfitting of the data even if geometric restraints are imposed. As a high resolution (2.38 Å) structure exists of teneurin[22] [PDB ID: 6FB3], this was used as a reference structure to further restrain the lower quality data collected and presented in this thesis. Stereochemical restraints were used throughout the refinement process, to ensure the model bond lengths, angles, and torsion angles were kept within physically reasonable ranges appropriate for a protein. In datasets where two or more copies of a protein are present in the ASU, additional restraints can be generated using non-crystallographic symmetry (NCS). Using NCS allows both copies of the protein to be treated as identical, restraining them to remain similar to each other.
When small molecules or ligands are added into the crystallographic model, they must also be refined alongside the protein. This requires these molecules to be restrained stereochemically. Many common ligands (N-linked glycans, substrates, co-factors, etc.) have stored restraints within Coot or other refinement programs. For those without pre-existing restraints, the HIC-Up database[186] was used to obtain a .cif file that contains the restraints from either: the ideal coordinates from the chemical formula, taking into account bond lengths and angles; or the experimental coordinates from a PDB file.

Each refinement cycle was run until convergence, where a new electron density map was calculated and generated. To minimise potential model bias, two Fourier transforms are used to calculate the electron density. The most commonly used are a $2F_{obs} - F_{calc}$ map, and a $F_{obs} - F_{calc}$ map, known as a difference map. This

difference map displays un-assigned electron density and negative density where atoms have been placed incorrectly. Thus allowing for the correct orientations of side chains, main chain, or ligands to be assigned successfully. This process is performed iteratively after each refinement cycle throughout the whole refinement process.

Statistically, the process of model building and refinement aims to minimise deviation between the experimental structure factors ($|F_{obs}|$) and those calculated from the model ($|F_{calc}|$), an assessment of this relationship is represented by the R-factor (R).

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

Before refinement of the first model is started, the reflections are randomly divided into two sets, termed the working and test sets. Each set has an independent R-factor generated for it after each cycle of refinement. However, only the working set is used in the refinement process. The sub-population of reflections in the test set (the smallest of either 5% of the total reflections, or 2000 reflections) is never used in refinement, and is only used to calculate an R_{free} value. This R_{free} is used for cross validation of the model quality. As the model becomes more complete throughout refinement, more parameters are introduced and the R-factors should fall. If refinement is continued after an optimal point has been reached, incorrect parameters can be added which do not improve the model, causing the R_{free} to increase as the R_{work} continues to decrease. This is indicative of an over-fitted model, and is more commonly seen with moderate to low resolution data where molecular details are over-worked outside of the scope of the data quality.

Ligand Refinement

There are a number of approaches to refine ligands placed in X-ray crystal structures. Many of these approaches attempt to balance improving the density at a specific ligand site, without introducing bias into the map and potentially boosting signal erroneously. With flexible ligands such as the those crystallised in this thesis, there are some pitfalls to avoid when placing and refining ligands into a structure. With

low to moderate resolution, many scientists decide to forgo a refined ligand, as it is too difficult to either place the molecule in a certain conformation, or the processing involved in improving the map introduces too much bias. Polder maps, particular omit maps produced in the Phenix suite[187] of programs, can be used to improve the density around a certain area where a ligand is known to be bound. This is achieved by excluding the bulk solvent in the selected area, boosting the signal from weak density that would otherwise be obscured by the bulk solvent. Another consideration is the occupancy of the proposed ligand. In many cases where the ligand does not fully occupy the crystal lattice, there can be a reduce density in the map. At moderate to low resolution, this can be almost indistinguishable from incorrect B-factors. B-factor refinement can be used to solve this ambiguity, especially with flexible ligands. Flexible ligands can also pose a problem in crystal structures. These molecules are often too flexible to be present in a single conformation in the dataset, contributing to blurred maps. As a result of this it can be difficult to place the ligand in a single position, so multiple conformations may be submitted.
In this thesis, some ligands were placable and were stringently assessed to ensure there was no introduced bias during their placement. The ligand was always added at the final stage of refinement, once the maps had been improved as much as possible. All of the ligands restraints used here were generated from already published files, as the sucrose octosulfate (SOS) ligand has a pre-defined ligand code within the Coot library (SCR) as well as each of the repeating monomers forming the longer heparin ligands (SGN, IDS).

2.3.7 Validation

Validation constitutes the final step in structure solution and is used to ensure the model agrees with known stoichiometric parameters. MolProbity[188, 189] is used to report incidents of disagreement within the protein model that has been built. These can include Ramachandran outliers, C^{β} deviations, incorrect rotamers or cis-peptides, protein clashes, and incorrect ligand or glycan geometries. These observations should be minimised throughout the model, unless there is alternative

evidence to show otherwise. Because of the presence of N-glycosylation in the protein structures described in this thesis, Privateer[190–192] was employed and used extensively to further validate the glycan geometry.

2.4 Biophysical Approaches

2.4.1 Heparin Column Purification

Purification using loose heparin chromatography beads was performed as a biophysical method to confirm the glycan binding site of teneurin. Binding site mutants were tested against WT protein to assess heparin binding capacity.

MoBicol classic mini-columns [*MoBiTec*] were assembled by inserting a 35 μm polyethylene filter [*MoBiTec*] and washing with 250 μl MilliQ H_2O. Heparin-agarose beads [*Sigma*] were added to the filtered column and lightly centrifuged (50 x g, 3 min) to form a settled bead layer on the filter. The beads were equilibrated with 10 mM Tris (pH 7.5) and the flow through discarded. The columns were plugged and 100 μl of cell supernatant containing the protein was diluted in 400 μl 10 mM Tris (pH 7.5) in the chromatography column. The column was incubated for 2 hours at RT, on a bench top rotating wheel. The flow through was collected by centrifugation (50 x g, 2 min) before 650 μl wash buffer (20 mM Tris, pH 7.5, 50 mM NaCl) was added and collected by centrifugation. Elution was performed in two stages. Initially, half the final volume of elution buffer (20 mM Tris, pH 7.5, 1 M NaCl, 1x SDS loading buffer) was added and the sample centrifuged such that the heparin beads were saturated in buffer. A second centrifugation step was performed with the remaining elution buffer and the eluted sample was collected in a sterile Eppendorf. Sample protein content was assessed by western blot (described in Section 2.2.7).

2.4.2 Surface Plasmon Resonance (SPR)

Principles of Surface Plasmon Resonance

Surface plasmon resonance (SPR) is an incredibly powerful approach used to investigate the binding kinetics of the interaction between an immobilised ligand, and an analyte. This method has a number of benefits over others, including the

ability to measure label-free protein–protein, or protein–small molecule, interactions in solution in real time[193, 194]. The surface plasmon resonance phenomenon refers to the ability of a metal surface, typically gold, to interact with polarised light. At a precise angle (θ) the free electrons on the surface of the metal absorb the incident light, resulting in a reduced intensity readout (Figure 2.3A). This angle of total internal reflection is dependent on a number of factors including properties of the metal film, the refractive index of the buffer either side of the film, and temperature, among others[195].

Typical biophysical SPR experiments utilise a ligand immobilised to the metal surface and a binding partner analyte in solution (Figure 2.3B). Binding of the analyte to the immobilised ligand changes the refractive index of the surface, and thus the angle (θ) required to achieve plasmon absorption. The change in angle is measured in arbitrary response units (RUs), which can be used to calculate binding kinetics of the interaction.

There are a number of considerations to be made when designing an SPR experiment. Specifically relevant for the work outlined in this thesis is the fact that the change in RUs upon analyte binding is limited by the relative molecular weight of the analyte. The binding of a relatively small sugar ligand in solution (∼10kDa) to a large immobilised protein (∼ 200 kDa) would give a very small change in response units (1/20 of the baseline) as the increase in weight of the complex compared to the unbound protein would be minimal. Because of this consideration, for the work outlined in this thesis, the analyte and ligand were prepared to immobilise the smallest of the two components, such that binding of a larger protein in solution to a small immobilised ligand would result in a large change in θ and RU.

Construct Design and Biotinylation

Protein constructs to be used as SPR immobilised ligands were cloned into a pHLSec$_{\text{Avi}}$ vector [*Addgene plasmid #99845*][159] to facilitate downstream biotinylation. This Avi tag comprises of a biotin ligase (BirA) recognition site (GLN-

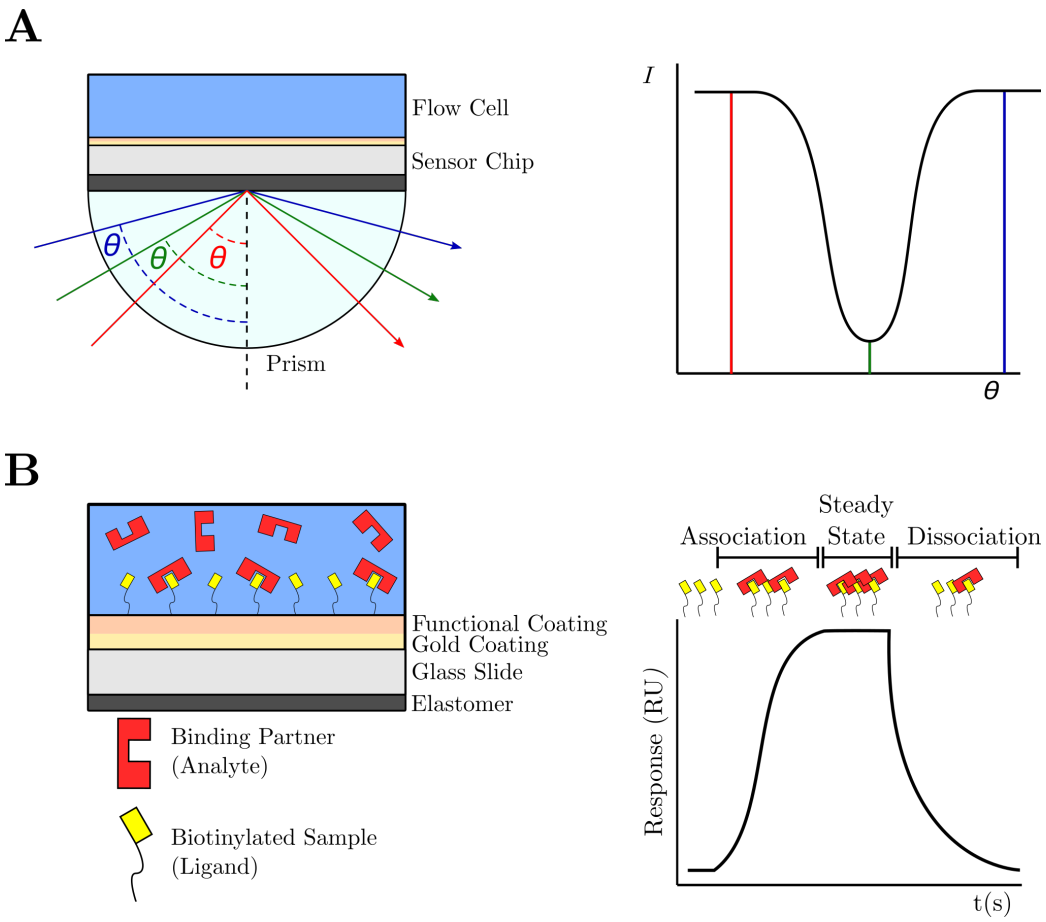


Figure 2.3: Overview of Surface Plasmon Resonance. A. Incident light shone at a surface the correct angle results in the plamson resonance phenomenon where the energy from the photon in absorbed into the surface (*green*). **B.** Binding of an analyte (*red*) to an immobilised ligand (*yellow*) results in a change in refractive index of the surface and thus a change in angle (θ) to produce the resonance. Binding Curves generated by SPR show three distinct stages of binding: association, steady state binding, and dissociation. Statistical analysis of these curves allows for the kinetics of the interaction to be calculated.

DIFEAQKIEWHE). Original constructs (Appendix A Table A.1) and primers (Appendix A Table A.2) were used to clone mouse Latrophilin linkers, as described in Sections 2.1.2 – 2.1.5. An N-terminal Myc tag was introduced into the construct to distinguish these from teneurin for western blotting. Small scale expression tests (Section 2.2.3) and Western Blots (Section 2.2.7) using both α-His and α-Myc antibodies were used to confirm expression of the glycosylated linkers.

For biotinylation of these SPR ligands, a small scale transfection was repeated with the addition of 1 μg BirA vector to the transfection cocktail, as well as

the replacement of PEI with Fugene [*Promega*]. The cell culture medium was supplemented with 100 mM biotin [*Sigma*] prior to addition of the transfection mix, the cells were harvested after 72 hours, as standard. An α-Strep western blot was performed to confirm biotinylation of the protein, after which the remaining harvested supernatant was dialysed in 1x PBS for 48 hours to remove any free biotin. Dialysed samples were frozen and kept at -80℃ until SPR was performed. For direct glycan interaction SPR biotinylated glycan samples were ordered from Iduron [*Manchester*]. Each sample was approximately 30 kDa, and synthesised with end biotinylation, supplied in powder form purified by SEC and dialysed to remove any free biotin. Each biotinylated glycan sample was diluted to 5 mg/ml in 1x PBS, and stored at -20℃. Each glycan was immobilised on the SPR chip (described below) at a working concentration of 100 μg/ml.

Chip Preparation

All SPR experiments were performed on a Biacore T200 SPR machine [*GE Healthcare*], with Bia Evaluation v1.0 software installed. Before a new SPR run was begun, the system was run through with a desorption protocol to wash the system, this was performed with a maintenance chip installed as to not damage assay chips. Activation of the CM5 [*GE Healthcare*] assay chip surface was performed with an application of EDC/NHS solution, resulting in reactive succinimide esters. This reactive surface allowed for streptavidin to be covalently attached to the chip surface via amine coupling. A final injection of ethanolamine deactivated any un-bound reactive sites on the chip surface. For all experiments, approximately 9000 response units (RUs) of streptavidin were immobilised per flow cell.

Biotinylated ligand (protein and glycan) were immobilised using the same protocol. The ligand was injected over the prepared chip at a flow rate of 5 μl/min for 25 s to gauge initial ligand immobilisation. This was repeated with subsequent injections of ligand, if needed, to reach an immobilised ligand density of 300–400 RU. For immobilisation of protein samples onto the SPR chip, the dialysed protein was

diluted 1/10, and with reference to the α-Strep western blot, to ensure a correct density of sample was immobilised.

SPR Injections and Analysis

Once the ligand had been successfully immobilised, test injections were performed with varying high concentrations of the protein analyte. This ensured an acceptable maximal binding response was seen in the system, from which serial dilutions could be made to diminish binding and calculate accurate binding kinetics. For the SPR experiments performed in this thesis, a high concentration of 5 μM for all protein analytes was used. This was injected over each chip flow cell, and binding (represented by an increase in RU compared to the blank flow cell) was recorded. After each test injection, a short dissociation step was performed to ensure that the system recovered to the baseline before each subsequent protein analyte injection. Dissociation buffer (2 M $MgCl_2$) was flown through the system at 5 μl/min for 25 s to remove all bound analyte.

After the high concentration test injections were performed, 10 serial dilutions (1 : 2 protein : buffer) were made of each protein analyte in appropriate running buffer (1x PBS, 0.005% Tween). The lowest protein concentration dilution of each analyte was again test injected to ensure a zero binding response in each case.

With a full dilution series of each analyte, and test injections for the high and low concentrations confirming maximal and zero binding, respectively, a full binding run was performed. Each protein analyte was injected over each flow cell at 5 μl/min, starting from the lowest to highest concentration. After each 250 s sample injection, a 25 s dissociation step was performed.

Analysis of the SPR data was performed with the Bia Evaluation v1.0 software. Dose response curves were analysed by Langmuir 1 : 1 binding kinetics, assuming a 1 : 1 protein–ligand stoichiometry to calculate the maximal response (B_{max}) and dissociation constant (K_D). Steady state analysis was not used for data processing

as a number of the SPR curves indicated the interaction had not equilibrated and reached a steady state.

2.4.3 Streptavidin Bead Pulldown

Binding specificity between teneurin and separate glycan species investigated in this thesis was assessed via streptavidin bead pulldown. 10 μl blank Streptavidin Agarose Ultra Performance beads [*TebuBio*] were washed with running buffer (20 mM Tris, pH 7.5; 150 mM NaCl) by centrifugation (700 rpm, 2 min) and incubated with the 35 kDa biotinylated glycans, used for SPR above, at room temperature for 1 hour per experiment. Following glycan immobilisation, the sample was again washed with buffer to remove excess unbound glycan. Purified protein of interest (either wild type *Gg*Ten2$^{\Delta EGF}$ or mutant protein) was then incubated with the primed beads in a solution of buffer containing 1% BSA for 2 hours at room temperature. The flow through was collected by centrifugation and the sample was washed with buffer without BSA. Elution of the streptavidin beads was performed with 1x SDS loading buffer to remove the bound protein from the beads.
Samples were either processed by SDS-PAGE and InstantBlue™ Coomasie staining, or western blot, to reveal the protein content of each fraction.

2.4.4 Heparan Sulfate Multimerisation

Teneurin was incubated with a long heparan sulfate chain (100 monosaccharide units) to assess the ability of the protein to form multimers along a glycan scaffold. The full extracellular domain of teneurin 2 (Appendix A Table A.1) was expressed and purified as above (Section 2.2.5). Full length teneurin ecto-domain (*Gg*Ten2^{Ecto}) was chosen due to its ability to form cis dimers in solution. The N-terminal EGF repeats encoded within this construct form cis dimers, as is found when teneurin is expressed on the cell surface in vivo. This dimerisation increases the avidity of the heparan binding as there are two potential binding sites for the heparan chain.
Multimerisation was assessed by size exclusion chromatography (SEC). Either protein alone, heparan sulfate alone, or a mixture of both protein and glycan were

incubated at room temperature for 1 hour prior to be loaded on a Superose® 6 Increase 10/300 GL column [*GE Healthcare*]. Due to the expected low affinity of binding, 5x molar excess of heparan sulfate was first complexed with teneurin. Subsequent experiments utilised 25x molar excess of the glycan chain. Elution fractions were assessed for protein content by α-His western blot, under either reducing or non-reducing conditions (Section 2.2.7).

2.5 Cell Biology

2.5.1 Cell Surface Staining

Membrane expression of full length fluorescently labelled protein constructs was confirmed after transient transfection (Section 2.2.3) by cell surface staining. 6-well plates were prepared and seeded as described previously with the addition of two autoclaved coverslips at the base of each well. Transfections were performed as described previously and incubated for 24 hr. A metal block was pre-emptively cooled on ice and covered with a piece of wet parafilm on which each seeded coverslip was placed. 50 μl ice cold buffer (1% BSA, 20 mM HEPES in HBSS, pH 7.5) was added to each coverslip and incubated on ice for 30 min in the dark. The buffer was replaced with 50 μl primary antibody, recognising the expression tag on the protein of interest (Ms α-HA), and incubated for 60 min. Each coverslip was washed once with ice cold PBS before immobilising with 250 μl 4% PFA for 20 min. Immobilised cells were washed once with ice cold PBS supplemented with 5 mM NH_4Cl and then once more with PBS alone. 50 μl Cy3 labelled secondary antibody (α-Ms Cy3) was added and incubated for 60 min at RT. Excess antibody was removed and each coverslip washed twice with PBS. 30 μl 10 μg/ml DAPI was added to the cells and incubated for 5 min at RT before a final wash with PBS. Each coverslip was then rinsed in MilliQ H_2O before being mounted onto a Superfrost® [*Fisher Scientific*] using Epredia™ Immu-Mount™ [*Thermo Fisher*]. Imaging was performed on a Nikon® Eclipse TE2000U epifluorescent microscope with an exposure time of 486 ms. Intensities of filtered light were adjusted per run to depending on expression and to improve visualisation.

2.5.2 Glycan Dot Assay

To assess the effects of glycan binding in a cell based environment, a migration assay was developed utilising the same 30 kDa biotinylated glycans samples used in SPR (Section 2.4.2) and pulldown assays (Section 2.4.3). Streptavidin was pre mixed with a 5x molar excess of glycan in PBS buffer containing Goat α-human Alexa Fluor 568 for visualisation. 5 μl of this 'Glycan dot' was dispensed onto each well of a Nunclon™ Delta 24 well assay plate [*Thermo Fisher*] and incubated in the dark for 30 min to allow for immobilisation onto the treated surface. Excess glycan solution was removed, and the assay area was washed with PBS and left overnight at 4℃ in PBS to avoid drying the spot. HEK 293T or N2A cells were prepared and transfected with either full length WT teneurin, R2092E-R2094E mutant, or FLRT2 (Appendix A Table A.2) as described in Section 2.2.3 and incubated for 24 hr.

After incubation, each cell condition was harvested and counted to ensure consistent cell numbers per assay repeat. Dead cells were removed from the transfected plate by vacuum, and each well was washed with 2 ml PBS before being vacuumed. The attached cells were dissociated by incubation with 5 mM EDTA for 5 min at 37℃ before being washed with 2 ml 2% medium and pelleted by centrifugation (200 x g, 5 min). The supernatant was carefully removed and the pelleted cells were resuspended in 0.5 ml 2% medium. This concentrated cell sample was counted using 20 μl cells and 20 μl Trypan blue in a Countess™ cell counting chamber slide [*Thermo Fisher*]. Dead cells were distinguished from live cells by their permeability of the stain. Counting was performed using a Countess™ II automated cell counter [*Invitrogen*], in duplicate, and the cells were diluted to 350,000 live cells/ml. 250 μl of cells were added to the immobilised glycan assay plate and incubated for the required time before imaging.

To image the assay, the cells were fixed and stained to improve visualisation. The assay media was removed from each well and replaced with 250 μl 4% PFA for 15 min to fix the cells. The PFA was then removed and each well was washed initially with PBS supplemented with 5 mM NH_4Cl, then again with PBS alone. The cells were

then stained with 200 μl 1 μg/ml DAPI to stain the nucleus before being washed again with PBS. Finally the cells were mounted using Epredia™ Immu-Mount™ [*Thermo Fisher*] and covered with a glass coverslip. Imaging was performed with a Nikon® Eclipse TE2000U epifluorescent microscope, and analysis was performed in Image J. Graphical data and ANOVA statistical analysis was performed in Prism.

3

The Structures of Teneurin–Glycan Complexes

3.1 Summary

This chapter encompasses the initial work carried out during my DPhil, investigating the interaction between the extracellular domain of teneurin, and glycosaminoglycans (GAGs) of physiological relevance. I initially set up crystallisation trials with a small inexpensive glycan mimic, sucrose octasulfate (SOS), which were collected at the Diamond Light Source to ~3.3 Å resolution. This revealed the binding site on the surface of the teneurin YD-shell, where heparin had been proposed to bind in a previous publication[40]. Molecular replacement utilising the previously solved teneurin crystal structure by Dr Verity Jackson[22, 158] was used to solve the phases. Refinement was carried out with autoBuster, prioritising the amino acid geometry and reference structure over my lower quality data maps. There were noticeable changes to the N-linked glycans, which were adjusted in Coot and validated with the Privateer software in CCP4[190–192]. Unfortunately, due to the limits of moderate resolution data and the inherent flexibility of the SOS ligand, a definitive conformation could not be confirmed and as such the SOS ligand was placed only at the end of refinement, and was not used for the refinement calculations.

The crystal data and ligand placement was corroborated with a number of separate approaches. Polder maps were used to boost the flattened signal at the protein surface to increase the density for the ligand. S-SAD data was collected at the I23 beamline to attempt to better place the Sulfur groups within the SOS ligand. However due to the inherent flexibility for the ligand this was not possible.

Following the SOS dataset, the physiologically relevant glycans heparan sulfate (HS) and chondroitin sulfate (CS) were co-crystallised. Data collections were performed in much the same way as with the original SOS structure, and were processed with the same pipeline. These structures revealed similar binding for the HS ligand, which contact the same 6 charged amino acids on the YD-shell. However, none of the CS containing crystals contained unassigned density at this location or any other, indicating no glycan binding in the crystal.
Glycan binding could also be authenticated using the ClusPro binding server, which can predict heparin binding sites computationally. This highlighted the same binding region as was discovered in the crystal structure, as well as one more on a loop in the NHL domain which was not present in the experimental data.

3.2 The Teneurin–SOS Complex

3.2.1 Crystallisation of the Teneurin–SOS Complex

The teneurin construct used throughout this chapter was provided by the lab, having been previously cloned by Dr Verity Jackson[158] (Appendix A Table A.1), and is the same construct published as the first teneurin structure[22]. The boundaries of this construct start immediately before the TTR (transthyretin-like) domain, and continue to the C-terminus of teneurin, where a hexahistidine tag has been added for purification purposes (Figure 3.1A). Protein was produced for crystallography as described in the methods above (Section 2.2). The purest SEC fractions (Figure 3.1B, C, fractions D2-D4) were pooled and the protein concentrated to 5.72 mg/ml in buffer containing 10 mM Tris, pH 7.5, 150 mM NaCl, 100 mM NDSB-256. To this, the glycan ligand sucrose octasulfate (SOS) [*Toronto Research Chemicals*]

was added to make a final concentration of 3 mM. The initial screens set up to be incubated at 18℃ were Morpheus, Morpheus II, JCSG+, PACT, PGA, MIDAS, and Index. A single Morpheus plate incubated at 4℃ was also prepared. All screens were performed using 1 : 1 protein : reservoir drop ratios.

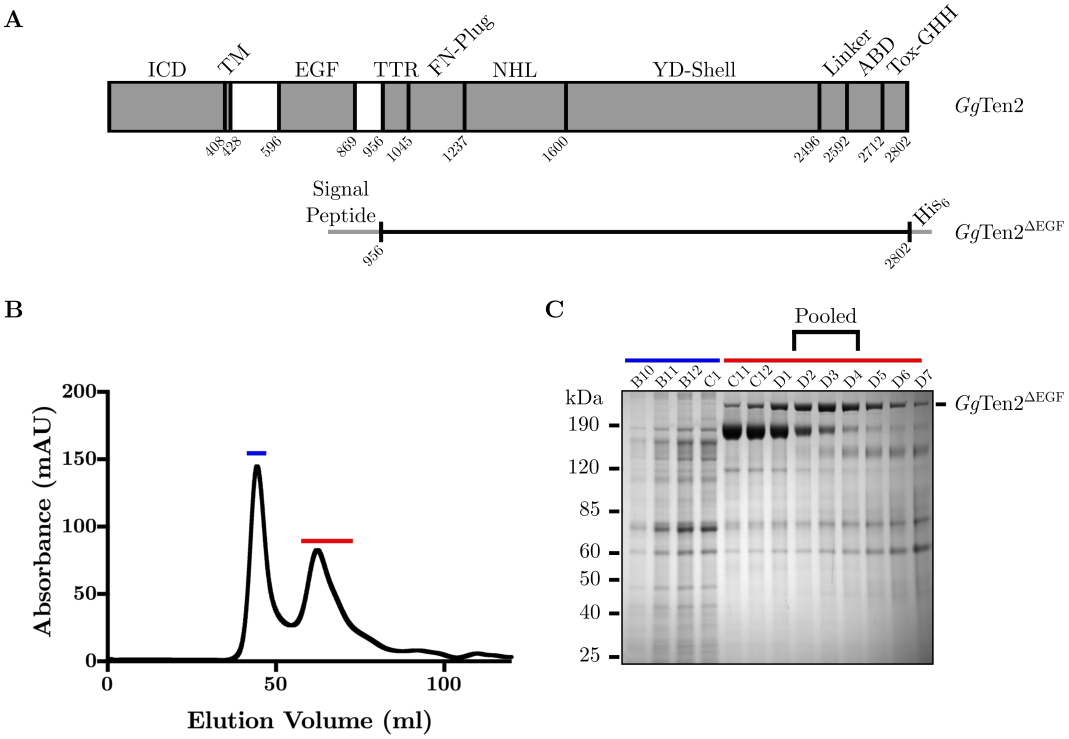


Figure 3.1: Design and Purification of *Gg*Ten2$^{\Delta EGF}$ for SOS co-crystallisation. **A.** Domain organisation of the whole *Gallus gallus* teneurin 2 protein. *ICD*, intracellular domain; *TM*, transmembrane; *EGF*, epidermal growth factor-like repeats; *TTR*, transthyretin-related; *FN plug*, fibronectin-plug; *NHL*, NCL, HT2A and Lin-41 like domain; *YD-shell*, tyrosine-aspartate repeat shell; *ABD*, antibiotic-binding domain-like; *Tox GHH* toxin like domain. The construct used for purification was cut downstream of the EGF repeats and contains an N-terminal signal peptide from the pHLSec vector, and a C-terminal hexahistidine tag. **B.** A SEC chromatogram profile for *Gg*Ten2$^{\Delta EGF}$. The first peak is the void volume and the protein elutes alongside a HEK293 cell contaminant. The blue and red lines indicates fractions analysed in C. **C.** Reducing SDS-PAGE gel of the peak fractions. *Gg*Ten2$^{\Delta EGF}$ is the upper band, and the contaminant is below in earlier fractions. The purest fractions were pooled and concentrated for crystallography.

Initial crystals grew across multiple screens, generally after 7-14 days incubation although this depended on the condition. Notably, a single crystal grew in the Morpheus II screen at 18℃ in a condition containing 40 mM Spermine tetrahy-

drochloride, Spermidine trihydrochloride, Diaminobutane dihydrochloride, Ornithine monohydrochloride; 0.1 M MOPSO, Bis-Tris, pH 7.5; 10% ($^{w}/_{v}$) PEG 8000, 20% ($^{w}/_{v}$) 1,5- Pentanediol (Figure 3.2B). Multiple crystals were harvested and cryo-cooled after 44 days incubation, with cryoprotectant solution containing the mother liquor supplemented with 25% glycerol, if required. Crystals were stored in LN_2 until beam time at the Diamond Light Source was available.

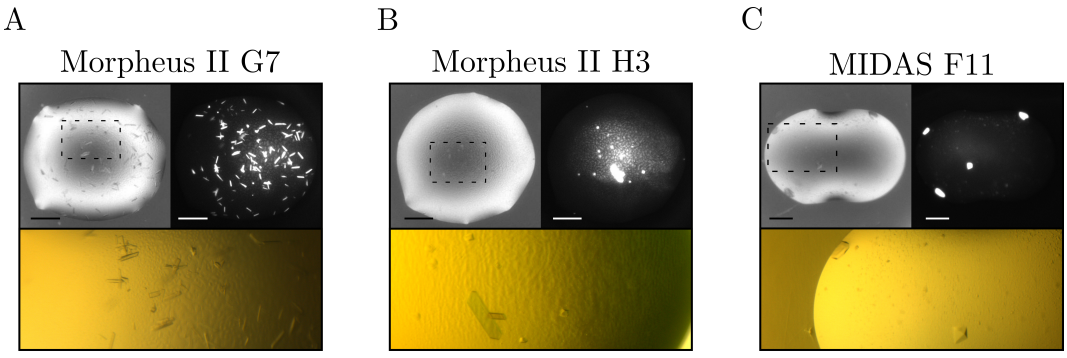


Figure 3.2: Crystal hits of the *Gg*Ten2$^{\Delta EGF}$-SOS co-crystallisation trials. Bright-field and UV images of each drop are shown above, with zoomed bright-field images of specific crystals (*black box*) below. All conditions were incubated at 18℃. Scale bars = 250 μm. **A.** Crystals grown in the Morpheus II G7 condition did not diffract. **B.** The single crystal grown in the Morpheus II H3 drop diffracted to ∼3.3 Å and was used for structure solution. **C.** The MIDAS F11 condition grew larger tetragonal crystals which only diffracted to ∼8.3 Å.

3.2.2 Data Collection and Processing

*Gg*Ten2$^{\Delta EGF}$–SOS crystals were screened at the Diamond Light Source synchrotron, at the I04 beamline. Data collection parameters are summarised in Table 3.1. Many of the parameters were determined from previously shot crystals during the visit. The crystal grown in the Morpheus II H3 condition described above produced the highest resolution diffraction (3.27 Å), and was later found to contain apparent density for the SOS ligand. The crystal and a representative diffraction image are shown in Figure 3.3. The only other diffracting crystals which allowed for automatic processing on ISpyB were also assessed but not further processed due to the poor

resolution, or the lack of ligand density in the map data.

Table 3.1: Parameters for the *Gg*Ten2$^{\Delta EGF}$–SOS data collections.

Beamline	i04
Wavelength (Å)	0.9795
Beam size (μm)	63x50
Oscillation (°)	0.1
Exposure (s)	0.1
Transmission (%)	18.74
Number of images	1800

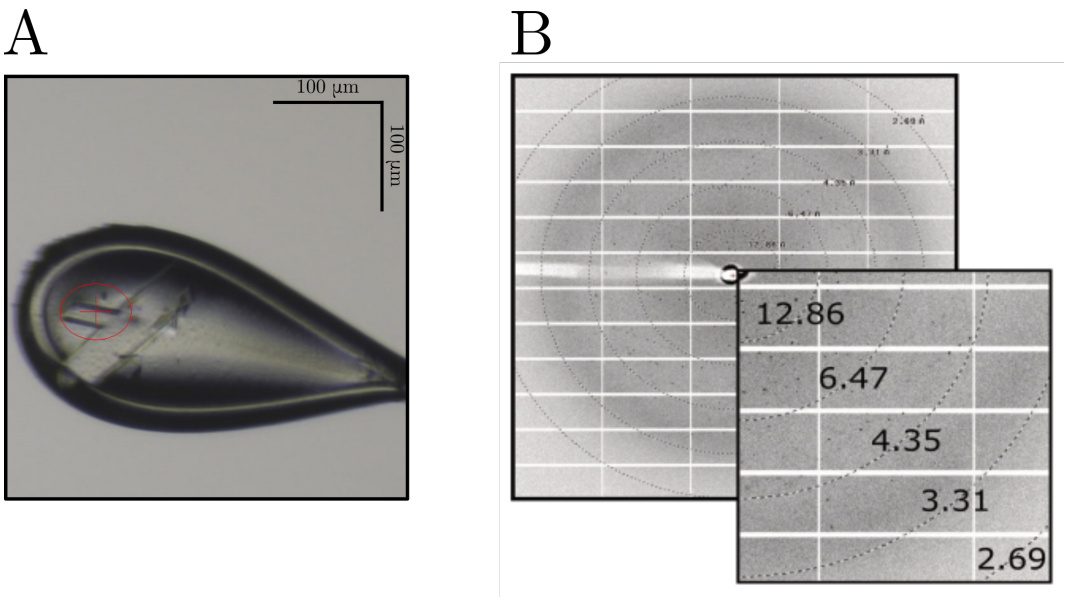


Figure 3.3: Harvested crystal and example diffraction image of the Morpheus II H3 sample. A. The harvested crystal in the loop, showing the beam size (*red circle*). **B.** A representative diffraction image of the data collection, the raw image is shown (*left*) alongside the thresholded image (*right*) with a zoomed section to show the spots more clearly. Resolution rings (Å) are shown for reference.

The collected data was automatically processed using the AutoPROC server[174]. Data was merged with Pointless and Aimless. The Staraniso plugin[175] was used to correct the slight anisotropy present in the sample. Due to the low transmission of the data collection, radiation damage was insignificant and was not a limiting factor when assessing the data quality. The main factor for resolution cut-off was

$CC_{1/2}$. This measures the correlation between two randomly selected halves of the data set, as is independent of bias. A $CC_{1/2}$ of between 30-50% was used as the high resolution shell cut-off for this data, the highest shell had a value of 30.3%. Prior to molecular replacement, the Matthew's coefficient was calculated to predict the number of molecules in the asymmetric unit, using the teneurin sequence. This predicted a pair of teneurin molecules per ASU, with a solvent content of 64.5%. The processed data was run through Phaser molecular replacement using the original *Gg*Ten2$^{\Delta EGF}$ crystal structure solved in 2018[158]. This model allowed for the phases of my glycan complex dataset to be solved with ease, and in this case produced a correct solution with two copies per ASU which could be confirmed by visualising the maps and MR model with Coot (Figure 3.4). The β-strands making up the teneurin YD-shell could be easily identified within the density and the placed MR output model fit well within these densities, as well as throughout the rest of the map. The Phaser program also outputs quality indicators in the form of a translational function Z-score (TFZ) which should be higher than 8 to indicate a correct solution, as well as a log-likelihood gain on intensity (LLG) which should be as high as possible. The single unique solution found by Phaser gave values of 91.98 and 16255.33, respectively. There was therefore no ambiguity in the calculated MR solution, which could be confirmed visually when inspecting the maps.

An initial refinement run with default parameters was performed with Refmac in the CCP4 suite to remove solvent noise after molecular replacement, and more easily identify unassigned areas of density which could be attributed to the SOS ligand bound to the protein surface. In the specific dataset described so far, this revealed a large unassigned density contacting the YD-shell of teneurin, around highly charged Arginine (R) and Lysine (K) residues. This putative binding site contacts 6 residues: K2078, R2092, R2094, K2102, R2106, and R2116 (Figure 3.5).

Refinement was carried out using the autoBuster program, making use of the high resolution teneurin ecto-domain structure as a reference restraint (PDB ID: 6FB3).

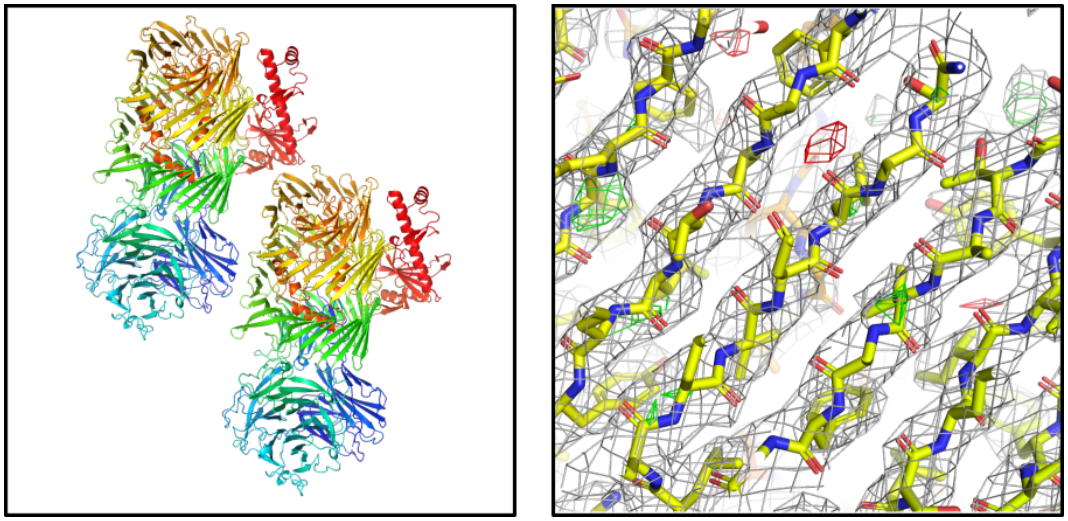


Figure 3.4: Molecular replacement results for the *Gg*Ten2$^{\Delta EGF}$–SOS data. Two copies of teneurin were proposed to be present in the crystal asymmetric unit. Successful molecular replacement could be seen by the distinct β-sheets visible from the teneurin YD-shell. The experimental 2Fo-Fc map is coloured *grey* and contoured to 1.0σ. The difference Fo-Fc map is contoured to $\pm$ 3.0σ coloured *red* and *green*.

