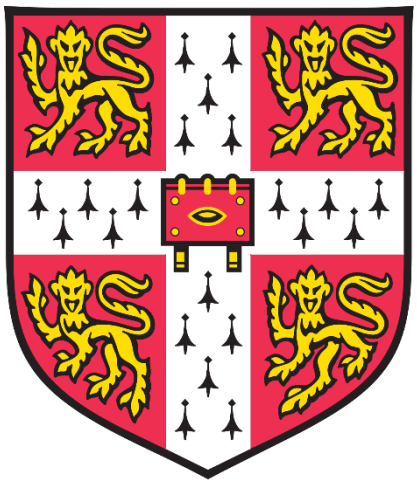
Characterisation of Prodomains in the GDNF Family



Matthew Ratcliff

University of Cambridge

Department of Biochemistry

St Edmund's College

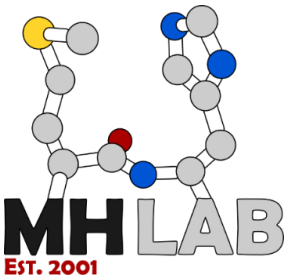
This dissertation is submitted for the degree of

Doctor of Philosophy

March 2021

Supervisor: Dr Marko Hyvönen - University of Cambridge

Co-supervisor: Dr Lutz Jermutus – AstraZeneca



Declaration of Originality

I confirm that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. This work has not been submitted in whole or in part for consideration for any other degree, diploma or other qualification at the University of Cambridge, any other University or similar institution. This dissertation does not exceed the prescribed word limit.

March 2021

Acknowledgements

Many thanks to my supervisor Marko, for his endless support, encouragement, optimism, and scientific guidance provided over the years. To all Hyvönen Group and Qkine members, past and present; Xuelu, Joe, Richard, Teo, Tom, Beata, Katharina, Luana, Monique, Catherine, Francesca, Marcelo, Paul, Aleksei, Marghe, Tonia, Andrew, Natalija, Katy, Chloe, Gerhard, and honorary lab member Michal. Thank you for your support, ideas, know-how and friendship in and out of the lab.

I would also like to thank the Biophysics and Crystallography facilities in the Department of Biochemistry. Especially, Dr Katherine Stott for help and support with CD and ITC, and expert advice and Dr Paul Brear for assistance with X-ray crystallography. In the department I would also like to express my gratitude to Clova, Janet and Tim for guaranteed entertainment and sustenance (on occasion). Thank you to our collaborators in the Stephen O'Rahilly research group for including us on the H202D project and performing key experiments for this thesis.

To my AstraZeneca supervisor Lutz, thank you for your support, guidance, and encouragement. Special, thank you to everyone at AstraZeneca, Gothenburg for welcoming me wholeheartedly and providing endless support - making my time in Sweden productive and enjoyable. Thank you for all the help and support from Dr Carina Johansson with HDX-MS and by Dr Anna Backmark with GDF15 bioactivity assays. At the Cambridge site I would particularly like to thank Dr Jacqueline Naylor for assistance and expertise in NRTN bioassays and excellent advice.

Thank you to all my friends at St Edmund's College, the University and beyond - especially Ben, Rachel, Harshita, Karson and Jon. Thanks to everyone in my BBSRC funded PhD cohort, and a special thanks to Aurora for keeping me sane and smiling, especially during coronavirus lockdowns. To my whole family, especially my Mam, Jenna, Caitlin, and little Freddie thank you for your love and support – I would not be anywhere without you all!

Finally, I would like to thank the BBSRC and AstraZeneca for funding and supporting this work.

Abbreviations

2-Aminoacridone Dye (AMAC)

Acetonitrile (ACN)

Activin Receptor-Like Kinase 1 (ALK1)

Amino acid (aa)

Anti-Müllerian Hormone (AMH)

Artemin (ARTN)

Beta-mercaptoethanol (β-ME)

Bio-Layer Interferometry (BLI)

Bone Morphogenetic Protein (BMP)

Bone Morphogenetic Protein 1/Tolloid (BMP1/TLD)

Bovine Serum Albumin (BSA)

Cadherin-Like Domain (CLD)

Chinese Hamster Ovary (CHO)

Circular Dichroism (CD)

C-Jun N-Terminal Kinase (JNK)

Column Volumes (CV)

Confidence Interval (CI)

Congenital Central Hypoventilation Syndrome (CCHS)

Cysteine Rich Domain (CRD)

Degree of Polymerisation (DP)

Deuterium Oxide/Heavy Water (D_2O)

Disulfide Bond (S-S)

Dithiothreitol (DTT)

Dopaminergic (DA)

Downstream of Tyrosine Kinases (DOK)

Dulbecco's Modified Eagle Media (DMEM)

Electrophoretic Mobility Shift Assay (EMSA)

Ellman's Reagent ((DTNB, 5,5-Dithio-Bis-(2-Nitrobenzoic Acid))

Endoplasmic Reticulum (ER)

Enzyme-Linked Immunosorbent Assay (ELISA)

Epidermal Growth Factor (EGF)

Escherichia Coli (E. Coli)

Ethylene Diamine Tetra Acetic Acid (EDTA)

Extracellular Matrix (ECM)

Fibroblast Growth Factor (FGF)

Fibroblast Growth Factor Receptor Substrate 2 (FRS2)

First Domain (D1)

Foetal Bovine Serum (FBS)

Follistatin (FST)

GDNF Family Receptor-A (Gfrα)

Glial-Cell-Line-Derived Neurotrophic Factor (GDNF)

Glial-Derived Neurotrophic Factor (GDNF)

Glial-Derived Neurotrophic Factor-Family Receptor A-Like (GFRAL)

Glucagon-Like Peptide-1 (GLP-1)

Glutathione S-Transferase (GST)

Glycoprotein A Repetitions Predominant (GARP)

Glycoprotein Hormone Family (GPH)

Glycosylphosphatidylinositol (GPI)

Growth and Differentiation Factor (GDF)

Growth and Differentiation Factor 15 (GDF15)

Half Maximal Effective Concentration (EC_{50})

Heparan Sulfate (HS)

Heterozygous (HZ)

Hirschsprung's Disease (HSCR)

Human Embryonic Kidney 293 Suspension (HEK293S)

Hydrochloric acid (HCl)

Hydrogen–Deuterium Exchange Mass Spectrometry (HDX-MS)

Hyperemesis Gravidarum (HG)

Inclusion Body (IB)

Insulin Receptor Substrate 1/2 (IRS1/2)

Ion-Exchange Chromatography (IEX)

Isopropyl-Beta-D-Thiogalactopyranoside (IPTG)

Isothermal Titration Calorimetry (ITC)

Lipopolysaccharides (LPS)

Low-Molecular-Weight Heparin (LMWH)

Lysis Buffer (LB)

Macrophage Inhibitory Cytokine 1 (MIC-1)

Maltose-Binding Protein (MBP)

Mature GDF15 (m-GDF15)

Mature NRTN (m-NRTN)

Mean Residue Ellipticity (MRE)

Medullary Thyroid Carcinoma (MTC)

Mitogen-Activated Protein Kinase (MAPK)

Molecular Weight (M_w)

Multiple Endocrine Neoplasia IIA (MEN2A)

Multiple Endocrine Neoplasia IIB (MEN2B)

Nerve Growth Factor (NGF)

Neural Cell Adhesion Molecule (NCAM)

Neurturin (NRTN)

Nickel-Immobilised Affinity Chromatography (Ni-IMAC)

Non-Detergent Sulfobetaines (NDSBS)

Non-Small Cell Lung Cancer (NSCLC)

NSAID Activated Gene (NAG-1)

Nuclear Magnetic Resonance (NMR)

Oncostatin M (OSM)

Online Mendelian Inheritance in Man (OMIM)

Papillary Thyroid Carcinoma (PTC)

Parkinson's Disease (PD)

Persephin (PSPN)

Persistent Müellerian Duct Syndrome 1 (PMDS1)

Phosphatidylinositol-3-Kinase (PI-3K)

Plasma Membrane (PM)

Platelet-Derived Growth Factor (PDGF)

Pro-Protein Convertase Subtilisin/Kexin (PCSK)

Protein Data Bank (PDB)

ProteinLynx Global Server (PLGS)

Pyridinium Propyl Sulfobetaine (PPS)

R Squared (R^2)

Rearranged During Transfection (RET)

Reversed Phase Chromatography (RPC)

Second Domain (D2)

Serum Response Factor (SRF)

Size Exclusion Chromatography (SEC)

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

Small Ubiquitin-Like Modifier (SUMO)

Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Surface Plasmon Resonance (SPR)

Tobacco Etch Virus (TEV)

Third Domain (D3)

Transforming Growth Factor Beta (TGF-B)

Trans-Golgi Network (TGN)

Transmembrane Domain (TMD)

Trifluoroacetic Acid (TFA)

Tyrosine Kinase (TK)

Vascular Endothelial Growth Factor (VEGF)

Wild Type (WT)

Zucker Diabetic Fatty (ZDF)

Abstract

Growth Differentiation Factor 15 (GDF15) and Neurturin (NRTN) have wide ranging roles in biology, with potential applications in diseases such as diabetes and obesity. As members of the glial cell line derived neurotrophic factor (GDNF) family, both GDF15 and NRTN signal through the RET receptor. RET signalling is essential in normal development and is facilitated by a distinct co-receptor for each GDNF family member. NRTN was originally identified as a neurotrophic factor and is associated with the normal development and maintenance of the nervous system. Additionally, NRTN has positive effects on hyperglycaemia in diabetic animal models. GDF15 is a stress regulator of energy homeostasis and body weight and is associated with numerous pathological conditions including cancer, obesity, and anorexia.

The GDNF family is part of the larger TGF-β superfamily of growth factors. Like all TGF-β superfamily members, proteins are produced as larger precursors consisting of an N-terminal prodomain and a C-terminal mature domain – pro-TGF-β proteins. The mature domains dimerise via disulfide bonds to form active ligands. Furthermore, the prodomain can be cleaved from the mature domain through proteolysis. GDNF family members are cleaved by furin, with a furin site between the pro and mature domains. Prodomains are known to alter TGF-β protein properties including reducing bioactivity and enhancing half-life.

To determine the impact of the prodomain on signalling, the different protein forms were evaluated using cell-based bioactivity assays. Both uncleaved pro-NRTN and pro-GDF15 have reduced RET signalling activity in comparison to their mature domains alone. However, cleaved pro-NRTN becomes as active as mature NRTN. Pro-GDF15 cleavage resulted in a hemi-cleaved protein – with only one prodomain of the dimer cleaved from its mature domain. This hemi-cleaved form of GDF15 was also less active than mature GDF15. Finally, H202D is a common variant of GDF15 in humans. Here, we demonstrate the H202D variant has no impact on GDF15 bioactivity but can affect the detection of the protein by anti-GDF15 antibodies.

Structural characterisation through circular dichroism (CD) reveals both the NRTN and GDF15 prodomains to be largely unstructured, unlike other TGF-β proteins. The prodomain of GDF15 also interacts with heparan sulfate (HS) as demonstrated in gel shift experiments. This may explain the localisation of pro-GDF15 at the cell surface where HS is common. Utilising hydrogen deuterium exchange mass spectrometry (HDX-MS) no interaction between the

NRTN prodomain and mature domain could be identified, again unlike other TGF-β proteins. This work highlights the diversity of prodomain structure and function in the TGF-β superfamily and suggests a common theme of unstructured and inhibitory prodomains for the GDNF subfamily.

TABLE OF CONTENTS

Chapter 1 - Introduction

1.0 Cell Signalling

Cell signalling is a broad term used to describe the transfer of information between or within cells, resulting in a biological response – an essential process in biology. Signalling can be induced by biochemical and physiological stimuli including, ions, small molecules, gases, macromolecules, osmolarity, light, temperature, mechanical force etc. Different categories of cell signalling are broadly based on the range of the signal, from between cells to between individuals. Cell signalling categories include Intracrine, juxtracrine, autocrine, paracrine, and endocrine.

Intracrine signalling is within a single cell whereas, in autocrine signalling the signalling molecule is first secreted before acting upon the secreting cell. Juxtracrine or contact-dependent signalling occurs between adjacent cells through cell-cell contacts such as membrane interactions and gap junctions. Paracrine and endocrine signalling both require secreted signalling molecules. Paracrine signalling affects the local environment while endocrine signalling is more systemic using the circulatory system to spread and act on distant cells and tissues. Beyond signalling within an organism, some signalling molecules can act on other individuals after secretion such as pheromones. A classic example is bombykol, a potent small molecule pheromone released by female silkworm moths to attract males[1].

Typically, cell signalling typically requires the transfer of the signal from outside of the cell, across the cell membrane, to the inside of the cell. This process of signal transduction is facilitated by signalling receptors such as, membrane ion channels, G-protein coupled receptors and enzyme-linked receptors. Receptor activation, or deactivation, relays the signal into the cell, typically triggering a signalling cascade. Biochemical cascades can amplify the signal, e.g., one signalling molecule can activate many downstream effectors such as enzymes. Signalling results in a specific biological response including changes to gene expression. Activation of cell signalling is tightly regulated, and so cells also have mechanisms to inactivate signalling pathways.

Cell signalling systems are essential to complex life. Signalling processes regulate diverse aspects of biology including development, growth, homeostasis, repair, metabolism,

reproduction and more. Pathologies such as cancer, autoimmune, metabolic, and degenerative diseases are often due to alterations in cell signalling because of the essential role it plays in biology.

Growth factors are a large group of signalling molecules containing hormones and cytokines. Growth factors are usually secreted and can act in an autocrine, paracrine, or endocrine fashion. Key growth factor families in humans include the vascular endothelial growth factor (VEGF) family, the epidermal growth factor (EGF) family, the fibroblast growth factor (FGF) family, and the transforming growth factor beta (TGF-β) superfamily. In this thesis, I will focus specifically on two members of TGF-β superfamily and the impact of their propeptides on cell signalling.

1.1 Propeptides and Prodomains in Biology

Proteins across biology contain propeptides, from bacteria to humans. Several biological roles have been identified for propeptides including protein folding, localisation, inhibition, and regulation. In addition, alterations to propeptide function are often associated with disease, highlighting the importance of these domains in biology [2,3].

UniProt defines a protein propeptide as "part of a protein that is cleaved during maturation or activation. Once cleaved, a propeptide generally has no independent biological function." Additional restrictions of the definition include sequence length and stability. Smaller sequences of 2 to 4 amino acids produced by proconvertase cleavage that are usually rapidly degraded would not be considered propeptides. Propeptides are also termed prodomains in the literature, for this thesis they will be considered synonymous. Lastly, prodomains can be N- or C-terminal as well as within protein sequences.

The classic human example of a propeptide containing protein is insulin. This prohormone is produced as a single peptide, proinsulin. Proinsulin is cross-linked by disulfide bonds and a central sequence, known as the C-peptide, is cleaved from within the protein producing mature insulin. The C-peptide is known to be biologically active, however its physiological role has yet to be defined[4].

The structural protein collagen ((alpha-1(I) chain)) has both N and C-terminal propeptides. Procollagen propeptide mutations are known to be detrimental, affecting processing and leading to skeletal abnormalities – this is thought be due to defects in collagen assembly [5].

Several proteases, such as bacterial subtilisin, possess a propeptide. The subtilisin propeptide functions to inhibit activity and regulate activation of the protease. Subtilisin protein folding is also aided by its propeptide[2]. Finally, the TGF-β (Transforming Growth Factor-β) superfamily members of signalling proteins also have propeptides which influence the regulation and control of cell signalling[6].

1.2 Growth Factors and the TGF- β Superfamily

Growth factors stimulate cells to proliferate and grow. However, most growth factors have further functions beyond this. One important family of growth factors is the transforming growth factor-β (TGF-β) Superfamily. These extracellular growth factors mediate a wide range of biological processes, playing key roles in homeostasis, development, the immune system, cancer and more[7]. There are over 30 TGF-β superfamily members in humans, which can be further categorised into subgroups based on similarity of function and clustering based on sequence conservation (figure 1.1).

Subgroups of the TGF-β superfamily include bone morphogenetic proteins (BMPs), the activins/inhibins, TGF-β isoforms (TGF-β 1,2 and 3), growth and differentiation factors (GDFs), and the glial-derived neurotrophic factors (GDNFs). Some family members are not included in any subgroups, such as nodal and anti-müllerian hormone (AMH)[8]. In this thesis, I will be looking at two members of the GDNF subfamily – GDF15 and NRTN.

The wide range of pathologies and developmental disorders linked to TGF-β proteins highlights the diverse and important range of biological roles of superfamily members. Myostatin (also called GDF8) is a negative regulator of muscle growth, with loss of function mutations leading to dramatically enhanced muscle mass [9]. Activin A and B play key roles in the development of the reproductive system with knockout experiments leading to infertility in female mice[10]. Clinically, several human skeletal disorders are known to arise from mutations in BMP proteins and alteration of signalling pathways affecting bone, cartilage and joint biology[11].

Many growth factors play multiple roles, such as glial-derived neurotrophic factor (GDNF). GDNF is an important neurotrophic factor for the development and maintenance of neuronal cells. The growth factor is also essential in renal development, as GDNF knockout mice fail to develop kidneys and an enteric nervous system [12,13]. Beyond development, TGF-β proteins and pathways are also implicated in several cancers and have become targets for therapeutic interventions[14].

Beyond the cytoplasm the oxidative environment facilitates the formation of disulfide bonds between protein cysteine residues. Disulfide bonds have been exploited by evolution to stabilise protein folds and cross-link components. Mature TGF-β family ligands are disulfide-linked dimeric proteins. Structurally, they share a characteristic cystine-knot motif that links two protomeric subunits via an inter-chain disulfide bond. In addition, each protomer has at least three intra-chain disulfides. Together, the disulfide bonds form a highly conserved cystine-knot motif and stabilise the shared butterfly-shaped structures of TGF-β protein mature domains (see figure 1.1B and 1.2). Activins, TGF-β-1 to 3, GDF11, GDF8 (myostatin) and GDF15 have a further two cysteine residues at the N-terminus of the mature domain, forming an extra intra-chain disulfide (see figure 1.1B and 1.2A)[8]. The term cystine-knot was originally coined in 1991 from a high resolution structure of nerve growth factor (NGF) and was used to describe the novel fold identified[15].

Full-length TGF-β proteins are comprised of an N-terminal prodomain followed by a C-terminal mature domain. As they are secreted proteins, an N-terminal signal peptide is also encoded, but it removed during synthesis. Disulfide bonds can also be seen in the prodomains of certain family members, such as TGF-β-1. The two prodomains of TGF-β-1 are connected by an inter-chain disulfide bond, linking two beta-hairpin regions to form a bow-tie-like structure (discussed in more detail later, and shown in figure 1.4C)[16].

The highly conserved cystine-knot motif is also shared by other proteins: the DAN family of BMP antagonists, the glycoprotein hormone family (GPH), neurotrophic factors and the platelet-derived growth factor (PDGF) family including the VEGF subfamily[17]. Furthermore, structurally related cystine-knot domains can also be found in the multi-domain CCN family of proteins[18].

The structures of the TGF-β mature domains are often described as ‘hand-in-hand’, with the palms of each hand forming the dimeric interface. The wrist, knuckle and finger regions are

solvent exposed creating the binding sites for appropriate receptors. Receptor binding sites of the mature domains can also be subdivided into concave and convex surface interactions (see figure 1.2)[8].

Homodimerisation is most common in the TGF-β superfamily however, several heterodimeric proteins have been reported in the literature with biologically relevant examples known for inhibin and BMP proteins. Heterodimerisation of TGF-β proteins can alter activity, signalling and roles of family members as well as increasing the diversity of possible ligands[19,20]. For example, Heterodimeric BMPs have been isolated and possess enhanced activity[21]. These heteromeric BMP ligands, can assemble signalling complexes of different type I receptors and have been shown to play a role in zebrafish development[22].

Most TGF-β family members signal though transmembrane type I and type II receptors. Ligand binding induces the formation of a heteromeric signalling complex of two type I and two type II transmembrane receptors bound to a single dimeric growth factor. Extracellular complex formation and conformational changes lead to intracellular receptor serine/threonine kinase domain phosphorylation, and activation. Once active these kinase receptors signal through intracellular mediators, such as Smads, leading to changes in gene expression[23].

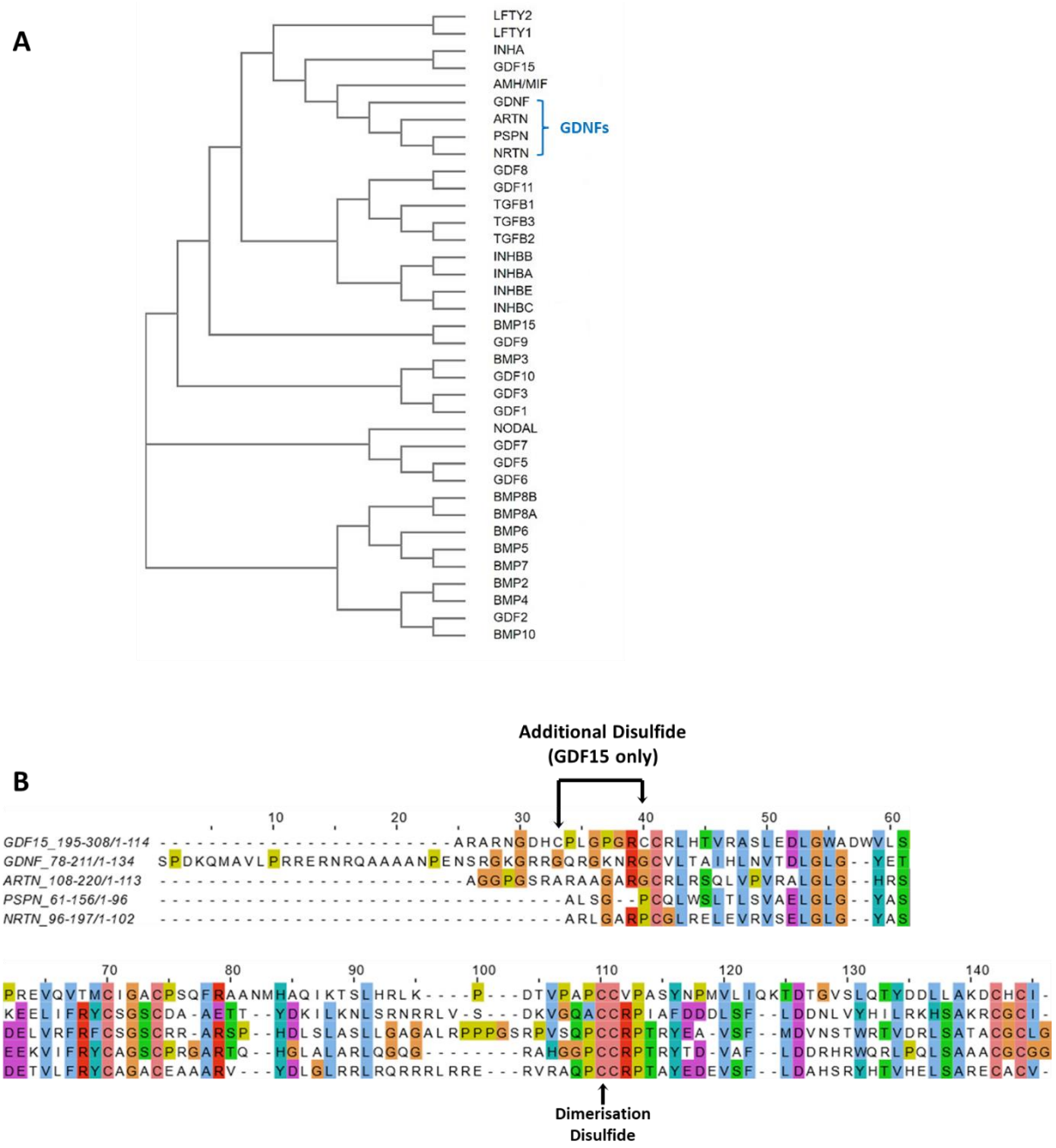


Figure 1.1. Sequence analysis of the TGF-β superfamily and the GDNF subfamily. A) Phylogenetic tree of human TGF-β superfamily members. Cladogram generated from multiple sequence alignment of the TGF-β superfamily (full-length protein sequences, neighbour-joining method without distance corrections using MUSCLE[24] (available at https://www.ebi.ac.uk/Tools/msa/muscle/) with the GDNF subfamily highlighted. B) Sequence alignment of GDNF subfamily members, including GDF15. Protein sequences for mature growth factors aligned in MUSCLE and visualised using Jalview[25].

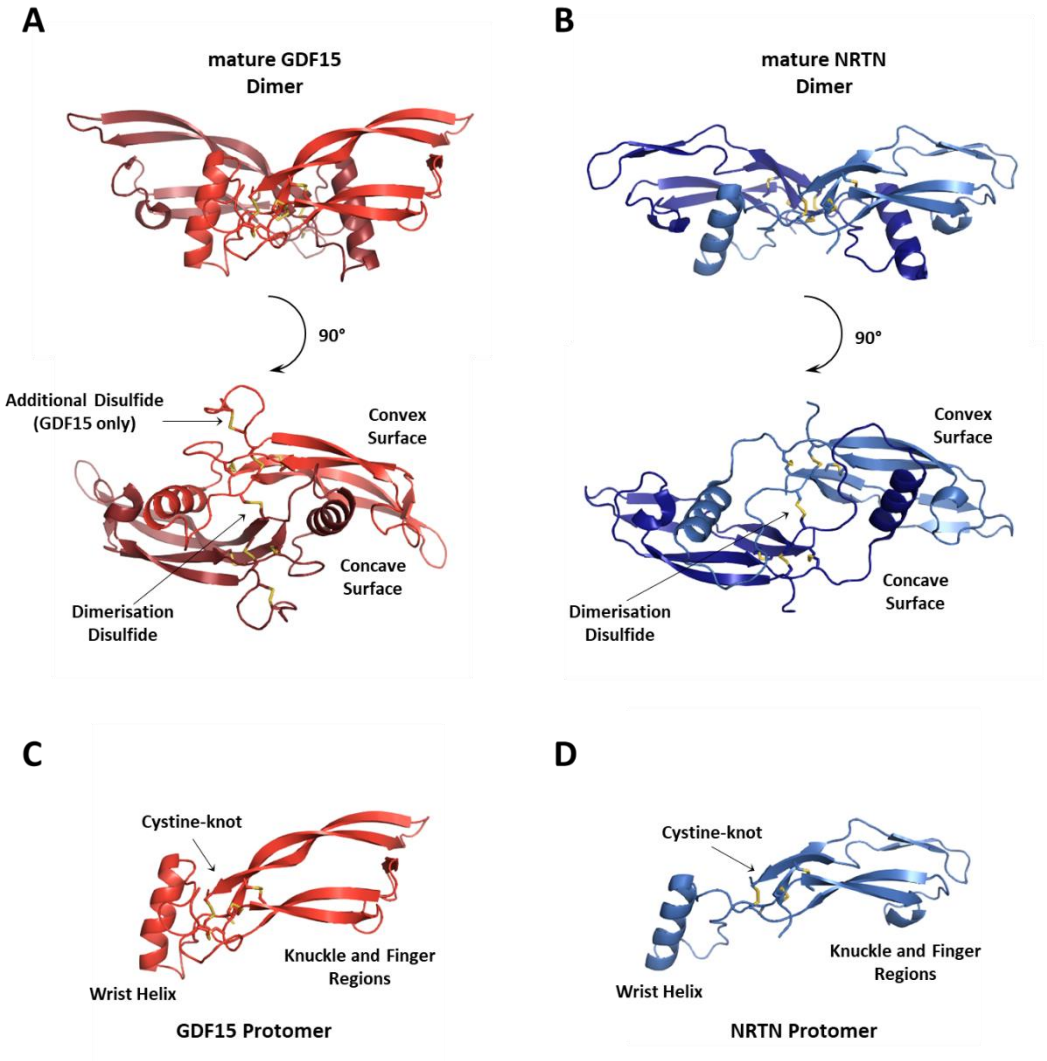


Figure 1.2. Structures of GDF15 and NRTN. X-ray crystal structures of mature GDF15 and NRTN proteins with structural elements indicated. A and B) Homodimers of mature GDF15 and mature NRTN crystal structures in different orientations. C and D) GDF15 and NRTN protomers. GDF15 shown in red and NRTN shown in blue, with different shades used for each protomer of the dimer and disulfide bridges shown in yellow. Figures generated using PyMOL (http://www.pymol.org), with structures from the Protein Data Bank (PDB codes – NRTN:5MR5 and GDF15:5VT2).

1.3 Biosynthesis and Processing of the TGF-β precursor

TGF-β superfamily proteins are synthesised as large precursors and processed before secretion. The precursor consists of an N-terminal signal peptide followed by a prodomain and a C-terminal mature domain (see figure 1.3). The mature domain is relatively small at approximately 100 amino acids and is the bioactive ligand of the protein. The prodomains vary in size but are typically larger at around 200 to 300 amino acids. The AMH prodomain is over 400 amino acids whereas members of the GDNF subfamily have atypically small prodomains under 100 amino acids. The precursor protein is translocated into the endoplasmic reticulum (ER) during translation directed by the signal peptide. The signal peptide is then removed via cleavage and the precursor protein dimerises through disulfide bonds, covalently linking the two mature domains[8,26]. The dimeric precursor then follows the secretory pathway, passing through the trans-Golgi network (TGN) before being secreted by the cell[27].

Many TGF-β superfamily precursors are known to be biologically inactive, with further processing required for activation. Proteolytic cleavage of the N-terminal prodomain can produce a bioactive ligand, leading to initiation of the relevant signalling pathway (see figure 1.3). TGF-β superfamily members are cleaved by members of the pro-protein convertase subtilisin/kexin (PCSK) family such as furin (PCSK3). The PCSK family of serine endoproteases are calcium-dependent, facilitating maturation of many proproteins through the cleavage at specific dibasic motifs within their substrates. For example, the furin recognition site is R-X-X-R, with X being any amino acid. However, a R-X-K/R-R site is preferred by the enzyme. Secondary structure also influences furin cleavage, with flexible and exposed regions being more susceptible to cleavage[28,29]. In general, prodomain processing occurs as part of the secretary pathway within the TGN[27] but extracellular cleavage by secreted proteases is known for myostatin and nodal[30,31].

TGF-β prodomains typically remain associated to mature domains post cleavage and this is often referred to as the pro-complex[26] - as is the case for activin A, myostatin and others[32,33]. Further processing steps are required for certain TGF-β superfamily proteins. For example, cleavage of myostatin and GDF11 by furin is necessary but not sufficient for full activation and release of mature growth factor. Both proteins also contain the recognition site for the bone morphogenetic protein 1/tolloid (BMP1/TLD) family of metalloproteases[34,35]. Additional

cleavage within the myostatin prodomain by tolloid releases the bioactive ligand through destabilisation of the pro-protein complex[36].