This improved the geometry of the teneurin-glycan complex map by imposing local structure similarity restraints (LSSR) throughout the refinement process, limiting the divergence from the higher quality original data and reducing the influence of my lower quality results. This restraint allowed the glycan containing structure to be easily refined as close as possible to the original, and left just obvious changes in certain flexible areas of the protein component, as well as differences in N-glycosylation to be refined. Throughout refinement, the maps covering the SOS binding site were iteratively improved as the rest of the model was better fit to the data. Initially, rigid body refinement was implemented, constraining the main domains within the *Gg*Ten2$^{\Delta EGF}$ (TTR, FN-Plug, NHL, YD-shell–Tox-GHH) individually. Throughout, translation-libration-screw-rotation (TLS) refinement was also employed, with each chain modelled as a single rigid group. Initially an R/R$_{free}$ of 0.2781/0.2820 was generated, this was reduced to 0.2431/0.2736 over the course of refinement of N-linked glycans and obvious protein conformational errors. These values are able to be so low due to the use of the high resolution reference structure throughout refinement, among other factors. N-linked glycans

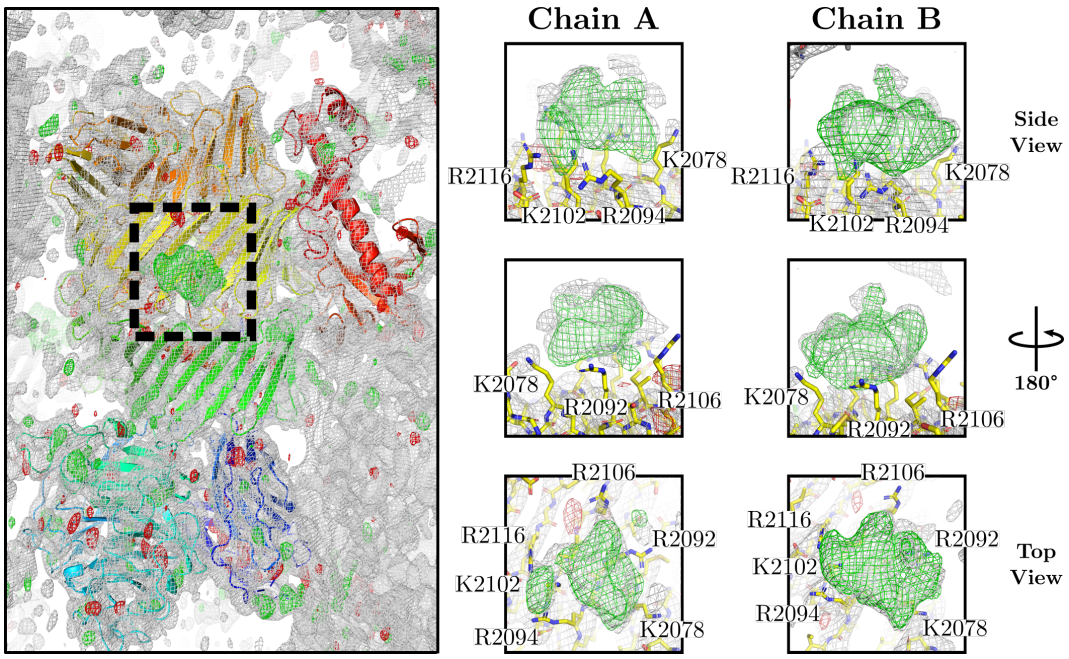


Figure 3.5: Initial density for the sucrose octasulfate (SOS) ligand. A large unassgined density can be seen on the surface of the teneurin YD-shell (*dashed box*). Un-assigned density was present in both chain A and chain B. The interacting charged residues (K2078, R2092, R2094, K2102, R2106, and R2116) were identified. The apparent SOS density is larger than a single SOS molecule, indicating flexibility or multiple conformations for the glycan. 2Fo-Fc map is coloured *grey* and contoured to 1.0σ. Fo-Fc difference map is contoured to $\pm$ 3.0σ coloured *red* and *green*.

were adjusted, added or trimmed, depending on the best fit at each step. When adding to existing, or building new, glycan chains the Coot N-linked glycan builder plugin was utilised, ensuring the correct glycan structure was present. Subsequent refinements required B-factor alterations to the glycans. Initially the B-factor for many N-linked glycans was too low, and was therefore producing poor maps with areas of incorrect density. To combat this the B-factors of the sugar molecules were set to 150, slightly above that of the average protein B-factor, and a round of refinement with Buster was allowed to refine these values to fit the data.

The resulting maps were noisy but clear, with obvious aligning density for the main protein, and a more sculpted unassigned density at the putative binding site described above. At the binding site, the charged side chains were removed from the structure and a round of refinement was run to attempt to assign their

orientation. This worked on the assumption that unassigned density would appear where the trimmed side chains were actually position, however this did not resolved the location of the flexible side chains well, and as such the conformations were assigned as best as possible into the density, whilst favouring good geometry, without a definitive conformation (due to both data quality and the inherent flexibility of the amino acid sidechains). Due to the moderate to low resolution of the data, and the inherent flexibility in the SOS ligand, it was decided to not place the sugar in the density during refinement. It was clear that a single conformation of the sugar could not be placed, and due to high B-factors caused by the resolution, coupled with the likelihood of multiple binding conformations, the sugar was illustratively placed as a final step after validation, without biasing the refinement with an exact conformation (Figure 3.6). Occupancy refinement was not performed, as at this resolution it would be impossible to distinguish from B-factor variations.

Validation of this structure followed the standard methods for low-moderate resolution data. Because the refinement process was completed with constant reference to the high resolution teneurin structure (PDB ID: 6FB3), minimal adjustments were made to the protein model itself, resulting in minimal validation of both Ramachandran and rotamer outliers. Where changes were made, the outliers were restored back to that of the reference structure. Due to the relatively high amount of N-linked glycan adjustments made during refinement, validation of the adjusted and added glycans was paramount. This was achieved using the Privateer software within the CCP4i2 suite[190–192]. The glycans built using the semi-automated module within Coot were found to often contain incorrect stereochemistry and conformation. These were easily adjusted, to one; ensure the correct anomer was present (adjusting the α or β linkages between monomers), and two; ensuring the glycan ring was in the lowest energy chair conformation ($^4C_1$) (Figure 3.7). The final Privateer output for this dataset is shown in Appendix B, Table B.1.

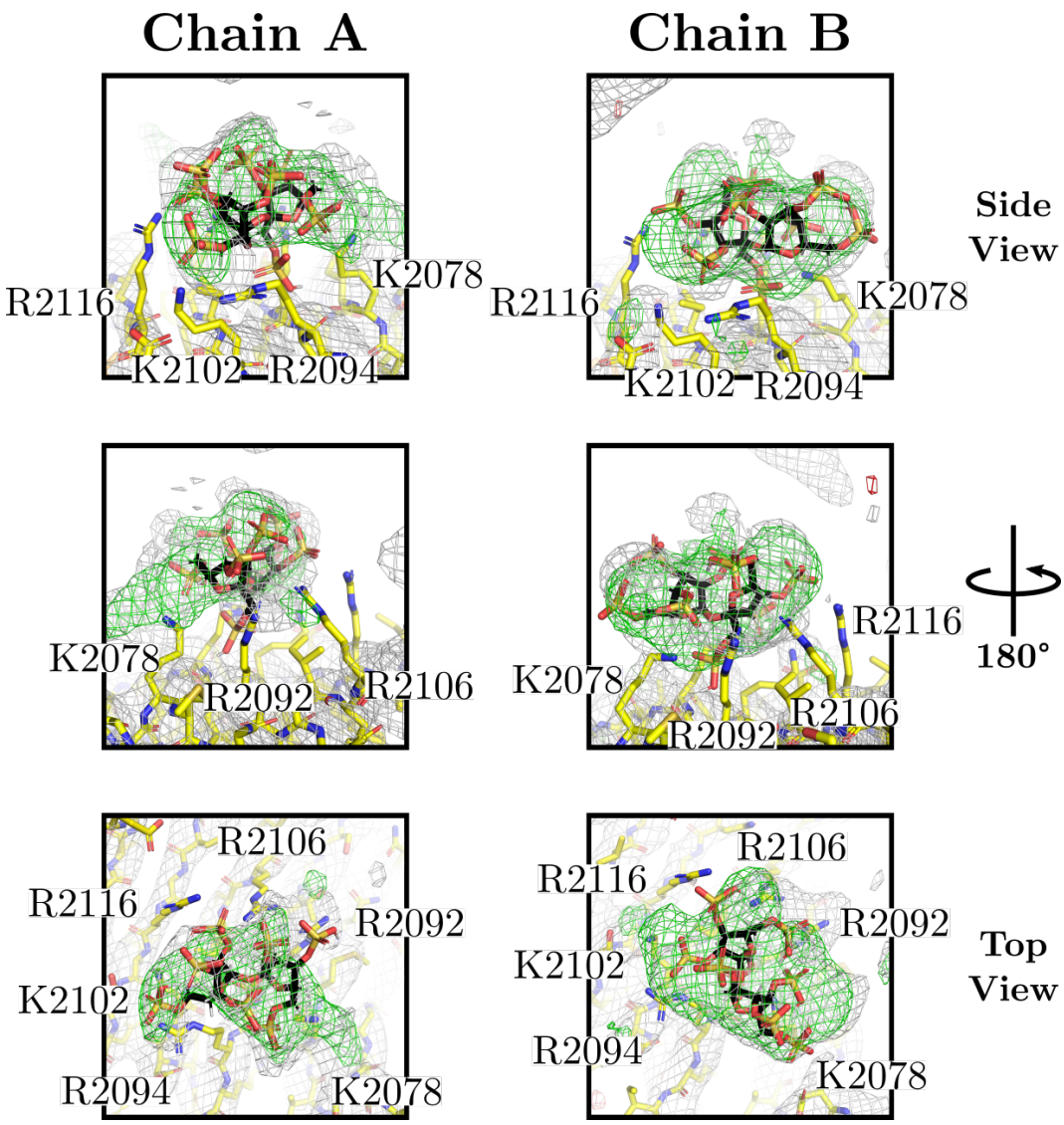


Figure 3.6: Sucrose octasulfate (SOS) placed in its unassigned density. The flexible SOS ligand was placed in the density after refinement and validation of the structure was completed, as to not bias the refinement. The final 2Fo-Fc map generated after refinement is coloured *grey* and contoured to 1.0σ. The final Fo-Fc difference map is contoured to $\pm$ 3.0σ coloured *red* and *green*.

A final check of the model quality for this dataset was performed in Molprobity, a program within the Phenix suite which performs an all-atom validation. This assesses the main statistics of the dataset (R_{work}, R_{free}, RMSD (Bonds), RMSD (Angles), Clashscore, and B-factors). A single Molprobity score is generated for the structure, which takes into account a combination of the assesses data quality indicators. A score lower than the data resolution indicates that the model is of

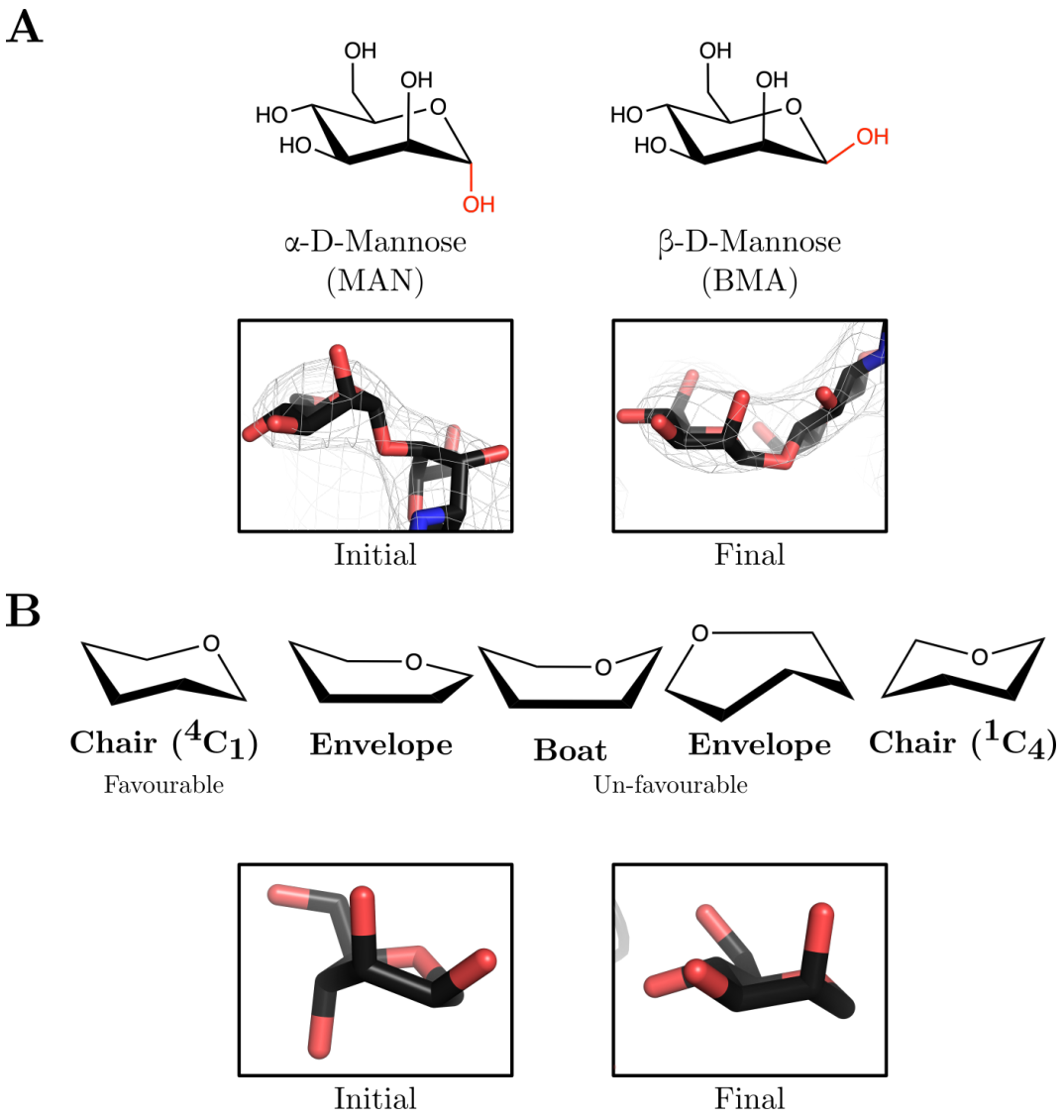


Figure 3.7: Validation of N-linked glycans with Privateer. A. Glycan anomers were adjusted to conform to the dictionary linkages. **B.** The lowest energy ring conformation ($^4C_1$ chair) was used for each pyranose sugar. 2Fo-Fc map is coloured *grey* and contoured to 1.0σ.

higher quality than the average structure at that resolution[188]. The Molprobity score for this dataset, which diffracted to ∼3.27 Å, was 2.08. The overall statistics for this dataset are represented in Table 3.2.

Table 3.2: Data collection, model refinement, and geometry statistics for the *Gg*Ten2$^{\Delta EGF}$–SOS structure. Values for the high resolution shell are shown in parentheses.

Protein Construct	*Gg*Ten2$^{\Delta EGF}$
Ligand	Sucrose octasulfate (SOS)
Condition	Morpheus II H3
Resolution Range (Å)	
Low Resolution	153.80
High Resolution	3.274 (3.391–3.274)
Space Group	P 1 2_1 1
Cell Dimensions	
a, *b*, *c* (Å)	140.58, 133.82, 158.37
α, β, γ (°)	90.00, 103.79, 90.00
R_{meas} (%)	20.44 (82.95)
Observations	94239 (4262)
Unique Observations	27335 (1367)
$I/\sigma I$	3.56 (1.16)
$CC_{1/2}$ (%)	96.6 (30.3)
Completness (%)	97.7 (1.3)
Multiplicity	3.4
Refinement	
No. Reflections	27306
R_{Work}	0.2431
R_{Free}	0.2736
No. Atoms	
Protein	29026
Ligand/ Ion	778
R.M.S.D	
Bond Lengths (Å)	0.011
Bond Angles (°)	1.57
Validation	
MolProbity Score	2.08
Rotamer Outliers (%)	1.31
Clashscore	14.45
Ramachandran Plot	
Favoured (%)	95.36
Allowed (%)	4.20
Outliers (%)	0.44
Average B Factors (Å^2)	
Protein	117.83
Ligand	134.54

3.2.3 The Teneurin–SOS Binding Site

Another method that can be used to clear up density for known ligand locations are Polder maps. These are specific omit maps which excludes bulk solvent around a specific region, and as such can boost weak signal for a ligand, protruding side chain, or flexible protein region. In this case an omit map could be used to increase the signal at the SOS ligand binding site which may be being obscured by the ligand interface surrounding the teneurin fold. There are considerations to be made when using a Polder map however, as the omitted area is calculated as a buffer around a specified number of atoms in a specific region. As such it can be incredibly easy to add bias into the calculations by specifying an exact ligand and conformation, which is then boosted by omitting surrounding bulk solvent, only increasing its potentially incorrect signal. Due to these considerations, a polder map should only be use where a ligand identity can be certain, and other approaches should be used when there is more uncertainty over the location or identity of unassigned density. In the case of structure in this thesis there has been considerable work outside of crystallographic processing, in the subsequent chapters, to identify the YD-shell binding density as that of the SOS ligand. Due to the unambiguousness of the ligand and its binding location, a polder map was generated around its binding site (Figure 3.8). Unfortunately this map does not increase the fidelity of the SOS density, instead showing a large broad area of density compared to the standard omit map. This is again indicative of a flexible ligand, which is able to bind in more conformations than previously presented in the original omit map. Because of this increased density area, the polder map does not help to place the ligand in this case, instead bolstering the hypothesis that this incredibly flexible ligand is bound, but in multiple confirmations contacting multiple amino acid side-chains on the teneurin YD-shell surface.

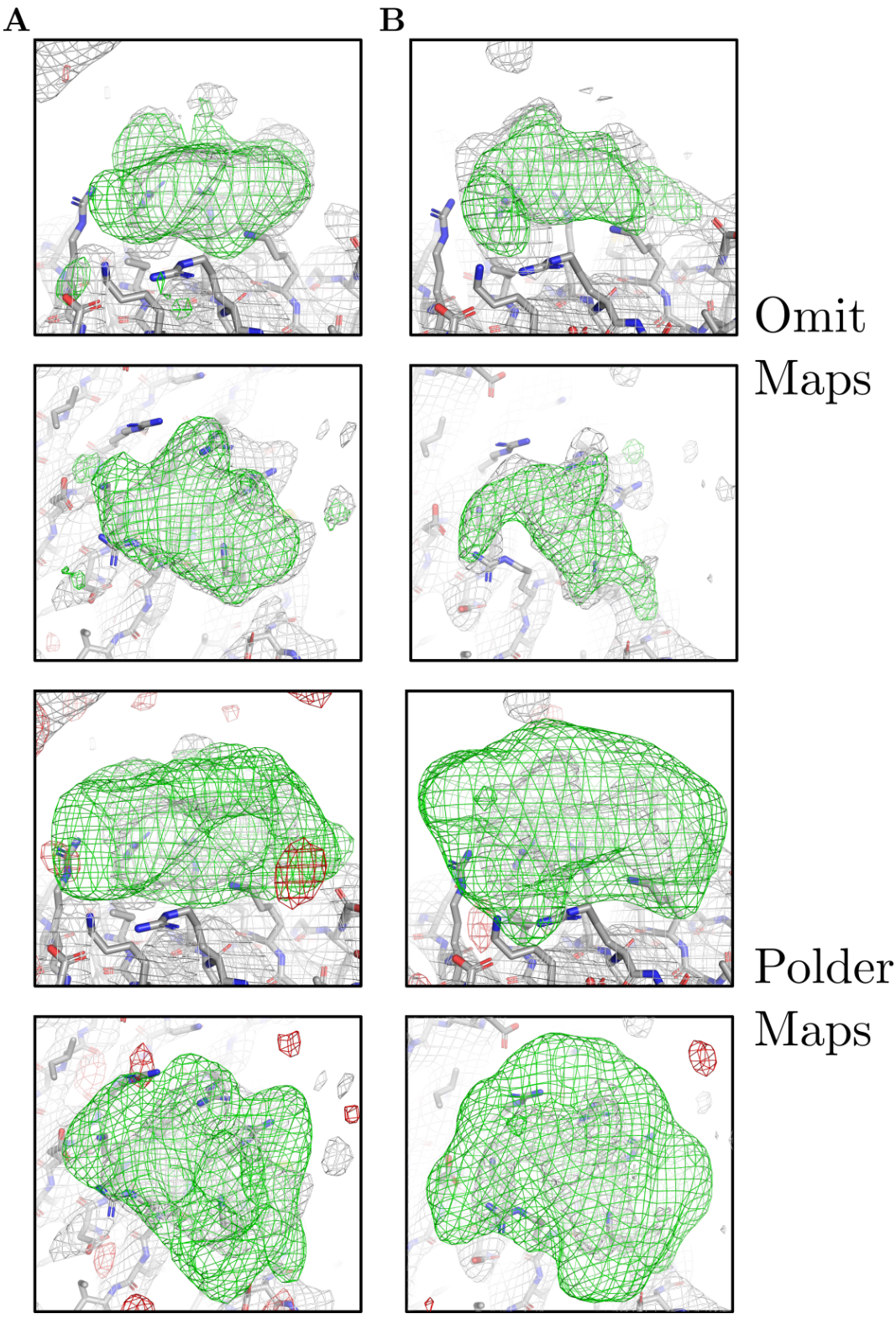


Figure 3.8: A Polder map of the SOS binding site compared to a standard omit map. **A.** Chain A of the *Gg*Ten2$^{\Delta EGF}$–SOS structure. **B.** Chain B of the *Gg*Ten2$^{\Delta EGF}$–SOS structure. The experimental 2Fo-Fc map is coloured *grey* and contoured to 1.0σ. The difference Fo-Fc map is contoured to $\pm$ 3.0σ coloured *red* and *green*. 2Fo-Fc map is coloured *grey* and contoured to 1.0σ for both representations (generated after final refinement in Buster). The difference Fo-Fc map is contoured to $\pm$ 3.0σ coloured *red* and *green* (*above*). The polder map generated from the Phenix suite (*below*) is also contoured to $\pm$ 3.0σ.

Single-Wavelength Anomalous Diffraction Data

In an attempt to help improve placement of the SOS ligand, single-wavelength anomalous diffraction (SAD) data was collected at Diamond Light Source, beamline I23. This approach differs to conventional X-ray diffraction experiments as it utilises the specific wavelength at which a specific element has a strong anomalous diffraction. Sulfur atoms, present in the amino acids Methionine and Cysteine, can be targeted using sulfur SAD (S-SAD) at a wavelength between 1.1 - 5.9 Å. At the I23 beamline, an optimum resolution of 2.75 Å is used. This was chosen as a balance between anomalous signal strength from the sulfur atom and adverse absorption effects. This can solve the phase problem encountered during crystallographic experiments and aid structure solution. This approach can be particularly powerful in assisting with low resolution model building by locating sulfur atoms within the native macromolecule. With respect to the work carried out in this thesis, S-SAD was used to help locate specific sulfur atoms from the flexible sucrose octasulfate (SOS) ligand, and potentially confirm binding in a specific orientation that could not be conclude from the original diffraction data alone. The statistics for this data collection are presented in Appendix B Table B.2. The resulting maps from the S-SAD experiment can be seen in Figure 3.9, the sulfur density from the nearby Met2079 residue can be seen in yellow, which indicates successful data collection and processing. The lack of clear S-SAD density in the ligand binding site indicates that the sulfur positions from the SOS ligand could not be accurately placed. This is not surprising, as the SOS ligand is known to be highly flexible and may be contacting different protein side chains in different orientations, giving rise to the dispersed density seen, and the lack of specific S positioning. However, when working with S-SAD data there is no specific σ level at which to present and analyse the sulphur density, instead each case should be assessed individually to ensure that the correct threshold is chosen to not include large amounts of noise. For this case, there is obvious density for the protein sulphur signal at 4.0σ and the lower cut-off 3.0σ. Therefore it would be appropriate to predict the small sulphur density seen in the ligand map (*) in Figure 3.9 was in-fact a S-group from the ligand. However, as

there is only a very small sulphur density, this does not particularly help with the placement of the final ligand outside of confirming a likely position of one of eight sulphur groups on the SOS ligand.

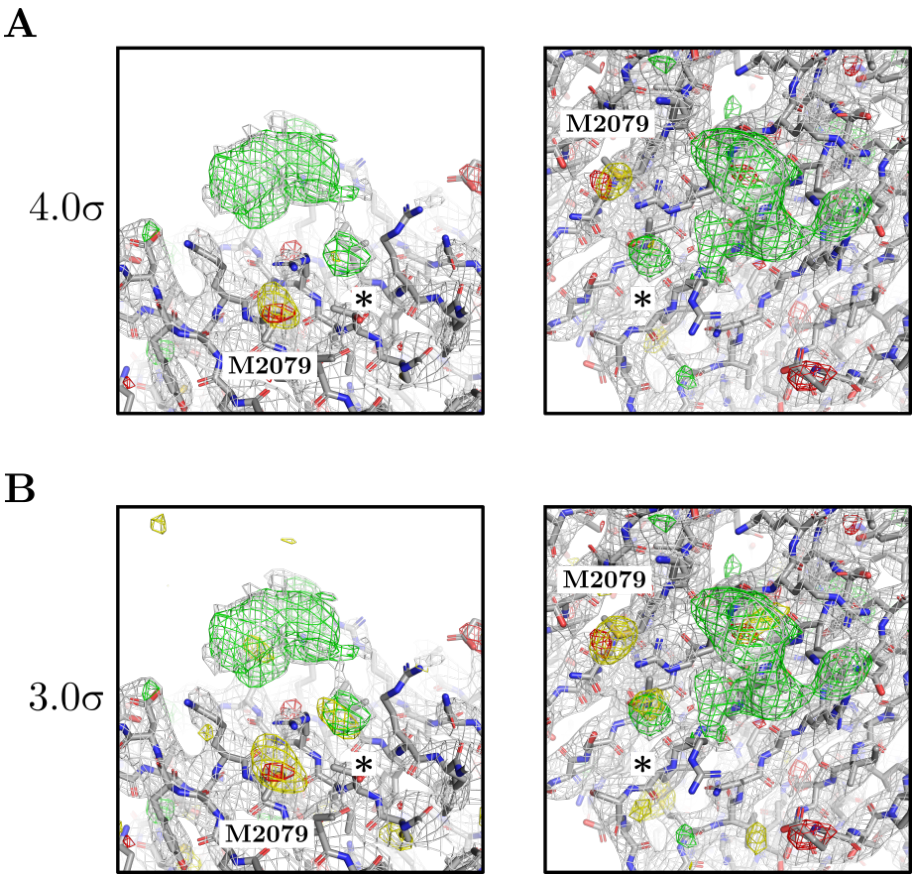


Figure 3.9: S-SAD density from the *Gg*Ten2$^{\Delta EGF}$–SOS complex. Electron density maps from the I23 S-SAD dataset collected for the SOS-complex crystal structure. **A.** Contouring the map to 4.0σ there is clear density for a nearby methionine side chain (*M2079*), and a very small signal in the ligand density (*). At 3.0σ, more S-SAD signal can be detected in the ligand binding pocket, however at this contour level it cannot be distinguished from background noise. The experimental 2Fo-Fc map is contoured to 1.0σ (*grey*). Difference maps are shown in *green* and *red*, contoured to 3.0σ. Anomalous difference maps generated from the S-SAD experiment are coloured *yellow*. All maps were generated after two rounds of refinement with Buster, to remove the majority of the noise and allow for S-SAD density to be seen more clearly.

3.3 The Teneurin–Glycan Complexes

Following the initial elucidation of the Teneurin–SOS complex described above, work was immediately moved to focus on the more physiologically relevant ligands chondroitin sulfate (CS) and heparan sulfate (HS). Due to their presence in the brain extracellular matrix, and their known roles in adhesion and cellular signalling[113, 196, 197], they were believed to be interesting targets for a potential role in teneurin adhesion, either by providing a novel binding partner, or by increasing the affinity between teneurin and a proteoglycan partner.

3.3.1 Crystallisation of the Teneurin–Glycan Complexes

Due to the heterogeneity resulting from differing sulfation patters of the main glycosaminoglycan (GAG) families (Chapter 1.3.1), multiple chondroitin sulfate (CS) and heparan sulfate (HS) species were investigated for crystallisation, including different oligosaccharide chain lengths, and differently sulfated species. Chondroitin sulfates were purchased as dimers with different sulfation positions (GalNAc, 4S (CS-A) and GalNAc, 6S (CS-C)), as well as a double (GlcA, 2S-GalNAc, 6S) sulfated tetrasaccharide species (CS-D) [*Iduron*]. Heparan sulfates were ordered in varying oligosaccharide lengths. The sulfated HS chains were ordered as a tetra-, and hexa-saccharide for crystallographic approaches (HS_{4mer} and HS_{6mer}), resembling the structure of heparin due to their fully sulfated functional groups.

Protein was produced for these assays as described above for the SOS crystal complex, initially supplemented with 3 or 9 mM of the desired ligand in each case. A crystal plate was set up with a dilution series of the same successful buffer conditions that produced the Teneurin–SOS crystal (Morpheus II, H3). This plate contained a diluted mother liquor (100-78%) across each of the 12 plate columns, to replicate the original conditions from the SOS dataset, whilst also optimising the condition for this new ligand. Alongside this dilution series, the core screens used previously were trialled (Morpheus, Morpheus II, JCSG+, MIDAS, Index, PGA, and PACT) as well as a PEG/Ion screen [*Hampton*] where protein yield allowed.

3.3.2 Heparan Sulfate (HS) Complex Structures

Heparan sulfate (HS) binding was investigated structurally due to the overwhelming evidence that the glycan chain plays major roles in cellular signalling[81, 82] and neuronal development and cell migration[198, 199]. The degree of sulfation of this chain is highly diverse, with different chains ordinarily containing regions of high and low sullfur groups studded along the whole. Naturally, using long sugar chains for crystallographic approaches is counter intuitive, as the longer the chain length the more degrees of freedom and flexibility is possible, and therefore the less likely a single conformation is to be present in the crystal lattice. Because of this, the shortest HS chains available were ordered. Two lengths, a tetrasaccharide (HS_{4mer}) and hexa-saccharide (HS_{6mer}) were ordered from Iduron.

*Gg*Ten2$^{\Delta EGF}$ protein was expressed and purified as described previously for the SOS complex crystallisation trial, and as is described in detail in Section 2.2.5. Coarse screen plates were set up with 5 mg/ml protein incubated with 9 mM HS ligand, at a 1:1 protein:reservoir drop ratio, incubated at 18℃. Crystals began to appear after around 10 days incubation, and grew for up to another week before being harvested and stored in LN_2. Data collection was performed at the Diamond Light Source synchrotron, beamline I03. Of the multiple pucks of crystals fished and shot at the beamline, 3 samples are noteworthy as they diffracted to a moderate resolution and were automatically processed through either the AutoPROC or DIALS pipelines (Figure 3.10).

Post processing of these datasets was successful, with molecular replacement placing one copy of teneurin in the HS_{4mer} dataset, and two copies per ASU for both the HS_{6mer} datasets, based on the results from a previous Matthew's coefficient calculation. An initial refinement run was performed to weed out datasets that contained only the protein, and had no density for the heparan sulfate ligand. The first two datasets, with HS_{4mer} and HS_{6mer}, respectively, showed an area, albeit small, of positive density that could be caused by a bound glycan ligand. Interestingly, the HS_{4mer} looks to bind at a unique location, close to residue R1893, a singular charged

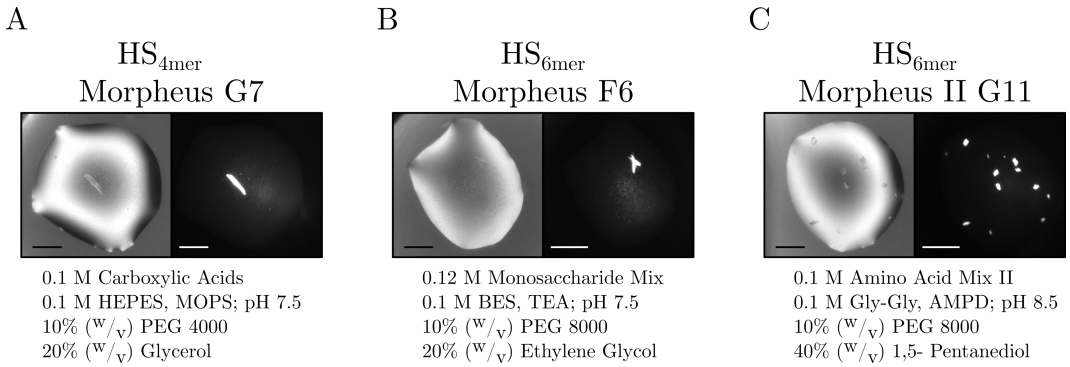


Figure 3.10: Diffraction quality crystal samples of the *Gg*Ten2$^{\Delta EGF}$-HS co-crystallisation trials which could be automatically processed by IspyB. Bright-field and UV images of each drop are shown. All conditions were incubated at 18°C. Scale bars = 250 μm. **A.** A cluster of long sheet crystals with the shorter HS_{4mer} ligand grew in Morpheus G7 and was auto-processed to ∼4.43 Å resolution. **B.** HS_{6mer} containing crystals grew in Morpheus F6 and diffracted to ∼4.14 Å resolution. **C.** HS_{6mer} containing crystals grew in Morpheus II G11 and the data was auto-processed to ∼4.33 Å resolution.

residue protruding from the teneurin YD-shell (Figure 3.11A). Interestingly, this smaller ligand shows no density at the previously identified binding site, centred around R2106. This newly identified potential binding site is intriguing, as the potentially bound ligand contacts only a single arginine residue protruding into the solvent space, which would normally not be enough to capture the relatively long and flexible HS ligand in a consistent enough conformation to be visible in the crystal electron density. However this binding site is located close to a region of crystal contacts, which artificially increase the order at that region. This patch of close contact between the two symmetry mates may be locking a potential ligand into place at a site not normally able to bind heparan sulfate (Figure 3.11A). The longer HS_{6mer} glycan does show apparent density at the 'traditional' binding site for one of the two datasets, (Figure 3.11B), although only a small region. When inspecting the maps of the third dataset visually using Coot, it was apparent that the crystal grown in Morpheus II, well G11, did not contain density for the HS chain. No discernible regions of unassigned density were present on the surface of the whole teneurin protein. This dataset, which diffracted to ∼4.33 Å resolution, was therefore not progressed further through refinement (Figure 3.11C).

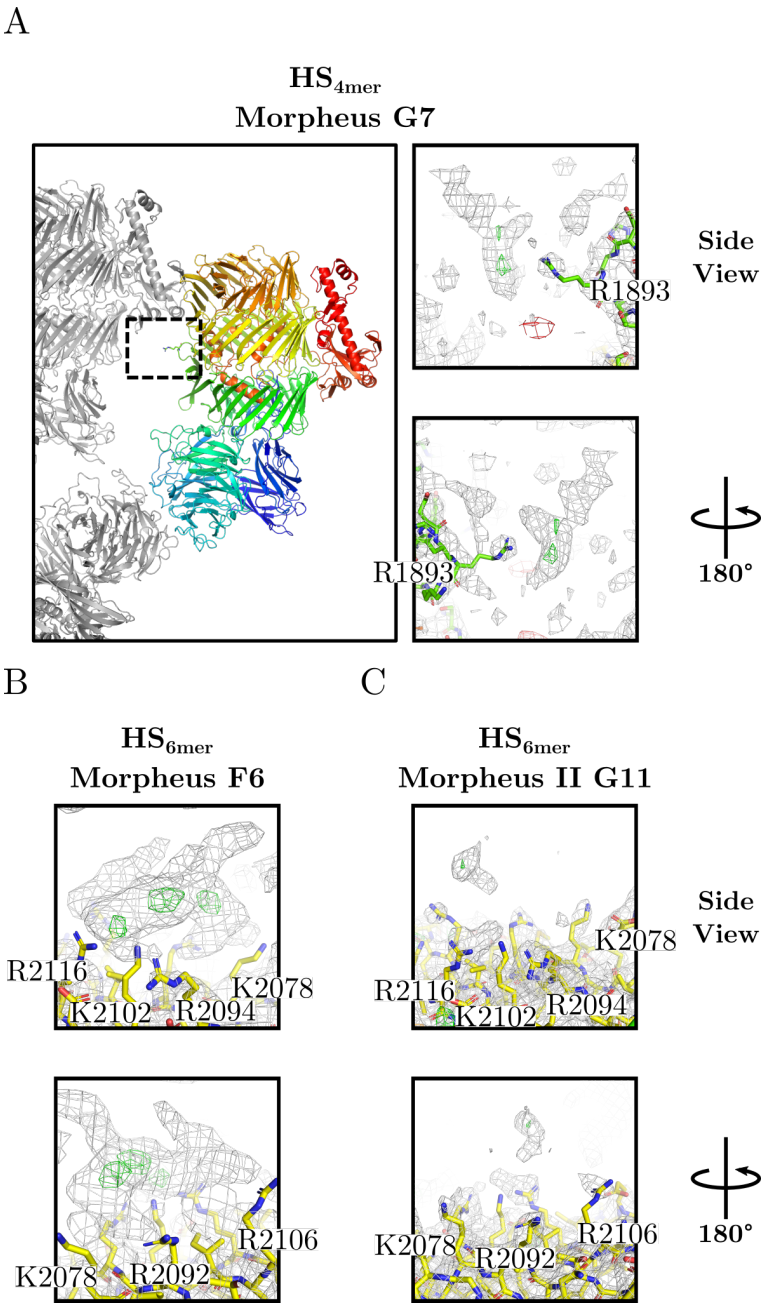


Figure 3.11: Unassigned density at the glycan binding site of the *Gg*Ten2$^{\Delta EGF}$–HS datasets. A. Unassigned density for the teneurin–HS_{4mer} structure reveals a potential new binding site at residue R1893 (*Dashed square*). **B.** Automatically processed data from ISpyB of a crystal containing the hexa-saccharide ligand (HS_{6mer}) revealed density at the original binding site on the YD-shell. **C.** A second dataset set up with the longer HS chain did not reveal any density on the protein surface. 2Fo-Fc map is coloured *grey* and contoured to 1.0σ. Fo-Fc difference map is contoured to $\pm$ 3.0σ coloured *red* and *green*. All maps were generated after an initial single refinement run with Refmac, to remove noise and reveal un-assigned denisty at the solvent accessible protein surface.

Further refinement of the shorter HS_{4mer} dataset subsequently eliminated any unassigned density at the R1893 binding site, indicating that the initial density was in-fact noise, not a bound HS chain. This is not too surprising as the single arginine residue at this site was unlikely to be able to bind a sugar repeat on its own, and I believe the initial appearance of any density was due to the artificial environment caused by the nearby crystal contacts.

Improving the contouring of the potential ligand density in the HS_{6mer} dataset was achieved through much the same means as the original sucrose octasulfate data. As with that model, the resolution of the HS bound models were of lower resolution than the native teneurin structure. As such this was used as a reference restraint throughout, and refinement of the protein backbone was minimal, with only small obvious changes being made with a bias towards satisfying the chemical and physical restraints of specific amino acids (rotamers and Phi/Psi angles) over the information from the electron density maps. Instead of improving the protein placement therefore, the same strategy as before was employed to improve the incorrect placement of the N-linked glycan chains (Figure 3.12). Modifications were made to the conservatively placed N-linked glycans throughout both protein chains of the long HS ligand dataset, often replacing a single or pair of N-Acetlyglucosamine (GlcNAc) monomers with a near complete branched high mannose tree, which was present due to the application of kifunensine during protein expression (Figure 2.1). Unfortunately, due to the fact that the atoms in the glycans make up a small amount of the total scattering of the whole dataset, there were minimal changes in the electron density map at the ligand binding sites in the HS complex structure. The contribution of the missing or differing glycans (and the very minor changes made to a handful of amino acid conformations) were not enough to improve the sculpting of the HS ligand binding site, and therefore the final maps after refinement closely resemble the initial data. This therefore again makes is difficult to accurately place the HS ligand chains in their respective density.

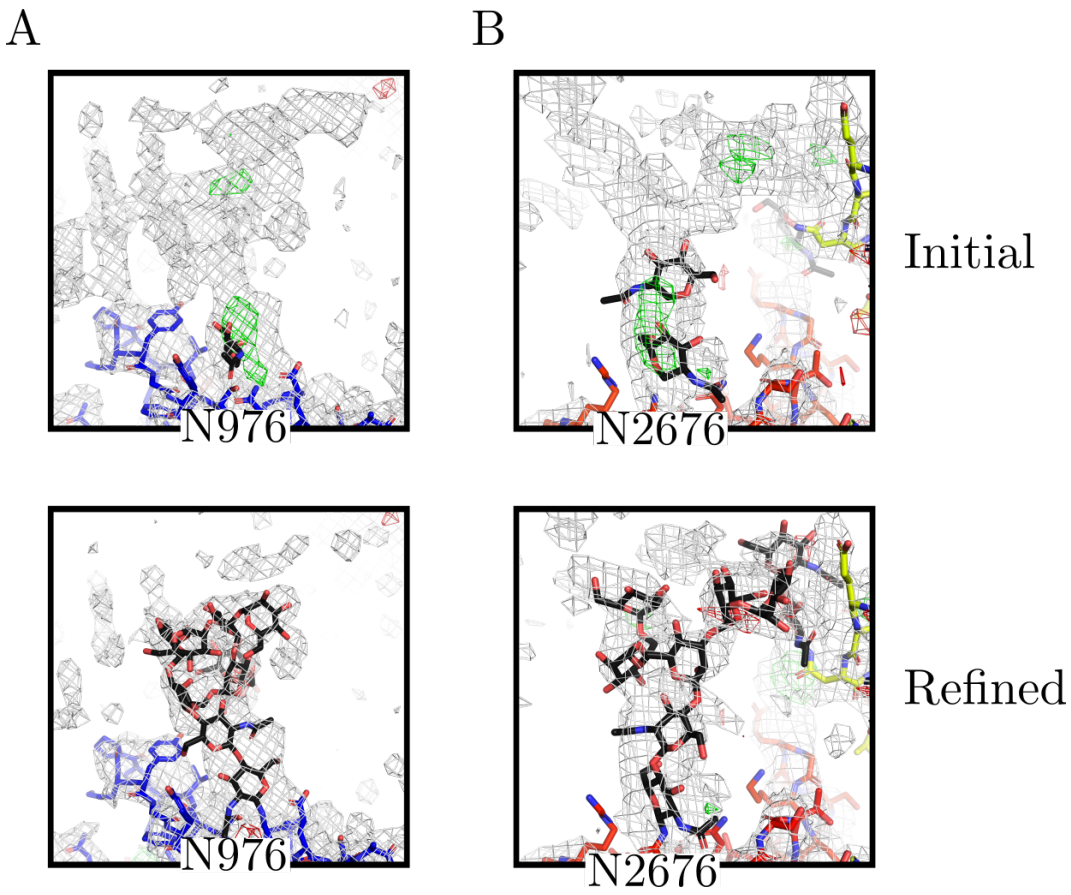


Figure 3.12: Glycan refinement carried out for the heparan sulfate containing dataset. The crystal structure of teneurin bound to HS_{6mer} had extensive unassigned density at the N-linked glycan sites. **A.** At this specific site (N976) A total of 10 (of a possible 11) glycans were able to be assigned to the electron density. **B.** At the N2676 site, 2 initially placed glycans were extended to 7 total residues in the glycan tree. Initial maps (*above*) were generated after the model was built and extensively refined in Buster. These maps were used to correct glycan density. Refined maps (*below*) were processed once more with identical Buster parameters to aseess for correct building of the glycans. 2Fo-Fc map is coloured *grey* and contoured to 1.0σ. Fo-Fc difference map is contoured to $\pm$ 3.0σ coloured *red* and *green*.

Another consideration with these ligand in particular, which was not an issue with the SOS disaccharide ligand solved previously, is the fact that the hexa-saccharide heparan sulfate ligand does not show complete density for its whole chain, only a small portion most closely associated with their respective protein binding residues. This is almost definitely a result of the flexibility of the glycan chain, which may only be pinned to the teneurin structure at one disaccharide, leaving the other repeats to exist flexibly in the solvent. This does make it is more difficult to place

the whole glycan chain in the correct place, and is impossible for the data presented here. Due to the fact that no density is visible in the dataset outside of the small region directly contacting the amino acid side chains, it is impossible to know where along the HS chain it is binding. There is no extra evidence in the data that, for example, the central disaccharide is binding to the teneurin surface over one end of the chain binding and leaving only one side with flexible, un-assignable, glycan density. Because of these issues, the placed ligand was treated as a disaccharide, and only the two repeating heparan sulfate sugar monomers (Iduronic acid and N-Acetlyglucosamine) were attempted to be assigned to the density (Figure 3.13). Polder maps, as were implemented for the SOS dataset above, were again generated for this dataset to improve the placement of the included disaccharide, as well as to potentially extend the density for the remaining glycan chain. However these maps did not improve the density at all, and as such were not used for this case. The final statistics for this dataset are shown in Table 3.3.

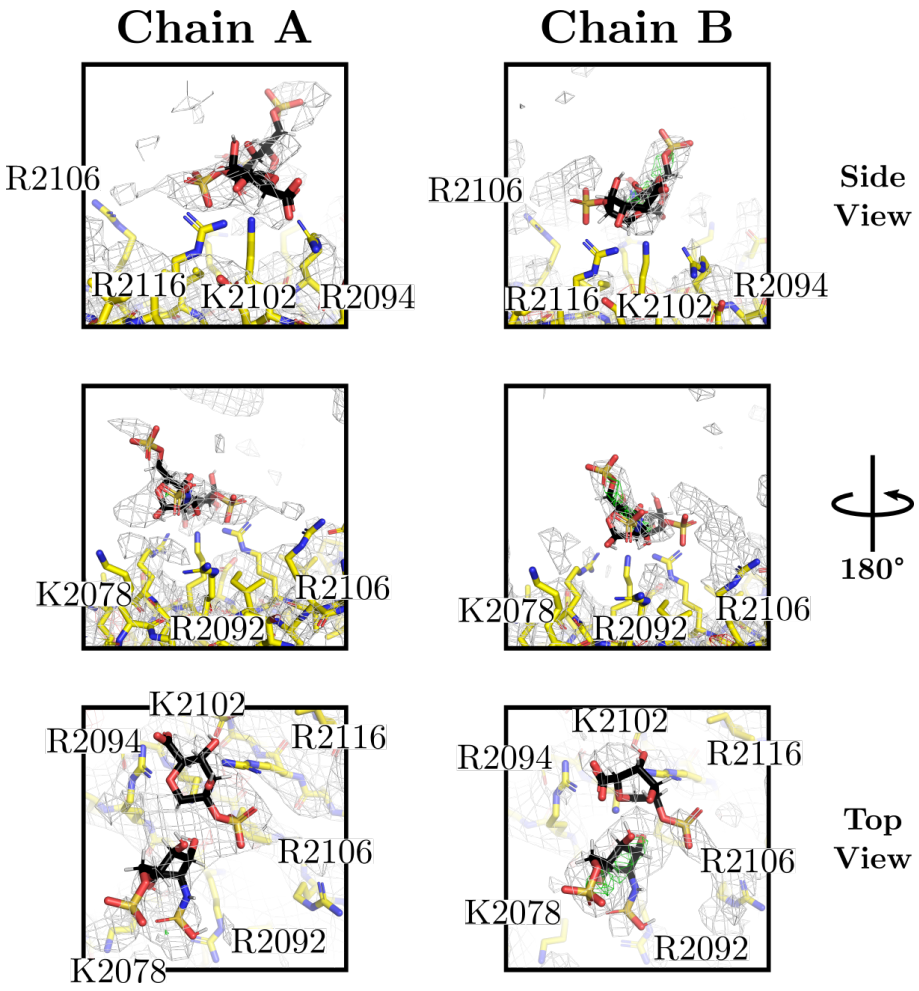


Figure 3.13: HS_{6mer} disaccharide ligand placed in its unassigned density. The long HS ligand was placed in the unidentified ligand density contacting the known binding pocket (K2078, R2092, R2094, K2102, R2106, R2116). Density in both chain A and chain B could place one disaccharide repeat of the whole glycan, with no extended density indicating the remaining sugar chain. Both maps were generated after final refinement in Buster and mpde; validation using Molprobity and Coot refinement indicators. 2Fo-Fc map is coloured *grey* and contoured to 1.0σ. Fo-Fc difference map is contoured to $\pm$ 3.0σ coloured *red* and *green*.

Table 3.3: Data collection, and model refinement and geometry statistics for the *Gg*Ten2$^{\Delta EGF}$–Heparan sulfate (HS_{6mer}) structure. Values for the high resolution shell are shown in parentheses.

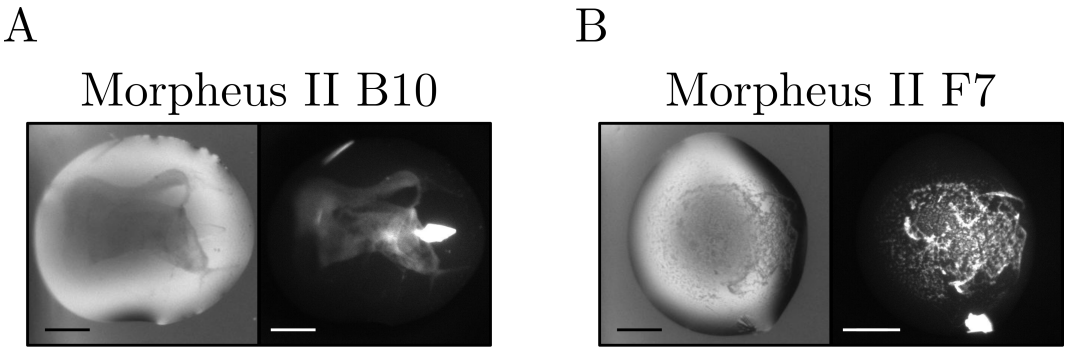
Protein Construct	*Gg*Ten2$^{\Delta EGF}$
Ligand	HS_{6mer}
Condition	Morpheus F6
Resolution Range (Å)	
Low Resolution	154.70
High Resolution	4.14 (4.270–4.143)
Space Group	P 1 2_1 1
Cell Dimensions	
a, *b*, *c* (Å)	140.40, 132.91, 158.82
α, β, γ (°)	90.00, 103.14, 90.00
R_{meas} (%)	0.955 (7.089)
Observations	297021 (14351)
Unique Observations	43457 (2016)
$I/\sigma I$	1.7 (0.2)
$CC_{1/2}$ (%)	100 (20)
Completness (%)	99.8 (94.8)
Multiplicity	6.8
Refinement	
No. Reflections	39816
R_{Work}	0.3153
R_{Free}	0.3179
No. Atoms	
Protein	29006
Ligand/ Ion	1151
R.M.S.D	
Bond Lengths (Å)	0.010
Bond Angles (°)	1.47
Validation	
MolProbity Score	1.64
Rotamer Outliers (%)	2.22
Clashscore	3.46
Ramachandran Plot	
Favoured (%)	96.40
Allowed (%)	3.33
Outliers (%)	0.27
Average B Factors (Å)	
Protein	118.05
Ligand	136.78

3.3.3 Chondroitin Sulfate (CS) Complex Structures

Due to the apparent importance of glycan sulfation pattern on function, particularly with respect to the chondroitin sulfate familiy of GAGs, different species were purchased and tested for crystallographic approaches. The two main sub-species of CS: chondroitin 4 sulfate (CS–A, GlcA–GalNAc 4S) and chondroitin 6 sulfate (CS–C, GlcA–GalNAc 6S) were purchased in their smallest repeating disaccharide, as well as the double sulfated form chondroitin 2,6 disulfate (CS–D, GlcA 2S–GalNAc 6S). These glycan forms are of particular interest due to their expression pattern in the developing brain, where the ratio of 4S : 6S increases during embryonic development and into adulthood, coinciding with growing plasticity in the brain. As with the crystallisation trials previously described above, the same *Gg*Ten2$^{\Delta EGF}$ protein construct was expressed in HEK293T cells with additional kifunensine, purified by IMAC and SEC, and concentrated to 5 mg/ml. The CS samples were added at a concentration of 9 mM and sitting drop vapour diffusion plates were stamped out using a Mosquito liquid handling robot. A range of crystallisation trial coarse screens were set up. Morpheus, Morpheus II, JCSG+, PACT, PGA, MIDAS, and Index plates were prepared with a 1 : 1 protein : reservoir drop ratio and incubated at 18℃.

Crystals appeared between 7 and 21 days after incubation, with mostly needle or sheet morphology which were often birefringent. Notably chondroitin 6 sulfate complex crystals grew more readily than those containing either chondroitin 4 sulfate or chondroitin 2,6 disulfate. A total of 20 crystals spanning the three CS ligand assay setups were harvested and cooled in LN_2 for data collection. Crystals that diffracted and were able to be automatically processed with the IspyB DIALS or AutoPROC pipelines to a moderate resolution were found in two conditions of a Morpheus II coarse screen, from wells B10 and F7 (Figure 3.14).

Molecular replacement was successful, and showed a single copy of teneurin per ASU in each case. As before, an initial refinement cycle was performed to clear up the majority of the noise before assessing for unassigned density which could be



2 mM Divalent Mix II
0.1 M Gly-Gly, AMPD; pH 8.5
12.5% ($^{w}/_{v}$) PEG 4000
20% ($^{w}/_{v}$) 1,2,6- Hexanetriol

0.1 M Monosaccharide Mix II
0.1 M BES, TEA; pH 7.5
10% ($^{w}/_{v}$) PEG 8000
20% ($^{w}/_{v}$) 1,5- Pentanediol

Figure 3.14: Diffraction quality crystal samples of the *Gg*Ten2$^{\Delta EGF}$-CS co-crystallisation trials which could be automatically processed by IspyB. Bright-field and UV images of each drop are shown. All conditions were incubated at 18°C. Scale bars = 250 μm. **A.** The large sheet crystal grown in Morpheus II B10 was autoprocessed to ∼3.88 Å. **B.** The cluster of small sheet crystals in the Morpheus II F7 drop diffracted to ∼4.1 Å.

the ligand of interest. With the first dataset, there were no large areas of density around any portion of the protein surface, and there was no density at the SOS binding site (K2078, R2092, R2094, K2102, R2106, R21116) (Figure 3.15A). The second dataset, with the crystal originating from well F7, had a large density just offset from the traditional binding site (Figure 3.15B). This was investigated as potential density of the CS–C (6S) ligand, and refinement was performed to improve the map to allow for some placement of the glycan. However, when refinement was performed in Buster, this large density disappeared, indicating that it is in-fact noise.

A final attempt at refinement was performed in Phenix due to the possibility of solvent flattening of the ligand density. The Phenix program inherently handles bulk solvent flattening better than autoBuster, so it was used to try and reduce any reduction of density in this case. Unfortunately, as with Buster the Phenix refinement output eliminated all of the density, corroborating that this was just noise and not contributed by a bound glycan ligand (Figure 3.16). The degree of solvent flattening can be assesed visually by inspecting the maps, specifically at

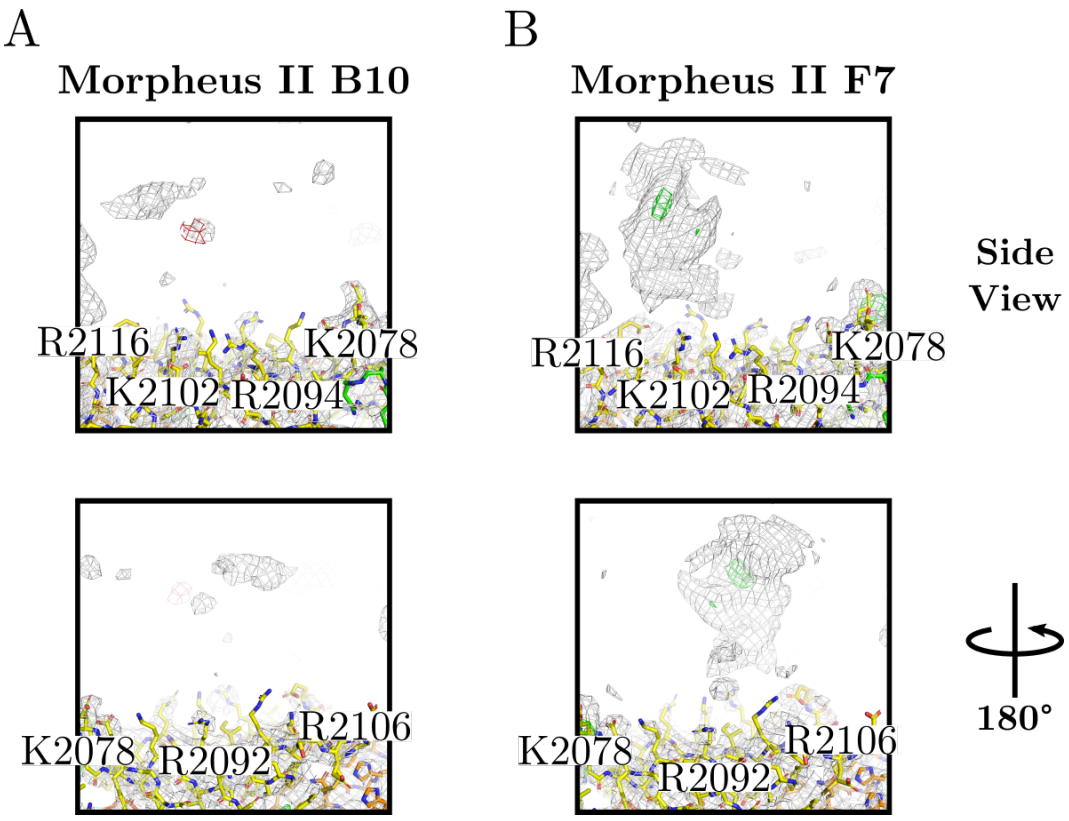


Figure 3.15: Unassigned density at the glycan binding site of the *Gg*Ten2$^{\Delta EGF}$–CS datasets. A. The first dataset had no discernible density present at the original SOS binding site or anywhere else along its surface. **B.** The second dataset has a large density patch just offset from the binding site. However after refinement in Buster and Phenix it was eliminated as noise. 2Fo-Fc map is coloured *grey* and contoured to 1.0σ. Fo-Fc difference map is contoured to $\pm$ 3.0σ coloured *red* and *green*.

the interface between protein and solvent, but also with the K_{Solv} value generated during refinement. This is a measure of the ratio of solvent density to that of the protein, where the average values are approximately 0.9 (indicating more density in the protein region comapred to the solvent). It can be seen that the K_{Solv} values differ slightly depending on the program used, with Buster exhibiting more flattening (a higher K_{Solv}) compared to both Refmac and Phenix refinement strategies. These slight differences in solvent can also be seen with the relative noise seen in the maps in the solvent space.

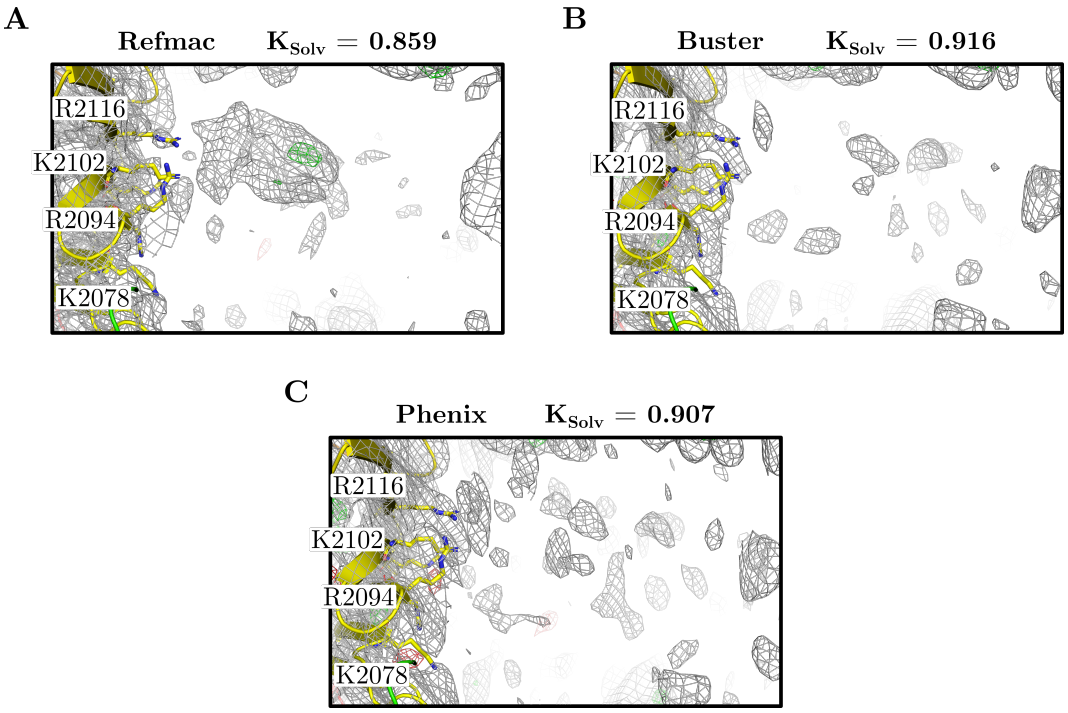


Figure 3.16: Solvent flattening in the three main refinement programs. Electron density maps for the *Gg*Ten2$^{\Delta EGF}$–CS dataset shown in Figure 3.15B in Refmac (*top*), Buster (*middle*), and Phenix (*bottom*) are represented. The 2Fo-Fc map is coloured *grey* and contoured to 1.0σ. Fo-Fc difference map is contoured to ± 3.0σ coloured *red* and *green*. K_{Solv} values from each refinement output are also quoted.

3.3.4 The Teneurin–Glycan Binding Site

The SOS and glycan binding site reported here can be corroborated by bioinformatic binding servers. The ClusPro binding server[200] is one such program. Originally built to simulate protein–protein docking, a heparin binding mode has been added. This mode predictively docks heparin tetramer molecules onto a submitted PDB structure. When the *Gg*Ten2$^{\Delta EGF}$ protein was submitted, binding was predicted to occur at one of two possible solvent accessible patches (other physiologically impossible locales were discarded). These two binding sites are shown in Figure 3.17. The same highly charged site shown above to bind SOS and HS was proposed to be a candidate for heparin binding. The server predicted interaction with the same protein side chains, showing the charged sulphur groups from the heparin interacting with the nitrogen from arginine and lysine residues. It also approximated a number of conformations for the heparin chain to bind in, again highlighting the potential

flexibility in glycan binding, as has been seen with the blurred diffraction maps in the structures above. Interestingly, another site, located on a highly charged flexible loop in the NHL domain, was proposed as a second binding site for heparin (R1247–R1249). This charged patch was previously overlooked, as no electron density appeared for the SOS or HS crystal datasets. It is possible that this binding site is not strong enough to capture the SOS ligand, and requires a longer chain GAG to contact the relatively linearly arranged charged side-chains.

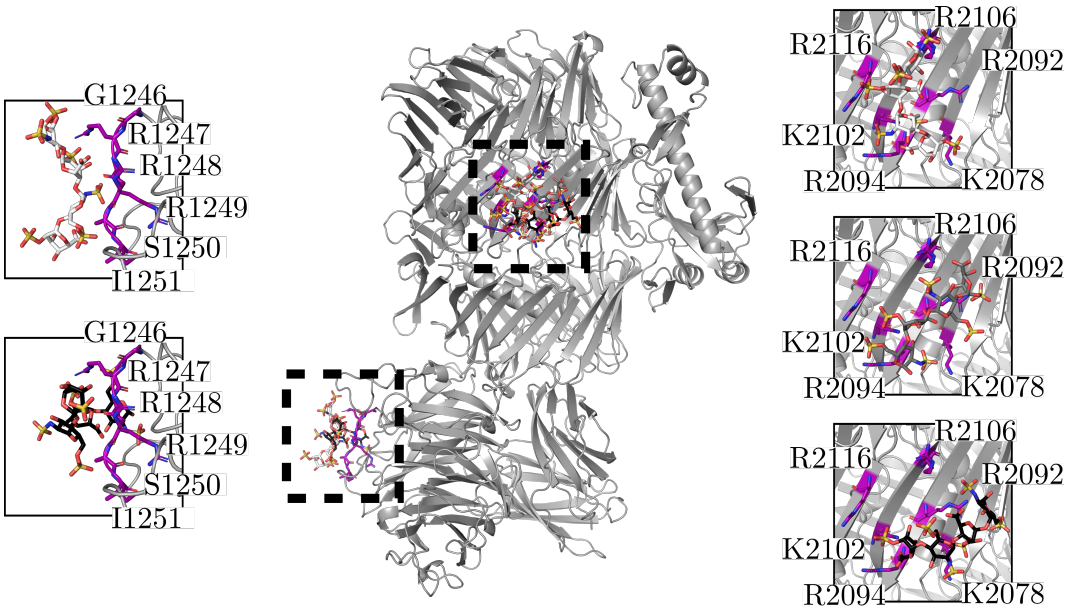


Figure 3.17: ClusPRO predicted binding between Teneurin and a tetrameric heparin chain. The ClusPRO binding server was used to corroborate the teneurin–SOS X-ray data set. The protein fragment (*grey*) was deposited, and the predicted heparin binding sites are highlighted (*purple*). The same binding site in the SOS structure was also identified with the ClusPRO server (*right*). Another possible binding site (*left*) on a flexible loop of the NHL domain was also found.

4

Biophysical Characterisation of Teneurin-Glycan Complexes

4.1 Summary

The crystal structures of the teneurin ecto-domain in complex with various glycan species revealed a compact binding site located on the YD-shell of the core protein fold. This chapter investigates the binding site in more detail biophysically, outlining the process of mutating the amino acids to abolish binding, proving a reduced binding via a pulldown assay utilising commercial heparin beads. Glycan binding was reduced to around 50% of the wild type capacity with the mutants cloned in this chapter.

I then moved on to investigate more physiologically relevant glycans, heparan sulfate (HS), and two sub-species of chondroitin sulfate (CS), CS–A and CS–C. I purchased biotinylated forms of these glycan species to probe the binding affinities between teneurin and the physiologically relevant ligands via SPR, however this proved to be difficult due to the heterogeneity of the glycan sample, and the potential for multiple binding events on a single immobilised glycan chain. I instead utilised the biotinylated samples for a pulldown assay. This involved coupling the modified glycans to streptavidin beads and assessing protein binding via western blot of the

appropriate samples.
Analysis of this assay reveals a preferential binding of teneurin to the HS glycan sample, and no statistically significant binding to either CS species. The mutants developed in this chapter show reduced binding to all of the glycan species, to the level of the negative controls.

The final experiment in this chapter outlines attempts to identify the presence of higher order protein complexes forming on a long glycan chain. This involved incubation of the whole extracellular domain of teneurin with a long heparan sulfate glycan (made up of around 100 repeating disaccharides). These samples were subjected to SEC to identify the presence of higher molecular weight species within the teneurin–HS complex, with the intention to utilise downstream SEC-MALS to estimate the molecular weight of the large oligomeric complexes.

4.2 Mutational Disruption of the Glycan Binding Site

To further identify and validate the potential GAG binding site to Teneurin, as shown from the initial Teneurin–SOS crystal structure (Section 3.2), mutational studies were centred around these binding residues. Primers were designed to either: reverse the charge of the positively charged Arginine and Lysine residues contacting the glycan to Glutamic Acid, or to introduce a bulky *N*-linked glycan into the GAG binding site (Appendix A Table A.2). Charge swap mutants were chosen instead of the commonly used approach of Alanine scanning[201, 202] due to the amino acids location on the solvent accessible surface of the protein and the highly charged GAG ligands. Of the six amino acids involved in binding, and therefore targeted for charge swap mutagenesis (K2078, R2092, R2904, K2102, R2106, and R2116) (Figure 4.1A, B), only two constructs, R2116E, and the double mutant R2092E+R2094E were successfully cloned and expressed. Alongside these charge swap mutants, the glycosylation mutant R2106N was successfully produced. This position was chosen

due to its central binding site location, and the pre-existing ^{2106}Arg-Phe-Ser (R-F-S) motif, allowing a single codon change to produce the Asn-Phe-Ser (N-F-S) *N*-linked glycan recognition sequence (Figure 4.1A).

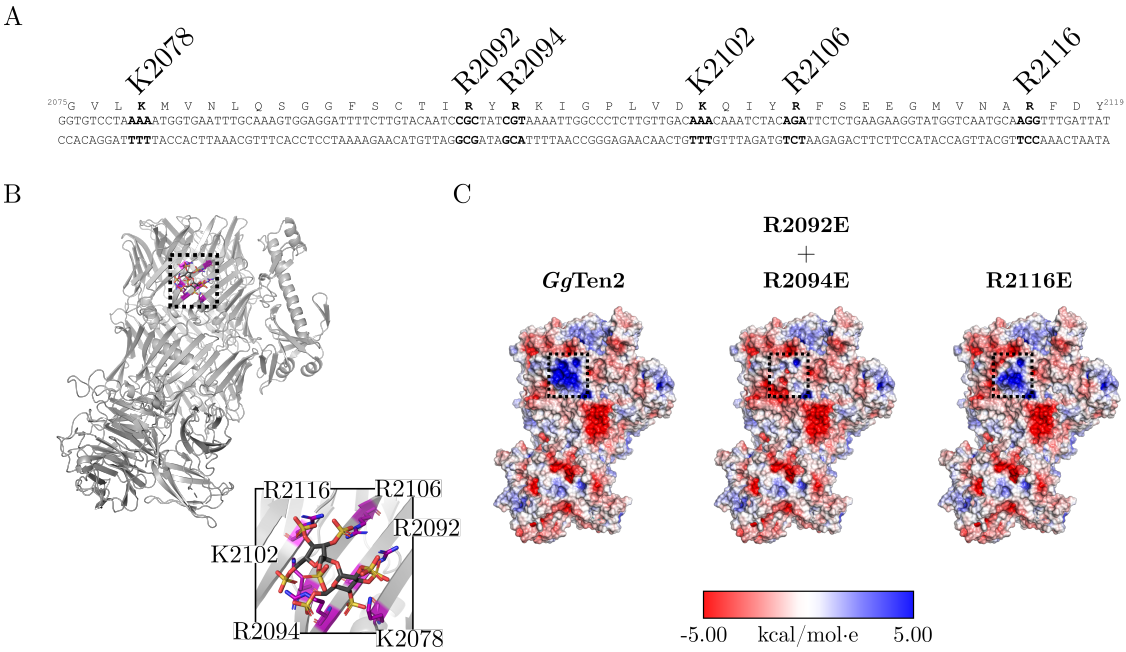


Figure 4.1: *Gg*Ten2$^{\Delta EGF}$ glycan binding mutant selection. A. The amino acid and DNA sequences of *Gg*Ten2 with each mutant position highlighted. **B.** Positions of each glycan binding mutant (*purple*) highlighted with respect to the whole *Gg*Ten2$^{\Delta EGF}$ structure. *Insert* Zoomed figure of the binding site, with bound sucrose octasulfate ligand (*black sticks*) corresponding to the binding site (*dashed square*). **C.** Surface representations of *Gg*Ten2$^{\Delta EGF}$ coloured by Coulomb charge from -5.00 to 5.00 kcal/mole·e. *Left* represents the original surface charge, *right* the R2092E+R2094E double mutant surface, and the R2116E surface charge.

Small scale expression tests were verified with each mutants prior to their use. Figure 4.2A shows one such expression test, which confirmed each mutant expressed at a similar level to that of the WT teneurin, with the exception of the combined triple mutant containing the R2116E and R2092E+R2094E mutations. This western blot was produced from a small 6-well plate format, and produced enough protein to complete small scale heparin bead purifications to assess glycan binding efficacy, which are described below.

Small scale heparin bead purification (Section 2.4.1) was carried out with these expected binding mutants, comparing the binding efficacy against the WT Teneurin. Comparing protein flow through and elution via western blot, mutation of the binding site reduced column binding compared to the native teneurin in all cases (Figure 4.2B). Statistical analysis of the elution percentage of each heparin purification showed a significant reduction in binding from each mutant protein. One-way analysis of variance (ANOVA) generates significant P values for all mutant samples compared to the WT teneurin (Table 4.1). Notably, there is no apparent statistical reduction in glycan binding when comparing any mutant to another, signifying that the single R2116E mutant, double R2092E+R2094E, and glycan R2106N mutants reduce biding to the same baseline, and as such the ability for some retention of the protein to the heparin beads may be due to another binding site, or retention of some binding at this site which would only be completely abolished if the binding site was further mutated or completely altered.

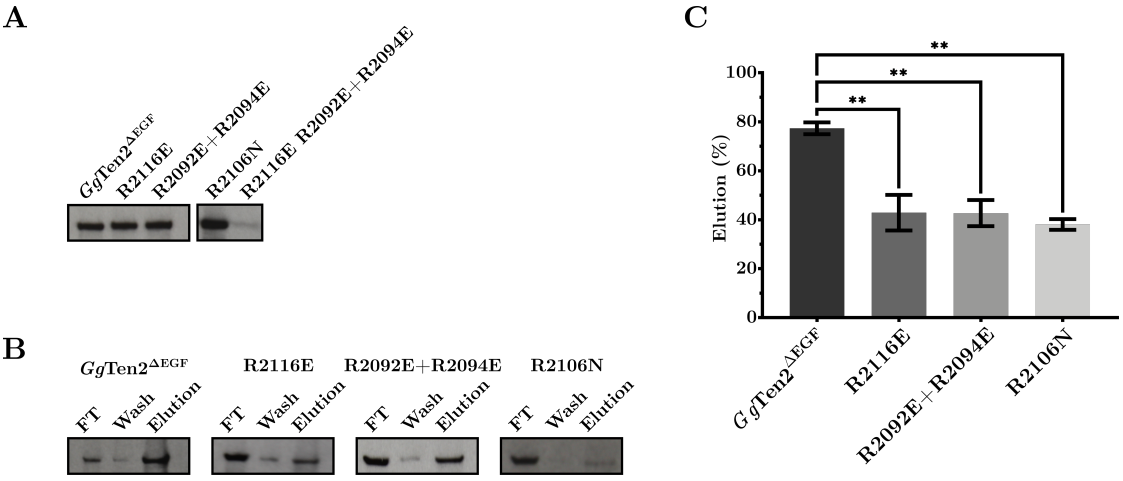


Figure 4.2: Small scale expression and Heparin purification of *Gg*Ten2$^{\Delta EGF}$ glycan binding mutants. A. Small scale expression tests of WT teneurin, and each mutant.The R2116E R2092E+R2094E expression was poor and as such was not carried forward. **B.** Small scale heparin purification of WT teneurin and each mutant. **C.** Graphical representation and statistical analysis ofcthe heparin purification assays. Error bars show SEM. One-way ANOVA test with Tukey's post hoc analysis, $p > 0.05$, n=3.

Table 4.1: One way ANOVA statistical analysis of the Teneurin–Heparin pulldowns.

Sample	R2092E+R2094E	R2116E	R2106N
***Gg*Ten2$^{\Delta EGF}$**	0.0042 (**)	0.0041 (**)	0.0018 (**)
R2092E+R2094E		0.9999	0.8899
R2116E			0.9001

4.3 Surface Plasmon Resonance (SPR)

Surface plasmon resonance (SPR) was employed to investigate the interaction between teneurin protein and glycan ligand in more depth than previously studied. To allow for the experiments to be carried out, a large scale purification of each glycan binding mutant was performed, yielding high enough amounts of pure mutant protein to flow over the SPR chip during testing, Figure 4.3 shows the final purified proteins in a Coomasie stained SDS-PAGE gel, after nickel affinity chromatography and size exclusion chromatography (SEC). These gels were used to determine which elution fractions to pool and concentrate, with only the purest fractions used for SPR.