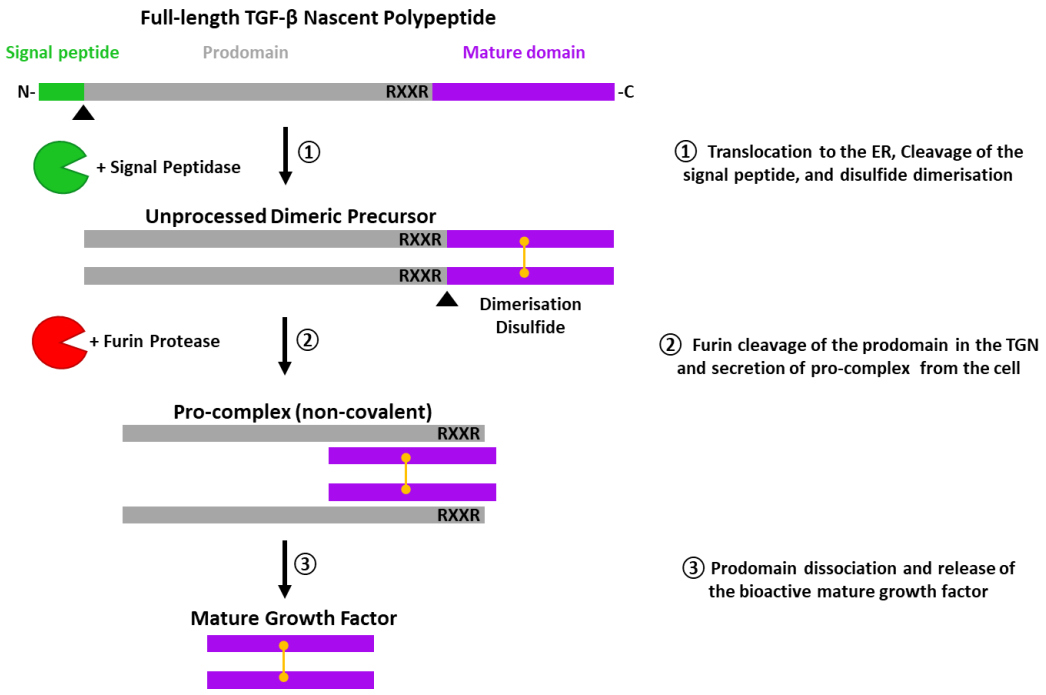


Figure 1.3 Overview of TGF-β superfamily proteins biosynthesis and maturation. Schematic depicting the biosynthesis, maturation, and secretion for the typical TGF-β protein. TGF-β superfamily members are synthesised as large precursor proteins. Precursor proteins are then further processed by protease enzymes, within the endoplasmic reticulum (ER) and trans-Golgi network (TGN), as the traffic through the secretory pathway. The nascent polypeptide chain for a typical TGF-β protein consists of an N-terminal signal sequence (shown in green), followed by a prodomain (shown in grey) and a C-terminal mature domain (shown in purple). Cleavage and processing are performed by two (or more) enzymes, a signal peptidase (shown in green) and furin protease (shown in red). The furin recognition site is annotated as RXXR with X being any amino acid. Mature domains dimerise via an inter-chain disulfide bridge, shown here in yellow. Pro-complexes between pro and mature domains post cleavage are common, associated via non-covalent interactions. Dissociation and release of the active growth factor (the mature disulfide-linked dimer) may be spontaneous or require further mechanisms of activation.

1.4 TGF-β Prodomain Roles

Originally TGF-β prodomains were identified to aid folding and secretion of the mature active growth factors. Early experiments demonstrated the prodomains of both activin A and TGF-β1 to be required for folding, assembly and secretion of the growth factors from cells[37]. Today, TGF-β family prodomains are known to have a wide range of functions in biology. Prodomains in the TGF-β superfamily are diverse in sequence, size, and structure as well as functionality. The functional role of the prodomain for several family members has been well characterised. Common themes of TGF-β prodomain function include aiding protein folding and secretion, inhibition and regulation, localisation, and storage[6]. Similar functions of prodomains are seen across many proteins and processes from protease inhibition, regulation of the coagulation cascade proteins and activation of the immune system in the complement cascade[38,39].

To date four crystal structures of TGF-β protein precursors have been determined, pro-myostatin, pro-activin A, pro-BMP9 and pro-TGF-β1 (see figure 1.4)[16,32,33,40]. Each precursor shown has a distinct prodomain sequence and architecture, however common elements such as the arm domain and forearm domain can be identified. These prodomains interact with key regions of the mature domains occluding receptor binding sites. BMPs 4,5,7,9,10 and GDF5 form the same non-covalent pro-complexes[41].

Prodomains of both TGF-β1 and myostatin are inhibitory, preventing cell signalling and activation of the relevant receptors. Both prodomains remain associated post cleavage with myostatin requiring cleavage by two enzymes for release and activation of the mature growth factor as previously described[16,32]. The TGF-β1 prodomain has an additional 'bow-tie' motif that dimerises the prodomain through a disulfide bridge. The dimeric prodomain encases the TGF-β1 mature growth factor in a straitjacket-like conformation, forming a latent complex (see figure 1.4C). This latent complex of pro-TGF-β1 then requires mechanical force for removal of its prodomains and subsequent activation. Force may be generated between cells or within the same cell via the cytoskeleton[16,42].

Mechanical force and release of the TGF-β1 mature growth factor is achieved through additional interaction between the prodomain and other proteins. Together these interacting partners stabilise, localise, and store the pro-TGF-β1 latent pro-complex at the extra cellular

matrix ready for activation through force. The arm domain, at the top of the complex, binds to specific integrins such as αvβ6. Integrin binding is facilitated via an RGD motif in each prodomain. At the base of the complex the α1-helix of the forearm domain is connected to LTBPs though disulfide bridges. LTBPs are connected to cell ECM and allow the generation of a resistant counterforce[16].

Alternative binding partners for the forearm domain are known, such as the membrane bound protein GARP (glycoprotein A repetitions predominant). Similarly, GARP binds through disulfide bonds as well non-covalent interactions which present and orient the pro-complex for activation[43]. Further studies demonstrate integrin-mediated conformational changes may permit activation and signalling without the complete dissociation of the prodomain[44].

The myostatin prodomain makes extensive interactions with the mature domain blocking key receptor binding sites of the protein (see figure 1.4 A). The uncleaved pro-myostatin precursor has no signalling activity and once cleaved, the prodomain is known to remain associated to the mature domain (pro-myostatin complex). This pro-myostatin complex is active but still displays reduced activity when compared to the mature active growth factor alone[32,45].

Similarly, to myostatin, the pro-activin A precursor has no signalling activity and the prodomain is known to remain associated to the mature domain post cleavage (pro-activin A complex). However, once cleaved, the pro-activin A complex has signalling activity comparable to mature activin A. Additionally, the prodomain can be displaced by other activin A binding partners, such as the antagonist protein follistatin (FST)[33]. In contrast, the pro-myostatin complex is resistant to FST binding[32].

The pro-BMP9 complex also signals with comparable activity to the mature ligand, with the prodomain either dissociating or being displaced by receptor binding. The prodomain of BMP9 does not display any dimerisation interactions in the complex structure, which may explain a lower level of affinity for the mature domain (see figure 1.4D)[40].

BMP9 circulates *in vivo* as pro-complex, associated with its mature domain[46]. Activin receptor-like kinase 1 (ALK1) is a type I receptor and binds BMP9. The BMP9 prodomains remains bound to the mature domain upon ALK1 receptor binding however, the α5-helix of the prodomain is displaced to accommodate the receptor (See figure 1.4D and E). In contrast, the

BMP9 prodomain is completely displaced by the regulatory co-receptor endoglin. Finally, pro-BMP9 binds ALK1 with similar affinity as mature BMP9[47]. Notably, not all TGF-β proteins form stable pro-complexes between pro and mature domains. The BMP2 prodomain is known to dissociate following furin cleavage, with a low affinity for the mature growth factor [41,48].

Prodomains have also been found to impact the stability and localisation of certain growth factors *in vivo* using animal models. Studies with pro-forms of TGF-β1 and activin A have been shown to have an increased the half-life of the proteins in the circulation, compared to mature proteins[49,50]. In addition, pro-TGF-β1 displayed an altered tissue distribution to the mature protein[49].

Prodomain mutations may also impact the function of growth factors. Anti-Müllerian hormone (AMH) has a relatively large N-terminal prodomain of over 400 amino acids. Post cleavage AMH remains bound non-covalently between n-prodomain and c-terminal mature domains (referred to as $AMH_{N,C}$). Cleavage is required for receptor activation, with the mature domain being receptor binding. Although, prodomain presence increases bioactivity, this is thought to be due to improved solubility and diffusion[51]. Several AMH prodomain mutations are known inhibit the developmental role of Müllerian duct regression in males. Mutations lead to a form of pseudo-hermaphroditism termed persistent Müllerian duct syndrome 1 (PMDS1). Mutations in the AMH mature domain, n-terminal signal sequence and the AMH receptor are also known to cause PMDS[52].

Prodomain roles for the GDNF subfamily remain largely unknown. However, the prodomain of GDF15 may have a role in protein localisation. Pro-GDF15 has been observed to remain at the extracellular matrix at the surface of cells in contrast to the mature form, suggesting a novel interaction partner for the prodomain. Additionally, the rapid increase of GDF15 levels under certain stress conditions suggests a storage and release mechanism which may be facilitated by the prodomain[53,54] – as it the case for TGF-β1.

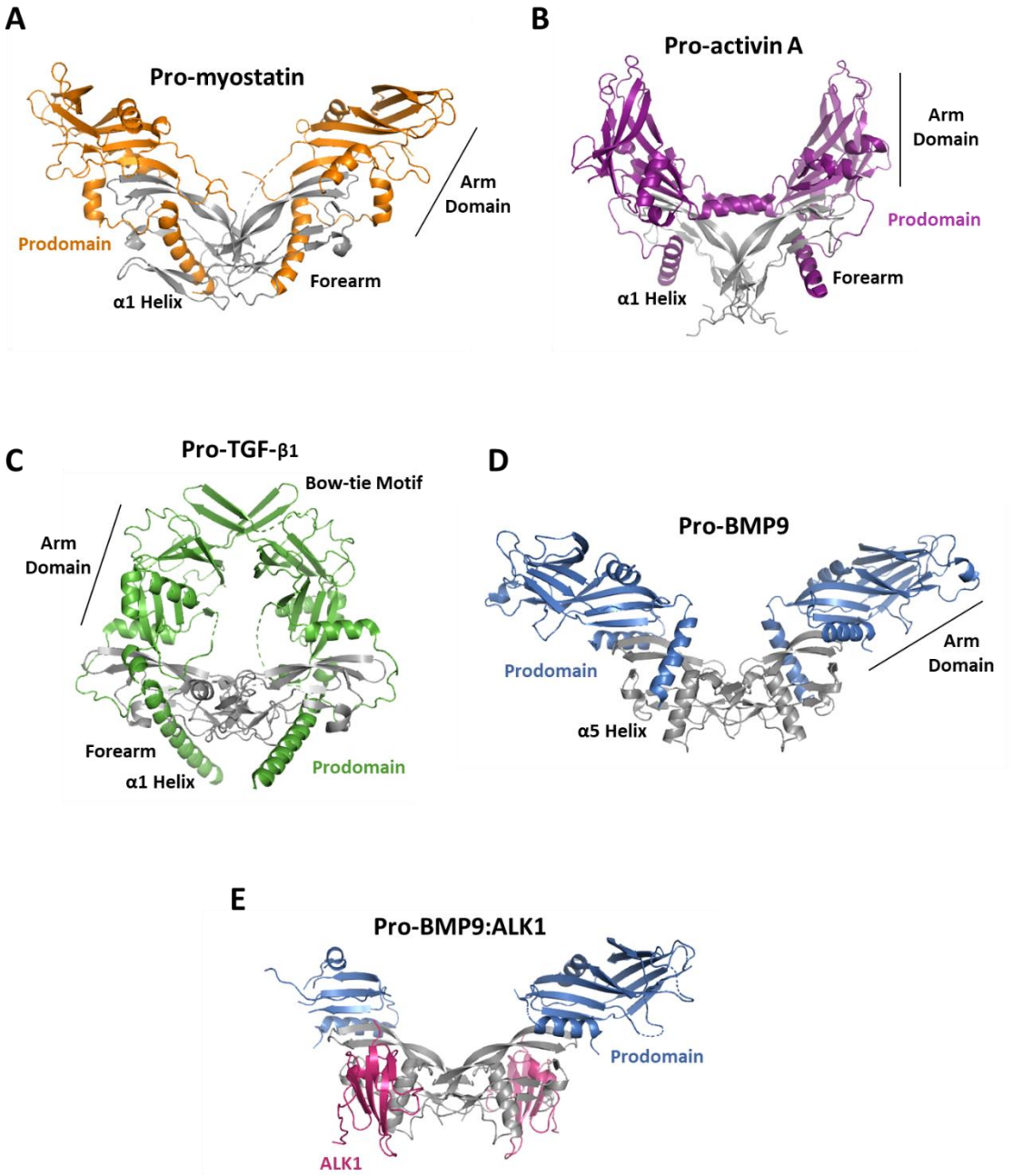


Figure 1.4. Structures of pro-TGF-β superfamily members. X-ray crystal structures of pro-TGF-β superfamily members, with key structural features of the prodomains annotated. The mature domains are shown in grey with the prodomains shown in various colours as indicated. The prodomains make extended interactions with the mature growth factor via N-terminal forearm domains and C-terminal arm domains. A) Pro-myostatin, PDB code: 5NTU, B) Pro-activin A, PDB code:5HLY, C) Pro-TGF-β1, PDB code: 3RJR D) Pro-BMP9, PDB code:4YCG and E) Pro-BMP9 with ALK1 receptor (shown in pink), PDB code:6SF2. The C-terminal alpha-5 helix of the prodomain is missing in the pro-BMP9:ALK1 structure. Structures from the Protein Data Bank.

1.5 The Extracellular Matrix and TGF-β Biology

The extracellular matrix (ECM) surrounds cells and is composed of many proteins, water, and polysaccharides. The ECM plays an essential role in scaffolding, homeostasis, and differentiation. Several components of the ECM are relevant to TGF-β biology including integrins, fibrillins and heparan sulfate (HS). The ECM plays a key role in growth factor storage, release, diffusion, and signalling[55]. Examples include BMP-4/5/7/10 fibrillin association at the ECM, which aids in the localisation and concentration of the growth factors at target cells and TGF-β1-ECM protein interactions for storage and release as previously described[16,56].

Approximately one third of TGF-β superfamily members bind heparin or HS oligosaccharides - including TGF-β1, TGF-β2, certain BMPs, GDFs and GDNFs. Many TGF-β antagonists also interact with HS, such as FST. Length and sulfation patterns of HS molecules are highly heterogeneous, and some protein interactions can be highly specific for certain modifications. For example, GDNF binding is dependent on the presence of 2-O-sulfates[57]. ECM-growth factor interactions can also aid signalling, by localising and concentrating ligands at the cell surface. NRTN is known to bind with HS, this interaction has also be proposed to play a role in signalling[58].

Manipulation of growth factor-ECM component interactions has been pursued to improve the therapeutic potential of growth factors. For example, in the treatment of Parkinson's disease neurotrophic factors are of keen interest. However, limitations in their effectiveness are thought to due to extensive ECM interactions and limited diffusion in the brain. To overcome this issue, Runeberg-Roos *et al.,* 2016 modified NRTN to reduce heparan sulfate binding in the ECM. They went on to demonstrate the modified NRTN had improved distribution in brain tissue[59]. Conversely, the addition of heparan-sulfate-binding sites to VEGF and PDGF has been shown to enhance their tissue regeneration properties in mice, by permitting controlled and sustained release from artificial scaffolds[60].

1.6 GDNF Subfamily and Signalling

1.61 GDNF subfamily members

Neurturin (NRTN) was first identified in 1996 for its ability to support the survival of neuronal cells in culture[61]. For this reason, and due to structural similarities, NRTN was then grouped with another neurotrophic factor named glial-cell-line-derived neurotrophic factor (GDNF) forming the GDNF subfamily of the TGF-β superfamily of growth factors. The GDNF subfamily has since grown to include, persephin (PSPN), artemin (ARTN). Recently growth and differentiation factor 15 (GDF15) has been included in this subfamily as well, following demonstration that it uses similar receptors to other members of this subfamily [62–66]. To date, structures of all mature GDNF subfamily members have been solved, except PSPN. The mature GDNF subfamily growth factors share similar features as other TGF-β proteins as described earlier (see figure 1.6)[58,62–65,67,68]. The GDNF subfamily has a broad range of biological roles from development to regulation of metabolism and the immune system[69].

1.62 GDNF family Signalling Receptors

Unlike the majority of the TGF-β proteins, the GDNF subfamily do not signal through type I and type II receptors via serine-threonine kinases. Instead, GDNF subfamily members signal through the RET tyrosine kinase receptor. Signalling also requires the aid of a co-receptor, with each family member having a specific GDNF family receptor-α (GFRα) co-receptor[66]. In 2017 the orphan receptor glial-derived neurotrophic factor-family receptor α-like (GFRAL) was identified as the co-receptor for GDF15 by four groups independently. GDF15 was also determined to signal via the RET receptor, with the aid of GFRAL and therefore GDF15 can now be thought of as the fifth GDNF subfamily member, even if it is more divergent in sequence [62–65].

The five co-receptors, GFRα1 to 4 and GFRAL each have specificity for one of the GDNF family members. GDNF uses GFRα1, NRTN uses GFRα2, ARTN uses GFRα3, PSPN uses GFRα4, and GDF15 uses GFRAL (see figure 1.5). However, alternative interactions have been reported in the literature. The co-receptors are all modified with a glycosyl phosphatidylinositol (GPI) anchor linking them to the membrane, except GFRAL which has a transmembrane domain and short intracellular region. The co-receptors GFRα 1,2,3 and GFRAL have three cysteine-rich globular domains with GFRα4 only having two. Dimeric GDNF ligands bound to specific

co-receptors can form a signalling complex with RET receptor. RET is then activated and undergoes autophosphorylation. Phosphorylated RET can then interact with adaptor and signalling proteins, initiating signalling pathways and altering growth, metabolism and gene expression[66,69,70]. Alternative receptors for certain GDNF subfamily members are known. For example, GDNF can signal through neural cell adhesion molecule (NCAM) with GFRα1 and the heparan sulfate proteoglycan syndecan-3[71,72].

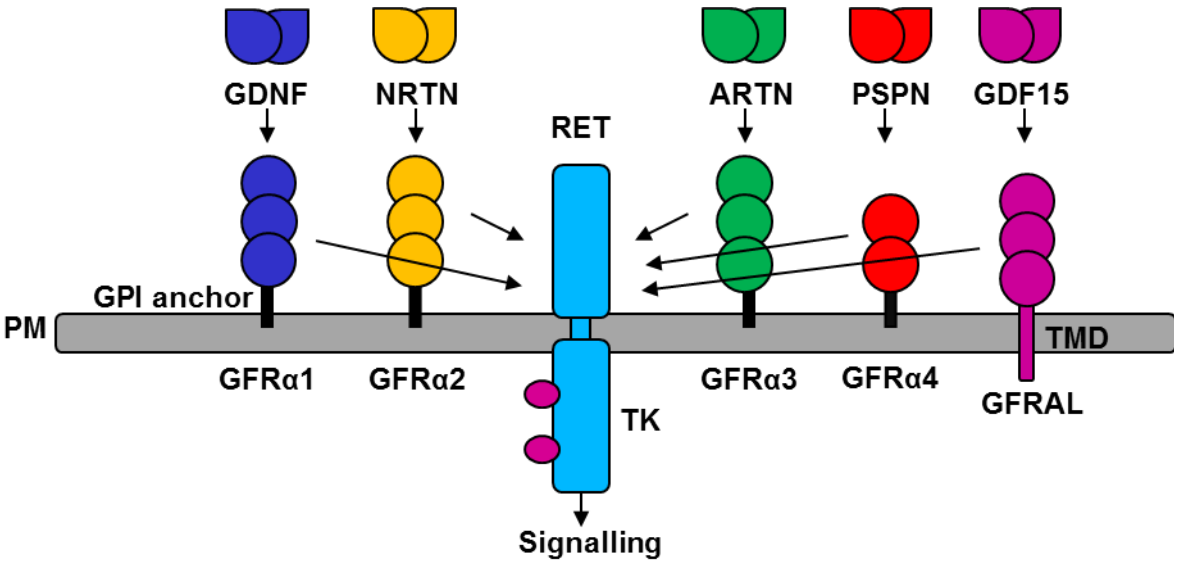


Figure 1.5. GDNF Subfamily Co-receptors and Signalling through RET. Simplified schematic of GDNF family signalling. Homodimeric GDNF subfamily ligands first bind to a preferred co-receptor of matching colour before recruiting and activating the RET receptor (cyan). Once activated the intracellular tyrosine kinase (TK) domain becomes phosphorylated (shown in pink) and the signal can then be transduced though the plasma membrane (PM). The GFRAL co-receptor is localised to the PM via a transmembrane domain (TMD), as is RET. The remaining GFRα co-receptors localised using glycosylphosphatidylinositol (GPI) anchors, modifications that interact with the PM.

The GDNF subfamily receptor RET plays an essential role in normal development and is expressed during embryogenesis. In adults RET expression has been detected in the brain, thyroid, kidneys, and lungs[69]. Phenotypes of mice that lack GDNF family ligands, co-receptors or the RET receptor vary. Knockouts of RET, GDNF and the GDNF co-receptor GFRα1 are all lethal and lead to breathing problems, neuronal abnormalities, and the absence of kidneys. Milder phenotypes are seen with other family members and their co-receptors[62–65,70].

1.63 RET Signalling Complex

The GDNF subfamily of ligands require the RET (rearranged during transfection) receptor, as well as a co-receptor for signalling activity. Originally identified as an oncogene, RET is a receptor tyrosine kinase (RTK). RET is a single-pass type I transmembrane signalling receptor with an intracellular tyrosine kinase (TK) domain and C-terminal tail[66]. Differential splicing of RET alters the length of the protein and generates three isoforms (RET9, 43, and 51)[69]. The extracellular N-terminal region of RET includes four cadherin-like domains (CLD1 to CLD4). RET requires calcium to fold correctly and function, with three calcium binding sites located between CLD2 and CLD3. A cysteine rich domain (CRD) is also present between CLD4 and the hydrophobic transmembrane domain[66]. Recent Cryo-EM structures of RET extracellular domain also identified the structure of the CRD as having a novel fold with an additional calcium binding site[73,74].

Evidence from AUC, Cryo-EM together with the crystal structure of CLD1-2 domains indicate RET to exist as an inactive dimer. Interactions between CLD1 and CLD2 domains from each monomer form a homodimer, holding RET in an inactive dimer with the intracellular kinase domains separated by the extracellular structure. Conformational changes upon ligand and co-receptor binding can then lead to activation of the RET receptor, by bringing the TK domains into proximity. The intracellular TK domains can then catalyse the phosphorylation of key tyrosine residues[73,75]. This model of inactive RET dimers could also apply to the surface of cells at the plasma membrane, although ligand and co-receptor based dimerisation has also been proposed[76].

The RET receptor forms a flexible C-shaped clamp surrounding both the mature growth factor and co-receptor creating a batwing-like architecture (see figure 1.6)[73]. The co-receptors second domain (D2) interacts with the growth factors' convex face knuckle epitope as seen in crystal and cryo-EM structures[58,63,67,68]. RET contacts both the growth factor and the co-receptor to form the active signalling complex. The membrane-proximal CRD contacts the growth factor bringing the C-terminal regions of RET into proximity. The co-receptor interacts with RET via its third domain (D3), interfacing with the N-terminal region of RET (CLD1 and CLD2 domains). These interactions share a common architecture among all four GDNF subfamily members analysed, however differences in specific interactions at each site can be seen. Additionally, the angle between the two RET receptors varies to accommodate different

ligands and co-receptors again demonstrating the flexibility of the RET receptor (See figure 1.6B). Finally, simultaneous binding between RET receptor, the co-receptor and growth factor explains why all three are required for signalling[73,77].

The RET extracellular signalling complex resembles other TGF-β family receptor complexes, each binding to similar convex and concave surfaces of mature growth factors as described earlier. Structural studies reveal RET binding to be equivalent to the concave type I receptor binding site in canonical TGF-β family ligands whilst the co-receptor binding site is equivalent to the convex type II receptor interaction[8,73].

Higher-order assemblies of RET complexes have also been observed, the NRTN/GFRα2/RET complex can further interact to form a tetrameric structure of two signalling complexes as seen by cryo-EM and cellular assays. These higher order assemblies may also play a role in regulation and endocytosis of the complex[73]. Similar multimeric organisations have also been reported for the GDNF/GFRα1/RET complex [74].

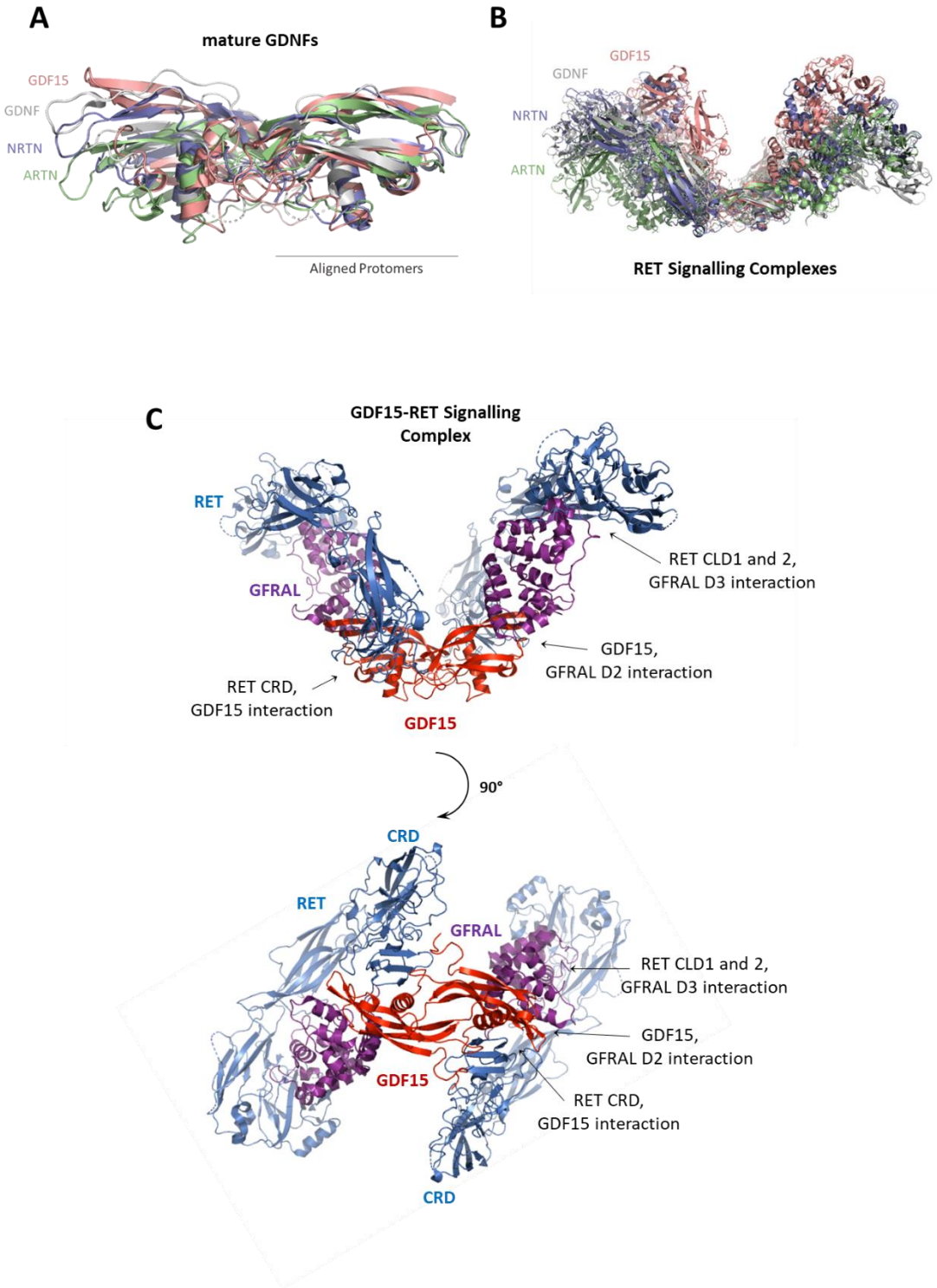


Figure 1.6. Cryo-EM structures of GDNF family member signalling complexes. A and B) protomer aligned cryo-EM structures of either mature GDNF family members or members in signalling complexes with RET and the relevant co-receptor, members shown in various colours as indicated. C) Cryo-EM structure of the GDF15/GFRAL/RET receptor signalling complex, with key interactions between all three components highlighted. The C-shaped clamp RET receptor shown in blue, the co-receptor GFRAL shown in purple and the dimeric

mature GDF15 ligand shown in red. Structures taken from the PDB – (PDB codes: 6Q2J, 6Q2N,6Q2O, 6Q2S)[73].

1.64 RET Signalling Pathways

Formation of the RET receptor signalling complex results in trans autophosphorylation of tyrosine residues within the RET intracellular domain. The intracellular domain has over 10 autophosphorylation sites with distinct downstream consequences. RET signalling can activate a variety of downstream signalling pathways and numerous interacting partners have been identified (see figure 1.7). As with all cell signalling pathways, signalling is cell type and context dependent[69]. Distinct regions of the membrane known as lipid rafts or microdomains also play a role in RET signalling by localising signalling components[78,79].

Once phosphorylated, intracellular RET tyrosine residues become recruitment sites for numerous adapter proteins, which in turn propagate intracellular signalling pathways. Differential binding preferences for each tyrosine residue are known, for example Y1015 is the docking site for phospholipase C. Other adaptor proteins include, Grb2/7/10, c-Src, Enigma, Shc, fibroblast growth factor receptor substrate 2 (FRS2), insulin receptor substrate 1/2 (IRS1/2) and downstream of tyrosine kinases (DOK1/4/5). Y1062 phosphorylation activates several downstream signalling pathways including the phosphatidylinositol-3-kinase (PI-3K)/AKT pathway, the c-Jun N-terminal kinase (JNK) pathway, the mitogen-activated protein kinase (MAPK) pathway[69].

1.65 Biology and Pathology of RET

RET signalling mediates a wide range of essential biological processes including cell survival, proliferation, differentiation, and metabolism. In development, RET plays a key role in the formation of the kidneys and nervous system. Given the importance of RET in development and its key role in signalling for several growth factors it is unsurprising that RET mutations are often associated with developmental disorders and pathologies including cancer. Numerous mutations and alterations in the RET gene, both gain and loss of function, have been identified with deleterious results[69] (examples listed in table 1).

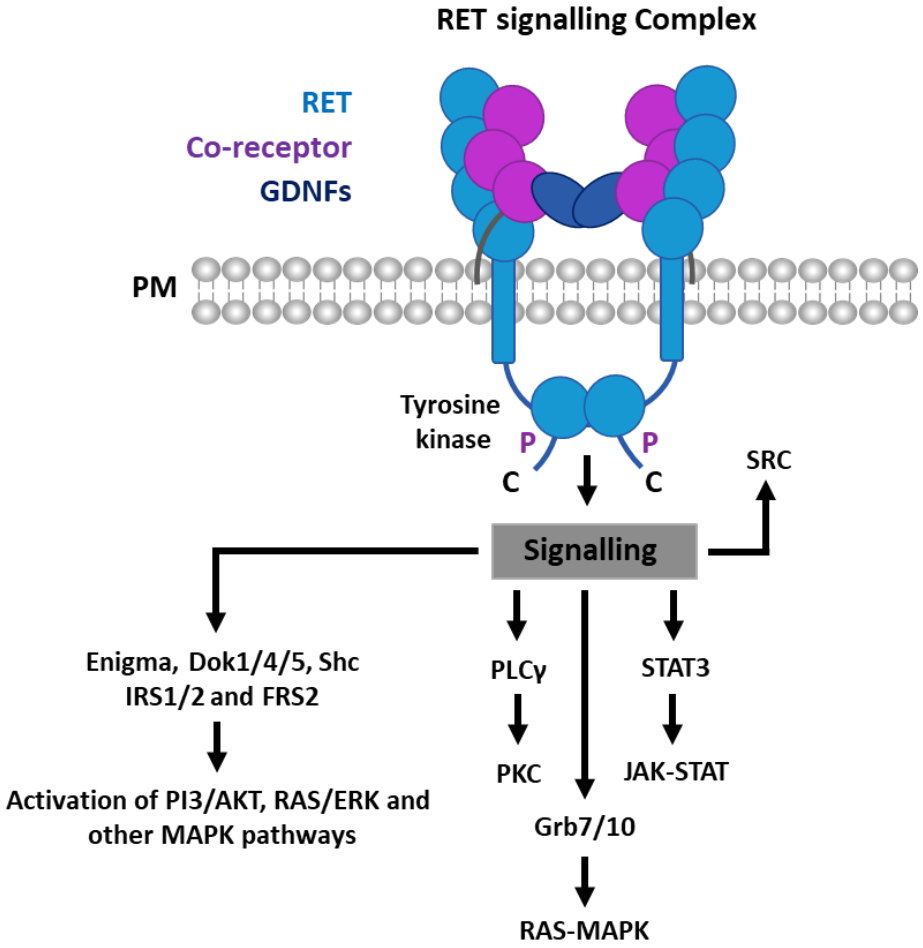


Figure 1.7. Activation intracellular signalling pathways by the RET receptor. The RET signalling complex is formed of two RET receptors, two co-receptors and a dimeric GDNF subfamily ligand. Interactions between all three components in the extracellular space activates the intracellular C-terminal tyrosine kinase domains of RET, transducing the signal through the plasma membrane (PM). RET receptor is then phosphorylated at tyrosine residues, as indicated by P. Phosphorylation enables docking of numerous adaptor proteins and subsequent activation of various downstream signalling pathways as indicated.