4.3.1 Glycan Binding to *Gg*Ten2$^{\Delta EGF}$

Direct glycan binding to the chicken teneurin 2 ecto-domain was probed by SPR using immobilised biotinylated glycan ligand samples, and the purified glycan binding mutants, described above, as analytes. To identify potential structural preferences for teneurin binding, multiple glycan species were ordered from Iduron, each approximately 30 kDa in weigt and end-biotinylated to allow for SPR chip immobilisation. Both chondroitin sulfate-A (4S) and chrondroitin sulfate-C (6S) were purchased, as well as a single heparan sulfate chain. These SPR reagents were purchased to match those used for crystallography (Section 3.3.1). Each glycan was immobilised onto the SPR chip surface at a density of approximately 300 RU. Purified wild type *Gg*Ten2$^{\Delta EGF}$ and each binding mutant was flown over to assess binding. A positive control of the murine FLRT2 Leucine-rich repeat (LRR) domain

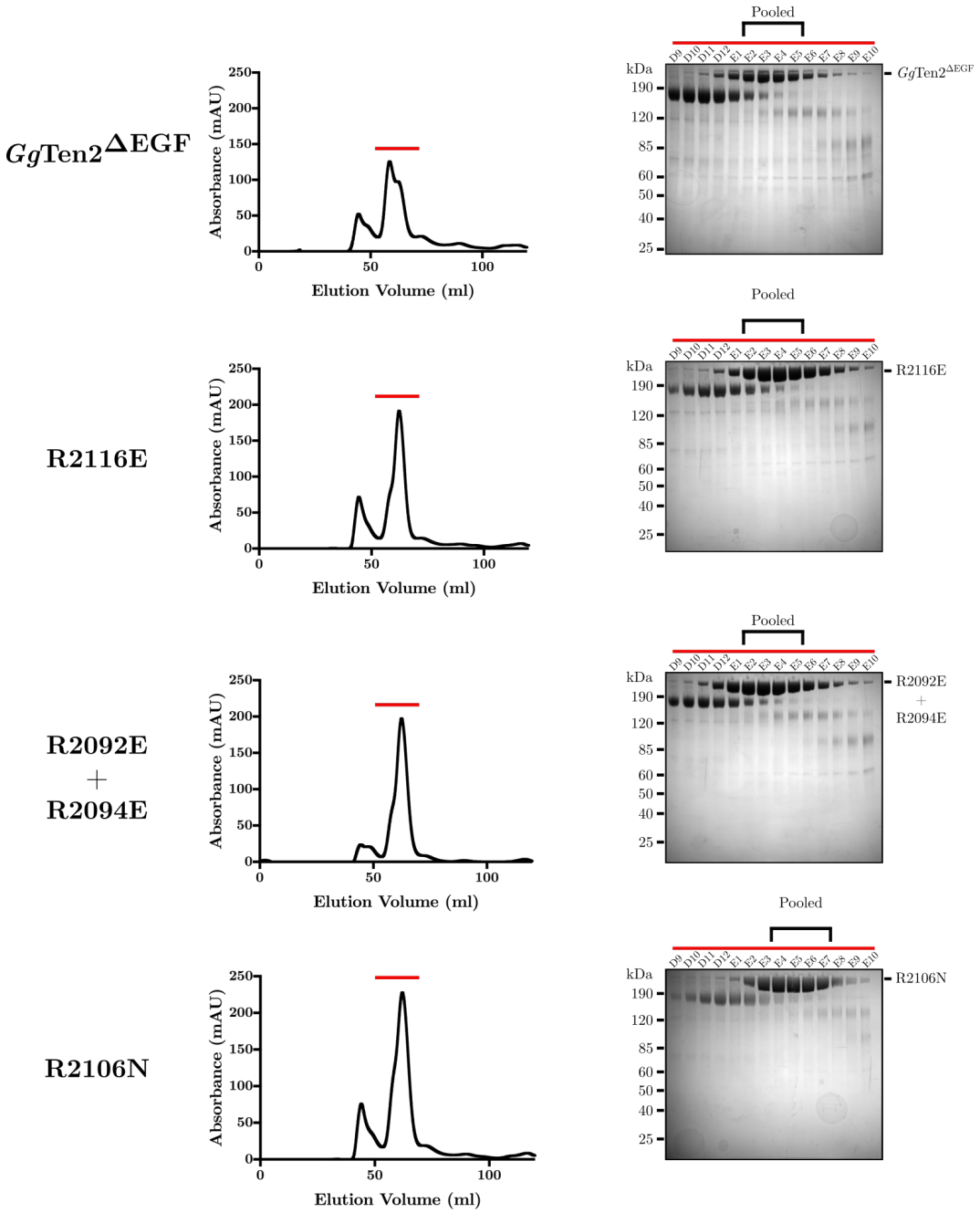


Figure 4.3: Purification of *Gg*Ten2$^{\Delta EGF}$ and mutants for use in SPR experiments. SEC traces and Coomasie stained gels are shown for each protein construct. The fractions run on the SDS-PAGE gels are highlighted above each elution profile (*red line*). The purest fractions pooled and concentrated for SPR experiments are indicated.

was used, as this construct has been show to bind to glycans in the literature[203, 204], as well as in our lab [*Unpublished data*].

When performing the SPR experiment, there were issues that arose during the immobilisation of the glycans, particularly when handling the CS–C sample. This glycan seemed to immobilise correctly, resulting in around ~300 RU of sample registered on the chip. However, whilst the SPR system was constantly flowing with buffer, the response units for the CS–C lane dropped throughout. The experiment was retried a second time and the same phenomenon occurred. As well as the issue with immobilisation of one of the glycans, the resulting SPR curves and equilibrium analysis (Figure 4.4, Table 4.2) generated peculiar results. Due to the length of the glycan chains used for the SPR experiment, and the nature of the teneurin binding, it is possible that the immobilised glycans contain multiple binding sites for the protein, and as such analysis with traditional $1:1$ binding models is futile. The analysis that was performed from this data (Table 4.2) also required equilibrium to have been achieved, which was not the case for many of the SPR curves acquired. A longer SPR injection time in future experiments would allow for equilibrium to be achieved in more of the samples. This equilibrium data is therefore only an estimate, and can not be convincingly used to assess the equilibrium of the teneurin-glycan interactions. The data does follow a similar trend throughout, consistently showing a weaker binding with each mutant construct compared to the wild type sample. However, due to the weak data from which the steady state analysis was generated, it cannot be used as conclusive evidence of reduced binding with this setup. The differences seen between glycan samples can also not be accurately assessed, as again the difficulties with the experimental set-up and resulting generated data throws doubt over the result.

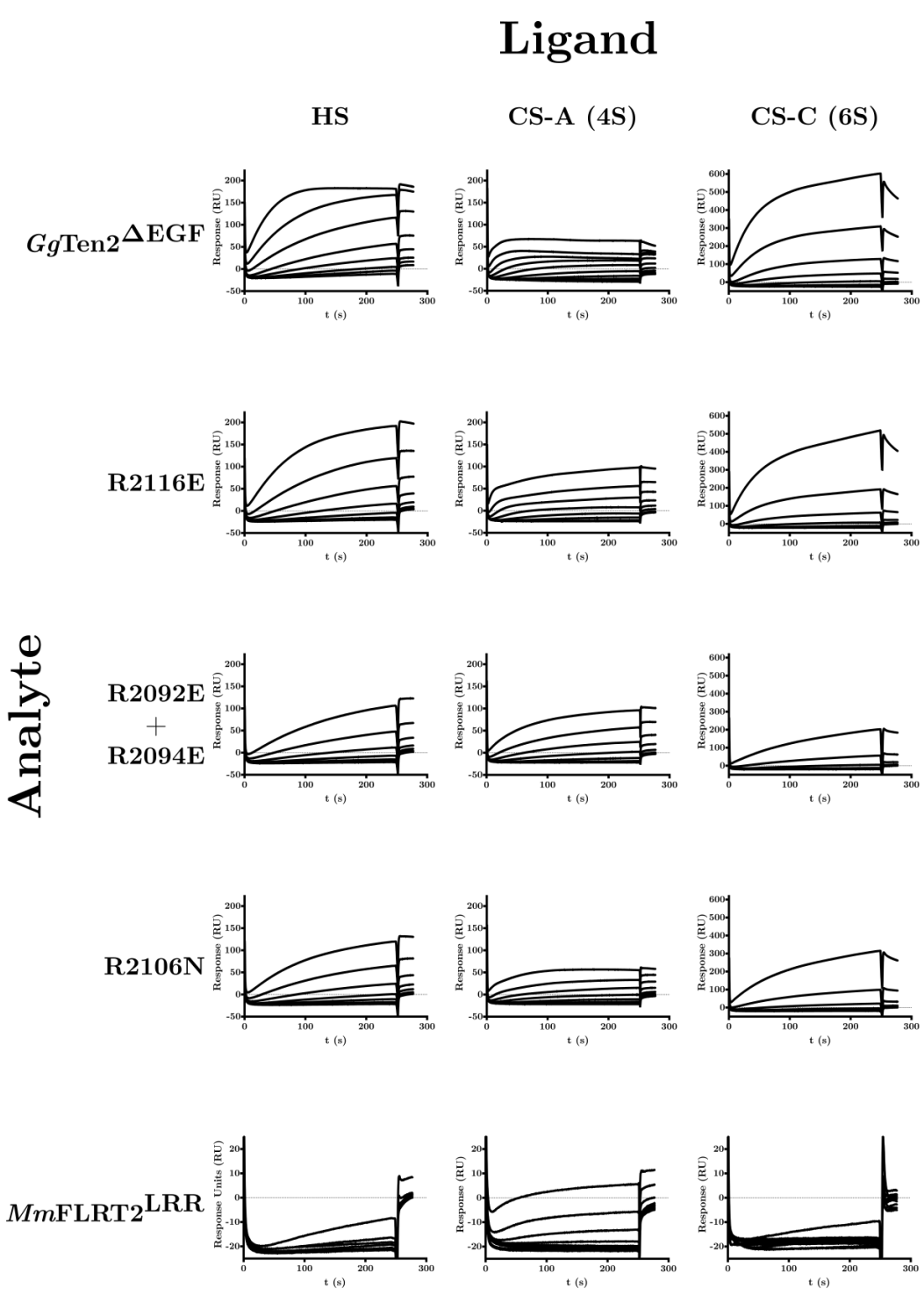


Figure 4.4: SPR binding curves of *Gg*Ten2$^{\Delta EGF}$ and mutants binding to glycans. *HS*, Heparan Sulfate; *CS-A*, Chondroitin Sulfate-A; *CS-C*, Chondroitin Sulfate-C. Axis are shown intersecting below the origin for clarity, Y=0 is represented by a dashed line.

Table 4.2: Calculated equilibrium analysis data of the Glycan–*Gg*Ten2$^{\Delta EGF}$ interactions.

Ligand	Analyte	K_D (M)	B_{max} (RU)
HS	*Gg*Ten2$^{\Delta EGF}$	9.29x10^{-8}	195.00
	R2116E	6.78x10^{-7}	333.60
	R2092E+R2094E	2.86x10^{-6}	538.30
	R2106N	6.31x10^{-7}	232.2
	*Mm*FLRT2^{LRR}	4.92x10^{-7}	40.75
CS-A (4S)	*Gg*Ten2$^{\Delta EGF}$	3.18x10^{-7}	54.64
	R2116E	1.94x10^{-6}	124.30
	R2092E+R2094E	2.06x10^{-6}	150.00
	R2106N	1.28x10^{-6}	65.87
	*Mm*FLRT2^{LRR}	3.58x10^{-6}	28.82
CS-C (6S)	*Gg*Ten2$^{\Delta EGF}$	1.68x10^{-5}	2467
	R2116E	2.50x10^{-4}	25200
	R2092E+R2094E	5.82x10^{-4}	27800
	R2106N	2.14x10^{-4}	13800
	*Mm*FLRT2^{LRR}	6.07x10^{-10}	0.03831

4.3.2 Glycan Dependancy of the Teneurin–Latrophilin Interaction

The protein interaction between teneurin and its binding partner latrophilin has been studied in detail[39, 46, 158], with recent structural results highlighting its binding site[36] (Figure 4.5A). To asses the effect of glycan chains on the interaction between teneurin and latrophilin (Lphn), constructs were cloned to isolate the glycan rich linker region situated between the extracellular olfactomedin and Horm/GAIN domains. This region has previously been reported to be studded with glycan attachments, both N- and O-linked to the polypeptide chain[197, 205]. Primers (Appendix A Table A.2) were used to isolate the linker region and clone into a pHLSec vector containing a C-terminal Avi tag, to allow for subsequent biotinylation (Figure 4.5B). Expression was confirmed with small scale expression tests as described before. Following the expression test, the same DNA prep was again used to transfect, this time with a BirA vector and supplementary free biotin. Successful biotinylation was confirmed by α-Strep western blot (Figure 4.5C), and free biotin was removed by dialysis.

SPR Chips were prepared as described in Section 2.4.2. Each *Mm*Lphn linker construct was immobilised in a separate flow cell, with approximately 350 RU immobilised. Upon test injection, it was discovered that the short linker sequences did not bind to the WT teneurin ecto-domain in the flow over, or any of the mutants. This result indicates that the minimal binding region for teneurin-latrophilin interactions is the latrophilin Lec-Olf protein domains, and the glycan studded linker downstream of this does not constitute a binding region. The glycosylated linker could be investigated in more detail. With treatment of glycosidic enzymes such as heparinase and chondroitinase, the glycan makeup of this small protein region could be probed in more detail, potentially unearthing a possible function and informing downstream approaches.

In previously published work from the lab in which teneurin-latrophilin interactions were probed in depth[36], the protein-protein interaction was identified by SPR to bind with nano-molar affinity, with the *Mm*Lphn3–*Gg*Ten2$^{\Delta EGF}$ interaction showing a slightly lower binding compared to other isoform combinations. Due to the tight binding seen between protein domains, it was decided to not continue investigating the effect of glycan binding on the teneurin–latrophilin interaction. Attempts to investigate the effect of glycans on this complex would have very little difference on the previous result, and such a small difference in binding could not be accurately attributed to the effects of glycans via SPR.

4.4 Streptavidin Bead Pulldowns

Following the difficulty using SPR to investigate different glycan interactions to the teneurin ecto-domain, it was decided to switch approaches and use a pulldown assay with purified protein to look more closely at the differing interactions. I utilised the same 30 kDa biotinylated glycans as were used in SPR, however this time immobilising them on loose Streptavidin beads. This would be used as an alternative for SPR, hopefully investigating the difference in binding between teneurin and each

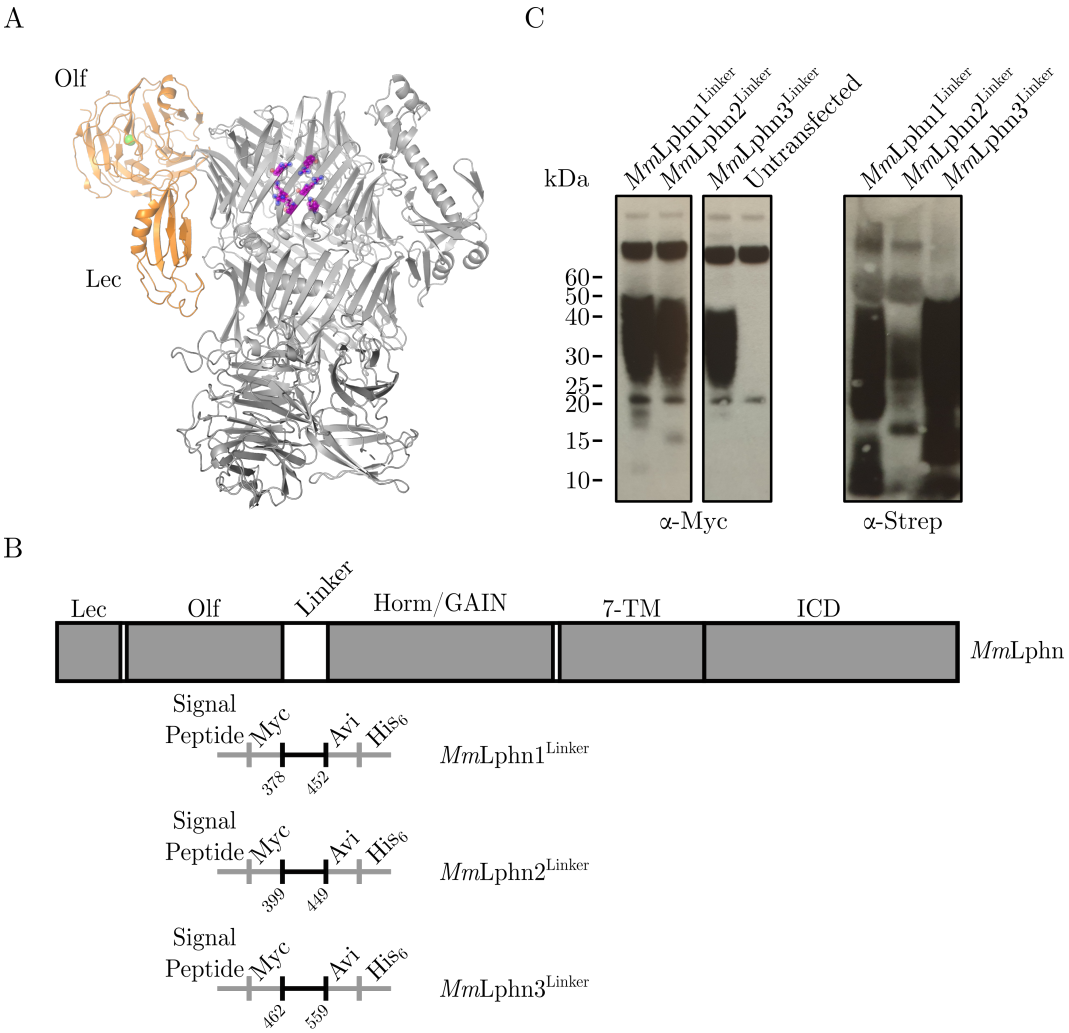


Figure 4.5: Latrophilin Constructs used for SPR probing of the teneurin-latrophilin interaction. A. The crystal structure of the teneurin-latrophilin interaction (PDB ID: 6SKA). Teneurin is shown as a cartoon representation (*grey*) with glycan binding residues highlighted (*purple*). Latrophilin Lec-Olf domains are shown (*orange*), with a Calcium ion (*green*). **B.** The domain organisation of *Mus musculus* Lphns. Each construct designed for SPR experiments including A.A boundaries are shown below. *Lec*, lectin domain; *Olf*, olfactomedin domain; *Horm*, hormone domain; *GAIN*, GPCR autoproteolysis-inducing domain; *7-TM*, seven-transmembrane domain; *ICD*, intracellular domain; *Myc*, Myc tag; *Avi*, Avitag; *His_6*, 6xHis tag. **C.** α-Myc and α-Strep western blots of the constructs shown in B. showing expression and subsequent successful biotinylation, respectively.

glycan species. Negative controls of streptavidin beads incubated with a biotinylated nanobody, which showed no binding to teneurin, in equal molar amounts to the test glycans, as well as the bare streptavidin beads alone were used. Biotinylated glycan or nanobody samples were prepared and immobilised on washed streptavidin

beads as described in Section 2.4.3. A known amount of either purified wild type *Gg*Ten2$^{\Delta EGF}$, or one glycan binding mutant, was incubated with the prepared streptavidin bead sample for 2 hr at room temperature before being washed and eluted. Protein binding to the respective glycan species was assessed by western blot, and band intensities were quantified. Figure 4.6 outlines the workflow for this assay.

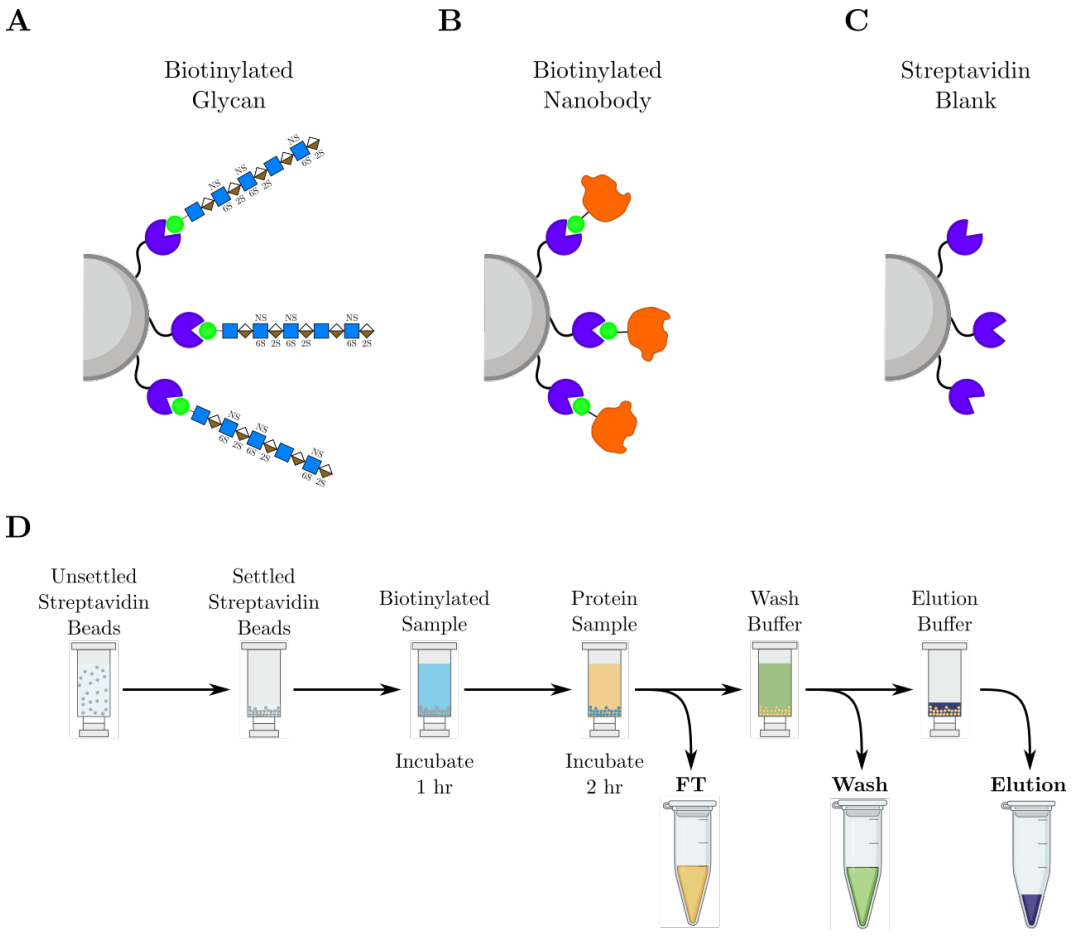


Figure 4.6: Workflow for the Streptavidin–Glycan–Teneurin pulldown assay. **A.** Schematic of the biotinylated glycans attached to streptavidin beads. **B.** Biontylated nanobody (*orange*) used as a negative control. **C.** The streptavidin blank sample used as a negative control. Agarose beads (*grey*) are studded with streptavidin (*purple*) which allows biotin (*green*) to immobilised the sample.**D.** Workflow for the pulldown assay. Beads were settled by centrifugation before being washed and the biotinylated sample incubated at room temperature. Excess glycan or nanobody was washed before teneurin incubation. The flow through, wash, and elution were assessed subsequently by western blots.

Western blots were used for assessment of protein content due to the small amounts of protein used in each pulldown, which itself was required due to the limited supply of biotinylated glycans. The proteins samples were previously purified for SPR, and as such were pure enough to produce clean SDS-PAGE gels stained with Coomasie (Figure 4.3). Western blots were imaged and assessed by Image J with each sample analysed in the same method as previously described for the Heparin bead pulldown above (Figure 4.7).

From these blots it can be seen that the wild type protein binds most strongly to the heparan sulfate (HS) glycan, which is supported by the graphical analysis (Figure 4.8). Two way ANOVA analysis was carried out between proteins for each corresponding sample, this showed that none of the equivalent samples (FT v FT or Bound v Bound) were statistically significant from each other. There is still a considerable proportion of the protein present in the wash sample for the HS complex, indicative of weak binding between the protein and glycan as these pulldowns were calculated to approximately 30x molar excess of glycan in each case. It is therefore a sign that the binding is not strong between the two macromolecules, perhaps unsurprisingly as the binding site only occupies a small fraction of the YD-shell surface. Unfortunately, the binding to both CS samples by the wild type protein is poor. With both glycans, the wild type is mostly lost in the experimental flow through, indicating that the protein is not bound to the glycosylated beads. All of the mutant samples show considerable amounts of protein in the flow through for each repeat, matching the blank nanobody and blank streptavidin controls. There is some retained binding to the HS beads however, where only about 50% of the protein is unbound, with the rest retained on the column. This, like the previously described initial Heparin bead pulldown (Section 4.2), may indicate an additional binding site for this glycan that was not disrupted by the mutations made, or the ability for the glycan to continue to bind even after a section of the charged teneurin binding site was mutated. More repeats of this assay would

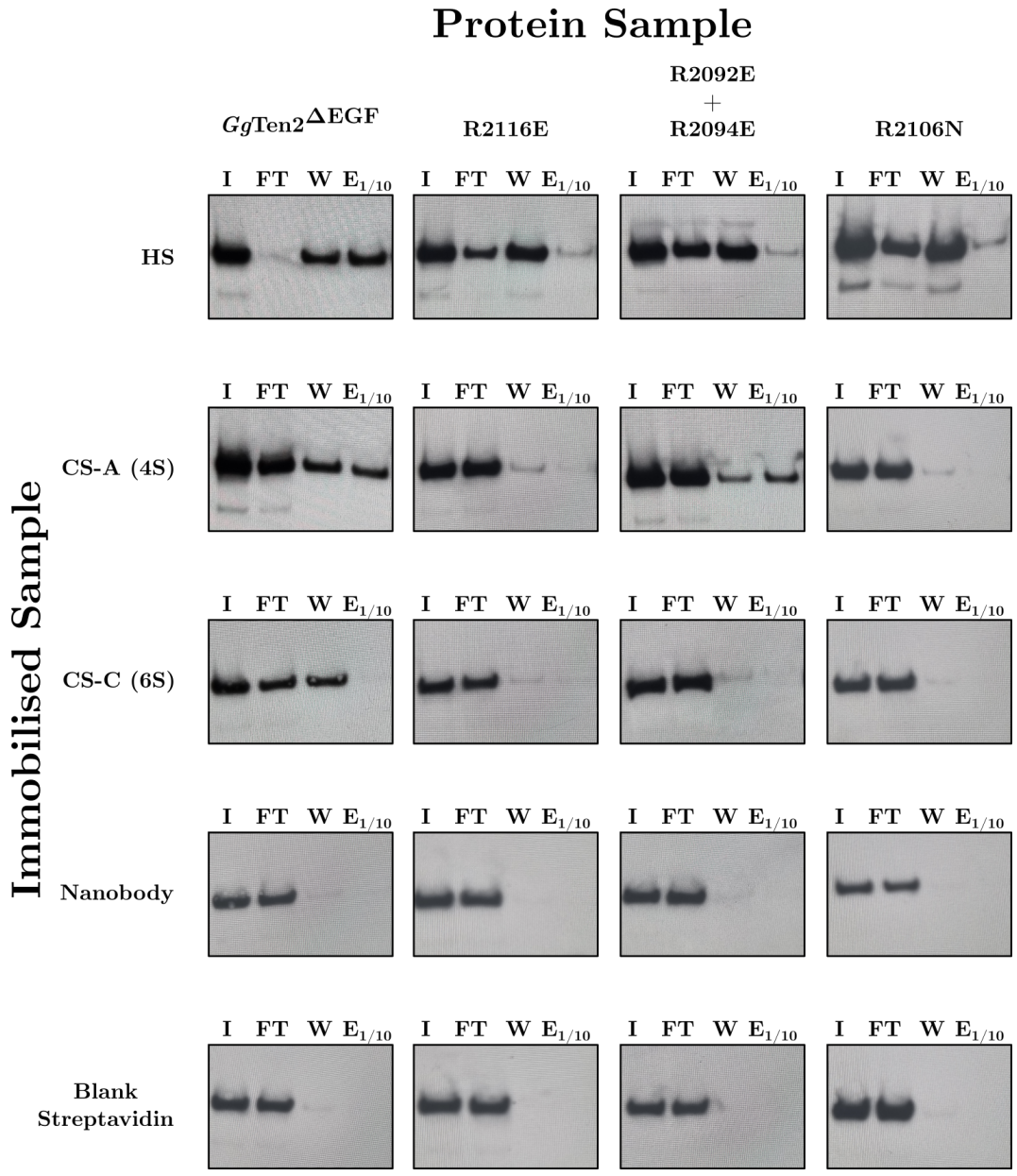


Figure 4.7: Biotinyalted glycan pulldowns comparing *Gg*Ten2$^{\Delta EGF}$ and mutants. Western blot images comparing Input [I], Flow-through [FT], Wash [W], and a diluted Elution sample [$E_{1/10}$], diluted to match the volume of the other sample to allow for comparable analysis.

be required to elucidate differential binding between protein and mutant for each glycan sample immobilised here.

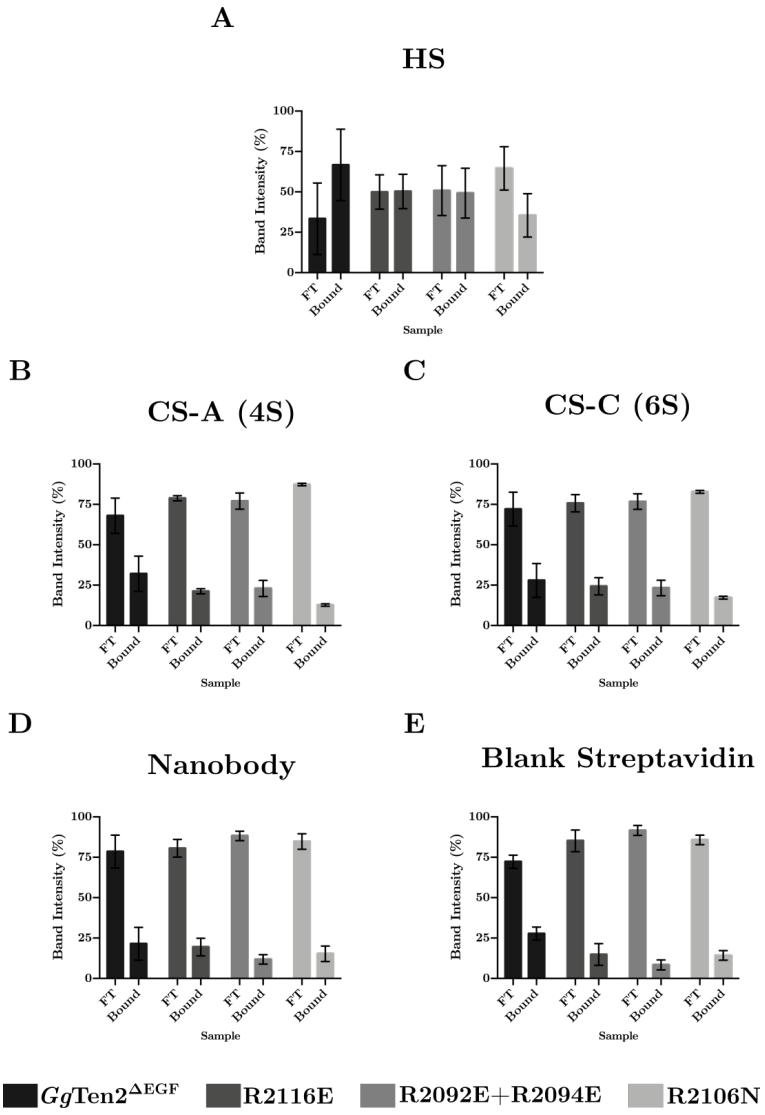


Figure 4.8: Graphical representation of the biotinylated-glycan pulldown assay with statistical analysis. Protein samples are distinguished by colour of the bars (*below*). Flow-through [*FT*] and bound sample (comprising of the Wash and Elution combined) are plotted as a percentage of the total band intensity. **A.** Immobilised Heparan Sulfate (HS) glycan. **B.** Immobilised Chondrotin Sulfate A (CS–A) glycan. **C.** Immobilised Chondrotin Sulfate C (CS–C) glycan. **D.** Immobilised biotinylated nanobody sample. **E.** Blank streptavidin control sample. Error bars show SEM. Two-way ANOVA test with Tukey's post hoc analysis, $p > 0.05$. All analyses were NS and didnt not meet the confidence threshold.

4.5 Teneurin–Heparan Sulfate Multimerisation

The ability for teneurin to multimerise along a long heparan sulfate chain was probed by SEC. This is an approach is commonly used as SEC-MALS, where multi-angle

light scattering is coupled to the size exclusion chromatography to estimate the molecular mass of the sample peaks. Heparan sulfate multimerisation has been assessed previously with this approach[86] where protein was seen to multimerise on long heparan sulfate chains, but not chondroitin sulfate. Purified full length teneurin ecto-domain ($Gg\text{Ten2}^{\text{Ecto}}$) was used due to its ability to dimerise owing to intermolecular disulphide bonds from each pair of EGF repeats. This dimeric form of teneurin is more closely matched to the native form of teneurin found on the cell, and has an increased avidity towards the heparan ligand due to the presence of two potential binding sites. The exact orientation of teneurin cis-dimer is not yet known however, so it may be possible that the dimeric form of teneurin blocks the glycan binding site for one or both of the proteins. Varying amounts of purified teneurin was incubated with a long 100-mer repeat of heparan sulfate (at 25x molar excess ligand) and incubated with the teneurin at room temperature for 1 hour before being loaded onto a Superose 6 SEC column. The resulting complex could be measured by migration time and assessed compared to the same amount of teneurin protein alone (Figure 4.9). A shift in the SEC profile of the complex sample towards a quicker retention time in the column therefore represents potential multimerisation.

The overlaid SEC profiles of increasing amounts of teneurin incubated with 25x molar excess of heparan sulfate can be seen in Figure 4.10A–D. Western blotting was used to assess the protein content of each elution fraction, and was used to compare the lone protein fraction with its corresponding complex sample (Figure 4.10E). As can be seen from both the chromatograms and western blots, teneurin does seem to interact with the long heparin chain. However, only a barely visible shift occurs. The shifted protein peak moves into the void volume, indicating that the protein actually aggregates when in complex with the heparan chain, and does not form distinct oligomeric forms. This may be a result of the protein failing to form a higher order state along the heparan chain, or due to an inherently weak interaction which requires a less harsh environment to survive. The experimental setup with 750 μg was repeated with 25x molar excess of teneurin, as apposed to

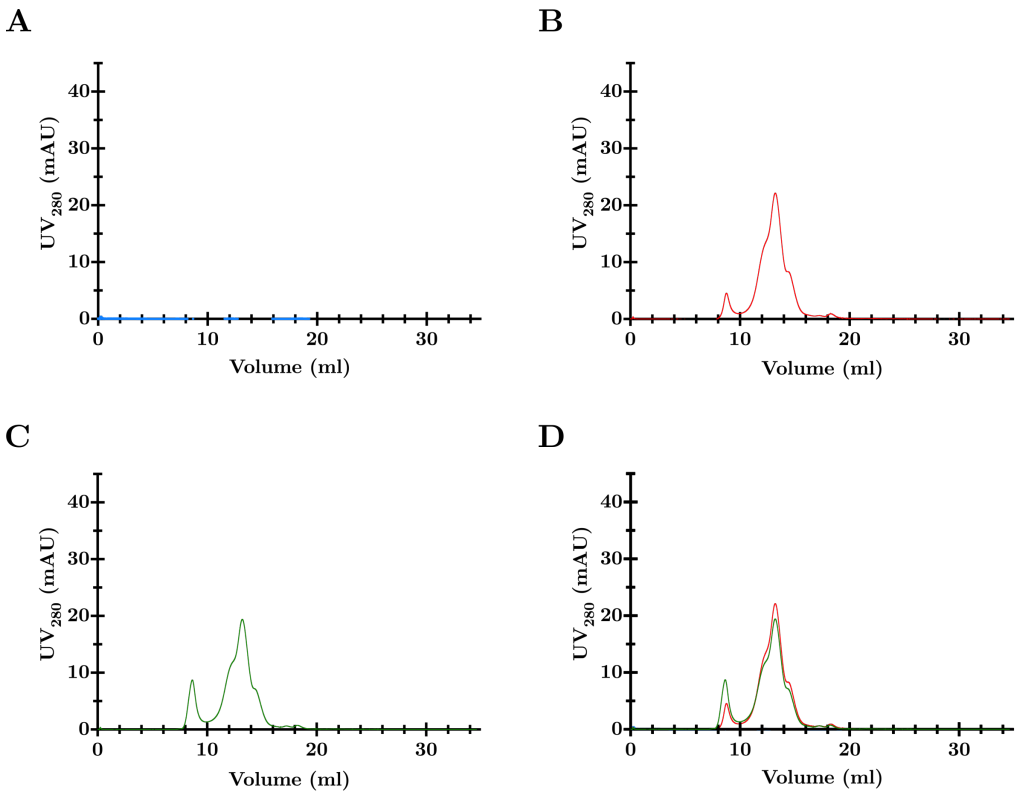


Figure 4.9: Multimerisation of *Gg*Ten2$^{\mathbf{Ecto}}$ on a long HS chain. A. HS alone run on a Superose 6 SEC column. **B.** 500 μg *Gg*Ten2^{Ecto} run on the same column alone, used as a reference point for the complex sample. **C.** 500 μg *Gg*Ten2^{Ecto} incubated with 25x molar excess HS. **D.** Overlay of all three graphs showing the increase in high molecular weight peak in the complex sample (*green*) compared to the protein alone (*red*).

excess HS glycan, to assess whether multimerisation occurs with excess protein, however there was no visible shift in the protein molecular weight [*Data not shown*]. Because of the unclear multimerisation and apparent aggregation of potentially sugar bound protein, it was decided not to progress this assay further by coupling MALS.