RET Associated Pathology	Abbreviation
Hirschsprung's Disease (Aganglionic Megacolon)	HSCR
Congenital Central Hypoventilation Syndrome	CCHS
Medullary Thyroid Carcinoma	MTC
Multiple Endocrine Neoplasia IIA	MEN2A
Multiple Endocrine Neoplasia IIB	MEN2B
non-small cell lung cancer	NSCLC
papillary thyroid carcinoma	PTC
Pheochromocytoma	-

Table 1. RET Mutation Associated Pathologies. Table of known pathologies associated with mutations and alteration of the RET receptor gene in humans. MEN2A and MEN2B are cancer syndromes. Source OMIM (available at: https://omim.org/entry/164761) and Li et al., 2019[80].

CDCC6-RET and NCOA4-RET are common examples of a RET fusion mutations in cancers. Rearrangements lead to the N-terminal region of one protein being fused to the RET C-terminal TK domain. Fusion of the RET gene results in aberrant activation of the TK domain and downstream signalling pathways leading to excessive growth and cancer[80]. Activating RET point mutations are also known to cause disease, for example M918T resulting in Multiple Endocrine Neoplasia IIB (MEN2B)[69].

Hirschsprung's Disease (HSCR) is also known as aganglionic megacolon and is characterised by the absence of specific neuronal cells in the gut, this often leads to intestinal symptoms. Studies indicate that defects in the neural crest development leads to the HSCR phenotype[81]. Defects in RET signalling are associated with the development of the disease. Inactivation mutations of RET are associated with HSCR in humans and downregulation of GFRα1, the GDNF co-receptor, has also been shown to lead to HSCR in animal models[69]. Additionally, NRTN mutations may contribute to HSCR when coupled with RET mutations[82].

Research into new therapeutics has targeted RET signalling for the treatment of cancers, neurodegeneration, and obesity. Inhibition of the intracellular TK domain has been the target

for cancer therapy using a small molecule approach[83]. Signalling activation has been the aim for neuroprotective treatments, using the natural ligands of the RET receptor such as GDNF and NRTN[69]. Obesity and neurodegeneration will be discussed in more detail later in relation to NRTN and GDF15.

1.7 Biology, Pathology and therapeutic potential of GDF15 and NRTN

NRTN was originally identified as a neurotrophic factor and is associated with the normal development and maintenance of the nervous system[61,84]. Additionally, NRTN has positive effects on hyperglycaemia in diabetic animal models[85]. GDF15 has key roles in the regulation of metabolism and weight, with associations to several pathologies including cancer, heart failure, anorexia, and severe nausea[86].

1.71 NRTN, GDNF and Neurodegenerative Disease

The neurotrophic factors NRTN and GDNF were both originally discovered for their ability to support the survival of neurons cells in culture, including dopaminergic (DA) neurones[61,87]. This ability has made them attractive therapeutic agents in the treatment of neurodegenerative diseases such as Parkinson's Disease (PD). PD is characterised by a loss of DA neurones in the brain, leading to motor impairments and dysfunction. Currently, treatments are available for PD but there is no cure[88]. Many studies have researched NRTN and GDNF as a novel therapy for PD, with mixed results[88,89]. Initial animal studies yielded positive results in both protection and regeneration of neurones in PD[90–92]. However, clinical trials in human were less successful. Trials were carried out using recombinant protein and viral vectors as delivery mechanisms[89]. Lack of effectiveness, in clinical trials, may be due to limited diffusion of the growth factor. As mentioned previously diffusion may be modified by manipulation of ECM interactions, as has been demonstrated for several growth factors - including engineering of NRTN to reduce its binding to heparan sulfates. This may improve the use of growth factors in PD therapy[59]. Additionally, to avoid delivery issues, small-molecule approaches for RET activation are being pursued[89].

1.72 GDF15 in Cancer, Pregnancy and Nausea

GDF15 is a stress regulator of energy homeostasis and body weight and is associated with numerous pathological conditions including cancer, obesity, and anorexia[86]. Expression of the GFRAL co-receptor is restricted to specific regions of the brainstem, the area postrema and nucleus tractus solitarius[62–65]. GDF15 is expressed more widely at low levels in most tissues, with higher levels of expression at the placenta and prostate[93]. Elevated GDF15 levels are associated with many disease states, including several cancers[86]. Cancer patients can develop anorexia or cachexia syndrome, causing extreme weight loss and muscle wasting. Animal model studies demonstrate GDF15 to be the driver of cachexia, with anti-GDF15 antibody treatment preventing weight loss[94]. In pregnancy, higher levels of GDF15 are associated with nausea and vomiting and hyperemesis gravidarum (HG). HG is characterised by sever nausea, vomiting and weight loss[95]. Phenotypes seen in cancer cachexia and HG are consistent, highlighting the role of GDF15 in weight loss and nausea. This also raises the possibility for GDF15, and/or GFRAL antagonism as a novel treatment for these conditions [86].

1.73 GDF15 and Obesity

Conversely, GDF15 signalling induction has been proposed in the treatment of obesity and as mentioned previously, elevated levels of GDF15 are associated with weight loss in humans[86,94]. The orphan receptor GFRAL was recently identified as the co-receptor for GDF15[62–65]. Furthermore, GDF15 effects on body mass and food intake were found to be mediated by the GFRAL co-receptor, as GFRAL knockout mice do not respond to GDF15[62]. GDF15 mediated weight loss is largely due to a decrease in food intake[62]. Similarly, administration of recombinant GDF15 to non-human primates results in decreased obesity and food intake[64]. These results highlight the role of GDF15 and GFRAL in regulation of body weight and homeostasis[62–65].

1.74 GDF15 H202D Variant

Of interest is the histidine to aspartic acid substitution (H202D) of GDF15, a common variant of GDF15 in humans[86]. The H202D mutation is located in the mature domain of GDF15 and is also referred to as H6D in the literature (when using residue numbering starting from the N-terminus of the mature domain, after furin cleavage). Studies have implicated the variant in

cancer and suggest potential biological differences between the variant and WT protein, however this is disputed[86,96].

Clinical studies suggest the H202D GDF15 variant is associated with lower prostate cancer incidence in humans[97] but is also associated with more aggressive growth if cancer develops[98]. The role of the variant in prostate cancer is not fully understood. To investigate this association cancer cells modified to express either WT or H202D GDF15 were injected into mice. The variant was reported to be more potent, significantly inhibiting tumour growth compared to the WT control[96]. This then led to the hypothesis that the H202D variant is more potent than WT[96] however, the signalling activity of the variant though RET and GFRAL has not been formally tested. Additionally, reported potency differences may be due to variable detection rates between the two forms of GDF15 and/or differences in tumour sizes between mice – raised by in Lockhart *et al.,* 2020[86].

1.75 Protective and Immunomodulatory Effects of GDF15

More broadly, studies utilising injury or stress-induced inflammation models have demonstrated GDF15 to provide a protective effect in multiple organ systems, regulating inflammation, immune cell infiltration and expansion. Early studies in the literature also claim GDF15 to be a TGF-β-like ligand, however this is likely due to TGF-β protein contamination in samples tested[86].

In the liver, GDF15 has been shown to have an anti-inflammatory role. For mouse models of liver fibrosis, recombinant GDF15 decreases the expression of pro-inflammatory cytokines and fibrotic mediators such as TNF-α and type 1 collagen 1 alpha 1[99]. In the hearts of both mice and humans GD15 expression is induced post-myocardial infarction. Furthermore, the anti-inflammatory properties of GDF15 are required for myocardial infarction survival with GDF15 knockout mice displaying a higher incidence of cardiac ruptures[100]. Additionally, over-expression of GDF15 has also been shown to be protective of heart tissue in transplant experiments in mice[101]. Similar findings have been seen in kidney transplantations, with low circulating levels of GDF15 being associated with organ rejection. Lastly, GDF15 deficiency exacerbates induced kidney injuries and enhances inflammation in mice, consistent with GDF15 having a protective effect on kidney tissue[102,103]. Together, these studies highlight the protective and immunomodulatory effects of GDF15.

1.76 GDF15 and NRTN in immunology & Infection

Beyond metabolism and weight, GDF15 is also implicated in immune modulation. GDF15 is induced in inflammatory states with increases in expression and serum levels detectable[104]. GDF15 was originally identified as an immunosuppressive factor, and termed macrophage inhibitory cytokine 1 (MIC-1)[105]. Additionally, GDF15 expression was also found to be increased by nonsteroidal anti-inflammatory drugs and was also designated as NSAID activated gene (NAG-1) with antitumorigenic and proapoptotic properties[106]. In severe inflammatory states such as sepsis GDF15 stimulates production of protective triglycerides, promoting immune tolerance and survival[104]. Interestingly, NRTN has also been highlighted to have an immunomodulatory role, regulating the macrophage cells of the lung during viral infection[107].

1.77 Diabetes, GDF15 and NRTN

Both mature GDF15 and NRTN have been examined in relation to type 2 diabetes[85,108,109]. NRTN has been demonstrated to have anti-diabetic activity, preventing hyperglycaemia in Zucker diabetic fatty (ZDF) rats, without reducing body mass or appetite[85]. ZDF rats are commonly used as a model for diabetes as they develop hyperglycaemia and hyperinsulinemia. When treating hyperglycaemia, NRTN alone was shown to be less effective than the anti-diabetic drug liraglutide[85]. The anti-diabetic drug, liraglutide is a glucagon-like peptide-1 (GLP-1) receptor agonist, increases insulin levels and decreases glucagon release[110]. However, co-administration of mature NRTN with liraglutide was shown to normalise hyperglycaemia in ZDF rats – highlighting the potential of combination therapies[85]. For GDF15, two groups individually demonstrated increased levels of the circulating GF in response to metformin, another anti-diabetic drug. Both studies indicate the therapeutic benefits of metformin are due to GDF15 including reduction in appetite and body mass[108,109].

1.8 Other Neurotrophic Factors and Pro-Forms

Beyond the TGF-β superfamily other neurotrophic factors with relevant pro-forms are known. For example, brain-derived neurotrophic factor (BDNF) and nerve growth factor (NGF) both have prodomains that aid secretion of the proteins[111]. The prodomain of BDNF has been demonstrated to interact with the mature domain at nanomolar affinity[112]. For NGF its prodomain is intrinsically disordered in structure and does not appear to make well defined

interactions with the mature domain – instead forming transient intra-molecular contacts[113]. Pro-forms of BDNF and NGF promote apoptosis through p75 and sortilin receptors, whereas mature proteins signal through either TrkA and B tyrosine kinase receptors[114,115]. The prodomain of NGF has been shown to be independently biologically active, inducing growth cone collapse in neuronal cells, however the mechanism of action is not yet fully understood[113].

Aims and Objectives

The main aim of this thesis is to characterise pro-GDF15 and pro-NRTN and determine the impact of the prodomain on signalling activity. Additionally, for GDF15, a H202D variant will be characterised in terms of signalling activity and impact on GDF15 protein quantification via ELISA. These aims can be divided into the following objectives:

Objective 1 - Production of GDF15 and NRTN pro-forms, mature forms and prodomains

Mature, pro-forms and prodomains alone will be produced recombinantly using bacterial expression systems and refolding where appropriate. Proteins will be purified to suitable levels of homogeneity and purity for characterisation of activity and interactions.

Objective 2 - Characterisation of pro-GDF15 and pro-NRTN signalling activity

The bioactivity of both growth factors will be tested using luciferase reporter-gene based cell-signalling assays via appropriate receptors and RET signalling. Mature and pro-forms of both growth factors will be tested.

Objective 3 - Characterisation of pro-NRTN and pro-GDF15 cleavage by furin and TEV

Furin cleavage by for both growth factors will be characterised, and engineered TEV protease cleavable constructs will be cleaved and tested for signalling activity.

Objective 4 - Structural characterisation of NRTN and GDF15 prodomains and exploration of interaction partners

The secondary structure of the prodomains will be characterised by circular dichroism (CD). Interactions between mature and prodomains will be assessed using biophysical techniques. Additionally, for the GDF15 prodomain initial experiments will be performed to examine a putative interaction with heparan sulfate-like molecules using CD and gel-shift based methods.

Objective 5 - Production and characterisation of the GDF15 H202D variant

The H202D GDF15 variant will be produced in the same way as the WT mature GDF15 as a homodimer. Furthermore, a heterozygous mix of homodimeric and heterodimeric GDF15 WT/H202D protomers will be produced to mimic heterozygous individuals. The bioactivity

will then be compared to WT using the same cell-signalling assays as previous. Enzyme-linked immunosorbent assay (ELISA) based methods can then be used to determine the impact of the variant on GDF15 quantification.

Completion of these aims and objectives will provide insight into the role of the prodomains and the H202D variant for GDNF family members NRTN and GDF15.

Chapter 2 - Materials and Methods

2.1 Cloning

Sequences corresponding to human NRTN and GDF15 amino acids were cloned into various protein expression vectors (see Chapters 3 and 4). Restriction enzymes, *Hind*III (NEB) and *Nco*I (NEB) were used to linearise vectors. To reduce self-ligation, linearised vectors were also treated with Shrimp Alkaline Phosphatase (Thermo Fisher Scientific). Polymerase chain reaction (PCR) was carried out using Phusion DNA Polymerase (Hyvönen Lab) to generate inserts, in accordance with Thermo Fisher Scientific guidelines. Inserts generated by PCR were amplified with overhanging sequences complementary to restriction enzyme sites of the desired vector, achieved through primer design (primer sequences listed in appendix, table 1 and 2). Inserts could then be incorporated into linearised vectors using Quick T4 DNA ligase (NEB). Plasmids were then transformed into competent *E. coli* DH5α cells before being isolated via a GeneJet Miniprep Kit (Thermo Fisher Scientific). Successful cloning reactions were then confirmed through Sanger dideoxy sequencing (Source Bioscience or Biochemistry Department Sequencing Facility, University of Cambridge). All enzymes and kits were used in accordance with manufacturer guidelines.

2.2 Protein expression

Expression of recombinant proteins was achieved in *Escherichia coli* (*E. coli*) BL21 (DE3) cells transformed with protein expressing constructs. Transformed cells were grown at 37°C, 200 rpm in 2xYT Broth (Formedium) with appropriate antibiotics. Expression was then induced at an optical density of >0.8, via the addition of 0.4 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) with a further 3-hour incubation. Alternatively, constructs expressed at 15°C were incubated overnight. The bacterial cell cultures were then pelleted by centrifugation and resuspended in MilliQ water or 50 mM TRIS-HCl pH 8.0, 500 mM NaCl (for soluble proteins). Cell pellets were then stored at -20°C. Additionally, constructs that displayed signs of degradation were lysed in the presences of cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Roche) at 1 tablet per 2 L of bacterial culture.

2.3 Inclusion Body Preparation

Full length pro-proteins and mature constructs of both NRTN and GDF15 were expressed as insoluble inclusion bodies in *E. coli* cells as well as the NRTN prodomain in isolation. Pelleted cells were resuspended in 30 ml lysis buffer (50 mM TRIS HCl pH 8.0, 5 mM ethylene diamine tetra acetic acid (EDTA) with 10 mM dithiothreitol (DTT)) per 1 L of cell culture. Cell lysis was achieved under high pressure using an Emulsiflex C5 homogenizer, with 0.5% Triton-X being added to the sample during lysis. The lysate was then incubated at room temperature with 0.2 mg of DNase and 4 mM $MgCl_2$ for 30 minutes. The lysate was centrifuged at 15000 g for 20 minutes and the supernatant discarded. Centrifugation and removal of the supernatant was repeated after resuspension with the following wash buffers; Buffer A (50 mM tris pH 8.0, 5 mM EDTA, 10 mM DTT, 0.5% v/v Triton-X 100), Buffer A with 2 M NaCl and finally Buffer A without Triton-X 100. The purified inclusion body was pelleted by centrifugation and stored at -20 °C.

2.4 Protein Refolding from Inclusion Bodies

Inclusion body pellets (from 1 L of bacterial cell culture) were re-suspended 5 to 10 ml in 100 mM TCEP pH 7.2, then solubilised with the addition of 15 to 30 ml 8 M guanidine hydrochloride, 50 mM Tris-HCl pH 8.0 and 10 mM EDTA. Samples were then incubated at room temperature for 20 minutes. Before refolding, the resolubilised inclusion bodies were buffer exchanged into deionised 6 M urea plus 20 mM HCl using Sephadex G25 fine beads (GE Healthcare). High concentration protein samples were adjusted to a concentration of ≤ 1 mg/ml before refolding with additional urea buffer. Aggregates were removed by centrifugation with supernatant taken for refolding. For the NRTN prodomain constructs, resolubilised inclusion bodies were taken directly for purification by RPC without refolding.

Refolding conditions were identified by small scale screening of 1 to 2 ml (for details see table 10.1) with the presence of disulphide linked dimers verified by non-reducing SDS-PAGE. Successful conditions were then upscaled to ≥ 1 L. For 1 L, 100 ml of protein in urea was added to 900 ml refolding solution at 4 °C while mixing, in the presence of cystine/cysteine redox pairs to facilitate thiol-disulfide exchange. The refolding process occurred over several days at 4 °C. Details of refolding solutions and conditions for each construct in table 10.2.

Condition Number	Buffer and pH (100 mM)	Salt (500 mM)	Additives
1	Sodium Phosphate pH 7.5	-	-
2	Tris-HCl pH 8.0	-	-
3	Tris-HCl pH 8.5	-	-
4	Tris-HCl pH 9.0	-	-
5	CHES pH 9.5	-	-
6	CAPS pH 10.0	-	-
7	Sodium Phosphate pH 7.5	NaCl	-
8	Tris-HCl pH 8.0	NaCl	-
9	Tris-HCl pH 8.5	NaCl	-
10	Tris-HCl pH 9.0	NaCl	-
11	CHES pH 9.5	NaCl	-
12	CAPS pH 10.0	NaCl	-
13	Sodium Phosphate pH 7.5	-	1 M PPS
14	Tris-HCl pH 8.0	-	1 M PPS
15	Tris-HCl pH 8.5	-	1 M PPS
16	Tris-HCl pH 9.0	-	1 M PPS
17	CHES pH 9.5	-	1 M PPS
18	CAPS pH 10.0	-	1 M PPS
19	Sodium Phosphate pH 7.5	NaCl	1 M PPS
20	Tris-HCl pH 8.0	NaCl	1 M PPS
21	Tris-HCl pH 8.5	NaCl	1 M PPS
22	Tris-HCl pH 9.0	NaCl	1 M PPS
23	CHES pH 9.5	NaCl	1 M PPS
24	CAPS pH 10.0	NaCl	1 M PPS

Table 2.1. Refolding Screen – Chemical Conditions. Final constituent concentrations of refolding conditions after addition of solubilised inclusion bodies. All conditions also include 0.2 mM L-Cystine and 2 mM L-Cysteine as a redox pair to catalyse disulfide exchange. Pyridinium propyl sulfobetaine (PPS) additive used to reduce aggregation.

NRTN constructs	Refolding Condition
Pro-NRTN C23S and Pro-TEV-NRTN C23S	100 mM TRIS pH 8.5, 1M PPS, 1.2 M urea with 0.2 mM L-cystine and 8 mM L-cysteine
Pro-NRTN WT	100 mM TRIS pH 8.5, 1M PPS, 0.6 M urea with 0.2 mM L-cystine and 2 mM L-cysteine
Mature NRTN	100 mM TRIS pH 8.5, 1M PPS, 100 mM NaCl, 1.2 M urea with 0.2 mM L-cystine and 8 mM L-cysteine
GDF15 constructs	
Pro-GDF15 and Pro-TEV-GDF15	100 mM TRIS pH 9.0, 1.2 M urea, 1 M PPS with 0.2 mM L-cystine and 3 mM L-cysteine
Mature GDF15 constructs (WT, HZ, H202D and heterodimers)	100 mM TRIS pH 8.0, 1.2 M urea, 1 M PPS with 0.2 mM L-cystine and 2 mM L-cysteine.

Table 2.2. Refolding solution composition for GDF15 and NRTN constructs. Wild type (WT) and heterozygous (HZ) as indicated (see Chapter 4 for HZ description).

2.5 Protein Purification

Refolded and soluble proteins were purified using various chromatographic methods, summarised in table 10.3 and detailed below. All proteins samples were analysed by SDS-PAGE at each stage and final samples of purified protein were then flash frozen in liquid nitrogen and stored at -80°C.

Construct	Purification Steps
His-NRTN Prodomain C23S	IB, RPC
NRTN Prodomain C23S	IB, RPC

Mature NRTN	IB, Refolding, IEX and RPC
Pro-NRTN WT	IB, Refolding, IEX and RPC
Pro-NRTN C23S	IB, Refolding, IEX and RPC
Pro-TEV-NRTN C23S	IB, Refolding, IEX and RPC
GDF15 Prodomain	Sol, Ni-IMAC, IEX
Mature GDF15 WT, HZ and H202D	IB, Refolding, IEX and RPC
Pro-GDF15	IB, Refolding, IEX and SEC
Pro-TEV-GDF15	IB, Refolding, IEX and SEC
Mature GDF15 Heterodimers	IB, Refolding, IEX and RPC

Table 2.3. Protein Purification Summary. Details of purification steps for purified protein constructs. Ion-exchange chromatography (IEX), Reversed-phase chromatography (RPC), Size-exclusion chromatography (SEC), Nickel-immobilised affinity chromatography (Ni-IMAC). Proteins purified from either Inclusion bodies (IB) and refolded or the supernatant as a soluble (Sol) protein. Wild type (WT) and heterozygous (HZ) as indicated (see Chapter 4 for HZ description).

2.6 Nickel Affinity Chromatography

Supernatant from lysed bacterial cell cultures was applied to Ni-NTA resin (Cube Biotech) equilibrated in lysis buffer (LB) 50 mM TRIS-HCl pH 8.0, 500 mM NaCl. The resin was then washed with 3 column volumes (CV) of LB, three CV of LB with 20 mM imidazole and finally three CV 200 mM imidazole. Fractions from each stage collected and analysed by SDS-PAGE.

2.7 Ion-exchange chromatography

Ion-exchange chromatography (IEX) performed using either a HiTrap SP HP (GE Healthcare) or SOURCE 30S column (GE Healthcare). Bound protein samples were then eluted via a sodium chloride gradient and analysed by SDS-PAGE.

2.8 Reversed-Phase Chromatography

Reversed-phase chromatography (RPC) was performed using an ACE5 C8-300, size 4.6×250 mm column (HiChrom) or Source 15 RPC column (GE Healthcare) equilibrated in 10% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA). Before loading, protein samples were adjusted to 10% ACN and 0.1% TFA. Elution of proteins were then carried out over a gradient

of 10% to 90% ACN and collected by fractionation. Eluted fractions were then analysed by SDS-PAGE and lyophilised to for storage at -80°C.

2.9 Size-Exclusion Chromatography

Size-Exclusion Chromatography (SEC) was performed for pro-GDF15 constructs using a Superdex 200 increase 10/300 GL or HiLoad 16/60 Superdex 200 column (both GE Healthcare/Cytiva) equilibrated in 50 mM TRIS HCl pH 8.0, 2 M urea and 500 mM NaCl. Fractions from SEC were then analysed by SDS-PAGE before being flash frozen in liquid nitrogen and stored at -80°C.

2.10 SDS-PAGE

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse all protein samples using 15% polyacrylamide gels. Samples were mixed with an equal volume of loading dye ((125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 0.02% bromophenol blue and +/- 8% β-mercaptoethanol (β-ME) as indicated)). Reduced samples, with added β-ME, were then heated for 5 minutes at 95°C before being loaded onto the gel. Gels were run at 200 V for 50 minutes in 25 mM Tris-HCl, 192 mM glycine with 0.1% w/v SDS before staining. Protein visualization was achieved using Coomassie blue staining (0.1% w/v Coomassie Brilliant Blue R, 10% ethanol and 10% acetic acid). Alternatively, 0.3 M $ZnCl_2$ was added after a 0.2 M imidazole incubation for zinc-stained gels. NEB Unstained Protein Standard, Broad Range (10-200 kDa – P7704S or P7717S) was used as a molecular weight (M_w) marker for all gels.

2.11 UV Spectrophotometry

UV spectra for all protein samples were measured using either Nanodrop (Thermo Scientific) or Varian Cary 50 Bio UV-Visible Spectrophotometer as appropriate. Protein concentration was then calculated using molecular weight and extinction coefficient values from the ExPASy ProtParam tool (available at https://web.expasy.org/protparam/).

2.12 Furin Cleavage

Furin cleavage was performed in 100 mM HEPES pH 7.2, 1 mM $CaCl_2$, 1 mM β-ME, 0.1% V/V Triton X-100 and incubated overnight at 25°C. 12.5 µg of protein substrate was incubated with one unit of furin enzyme (NEB). Cleavage of GDF15 protein constructs was performed

with an additional 500 mM Urea and 250 mM NaCl present to aid solubility. Samples were then analysed by SDS-PAGE.

2.13 TEV cleavage

Protein cleavage was carried out using TEV protease (produced by Beata Blaszczyk, Hyvönen lab) for TEV cleavable constructs pro-TEV-NRTN and pro-TEV-GDF15. Pro-TEV-GDF15 was cleaved in 2 M Urea, 500 mM NaCl, 50 mM TRIS pH 8.0, and pro-TEV-NRTN was cleaved in 250 mM HEPES pH 7.2 at room temperature overnight. Samples were then analysed by SDS-PAGE.

2.14 Ellman's Assay

Ellman's Reagent ((DTNB, 5,5-dithio-bis-(2-nitrobenzoic acid)) Thermo Scientific was used to measure reduced cysteines/free sulfhydryls in solution following the manufacturers guidelines. Briefly, a standard curve of known reduced cysteine concentration was produced and then used to quantify the level of free cysteines in protein samples (pro-NRTN WT and pro-NRTN C23S mutant).

2.15 GDF15 Bioactivity Reporter Gene Assay

GDF15 bioactivity assays performed in HEK293S-SRF-RET/GFRAL cells at AstraZeneca, Gothenburg, Sweden. This stable cell line was modified to express full length human RET and GFRAL receptors as well as serum response factor with bioassay developed by AstraZeneca. Assay specificity determined during development stage with suitable positive and negative controls. Cells were grown and seeded by Dr Anna Backmark, in T75 flasks in DMEM GlutaMAX (Gibco) supplemented with 10% foetal bovine serum (FBS) (Gibco) and incubated for 30 hours at 37°C, with 5% carbon dioxide. Post incubation, media was switched to serum free DMEM and cells were then Incubated overnight under the same conditions. The following day cells were detached using Accutase enzyme, neutralised in DMEM, 1% FBS, pelleted by centrifugation and resuspended in in serum free DMEM with an added 25 mM HEPES pH 7.2 (GDF15 bioassay media). Cells in suspension were counted using a Cedex automatic cell counter (Roche) and Trypan Blue (Sigma) staining as standard. After counting, cells were then diluted to 500 k cells/ml in GDF15 bioassay media before plating in 384-well plates (15 µl/well).

After plating, 5 µl of GDF15 in GDF15 bioassay media was then added to the cells at 4x the final concentration (as wells contained already 15 µl of media). Cells with protein samples were then incubated for 5 hours at 37°C, with 5% carbon dioxide. After 5 hours, the plate was then equilibrated to room temperature before the addition of Steady-Glo Luciferase Assay System solution (Promega) at 12µl/well and incubated for a further 20 minutes. The luminescence was measured using an EnVision plate reader (Perkin Elmer) with crosstalk correction applied. Chinese hamster ovary (CHO) cell derived recombinant Human GDF-15/MIC-1 (Peprotech 120-28C) protein was used a positive control. Negative controls of protein sample buffers only were also screened to confirm absence of signalling activity. Triplicates of each dose were performed, and data was analysed in GraphPad 7.0 (Prism). In each assay two technical replicases were performed.

2.16 NRTN Bioactivity Reporter Gene Assay

NRTN bioactivity assays were performed using a TGW-SRE Luciferase cell line at AstraZeneca, Cambridge. The cell line and bioassay development were done previously by AstraZeneca and assay specificity was determined during development stage with suitable positive and negative controls. The cell line used is a stable transfection of the human neuroblastoma TGW cell line JCRB0618 (Japanese Collection of Research Bioresources, Osaka, Japan) with pCignal Lenti-SRE-Reporter gene (SABiosciences, CLS-010L) selected previously through puromycin selection with the final cell line containing a reporter gene construct with a luciferase gene under the control of serum-response elements, as previously described[58,85].

Cells incubated at 37°C, with 5% carbon dioxide in DMEM:F12 (Invitrogen) and 15 % FBS (Invitrogen), with 350 µg/ml geneticin (PAA), and 1% penicillin-streptomycin (10,000 U/ml Invitrogen). Cells were grown and seeded in 96-well plates by Dr Jacqueline Naylor. Cultured cells were detached using 0.05% trypsin/EDTA (Invitrogen) and seeded in 96-well plates (at 40,000 cells per well in 90 µl of serum free media). Seeded cells were then grown overnight before incubation with NRTN constructs at 37°C and 5% CO_2.

Serial dilutions of NRTN constructs in serum free media were added to the cells (10 ul per well) at 10x the desired final concentration and incubated for 4-5 hours. Triplicates of each dose were performed. Luciferase activity was measured using Steady-Glo® Luciferase Assay Substrate (Promega, Cat# E2550) following the manufacturer's instructions. After addition of

the substrate, plates were then protected from light and incubated for 5–10 min at room temperature. Luminescence was then read using an EnVision plate reader with data analysed in GraphPad 7.0 (Prism).

2.17 Circular Dichroism Spectroscopy

Circular dichroism (CD) data were collected using an Aviv model 410 spectropolarimeter (Biophysics Facility, Department of Biochemistry, University of Cambridge). Purified NRTN and GDF15 constructs (mature NRTN, pro-NRTN, NRTN prodomain and the GDF15 prodomain) were analysed between 0.16 and 0.2 mg/ml in water using 1 mm path-length quartz cuvettes. GDF15 prodomain samples were analysed in 10mM Na_2HPO_4/ NaH_2PO_4 pH 7.5, in the presence or absence of 0.2 mg/ml Low-molecular-weight heparin (LMWH, Iduron). Sample concentration listed in table 10.4, with UV spectrophotometry used to determine protein concentration. CD spectra were recorded between 185 and 250 nm at 25 °C, with 1 nm increments, 1 nm bandwidth and 3-5 seconds averaging time. Three measurements per sample were averaged, buffer signal corrected and smoothed before being converted into mean residue ellipticity ([θ]) (x1000 deg.cm^2.dmol^{-1}.residue^{-1}). In each assay three technical replicases were performed.

Sample	Concentration (mg/ml)	Residues
NRTN Prodomain	0.17	20-95
His-mature NRTN	0.16	96-197
Pro-NRTN	0.20	20-197
His-GDF15 Prodomain	0.20	30-196

Table 2.4. Far-UV CD Spectroscopy. Purified protein sample construct information and concentration for CD analysis. His-tagged constructs indicated.

2.18 Hydrogen deuterium exchange mass spectrometry (HDX-MS)

Deuterium exchange reactions were conducted at room temperature for 0, 30, 60, 600 or 1800 seconds. Protein samples of mature NRTN, pro-NRTN or NRTN prodomain were diluted to 10 µM in 50 mM MOPS pH 7.2. 5 µL of each protein sample was added to 57 µL of 50 mM MOPS pD 7.2 in D_2O. The addition of 62 µL of 3 M urea, 400 mM TCEP pH 2.5 in H_2O was used to quench the exchange reaction. Quenching was conducted at 0 °C to prevent back exchange.

Non-deuterated protein samples were generated for reference by following the same method in H_2O. Exchange and quenching reactions were performed in triplicate and automated by using a HDx-3 PAL™ robot (LEAP Technologies).

Quenched protein samples were then injected onto a Waters Enzymate™ BEH Pepsin Column (2.1 x 30 mm) for protein digestion. Post digestion, peptides were injected into a Waters nanoACQUITY UPLC system and desalted using a VanGuard Pre-Column Acquity UPLC BEH C18 (2.1 × 5 mm). Peptide elution and separation was then achieved by acetonitrile gradient with 0.1% formic acid. Eluted peptides were then analysed on a Waters Synapt G2-Si LC-MS system.

Peptides from three replicates of non-deuterated protein samples were identified using the ProteinLynx Global Server (PLGS v3.0.2, Waters) and subsequently imported into DynamX v3.0 data analysis software (Waters) for processing. Only peptides with a minimum intensity of 1000, present in at least two out of three replicates, and with a sequence length between 3 and 25 amino acids were retained for analysis. HDX-MS analysis was performed on the NRTN prodomain as sufficient peptide coverage for mature domain of NRTN could not be achieved. Following the first round of automated spectral data processing by DynamX v3.0, each peptide was manually examined for suitability and with any errors corrected or edited as appropriate. HDX-MS experiments were performed at AstraZeneca, Gothenburg, Sweden.

2.19 EMSA

The GDF15 prodomain was incubated at 2 mg/ml with and without 1 µM of the fluorescently tagged heparan sulfate, DP10-AMAC in 25 mM CHES pH 9.0, 500 mM NaCl. Incubated protein samples were the analysed by electrophoretic mobility shift assay (EMSA) on a 1% agarose gel, at 90 volts for 45 minutes. Sample incubation and EMSA performed at room temperature.

2.20 ITC

Isothermal titration calorimetry (ITC) experiments were performed at 25 °C using a MicroCal iTC200 instrument (GE Healthcare) - Biophysics Facility, Department of Biochemistry, University of Cambridge. The GDF15 prodomain buffer exchanged into 100 mM TRIS pH 8.0, 5 mM EDTA (buffer A) via a NAP-5 column (GE Healthcare) and adjusted to a final concentration of 50 µM. Lyophilised NRTN prodomain was resuspended in water was

adjusted to buffer A by dilution to a final concentration of 100 µM. Lyophilised mature GDF15 and mature NRTN were resuspended in 100 µl of 10 mM HCl, before being diluted to a final concentration of 10 µM in buffer A with 3.2 µM bovine serum albumin (BSA). The cell contained either mature GDF15 or mature NRTN with the syringe containing the corresponding prodomain. Post insertion of the syringe into the cell, the stirring speed was set to 750 rpm and cell temperature to 25°C. After equilibration, the first injection was 0.2 µl and 2 µl thereafter. Spacing between injection was set to 160 seconds. Data was then processed and analysed using Origin 7.0 software.

2.21 Turbidity Assay

50 µl of the GDF15 prodomain at 5µM in 10 mM Sodium Phosphate pH 7.5, +/- 150 mM NaCl was assessed for turbidity in the presence of LMWH (Iduron). Turbidity assessed via the measurement of absorbance at 340 nm. LMWH at 0.1 mg/ml (estimated to be ~50 µM) was added progressively, with mixing and measurement of absorbance. Absorbance measurements collected using a Varian Cary 50 Bio UV-Visible Spectrophotometer. Data collected were then analysed using GraphPad 7.0 (Prism) software.

2.22 Initial Protein Crystallisation Screening

The GDF15 prodomain, pro-NRTN or pro-TEV-NRTN were all screened between 6 and 9 mg/ml for suitable crystallisation conditions using a Mosquito Crystal (TTP Labtech) crystallisation robot in MRC 2 Well Crystallization Plates in UVP (Jena Bioscience) and incubated at 18°C. Crystallisation method selected was vapour diffusion in sitting drops, with a 1:1 or 2:1 protein ratio to well solution (200 nl). Imaging and UV-imaging was used to determine the presence of protein crystals. Initial screening of crystallisation conditions was carried out using commercially available screens.

Chapter 3 – The Production and Characterisation of NRTN pro-forms

In this chapter we will examine the production and characterisation of NRTN pro-forms, through recombinant protein expression and purification, cell-signalling assays, enzyme cleavage and biophysical analysis.

3.0 Production of NRTN Constructs

To assess the role of the prodomain on NRTN activity and signalling protein samples must first be produced. The production of pure and homogenous protein is essential for the functional characterisation of any protein. In this section we will address the production of recombinant mature NRTN, pro-NRTN, the NRTN prodomain and a Tobacco Etch Virus (TEV) protease cleavable form pro-NRTN. In future sections purified NRTN proteins will be used for bioactivity assays and biophysical experiments.

Several protein expression systems are widely used in recombinant protein expression including bacterial, mammalian and insect cell systems. Here we chose a bacterial expression system using *Escherichia coli (E. coli)*. Growth factor expression from *E. coli* is commonplace and has been successfully employed for many growth factors in the field and is well established in the Hyvönen lab. Bacterial expression systems can be relatively inexpensive, are easily manipulated and can generate high yields of proteins. Additionally, *E. coli* expression systems do not express furin proteases and so uncleaved forms of pro-NRTN can be easily generated for testing. Bacterial based expression systems also avoid issues of cross-contamination with other growth factors. This has been area of concern in studies using mammalian derived growth factors such as GDF15[86], as discussed in the introduction.

As described in the introduction, NRTN has several disulfide-bonds which are essential to the correct folding and production of an active dimeric growth factor. However, the bacterial cytoplasm has a reductive environment that does not facilitate the formation of disulfide-bonds from cysteine residues. This often leads to misfolding and aggregation of cysteine-rich proteins when over-expressed in *E. coli* – these structures are known as inclusion bodies.

Inclusion bodies can be isolated easily from bacterial cell lysate and are often relatively pure as they can be separated from the majority of soluble *E. coli* protein contaminants. To achieve correctly folded protein from isolated inclusion bodies, they must be denatured and refolded in suitable conditions to produce the native structure. The in vitro refolding process requires the screening and optimisation of suitable chemical and physical variables to identify a suitable condition. Once identified, refolded protein can then be purified using standard protein purification techniques. In this chapter we first identify a suitable refolding and purification method for mature and pro-NRTN that generates a high yield of pure protein suitable for further studies.

3.1 NRTN Construct Design and Prodomain Structure

3.11 NRTN construct design

Human NRTN is 197 amino acids (aa) in length and consists of an N-terminal signal sequence (1-19), a prodomain (20-95) and a mature domain (96-197) the active growth factor (UniProt: Q99748). A furin cleavage site is located between the prodomain and mature domain (see figure 3.1A). Plasmids encoding full-length human pro-NRTN (20-197) and mature NRTN (96-197) were available in the group and were used as the templates for NRTN constructs (see table 3.1). Furthermore, all inserts were codon optimised to facilitate protein expression in *E. coli*.

Certain TGF-β prodomains, like those of TGF-βs, activins and myostatin, contain additional cysteine residues within the prodomain. In the case of TGF-β1 prodomain cysteine residues form an additional disulfide-bond between the prodomains and form disulfide-bonds with other proteins, as described earlier. Unpaired cysteine residues can be problematic in refolding approaches and interfere with the formation of the cystine-knot – essential for TGF-β protein structure. NRTN also has a cysteine residue close to the N-terminus of the prodomain, but the function of this residue is unknown.

Mutation of prodomain cysteine residues to serine in the prodomains can avoid the problem of additional unpaired cysteines forming incorrect disulfides and disulfide-linked aggregates. This approach has been successfully utilised to improve the production of pro-TGF-β1 and refolded pro-activin A[16,33]. Therefore, an additional pro-NRTN construct was generated using

PCR site-directed mutagenesis to mutate the N-terminal cysteine of the prodomain to serine (C23S) and cloned in the vector pOP1 (see table 3.1).

Vector - Insert	Recombinant Protein tag
pBAT4 pro-NRTN 20-197*	n/a
pHAT2 pro-NRTN 20-197*	His
pHAT2 mature NRTN 96-197*	His
pOP1 pro-NRTN 20-197 C23S	n/a
pOP1 pro-TEV-NRTN 20-197 C23S**	n/a
pHAT4 NRTN 20-95 C23S prodomain	His
pMAT11 NRTN 20-95 C23S prodomain	His-MBP
pGAT3 NRTN 20-95 C23S prodomain	His-GST
pOP3Su NRTN 20-95 C23S prodomain	His-Sumo
pOP1 NRTN 20-95 C23S prodomain	n/a

Table 3.1. Cloned NRTN Constructs. Summary human NRTN constructs cloned for protein expression. All constructs sequenced to confirm 100% identity to the desired sequence, at the amino acid level. Inserts were also confirmed to be within the correct reading frame within the vectors. Vectors used; pBAT4[116], pHAT2[116], pMAT11 - Sheena Gordon, unpublished, pGAT3 - Peränen and M. Hyvönen, unpublished, pOP3Su and pOP1 plasmids - Hyvönen lab, unpublished. *Indicates donated material, cloned by Dr Miglė Kišonaitė. **Pro-TEV-NRTN construct with furin site replaced by recombinant Tobacco Etch Virus (TEV) protease site for TEV cleavage *in vitro.*

3.12 NRTN prodomain conservation and secondary structure prediction

From sequence alignment we can see the prodomain of NRTN is relatively well conserved, with a conserved stretch of serine residues and sections of conservation that map to the predicted helical secondary structure elements (See figure 3.1 B and C). At the C-terminal we can see high levels of conservation of several arginine residues, this corresponds to the furin cleavage site, which is characterised by two conserved arginine residues, separated by two residues (RXXR) figure 3.1 A and B). From secondary structure prediction we can see that the prodomain is largely unstructured. Approximately 80% of the prodomain contains no

predicted structural elements. Three α–helical segments are predicted, and they seem to be well conserved in sequence as well across species. However, only one of the α-helices is predicted with high confidence (figure 3.1 B and C). A high level of conservation in the NRTN prodomain suggest that it carries a specific, as yet unknown function.

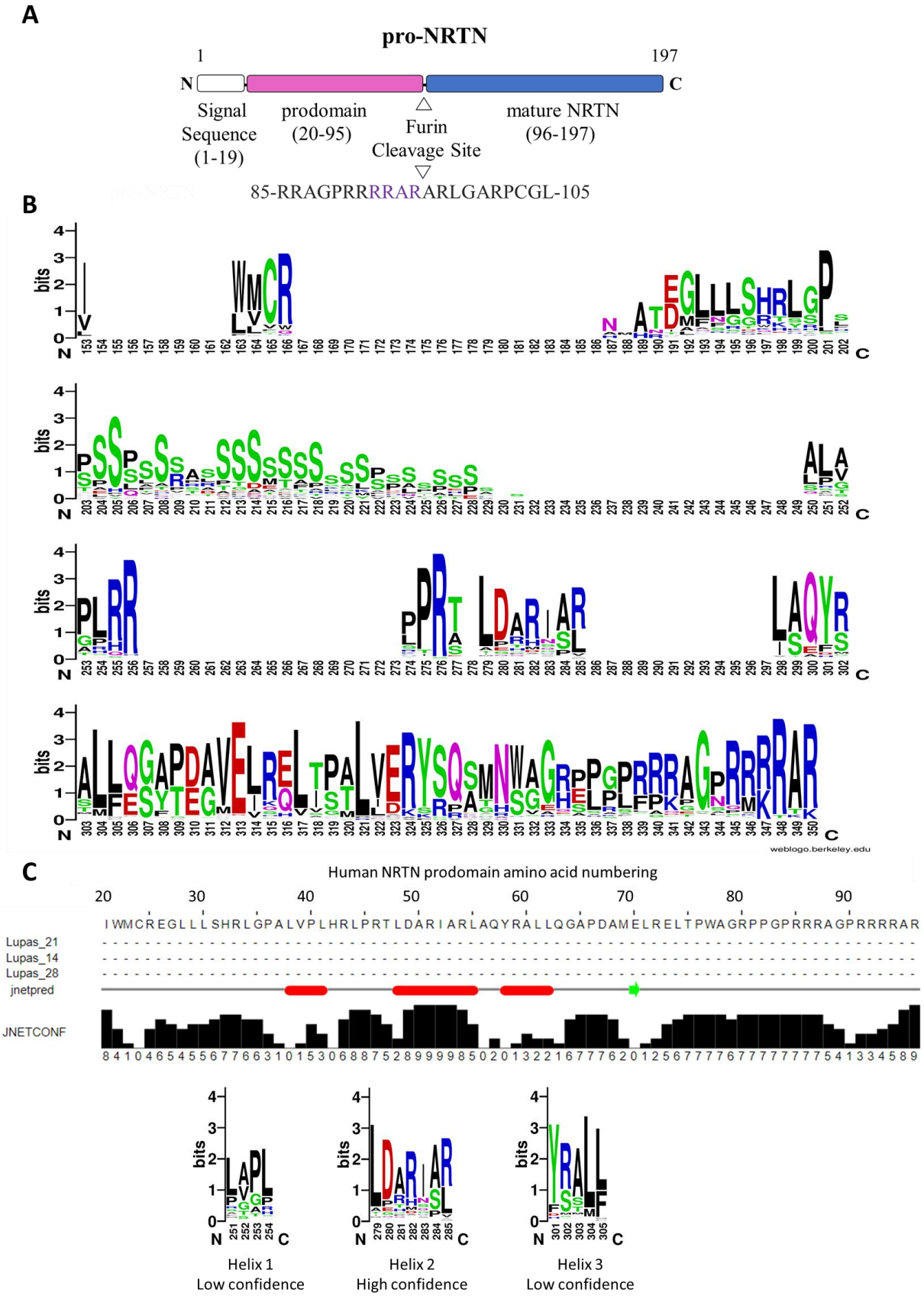


Figure 3.1. Pro-NRTN, Prodomain Conservation and Secondary Structure. A) Schematic of human pro-NRTN protein, signal sequence in white, prodomain in pink and mature domain in

blue. Position of the furin cleavage site indicated, and recognition sequence shown in purple. Domains of Pro-NRTN are proportionate to length in amino acids. B). Conservation of the NRTN prodomain. All sequences for NRTN orthologues taken from UniProt (https://www.uniprot.org/) (113 sequences, from fish to mammals) and aligned in MUSCLE[24] (https://www.ebi.ac.uk/Tools/msa/muscle/). Figures were then created using WebLogo3[117] (http://weblogo.threeplusone.com/create.cgi). C) NRTN prodomain (aa 20-95) secondary structure predication, α–helical regions shown in red with conservation of each helical region below. Prediction confidence level is indicated by the back bars. Created using WebLogo3. and Jpred4[118] (https://www.compbio.dundee.ac.uk/jpred/).

3.2 Expression and refolding of pro-NRTN C23S

Previous work by Katharina Ravn (Hyvönen lab) identified suitable conditions for the expression, refolding, and purification of pro-NRTN wild type. Following this method pro-NRTN could be produced, however problems with aggregation, low yield was observed. Additionally aggregates, such as disulfide-linked oligomers were found to co-purify with the dimeric from and became major contaminants (see figure 3.2). The pro-NRTN wild type was found to have unpaired cysteine residues by Ellman's assay (data not shown – for full details see materials and methods). To overcome issues of aggregation due to unpaired cysteine residues we switched to pro-NRTN C23S, as described previously. The wild type pro-NRTN method was then used as the starting point to produce pro-NRTN C23S. From this point on, pro-NRTN C23S will also be referred to as pro-NRTN.

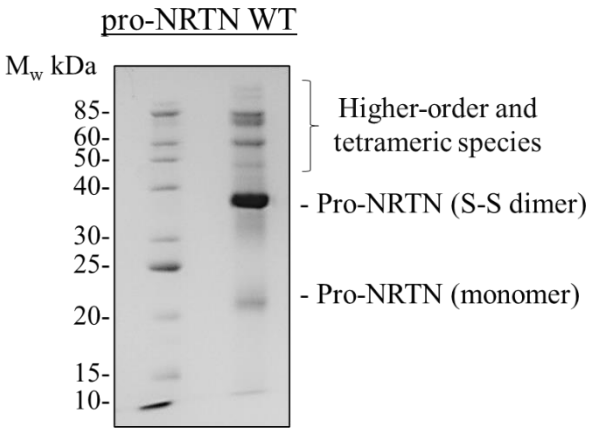


Figure 3.2 Wild type pro-NRTN. Refolded and purified wild type pro-NRTN analysed by non-reducing SDS-PAGE, to maintain disulfide bonds (S-S). Tetrameric, dimeric, monomeric, and higher-order species indicated.

As expected, pro-NRTN was insoluble in bacterial expression tests and resided entirely in the insoluble pellet fraction and runs at ~20 kDa as a monomer (see figure 2.3A). Therefore, the protein was purified from inclusion bodies. Inclusion body preparation yielded relatively pure pro-NRTN, with the desired protein as the major species in the sample and major contaminants removed in the wash stages (figure 3.3B). Inclusion bodies can be easily resolubilised – here we use reducing and denaturing agents to dissolve the purified inclusion bodies (see materials and methods). As solubilised inclusion bodies would be denatured and monomeric, we then aimed to identify the chemical conditions to refold the protein – to produce the native dimeric growth factor.

Numerous parameters can be screened and optimised in search of a successful refolding condition including, temperature, protein concentration, pH, detergents, Ionic strength, additives, non-detergent sulfobetaines (NDSBs), protein chaperones/binding partners and more. Denatured protein may be exchanged into refolding conditions using dialysis or dilution. The formation and shuffling of disulfide bonds are promoted through the addition of a redox pair, such as cystine (oxidised) and cysteine (reduced). Protein aggregation can often be an issue during protein refolding, due to exposed hydrophobic regions. Additives to prevent aggregation are often included, such as pyridinium propyl sulfobetaine (PPS). PPS is thought to stabilise hydrophobic surfaces of denatured protein and so improves solubility throughout the refolding process[119,120]. Urea and guanidine chloride are chaotropic agents and, at high concentrations will denature proteins. This is a useful property when resolubilising inclusion bodies. The denaturants can then be diluted or removed to enable the protein to refold under different conditions. Lower concentrations of denaturant and/or amino acids, such as urea and arginine, may also be beneficial, helping to stabilise the refolded target protein and prevent aggregation – as used for activin A refolding [33,121]. Urea denatures protein structure though interaction with the peptide backbone, altering the solvation pattern as well as hydrogen bonding [121]. In this thesis we will use PPS and urea to aid refolding.

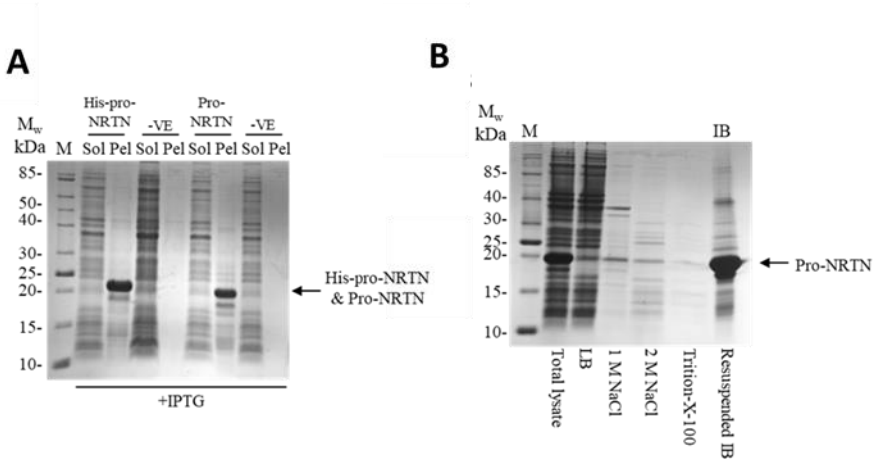


Figure 3.3. Expression and Inclusion Body Preparation of pro-NRTN. A) His-pro-NRTN & pro-NRTN expression tests in transformed *E. coli* BL21 DE3 cells at 37°C for three hours + IPTG (Sol = soluble fraction, Pel = insoluble pellet fraction and -VE = negative control – empty vector). B) pro-NRTN inclusion body preparation stages – full details in materials and methods (LB = lysis buffer, IB = inclusion body). All protein samples analysed by SDS-PAGE.

The previously identified refolding conditions for pro-NRTN wild type as a starting point for refolding (100 mM TRIS pH 8.5, 1M PPS, 0.6 M urea with 0.2 mM L-cystine and 2 mM L-cysteine). In this refolding condition, the C23S protein was also found to refold, forming a disulfide-linked dimer and overall solubility partially improved. However, aggregation and precipitation remained issues when refolding pro-NRTN, reducing the amount of soluble material that could be purified and reducing the overall yield. Further optimisation of refolding conditions reduced visual aggregation and overall yield significantly.

Further screening of temperature, composition and concentration of refolding solution and incubation time identified three parameters that reduced overall aggregation during refolding optimisation. Firstly, increasing the level of L-cysteine from 2 to 8 mM, presumably by altering disulfide formation dynamics. Secondly, increasing the concentration of urea in the refolding solution. Screening of increasing urea concentrations from the initial 0.6 M reduced visible aggregation up to 3.0 M. Beyond a concentration of 3.0 M urea, the refolding process was inhibited with a reduction in dimeric pro-NRTN visible by SDS-PAGE (See figure 3.4). Finally, reducing the overall protein concentration to ~0.1 mg/ml in the refold solution also improved solubility. Pro-NRTN was then refolded and purified from this optimised condition - The final condition used for pro-NRTN refolding was 100 mM TRIS pH 8.5, 1M PPS, 1.2 M urea with 0.2 mM L-cystine and 8 mM L-cysteine, with refolding carried out at 4°C for 5 days.

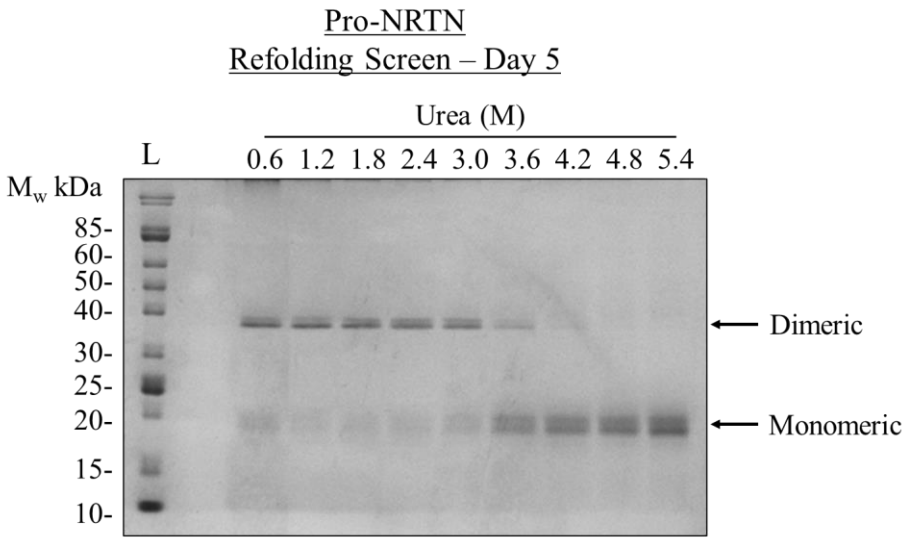


Figure 3.4. Refolding optimisation of pro-NRTN. Refolding condition screening of purified pro-NRTN inclusion bodies in 100 mM TRIS pH 8.5, 1M PPS, 0.2 mM L-cystine and 8 mM L-cysteine, with refolding carried out at 4°C. Varying concentrations of urea included in the refold solution as indicated. Pro-NRTN dimeric and monomeric bands indicated by arrows. Refolding carried out at 4°C for 5 days with samples analysed by non-reducing SDS-PAGE.

3.3 Purification of refolded pro-NRTN

With refolded material produced, the next stage is to purify the protein. Cation exchange chromatography and reversed phase chromatography (RPC) were successful in past purifications of wild type pro-NRTN and so were used here also. Cation exchange was successful in binding and eluting the refolded pro-NRTN dimer, with bands viable at ~40 kDa (see figure 2.4A). Pro-NRTN contains many positively charged residues, with a theoretical pI of 11.04 (calculated using ExPasy ProtParam tool) and binds well to the negative matrix of the column. Once bound protein can then be eluted under a gradient of sodium chloride. Eluted fractions containing the dimer were then taken for further purification by RPC.

The pro-NRTN disulfide-linked homodimer runs lower than its theoretical molecular weight of 41 kDa, with a band seen below the 40 kDa (see figure 3.5 A). Disulfide-linked proteins can exhibit anomalously fast migration when analysed via non-reducing SDS-PAGE. Disulfide bonds link two peptides together creating a more compact structure, and so the protein migrates faster through the gel in comparison to fully linear and monomeric denatured protein[122]. Pro-NRTN was confirmed to be a disulfide-linked dimer by SDS-PAGE (see figure 3.5 B). The presence of disulfide-linked dimers can be confirmed by non-reducing SDS-PAGE

using reduced and non-reduced samples. Here we use the reducing agent β-mercaptoethanol (β-ME) and heat to reduce disulfides creating monomers from dimers as seen in figure 3.5 B.

Bands at ~20 kDa can be seen after cation exchange (figure 3.4 A), these contaminants are likely monomeric forms of pro-NRTN and were excluded from further purification. RPC was then used to further purify and concentrate the protein. RPC is carried out under acidic conditions using organic solvents, such as acetonitrile to separate molecules based on hydrophobicity. Fractions from cation exchange were acidified and loaded onto the column before elution under an increasing concentration of acetonitrile. Clean bands of pro-NRTN, at ~40 kDa can be seen from RPC (see figure 3.5 C). RPC enabled the concentration and further separation of pro-NRTN dimer from aggregates and monomeric species – earlier fractions in figure 3.5 C were selected and were pooled and concentrated for storage. As can be seen in the final sample, pro-NRTN is relatively pure and homogenous with a clear band at ~40 kDa (see figure 3.5 D). Faint smaller molecular weight bands can be seen under pro-NRTN, these are likely degradation products with full-length pro-NTRN as the majority species (see figure 3.5 D). This degradation is likely N-terminal degradation as no degradation is visible for mature NRTN with the same C-terminal as pro-NRTN (see figure 3.7) A flow-chart of protein expression and purification method is summarised in figure 3.5 E.

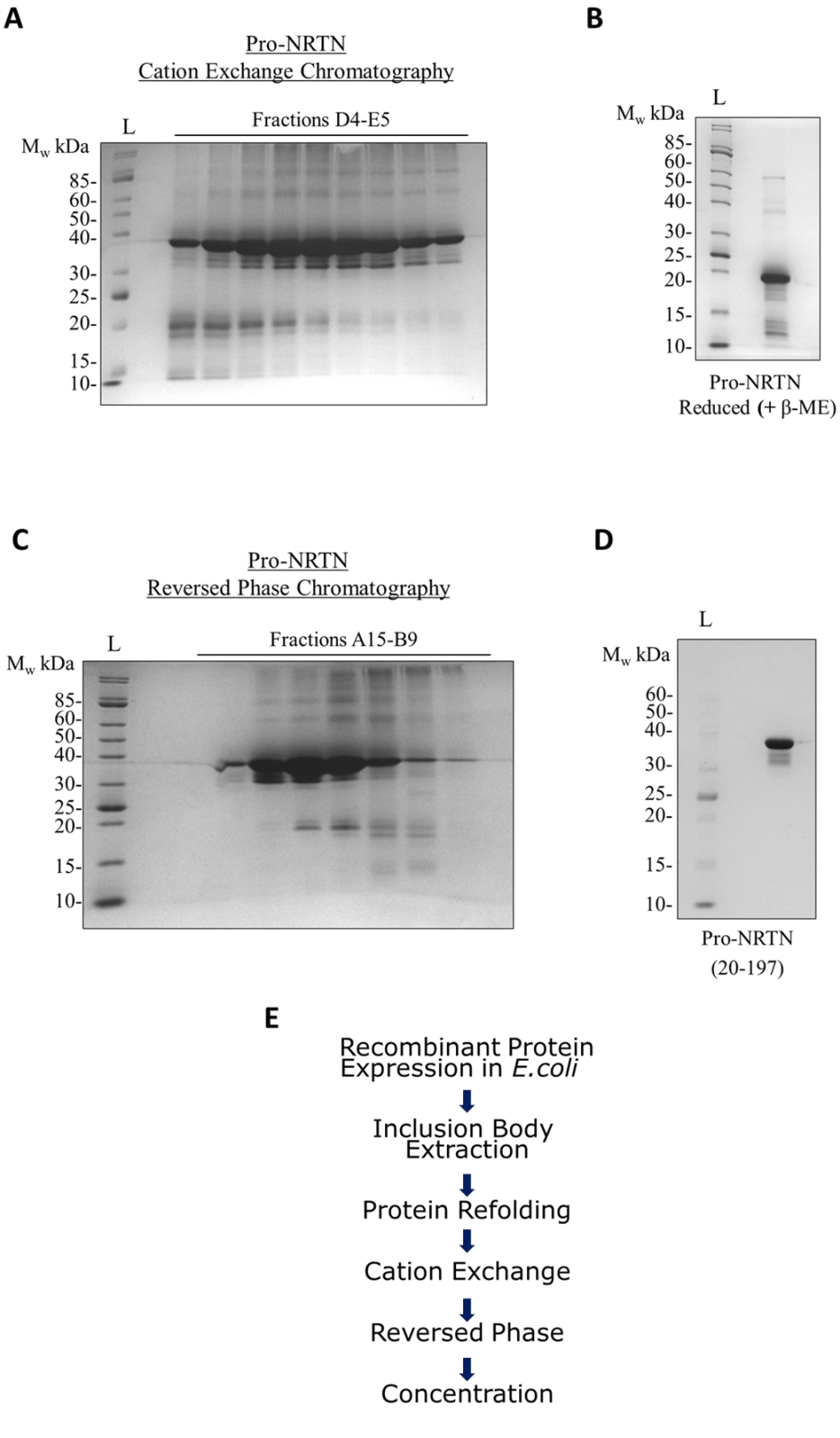


Figure 3.5. Purification of pro-NRTN. Protein purification stages for the purification of pro-NRTN (20-197) as analysed by SDS-PAGE. A) Cation exchange chromatography performed on refolded pro-NRTN. B) Dimeric pro-NRTN reduced with β-ME (β-mercaptoethanol) to monomeric form. C) Reversed phase chromatography performed on fractions from cation

exchange chromatography. D) Purified and concentrated pro-NRTN sample. E) Summary flow-chart of protein expression and purification method.

3.4 Refolding and Purification of pro-TEV-NRTN

3.41 Pro-TEV-NRTN construct design and refolding

As described earlier, NRTN is cleaved by the protease furin, at a recognition site between the prodomain and mature domains (see figure 3.1). To examine the impact of cleavage on NRTN bioactivity a Tobacco Etch Virus (TEV) protease cleavable pro-NRTN construct (pro-TEV-NRTN) was created, this construct will allow specific and controlled cleavage *in vitro* by TEV protease. This will enable analysis of the cleaved form of pro-NRTN, by bioactivity assay (further details and the use of this protein later in this Chapter – see figure 3.6). The engineered pro-TEV-NRTN protein has the natural furin cleavage site replaced by a TEV protease site. In order to retain the length between NRTN domains no amino acids were inserted or deleted, as this had been found to cause problems when refolding other pro-TGFβ proteins in the Hyvönen lab (data unpublished). This also ensured that no artefacts were introduced due to increased length between the mature domain and the rest of the pro-domain.