One final experiment was performed using the *Gg*Ten2$^{\Delta\mathrm{EGF}}$ construct used throughout the crystallography and other binding studies described in this thesis. Incubating this construct with the long heparan sulfate chain was hoped to form the dimeric form of protein, the most common multimeric arrangement of teneurin in the cell. Unfortunately, there was no visible shift in the protein peak during the SEC run when the teneurin protein was incubated with the long glycan chain [*Data not shown*],

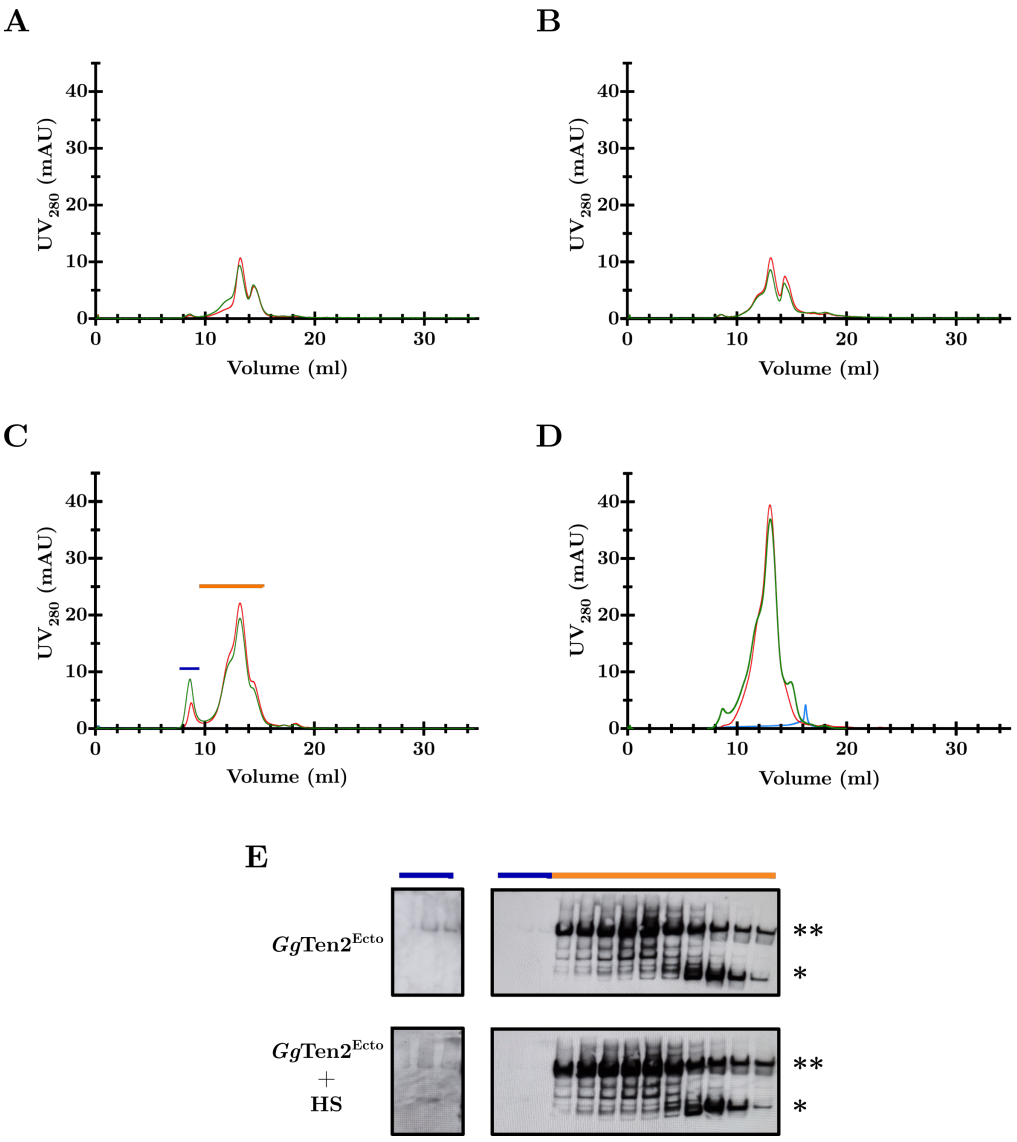


Figure 4.10: Multimerisation of *Gg*Ten2^{Ecto} on a long HS chain. *Red* denotes protein run alone through the SEC pipeline, and *green* represents the same protein incubated with 25x molar excess of HS. The glycan peak (*blue*) can only be seen at the highest concentrations. **A.** 150 μg *Gg*Ten2^{Ecto}, and the same protein incubated with 25x molar excess of HS. **B.** 200 μg *Gg*Ten2^{Ecto}, and in complex with the glycan. **C.** 500 μg *Gg*Ten2^{Ecto} incubated with 25x molar excess HS. **D.** 750 μg *Gg*Ten2^{Ecto} and the same amount of protein complexed with HS. **E.** Western blots of the 500 μg protein samples, with blot fractions represented by bars, *blue* represents potential aggregate protein, and *orange* the purified protein running at a volume expected of protein alone. * and ** denote the sizes of monomeric and dimeric *Gg*Ten2^{Ecto}, respectively. An insert of the aggregate peak from longer developed western blots is shown for clarity.

indicating that the heparin chain was either too short to allow multimerisation, or that the interactions themselves to do not form along long glycans.

5

Cellular Effects of Teneurin-Glycan Binding

5.1 Summary

This chapter develops on the work previously presented, investigating the cellular effects of glycan binding by migration assay, whilst combining the binding mutants characterised in Chapter 4. These mutants were cloned in full length teneurin constructs to begin to investigate the affect of glycan binding on cell migration. A dot assay was developed, similar to the conventional stripe migration assay, which utilised the biotinylated glycan samples used in the previous chapter. These glycans were immobilised on the culture plate surface and transfected cells seeded ontop. The results of these assays, both with HEK293 and N2A cells were not conclusive. Two way ANOVA analysis showed a non statistically significant effect of glycan binding on migration, and as such more work is needed to properly characterise the effects of glycan binding on cell migration.

5.2 *Gg*Ten2^{Full} Teneurin Cell Surface Expression

Prior to investigating the in vitro cellular effects of teneurin–glycan complexes, test expression of full length teneurin, and the full length mutants created in this thesis, were performed. The methodology is described in Section 2.5.1, and was used for HEK293, Hela, and Neuro 2A (N2A) cell lines (Figure 5.1).

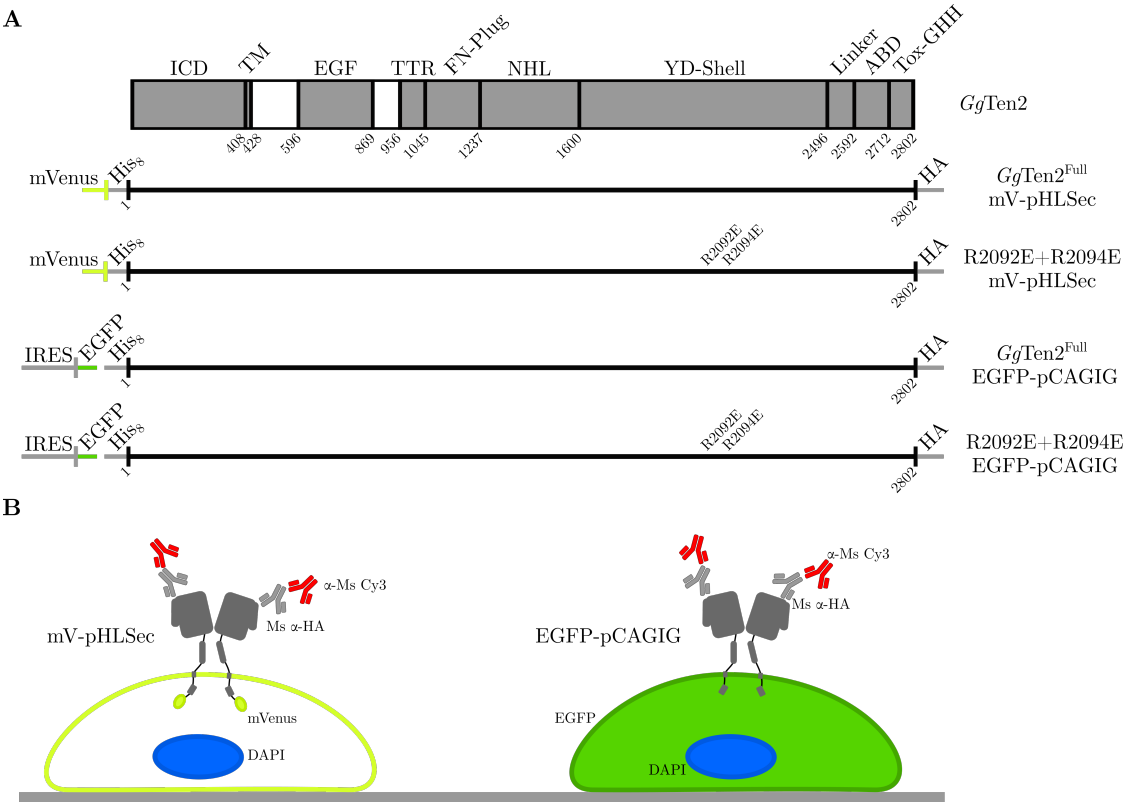


Figure 5.1: Full length teneurin constructs and Cell surface staining. A. The constructs cloned for in vitro assays used in this chapter, Wild type and mutant teneurin was cloned in pHLSec and pCAGIG plasmids which contained mVenus and EGFP for visualisation, respectively. **B.** Schematic of the cell surface staining setup. Red signal was produced by the antibody sandwich using Cy3 2° antibody (*red*), Green signal was produced by mVenus or EGFP, DAPI stained the cell nucleus blue.

This preliminary test showed that each of the constructs produced expressed successfully, as confirmed by both binding of the antibody sandwich resulting in a red fluorescent signal, and the green signal resulting from the associated fluorescent protein included in their respective vectors. The inclusion of DAPI

staining highlighted the nucleus of each cell, and was used to qualitatively inform the extent of transfection compared to the total cells. Only a small population of the total cells were transfected in both the HEK293 and N2A cell populations (Figure 5.2), and unfortunately very weak signal was seen with the Hela cell testing and as such were not continued. It was therefore decided to perform preliminary migration assays with HEK cells, and then to progress to N2A neuronal cells once these were kindly donated by Prof. Esther Becker. As expression testing was used only as an initial visual check for presence of the full length teneurin, no quantification was performed.

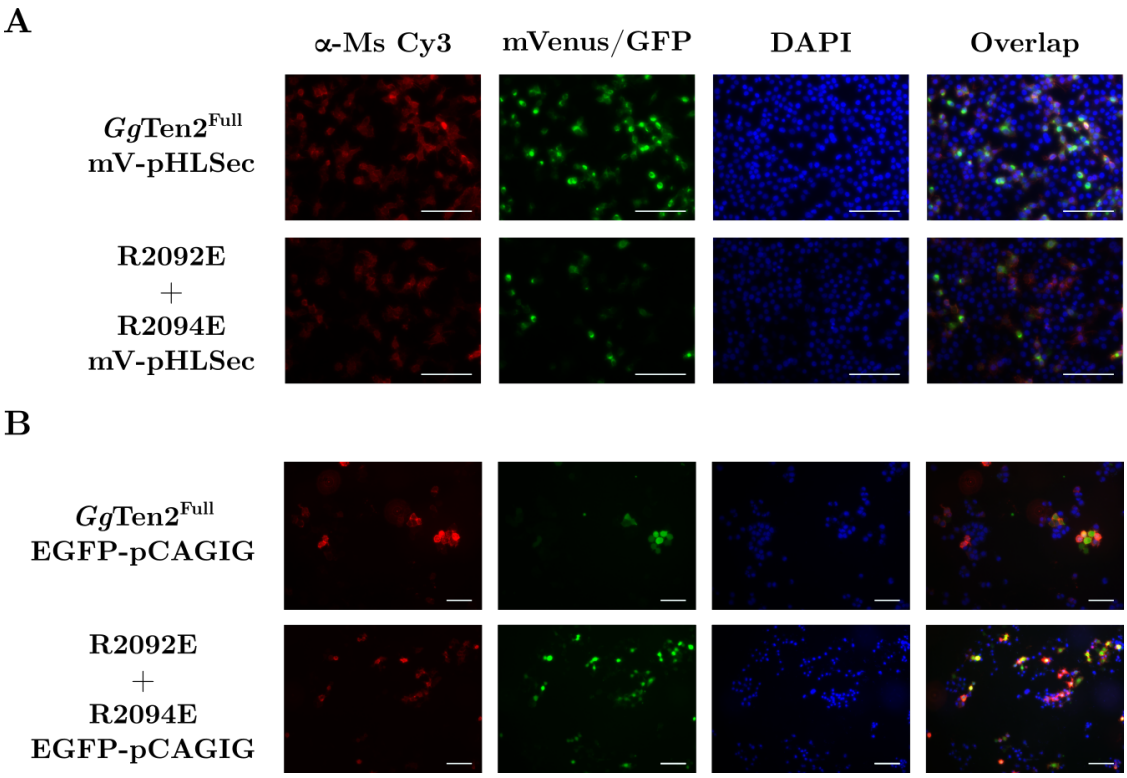


Figure 5.2: Full length teneurin expression testing. Representative expression testing of full length wild type and mutant teneurins used throughout the cell migration experiments in this chapter. Red signal is produced by a Cy3 antibody sandwich specific to teneurin, Green as a result of fluorescent mVenus or EGFP expression, and Blue from DAPI staining of the cell nuclei. **A.** Expression of the mVenus conjugated teneurin constructs in HEK293 cells. **B.** N2A cell expression of the EGFP and teneurin co-expression. Scale bar (*White*) = 100 μm.

5.3 Glycan Drop Assays

To investigate the effects of glycan binding on cell migration, a glycan drop assay was developed. This assay (described in detail in Section 2.5.2) worked similarly to the conventional stripe assay, however the immobilised ligand was instead added a a single drop on the culture plate surface. Specifically for the work described here, the biotinylated glycans used throughout SPR and pulldown experiments (Sections 2.4.2 and 4.4) were utilised and immobilised onto a Nunc tissue culture plate, by way of streptavidin binding. The tissue culture dish is treated with a proprietary Nunclon™ hydrophobic coating, which allows for passive adsorption of streptavidin onto the surface. For the work here, the streptavidin was pre-mixed with 5x molar excess of the biotinylated glycan of choice. This strep-glycan complex was allowed to adsorb onto the culture surface with added fluorescent antibody to allow for visualisation fo the glycan region. Transfected cells were then applied on top of the prepared surface and incubated for varying time points before being immobilised, stained, and imaged. A schematic of the assay set-up is shown in Figure 5.3.

5.3.1 HEK Cell Migration

Due to the high expression of the mVeus-teneurin-pHLSec construct seen with HEK293 cells above, this cell line was utilised as an initial test to begin to probe the effects of migration and to indicate a potential dose response with respect to a range of immobilised glycans.

To properly ascertain the concentration dependant manner of glycan migration affects, multiple series of immobilised glycans were tested. An initial amount of 1.00 μg (based on previous protein stripe assays performed in the lab) glycan was chosen to immobilise per 'dot', at 5x molar excess compared to the streptavidin. Alongside this, lower amounts of 0.10 and 0.01, and a higher 2.00 μg were tested, with their respective adjusted streptavidin amounts, each at three time points. Representative images (artificially adjusted to allow for easier visualisation by eye) are presented

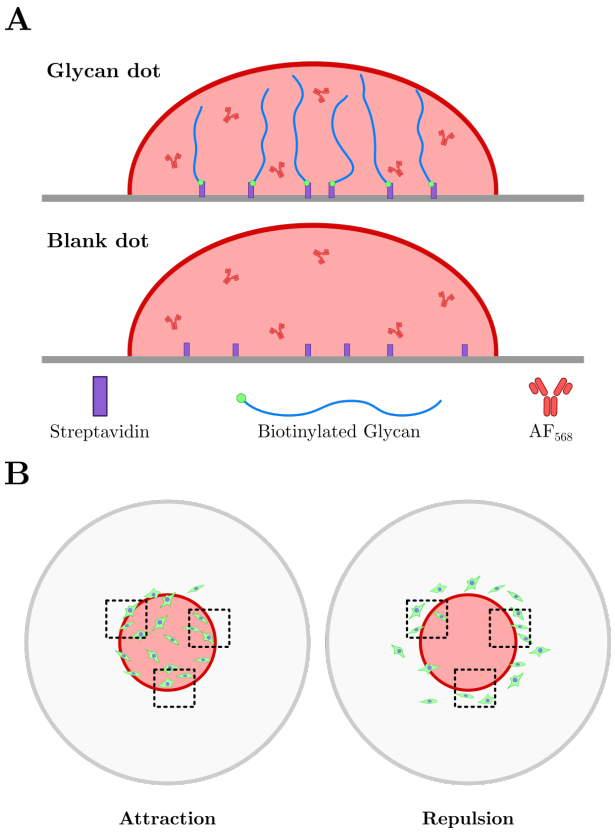


Figure 5.3: Schematic representation of the glycan dot assay. A. The assay was set up by pre mixing the streptavidin (*purple*) and biotinylated glycans (*blue*), along with the Alexa Fluor 568 used only to visulaise the glycan containing region. Blank dots were set up in the same way, without the addition of biotinylated glycan. **B.** Once the assay was incubated for the required time, the cells were fixed with PFA and imaged with a fluorescent microscope. Images were taken at the boundry between glycan environment and untreated culutre dish (*black square*), and cells were counted to distinguish between attractive (*left*) and repulsive (*right*) glycan environment.

in Figure 5.4.

For analysis, the glycan amounts are compared to their corresponding 'blank dots' containing the same fluorescent antibody and streptavidin amounts but without the glycan. The initial measurement at 4 hr was at approximately the time at which the cells are fully adhered to the assay surface, before a considerable amount of migration was able to take place. As well as acting as an initial time point, this could also inform whether cells are able to adhere to the glycan surface, as un-adhered cells would be washed off before the PFA application and subsequent

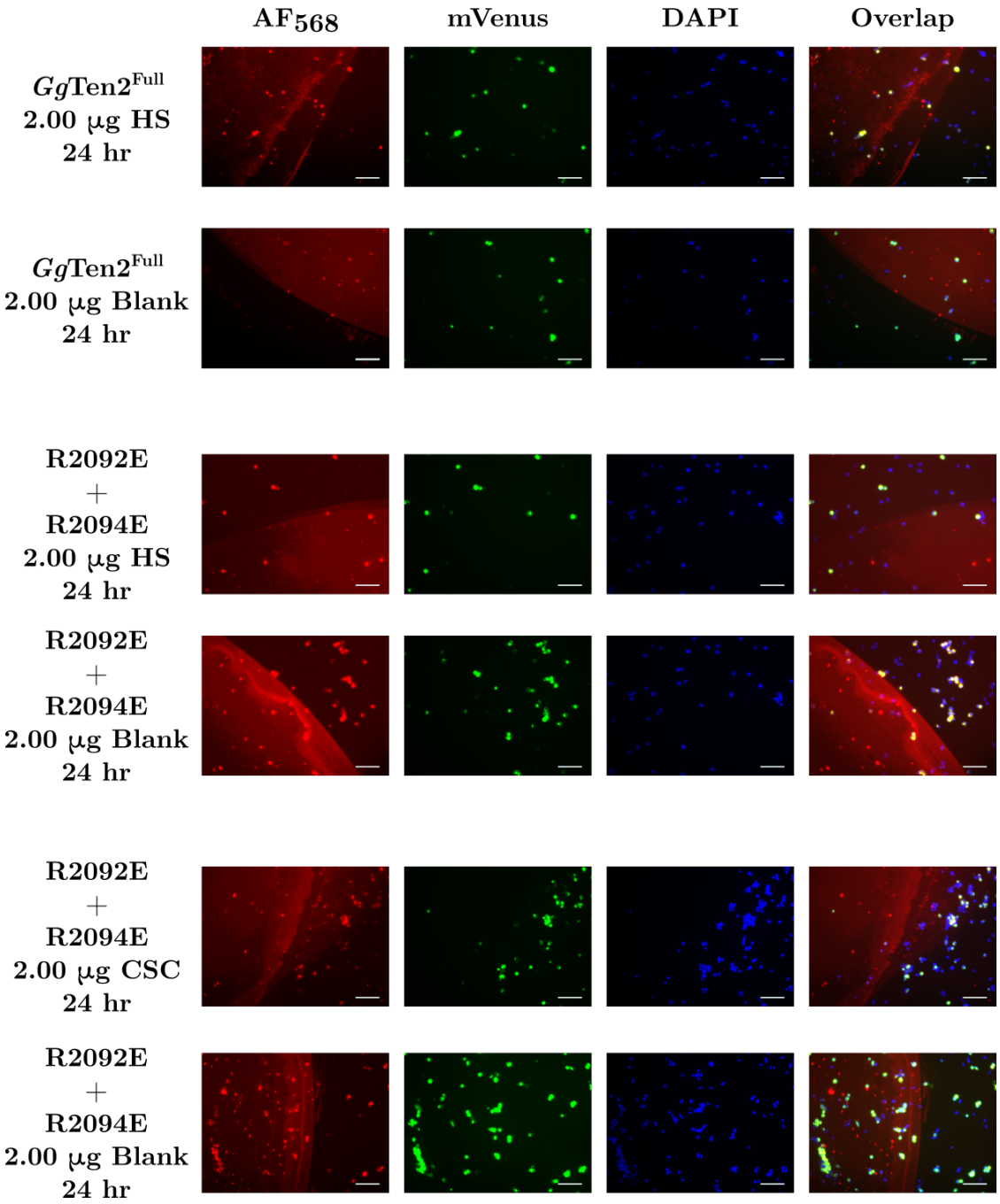


Figure 5.4: Example images from the HEK293 cell migration assay. Regions of immobilised glycan were visualised with the addition of Alexa Fluor™ 568 (AF_{568}) (*red*). Transfected cells can be identified by the conjugated mVenus-teneurin construct of interest (*green*). The nuclei were stained with DAPI (*blue*). An overlay of all three channels is shown for clarity. Image intensities have been adjusted for ease of visualisation but were not adjusted for analysis. Scale bar (*White*) = 100 μm.

staining steps. In all cases, there were no reduced cell numbers found on the glycan region compared to the blank coverslip. Images were also taken at 24 hr and 48 hr post incubation on the pre-prepared assay plate to assess more long term effects on migration. Analysis of the cell migration was performed by assessing the proportion of transfected cells that were present on the Alexa Fluor™ 568 stained glycan dot compared to that of the stained blank (non-glycan) repeat. The calculations were adjusted to take into account the area of each image containing the glycan region, compared to the untreated surface. As a result, the data is presented as relative migration as a proportion of the blank, where a value of 1.0 represents the same number of cells in the stained region of the glycan plate as the blank dot (Figure 5.5), and values above 1.0 indicate migration into the glycan region. It can be seen that migration is very similar between glycan condition and blank, in almost all cases. This indicates the glycans do not affect migration, except with the CS-C (6S) mutant data where a lack of glycan binding allows for the cell to be halted in a high glycan environment. This may be explained by teneurin playing a role in positive migration through the CS-C environment, but with its removal from the signalling, migration is reduced. However, it would therefore be expected that the wild type *Gg*Ten2^{Full} would show an opposing trend, where a high concentration is beneficial for migration, and a lower concentration inhibits the cell movement which is not seen here. Statistical ANOVA analysis of all data points showed that only two comparisons were statistically significant. The heparan sulfate immobilised sample at the highest concentration (2.0 μg) caused significantly increased migration compared to the corresponding 0.1 μg sample (P=0.0171) and 1.0 μg immobilised data (P=0.0498). This suggests that Heparan sulfate is a positive regulator of migration, and at high amounts its interaction with teneurin enchances cell motility.

5.3.2 N2A Cell Migration

Following the migration assay with HEK 293 cells, a new, more neuronal appropriate, cell type was used. These Neuro-2A (N2A) cells are a model cell line used for

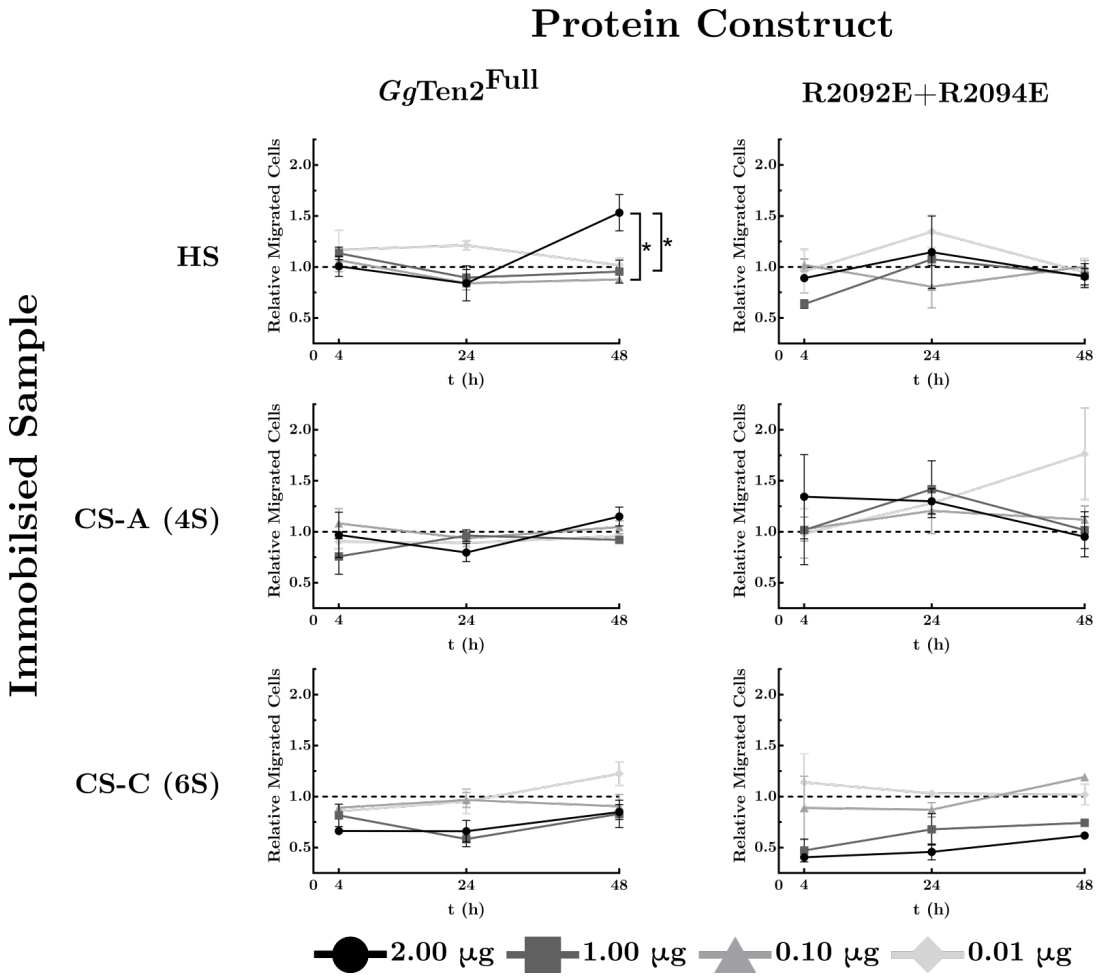


Figure 5.5: HEK cell glycan migration results. Migration was assessed on the glycan dot as a proportion of the blank, which did not contain the immobilised glycan. Four glycan amounts were immobilised (*below*), at 3 time points for both wild type *Gg*Ten2^{Full} and R2092E+R2094E transfected HEK293 cells. Error bars show SEM, n=3. Two-way ANOVA test with Tukey's post hoc analysis, $p > 0.05$.

investigating mouse brain as they have the capability to differentiate into neuronal cells over time. This cell line does not express any of the teneurin, latrophilin, or FLRT protein family members[206–208] and as such is an ideal candidate to investigate the role of teneurin on cell motility and migration. One characteristic of this cell line is their long processes, caused by the synthesis of large amounts of microtubules. These processes can be seen under low magnification microscopy (Figure 5.6) and are studied as a model for axonal growth in vitro. N2A cells are used in the study of neuronal outgrowth[209], neurodegenerative diseases such as Alzheimer's disease[210, 211], and neurotoxicity due to their susceptibility to viral infection[212, 213].

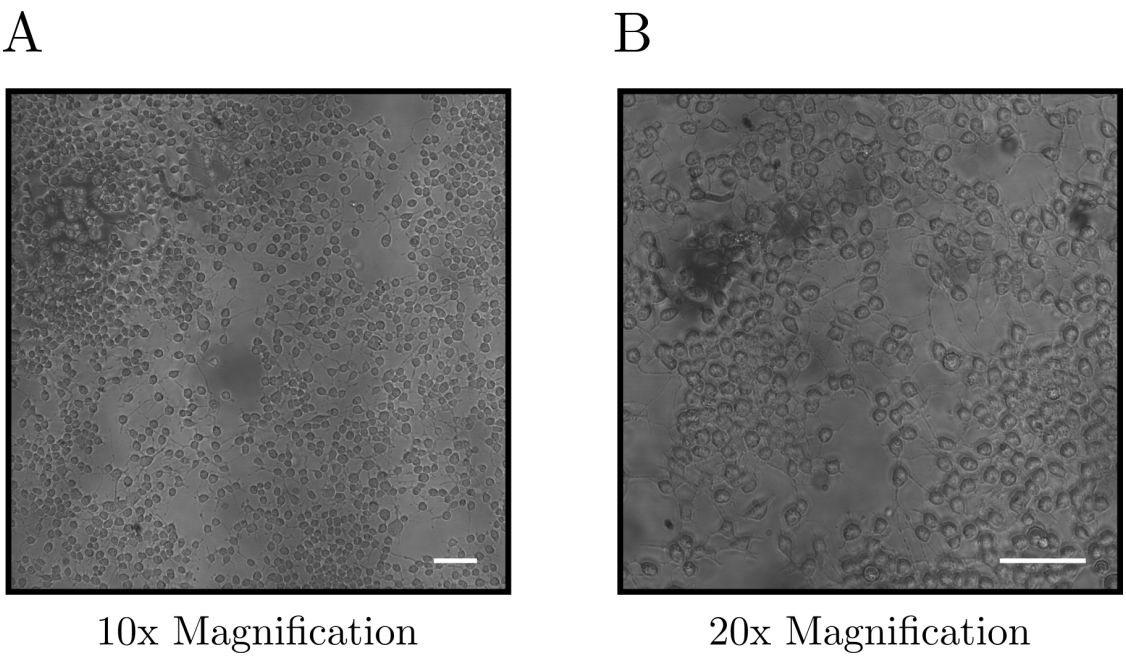


Figure 5.6: Light microscopy image of N2A cells. A. 10x magnification image of cultured N2A cells. **B.** 20x magnification image of N2A cells. The microtubule rich processes can be seen as thin extensions originating from the cell body. Scale bar (*White*) = 100 μm.

Migration was studied over a 24 hr period, with the cells set up in the same method as above. Due to the time restraints towards the end of my DPhil, only the highest glycan immobilisation amount was tested (2.0 μg). Three replicates were measured. The analysed data is shown in Figure 5.7, with migrated transfected cells on the glycan area compared to the un-transfected population. Two way ANOVA analysis was performed in Prism. The statistical analysis highlights that the difference in migration seen between each transfected population compared to their un-transfected population in each condition is not significant, as all comparisons were under the significance threshold of 95%. Therefore, no conclusions can be made from this data that the highest glycan dose affects N2A migration. This may require more repeats to eliminate random migration and overall noise in the data, a could be performed again with low passage number cells as the transfection efficiency was dramatically reduced with ageing cells. This is a common occurrence with all cell types and transfection reagents[214], however there was a very severe drop off in

transfection efficiency even within the cell ages used in this test, with a much higher transfection efficiency with the youngest (P5) compared to oldest (P15) cell populations. Repeating, or increasing the replicates with better behaved cells would ultimately improve the fidelity with which the effect of teneurin binding could be assessed. Representatives images from the assays are shown in Figure 5.8.

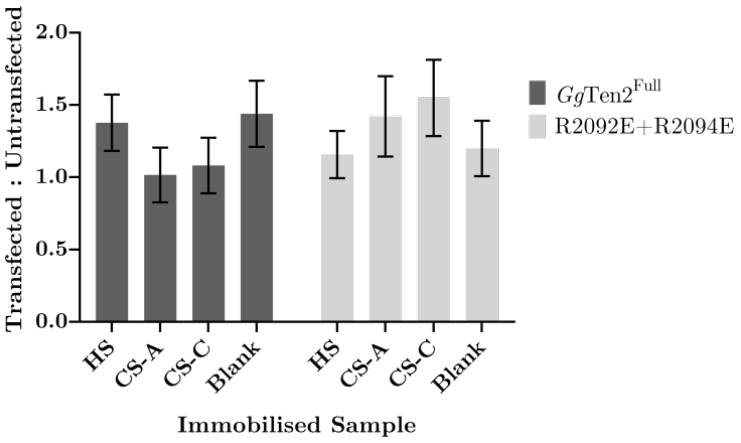


Figure 5.7: Graphical representation and analysis of N2A cell migration. All assay plates were immobilised with 2.0 μg of their respective glycan, except the blank. N2A cells were transfected with Ten2^{Full} or R2092E+R2094E. Migration was assessed as a comparison of transfected and un-trasnfected cell populations. Error bars show SEM, n=3. Two-way ANOVA test with Tukey's post hoc analysis, $p > 0.05$. Analysis showed that all data was NS, and as such conclusions cannot be drawn as to the effect of glycans on migration without more repeats.

Considering both migration assay results, conclusions can be made about the effects of glycan species on migration. Broadly it can be seen that the HS glycan species act as positive migration cues in the HEK cell system, an effect that is diminished with the N2A cells. This reduction in migration for teh more neuronal cell type may be as a result of differing endogenous protein binding partner expression. The CS species, specifically CS–C shows a reduced migration in both the HEK and N2A migration assays. Although not statistically significant in either populations, a reduction in the number of cells migrating into the glycan dot region can be seen from the raw images and analysed data. This suggests that the CS sugars are playing an inibitory role, potentially acting directly via teneurin intracellular

signalling, or by grouping inhibitory signalling players more closely to intensify the internal adhesive signal and diminish the ability to form transient migratiory adhesions. The flipping of this inhibitory effect with the N2A cells transfected with the R2092E+R2094E double mutant points to a function of teneurin specifically. The diminished potential for glycan binding with this construct coupling to an increase in migration above that of the blank sample suggests that in this case teneurin may be involved in positive signalling, which is reversed when CS is present and bound to the teneurin YD-shell, either by the first-hand teneurin signalling or its interaction with other signalling partners.

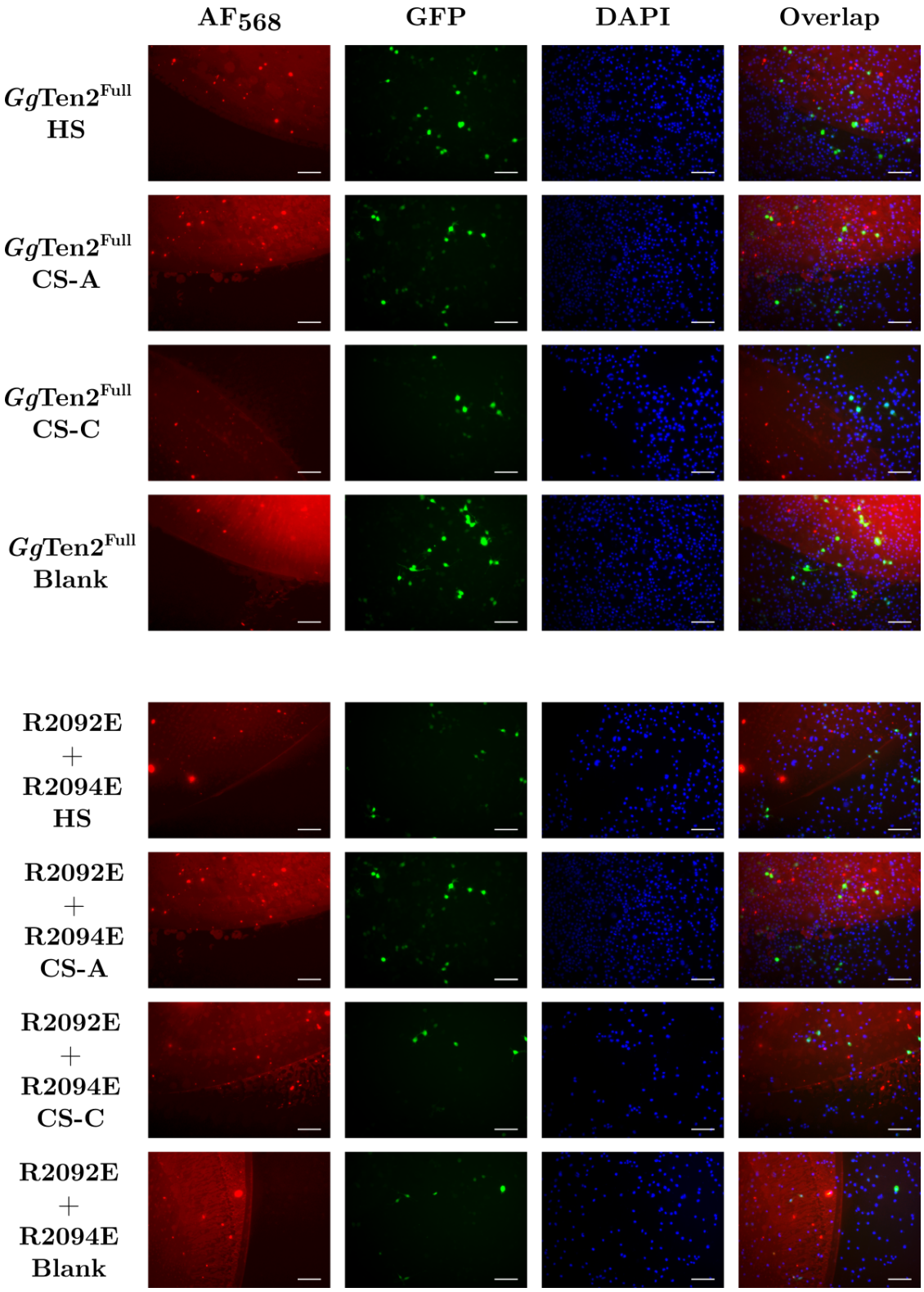


Figure 5.8: Example images from the N2A cell migration assay. Regions of immobilised glycan were visualised with the addition of Alexa Fluor™ 568 (AF_{568}) (*red*). Transfected cells can be identified by the GFP co-expressed alongside the teneurin construct of interest (*green*). The nuclei were stained with DAPI (*blue*). An overlay of all three channels is shown for clarity. Image intensities have been adjusted for ease of visualisation but were not adjusted for analysis. Scale bar (*White*) = 100 μm.

6 Discussion and Future Work

6.1 The Teneurin–Glycan complexes

The crystal structures elucidated in this thesis have proven for the first time that teneurin binds to extracellular glycan families, and does so via a charged patch on its YD-shell. Although a binding site had been predicted before[40], this work is the first structural proof of glycan binding. The 6 residues on the YD-shell; K2078, R2092, R2094, K2102, R2106, and R2116 interact with the charged sulfur groups of selected GAG species. This may suggest a promiscuity for the interaction, allowing teneurin to interact with other charged GAG species, although this has not been studied and would need to be confirmed with assays involving said glycans. As indicated by the kinetics and binding data in this thesis, there seems to be a preference of higher sulfation heparan sulfate (HS) glycan species over the chondroitin sulfate (CS) ligands. Specifically, the HS and CS pulldown assay (Section 4.4) indicated that the HS binding was significant compared to the blank, however both CS–A (4S) and CS–C (6S) did not have statistically higher amounts of teneurin protein retained in the column. Due to this inconclusive result, more rigorous testing would be required to confirm this preferential binding model, with reliable kinetics data allowing for more accurate comparisons of the differential binding between teneurin and glycan species.

The glycan binding mutants designed in this work undoubtable reduce binding when incubated with loose heparin beads, showing approximately 50% reduced binding capacity compared to wild type teneurin (Figure 4.2). The remaining background binding of teneurin to the heparin beads may be due to the presence of a second potential glycan binding site on teneurin, as predicted by the ClusPro binding server (Figure 3.17). This was discovered after mutagenesis and biophysical assays were performed, and as such the additional potential triplicate binding site proposed from the server (R1247, R1248, and R1249) was not mutated in the later work. There was no evidence for this binding site in the initial teneurin–SOS crystal structure. This potential second glycan binding site may be the cause of the latent binding still seen with the SOS binding site mutants, as well as the non-zero binding seen later with the highest concentration SPR runs, and biotinylated glycan pulldowns. If this binding site is genuine, it may be the cause of interrupted teneurin dimerisation. It is known that a splice site within a loop of the teneurin NHL domain is responsible for modulating teneurin trans-interactions, and also plays a role in the formation of cis-dimers of teneurin molecules originating from the same cell[22, 38]. This 7 residue string (^{1303}RNKEFKH^{1309}), lies directly adjacent to the predicted second heparin binding site (Figure 6.1). Although there are hitherto no structural trans-dimeric teneurin solutions, it is known that this region is important, and as such prior binding of a longer glycan chain (longer than the tetrasaccharide used in the ClusPro prediction) may play a role in the formation of the large homodimer. The longer constructs used in this section are able to form cis-dimers owing to the free cysteine residues in the upstream EGF repeats, which form two disulphide bonds connecting the two teneurin molecules. However the apparent lack of further teneurin molecules on the long heparan sulfate chain indicates either a lack of space along the sugar, or blocking of a potential dimerisation region, adjacent to the second predicted glycan binding site. Interestingly, Ten-1 isoforms, which do not contain this potential binding site in the NHL domain, have been shown to be unable to form trans-dimers in aggregation assays performed by collaborators

[*Unpublished data*].

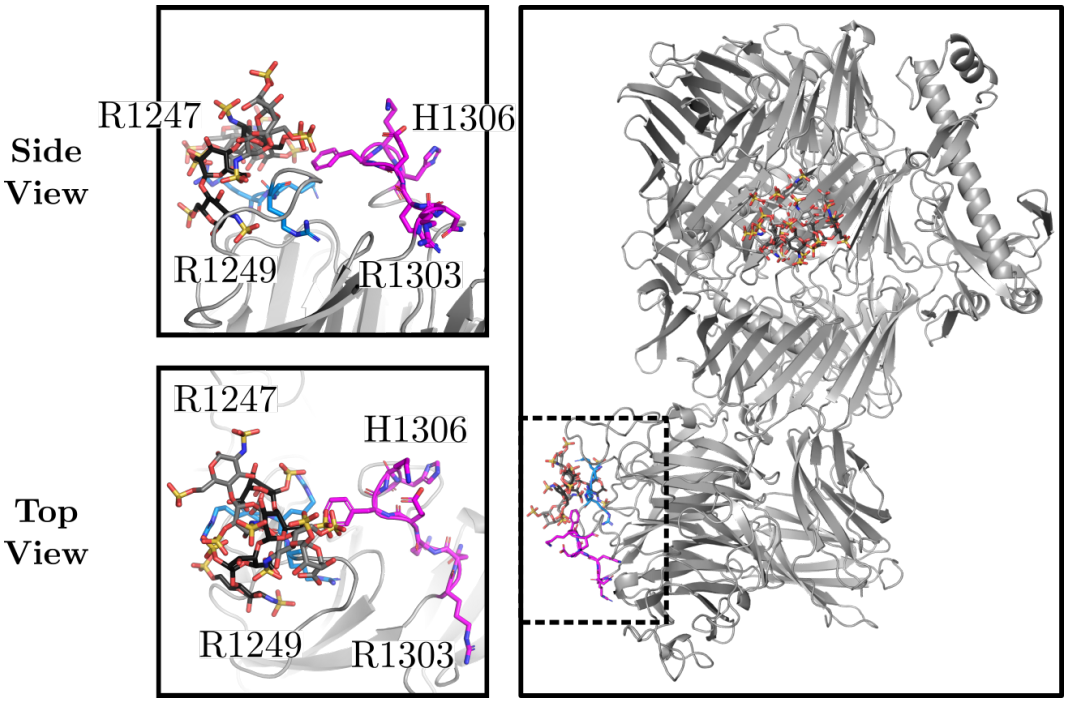


Figure 6.1: The predicted HS binding site lies adjacent to the teneurin splice site. The HS binding site predicted by the ClusPRO binding server (R1247–R1249) (*blue*) is located next to the NHL splice site (R1303–H1309) (*purple*) which is known to modulate teneurin–teneurin trans interactions. This may implicate glycan binding in the ability for teneurin to form higher order structures.

Conservational analysis of teneurin sequences was performed comparing 29 total teneurin sequences (Appendix C Table C.1). Due to the high sequence homology between vertebrate teneurin orthologues, sequences of all four teneurin isoforms were used where possible. Analysis of the teneurin residues reveals that the original binding site revealed from the SOS complex structure is very highly conserved among all vertebrate teneurin homologues, and invertebrate species. Of the 6 residues involved in binding, 5 are 100% conserved amongst the mammalian organisms tested (Figure 6.2). The only outlier is K2102, which is mutated to R in all but the mammalian Ten-2 isoforms (Appendix C Figure C.1). However this K–R mutation retains the charged nature of the binding pocket and as such glycan binding, if it is present in other teneurin isoforms, would be expected to

be retained. The invertebrate *Drosophila melanogaster* Ten-m protein shows a moderately conserved charged patch, with 3 residues containing positive Arginine or Lysine amino acids. However, 2 of the remaining 3 residues are glutamic acid, which would electrostatically repel the sulfur groups from GAGs. The second Drosophila teneurin isoform, Ten-a, contains a single positively charged residue in this patch (K2254) flanked by mainly aspartic and glutamic acid residues. These opposite charges between drosophila teneurin isoforms could point to segregated functions for the two proteins, with Ten-a functioning without glycan binding, and the Ten-m evolving to develop glycan interactions. This could function that is then retained and evolved further without more complex organisms, which evolve multiple glycan binding teneurin isoforms to account for their increase in molecular complexity. The nematode model organism *Caenorhabditis elegans* shows no conservation of the binding site, with R2031 the only charged residue in the pocket.

The second binding site made up of the three arginine residues R1247, R1248 and R1249, are conserved in all mammalian teneurin isoforms, apart from teneurin-1 (Appendix C Figure C.1), which is unable to form trans-dimers irrespective of the presence of the adjacent splice site (Figure 6.1). This points to a potential importance for these three residues and the resulting glycan binding. Although not seen in the mutimerisation experiment carried out as part of this thesis (Section 4.5), it may be the case that the extracellular glycans are important modulators of teneurin-teneurin interactions across cells in mammalian systems. Where glycan binding is not possible due to the mutation of this splice site glycan binding triad (RRR–HQR in Ten-1), there is an accompanied loss of trans interaction. However, the same triad of residues (HQR) is seen in Ten-m Drosophila teneurin which is able to form trans dimers[45, 215, 216], alongside Ten-A which contains a LQR motif. The *C. elegans* teneurin protein (containing the residues VKR) are able to form dimeric teneurin. This illustrates the potential for other factors to instead control the ability of teneurin to form trans-dimers.

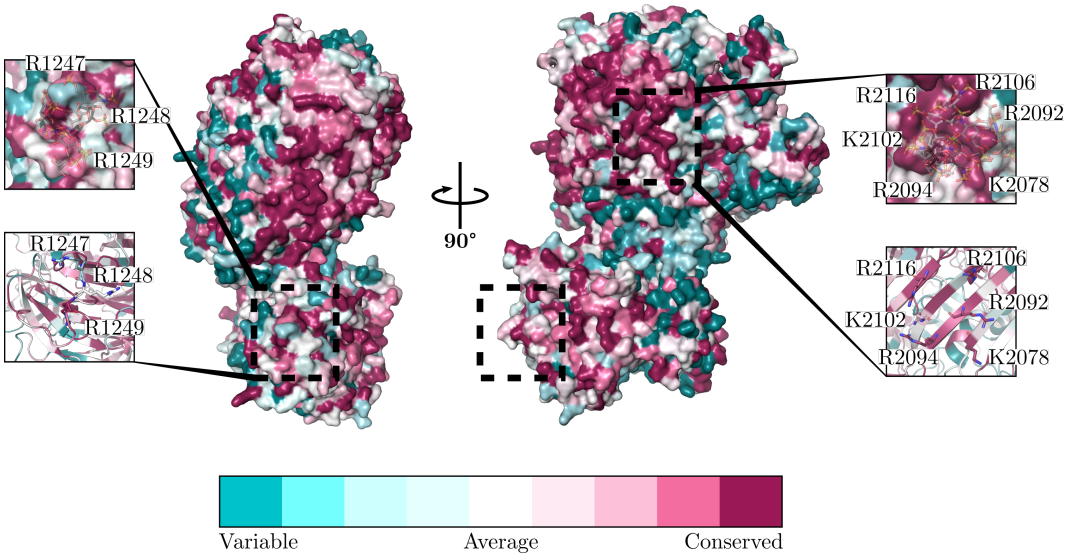


Figure 6.2: Conservational analysis of the teneurin sequence. Conservation was performed with 29 total teneurin sequences including all four teneurin isoforms where possible (Appendix C Fig C.1). Analysis shows the original binding site (K2078, R2092, R2094, K2102, R2106, R2116) shows high sequence conservation among vertebrates. The second binding site predicted by ClusPRO (R1247, R1248, R1249) shows poor conservation in comparison.

One of the key questions still to be answered in relation to the work investigated in this thesis is the origins of the interacting glycans. Both heparan sulfate and chondroitin sulfate, not to mention other potential GAG species that haven't been tested here but may interact with teneurin, are expressed on the cell surface as a proteoglycan, that is attached to a core polypeptide. Of the known protein binding partners of teneurin, none are known to harbour their own GAG chains on their surface. However, it is known from work produced in our lab, that FLRT2, a binder to teneurin in the inhibitory complex used in neuronal migration, is also able to bind sucrose octasulfate [*unpublished data*]. It is also known to bind to the heparan sulfate chains supplied by neurexin[197], and that another member of this family of proteins, LRRTM4, can bind to the HS chains supplied by glypican. These glycans are in-fact required for synapse formation[203, 204]. However, the addition of teneurin into this interactome has not yet been studied. The glycans may be playing a role in the recruitment of protein binding parters to the locale, increasing the density of signalling in defined areas of the membrane. This is a common role of glycans

with other signalling proteins, however has not yet been studied with a focus on the teneurin–lphn–FLRT complex and its activity.

6.2 Glycan Binding Candidate Screening

One of the main difficulties in studying the binding of glycosaminoglycan (GAG) interacting proteins is the potential preferential binding that can be seen between glycan species. There are multiple reports of glycan binding investigation where binding is only detected for one of the specific family, or even specificity within a glycan family[217, 218]. The work described here centres around the most common GAG, heparan sulfate (HS), as an indicator of general binding potential. The simultaneous investigation of chondroitin sulfate (CS) was pursued due to the role of CS in brain development[90, 128] and axon regeneration[140, 144, 219]. Due to mostly time, and financial restraints, these were the only glycan families that could be investigated as part of this thesis. It is by no means a comprehensive characterisation of teneurin binding to extracellular glycans, and more work should be subsequently carried out to investigate the missed samples outside of the scope of this work. There is very little research available into the binding partners of the other glycan families (DS, KS, and HA). Therefore it would be interesting to investigate their potential as a lesser studied ligand, which could shed some light onto the mechanism behind some of their functions in the body.

The biggest limitations of the biophysical characterisation of the teneurin–glycan binding in this thesis is the lack of quantitative kinetic data. This is something that was hoped to be completed with the SPR experiments carried out in Section 4.3. However, due to the difficulties with this experiment it was unable to be achieved. The issues with the SPR experiment attempted here are difficult to dissect. Immobilisation of the glycans on the SPR chip was successful, indicating that they are correctly biotinylated. However, the CS–C sample did not stay attached to the gold surface, as the RU dropped whilst the protein samples were

prepared and run on the system. This was potentially due to poor biotinylation of the glycan, and as such it was gradually washed from the chip by the constant buffer flow. The quality of the biotinylated samples from Iduron could not be easily confirmed, and as such I relied solely on their in house quality control checks. Of the two glycans that did retain binding, there is the potential that the glycans are long enough to contain multiple binding site, and as such a $1:1$ binding mode is unlikely. In reality, it is likely that the teneurin protein can bing to multiple sites on a single long glycan chain, multiple proteins may be able to bind simultaneously to a single glycan, and the interaction is transient enough that multiple binding events can occur whilst a single protein flows over the chip. It may have been more appropriate to attempt the SPR experiments here with smaller glycan chains potentially using the disaccharide repeats for each glycan. However this was decided against due to the potential for steric clashes. With such a small molecule being immobilised and a large protein washing over and binding, there was a concern that the teneurin would not be able to orient itself in the correct binding orientation without clashing with either the chip surface, or by blocking the neighbouring ligands on the surface. Because of these multitude of potential issues, SPR has proven to be unsuitable for glycan binding investigation in this case.

There are a number of alternative methods that can be employed to investigate the kinetic binding of protein–macromolecule complexes. One of the most common approaches is Isothermal titration calorimetry (ITC). This is commonly used to study interactions between different sized macromolecules in their native aqueous environment, and has been used to study protein–glycan interactions previously[220–222]. The technique measures the temperature change between a sample and reference cell when the ligand of choice is introduced into the sample system. ITC is unique in the fact that it is one of the only techniques that does not require modifying the protein or ligand with tags or through immobilising, and can determine dissociation constant (K_D), stoichiometry, enthalpy (ΔH), and entropy (ΔS). Through sequential injections of the ligand of interest into the assay chamber, the temperature change

caused by binding (either exothermic or endothermic) is measured, and slowly plateaus as the system becomes more saturated with bound ligand, resulting in less of a response for new binding events. There are some concerns with respect to this method, chiefly that it is best suited to high affinity interactions, where the possibility of dissociation and re-association of a single protein molecule are minimised, however this limitation can be discounted with proper experimental design and sample preparation. The approach is also quite sample intensive, and it can be more difficult to measure non covalent interactions due to their low binding enthalpies.

Another common biophysical approach used in conjunction or as a replacement for SPR is microscale thermophoresis (MST)[223, 224]. This approach also records a change in temperature, but unlike ITC, this is induced by an infrared laser shone at a fluorescently labelled sample. This microscopic temperature change in the sample environment changes two intrinsic properties of the fluorescent target molecule (either an additional tag[225] or native aromatic amino acids[226]), phenomena known as temperature related intensity change (TRIC) and thermophoresis. The TRIC response of a specific macromolecule is sensitive to its microenvironment, which changes upon ligand binding, and the thermophoresis effect is proportional to size, charge, and solvation entropy of the molecule, which differs between native and bound samples. Overall MST readings are plotted against ligand concentration to obtain a dose response, from which binding affinities (K_D) can be obtained. When relying on the inherent fluorescence of aromatic amino acids, there can be difficulty within MST to generate a signal above the noise. This is the reason for the, at-least historical, reliance on fluorescent modification of the sample. This can be perform on either the protein or ligand, and in the case of work described here would be most appropriate to label the teneurin protein. However there are then considerations about location of the fluorophore, as the change in microenvironment upon binding would necessitate the fluorescent group be close in proximity to the binding site, without hindering the association of the ligand. Heparan sulfate and other GAGs have been modified with fluorescent groups[227, 228], however I believe in the case of

the small chain glycans used in this kind of assay, the possibility of the fluorophore interfering with binding would be likely. Using a fluorescently labelled glycan ligand would also require the expertise and collaboration of pure chemical research groups, whereas the addition of fluorescent tags on the teneurin protein can, and has been, performed by the lab and in this thesis.

An interesting avenue for investigation, related more to the structural studies performed earlier in this thesis, as well as the multimerisation experiments performed at the end of Chapter 4, is the possibility of using teneurin–glycan complexes in negative staining and cryo-EM. This assay could be used as an alternative to SEC MALS to visually identify the presence of multiple teneurin molecules on a long glycan chain. This could shine a light on the minimum length of teneurin required for mutlimeristiaon by potentially allowing higher order multimers to attach to longer chains and not those of an incorrect length. There could also be investigations into the glycan binding promiscuity of teneurin with this approach. This approach could be used to structurally confirm the presence and orientation of potential cis- and trans-multimer arrangements of teneurin, as have been proposed thus far without structural insight. The ability for proteins to multimerise on increasing lengths of glycan chain has been studied previously[86]. And has been studied extensively with respect to Wnt–Hedgehog signalling where the glycan chains have been shown to aggregate the signalling molecule, and expression of the HSPG responsible for this overlaps with Hedgehog expression[229, 230], an effect which is itself modulated by the composition of the HS glycan chains presented in the ECM[231]. These interactions have not been studied directly by cryo-EM, potentially due to the lower affinity interactions seen between many protein–glycan complexes, and this potential incompatibility with the approach.

6.3 In Vivo Effects of Glycan Complexes

Investigating the cellular effects of glycan binding has formed the final results chapter in this thesis, and as such more work is definitely needed to fully investigate the effects. Due to time constraints during my DPhil, a full investigation of the cellular effects was not able to be conducted. Nonetheless, the results presented here form a good basis for explaining the relationship between teneurin and cell migration in the developing brain, alongside the existing evidence of protein complex formation. Glycan binding has been shown here to affect both HEK 293 cell movement, as well as the neuronal N2A cell line.

The results presented here, although statistically insignificant from each other with respect to glycan species, indicate a potential effect of glycan binding on cell migration. The initial HEK cell migration assay (Section 5.3.1) indicated a fairly neutral migratory response to glycan binding with both HS and CS–A immobilised glycans. The only effect seen was that with CS–C. With this specific glycan, there seemed to be a negative migratory effect when the cells came into contact with the glycan compared to the blank this was more visible at the higher immobilised glycan level than the lower, indicating an inhibitory effect of migration. This somewhat follows the published literature, as CS as a whole is known to be inhibitory to migration, however I would expect to see a similar result with the CS–A (4S) assay, as this is known to be the more inhibitory sub-species of CS glycan compared to the CS–C (6S) form[147].

The N2A migration assay (Section 5.3.2) again shows a fairly neutral effect of glycan binding on migration. However it is clear that there is some effect of CS binding for this cell line. The wild type *Gg*Ten2^{Full} transfected cells do migrate less favourably into the chondroitin sulfate containing region of the assay area compared to the blank, in both cases. The heparan sulfate assay population shows a near identical migration to that of the blank, indicating that the immobilised HS glycan is not having an effect of increasing migration to that environment, even though binding is confirmed by both the crystallography and biophysical assays performed previously. The double mutant (R2092E+R2094E) transfected N2A

cells exhibit a mirrored migration to the wild type populations. Both the blank and HS populations are again close to identical, however the CS containing assay populations actually show increased migration compared to the untransfected cells in the population. This indicates that teneurin specifically is having an effect on migration, as the disrupted binding seen with these constructs results in a recovered, and actually improved, migration. This may point to a deeper function of glycan binding to the teneurin extracellular domain, where abolishing this interaction also disrupts further inhibitory processes that are normally in place. For example, the native role of teneurin with respect to migration is to form a repulsive complex with its two main protein binding partners, Lphn and FLRT[36]. When glycan binding is disrupted, it may destabilise this native complex, and therefore increase migration above and beyond the natural levels. This could have knock on effects with respect to synaptogenesis, which requires precise timing of the respective cells to form correct connections. The multiple isoforms of teneurin (Ten-1–4) in an organism, coupled with the multiple glycan species, sub-species, and spatio-temporal expression patterns gives rise to a complex system with more fidelity than can be assayed in vitro.

The work here is predicated on using the correct amount of immobilised glycan to see an effect on migration. This proved to be a difficult range to find, as the amounts of glycan in the developing brain and in specific in vivo environments is difficult to study, and incredibly dynamic. The amounts of immobilised glycan used throughout this thesis were based on previous immobilisation experiments performed in the lab, where protein binding partners were attached to the tissue culture surface. In these cases an equivalent of 1.0 μg of protein was assessed, and as such this was the target for glycan immobilisation. Higher and lower concentrations were tested to see if there was an effect when varying the amount of glycan present in each case. Immobilisation amounts higher than the 2.0 μg used in this work would have been preferable, however the price to custom order this product from Iduron, as well as the time constraints for the lab work presented here unfortunately

prohibited its exploration.

Many methods have been used to measure GAG composition directly from biological samples, including high performance liquid chromatography (HPLC), mass spectrometry (MS), tandem MS (MS/MS) and fluorophore-assisted capillary electrophoresis (FACE)[232–236]. However these approaches require an abundance of sample and large quantities of biological tissue. Approaches with fixed tissues are reliant of immunohistochemcial (IHC) characterisation of CS disaccharides, however there is often cross reactivity between CS antibodies. These approaches are targeted at investigating the ratios of differing glycan species, for example the ratio changes of CS–A (4S) to CS–C (6S) in the developing brain[138, 139, 237], rather than the total amount of glycan present in a tissue. To my knowledge, there are no such studies, and the scientific consensus is that glycans create a dense net of sugar chains, differing in their sulfation composition rather than total expression. These nets are in near constant interaction with expressed proteins harbouring a binding motif. Evidence of overlapping expression of teneurin and the key GAG polymerisation and modifying enzymes (sulfotransferases) would act to confirm the likely interaction between teneurin and glycosaminoglycans. The relative fluctuations of glycan species during development requires expression of specific modifying enzymes, the sulfotransferases, with specific enzymes responsible for specific modifications of the glycan backbone. The Allen brain atlas[238, 239] is an invaluable tool for investigating overlapping expression patterns, with extensive expression studies looking at the developing mouse brain. Unfortunately, this is not an exhaustive list of gene expression, and in this case only one of the teneurin isoforms (Ten-3) has been mapped in the developing mouse brain. Alongside this, the sulfotransferase enzymes responsible for adding specific sulfur groups to create the distinct glycan sub species have also not been mapped to the developing mouse brain. The Ndst4 gene, responsible for the expression of the N-deacetylase/N-sulfotransferase (heparin glucosaminyl) 4 enzyme, has been mapped. This bifunctional enzyme removes an acetyl group from the glycosamine (GlcNAc) sugar, and adds a sulphur group in its

place. This enzyme is therefore essential for correct heparan sulfate biosynthesis, and is a marker for increased expression of the glycan in the ECM. Looking at E18.5 mouse brain in situ hybridisation (ISH) data from the Allen brain atlas, the expression patterns of teneurin-3 (Tenm3) and Ndst4 can be seen to overlap relatively poorly. Both genes are expressed at their highest level in the midbrain, however in distinct subregions of this brain structure. The teneurin gene being expressed almost ubiquitously (with highest expression in regions associated with the visual system) except the alar plate region of the midbrain (Figure 6.3A), where Ndst4 is expressed most intensely (Figure 6.3B). There is overlapping expression in much of the hindbrain, albeit at lower, less localised, levels. This expression pattern is not a complete picture of teneurin and GAG expression, as we know from other sources that the other teneurin isoforms (Tenm1, 2, 4) express in other portions of the developing mouse brain[35, 240]. As such there is teneurin present in the same regions as the HS modifying enzymes during development, as well as the likelihood that the GAG modifying enzymes have distinct expression patterns also. This may point to a more complex interaction system than simply all teneurins binding to all glycans, instead pointing to a network of interactions between differing teneurin isoforms and GAG species acutely and accurately interacting in the developing brain.

Another valuable database is the Single Cell Portal (SCP). This is an open source repository of individual studies that have investigated gene expression. One such study, and part of the SCP, investigated gene expression in the developing mouse neocortex[241]. This study used single-cell RNA sequencing over the entire embryonic and into postnatal development periods to track gene expression in all cell types of the developing cerebral cortex. The data collected in this study can be used here to investigate expression levels of teneurin across embryonic brain development stages, as well as glycan synthesis and modifying genes. Using this database, the teneurin genes (Tenm1–4) could be assessed, as well as a wide range of enzymes involved in glycan synthesis and modification. A total of 20 Heparin/HS interacting genes were identified, as well as 19 genes involved in chondroitin sulfate generation

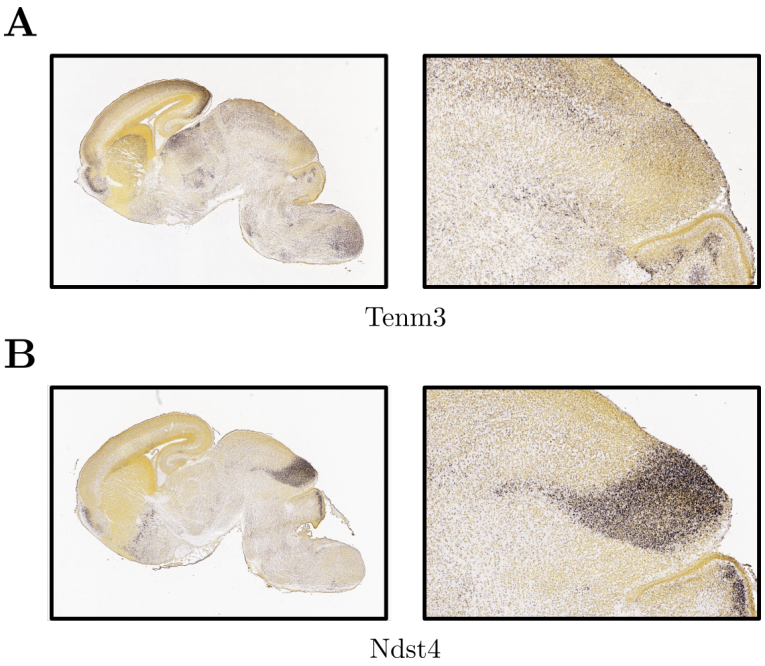


Figure 6.3: E18.5 mouse developing brain expression of Tenm3 and Ndst4 from the Allen brain atlas. A. Expression of teneurin-3 (Ten3) is ubiquotuois throughhout the brain, with highest expression in the regions associated with visual system organisation. **B.** Ndst4 is less highly expressed throughout the brain and has a concentrated patch in the alar plate.

(Figure 6.4). Gene expression was compared from embryonic day (E)10 to postnatal day (P)4. The teneurin expression can be seen to switch from mostly Tenm4 in the early stages of development, to a more evenly dispersed expression across the three other isoforms during development. The HS modifying enzymes are generally widely expressed throughout the stages of development, however there are obvious phases of expression of specific genes. For example, the HS sulfotransferase gene Hs2st1, responsible for the transfer of a sulfur group to the C2 position of the Iduronic acid backbone, is expressed relatively early, around day E11. This expression wanes however, and other enzymes, specifically Hs3st1 and Hs6st2, increase expression in later embryonic and postnatal time points. This could point to the generation of specifically sulfated HS glycans at different developmental points, which could influence potential signalling pathways by modulating binding through the differentially sulfated glycan chains.

The prevalence of many of the chondroitin sulfate synthesising and modifying genes

seems to occur at later stages of development compared to their HS associated counterparts. The maximal expression of the 4-*O*-sulfotransferase genes (Chst8–11), responsible for the generation of CS–A (4S), occurs broadly later during development. This may be a cause of neuronal plasticity, as this sub-class of chondroitin sulfate has been shown previously to be more inhibitory than other forms, and has been studied during development where its prevalence grew during late embryonic time points and peaked during adulthood[138, 139]. The enzymes responsible for CS–C (6S) production (Chs1–7), are also most intensely expressed postnatally, where Chst1 and Chst2 show maximal prevalence. However, the other sulfotransferase enzymes seem to be expressed in individual patters. Such that at every time point there is 6-O transferase activity. This multifaceted approach and apparent redundancy in the generation of specific CS sub species must have a function within brain development, potentially acting as feedback for signalling responses, or as a potential safeguard for such an essential role for these glycans.

The expression of specific teneurin isoforms seem to match with specific groups of HS/CS modifying enzymes. For example, early expression of Hs2st1 and Hs3st1 enzymes responsible for location specific sulfation of the sugar backbone corresponds with maximal relative expression of Tenm4, indicating a potential link between this specific teneurin isoform and the heparan sulfate chain being produced at the same stage of development. The Hs6st2 enzyme is expressed more broadly over a later embryonic stage and after birth, which also corresponds to the expression of Tenm1–3. These differentially sulfated HS chains are all thought to bind to the teneurin surface irrespective of their sulfur group position, and it may be the proteins that are presenting these glycans into the ECM that play a role in modulating the effects of binding through their signalling pathways.

6.4 Future Functional Assays

An immediate improvement to the glycan drop assay used in this thesis would be the utilisation of the more conventional ‘stripe assay’[242]. This relies on the same

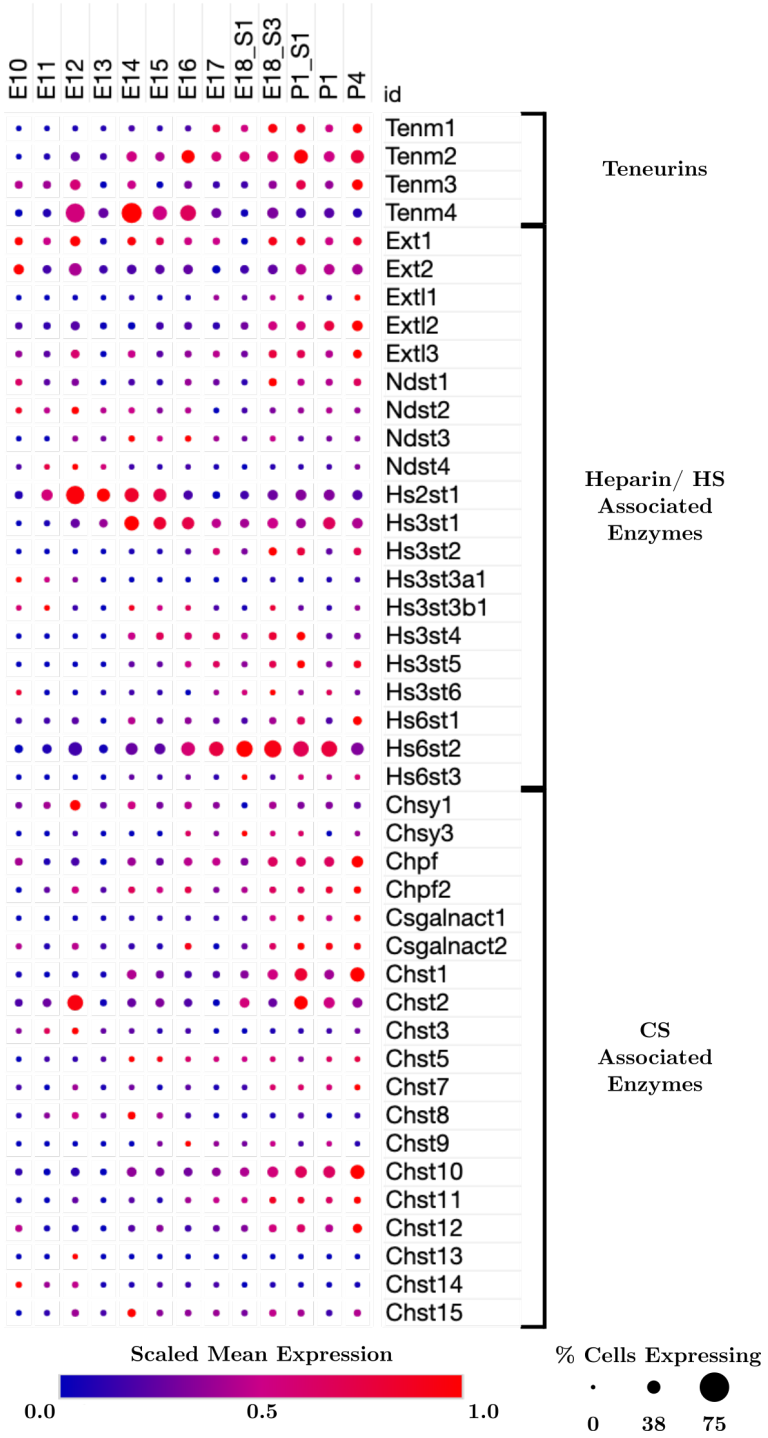


Figure 6.4: Single-cell RNA seq atlas of teneurin and glycan generating enzymes from the developing mouse neocortex. The heat-map shows normalised expression level from 0 *blue* to 1.0 *red*. The size of each entry is proportional to the percentage of all cells expression the gene. Data generated from[241].

principle: immobilising the ligand of interest, in this case biotinylated glycans, on the cell culture surface, before seeding the transfected cells on-top. However, the big difference between this assay and the glycan drop assay developed for this thesis is the size of the immobilised–blank region. The stripe matrices contain alternating 90 μm wide channels that allow for much smaller alternating environments compared to the dot assay. This allows the incubated cell to sense both environments, and migrate depending on cues from the two regions[243]. It also allows for the cell body to adhere to one stripe, and potentially extend protrusions into the second condition, rather than only being surrounded by one environment. This assay requires no further preparation outside of that already completed for the glycan assays already performed. The assays performed as part of this work were completed at the end of my time in the lab, after considerable efforts were made to ensure the full length teneurin construct and mutant expressed well in the required cell types. With more time, the stripe assay described here would have been performed as the next step to investigate the effect of glycan binding.

One additional consideration that could be used to better confirm the specificity of glycan binding and its effects would be the use of a negative control immobilised glycan. Using, for example, a biotinylated hyaluronic acid (HA) sample and immobilising it in the same way as the assay sample of interest, this could be used to better understand the effect of sulfation in general on the cell migration when bound, as well as giving an insight into the effects of other glycan species. This was not originally performed as part of the migration assays in this work due to both financial and time constraints.

The ultimate goal of investigating the cellular effects of glycan binding is to identify the function of teneurin–glycan binding with respect to neuronal migration and pathfinding in the developing brain, as well as synaptogenesis. To this end, more in depth assays can be utilised to directly measure the interaction in the developing brain. One of the most useful for the approach, and one that has been performed with collaborators previously[36], is in utero electroporation[244–246] (Figure 6.5A).

This is an approach that transfects a gene of interest into neural progenitors of specific brain areas of a mouse embryo[247, 248]. Electrodes are used to deliver the DNA into the cell, by inducing the formation of electropores. This approach has been used extensively in the field since it was first described in 2001, due to its reliability and relative simplicity, and has been used to target many different brain regions by fine tuning the electrode placement in the developing embryo. Of particular interest to the work carried out here is the ability to transfect fluorescently labelled protein, in this case wild type teneurin and the glycan binding mutants developed in this work. Neurons can be tracked by their fluorescent labelling, and the effect of in utero glycan binding can be studied with respect to neuron migration, speed, and destination location. The existing knowledge of teneurin in the developing brain, before synaptogenesis, describes how formation of the teneurin–Lphn–FLRT complex retards neuronal migration, slowing the speed at which the transfected cells move to their final destination and form synapses[36]. Assessing the role of glycan binding mutants, which contain mutations that are not thought to affect the teneurin–Lphn–FLRT complex directly, may result in affected migration, possibly resulting in the migrating neuron arriving at its destination earlier than expected. Due to the high prevalence of both HS and CS at the synapse, the lack of binding at this structure could destabilise the connection between cells and result in poor synapse formation.

There are also a number of assays which can be used to investigate the second interesting avenue of glycan binding research, axonal regeneration. As is well know, extracellular glycans define the inhibitory environments that are present post CNS injury[141, 144]. Specifically, the presence of chondroitin sulfate, and its transition of expression from the more permissive CS–C (6S) to the inhibitory CS–A (4S) is a current target for therapeutics that act to make the extracellular environment around an injury site more tolerant for migrating axons recovering through the injury. Recently, many treatments with chondroitinase enzymes (ChABC) have been utilised to degrade these inhibitory glycan chains following trauma[86, 131].

Neurite outgrowth assays can be used to investigate this phenomenon from the viewpoint of teneurin and its ability to bind glycans in a permissive or inhibitory environment. These assays are performed in much the same way as the glycan spot assay in this thesis and the stripe assays described above. However they can also be tuned to investigate neurite outgrowth by restricting the migration of the whole cell, instead only allowing the axons to migrate throughout the culture dish. Common microfluidic devices contain two compartments[249], one in which the cells are placed (somal side) and a second side (neuritic side) separated by microgrooves around 150 μm in length[250] (Figure 6.5B). These grooves are small enough for axons to grow through, but too thin for the whole cell to migrate to the other side of the chamber. Each half can be filled with culture media containing growth cues, or a gradient of cues across the chamber, and the amount of neurite outgrowth is measured[251]. This approach can give important information regarding the effect of specific cues, and can be used for both repulsive and attractive compounds of interest. These assays are often used with cultured neurons or explants, something that is not possible currently in out lab. However, the N2A cells used in this thesis may be of use here, as they produce long protrusions from their cell body, and are a common cell line used to investigate axonal growth[209].