Pro-TEV-NRTN was screened for a suitable refolding condition in various conditions (see materials and methods). Despite a change in sequence, refolding was successful using a similar condition as pro-NRTN. Pro-TEV-NRTN was found to refold successfully in 100 mM TRIS pH 8.5, 1.2 M urea, 1 M PPS, 0.2 mM L-cystine and 4 mM L-cysteine, with refolding carried out at 4°C over several days. Presence of the pro-TEV-NRTN dimer was confirmed by SDS-PAGE, as previous.

3.42 Pro-TEV-NRTN purification

Once refolded, pro-TEV-NRTN was then refolded and purified using the same method as pro-NRTN (summarised in the flow chart of figure 3.6). The final purified protein sample was found to be pure, without major contaminants as seen by SDS-PAGE (See figure 3.6). A faint band can be seen under the main band of pro-TEV-NRTN, this is thought to be degradation products as seen previously with pro-NRTN preparations. Pro-TEV-NRTN gives a clean band between 30 and 40 kDa under non-reducing conditions and under 20 kDa when reduced,

demonstrating the protein to be a disulfide-linked dimer as expected (See figure 3.6) (often a partial reduction of dimer is achieved, and a small amount of dimer can be seen in the reduced sample).

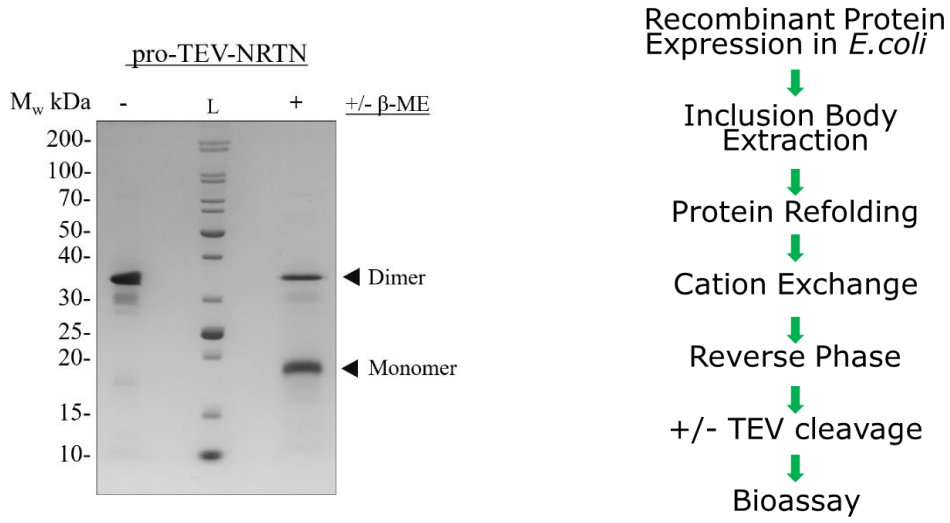


Figure 3.6. Purified pro-TEV-NRTN Sample and Method. Sample of purified pro-TEV-NRTN analysed by SDS-PAGE, in non-reducing and reducing conditions (+β-ME). Dimeric and monomeric forms indicated with flow chart depicting purification method and future uses.

3.5 Refolding and Purification of mature NRTN

Production of mature NRTN was started using the conditions Katharina Ravn had identified. The protein was found to express well in the insoluble fraction when induced at 37°C for three hours. Again, after purification of inclusion bodies screening for appropriate refolding conditions could be completed. Screening of refolding conditions demonstrated mature NRTN to refold in 100 mM TRIS pH 8.5, 1M PPS, 100 mM NaCl, 0.6 M Urea with 0.2 mM L-cystine and 2 mM L-cysteine producing a dimeric protein of expected M_w by non-reducing SDS-PAGE. This was used as the starting point for mature NRTN purification, and a similar approach was used as for pro-NRTN (see figure 3.5 E).

As for pro-NRTN, problems with aggregation were present in the refolding stages - reducing overall yield and homogeneity of the purified protein. Secondary screening of refolding conditions reduced aggregation and improved the overall yield of protein purifications. This was achieved by increasing the concentration of urea, L-cysteine and reducing the final concentration of protein in the refolding solution to ~0.1 mg/ml. The final condition used for

mature NRTN refolding 100 mM TRIS pH 8.5, 1M PPS, 100 mM NaCl, 1.2 M urea with 0.2 mM L-cystine and 8 mM L-cysteine, with refolding carried out at 4°C for 5 days. Under these conditions a faint band at ~25 kDa can be seen by non-reducing SDS-PAGE, consistent with the predicted M_w for dimeric mature NRTN of 26 kDa (see figure 3.7 A).

Once refolded, the protein must then be purified and concentrated. Mature NRTN has a high number of positively charged residues and the theoretical pI of the protein is 9.01 (ExPASy ProtParam). Therefore, cation exchange chromatography was selected as a suitable first step for protein capture and purification from the refolding solution. Cation exchange was performed using a SOURCE 30S column (GE Healthcare) in 6 M urea, 20 mM HCl. Fractions containing dimeric mature NRTN were then pooled for further purification (see figure 3.7 A). Samples were then confirmed to contain the disulfide-linked dimers by the addition of the reducing agent β-ME. Reduced samples run the run below 15 kDa as seen by SDS-PAGE (figure 3.7 B). This matches the predicted size of the mature domain monomer at 13 kDa.

To further purify the sample and remove contaminants, the protein was then purified by RPC (see figure 3.7 C). The final purified protein gives a clear band at 25 kDa consistent with dimeric mature NRTN. Additionally, the purified protein is free of any visible contaminants or degradation products as seen by SDS-PAGE (See figure 3.7 D).

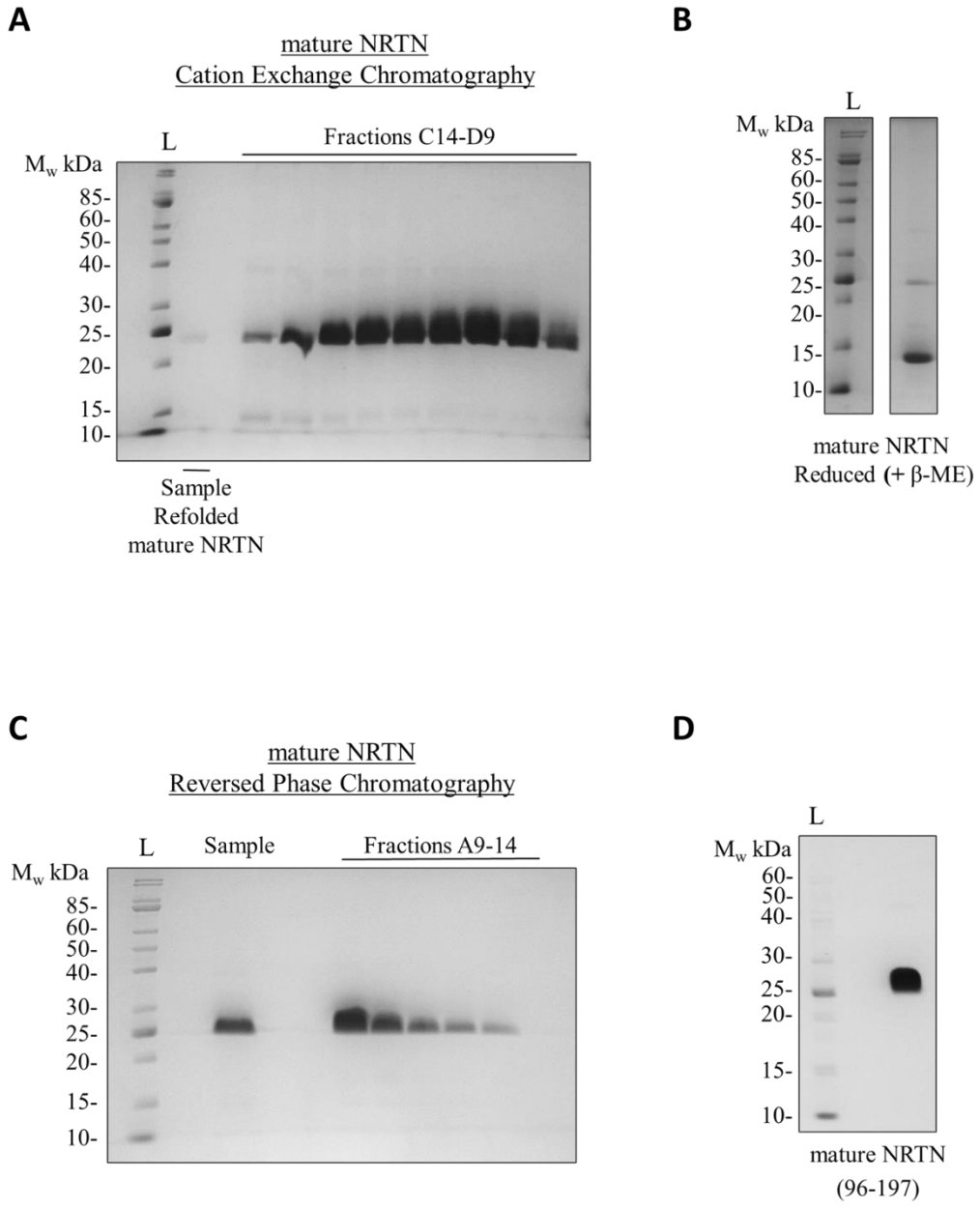


Figure 3.7. Purification of mature NRTN. Protein purification stages for the purification of mature NRTN (96-197) analysed by non-reducing SDS-PAGE. A) Cation exchange chromatography performed on refolded mature NRTN. B) mature NRTN reduced with β-ME C) Reversed phase chromatography performed on fractions from cation exchange. D) Purified and concentrated mature NRTN sample. Panel B is a composite image from one gel with unrelated lanes removed for clarity.

3.6 NRTN Prodomain purification

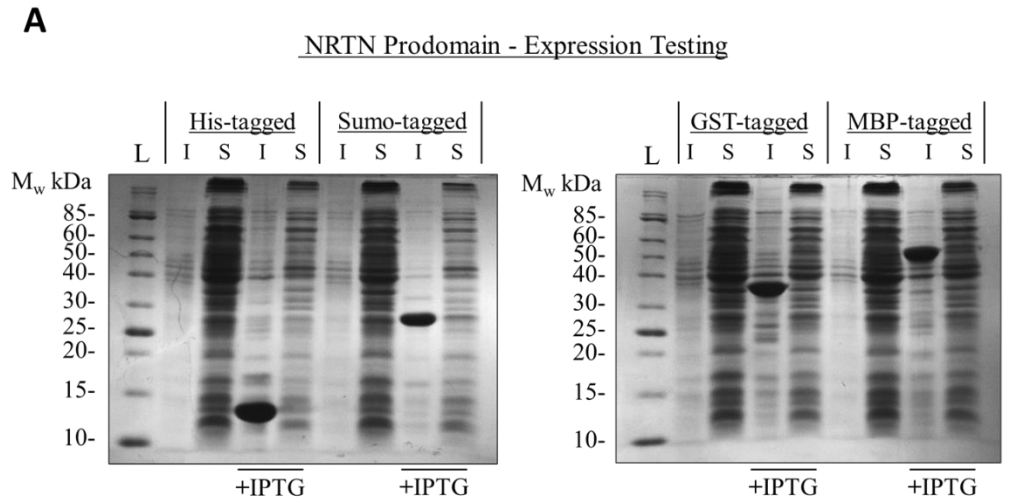
The sequence corresponding to human NRTN amino acids 20-95 was cloned into various protein expression plasmids (See table 3.1) – again the N-terminal cysteine residue was changed to serine for consistency. From here on out the NRTN prodomain 20-95 C23S construct will be referred to as the NRTN prodomain.

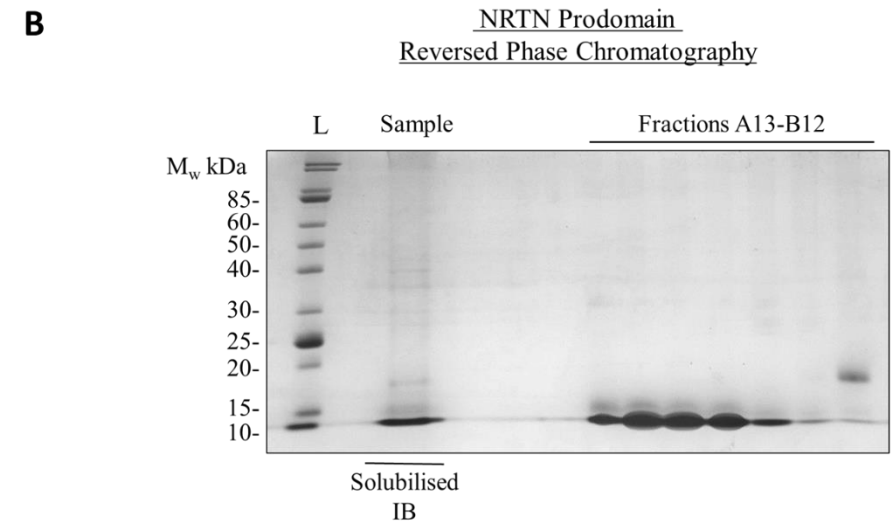
Several commonly used protein N-terminal purification tags were trialled to aid expression and purification of the construct as a fusion protein, including SUMO, GST, MBP and a His-tag. However, expression testing demonstrated all constructs to be insoluble, with relatively large bands corresponding to the predicted NRTN prodomain molecular weight (9 kDa) plus tag seen in the insoluble fraction (see figure 3.8 A - M_w of NRTN prodomain modified to be His-tagged at 12 kDa, SUMO-tagged at 21 kDa, GST-tagged at 35 kDa and MBP-tagged at 54 kDa). Although SUMO is around 11 kDa it is known to run at approximately 15-17 kDa by SDS-PAGE[123], this explains why the SUMO-tagged prodomain runs at almost 30 kDa despite having a M_w of 21 kDa (see figure 3.8 A).

Given the lack of predicted secondary structure (see figure 3.1) and high expression as an insoluble protein the prodomain was then purified from inclusion bodies. We assume formation of α-helices under the right conditions after purification, i.e., the absence of denaturants such as urea and guanidine hydrochloride used for inclusion body extraction and acetonitrile used in protein purification. As protein expression tags made no difference to solubility, an untagged construct was selected for further purification. This also avoids problems with misfolded or aggregated tags such as MBP in the inclusion bodies. Affinity tags, such as MBP, were also deemed unnecessary for purification of pure protein.

The untagged NRTN prodomain construct expressed well level under the same conditions as the fusion constructs and could be purified from inclusion bodies successfully. Solubilised inclusion bodies were found to be relatively pure with a distinct band corresponds to the predicated M_w of the NRTN prodomain - 9 kDa (see figure 3.8 B). Further purification was carried out using RPC. The construct eluted well from RPC with the prodomain visible in fractions collected. Additionally, a larger contaminant between 15-20 kDa seen in the solubilised inclusion body (RPC sample) could be separated from the majority of the protein of interest (see figure 3.8 B). Lyophilised prodomain could be readily resuspended and was

soluble. The final purified protein was pure with no visible contaminant or degradation products visible by SDS-PAGE (see figure 3.8 C).





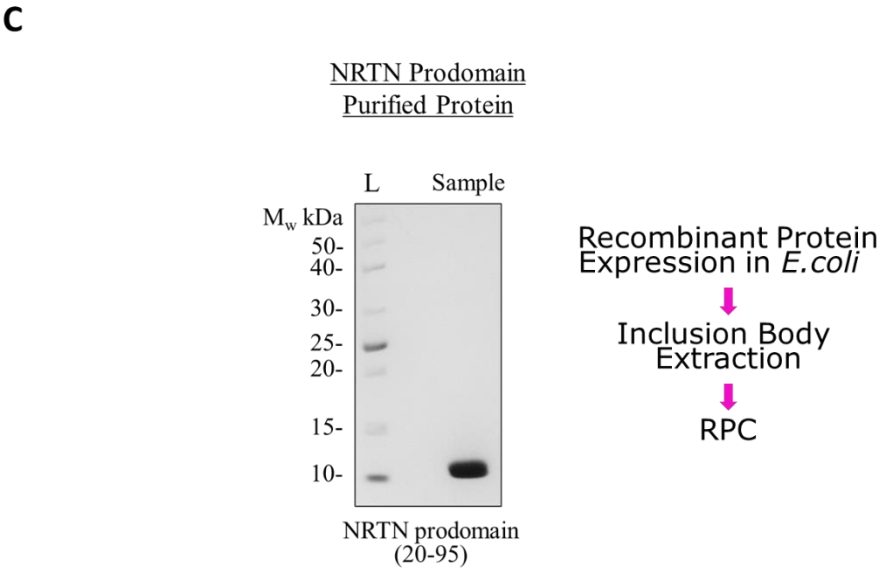


Figure 3.8. The expression and purification of the NRTN prodomain. Expression testing and protein purification stages of the NRTN prodomain (20-95) C23S, analysed by SDS-PAGE. A) NRTN prodomain construct expression testing as a fusion protein with various tags as indicated - glutathione S-transferase (GST), Small ubiquitin-related modifier (SUMO), histidine-tagged (His) and Maltose binding protein (MBP). Insoluble (I) and soluble (S) fractions collected and analysed for all constructs +/- IPTG. Expression of protein constructs induced by the addition of IPTG, in transformed *E. coli* BL21 DE3 cells at 15°C overnight. B) Purification of untagged NRTN prodomain by RPC from solubilised inclusion bodies (IB). C) Purified and concentrated NRTN prodomain sample with summary of purification procedure.

3.7 Pro-NRTN Signalling

NRTN signalling has been characterised previously[58,77,85]. However, past work has focussed solely on the NRTN mature domain and so the signalling activity of the pro-NRTN precursor and cleaved pro-NRTN remain unknown. As mentioned previously, the prodomains of several TGF-β family members affect bioactivity - such as activin A[33]. Therefore, the different forms of NRTN produced were analysed to determine the impact of the prodomain on NRTN signalling activity. The testing of cleaved pro-NRTN was facilitated by using a TEV cleavable pro-NRTN construct (pro-TEV-NRTN). This engineered pro-TEV-NRTN protein has the natural furin cleavage site replaced by a TEV protease site. The TEV site enables controlled cleavage of the pro-protein and the analysis of the cleaved form. In this section I will examine the signalling activity of the various NRTN constructs produced using a cell-based reporter gene assay.

3.8 The Signalling Activity of Mature and pro-NRTN

3.81 Luciferase Based Reporter Gene Assay and Analysis

In order to determine the impact of the prodomain on NRTN signalling a cell-based luciferase reporter gene assay was selected (used by Sandmark *et al.,* 2018[58], Bigalke *et al.,* 2019[77]). This assay uses modified human neuroblastoma TGW cells responsive to NRTN signalling, expressing both RET and GFRα2. As well as expressing the relevant receptors, The TGW cell line was selected by AstraZeneca as it is a robust neuronal cell line and NRTN is a neuronal

factor. The transfected cell line contains a reporter gene construct with a luciferase gene under the control of repetitive serum-response elements (SRE). Activation of the RET receptor by NRTN triggers transphosphorylation and activation of downstream signalling pathways, such as the MAPK pathway. MAPK pathway activation then leads to the expression of firefly luciferase enzyme from the reporter gene construct. Luciferase expression can then be measured by the addition of its substrate luciferin. Luciferase activity generates light, which can be measured and used to determine the level of expression and signalling by NRTN (full details in materials and methods).

Dose-response curves could be obtained for all NRTN constructs measured (see figure 3.1 and 3.2). The enabled the determination of half maximal effective concentration (EC_{50}) values - the dose at which half of the maximum activity is achieved from a baseline. Fitting quality can be assessed by the R^2 value, a statistical indicator of how close the data matches the model fitting. All dose-response curves for NRTN constructs have an R^2 value above 0.95 indicating a good model fitting from the data (see table 3.1 and 3.2).

The Hill equation is used to model ligand binding to macromolecules e.g., receptors as a function of ligand concentration and has been used here to plot dose-response curves and quantify ligand interaction/activity. In this case binding is inferred by signalling activity (output of the reporter-gene assay). The hill slope quantifies the steepness of the curve plotted in the dose response curves between baseline and maximal response (see figure 3.9 and table 3.2 for an examples). The steepness of curves can vary depending on the activity of the ligand - and inhibitory ligands have negative hill slope values instead of positive values. A higher hill slope value e.g., 1.5 for an activating ligand means that smaller increases in ligand concentration are needed to reach a higher signalling activity e.g., maximum signal from baseline. Equally, a smaller hill slope value e.g., 1.0 will require greater concentration increases to increase signalling activity.

Bioactivity using reporter gene assays has been reported in several ways for growth factors e.g., EC_{50} in weight/volume or EC_{50} in molar units. EC_{50} values for growth factors are also reported variably as either for monomer or dimer concentration in the literature. In this work we will use the dimeric protein M_w to determine molarity and report EC_{50} for the dimer as this is the biologically relevant form and the active ligand for NRTN. Additionally, due to

significant M_w differences between NRTN constructs we wish to compare molarity will be used to eliminate any apparent bioactivity differences due to size/M_w.

3.82 Mature NRTN Signalling Activity

Firstly, we set out to demonstrate the growth factors produced were active and comparable to control samples using the luciferase reporter gene assay. Control mature NRTN is known to be bioactive and was provided by AstraZeneca (produced using bacterial expression systems – as used in Trevaskis *et al.,* 2017[85]). As can be seen in figure 3.9A, mature NRTN generates a clear dose-response curve in the bioactivity assay the same as the control. Furthermore, mature NRTN has an EC_{50} of 0.39 nM which is equivalent to the control mature NRTN at 0.34 nM (see table 3.2). Some assay variability is known, for example in subsequent assays with mature NRTN the EC_{50} values varied between 0.25 and 0.40 nM. Additionally, given the EC_{50} values for both mature NRTN and the control are equivalent we can conclude that the protocol used here produces protein of comparable quality as that used by AstraZeneca. Similar EC_{50} values for mature NRTN have been reported in the literature of 0.24 nM and 0.09 nM [58,77] (see table 3.3).

3.83 Pro-NRTN Signalling Activity

Now that we have established the method of NRTN production to be suitable, we can assess the impact of the prodomain on NRTN signalling and activity. Pro-NRTN and mature NRTN signalling was assessed using the same reporter gene assay as previously described and both proteins were tested at equimolar concentrations. The unexpected finding was that the pro-form retains significant bioactivity. In figure 3.9B we can see pro-NRTN dose-response curve is shifted to the right due to its reduced activity, this generates a lower EC_{50} value for the pro-form when compared to mature NRTN. This difference in EC_{50} values between pro-NRTN and mature NRTN is approximately 10-fold at 2.8 nM and 0.25 nM respectively (see table 3.2). This demonstrates the pro-form to be less active than the mature form of NRTN, but we observe a relatively mild inhibitory effect for the prodomain when compared to other pro-TGF-β proteins. For example, pro-activin A and pro-myostatin signalling is completely inhibited by the uncleaved prodomain[32,33].

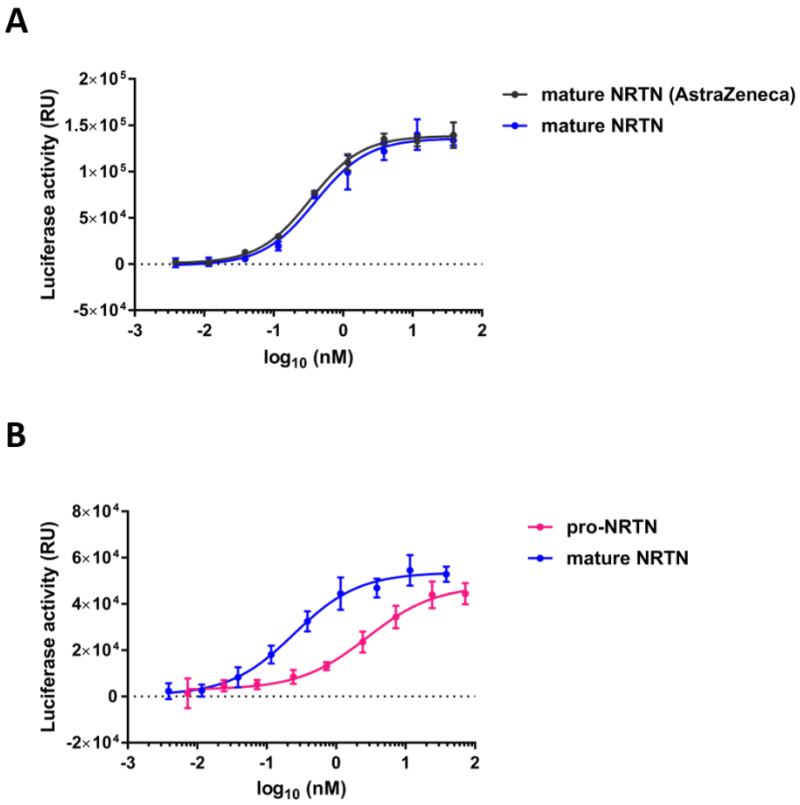


Figure 3.9. Bioactivity of mature and pro-NRTN forms. Fitted dose response curves for NRTN mature and pro-forms from luciferase reporter gene assay in human neuroblastoma TGW cells. A) Refolded and purified mature NRTN shown in blue and the control mature NRTN (AstraZeneca) shown in black. B) mature NRTN shown in blue and pro-NRTN shown in pink. Data shown are the means of triplicate measurements ± SD, n=3.

Sample	Figure	EC_{50} (nM)	EC_{50} (nM) 95 % CI*	Hill Slope	Hill slope 95% CI*	R^2
Mature NRTN	3.1A	0.39	0.27 to 0.59	1.14	0.77 to 1.738	0.98
Mature NRTN	3.1B	0.25	0.16 to 0.38	0.93	0.61 to 1.371	0.96
Mature NRTN AstraZeneca	3.1A	0.34	0.27 to 0.42	1.17	0.93 to 1.48	0.99
Pro-NRTN	3.1B	2.80	1.68 to 6.15	0.88	0.51 to 1.399	0.95

Table 3.2: NRTN construct bioactivity values in human neuroblastoma TGW cells. Pro-NRTN, mature NRTN and control mature NRTN (AstraZeneca). All values presented at two decimal places, n=3. *CI Confidence interval.

Assay type	Published EC_{50}	EC_{50} (nM)*	Protein source	Reference
Luciferase based reporter gene assay	5.6 ng/ml	0.24	Bacterial expression system	Sandmark *et al.*, 2018[58]
Luciferase based reporter gene assay	2.1 ng/ml	0.09	Bacterial expression system	Bigalke *et al.*, 2019[77]

Table 3.3: Published EC_{50} bioactivity values for mature NRTN. EC_{50} vales determined from luciferase-based reporter gene assays in human neuroblastoma TGW cells. All values presented at two decimal places. *EC_{50} (nM) calculated from published values for dimeric mature NRTN.

3.9 Pro-NRTN Cleavage and Signalling Activity

NRTN contains a furin cleavage site located between the prodomain and mature domain. Here we aim to examine the impact of cleavage on cell signalling and activity. To do this we created a TEV cleavable form of NRTN, replacing the natural furin site with a TEV cleavage site. The furin cleavage site is typically just four residues "RXXR". In order not to introduce an artefact by extending the prodomain, eight residues of pro-domain of NRTN were substituted with the eight-residue recognition sequence of TEV (see figure 3.10 A). Cleavage by TEV protease should be highly sequence-specific, relatively efficient *in vitro* and can be easily manipulated. Additionally, with the native furin site replaced the pro-TEV-NRTN construct becomes resistant to furin cleavage, minimising the change of cleavage in cell culture medium. To determine if furin cleavage in the mammalian cell-based bioassays are contributing to pro-NRTN activity, the activity of pro-TEV-NRTN can be used as a control. Similar approaches have been used in the study of activin A and myostatin signalling with HRV-3C (PreScission) protease[32,33].

3.91 TEV Cleavage of pro-TEV-NRTN

To generate cleaved pro-NRTN, purified pro-TEV-NRTN was incubated overnight with TEV protease. TEV protease is a highly sequence-specific cysteine protease and requires reducing conditions to maintain activity. As NRTN contains several disulfide bonds that can be reduced cleavage conditions were also screened to ensure dimeric NRTN remained. Cleavage efficiency was shown to be variable in several conditions, with pro-TEV-NRTN in condition 7

(HEPES pH 7.2 with no salt) shown to be the most efficient at cleavage as analysed by SDS-PAGE (figure 3.10 B). Curiously conditions 5 and 6 in figure 3.10 B result in a higher molecular weight band above pro-NRTN, this may be some form of protein aggregation - pro-NRTN plus a cleavage product for example. These conditions were not used in further experiments. Although some uncleaved pro-TEV-NRTN species remain (at 40 kDa) the majority is cleaved, with bands corresponding to the mature NRTN dimer at 25 kDa and the NRTN prodomain at 10 kDa visible. Condition 7 was then used to produce cleaved pro-TEV-NRTN for testing in reporter gene assays.

3.92 The signalling activity of cleaved pro-TEV-NRTN

With cleaved pro-TEV-NRTN produced we can now examine the impact of cleavage on signalling activity using the same luciferase-based reporter gene assay. Both uncleaved pro-TEV-NRTN and cleaved pro-TEV-NRTN are capable of signalling though RET, generating dose-response curves in the bioassays. Cleaved protein demonstrates increased activity, as can be seen by a shift in the dose-response curve (see figure 3.10C). Uncleaved pro-TEV-NRTN has an EC_{50} of 2.38 nM and the cleaved protein has an EC_{50} of 0.35 nM (see table 3.4). This demonstrates a similar level of inhibition as seen between mature and pro-NRTN in the previous section. TEV cleavage restores the full activity of NRTN with the cleaved pro-TEV-NRTN EC_{50} equivalent to the mature NRTN constructs tested previously (see table 3.2 and 3.4). Furthermore, pro-NRTN, and pro-TEV-NRTN have similar EC_{50} values in the nanomolar range (see table 3.2 and 3.4).

As a negative control, TEV enzyme and/or sample buffer alone was found to induce no signalling activity (data not shown). A less invasive control to prevent furin cleavage could have been used - such as mutation of an Arginine residue to Alanine, in the furin site.

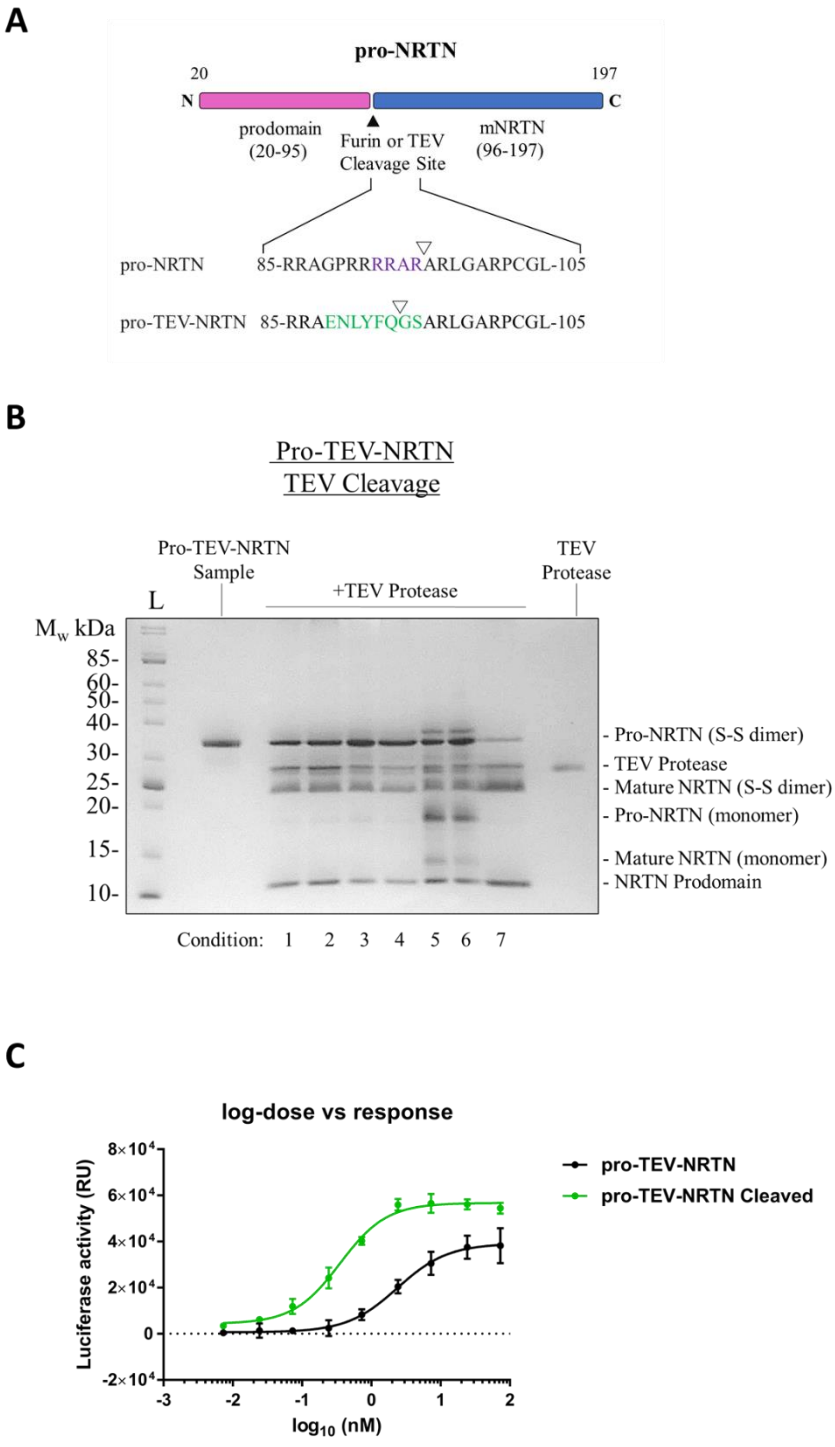


Figure 3.10. Cleavage and bioactivity of pro-TEV-NRTN by luciferase reporter gene assay. A) Schematic of pro-TEV-NRTN construct. NRTN prodomain in pink and mature domain in blue. Natural furin site amino acids in purple and engineered TEV site in green. Cleavage site indicated by white triangle. B) Cleavage of pro-TEV-NRTN by TEV protease under various conditions – 1) 50mM HEPES, pH 7.2, 50mM NaCl, 2) 100mM HEPES, pH 7.2, 50mM NaCl, 3) 50mM HEPES, pH 7.2, 150mM NaCl, 4) 100mM HEPES, pH 7.2, 150mM NaCl, 5) 50mM TRIS, pH 8.5, 50mM NaCl, 6) 50mM TRIS, pH 8.5, 150mM NaCl and 7) 250mM HEPES, pH 7.2.

Analysis of cleavage conditions by SDS-PAGE (non-reducing conditions). S-S (disulfide-linked). C) Bioactivity of pro-TEV-NRTN by luciferase reporter gene assay using human neuroblastoma TGW cells-line. Fitted dose response curves for uncleaved (black) and TEV cleaved (green) pro-TEV-NRTN. Data shown are the means of triplicate measurements ± SD, n=3.

Sample	Figure	EC_{50} (nM)	EC_{50} (nM) 95 % CI*	Hill Slope	Hill slope 95% CI*	R^2
Pro-TEV-NRTN	3.2C	2.38	1.61 to 3.71	1.19	0.78 to 1.86	0.95
Pro-TEV NRTN Cleaved	3.2C	0.35	0.28 to 0.44	1.29	1.01 to 1.65	0.98

Table 3.4: pro-TEV-NRTN construct bioactivity values by luciferase reporter gene assay in the human neuroblastoma TGW cell-line. Cleaved pro-TEV-NRTN cleaved with TEV protease. All values presented at two decimal places, n=3. *CI Confidence interval.

3.10 Furin Cleavage of pro-NRTN

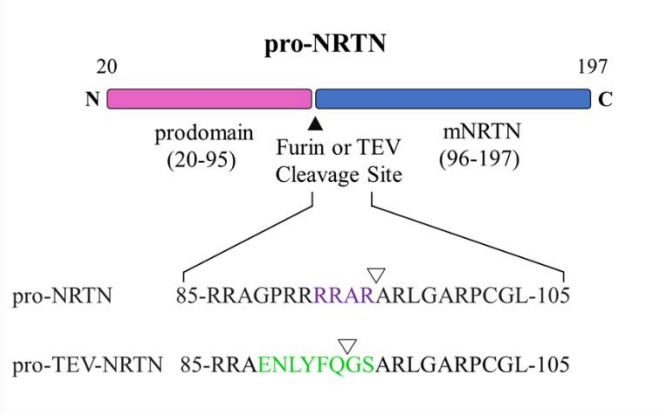
Previously we have used engineered TEV cleavable constructs to cleave pro-NRTN. In this section we will examine the cleavage of pro-NRTN by its native protease, furin *in vitro*. As shown previously pro-TEV-NRTN is fully cleaved by TEV enzyme Now we will determine if furin cleavage produces the same product as TEV protease. We will also determine the specificity of furin cleavage, using non-cleavable constructs.

To confirm furin cleavage pro-NRTN purified proteins were incubated with furin enzyme. The engineered TEV cleavable construct was also subjected to furin incubation. For pro-TEV-NRTN, the natural furin site has been abolished and replaced by the TEV cleavable sequence ENLYFQGS, without altering the length of the protein - avoiding any artefacts an increase in the pro-domain length might introduce (figure 3.11). As TEV cleavable constructs has the furin site abolished, it will be used as a negative control to determine the specificity of furin cleavage *in vitro.*

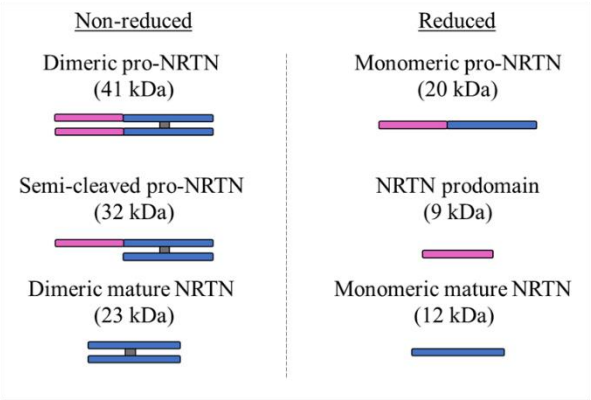
Pro-NRTN can be cleaved by furin enzyme as expected. Furin cleavage results in the production of a mature dimeric NRTN ~ 25 kDa and a monomeric prodomain ~10 kDa, as analysed by non-reducing SDS-PAGE (see figure 3.11 C). This is also in agreement with TEV cleavage products of pro-TEV-NRTN discussed previously in this chapter. Interestingly, the

negative control with the furin site abolished, the pro-TEV-NRTN construct can also be cleaved by furin. A smaller band for pro-TEV-NRTN is seen after furin incubation in non-reducing conditions (see figure 3.11 A and D). No secondary cleavage product is seen, as it is likely to small to remain on the gel. An alternative RXXR furin cleavage site was found within the sequence of the NRTN prodomain – RIAR. Searches of a database of furin cleavage site motifs called FurinDB[124] identified another protein, human collagen type XXV alpha 1, with the same RIAR site known to be cleaved by furin[125]. However, as the prodomain band remains at 10 kDa in wild-type pro-NRTN furin cleavage we can conclude this is not the preferred cleavage site for furin (see figure 3.11 C).

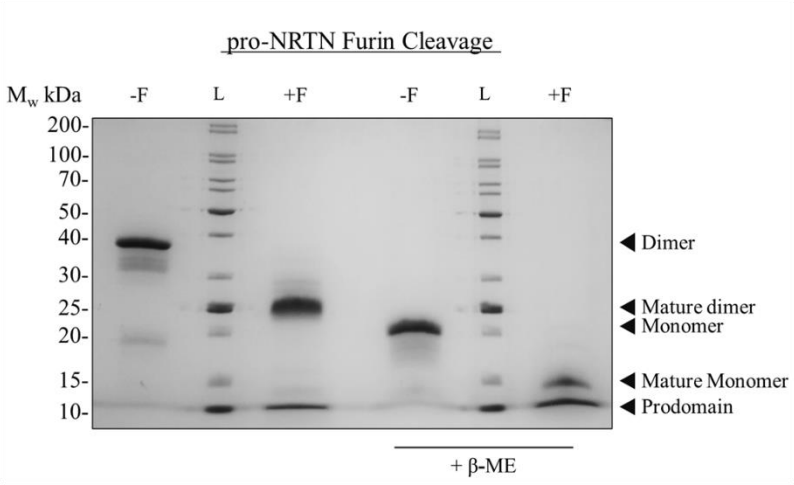
A


pro-NRTN
20
197
N
C
prodomain
(20-95)
Furin or TEV
Cleavage Site
mNRTN
(96-197)
pro-NRTN 85-RRAGPRRRRARARLGARPCGL-105
pro-TEV-NRTN 85-RRAENLYFQGSARLGARPCGL-105

B


Non-reduced
Reduced
Dimeric pro-NRTN
(41 kDa)
Monomeric pro-NRTN
(20 kDa)
Semi-cleaved pro-NRTN
(32 kDa)
NRTN prodomain
(9 kDa)
Dimeric mature NRTN
(23 kDa)
Monomeric mature NRTN
(12 kDa)

C


pro-NRTN Furin Cleavage
Mw kDa
-F
L
+F
-F
L
+F
200-
100-
70-
50-
40-
30-
25-
20-
15-
10-
Dimer
Mature dimer
Monomer
Mature Monomer
Prodomain
+ β-ME

D

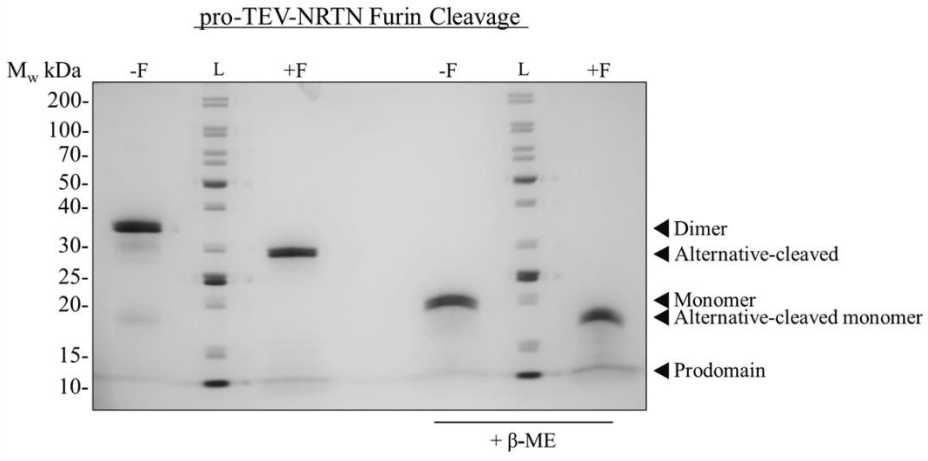

pro-TEV-NRTN Furin Cleavage
Mw kDa
-F
L
+F
-F
L
+F
200-
100-
70-
50-
40-
30-
25-
20-
15-
10-
Dimer
Alternative-cleaved
Monomer
Alternative-cleaved monomer
Prodomain
+ β-ME

Figure 3.11. Furin cleavage of pro-NRTN. A) Schematic of pro-NRTN constructs – with a natural furin site (purple) and pro-TEV-NRTN with an artificial TEV site (green). The location of cleavage site indicated by white arrows. B) Schematic of theoretical NRTN furin cleavage products. Prodomain shown in pink, mature domain in blue and with disulfide bonds shown in grey. Molecular weight of each product indicated, under non-reducing and reducing conditions. C and D) Non-reducing and reducing SDS-PAGE analysis of NRTN constructs post furin incubation.

3.11 The NRTN Prodomain: Structure and Interactions

In this section we will examine the secondary structure and interactions of the NRTN prodomain using biophysical methods. The structure of the mature domain of NRTN has been previously characterised, however the structure of the prodomain remains unknown[58,73,77]. As discussed in the introduction the prodomains of other TGF-β family members can reduce bioactivity, as seen with pro-NRTN in this chapter. This is the case for activin A, myostatin and others with their prodomains interacting with the mature domains to block receptor binding sites and activation of cell signalling[16,32,33,45]. An interaction between the prodomain and mature domain of NRTN could also explain a reduction in bioactivity observed for pro-NRTN. Therefore, we will aim to determine the structure and inter-domain interactions for NRTN. NRTN constructs produced previously will be analysed to determine secondary structure using CD and interactions between pro and mature domains will be examined by ITC and HDX-MS.

3.12 Structure of the NRTN prodomain

To analyse the structure of the prodomain we used circular dichroism (CD). CD uses polarised light to investigate the secondary structure of proteins. Chiral molecules, such as proteins, may respond to differently to polarised light with slight differences in absorbance. This forms the basis of CD, with the different secondary structural elements of proteins such as α-helix, β-sheet and disordered/unstructured regions generating different spectra. The characteristic shapes of spectra for secondary structural elements are known and can be used to determine the overall secondary structure of proteins[126]. However, many deconvolution methods do not accurately determine the secondary structure of largely unstructured proteins as these methods are mainly based on data generated from structured proteins[127,128].

Analysis of NRTN constructs by CD reveals the prodomain to be largely unstructured in solution. Spectra for the prodomain display a characteristically unstructured profile (see figure 3.12). However, the spectra of the full-length protein, pro-NRTN, is not a perfect summation of the two individual domains independently. This suggests a potential structural change for the prodomain as part of pro-NRTN (See figure 3.12). Deconvolution of the mature NRTN spectra was possible and performed as a positive control using the BeStSel web server[129]. Deconvolution resulted in 42% β-sheet, 13% turn, 3% helix and 42% other. This is as expected when compared to the known structure of mature NRTN[58] however, helical content is underestimated compared to the crystal structure.

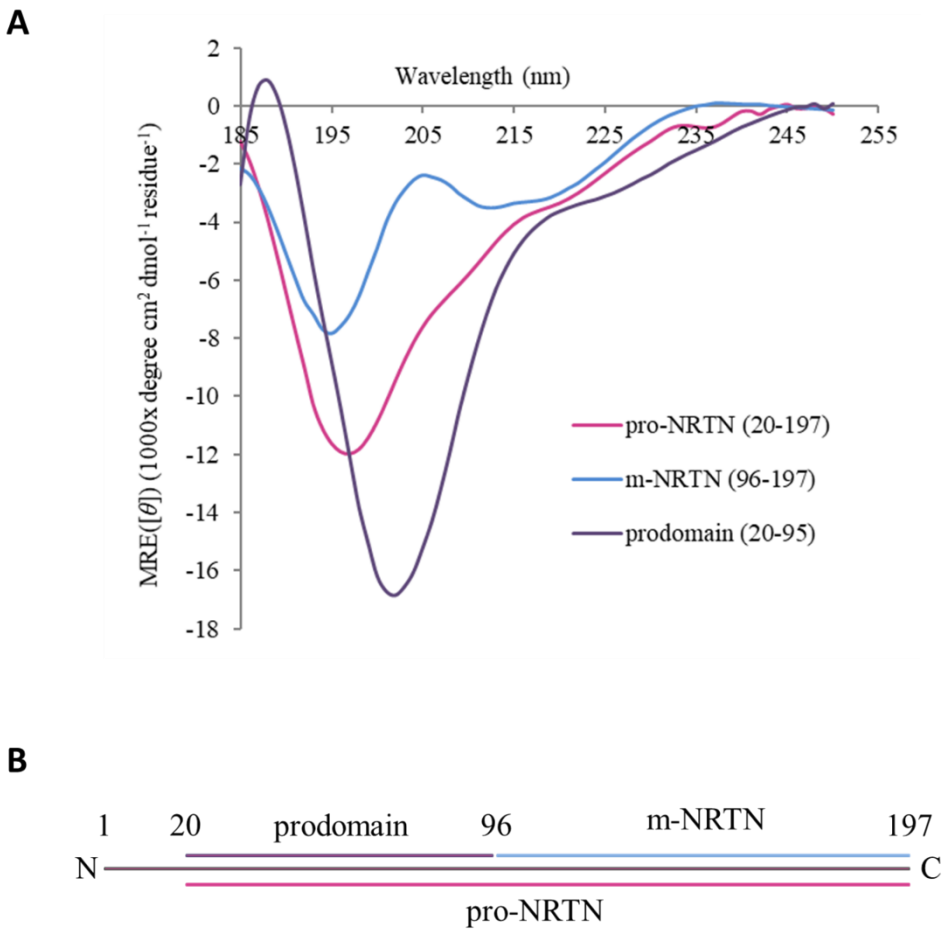


Figure 3.12. Circular dichroism analysis of NRTN constructs. A) Far UV circular dichroism (CD) spectra of pro-NRTN in pink, mature NRTN (m-NRTN) in blue and the NRTN prodomain shown in purple (amino acid residue numbers indicated). CD data recorded between 250 and 185 nm at 1 nm increments. Data shown in mean residue ellipticity, MRE ([θ]) (1000× degree cm^2 dmol^{-1} residue^{-1}). B) Schematic of pro-NRTN constructs CD.

3.13 Interaction between NRTN Pro- and Mature Domains

3.13.1 Isothermal titration calorimetric analysis

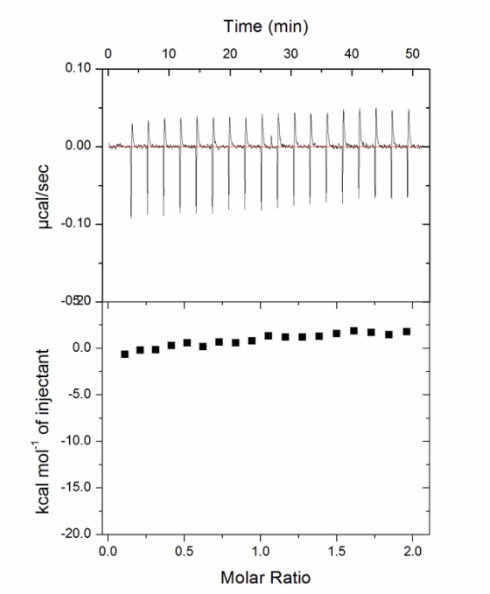


Figure 3.13. Prodomain and mature NRTN interaction via ITC. Thermogram and isotherm traces generated by isothermal titration calorimetry (ITC) for the titration of the NRTN prodomain into the mature NRTN.

Initial experiments to examine the potential interaction between the NRTN prodomain and mature domains used individual purified proteins. Isothermal titration calorimetry (ITC) can be used to examine the interaction between molecules and is often used to characterise protein-protein interactions. Heat changes upon interaction (or dissociation) between molecules are used to determine properties such as binding affinity. Binding between the prodomain and mature NRTN was not detected by ITC, as no significant heat changes are observed when injecting prodomain into mature domain (see figure 3.13). However, mixing the prodomain and mature domain together may not produce a genuine interaction by ITC. One component may be aggregated and so may prevent binding. Additionally, the protein interaction may be formed as both components are synthesised – i.e., it may require co-translation.

3.13.2 Interactions via HDX-MS

In the previous section we observed that no interaction could be detected between the prodomain and mature domain of NRTN in initial ITC experiments. However, these initial experiments used individual purified proteins. We know from previous sections in chapter 3, that the prodomain of NRTN is only inhibitory to bioactivity before cleavage. Therefore, we sought to determine if any interaction could be detected in uncleaved pro-NRTN between the prodomain part and the mature domain. When covalently linked, as in uncleaved pro-NRTN, the relative prodomain concentration will be increased and its orientation will be constrained by the covalent linkage. This would be expected to enhance any interactions between the domains.

In order to see if any interaction between the two domains could be detected when covalently linked as pro-NRTN, we chose to use Hydrogen–deuterium exchange mass spectrometry (HDX-MS). Hydrogen atoms on some chemical groups can be exchanged to deuterium, including amide hydrogens of the protein backbone (NH to ND). This is exploited in HDX-MS to gain information on protein structure, dynamics, and function. HDX-MS has been successfully employed to identify protein-protein interaction sites, protein conformational changes and folding, protein-membrane and nucleic acid interaction sites as well as small-molecule binding at single-amino acid resolution[130].

Once the protein is exposed to heavy water (D_2O), H-D exchange reactions with the solvent occur rapidly for main chain amines in particular. These reactions are also reversible and can occur from milliseconds to days. Exchange reaction time is dependent on the chemical environment i.e., the degree of protection from the surrounding solvent. Groups that are buried within a structure or involved in hydrogen bonding are exchanged less quickly than exposed groups. The exchange reaction can be rapidly quenched by low pH and temperature. Exposed proteins are then digested to peptides and analysed by mass spectrometry for changes to molecular weight (deuterium ~2 Da, hydrogen ~1 Da). Typically, proteins are sampled at a series of timepoints to gain information on the exposure, exchange rate and structure of the protein[130,131].

Good peptide coverage of the complete protein of interest is essential in HDX-MS. Unfortunately, coverage from mature NRTN was insufficient to analyse its interaction with

the prodomain. This is likely due to its high stability and resistance of heavily disulfide-linked mature NRTN to proteolysis in the chemical conditions required for HDX-MS. However, digestion and MS of the prodomain produced high levels of peptide coverage across the protein (see figure 4.14 A). Once this was established, we then sought to compare deuterium uptake for the prodomain in the presence and absence of the mature domain. This was achieved using the purified prodomain and pro-NRTN constructs (for full details of HDX-MS experiments see Materials and Methods).

Analysis by HDX-MS identifies no significant difference between deuterium uptake of the prodomain and the prodomain as part of pro-NRTN at the timepoints collected. This can be seen in the heatmap of deuterium uptake for the NRTN prodomain and pro-NRTN. Throughout the prodomain sequence at the four timepoints collected, both states display similar levels of relative fractional uptake (see figure 4.14 B). Individual uptake plots for all peptides analysed did not display a significant difference between states. The significance threshold for interaction changes was set at 0.5 Da as standard[130,131]. In summary no interaction between the prodomain and mature domain of NRTN could be detected by HDX-MS.

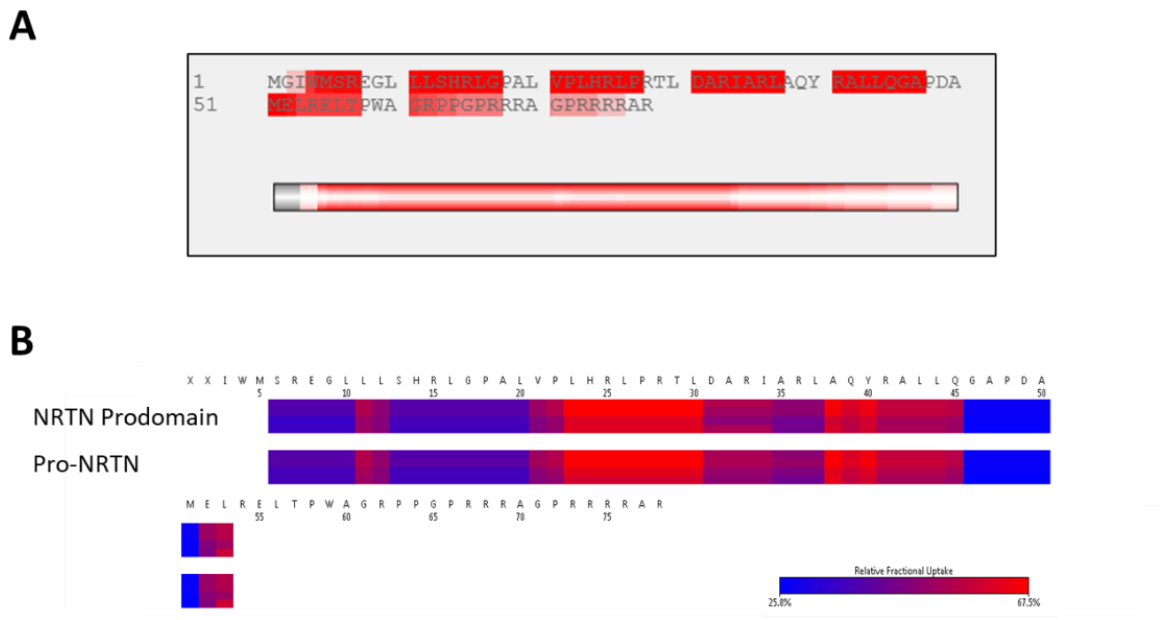


Figure 4.14. NRTN prodomain and mature domain interactions by HDX-MS. Hydrogen–deuterium exchange mass spectrometry (HDX-MS) data for NRTN constructs. Deuterium uptake compared between the NRTN prodomain and the prodomain covalently attached to the mature domain as uncleaved pro-NRTN. A) Peptide coverage of the prodomain by MS,

created using PLGS v3.0.2. Peptide coverage compared to theoretically possible peptide coverage and colour coded as a gradient according to detected peptide number from red to white (high coverage in red to low coverage in white). B) combined heatmap of uptake of deuterium uptake for the NRTN prodomain and pro-NRTN, across the prodomain sequence at the four timepoints collected. Created using DynamX v3.0. For full details see materials and methods.

3.14 Conclusions

This chapter describes the production and characterisation of NRTN pro-forms, through recombinant protein expression and purification, cell-signalling assays, enzyme cleavage and biophysical analysis.

Firstly, we describe the development and optimisation of protocols to produce human NRTN constructs; mature NRTN, pro-NRTN, pro-TEV-NRTN and the NRTN prodomain from bacterial expression systems. mature NRTN, pro-NRTN and pro-TEV-NRTN were all refolded and purified successfully and each protein was demonstrated to be a disulfide-linked homodimer as expected. Despite the lack of cysteine residues, the NRTN prodomain also proved to be insoluble in bacterial expression systems and so was purified from inclusion bodies also. All proteins were deemed to be relatively pure being free from major contaminants and degradation products, as analysed by SDS-PAGE. Purified NRTN constructs can now be used to examine the bioactivity, interactions, and structure of pro-NRTN and the prodomain.

Next, we examined the bioactivity of mature and pro-NRTN and determined the impact of cleavage on cell signalling using a TEV cleavable construct. Bioactivity was determined using a cell-based luciferase reporter gene assay. Mature NRTN was found to be equivalent to control mature NRTN (AstraZeneca), demonstrating the protein of equal quality with equivalent EC_{50} values to the control. The prodomain was found to be inhibitory to cell signalling, as mature NRTN is approximately 10-fold more active than pro-NRTN. Addition of the TEV cleavage site to pro-NRTN was determined to have no impact on bioactivity but enabled the production of cleaved pro-NRTN for testing. Once cleaved the prodomain has no effect on NRTN signalling because cleaved pro-NRTN has an equivalent level of signalling activity as mature NRTN.

As expected, furin can cleave pro-NRTN *in vitro*. Furin produces the same cleavage products as seen with TEV cleavage, producing a mature dimer and a free prodomain. Furin cleavage can be abolished by mutation of the furin site, preventing the cleavage between the pro and mature domains. However, pro-NRTN has a secondary or alternative furin site, that can be cleaved *in vitro* and is also found in a specific collagen. This alternative site is not preferred, as processing is only visible when the main furin site is mutated.

Finally, we examined the secondary structure and interactions of the prodomain using biophysical methods. Structural characterisation through CD reveals the NRTN prodomain to be largely unstructured and binding experiments with ITC display no interaction between the prodomain and mature NRTN. Through HDX-MS, no interaction between the NRTN prodomain and mature domain could be identified. In summary, the NRTN appears to be unstructured and does not interact with the mature domain.

Chapter 4 - Production and Characterisation of GDF15 pro-forms

In this chapter we will examine the production and characterisation of GDF15 variants and pro-forms, through recombinant protein expression and purification, cell-signalling assays, enzyme cleavage and biophysical analysis.

4.0 Production and Design of GDF15 Constructs

To assess the role of the prodomain on GDF15 activity and signalling protein samples must first be produced. Using the same approach as in the previous NRTN chapter, the GDF15 prodomain, mature domain and pro-GDF15 will all be produced as well as a TEV-cleavable form - pro-TEV-GDF15. Additionally, we will produce the GDF15 H202D variant of interest as a mature protein to examine the variants impact on bioactivity and quantification by ELISA. In this first section we will address the production of recombinant GDF15 constructs.

Human GDF15 is 308 amino acids (aa) in length and consists of an N-terminal signal sequence (1-29), a prodomain (30-196) and a mature domain (197-308) which is the active growth factor (UniProt: Q99988) (see figure 4.1). Plasmids encoding mature and pro-GDF15 were available in the group (Dr Thomas Cotton, unpublished) and were used as the templates for the cloning of additional GDF15 constructs (see table 4.1) with sequences codon optimised to facilitate protein expression in *E. coli*. From secondary structure analysis we can see that the GDF15 prodomain predicted to be more structured than the NRTN prodomain, with high confidence predictions for three α-helical segments. However, the majority of the protein contains no predicted structural elements (see figure 4.1). The GDF15 prodomain also contains a putative N-linked glycosylation site at amino acid residue 70 (UniProt - Q99988, see figure 4.1 B).

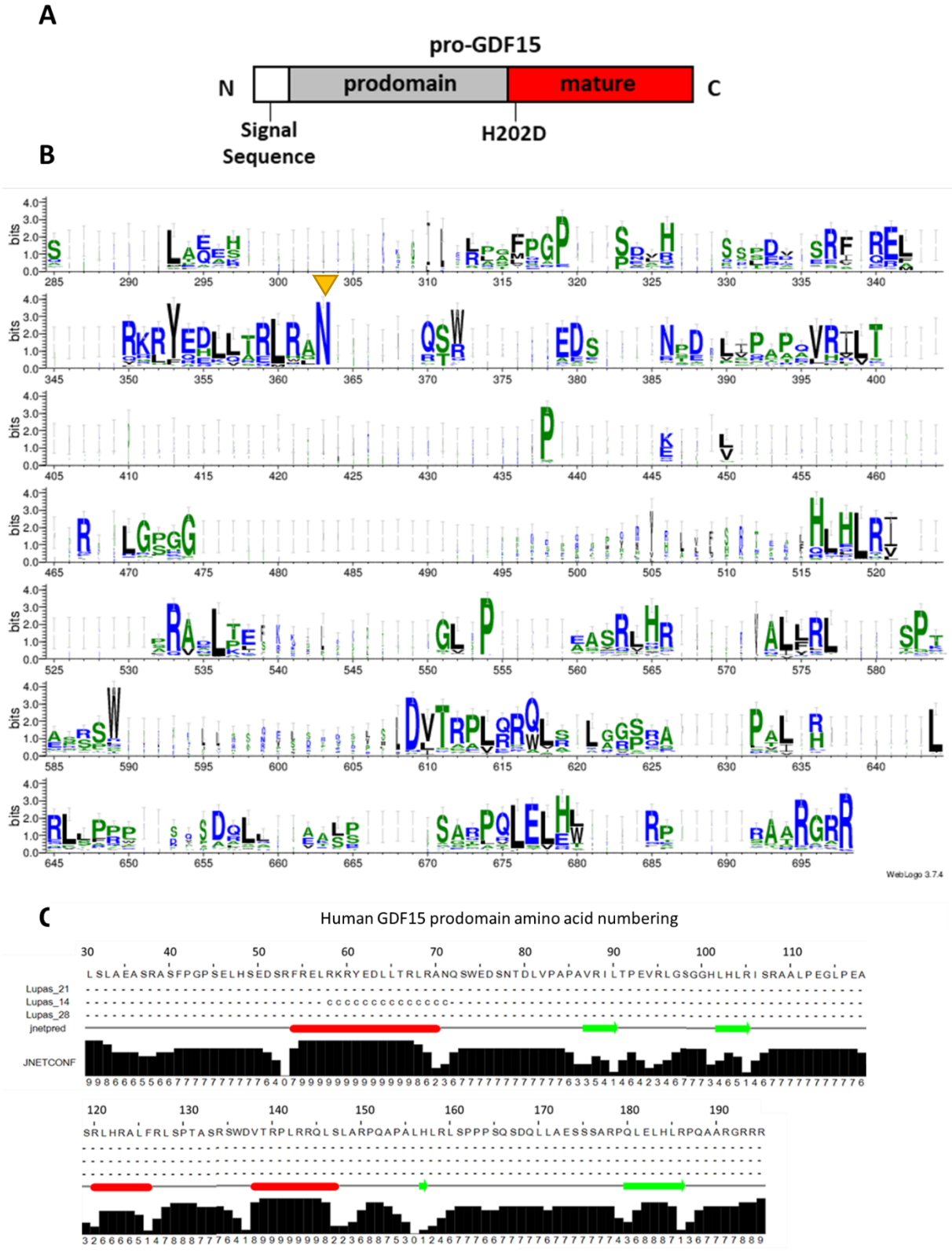


Figure 4.1. Pro-GDF15. A) Schematic of human pro-GDF15 protein, prodomain in grey and mature domain in red. Position of the H202D site indicated. Domains of Pro-GDF15 proportionate to size in amino acids. B). Conservation of GDF15 prodomain (amino acids 30-196). Sequences for GDF15 orthologues taken from UniProt (https://www.uniprot.org/) (109

sequences, from fish to mammals) and aligned in MUSCLE[24] (available at https://www.ebi.ac.uk/Tools/msa/muscle/) and created using WebLogo3[117] (available at http://weblogo.threeplusone.com/create.cgi). Yellow triangle signifies a conserved putative N-linked glycosylation site. C) GDF15 prodomain Secondary structure predication, created using Jpred4[118] (available at https://www.compbio.dundee.ac.uk/jpred/). Secondary structural elements, β-strands shown in green and α-helical regions shown in red with conservation of each helical region below. Prediction confidence level is indicated by the back bars.