6.5 Summary of Findings in this Thesis

The work presented in this thesis represents an initial investigation into the binding of teneurin and glycans, a topic that has only thus far begun to be studied. These initial investigations identified binding between the protein and glycan mimic discovering a binding site that had been proposed on the YD-shell previously[40]. The physiological ligands heparan sulphate, and chondroitin sulphate A and chondroitin sulphate C were also shown to bind via structural and biophysical approaches. I was able to disrupt this binding site by mutating the highly charged Arginine and Lysine residues in the binding pocket to Glutamic acid, as well as via N-linked glycan blocking, and proved that these mutants drastically reduced binding via pulldown

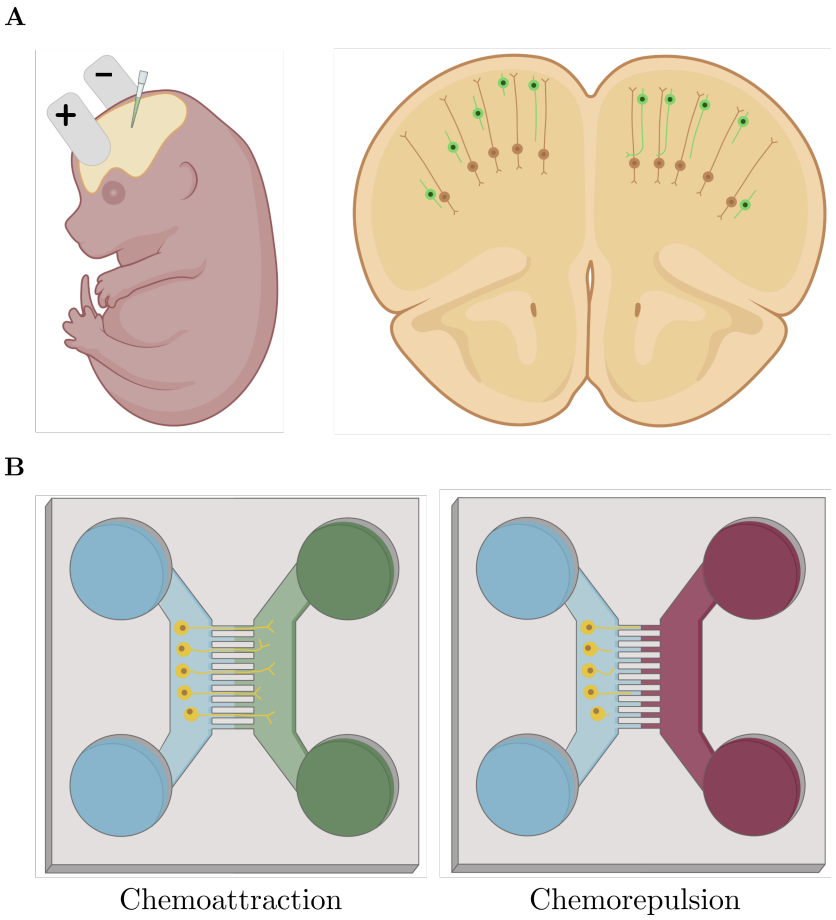


Figure 6.5: Chematic drawings of in utero electroporation and microfluidic experiments. A. In utero electroporation is performed by transfecting a specific region of the mouse embryo with a fluorescent target protein. The transfected neurons can then be tracked and migration studied. **B.** Microfluidic chambers can be used to assess axonal growth by allowing only neuronal axons to penetrate through small channels in the culture dish. Attractive or repulsive compounds can be supplemented in the culture media.

assays. Finally, the in vitro affects of teneurin–glycan complexes were investigated. Cell migration assays showed that these interactions do have a role in the cell, and that binding to specific glycan species can retard migration, whereas others seem to promote cell motility, mostly in agreement with the existing roles of these glycans. The immediate further work following from that described here would involve the continued characterisation of glycan binding, hopefully using other methods to ascertain the kinetics of binding. More work is needed to describe any potential glycan binding specificity, testing different families of glycans to understand if teneurin has any binding preference or if it binds promiscuously across all or

multiple glycan families. Finally, more research should be performed with respect to the investigation into the cellular effects of glycan binding. The cell migration assays used in the later part of this work serve as a crude example of the potential workflow that could be further investigated. An ultimate goal of inserting the glycan binding mutants produced here into the in utero developing brain would undoubtably give the best indication of the whole system effect.

Appendices

A

Constructs and Primers

Constructs

Table A.1: Protein constructs used in this thesis from previous sources. $A.A_N$, N-terminal amino acid; $A.A_C$, C-terminal amino acid; *Ref.*, Reference; *Gg*, Gallus gallus; *Mm*, Mus musculus; *His6*, hexahistidine.; *mV-pHLSec*, mVenus modified pHLSec; *HA*, Human influenza hemagglutinin.

Construct Name	Vector	Cloning Sites	Tags	$A.A_N$-$A.A_C$	Ref.
*Gg*Ten2$^{\Delta EGF}$	pHLSec	*AgeI* *KpnI*	$His6_C$	955 2802	[158]
*Gg*Ten2^{Ecto}	pHLSec	*AgeI* *KpnI*	$His6_N$ $Avi^{TM}Tag_N$	596 2802	[158]
*Gg*Ten2^{Full}	mV-pHLSec	*AgeI* *KpnI*	$His8_N$ HA_C	1 2802	[158]
*Mm*Lphn1^{Ecto}	pHLSec	*AgeI* *KpnI*	$His6_C$	1 829	[158]
*Mm*Lphn2^{Ecto}	pHLSec	*AgeI* *KpnI*	$His6_C$	1 841	[158]
*Mm*Lphn3^{Ecto}	pHLSec	*AgeI* *KpnI*	$His6_C$	1 936	[158]
*Mm*FLRT2^{LRR}	pHLSec	*AgeI* *KpnI*	$FLAG_N$ $His6_C$	1 362	[158]

Table A.2: Mutant constructs cloned as part of this thesis. Nucleotide polymorphisms are shown in *red*. $A.A_N$, N-terminal amino acid; $A.A_C$, C-terminal amino acid; *Dir.*, Direction; *Mm*, Mus musculus; *His6*, hexahistidine; *F*, Forward; *R*, Reverse; F_{Mut}, Forward mutagenesis primer; R_{Mut}, Reverse mutagenesis primer.

Construct Name	Mutation(s)	Vector	Tags	$A.A_N$-$A.A_C$	Dir.	Primer
R2078E	R2078E	pHLSec	$FLAG_N$ $His6_C$	955 2802	F R F_{Mut} R_{Mut}	GTAGCTGAAACCGGTGACTACAAGGACGACGACGACAAGGGTAGCCTTGTGTCTCTTATAAGAGGC GTGGTGCTTGGTACCCCTCTTTCCCATTTCATTCTGTCTTAAAAACTGG GATGAAACTACAGGTGTCCTAGAAATGGTG CCACTTTGCAAATTCACCATTTCTAGGAC
R2092E+R2094E	R2092E R2094E	pHLSec	$FLAG_N$ $His6_C$	955 2802	F R F_{Mut} R_{Mut}	GTAGCTGAAACCGGTGACTACAAGGACGACGACGACAAGGGTAGCCTTGTGTCTCTTATAAGAGGC GTGGTGCTTGGTACCCCTCTTTCCCATTTCATTCTGTCTTAAAAACTGG GAGGATTTTCTTGTACAATCGAGTATGAAAAAATTGG GTACAAGAAAATCCTCCACTTTGCAAATTC
K2102E+R2106E	K2102E R2106E	pHLSec	$FLAG_N$ $His6_C$	955 2802	F R F_{Mut} R_{Mut}	GTAGCTGAAACCGGTGACTACAAGGACGACGACGACAAGGGTAGCCTTGTGTCTCTTATAAGAGGC GTGGTGCTTGGTACCCCTCTTTCCCATTTCATTCTGTCTTAAAAACTGG GTAAAATTGGCCCTCTTGTTGACGAACAAATC CCATACCTTCTTCAGAGAATTCGTAGATTTGTTCGTCAAC
R2116E	R2116E	pHLSec	$FLAG_N$ $His6_C$	955 2802	F R F_{Mut} R_{Mut}	GTAGCTGAAACCGGTGACTACAAGGACGACGACGACAAGGGTAGCCTTGTGTCTCTTATAAGAGGC GTGGTGCTTGGTACCCCTCTTTCCCATTTCATTCTGTCTTAAAAACTGG GAAGAAGGTATGGTCAATGCAGAGTTTG GTGATATGTATAATCAAACTCTGCATTG
R2106N	R2106N	pHLSec	$FLAG_N$ $His6_C$	955 2802	F R F_{Mut} R_{Mut}	GTAGCTGAAACCGGTGACTACAAGGACGACGACGACAAGGGTAGCCTTGTGTCTCTTATAAGAGGC GTGGTGCTTGGTACCCCTCTTTCCCATTTCATTCTGTCTTAAAAACTGG CTTGTTGACAAACAAATCTACAACTTCTCTG GACCATACCTTCTTCAGAGAAGTTGTAG
*Mm*Lphn1 Linker		$pHLSec_{Avi}$	Myc_N Avi_C $His6_C$	378 452	F R	GTAGCTGAAACCGGTGAACAAAAACTCATCTCAGAAGAGGATCTGGGTCCAGATCCCAGTGCTGGCCCAGCCACTTCC CCACCGGAACCTCCGGTACCATGCAGATTGGGGGCTGGAGGCCTTCGGG
*Mm*Lphn2 Linker		$pHLSec_{Avi}$	Myc_N Avi_C $His6_C$	399 449	F R	GTAGCTGAAACCGGTGAACAAAAACTCATCTCAGAAGAGGATCTGGGTCCTGACCCTGCCCAAGTGCCTACCAC CCACCGGAACCTCCGGTACCCTTTGTCCCTCTGCTTCCGTCCTGAGGACC
*Mm*Lphn3 Linker		$pHLSec_{Avi}$	Myc_N Avi_C $His6_C$	462 559	F R	GTAGCTGAAACCGGTGAACAAAAACTCATCTCAGAAGAGGATCTGGGTCTGGACAGCAGATCTGGACCTGTGCATC CCACCGGAACCTCCGGTACCGATCTGAGAGGCGGCGCTAGGCAGATGG

B
Crystallographic Data

Table B.1: Privateer glycan validation for the *Gg*Ten2$^{\Delta EGF}$–SOS structure. *Q*, Total puckering amplitude (Å); *Phi* (0-360°); *Theta* (0-180°); *Conf*, Conformation; *RSCC*, Real space correlation coefficient; *Diag*, Diagnosis.

Chain	Name	Q	Phi	Theta	Anomer	D/L	Conf	RSCC	BFactor	Diag
A	NAG	0.572	64.7459	6.21293	beta	D	$^4C_1$	0.79	122.84	Ok
A	NAG	0.574	180.656	1.68495	beta	D	$^4C_1$	0.34	138.132	Ok
A	NAG	0.597	174.212	2.86437	beta	D	$^4C_1$	0.53	126.261	Ok
A	NAG	0.515	233.312	4.9355	beta	D	$^4C_1$	0.42	134.286	Ok
A	BMA	0.594	11.2733	12.0164	beta	D	$^4C_1$	0.57	145.983	Ok
A	MAN	0.568	316.984	4.46734	alpha	D	$^4C_1$	0.22	134.686	Ok
A	NAG	0.568	22.5496	11.6761	beta	D	$^4C_1$	0.67	80.275	Ok
A	NAG	0.576	46.0229	6.45142	beta	D	$^4C_1$	0.64	94.8921	Ok
A	NAG	0.588	353.423	5.81465	beta	D	$^4C_1$	0.57	109.915	Ok
A	NAG	0.616	191.043	1.83743	beta	D	$^4C_1$	0.46	139.408	Ok
A	NAG	0.555	182.801	1.64695	beta	D	$^4C_1$	0.48	136.114	Ok
A	BMA	0.575	44.2661	14.8817	beta	D	$^4C_1$	0.59	130.256	Ok
A	NAG	0.578	195.963	2.1556	beta	D	$^4C_1$	0.38	110.451	Ok
A	NAG	0.600	199.031	2.26359	beta	D	$^4C_1$	0.68	134.32	Ok
A	NAG	0.579	23.9218	9.58153	beta	D	$^4C_1$	0.28	172.759	Ok
A	NAG	0.517	190.384	1.58128	beta	D	$^4C_1$	0.54	141.569	Ok
A	NAG	0.567	174.84	1.28361	beta	D	$^4C_1$	0.60	146.971	Ok
A	BMA	0.501	166.274	1.35079	beta	D	$^4C_1$	0.64	150.415	Ok
A	MAN	0.632	168.733	1.68748	alpha	D	$^4C_1$	0.82	150.085	Ok

Chain	Name	Q	Phi	Theta	Anomer	D/L	Conf	RSCC	BFactor	Diag
A	MAN	0.559	181.019	1.78033	alpha	D	$^4C_1$	0.83	151.16	Ok
A	MAN	0.596	6.70466	2.0521	alpha	D	$^4C_1$	0.39	154.094	Ok
A	NAG	0.563	12.1988	12.6046	beta	D	$^4C_1$	0.73	107.516	Ok
A	NAG	0.575	4.58784	8.41715	beta	D	$^4C_1$	0.69	112.571	Ok
A	MAN	0.559	23.6256	9.68101	alpha	D	$^4C_1$	0.04	124.71	Ok
A	MAN	0.580	41.9374	5.35764	alpha	D	$^4C_1$	0.74	116.755	Ok
A	MAN	0.515	185.793	14.0285	alpha	D	$^4C_1$	0.38	155.062	Ok
A	MAN	0.597	186.356	1.49031	alpha	D	$^4C_1$	-0.2	167.781	Ok
A	BMA	0.574	63.7653	6.3243	beta	D	$^4C_1$	0.35	106.17	Ok
A	MAN	0.587	1.86214	5.4069	alpha	D	$^4C_1$	0.56	124.974	Ok
A	MAN	0.589	77.2101	5.39414	alpha	D	$^4C_1$	0.23	145.7	Ok
A	MAN	0.596	28.0305	12.7825	alpha	D	$^4C_1$	0.37	154.762	Ok
A	BMA	0.575	175.158	1.57337	beta	D	$^4C_1$	0.55	137.696	Ok
A	BMA	0.542	7.88627	13.4142	beta	D	$^4C_1$	0.69	127.519	Ok
A	MAN	0.565	109.442	10.2033	alpha	D	$^4C_1$	0.51	134.6	Ok
A	MAN	0.540	32.1981	13.6836	alpha	D	$^4C_1$	0.44	133.76	Ok
B	NAG	0.563	178.095	1.45957	beta	D	$^4C_1$	0.57	119.456	Ok
B	NAG	0.514	179.368	1.81842	beta	D	$^4C_1$	0.44	135.669	Ok
B	NAG	0.569	190.337	1.4823	beta	D	$^4C_1$	0.73	133.266	Ok
B	NAG	0.572	177.819	1.52697	beta	D	$^4C_1$	0.74	137.164	Ok
B	BMA	0.575	180.336	1.63839	beta	D	$^4C_1$	0.29	141.047	Ok
B	NAG	0.577	185.725	1.57127	beta	D	$^4C_1$	0.64	124.015	Ok
B	NAG	0.585	177.532	1.66852	beta	D	$^4C_1$	0.64	128.395	Ok
B	BMA	0.550	178.775	1.74442	beta	D	$^4C_1$	0.78	130.388	Ok
B	NAG	0.583	185.02	1.43514	beta	D	$^4C_1$	0.49	111.89	Ok
B	NAG	0.713	176.476	1.43263	beta	D	$^4C_1$	0.57	138.739	Ok
B	NAG	0.560	182.602	1.4994	beta	D	$^4C_1$	0.48	139.971	Ok
B	NAG	0.565	184.243	1.52873	beta	D	$^4C_1$	0.67	119.893	Ok
B	NAG	0.589	179.429	1.60732	beta	D	$^4C_1$	0.56	170.138	Ok
B	NAG	0.506	158.469	1.39175	beta	D	$^4C_1$	0.70	106.495	Ok
B	NAG	0.603	190.224	1.81266	beta	D	$^4C_1$	0.61	121.231	Ok
B	MAN	0.602	178.415	1.5897	alpha	D	$^4C_1$	0.20	142.635	Ok
B	MAN	0.599	175.993	1.50187	alpha	D	$^4C_1$	0.43	131.268	Ok
B	NAG	0.590	194.7	1.80272	beta	D	$^4C_1$	0.27	128.631	Ok
B	BMA	0.595	178.074	1.66373	beta	D	$^4C_1$	0.06	151.059	Ok
B	BMA	0.558	179.317	1.70322	beta	D	$^4C_1$	0.57	127.831	Ok
B	MAN	0.595	179.333	1.50382	alpha	D	$^4C_1$	0.66	151.967	Ok
B	NAG	0.622	195.762	1.60169	beta	D	$^4C_1$	0.68	139.604	Ok
B	NAG	0.614	198.506	2.19846	beta	D	$^4C_1$	0.75	151.534	Ok
B	BMA	0.537	182.465	2.53328	beta	D	$^4C_1$	0.52	158.486	Ok
B	MAN	0.604	173.773	1.72116	alpha	D	$^4C_1$	0.64	165.408	Ok
B	MAN	0.586	178.19	1.62354	alpha	D	$^4C_1$	0.73	165.332	Ok
B	MAN	0.571	180.435	1.65912	alpha	D	$^4C_1$	0.39	167.662	Ok

Table B.2: Data collection, and model and geometry statistics for the *Gg*Ten2$^{\Delta EGF}$–SOS structure S-SAD data collection. Data was collected at the Diamond Light Source I23 beamline at a higher wavelength of 2.75 (Å).

Protein Construct	*Gg*Ten2$^{\Delta EGF}$
Ligand	Sucrose-octasulfate (SOS)
Condition	Morpheus F6
Resolution Range (Å)	
Low Resolution	153.80
High Resolution	3.25
Space Group	P 1 2_1 1
Cell Dimensions	
a, *b*, *c* (Å)	140.31, 133.74, 158.18
α, β, γ (°)	90.00, 103.56, 90.00
$CC_{1/2}$ (%)	86.2
Completness (%)	97.0
Refinement	
No. Reflections	86896
R_{Work}	0.2417
R_{Free}	0.2448
No. Atoms	
Protein	29006
Ligand/ Ion	608
R.M.S.D	
Bond Lengths (Å)	0.05
Bond Angles (°)	0.86

C
Conservational Analysis

Table C.1: Conservational analysis of teneurin. The sequences used for conservation were obtained from Uniprot.

Isoform	Species	Uniprot Entry
Ten_1	*Homo sapiens* (Human)	Q9UKZ4
Ten_1	*Gallus gallus* (Chicken)	Q9W6V6
Ten_1	*Mus musculus* (Mouse)	Q9WTS4
Ten_1	*Callorhinus ursinus* (Northern fur seal)	A0A3Q7MN45
Ten_1	*Xenopus tropicalis* (Western clawed frog)	A0A6I8QJV4
Ten_1	*Danio rerio* (Zebrafish)	F1RA92
Ten_2	*Homo sapiens* (Human)	Q9NT68
Ten_2	*Gallus gallus* (Chicken)	Q9DER5
Ten_2	*Mus musculus* (Mouse)	Q9WTS5
Ten_2	*Callorhinus ursinus* (Northern fur seal)	A0A3Q7NMH6
Ten_2	*Xenopus tropicalis* (Western clawed frog)	A0A6I8SDQ5
Ten_2	*Pan troglodytes* (Chimpanzee)	A0A2I3RMK6
Ten_2	*Bos taurus* (Bovine)	F1MEX4
Ten_2	*Equus caballus* (Horse)	F6Q9M0
Ten_2	*Rattus norvegicus* (Rat)	Q9R1K2
Ten_3	*Homo sapiens* (Human)	Q9P273
Ten_3	*Gallus gallus* (Chicken)	A0A1L1RXD4
Ten_3	*Mus musculus* (Mouse)	Q9WTS6
Ten_3	*Callorhinus ursinus* (Northern fur seal)	A0A3Q7Q8H8
Ten_3	*Danio rerio* (Zebrafish)	Q9W7R4
Ten_3	*Alligator mississippiensis* (American alligator)	A0A151NNA2
Ten_4	*Homo sapiens* (Human)	Q6N022
Ten_4	*Mus musculus* (Mouse)	Q3UHK6
Ten_4	*Callorhinus ursinus* (Northern fur seal)	A0A3Q7NLF5
Ten_4	*Xenopus tropicalis* (Western clawed frog)	A6QL56
Ten_4	*Danio rerio* (Zebrafish)	Q9W7R3
Ten_A	*Drosophila melanogaster* (Fruit fly)	Q9VYN8
Ten_M	*Drosophila melanogaster* (Fruit fly)	O61307
Ten_1	*Caenorhabditis elegans*	G5EGQ6

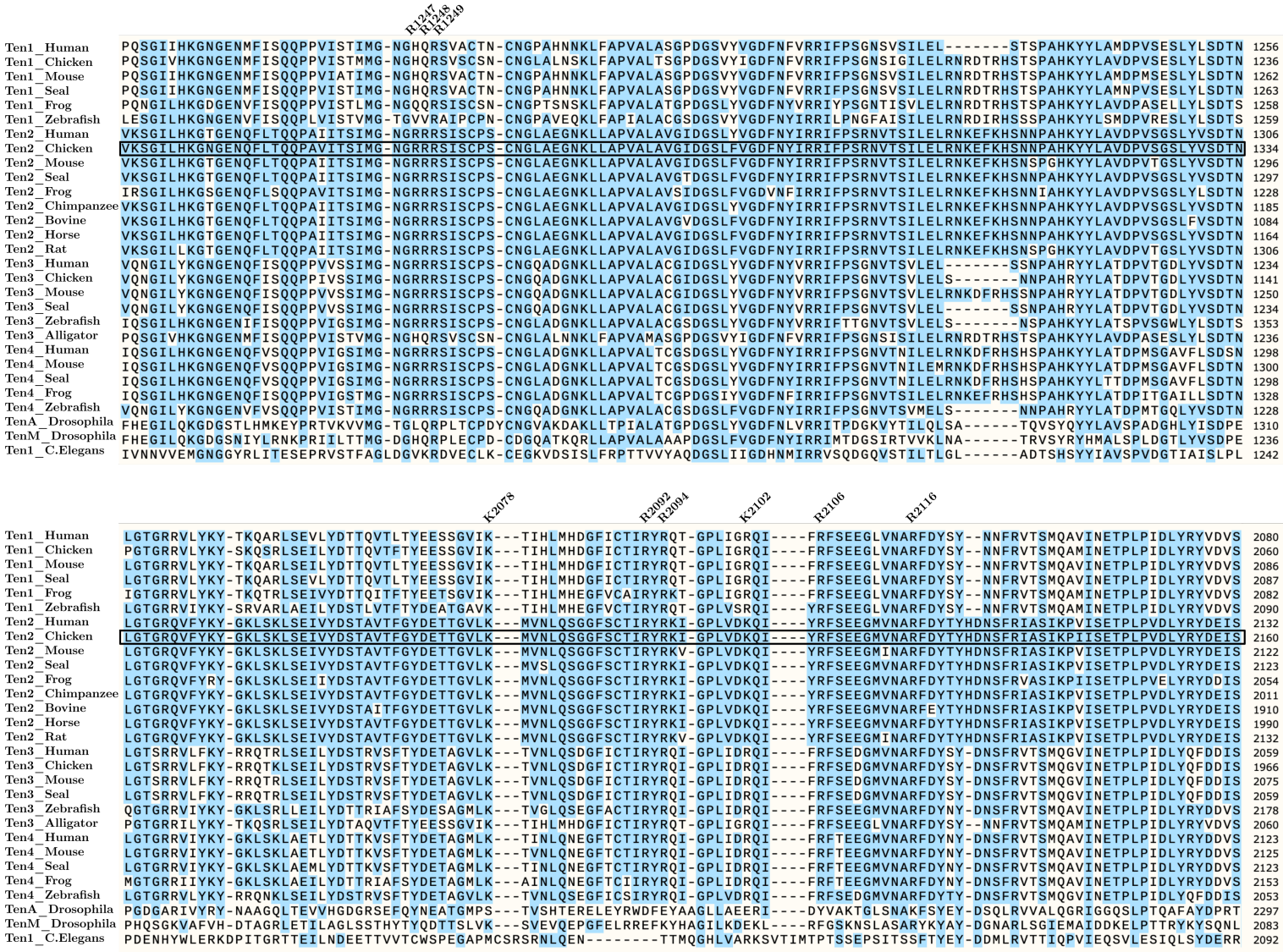


Figure C.1: Sequence alignment of teneurin proteins Alignment of 29 teneurin sequences showing the glycan binding sites predicted by the ClusPRO binding server (R1247, R1248, R1249) and resolved structurally (K2078, R2092, R2094, K2102 R2016, R2116).

References

[1] Ramon y Cajal. *Histology of the nervous system of man and vertebrates (translated from French)*. New York ; Oxford: Oxford University Press, 1995.

[2] R W Sperry. "Chemoaffinity in the orderly growth of nerve fiber paterns and connections". In: *Proc. Natl. Acad. Sci. USA* 50 (1963), pp. 703–710.

[3] Eric F. Schmidt, Hideaki Togashi, and Stephen M. Strittmatter. *Semaphorins and their Receptors in Vertebrates and Invertebrates*. Vol. 2-3. Elsevier Science (USA), 2003, pp. 877–881.

[4] K. Kullander et al. "Ephrin-B3 is the midline barrier that prevents corticospinal tract axons from recrossing, allowing for unilateral motor control". In: *Genes and Development* 15.7 (2001), pp. 877–888.

[5] Naoaki Ishii et al. "UNC-6, a laminin-related protein, guides cell and pioneer axon migrations in C. elegans". In: *Neuron* 9.5 (1992), pp. 873–881.

[6] Thomas Kidd, Kimberly S. Bland, and Corey S. Goodman. "Slit is the midline repellent for the Robo receptor in Drosophila". In: *Cell* 96.6 (1999), pp. 785–794.

[7] Patricia T. Yam and Frederic Charron. "Signaling mechanisms of non-conventional axon guidance cues: The Shh, BMP and Wnt morphogens". In: *Current Opinion in Neurobiology* 23.6 (2013), pp. 965–973.

[8] Daniel del Toro et al. "Regulation of Cerebral Cortex Folding by Controlling Neuronal Migration via FLRT Adhesion Molecules". In: *Cell* 169.4 (2017), 621–635.e16.

[9] Erik W. Dent, Stephanie L. Gupton, and Frank B. Gertler. "The growth cone cytoskeleton in Axon outgrowth and guidance". In: *Cold Spring Harbor Perspectives in Biology* 3.3 (2011), pp. 1–39.

[10] Sarah McFarlane. "Metalloproteases: Carving out a role in axon guidance". In: *Neuron* 37.4 (2003), pp. 559–562.

[11] Miguel Santiago-Medina et al. "Regulation of ECM degradation and axon guidance by growth cone invadosomes". In: *Development (Cambridge)* 142.3 (2015), pp. 486–496.

[12] M. Hattori, M. Osterfield, and J. G. Flanagan. "Regulated cleavage of a contact-mediated axon repellent". In: *Science* 289.5483 (2000), pp. 1360–1365.

[13] Marcus Singer, Ruth H. Nordlander, and Margaret Egar. "Axonal guidance during embryogenesis and regeneration in the spinal cord of the newt: The blueprint hypothesis of neuronal pathway patterning". In: *Journal of Comparative Neurology* 185.1 (1979), pp. 1–21.

[14] Jian Zuo et al. “Neuronal matrix metalloproteinase-2 degrades and inactivates a neurite- inhibiting chondroitin sulfate proteoglycan”. In: *Journal of Neuroscience* 18.14 (1998), pp. 5203–5211.

[15] Ana Gutierrez-Fernandez et al. “Plasminogen enhances neuritogenesis on laminin-1”. In: *Journal of Neuroscience* 29.40 (2009), pp. 12393–12400.

[16] Carrie L. Hehr, Jennifer C. Hocking, and Sarah McFarlane. “Matrix metalloproteinases are required for retinal ganglion cell axon guidance at select decision points”. In: *Development* 132.15 (2005), pp. 3371–3379.

[17] Yuanyuan Y. Chen et al. “Targeting of retinal axons requires the metalloproteinase ADAM10”. In: *Journal of Neuroscience* 27.31 (2007), pp. 8448–8456.

[18] Wei Hsuan Yu and J. Frederick Woessner. “Heparan sulfate proteoglycans as extracellular docking molecules for matrilysin (matrix metalloproteinase 7)”. In: *Journal of Biological Chemistry* 275.6 (2000), pp. 4183–4191.

[19] Ina Kalus et al. “Differential involvement of the extracellular 6-O-endosulfatases Sulf1 and Sulf2 in brain development and neuronal and behavioural plasticity”. In: *Journal of Cellular and Molecular Medicine* 13 (Nov. 2009), pp. 4505–4521.

[20] Richard P. Tucker et al. “Phylogenetic analysis of the teneurins: Conserved features and premetazoan ancestry”. In: *Molecular Biology and Evolution* 29.3 (2012), pp. 1019–1029.

[21] Richard P Tucker. “Horizontal Gene Transfer in Choanoflagellates”. In: *J. Exp. Zool* 320 (2013), pp. 1–9.

[22] Verity A. Jackson et al. “Structures of Teneurin adhesion receptors reveal an ancient fold for cell-cell interaction”. In: *Nature Communications* 9.1 (2018), pp. 1–9.

[23] Xiaochen Wang et al. “Caenorhabditis elegans transthyretin-like protein TTR-52 mediates recognition of apoptotic cells by the CED-1 phagocyte receptor”. In: *Nature Cell Biology* 12.7 (2010), pp. 655–664.

[24] Javier Garcia-Pardo et al. “Amyloid formation by human carboxypeptidase D transthyretin-like domain under physiological conditions”. In: *Journal of Biological Chemistry* 289.49 (2014), pp. 33783–33796.

[25] A. Alkelai et al. “A role for TENM1 mutations in congenital general anosmia”. In: *Clinical Genetics* 90.3 (2016), pp. 211–219.

[26] Verity A. Jackson et al. “Teneurin Structures Are Composed of Ancient Bacterial Protein Domains”. In: *Frontiers in Neuroscience* 13.March (2019).

[27] Christos Gatsogiannis et al. “Tc toxin activation requires unfolding and refolding of a β-propeller”. In: *Nature* 563.7730 (2018), pp. 209–233.

[28] David A. Lovejoy, Arij Al Chawaf, and M. Z.Alia Cadinouche. “Teneurin C-terminal associated peptides: An enigmatic family of neuropeptides with structural similarity to the corticotropin-releasing factor and calcitonin families of peptides”. In: *General and Comparative Endocrinology* 148.3 (2006), pp. 299–305.

[29] Liqun Wang et al. “Teneurin proteins possess a carboxy terminal sequence with neuromodulatory activity”. In: *Molecular Brain Research* 133.2 (2005), pp. 253–265.

[30] Laura A. Tan et al. “Teneurin C-terminal associated peptide (TCAP)-1 modulates dendritic morphology in hippocampal neurons and decreases anxiety-like behaviors in rats”. In: *Physiology and Behavior* 104.2 (2011), pp. 199–204.

[31] Laura A. Tan et al. “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”. In: *Behavioural Brain Research* 188.1 (2008), pp. 195–200.

[32] Jacqueline Ferralli, Richard P. Tucker, and Ruth Chiquet-Ehrismann. “The teneurin C-terminal domain possesses nuclease activity and is apoptogenic”. In: *Biology Open* 7.3 (2018), pp. 1–9.

[33] Xiao Hong Zhou et al. “The murine Ten-m/Odz genes show distinct but overlapping expression patterns during development and in adult brain”. In: *Gene Expression Patterns* 3.4 (2003), pp. 397–405.

[34] Olivia Bibollet-Bahena et al. “A dual-strategy expression screen for candidate connectivity labels in the developing thalamus”. In: *PLOS ONE* 12.5 (2017), pp. 1–43.

[35] Dominic S. Berns et al. “Teneurin-3 controls topographic circuit assembly in the hippocampus”. In: *Nature* 554.7692 (2018), pp. 328–333.

[36] Daniel del Toro et al. “Structural Basis of Teneurin-Latrophilin Interaction in Repulsive Guidance of Migrating Neurons”. In: *Cell* 180.2 (2020), 323–339.e19.

[37] Jingxian Li et al. “Alternative splicing controls teneurin-latrophilin interaction and synapse specificity by a shape-shifting mechanism”. In: *Nature Communications* 11.1 (2020).

[38] Jingxian Li et al. “Structural Basis for Teneurin Function in Circuit-Wiring: A Toxin Motif at the Synapse”. In: *Cell* 173.3 (2018), pp. 735–748.

[39] Richard Sando, Xian Jiang, and Thomas C. SÃ¼dhof. “Latrophilin GPCRs direct synapse specificity by coincident binding of FLRTs and teneurins”. In: *Science* 363.6429 (2019), eaav7969.

[40] Ariane D. Minet et al. “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”. In: *Journal of Cell Science* 112.12 (1999), pp. 2019–2032.

[41] Stefan Baumgartner and Ruth Chiquet-Ehrismann. “Tena, a Drosophila gene related to tenascin, shows selective transcript localization”. In: *Mechanisms of Development* 40.3 (1993), pp. 165–176.

[42] Stefan Baumgartner et al. “tenm, a Drosophila gene related to tenascin, is a new pair-rule gene”. In: *EMBO Journal* 13.16 (1994), pp. 3728–3740.

[43] Anna Levine et al. “odd Oz: A novel Drosophila pair rule gene”. In: *Cell* 77.4 (1994), pp. 587–598.

[44] Weizhe Hong, Timothy J. Mosca, and Liqun Luo. “Teneurins instruct synaptic partner matching in an olfactory map”. In: *Nature* 484.7393 (2012), pp. 201–207.

[45] Timothy J. Mosca et al. “Trans-synaptic Teneurin signalling in neuromuscular synapse organization and target choice”. In: *Nature* 484.7393 (2012), pp. 237–241.

[46] John Paul Silva et al. “Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.29 (2011), pp. 12113–12118.

[47] Matthew L. O’Sullivan et al. “LPHN3, a presynaptic adhesion-GPCR implicated in ADHD, regulates the strength of neocortical layer 2/3 synaptic input to layer 5”. In: *Neural Development* 9.1 (2014), pp. 1–11.

[48] Matthew L. O’Sullivan et al. “FLRT Proteins Are Endogenous Latrophilin Ligands and Regulate Excitatory Synapse Development”. In: *Neuron* 73.5 (2012), pp. 903–910.

[49] Paride Antinucci et al. “Teneurin-3 specifies morphological and functional connectivity of retinal ganglion cells in the vertebrate visual system”. In: *Cell Reports* 5.3 (2013), pp. 582–592.

[50] Judith S. Cruz-Ortega and Antony A. Boucard. “Actin cytoskeleton remodeling defines a distinct cellular function for adhesion G protein-coupled receptors ADGRL/latrophilins 1, 2 and 3”. In: *Biology Open* 8.4 (2019), pp. 1–11.

[51] Masahiro Mori et al. “Roles of Nectins and Nectin-Like Molecules in the Nervous System”. In: *Cell Adhesion Molecules: Implications in Neurological Diseases*. Ed. by Vladimir Berezin and Peter S. Walmod. New York, NY: Springer New York, 2014, pp. 91–116.

[52] Mary Packard et al. “The Drosophila Wnt, wingless, provides an essential signal for pre- and postsynaptic differentiation”. In: *Cell* 111.3 (2002), pp. 319–330.

[53] Claudia Miech et al. “Presynaptic local signaling by a canonical wingless pathway regulates development of the Drosophila neuromuscular junction”. In: *Journal of Neuroscience* 28.43 (2008), pp. 10875–10884.

[54] Anne Marie Luchtenborg et al. “Heterotrimeric Go protein links Wnt-Frizzled signaling with ankyrins to regulate the neuronal microtubule cytoskeleton”. In: *Development (Cambridge)* 141.17 (2014), pp. 3399–3409.

[55] Jingjun Li et al. “Crucial Role of Drosophila Neurexin in Proper Active Zone Apposition to Postsynaptic Densities, Synaptic Growth, and Synaptic Transmission”. In: *Neuron* 55.5 (2007), pp. 741–755.

[56] Daniel Banovic et al. “Drosophila Neuroligin 1 Promotes Growth and Postsynaptic Differentiation at Glutamatergic Neuromuscular Junctions”. In: *Neuron* 66.5 (2010), pp. 724–738.

[57] Daniel T. Pederick et al. “Reciprocal repulsions instruct the precise assembly of parallel hippocampal networks”. In: *Science* 372.6546 (2021), pp. 1068–1073.

[58] Artur Kania and Rudigger Klein. “Mechanisms of ephrin-Eph signalling in development, physiology and disease”. In: *Nature Reviews Molecular Cell Biology* 17.4 (2016), pp. 240–256.

[59] Franck Polleux, Theresa Morrow, and Anirvan Ghosh. “Semaphorin 3A is a chemoattractant for cortical apical dendrites”. In: *Nature* 404.6778 (2000), pp. 567–573.

[60] Zhuhao Wu et al. “A Combinatorial Semaphorin Code Instructs the Initial Steps of Sensory Circuit Assembly in the Drosophila CNS”. In: *Neuron* 70.2 (2011), pp. 281–298.

[61] Yi Fan Huang, Shuji Mizumoto, and Morihisa Fujita. “Novel Insight Into Glycosaminoglycan Biosynthesis Based on Gene Expression Profiles”. In: *Frontiers in Cell and Developmental Biology* 9.September (2021), pp. 1–13.

[62] Kristian Prydz. “Determinants of glycosaminoglycan (GAG) structure”. In: *Biomolecules* 5.3 (2015), pp. 2003–2022.

[63] A. Hoffer. “The discovery of heparin”. In: *Journal of Orthomolecular Medicine* 8.3 (1959), p. 133.

[64] K. M. Brinkhous et al. “The Inhibition of Blood Clotting: An Unidentified Substance Which Acts in Conjunction with Heparin to Prevent the Conversion of Prothrombin into Thrombin”. In: *American Journal of Physiology* 125.4 (1939), pp. 683–687.