Plasmid	Insert	amino acids	Mutations
pHAT2	His-Pro-GDF15*	30-308	-
pHAT2	His-Pro-TEV-GDF15*	30-308	TEV site**
pHAT2	His-GDF15 prodomain*	30-196	-
pHAT2	His-GDF15 prodomain*	47-196	-
pOP1B	mature GDF15*	197-308	-
pOP1B	mature GDF15*	197-308	H202D
pOP1	N-terminal-Strep-tag II + mature GDF15	197-308	-
pOP1	N-terminal-Strep-tag II + mature GDF15	197-308	H202D
pHAT4	mature GDF15	197-308	-
pHAT4	mature GDF15	197-308	H202D

Table 4.1 Cloned GDF15 Constructs. Summary of human GDF15 constructs used for protein expression. All constructs were sequenced to confirm 100% identity to the desired sequence. Inserts were also confirmed to be within the correct reading frame within the vectors. Vectors used; pHAT2[116], pOP1 and pOP1B plasmids (https://hyvonen.bioc.cam.ac.uk/pOP-vectors). pOP1B plasmid contains an upstream minicistron to aid expression. *indicates donated material, cloned by Dr Thomas Cotton and Dr Marko Hyvönen. ** furin protease site replaced by recombinant TEV protease site for TEV cleavage *in vitro.*

4.1 Refolding and purification of mature GDF15 WT and H202D

4.11 Refolding of mature GDF15 WT and H202D

Previous work in the Hyvönen Group, by Dr Thomas Cotton, determined a suitable protocol for the expression, refolding and purification of mature GDF15. This protocol has been adapted and followed here for the purification of mature GDF15 wild type (WT) and the H202D variant. In brief, GDF15 constructs were expressed in *E. coli* BL21 (DE3) cells at 37°C for 3 hours. Inclusion bodies were then harvested and purified as previous (see materials and methods) before being solubilised and the protein refolded in 100 mM TRIS pH 8.0, 1.2 M urea, 1 M PPS with 0.2 mM L-cystine and 2 mM L-cysteine. Final protein concentration was 0.1 mg/ml and refolding was performed at 4°C for 2 weeks. Lastly, presence of refolded dimeric species was confirmed by non-reducing SDS-PAGE as can be seen by the band between 20 and 25 kDa markers (predicted M_w 25 kDa) (see figure 4.2).

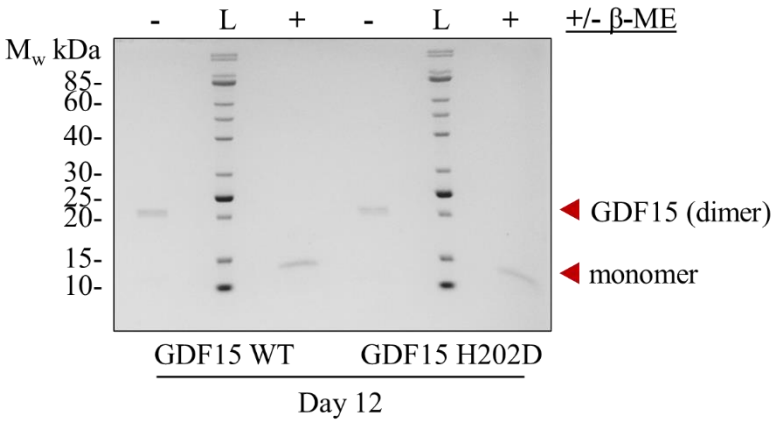


Figure 4.2. Refolding of GDF15 WT and H202D constructs. Refolding of mature GDF15 constructs analysed by SDS-PAGE, in non-reducing and reducing conditions (+β-ME). Dimeric and monomeric forms indicated. Proteins refolded at 0.1 mg/ml in 100 mM TRIS pH 8.0, 1.2 M urea, 1 M PPS with 0.2 mM L-cystine, 2 mM L-cysteine at 4°C for 12 days.

4.12 Purification of mature GDF15 WT and H202D

Once refolded, mature GDF15 constructs could be purified (method summarised in figure 4.4). RPC was selected as the initial capture step from the refolding solution and was able to bind and elute the dimeric GDF15 as seen by SDS-PAGE with bands at 20 kDa (see figure 4.3 A and B). However, a significant contaminant can be seen at 10 kDa, this is likely the monomeric form of GDF15 from the initial refolding solution. The high molecular weight bands, above 85 kDa, are thought to be gel artefacts, as they do not appear at every stage of

the purification. In this case the gel artefacts are likely due to high protein concentration, overloading the gel and preventing samples from migrating to the expected location on the gel. The same samples that exhibit this in the final RPC step (see figure 4.3 E and F) do not display this when diluted to lower concentrations for the final sample gels (see figure 4.4). The fractions containing dimeric GDF15 were further purified by cation exchange chromatography. Cation exchange proved successful in removing the majority of unwanted monomeric species from the sample, with most monomeric GDF15 eluting in earlier fractions under lower concentrations of salt – as seen in fractions C8 and C3 (see figure 4.3 C and D). RPC was performed as the final purification step, resulting in relatively pure protein for the dimeric GDF15 constructs (see figure 4.3 E and F). Although both proteins are relatively pure after ion exchange, RPC enables proteins to be buffer exchanged from urea into acetonitrile-based solutions for lyophilisation and storage as powder. Proteins can then be resuspended into more suitable buffers for cell-based assays, higher concentrations of urea may be detrimental to cells.

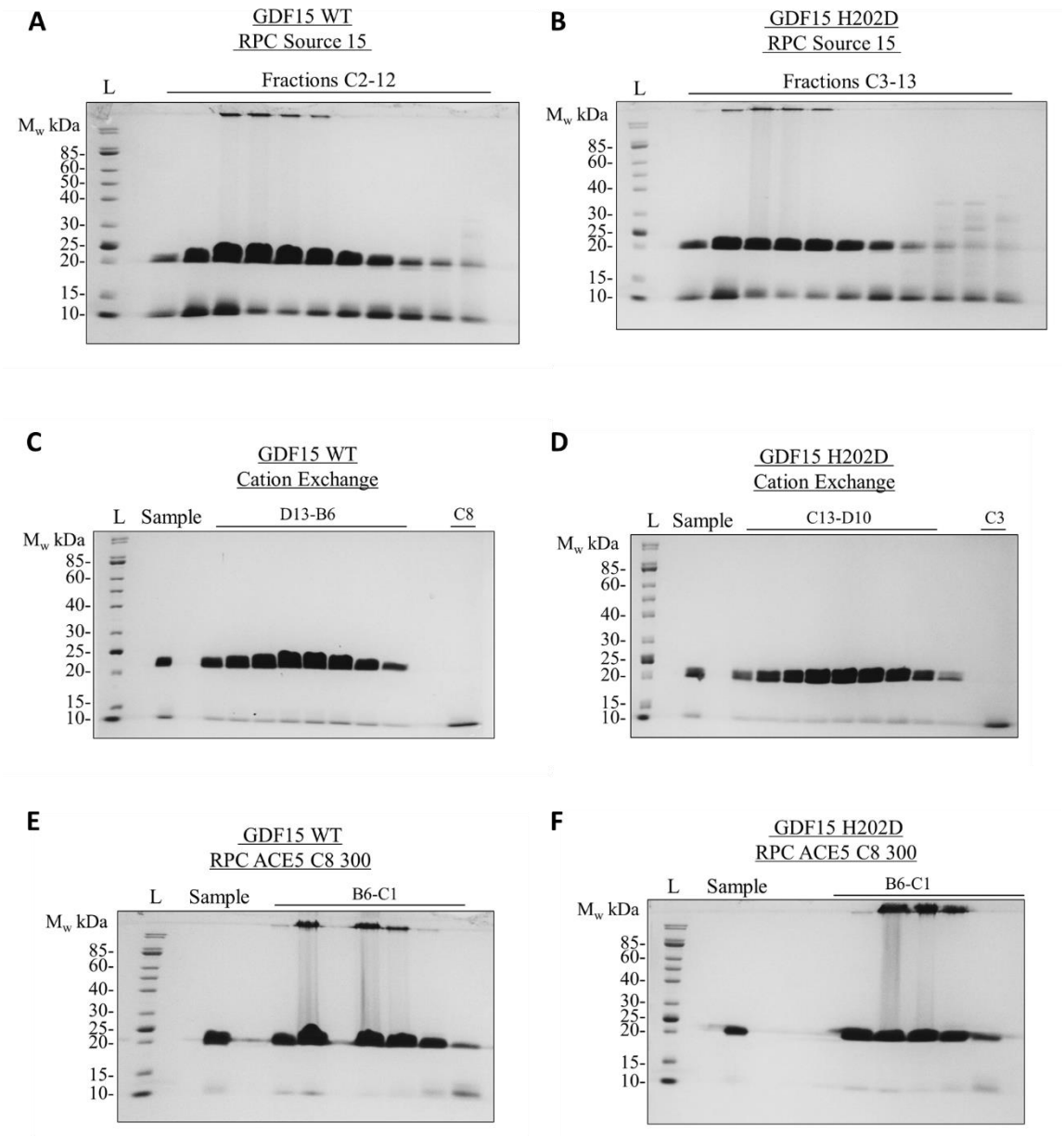


Figure 4.3. Purification of mature GDF15 constructs. Purification of mature GDF15 wild type (WT) and H202D variant proteins analysed by non-reducing SDS-PAGE. A and B) Reversed phase chromatography (RPC) of GDF15 WT and H202D. C and D) Cation exchange of GDF15 WT and H202D. E and F) Final RPC.

Final samples of purified GDF15 WT and H202D constructs were confirmed to be disulfide-linked dimers by the addition of the reducing agent β-ME. Both proteins give clean bands at approximately 20 kDa for the non-reduced dimeric form, and bands at 10 kDa for the reduced monomeric form. Samples are of high purity with no visible contaminants (see figure 4.4).

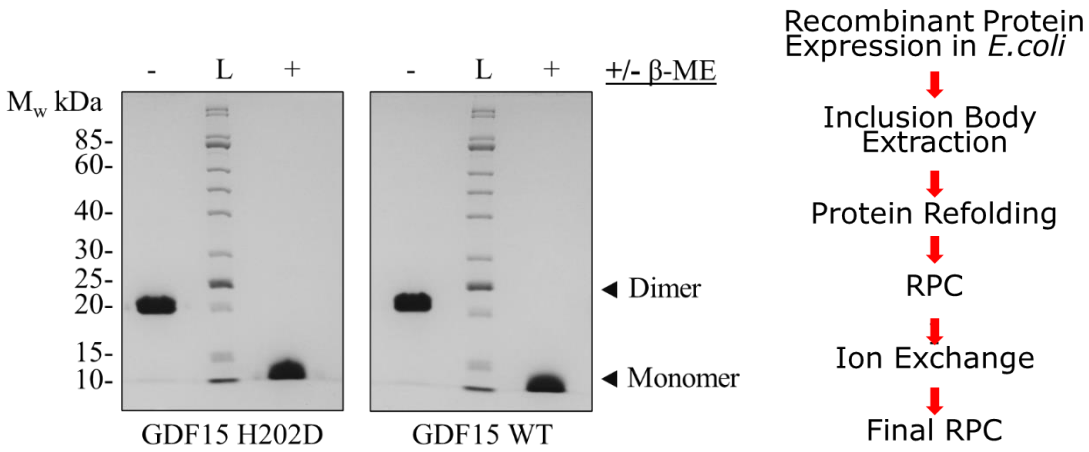


Figure 4.4. Purified mature GDF15 WT and H202D and Purification Summary. Samples of purified mature GDF15 proteins analysed by SDS-PAGE, in non-reducing and reducing conditions (+β-ME). Dimeric and monomeric forms indicated. Summary of protein purification with key stages listed.

4.2 Refolding and purification of pro-GDF15 and pro-TEV-GDF15

4.21 Refolding of pro-GDF15 and pro-TEV-GDF15

Initial work on pro-GDF15 in the Hyvönen Group, by Dr Thomas Cotton, determined suitable protocols for the expression, refolding and purification of pro-GDF15. Expression and purification of pro-GDF15 constructs were performed in the same way as mature constructs, as described earlier. Both pro-GDF15 and pro-TEV-GDF15 were refolded under the same conditions at a final protein concentration of ~0.15 mg/ml. Refolding was successful in 100 mM TRIS pH 9.0, 1.2 M urea, 1 M PPS with 0.2 mM L-cystine and 3 mM L-cysteine at 4°C for 3 to 5 days. Successful refolding was confirmed by the presence of dimeric pro-GDF15 constructs at ~60 kDa as analysed by SDS-PAGE (see figure 4.5). Monomeric species at ~30 kDa remain in both refolding samples and reduction of the dimer is incomplete.

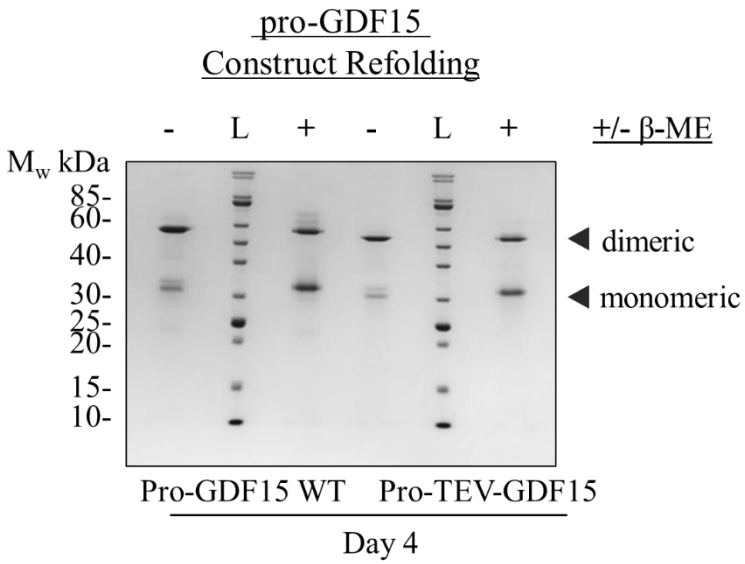


Figure 4.5. Refolding of pro-GDF15 constructs. Refolding of pro-GDF15 and pro-TEV-GDF15 constructs analysed by SDS-PAGE, in non-reducing and reducing conditions (+β-ME). Dimeric and monomeric forms indicated. Proteins refolded at ~0.15 mg/ml in 100 mM TRIS pH 9.0, 1.2 M urea, 1 M PPS with 0.2 mM L-cystine, 3 mM L-cysteine at 4°C for 4 days.

4.22 Purification of pro-GDF15 and pro-TEV-GDF15

For purification of refolded pro-GDF15, samples were adjusted to pH 8.0 to ensure column interaction during cation exchange chromatography (HiTrap SP HP). After elution via a salt gradient, fractions containing dimeric protein were then taken for further purification using size exclusion chromatography (SEC - HiLoad 16/60 Superdex 200). Previous work with pro-GDF15 identified precipitation and aggregation issues during SEC trials preventing the purification of a soluble and pure protein. These issues were improved using 2 M urea, 500 mM NaCl, 50 mM TRIS pH 8.0 to prevent aggregation and charge based interactions with the column resin. SEC was successful in separating dimeric pro-GDF15 from the majority of monomeric species and contaminants remaining after cation exchange (see figure 4.6 A and B). The final purified pro-GDF15 protein was shown to be pure, and a disulfide-linked dimer is the majority species (see figure 4.6 C). A faint contaminant at ~30 kDa is viable, this is likely to be residual monomeric material.

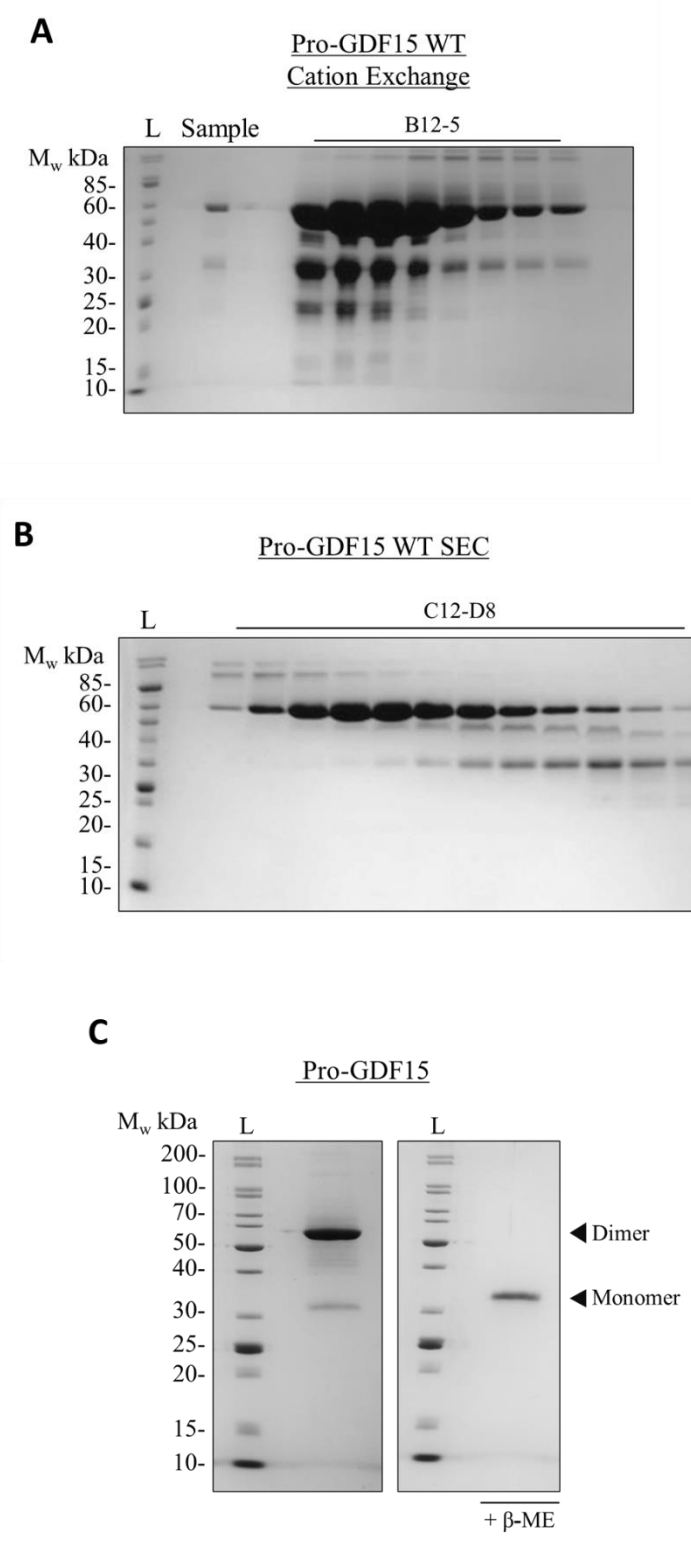


Figure 4.6. Purification of pro-GDF15. Purification of pro-GDF15 analysed by non-reducing SDS-PAGE. A) Cation exchange of pro-GDF15.B) Size exclusion chromatography (SEC) for pro-GDF15. C) Sample of purified pro-GDF15 analysed by SDS-PAGE, in non-reducing and reducing conditions (+β-ME). Dimeric and monomeric forms indicated.

Once refolded, Pro-TEV-GDF15 was produced using the same method as pro-GDF15 (summarised in figure 4.7). The protein was purified to homogeneity displaying a clean band of the correct molecular weight by SDS-PAGE for both oxidised disulfide-linked dimer (above 50 kDa) and reduced monomer (above 30 kDa) as expected (see figure 4.7). As previously described, the disulfide-linked dimer can run at a lower M_w than predicted due to its compact structure.

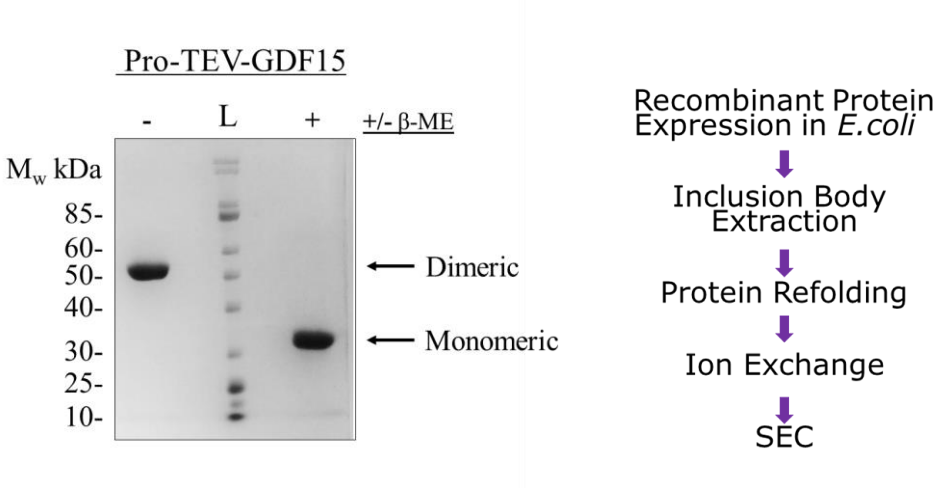


Figure 4.7. Purified pro-TEV-GDF15 and Purification. Sample of purified pro-TEV-GDF15 analysed by SDS-PAGE, in non-reducing and reducing conditions (+β-ME). Dimeric and monomeric forms indicated. Flow chart summarising method of protein purification with each stage highlighted.

4.3 Purification of the GDF15 prodomain

The GDF15 prodomain (amino acids 30-196) was found to be a soluble protein when expressed in *E. coli* BL21 (DE3) cells at 15°C overnight. The prodomain of GDF15 was N-terminal His-tagged protein enabling purification via Ni-affinity chromatography. Eluted material from Ni-affinity chromatography contained a band of 20 kDa as the main species – corresponding to the predicted molecular weight of the GDF15 prodomain. Ni-affinity-purified protein was then loaded onto an ion exchange column for further purification (see sample lane - figure 4.8 A). Cation exchange was successful in concentrating the protein, as expected. However, a smaller band can be seen under 20 kDa (see figure 4.8 A). The band is likely a degradation product of the GDF15 prodomain as it becomes more prominent during the purification. The degradation product co-purifies with the full-length prodomain and

remains present in the final sample (see figure 4.8 B). This implies that the GDF15 prodomain may be dimeric, preventing the separation of the degradation product by cation exchange. Techniques such as size exclusion chromatography - multi-angle light scattering (SEC-MALS) could demonstrate the protein to be dimeric in solution through the determination of the proteins molecular weight.

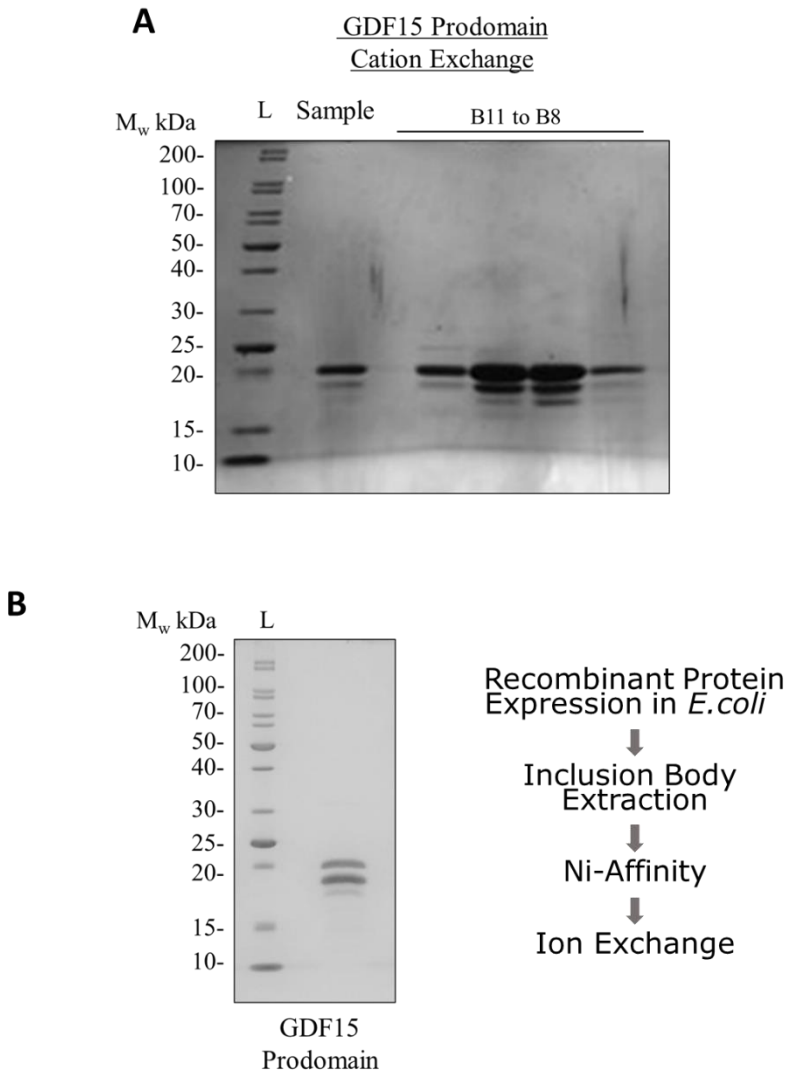


Figure 4.8. Purification of the GDF15 prodomain. A) Purification of the GDF15 prodomain by cation exchange chromatography. Sample is 200 mM Imidazole elution from initial Ni-affinity chromatography. B) Sample of purified GDF15 prodomain and summary of purification. All protein samples analysed by SDS-PAGE.

4.4 Pro-GDF15 Signalling

In this section we will examine the signalling activity of the various GDF15 constructs produced using a cell-based reporter gene assay. Several publications have characterised GDF15 signalling since the recent discovery of its co-receptor GFRAL[62–65]. Again, past studies have focussed mainly on mature domain of GDF15, with pro-GDF15 signalling remaining poorly understood. Using a similar approach as NRTN in chapter 3, here we aim to determine the impact of the prodomain on GDF15 signalling comparing mature and pro-GDF15. To examine cleavage and activation of pro-GDF15 a TEV-cleavable construct will be used (pro-TEV-GDF15). The engineered pro-TEV-GDF15 protein has the natural furin cleavage site replaced by a TEV protease site.

4.5 The Signalling Activity of Mature and pro-GDF15

4.51 Luciferase Based Reporter Gene Assay

To determine the impact of the prodomain on GDF15 signalling we used a cell-based reporter gene, developed by AstraZeneca. This assay uses modified HEK293S-SRF-RET/GFRAL cells responsive to GDF15 signalling. This stable cell-line has been modified to express human RET and GFRAL receptors as well as serum response factor (SRF). The HEK293S (human embryonic kidney 293 suspension) cell-line is adapted to growth in suspension. The bioactivity assay then works in the same was as described for the NRTN bioactivity assay (Chapter 3). GDF15 induced RET signalling activates, downstream signalling pathways leading to the expression of the reporter gene – firefly luciferase. Luciferase levels can be measured by its enzyme activity and used to determine the level of signalling activity (full details in materials and methods). Dose-response curves and EC_{50} values were calculated for the dimeric form of all GDF15 proteins (as described in Chapter 3).

4.52 Signalling Activity of mature and pro-GDF15

Using the luciferase-based reporter gene assay, both mature GDF15 and the control generate similar dose-response curves (see figure 4.9). The control used here is commercially available human recombinant mature GDF15, produced in mammalian cells (PeproTech). Both proteins are equivalent in terms of signalling activity with EC_{50} values of 0.39 nM for mature GDF15

and 0.60 nM for the control (see table 4.2). From this, we can conclude that the refolded and purified mature GDF15 from bacterial expression systems produced in this work is of equivalent quality to the mammalian derived protein.

The GDF15 pro-form remains active and is able to produce a signal, as seen from the dose response curve (see figure 4.9). However, its activity is reduced compared to the mature form of GDF15 the curve is shifted to the right and the EC_{50} is higher. As seen with NRTN in chapter 3, we see an approximately 10-fold difference in bioactivity with pro-GDF15 having reduced activity. Pro-GDF15 has an EC_{50} of 3.09 nM and the mature GDF15 0.39 nM (see table 4.2). This demonstrates the GDF15 prodomain to have inhibitory properties for GDF15 signalling.

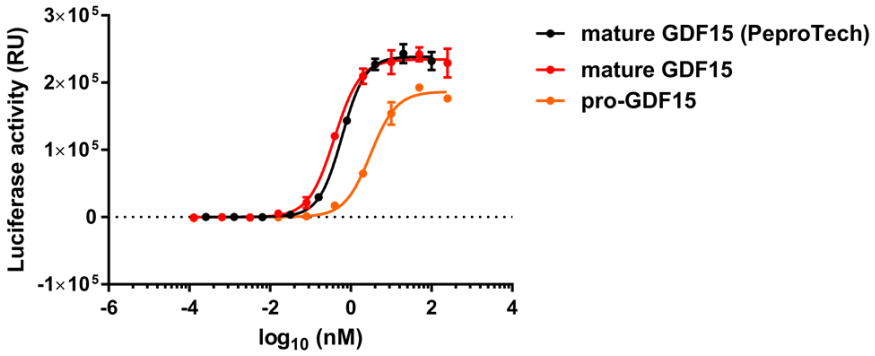


Figure 4.9. Bioactivity of mature and pro-GDF15 forms by luciferase reporter gene assay. A) Fitted dose response curves for GDF15 mature and pro-forms. Refolded and purified mature GDF15 shown in red and the control mature GDF15 (PeproTech – mammalian cell-line derived) shown in black. Pro-GDF15 shown in orange. Data shown are the means of triplicate measurements ± SD, n=2.