[65] John M. Whitelock and Renato V. Iozzo. “Heparan sulfate: A complex polymer charged with biological activity”. In: *Chemical Reviews* 105.7 (2005), pp. 2745–2764.

[66] Keiichi Takagaki et al. “Domain structure of chondroitin sulfate E octasaccharides binding to type V collagen”. In: *Journal of Biological Chemistry* 277.11 (2002), pp. 8882–8889.

[67] Barbara Mulloy et al. *NMR and molecular-modelling studies of the solution conformation of heparin*. 1993, pp. 849–858.

[68] Barbara Mulloy and Mark J Forster. *Conformation and dynamics of heparin and heparan sulfate*. 2000, pp. 1147–1156.

[69] Gabriel Waksman and Andrew B Herr. “New insights into heparin-induced FGF oligomerization”. In: *Nature Structural Biology* 5 (7 1998), pp. 527–530.

[70] Samuel C. Griffiths et al. “Hedgehog-Interacting Protein is a multimodal antagonist of Hedgehog signalling”. In: *Nature Communications* 12 (1 Dec. 2021).

[71] Ramin Dubey et al. “R-spondins engage heparan sulfate proteoglycans to potentiate wnt signaling”. In: *eLife* 9 (May 2020), pp. 1–25.

[72] “Glycosaminoglycans”. In: *Handbook of Glycomics* (2010), pp. 59–80.

[73] Nikos K. Karamanos et al. “Ion-pair high-performance liquid chromatography for determining disaccharide composition in heparin and heparan sulphate”. In: *Journal of Chromatography A* 765.2 (1997), pp. 169–179.

[74] Jon A. Deakin and Malcolm Lyon. “A simplified and sensitive fluorescent method for disaccharide analysis of both heparan sulfate and chondroitin/dermatan sulfates from biological samples”. In: *Glycobiology* 18.6 (2008), pp. 483–491.

[75] Ola M. Saad and Julie A. Leary. “Compositional analysis and quantification of heparin and heparan sulfate by electrospray ionization ion trap mass spectrometry”. In: *Analytical Chemistry* 75.13 (2003), pp. 2985–2995.

[76] Yung Jen Chuang et al. “Heparin enhances the specificity of antithrombin for thrombin and factor Xa independent of the reactive center loop sequence. Evidence for an exosite determinant of factor Xa specificity in heparin-activated antithrombin”. In: *Journal of Biological Chemistry* 276.18 (2001), pp. 14961–14971.

[77] Ryan J. Weiss, Jeffrey D. Esko, and Yitzhak Tor. “Targeting heparin and heparan sulfate protein interactions”. In: *Organic and Biomolecular Chemistry* 15.27 (2017), pp. 5656–5668.

[78] Himanshu Gadgil and Harry W. Jarrett. “Heparin elution of transcription factors from DNA-Sepharose columns”. In: *Journal of Chromatography A* 848.1-2 (1999), pp. 131–138.

[79] David Anak Simon Davis and Christopher R. Parish. “Heparan sulfate: A ubiquitous glycosaminoglycan with multiple roles in immunity”. In: *Frontiers in Immunology* 4.DEC (2013), pp. 1–7.

[80] Anna M. Vogt et al. “Heparan sulphate identified on human erythrocytes: A Plasmodium falciparum receptor”. In: *Biochemical Journal* 381.3 (2004), pp. 593–597.

[81] Chun Han Han et al. “Drosophila glypicans control the cell-to-cell movement of Hedgehog by a dynamin-independent process”. In: *Development* 131.3 (2004), pp. 601–611.

[82] Xavier Franch-Marro et al. “Glypicans shunt the Wingless signal between local signalling and further transport”. In: *Development* 132.4 (2005), pp. 659–666.

[83] Keisuke Kamimura and Nobuaki Maeda. “Glypicans and Heparan Sulfate in Synaptic Development, Neural Plasticity, and Neurological Disorders”. In: *Frontiers in Neural Circuits* 15 (Feb. 2021).

[84] Yi Huei Linda Jen, Michele Musacchio, and Arthur D. Lander. “Glypican-1 controls brain size through regulation of fibroblast growth factor signaling in early neurogenesis”. In: *Neural Development* 4 (1 2009).

[85] S. Lugert et al. “Glypican-2 levels in cerebrospinal fluid predict the status of adult hippocampal neurogenesis”. In: *Scientific Reports* 7 (2017).

[86] Charlotte H. Coles et al. “Proteoglycan-specific molecular switch for RPTPσ clustering and neuronal extension”. In: *Science* 332.6028 (2011), pp. 484–488.

[87] Anni Hienola et al. “N-syndecan deficiency impairs neural migration in brain”. In: *Journal of Cell Biology* 174 (Aug. 2006), pp. 569–580.

[88] Tyson J. Edwards and Marc Hammarlund. “Syndecan promotes axon regeneration by stabilizing growth cone migration”. In: *Cell Reports* 8 (July 2014), pp. 272–283.

[89] K. Murakami et al. “Nerve injury induces the expression of syndecan-1 heparan sulfate proteoglycan in primary sensory neurons”. In: *Neuroscience* 300 (Aug. 2015), pp. 338–350.

[90] Hang Wang et al. “Chondroitin-4-sulfation negatively regulates axonal guidance and growth”. In: *Journal of Cell Science* 121.18 (2008), pp. 3083–3091.

[91] Rachel Lin et al. “6-Sulphated chondroitins have a positive influence on axonal regeneration”. In: *PLoS ONE* 6.7 (2011).

[92] Rahul Raman, V. Sasisekharan, and Ram Sasisekharan. "Structural Insights into biological roles of protein-glycosaminoglycan interactions". In: *Chemistry and Biology* 12.3 (2005), pp. 267–277.

[93] L. W. Fisher, J. D. Termine, and M. F. Young. "Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan II (decorin) and several nonconnective tissue proteins in a variety of species". In: *Journal of Biological Chemistry* 264.8 (1989), pp. 4571–4576.

[94] D. R. Zimmermann and E. Ruoslahti. "Multiple domains of the large fibroblast proteoglycan, versican". In: *EMBO Journal* 8.10 (1989), pp. 2975–2981.

[95] H. Greiling and JE. Scott. *Keratan sulphate-chemistry, biology, chemical pathology*. The Biochemical Society (London), 1989.

[96] James L. Funderburgh, Bruce Caterson, and Gary W. Conrad. "Keratan sulfate proteoglycan during embryonic development of the chicken cornea". In: *Developmental Biology* 116.2 (1986), pp. 267–277.

[97] Gavin M. Brown et al. "Human aggrecan keratan sulfate undergoes structural changes during adolescent development". In: *Journal of Biological Chemistry* 273.41 (1998), pp. 26408–26414.

[98] Birgit Meyer-Puttlitz et al. "Chondroitin Sulfate and Chondroitin/Keratan Sulfate Proteoglycans of Nervous Tissue: Developmental Changes of Neurocan and Phosphacan". In: *Journal of Neurochemistry* 65.5 (1995), pp. 2327–2337.

[99] Tomas Asplund et al. "Characterization of hyaluronan synthase from a human glioma cell line". In: *Biochimica et Biophysica Acta - General Subjects* 1380.3 (1998), pp. 377–388.

[100] Peter Prehm. "Biosynthesis of hyaluronan: Direction of chain elongation". In: *Biochemical Journal* 398.3 (2006), pp. 469–473.

[101] Tobias Schulz, Udo Schumacher, and Peter Prehm. "Hyaluronan export by the ABC transporter MRP5 and its modulation by intracellular cGMP". In: *Journal of Biological Chemistry* 282.29 (2007), pp. 20999–21004.

[102] Steven Bowman et al. "Recent advances in hyaluronic acid based therapy for osteoarthritis". In: *Clinical and Translational Medicine* 7.1 (2018).

[103] Caitlin O. McAtee, Joseph J. Barycki, and Melanie A. Simpson. *Emerging roles for hyaluronidase in cancer metastasis and therapy*. 1st ed. Vol. 123. Elsevier Inc., 2014, pp. 1–34.

[104] Dacai Liu et al. "Expression of hyaluronidase by tumor cells induces angiogenesis in vivo". In: *Proceedings of the National Academy of Sciences of the United States of America* 93.15 (1996), pp. 7832–7837.

[105] Masaru Inatani et al. "Mammalian Brain Morphogenesis and Midline Axon Guidance Require Heparan Sulfate". In: *Science* 302.5647 (2003), pp. 1044–1046.

[106] Yoshihiro Matsumoto et al. "Netrin-1/DCC signaling in commissural axon guidance requires cell-autonomous expression of heparan sulfate". In: *Journal of Neuroscience* 27.16 (2007), pp. 4342–4350.

[107] Patrick Steigemann et al. "Heparan Sulfate Proteoglycan Syndecan Promotes Axonal and Myotube Guidance by Slit/Robo Signaling". In: *Current Biology* 14.3 (2004), pp. 225–230.

[108] Fang Wang et al. “LAR receptor tyrosine phosphatases and HSPGs guide peripheral sensory axons to the skin”. In: *Current Biology* 22.5 (2012), pp. 373–382.

[109] Edda Kastenhuber et al. “Netrin-DCC, robo-slit, and heparan sulfate proteoglycans coordinate lateral positioning of longitudinal dopaminergic diencephalospinal axons”. In: *Journal of Neuroscience* 29.28 (2009), pp. 8914–8926.

[110] *Functional Requirements for Heparan Sulfate Biosynthesis in Morphogenesis and Nervous System Development in C. elegans*. Vol. 13. 1. 2017.

[111] Eillen Tecle, Carlos A. Diaz-Balzac, and Hannes E. Bulow. “Distinct 3-o-sulfated heparan sulfate modification patterns are required for kal-12dependent neurite branching in a context-dependent manner in caenorhabditis elegans”. In: *G3: Genes, Genomes, Genetics* 3.3 (2013), pp. 541–552.

[112] Hannes E. Bulow and Oliver Hobert. “Differential sulfations and epimerization define heparan sulfate specificity in nervous system development”. In: *Neuron* 41.5 (Mar. 2004), pp. 723–736.

[113] Hannes E. Bulow et al. “Extracellular Sugar Modifications Provide Instructive and Cell-Specific Information for Axon-Guidance Choices”. In: *Current Biology* 18 (2008), pp. 1978–1985.

[114] Simon L. Bullock et al. “Renal agenesis in mice homozygous for a gene trap mutation in the gene encoding heparan sulfate 2-sulfotransferase”. In: *Genes and Development* 12.12 (1998), pp. 1894–1906.

[115] Thomas Pratt et al. “Heparan sulphation patterns generated by specific heparan sulfotransferase enzymes direct distinct aspects of retinal axon guidance at the optic chiasm”. In: *Journal of Neuroscience* 26.26 (2006), pp. 6911–6923.

[116] James M. Clegg et al. “Heparan sulfotransferases Hs6st1 and Hs2st keep Erk in check for mouse corpus callosum development”. In: *Journal of Neuroscience* 34.6 (2014), pp. 2389–2401.

[117] Shouwen Wang et al. *QSulf1, a heparan sulfate 6-O-endosulfatase, inhibits fibroblast growth factor signaling in mesoderm induction and angiogenesis*. 2004, pp. 4833–4838.

[118] Keishi Narita et al. “HSulf-1 inhibits angiogenesis and tumorigenesis in vivo”. In: *Cancer Research* 66 (June 2006), pp. 6025–6032.

[119] Xingbin Ai et al. “QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling”. In: *Journal of Cell Biology* 162 (July 2003), pp. 341–351.

[120] Gurtej K Dhoot et al. *Regulation of Wnt Signaling and Embryo Patterning by an Extracellular Sulfatase*. 2001, pp. 1663–1666.

[121] Neil Dani et al. “A Targeted Glycan-Related Gene Screen Reveals Heparan Sulfate Proteoglycan Sulfation Regulates WNT and BMP Trans-Synaptic Signaling”. In: *PLoS Genetics* 8 (Nov. 2012).

[122] Rana El Masri et al. “Extracellular endosulfatase Sulf-2 harbors a chondroitin/dermatan sulfate chain that modulates its enzyme activity”. In: *Cell Reports* 38 (Mar. 2022).

[123] Hidenao Toyoda, Akiko Kinoshita-Toyoda, and Scott B. Selleck. "Structural analysis of glycosaminoglycans in Drosophila and Caenorhabditis elegans and demonstration that tout-velu, a Drosophila gene related to EXT tumor suppressors, affects heparan sulfate in vivo". In: *Journal of Biological Chemistry* 275.4 (2000), pp. 2269–2275.

[124] Tabea Dierker et al. "Nematodes join the family of chondroitin sulfate-synthesizing organisms: Identification of an active chondroitin sulfotransferase in Caenorhabditis elegans". In: *Scientific Reports* 6.July (2016), pp. 2–11.

[125] Diane M. Snow et al. "A chondroitin sulfate proteoglycan may influence the direction of retinal ganglion cell outgrowth." In: *Development* 113 4 (1991), pp. 1473–85.

[126] K. Y. Chung et al. "Axon routing at the optic chiasm after enzymatic removal of chondroitin sulfate in mouse embryos". In: *Development* 127.12 (2000), pp. 2673–2683.

[127] Tyler Schwend et al. "Corneal sulfated glycosaminoglycans and their effects on trigeminal nerve growth cone behavior in vitro: Roles for ECM in cornea innervation". In: *Investigative Ophthalmology and Visual Science* 53.13 (2012), pp. 8118–8137.

[128] Maki Ishii and Nobuaki Maeda. "Oversulfated chondroitin sulfate plays critical roles in the neuronal migration in the cerebral cortex". In: *Journal of Biological Chemistry* 283.47 (2008), pp. 32610–32620.

[129] Michael Schmalfeldt et al. "Brain derived versican V2 is a potent inhibitor of axonal growth". In: *Journal of Cell Science* 113.5 (2000), pp. 807–816.

[130] Yaojiong Wu et al. "Versican V1 Isoform Induces Neuronal Differentiation and Promotes Neurite Outgrowth". In: *Molecular Biology of the Cell* 15.5 (2004), pp. 2093–2104.

[131] K Nishimura et al. "Opposing functions of chondroitin sulfate and heparan sulfate during early neuronal polarization". In: *Neuroscience* 169 (Sept. 2010), pp. 1535–47.

[132] Samuel David and Albert J. Aguayo. "Axonal elongation into peripheral nervous system "bridges" after central nervous system injury in adult rats". In: *Science* 214.4523 (1981), pp. 931–933.

[133] Martin Benfey and Albert J. Aguayo. "Extensive elongation of axons from rat brain into peripheral nerve grafts". In: *Nature* 296.March (1982), pp. 150–152.

[134] J H Yi et al. "Alterations in sulfated chondroitin glycosaminoglycans following controlled cortical impact injury in mice". In: *Journal of Comparative Neurology* 520 (2012).

[135] Marcela Rincon-Restrepo et al. "Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes". In: *Nano Letters* 11.2 (2011), pp. 746–750.

[136] Albrecht M. Clement et al. “The DSD-1 carbohydrate epitope depends on sulfation, correlates with chondroitin sulfate D motifs, and is sufficient to promote neurite outgrowth”. In: *Journal of Biological Chemistry* 273.43 (1998), pp. 28444–28453.

[137] Albrecht M. Clement, Kazuyuki Sugahara, and Andreas Faissner. “Chondroitin sulfate E promotes neurite outgrowth of rat embryonic day 18 hippocampal neurons”. In: *Neuroscience Letters* 269.3 (1999), pp. 125–128.

[138] Kimberly M. Alonge et al. “Quantitative analysis of chondroitin sulfate disaccharides from human and rodent fixed brain tissue by electrospray ionization-tandem mass spectrometry”. In: *Glycobiology* 29.12 (2019), pp. 847–860.

[139] Daniela Carulli et al. “Animals lacking link protein have attenuated perineuronal nets and persistent plasticity”. In: *Brain* 133.8 (2010), pp. 2331–2347.

[140] Panpan Yu et al. “An In Vitro Model of Reactive Astrogliosis and Its Effect on Neuronal Growth”. In: *Astrocytes: Methods and Protocols*. Ed. by Richard Milner. Totowa, NJ: Humana Press, 2012, pp. 327–340.

[141] Daniela Carulli et al. “Composition of perineuronal nets in the adult rat cerebellum and the cellular origin of their components”. In: *Journal of Comparative Neurology* 494.4 (2006), pp. 559–577.

[142] L. Uhlin-Hansen et al. “Proteoglycan metabolism in normal and inflammatory human macrophages”. In: *Blood* 82.9 (1993), pp. 2880–2889.

[143] Rowan K. Hussein et al. “Role of Chondroitin Sulfation Following Spinal Cord Injury”. In: *Frontiers in Cellular Neuroscience* 14.August (2020), pp. 1–12.

[144] Clare M. Galtrey and James W. Fawcett. “The role of chondroitin sulfate proteoglycans in regeneration and plasticity in the central nervous system”. In: *Brain Research Reviews* 54.1 (2007), pp. 1–18.

[145] Yuntao Duan and Roman J. Giger. “A new role for RPTPσ in spinal cord injury: Signaling chondroitin sulfate proteoglycan inhibition”. In: *Science Signaling* 3.110 (2010), pp. 1–5.

[146] Yingjie Shen et al. “PTP s Is a Receptor for CSPG, an inhibitor of neural regeneration”. In: *Science* 326.October (2009), pp. 592–596.

[147] Simona Foscarin et al. “Brain ageing changes proteoglycan sulfation, rendering perineuronal nets more inhibitory”. In: *Aging* 9.6 (2017), pp. 1607–1622.

[148] Elizabeth J. Bradbury et al. “Chondroitinase ABC promotes functional recovery after spinal cord injury”. In: *Nature* 416.6881 (2002), pp. 636–640.

[149] Yu Shang Lee et al. “Nerve regeneration restores supraspinal control of bladder function after complete spinal cord injury”. In: *Journal of Neuroscience* 33.26 (2013), pp. 10591–10606.

[150] John D. Houle et al. “Combining an autologous peripheral nervous system "bridge" and matrix modification by chondroitinase allows robust, functional regeneration beyond a hemisection lesion of the adult rat spinal cord”. In: *Journal of Neuroscience* 26.28 (2006), pp. 7405–7415.

[151] James M. Massey et al. “Chondroitinase ABC digestion of the perineuronal net promotes functional collateral sprouting in the cuneate nucleus after cervical spinal cord injury”. In: *Journal of Neuroscience* 26.16 (2006), pp. 4406–4414.

[152] Guillermo Garcia-Alias et al. “Chondroitinase ABC treatment opens a window of opportunity for task-specific rehabilitation”. In: *Nature Neuroscience* 12.9 (2009), pp. 1145–1151.

[153] Stephanie C. Jefferson, Nicole J. Tester, and Dena R. Howland. “Chondroitinase ABC promotes recovery of adaptive limb movements and enhances axonal growth caudal to a spinal hemisection”. In: *Journal of Neuroscience* 31.15 (2011), pp. 5710–5720.

[154] Sujeong Yang et al. “Perineuronal net digestion with chondroitinase restores memory in mice with tau pathology”. In: *Experimental Neurology* 265 (2015), pp. 48–58.

[155] Luigi Corvetti and Ferdinando Rossi. “Degradation of chondroitin sulfate proteoglycans induces sprouting of intact Purkinje axons in the cerebellum of the adult rat”. In: *Journal of Neuroscience* 25.31 (2005), pp. 7150–7158.

[156] Jill R. Faulkner et al. “Reactive Astrocytes Protect Tissue and Preserve Function after Spinal Cord Injury”. In: *Journal of Neuroscience* 24.9 (2004), pp. 2143–2155.

[157] Gary Temple et al. “The completion of the Mammalian Gene Collection (MGC)”. In: *Genome Research* 19.12 (2009), pp. 2324–2333.

[158] Verity Jackson. “Mechanistic Studies of the adhesion-GPCR Latrophilin and its Interactions in Neural Guidance”. PhD thesis. Department of Biochemistry, University of Oxford, 2017.

[159] A. Radu Aricescu, Weixian Lu, and E. Yvonne Jones. “A time- and cost-efficient system for high-level protein production in mammalian cells”. In: *Acta Crystallographica Section D* 62.10 (2006), pp. 1243–1250.

[160] Veronica T Chang et al. “Glycoprotein Structural Genomics: Solving the Glycosylation Problem”. In: *Structure* 15.3 (Mar. 2007), pp. 267–273.

[161] A. D. Elbein et al. “Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I”. In: *Journal of Biological Chemistry* 265.26 (1990), pp. 15599–15605.

[162] J. Jancarik and S.-H. Kim. “Sparse matrix sampling: a screening method for crystallization of proteins”. In: *Journal of Applied Crystallography* 24.4 (1991), pp. 409–411.

[163] Fabrice Gorrec. “The Morpheus protein crystallisation screen”. In: *Journal of Applied Crystallography* 42.6 (2009), pp. 1035–1042.

[164] Fabrice Gorrec. “The Morpheus II protein crystallisation screen”. In: *Acta Crystallographica Section F* 71.7 (2015), pp. 831–837.

[165] Rebecca Page and Raymond C. Stevens. “Crystallization data mining in structural genomics: using positive and negative results to optimize protein crystallization screens”. In: *Methods* 34.3 (2004), pp. 373–389.

[166] Janet Newman et al. “Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy”. In: *Acta Crystallographica Section D* 61.10 (2005), pp. 1426–1431.

[167] Sergei Radaev and Peter D. Sun. “Crystallization of protein-protein complexes”. In: *Journal of Applied Crystallography* 35.6 (2002), pp. 674–676.

[168] Clemens Grimm et al. “A crystallization screen based on alternative polymeric precipitants”. In: *Acta Crystallographica Section D* 66.6 (2010), pp. 685–697.

[169] Ting-Chou Hu et al. “High-molecular-weight polymers for protein crystallization: poly-γ-glutamic acid-based precipitants”. In: *Acta Crystallographica Section D* 64.9 (2008), pp. 957–963.

[170] Alexander McPherson and Bob Cudney. “Searching for silver bullets: An alternative strategy for crystallizing macromolecules”. In: *Journal of Structural Biology* 156.3 (2006), pp. 387–406.

[171] R. Cudney et al. “Screening and optimization strategies for macromolecular crystal growth”. In: *Acta Crystallographica. Section D* 50.4 (1994), pp. 414–423.

[172] Stefan Horer et al. “Crystal-contact engineering to obtain a crystal form of the Kelch domain of human Keap1 suitable for ligand-soaking experiments”. In: *Acta crystallographica. Section F* 69.6 (2013), 592â”596.

[173] O. Aurelius et al. “Long-wavelength macromolecular crystallography – First successful native SAD experiment close to the sulfur edge”. In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 411 (2017), pp. 12–16.

[174] C. Vonrhein et al. “Data Processing and Analysis with the autoPROC toolbox”. In: *Acta Crystallographica Section D* 67 (2011), pp. 293–302.

[175] I.J. Tickle et al. *STARANISO*. Cambridge, United Kingdom: Global Phasing Ltd. 2019.

[176] Wolfgang Kabsch. “*XDS*”. In: *Acta Crystallographica Section D* 66.2 (2010), pp. 125–132.

[177] Philip Evans. “Scaling and assessment of data quality”. In: *Acta Crystallographica Section D* 62.1 (2006), pp. 72–82.

[178] Philip R. Evans and Garib N. Murshudov. “How good are my data and what is the resolution?” In: *Acta Crystallographica Section D* 69.7 (2013), pp. 1204–1214.

[179] P. Andrew Karplus and Kay Diederichs. “Assessing and maximizing data quality in macromolecular crystallography”. In: *Current Opinion in Structural Biology* 34 (2015), pp. 60–68.

[180] Kay Diederichs and P Andrew Karplus. “Improved R-factors for diffraction data analysis in macromolecular crystallography”. In: *Nature Structural Biology* 4.4 (1997), pp. 269–275.

[181] Airlie J. McCoy et al. “*Phaser* crystallographic software”. In: *Journal of Applied Crystallography* 40.4 (2007), pp. 658–674.

[182] George M. Sheldrick. “A short history of *SHELX*”. In: *Acta Crystallographica Section A* 64.1 (2008), pp. 112–122.

[183] M. D. Winn, M. N. Isupov, and G. N. Murshudov. “Use of TLS parameters to model anisotropic displacements in macromolecular refinement”. In: *Acta Crystallographica Section D* 57.1 (2001), pp. 122–133.

[184] G. Bricogne et al. *Buster Version 2.10.3*. Cambridge, United Kingdom: Global Phasing Ltd. 2019.

[185] Paul Emsley et al. “Features and Development of Coot”. In: *Acta Crystallographica Section D - Biological Crystallography* 66 (2010), pp. 486–501.

[186] Gerard J. Kleywegt and T. Alwyn Jones. “Databases in Protein Crystallography”. In: *Acta Crystallographica Section D* 54.6 Part 1 (1998), pp. 1119–1131.

[187] Paul D. Adams et al. “*PHENIX*: a comprehensive Python-based system for macromolecular structure solution”. In: *Acta Crystallographica Section D* 66.2 (Feb. 2010), pp. 213–221.

[188] Vincent B. Chen et al. “MolProbity: all-atom structure validation for macromolecular crystallography”. In: *Acta Crystallographica. Section D: Biological Crystallography* 66.Pt 1 (2010).

[189] Christopher J. Williams et al. “MolProbity: More and better reference data for improved all-atom structure validation”. In: *Protein Science* 27.1 (2018), pp. 293–315.

[190] Jon Agirre et al. “Privateer: software for the conformational validation of carbohydrate structures”. In: *Nature Structural & Molecular Biology* 22.11 (2015), pp. 833–834.

[191] Jon Agirre et al. “Carbohydrate anomalies in the PDB”. In: *Nature Chemical Biology* 11.5 (2015), p. 303.

[192] Haroldas Bagdonas, Daniel Ungar, and Jon Agirre. “Leveraging glycomics data in glycoprotein 3D structure validation with Privateer”. In: *Beilstein Journal of Organic Chemistry* 16 (2020), pp. 2523–2533.

[193] Matthew A. Cooper. “Label-free screening of bio-molecular interactions”. In: *Analytical and Bioanalytical Chemistry* 377.5 (2003), pp. 834–842.

[194] Christine L. Weingart et al. “Fluorescent labels influence phagocytosis of Bordetella pertussis by human neutrophils”. In: *Infection and Immunity* 67.8 (1999), pp. 4264–4267.

[195] G E Healthcare and Life Sciences. “Biacore Assay Handbook”. In: (2012).

[196] Xinhua Lin et al. “Heparan sulfate proteoglycans are essential for FGF receptor signaling during Drosophila embryonic development”. In: *Development* 126.17 (1999), pp. 3715–3723.

[197] Peng Zhang et al. “Heparan Sulfate Organizes Neuronal Synapses through Neurexin Partnerships”. In: *Cell* 174.6 (2018), 1450–1464.e23.

[198] Dingze Mang et al. “Heparan Sulfate-Instructed Self-Assembly Selectively Inhibits Cancer Cell Migration”. In: *ACS Applied Materials and Interfaces* 13.15 (2021), pp. 17236–17242.

[199] Tarja Kinnunen, Joanne Townsend, and Jeremy Turnbull. “Heparan sulfate is essential for neuron migration and axon outgrowth in Caenorhabditis elegans”. In: *International Journal of Experimental Pathology* 85.4 (2004), A69–A70.

[200] Dima Kozakov et al. “The ClusPro web server for protein-protein docking”. In: *Nature Protocols* 12.2 (2017), pp. 255–278.

[201] Fabrice Lefevre, Marie Helene Remy, and Jean Michel Masson. “Alanine-stretch scanning mutagenesis: A simple and efficient method to probe protein structure and function”. In: *Nucleic Acids Research* 25.2 (1997), pp. 447–448.

[202] Oded Edelheit et al. “Identification of the roles of conserved charged residues in the extra cellular domain of an epithelial sodium channel (ENaC) subunit by alanine mutagenesis”. In: *American Journal of Physiology - Renal Physiology* 300.4 (2011), pp. 887–897.

[203] Tabrez J. Siddiqui et al. “An LRRTM4-HSPG complex mediates excitatory synapse development on dentate gyrus granule cells”. In: *Neuron* 79.4 (2013), pp. 680–695.

[204] Joris DeWit et al. “Unbiased discovery of Glypican as a receptor for LRRTM4 in regulating excitatory synapse development”. In: *Neuron* 79.4 (2013), pp. 696–711.

[205] Thomas C. Sudhof. “Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits”. In: *Cell* 171.4 (2017), pp. 745–769.

[206] Andrew D. Rouillard et al. “The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins”. In: *Database : the journal of biological databases and curation* 2016 (2016).

[207] Andrew I Su et al. “Large-scale analysis of the human and mouse transcriptomes”. In: *PNAS* 99 (7 2002), pp. 4465–4470.

[208] Andrew I Su et al. “A gene atlas of the mouse and human protein-encoding transcriptomes”. In: *PNAS* 101 (16 2004), pp. 6062–6067.

[209] Rafael Salto et al. “β-Hydroxy-β-methylbutyrate (HMB) promotes neurite outgrowth in Neuro2a cells”. In: *PLoS ONE* 10.8 (2015), pp. 1–13.

[210] Zhuo Wang et al. “Catalpol inhibits amyloid-β generation through promoting α-cleavage of APP in Swedish mutant APP overexpressed N2a cells”. In: *Frontiers in Aging Neuroscience* 10.MAR (2018), pp. 1–11.

[211] Patrick Provost. “Interpretation and applicability of microrna datato the context of Alzheimer’s and age-related diseases”. In: *Aging* 2.3 (2010), pp. 166–169.

[212] Luigi A. De Girolamo, Alan J. Hargreaves, and E. Ellen Billett. “Protection from MPTP-induced neurotoxicity in differentiating mouse N2a neuroblastoma cells”. In: *Journal of Neurochemistry* 76.3 (2001), pp. 650–660.

[213] Xiaoling Yuan et al. “Respiratory syncytial virus prolifically infects N2a neuronal cells, leading to TLR4 and nucleolin protein modulations and RSV F protein co-localization with TLR4 and nucleolin”. In: *Journal of Biomedical Science* 25.1 (2018), pp. 1–13.

[214] Talita Giacomet de Carvalho et al. “A simple protocol for transfecting human mesenchymal stem cells”. In: *Biotechnology Letters* 40.3 (2018), pp. 617–622.

[215] Alison T. DePew, Michael A. Aimino, and Timothy J. Mosca. “The tenets of teneurin: Conserved mechanisms regulate diverse developmental processes in the Drosophila nervous system”. In: *Frontiers in Neuroscience* 13.JAN (2019), pp. 1–11.

[216] Marion Kusche-Gullberg et al. “Drosophila heparan sulfate, a novel design”. In: *Journal of Biological Chemistry* 287.26 (2012), pp. 21950–21956.

[217] Sylvain D. Vallet, Olivier Clerc, and Sylvie Ricard-Blum. “Glycosaminoglycan-Protein Interactions: The First Draft of the Glycosaminoglycan Interactome”. In: *Journal of Histochemistry and Cytochemistry* 69.2 (2021), pp. 93–104.

[218] John M Walker. *Carbohydrate Microarrays methods and Protocols*. Vol. 531. 1. 2009, p. 588.

[219] Leonard L. Jones, Dana Sajed, and Mark H. Tuszynski. "Axonal Regeneration through Regions of Chondroitin Sulfate Proteoglycan Deposition after Spinal Cord Injury: A Balance of Permissiveness and Inhibition". In: *Journal of Neuroscience* 23.28 (2003), pp. 9276–9288.

[220] Kavita Singh et al. "Structure of the DBL3x domain of pregnancy-associated malaria protein VAR2CSA complexed with chondroitin sulfate A". In: *Nature Structural and Molecular Biology* 15.9 (2008), pp. 932–938.

[221] Ferenc Zsila et al. "Heparin and Heparan Sulfate Binding of the Antiparasitic Drug Imidocarb: Circular Dichroism Spectroscopy, Isothermal Titration Calorimetry, and Computational Studies". In: *Journal of Physical Chemistry B* 122.6 (2018), pp. 1781–1791.

[222] Maria J. Hernaiz et al. "Characterization of heparin binding by a peptide from amyloid P component using capillary electrophoresis, surface plasmon resonance and isothermal titration calorimetry". In: *European Journal of Biochemistry* 269.12 (2002), pp. 2860–2867.

[223] David Ulbricht et al. "Basic residues of β-sheet a contribute to heparin binding and activation of vaspin (serpin A12)". In: *Journal of Biological Chemistry* 292.3 (2017), pp. 994–1004.

[224] Davide Angeletti et al. "Binding of subdomains 1/2 of PfEMP1-DBL1 α to Heparan sulfate or heparin mediates Plasmodium falciparum rosetting". In: *PLoS ONE* 10.3 (2015), pp. 1–15.

[225] Moran Jerabek-Willemsen et al. "Molecular Interaction Studies Using Microscale Thermophoresis". In: *ASSAY and Drug Development Technologies* 9.4 (2011), pp. 342–353.

[226] Moran Jerabek-Willemsen et al. "MicroScale Thermophoresis: Interaction analysis and beyond". In: *Journal of Molecular Structure* 1077 (2014), pp. 101–113.

[227] U. Warttinger et al. "A fluorescent probe assay (Heparin Red) for direct detection of heparins in human plasma". In: *Analytical and Bioanalytical Chemistry* 408.28 (2016), pp. 8241–8251.

[228] Melissa Rappold, Ulrich Warttinger, and Roland Kramer. "A fluorescent probe for glycosaminoglycans applied to the detection of dermatan sulfate by a mix-and-read assay". In: *Molecules* 22.5 (2017), pp. 1–11.

[229] Yusuke Mii and Shinji Takada. "Heparan Sulfate Proteoglycan Clustering in Wnt Signaling and Dispersal". In: *Frontiers in Cell and Developmental Biology* 8.July (2020).

[230] Neha Vyas et al. "Nanoscale Organization of Hedgehog Is Essential for Long-Range Signaling". In: *Cell* 133.7 (2008), pp. 1214–1227.

[231] Yusuke Mii et al. "Roles of two types of heparan sulfate clusters in Wnt distribution and signaling in Xenopus". In: *Nature Communications* 8.1 (2017), pp. 1–19.

[232] Imeobong U. Antia et al. “Analysis of procainamide-derivatised heparan sulphate disaccharides in biological samples using hydrophilic interaction liquid chromatography mass spectrometry”. In: *Analytical and Bioanalytical Chemistry* 410.1 (2018), pp. 131–143.

[233] Chun Shao et al. “Mass spectral profiling of glycosaminoglycans from histological tissue surfaces”. In: *Analytical Chemistry* 85.22 (2013), pp. 10984–10991.

[234] Francyne Kubaski et al. “Glycosaminoglycans detection methods: Applications of mass spectrometry”. In: *Molecular Genetics and Metabolism* 120.1-2 (2017), pp. 67–77.

[235] “Detection and quantification of the sulfated disaccharides in chondroitin sulfate by electrospray tandem mass spectrometry”. In: *Journal of the American Society for Mass Spectrometry* 11.10 (2000), pp. 916–920.

[236] Harumi Osago et al. “Quantitative analysis of glycosaminoglycans, chondroitin/dermatan sulfate, hyaluronic acid, heparan sulfate, and keratan sulfate by liquid chromatography-electrospray ionization-tandem mass spectrometry”. In: *Analytical Biochemistry* 467 (2014), pp. 62–74.

[237] Chie Mitsunaga et al. “Chondroitin sulfate/dermatan sulfate hybrid chains in the development of cerebellum: Spatiotemporal regulation of the expression of critical disulfated disaccharides by specific sulfotransferases”. In: *Journal of Biological Chemistry* 281.28 (2006), pp. 18942–18952.

[238] Elaine H. Shen, Caroline C. Overly, and Allan R. Jones. “The Allen Human Brain Atlas. Comprehensive gene expression mapping of the human brain.” In: *Trends in Neurosciences* 35.12 (2012), pp. 711–714.

[239] Reference Atlases et al. “ALLEN Developing Mouse Brain Atlas”. In: June (2013), pp. 1–10.

[240] Daniela Kenzelmann et al. “Teneurin-1 is expressed in interconnected regions of the developing brain and is processed in vivo”. In: *BMC Developmental Biology* 8 (2008), pp. 1–14.

[241] Daniela J. Di Bella et al. “Molecular logic of cellular diversification in the mouse cerebral cortex”. In: *Nature* 595.7868 (2021), pp. 554–559.

[242] Bernd Knoll et al. “Stripe assay to examine axonal guidance and cell migration”. In: *Nature Protocols* 2.5 (2007), pp. 1216–1224.

[243] Satoru Yamagishi et al. “Stripe assay to study the attractive or repulsive activity of a protein substrate using dissociated hippocampal neurons”. In: *Journal of Visualized Experiments* 2016.112 (2016), pp. 1–7.

[244] Tetsuichiro Saito and Norio Nakatsuji. “Efficient gene transfer into the embryonic mouse brain using in vivo electroporation”. In: *Developmental Biology* 240.1 (2001), pp. 237–246.

[245] H. Tabata and K. Nakajima. “Efficient in utero gene transfer system to the developing mouse brain using electroporation: Visualization of neuronal migration in the developing cortex”. In: *Neuroscience* 103.4 (2001), pp. 865–872.

[246] T. Fukuchi-Shimogori and E. A. Grove. “Neocortex patterning by the secreted signaling molecute FGF8”. In: *Science* 294.5544 (2001), pp. 1071–1074.

[247] Joseph Lo Turco, Jean Bernard Manent, and Faez Sidiqi. "New and improved tools for in utero electroporation studies of developing cerebral cortex". In: *Cerebral Cortex* 19.SUPPL. 1 (2009).

[248] Sharon Margriet Kolk, Annetrude Johanne de Mooij-Malsen, and Gerard Julianus Maria Martens. "Spatiotemporal Molecular Approach of in utero Electroporation to Functionally Decipher Endophenotypes in Neurodevelopmental Disorders". In: *Frontiers in Molecular Neuroscience* 4.November (2011), pp. 1–7.

[249] R. B. Campenot. "Local control of neurite development by nerve growth factor". In: *Proceedings of the National Academy of Sciences of the United States of America* 74.10 (1977), pp. 4516–4519.

[250] Sunja Kim et al. "Microfluidic systems for axonal growth and regeneration research". In: *Neural Regeneration Research* 9.19 (2014), pp. 1703–1705.

[251] Varun Balaji et al. "Pathological missorting of endogenous MAPT/Tau in neurons caused by failure of protein degradation systems". In: *Autophagy* 14.12 (2018), pp. 2139–2154.