Sample	Figure	EC_{50} (nM)	EC_{50} (nM) 95 % CI*	Hill Slope	Hill slope 95% CI*	R^2
mature GDF15	6.1A	0.39	0.34 to 0.46	1.36	1.12 to 1.71	0.99
mature GDF15 (PeproTech)	6.1A	0.60	0.54 to 0.67	1.51	1.31 to 1.78	0.99
Pro-GDF15	6.1A	3.09	2.64 to 3.62	1.35	1.13 to 1.63	0.99

Table 4.2: GDF15 construct bioactivity values by luciferase reporter gene assay in HEK293S cells. Pro-GDF15, mature GDF15 and control mature GDF15 (PeproTech). All values presented at two decimal places, n=2. *CI Confidence interval.

4.6 The Cleavage and Signalling Activity of the pro-TEV-GDF15 Precursor

4.61 TEV Cleavage of pro-TEV-GDF15

To examine the impact of cleavage on pro-GDF15 signalling cleaved material must be produced. Using the same approach as previously described, we incubated the engineered construct pro-TEV-GDF15 with TEV protease enzyme in 2 M Urea, 500 mM Nacl, 50 mM TRIS pH 8.0. TEV protease was reduced before addition to protein samples. Again, this engineered construct has the natural furin site replaced by a TEV cleavage site (see figure 4.10 A). Unlike pro-TEV-NRTN cleavage, the cleavage of pro-TEV-GDF15 was largely incomplete even with relatively high concentrations of TEV enzyme. Post incubation, cleavage products were then analysed by SDS-PAGE. In non-reducing conditions, we can see bands corresponding to uncleaved protein at 50-60 kDa, the prodomain (at 20 kDa) and a prominent band at approximately 40 kDa (see figure 4.10 C). The band at 40 kDa corresponds to the molecular weight of a semi-cleaved form of pro-TEV-GDF15, with only one prodomain of the dimer cleaved (see figure 4.10 B). It is also possible that the 20 kDa band seen in non-reducing conditions is the mature GDF15 dimer, as this protein also runs at 20 kDa (as previously demonstrated). However, under reducing conditions this 20 kDa band remains present at equivalent intensity (see figure 4.10 C). This demonstrates the protein is not the disulfide-linked dimeric mature GDF15. Additionally, the reducing conditions were confirmed to be effective in reducing disulfides, as monomeric species of pro-TEV-GDF15 can be seen at 30 kDa (see figure 4.10 C). Finally, a band corresponding to the mature GDF15 domain at 10 kDa can be observed under reducing conditions. Monomeric mature GDF15 is a product of the reduced semi-cleaved form (see figure 4.10 B and C). Curiously, protein cleavage seems to be produced only a semi-cleaved form of pro-TEV-GDF15. Incomplete proteolysis of substrates is always possible. However, in this case the only cleavage product that can be detected is the semi-cleaved form. This suggests that cleavage changes the properties of the protein and causing it to be resistant to cleavage *in vitro* such as conformational changes or aggregation. Semi-cleaved forms of pro-GDF15 have also been reported in the literature[53,54].

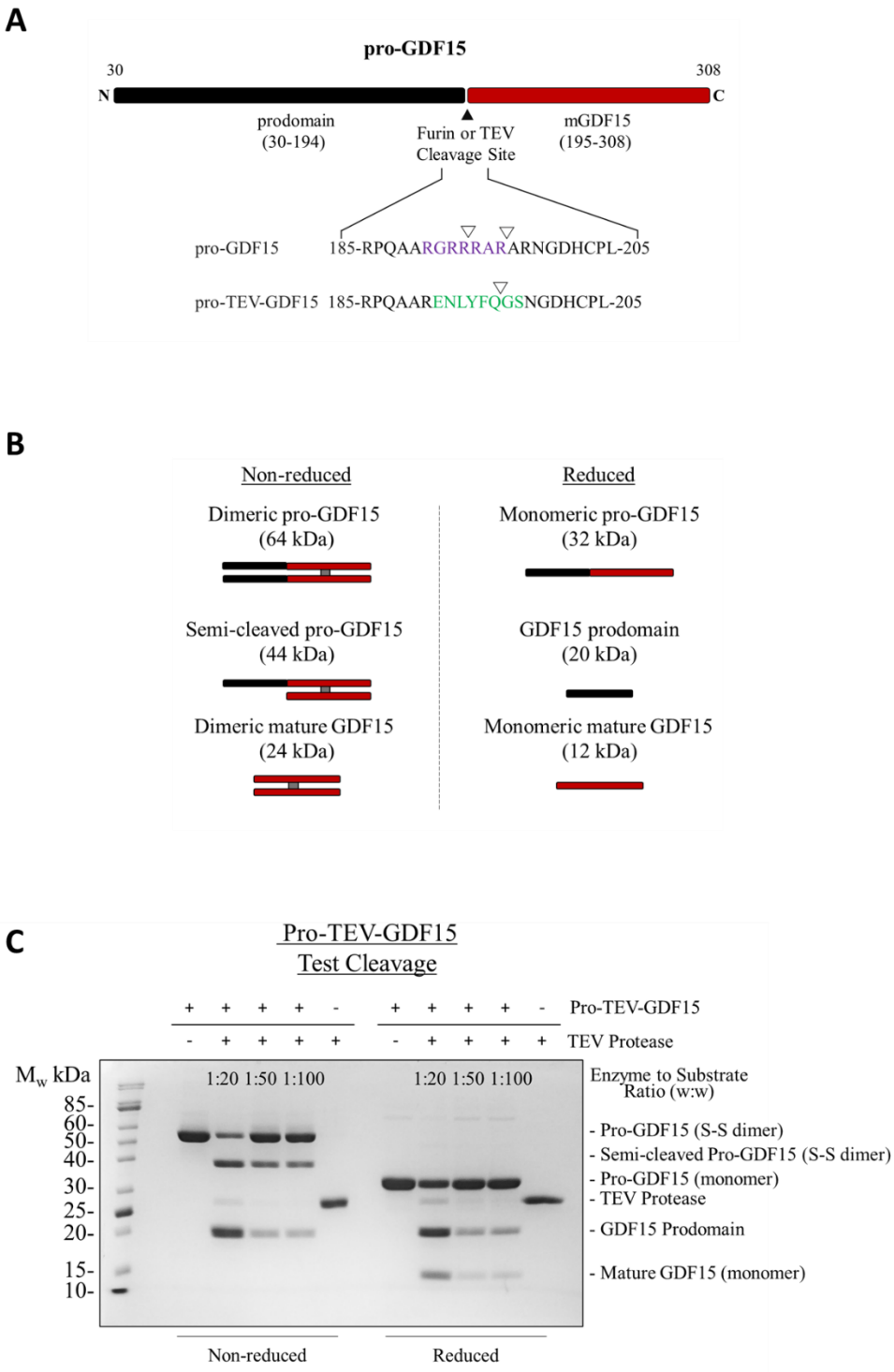


Figure 4.10. Cleavage of pro-TEV-GDF15 by TEV protease. A) Schematic of pro-TEV-GDF15 construct. GDF15 prodomain in black and mature domain in red. Natural furin site amino acids in purple and engineered TEV site in green. Cleavage sites indicated by white triangles. B) Schematic of theoretically possible pro-TEV-GDF15 TEV cleavage products under reducing and non-reducing conditions. Disulfide-linked dimeric proteins indicated with disulfides shown in

grey. C) Cleavage of pro-TEV-GDF15 by TEV protease and analysis of cleavage by SDS-PAGE, reducing and non-reducing conditions as indicated. Pro-TEV-GDF15 cleaved in 2 M Urea, 500 mM NaCl, 50 mM TRIS pH 8.0, with added reduced TEV protease. S-S (disulfide-linked).

4.62 The signalling activity of cleaved pro-TEV-GDF15

With cleaved protein produced, we can now examine the impact of cleavage on bioactivity using the same reporter gene assay. Dose response curves were obtained for pro-TEV-GDF15, cleaved and uncleaved. However, the maximal response could not be accurately determined as a maximal response plateau was not reached at concentrations tested and so EC_{50} values are partly based on extrapolation (see figure 4.11). Therefore, the calculated EC_{50} values for these constructs are used only as an indicator of activity. Comparison of shifting dose response curves has been used here to compare activity between proteins. Pro-TEV-GDF15 signal is similar to WT pro-GDF15, although with reduced bioactivity (see figure 4.11) with an approximately 2-fold difference in EC_{50} values (see table 4.2 and 4.3). The difference in activity between WT and TEV cleavable forms of pro-GDF15 may be due to furin presence within the bioassay. Semi-cleaved pro-TEV-GDF15 has increased activity compared to the uncleaved protein but is not as active as mature form, as might be expected from a partially cleaved species (see figure 4.11). The EC_{50} of semi-cleaved GDF15 is approximately 5-fold higher, at 2 nM than mature GDF15 0.39 nM (see table 4.2 and 4.3). From this, we can conclude that partial cleavage of pro-GD15 can restore partial activity.

As a negative control, TEV enzyme and/or sample buffer alone was found to induce no signalling activity (data not shown). A less invasive control to prevent furin cleavage could have been used - such as mutation of an Arginine residue to Alanine, in the furin site.

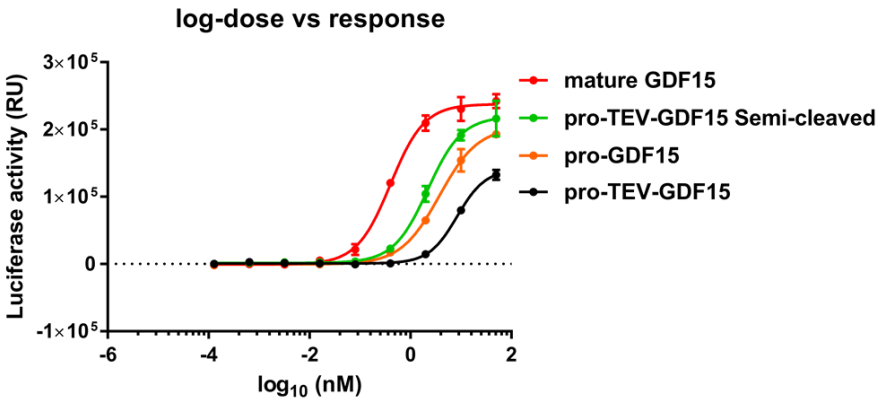


Figure 4.11. Cleavage and bioactivity of pro-TEV-GDF15 by luciferase reporter gene assay. Bioactivity of pro-TEV-GDF15 by luciferase reporter gene assay using in HEK293S cells. Fitted dose response curves for pro-TEV-GDF15 (black) and TEV cleaved pro-TEV-GDF15 (green). Data shown are the means of triplicate measurements ± SD, n=2. Dose response curves from figure 4.9 for mature GDF15 (red) and pro-GDF15 (orange) shown for reference.

Sample	Figure	EC_{50} (nM)	EC_{50} (nM) 95 % CI*	Hill Slope	Hill slope 95% CI*	R^2
Pro-TEV-GDF15	6.3 C	8.42	7.71 to 9.30	1.55	1.36 to 1.79	0.99
Pro-TEV-GDF15 Semi-cleaved	6.3 C	2.20	1.84 to 2.71	1.29	1.04 to 1.66	0.99

Table 4.3: pro-TEV-GDF15 construct bioactivity values by luciferase reporter gene assay in HEK293S cells – non-cleaved and cleaved with TEV protease. All values presented at two decimal places, n=2. *CI Confidence interval.

4.7 Furin Cleavage of pro-GDF15

In this section we will examine the cleavage of pro-GDF15 by furin protease using the same approach as pro-NRTN (see Chapter 3). As shown previously, pro-TEV-GDF15 cleavage results in a semi-cleaved form. Now we will determine if furin cleavage produces the same products as TEV protease *in vitro*. We will also determine the specificity of furin cleavage, using non-cleavable constructs (see figure 4.12).

As expected, pro-GDF15 can be cleaved by furin generating bands corresponding to the prodomain at 20 kDa and the semi-cleaved form at 40 kDa (see figure 4.12). The apparent bias towards this semi-cleaved form is evident by both furin and TEV cleavage as seen previously. It is possible that, post cleavage, the semi-cleaved form is more susceptible to aggregation or phase separation as shown previously for the prodomain in isolation. This could prevent further cleavage of the semi-cleaved form by furin enzyme, by preventing access of the enzyme to the secondary cleavage site. The pro-TEV-GDF15 remains uncleaved by furin, demonstrating the altered furin sequence to be only site for furin cleavage (see figure 4.12 D).

Interestingly, the semi-cleaved form of pro-GDF15 has also been reported to be secreted from mammalian cell-lines in the literature, along with uncleaved pro-GDF15 and fully cleaved mature GDF15[53,54]. For other TGF-β proteins semi-cleaved forms have also been observed. Furin processing of pro-myostatin *in vitro* also produces a partially cleaved form of the protein, due to incomplete proteolysis[132].

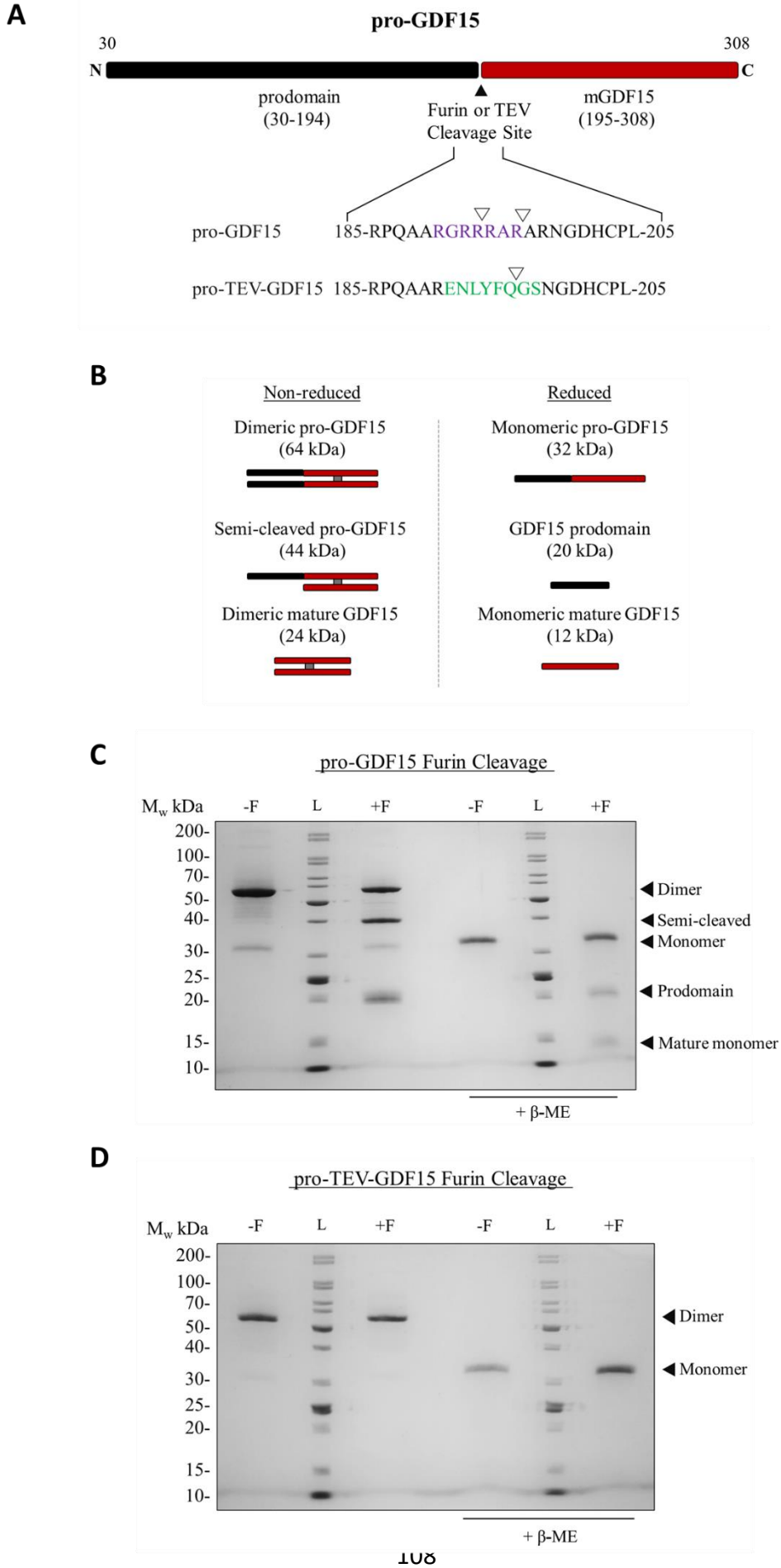

A
pro-GDF15
30
308
N
C
prodomain
(30-194)
Furin or TEV
Cleavage Site
mGDF15
(195-308)
pro-GDF15 185-RPQAARGRRRARARNGDHCPL-205
pro-TEV-GDF15 185-RPQAARENLYFQGSNGDHCPL-205
B
Non-reduced
Reduced
Dimeric pro-GDF15
(64 kDa)
Monomeric pro-GDF15
(32 kDa)
Semi-cleaved pro-GDF15
(44 kDa)
GDF15 prodomain
(20 kDa)
Dimeric mature GDF15
(24 kDa)
Monomeric mature GDF15
(12 kDa)
C
pro-GDF15 Furin Cleavage
Mw kDa
-F
L
+F
-F
L
+F
200-
100-
70-
50-
40-
30-
25-
20-
15-
10-
Dimer
Semi-cleaved
Monomer
Prodomain
Mature monomer
+ β-ME
D
pro-TEV-GDF15 Furin Cleavage
Mw kDa
-F
L
+F
-F
L
+F
200-
100-
70-
50-
40-
30-
25-
20-
15-
10-
Dimer
Monomer
+ β-ME

Figure 4.12. Furin cleavage of pro-GDF15. A) Schematic of pro-GDF15 constructs – pro-GDF15 with a natural furin site (purple) and pro-TEV-GDF15 with an artificial TEV site (green). Cleavage site positions indicated by white arrows. B) Schematic of theoretical GDF15 furin cleavage products under non-reducing and reducing conditions. Prodomain in black, mature domain in red with disulfide bonds shown in grey with the molecular weight of each product indicated below. C and D) Non-reducing and reducing SDS-PAGE analysis of pro-GDF15 or pro-TEV-GDF15 post furin incubation.

4.8 The GDF15 H202D Variant

Now we will examine the signalling activity and detection of the GDF15 H202D variant. As described in the introduction H202D is a common variant in humans, however it is uncertain if this variant changes the bioactivity of GDF15[86]. Here will determine the bioactivity of GDF15 H202D using the same GDF15 luciferase-based reporter gene assay as used previously. For purification of GDF15 H202D variant see figure 4.3. Detection and quantification of GDF15 is essential in the study of GDF15 biology and pathology. Additionally, we will determine the impact of the variant on GDF15 detection and quantification by ELISA. Finally, a heterodimeric protein of mature wild type and H202D GDF15 is produced to mimic GDF15 in heterozygous individuals. The heterodimeric protein will be used to determine the impact on quantification for GDF15 from heterozygous individuals. ELISA experiments in this section were performed by our collaborators as indicated.

H202D is the most common missense variant of GDF15 in humans, with an allele frequency of 23% - 1000 Genomes Project Phase 3, Ensemble (https://www.ensembl.org). Studies have implicated the variant in cancer and suggest potential biological differences between the variant and wild type GDF15, however this is disputed[86,96]. The H202D mutation is present in the N-terminal of the GDF15 mature domain, far from the co-receptor GFRAL interaction site (see figure 4.13 A). In RET containing structures no interactions can be observed between H202 and the RET receptor, however the RET CRD does bind the mature growth factor (see figure 4.13 B). From this, we can hypothesise that it is unlikely for the H202D variant to impact cell signalling. In this section we will test this hypothesis using previously purified mature GDF15 and H202D variant to determine bioactivity.

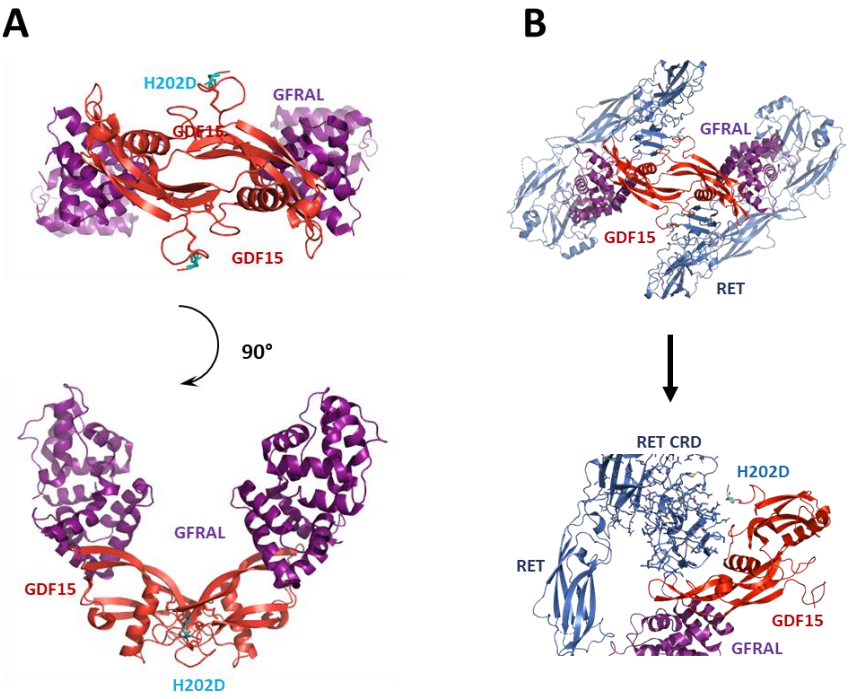


Figure 4.13. The GDF15 H202D variant. Crystal and Cryo-EM structures of the GDF15 extracellular signalling complex. Taken from Allan, *et al.* 2017 (PDB code: 5VZ4) and Li*, et al.* 2019 (PDB code: 6Q2J). RET and RET cysteine rich domain (CRD) shown in dark blue, GDF15 mature domain shown in red with H202D highlighted in cyan and GFRAL in purple.

4.9 Bioactivity of the GDF15 H202D variant

To assess the impact of the H202D variant on the activity of mature GDF15 we used the same GDF15 luciferase-bioactivity assay as previously described. Purified wild type and H202D variant homodimers of mature GDF15 both produce almost identical dose response curves in the bioassay, indicating no difference in bioactivity between the two (see figure 4.14). Furthermore, calculated EC_{50} values for the WT and H202D variant were 0.39 nM and 0.45 nM, again demonstrating no difference in bioactivity (see table 4.4).

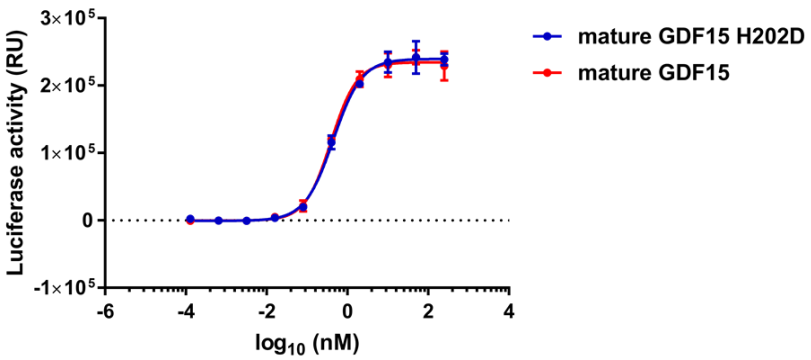


Figure 4.14. Bioactivity of the mature GDF15 WT and H202D variant by luciferase reporter gene assay. Fitted dose response curves for GDF15 constructs. WT mature GDF15 shown in red and the H202D variant shown in blue. Data shown are the means of triplicate measurements ± SD, n=2.

Sample	Figure	EC_{50} (nM)	EC_{50} (nM) 95 % CI*	Hill Slope	Hill slope 95% CI*	R^2
Mature GDF15 wild type	7.2	0.39	0.34 to 0.46	1.36	1.12 to 1.71	0.99
Mature GDF15 H202D	7.2	0.45	0.38 to 0.52	1.23	1.04 to 1.50	0.99

Table 4.4: WT and H202D variant mature GDF15 bioactivity values by luciferase reporter gene assay in HEK293S cells. All values presented at two decimal places. *CI Confidence interval and n=2.

4.8 Quantification of the GDF15 H202D variant

As mentioned previously, GDF15 has several biological roles and is implicated in several pathologies[86]. Given the prevalence and putative biological differences of the H202D variant investigations into its role in biology and pathology have been initiated by our collaborators - Professor Sir Stephen O'Rahilly research group, Department of Clinical Biochemistry, University of Cambridge. To measure GDF15 levels they used enzyme-linked immunosorbent assay (ELISA) based methods. During ELISA analysis of clinical samples, lower levels of GDF15 seemed to correlate with a H202D genotype. This led to the hypothesis that GDF15 H202D variant levels were underestimated using ELISA based detection methods. More specifically, antibodies used in certain widely used commercially available ELISA kits were thought to have a lower affinity for the variant. H202 is known to be on the surface of GDF15 as seen in several

structures (see figure 4.13) and so it could be included in an antibody binding site. This would explain the lower levels of GDF15 observed. If detection levels are altered by the presence of the H202D variant, a correction factor can be identified using purified proteins of known concentrations. Correction factors can then be and applied to samples based on known genotype. Corrected levels of GDF15 can then be examined in relation to relevant phenotypes or pathologies to determine correlation.

4.81 Production of the GDF15 heterodimer

For completeness we also sought to produce a heterodimer of WT and H202D GDF15 variant to mimic heterozygous individuals. We assume both alleles will be expressed equally. A combination of N-terminal Strep- and His-tags were used to purify heterodimeric species, as well as Strep- or His-tagged WT homodimers to be used as controls. Although, purification was successful the N-terminal tags used interfered with quantification by ELISA (data not shown). As the C-terminus of GDF15 is poorly accessible, C-terminal tags were not deemed to be realistic for use without compromising the integrity of the proteins[62–65].

Instead, an untagged heterozygous mixture of variants was produced by mixing denatured, solubilised WT and the H202D variant of mature GDF15 prior to refolding. The GDF15 heterozygous mix (HZ) was purified using the same method as previously described for mature GDF15 proteins. Assuming equal expression and likelihood of dimer formation this would produce 25% WT homodimer, 25% H202D homodimer and 50% heterodimer (see figure 4.15 A). Purification of mature GDF15 HZ was successful, yielding pure protein (see figure 4.15 B). Additionally, the predicted ratio of species and the presence of the heterodimer was confirmed by mass spectrometry (MS) analysis with shifts in Da consistent with the theoretical M_w of each variant (see figure 4.15 C).

The heterodimeric protein as part of the HZ mix was not subjected to the cell-signalling assay in this work. As bioactivity of both mature GDF15 WT and the H202D variant are equivalent (see figure 4.14, table 4.4) it is unlikely a heterodimer would vary in bioactivity as well, however this remains untested.

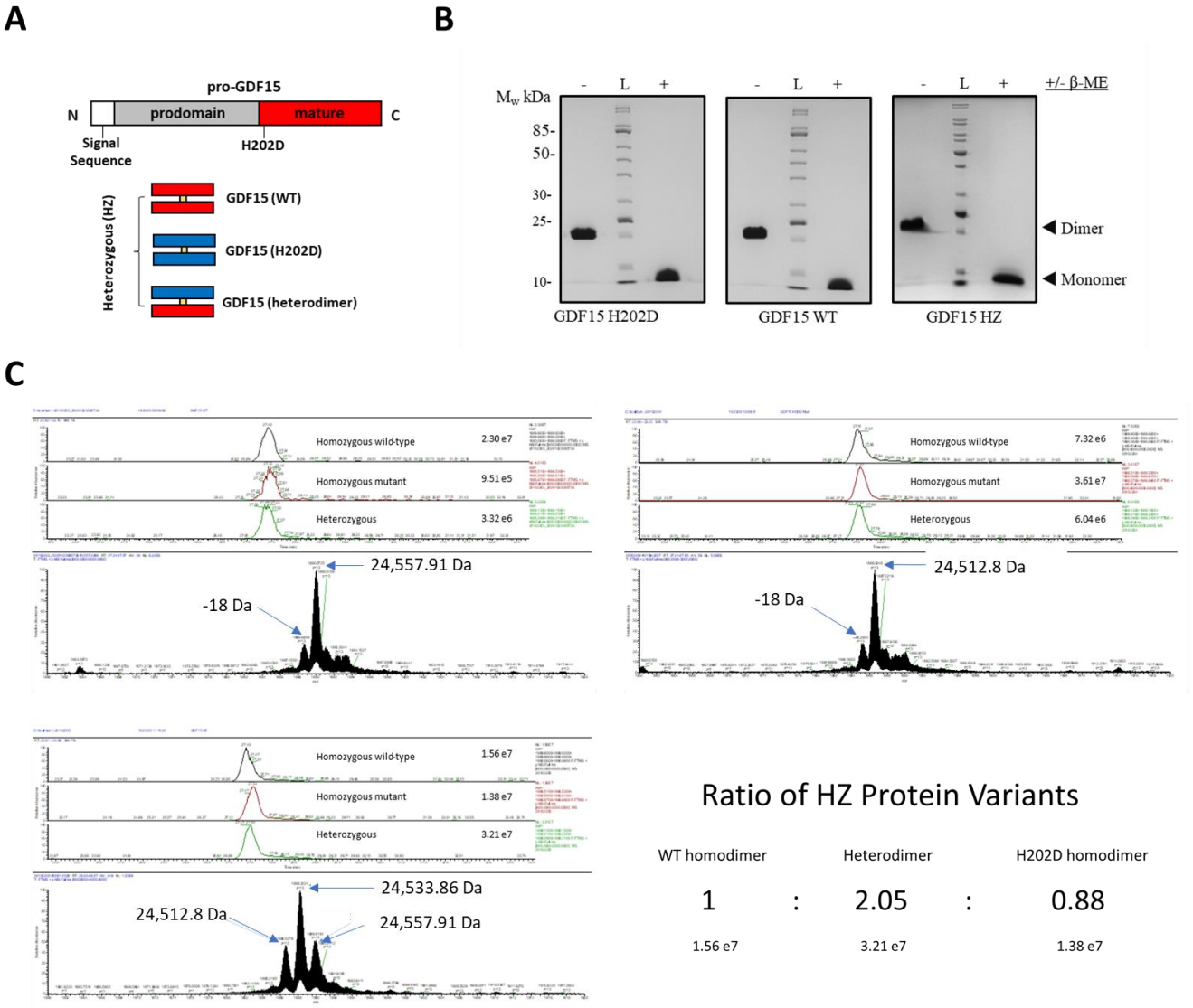


Figure 4.15. Purified mature GDF15 Heterozygous (HZ). A) schematic of GDF15 and heterozygous (HZ) protein heterogenous mix, pro-GDF15 prodomain in grey and mature domains in red for the wild type and blue for the H202D variant. Position of the H202D site indicated and disulfide-bridges between mature domains shown in yellow. Domains of Pro-GDF15 proportionate to size in amino acids. B) Sample of purified mature GDF15 HZ analysed by SDS-PAGE, in non-reducing and reducing conditions (+β-ME). Dimeric and monomeric forms indicated. Protein composed of GDF15 WT and H202D variant sequences. C) Intact Mass spectrometry (MS) analysis of purified mature GDF15 HZ and homodimeric constructs, performed by Dr Richard Kay. Weight of species and ratio from intact MS analysis indicated.

4.82 Quantification of GDF15 by ELISA

With purified mature GDF15 HZ, H202D and WT protein samples we could then proceed with ELISA testing. ELISA experiments were performed on purified proteins from this work by our collaborators in the Professor Sir Stephen O'Rahilly's research group in the School of Medicine. Initial data is presented here. ELISA measurements demonstrate the H202D variant concentration is underestimated. In both assays we observe ~50% recovery for the H202D variant homodimer relative to WT, with a calculated correction factor of ~2. GDF15 HZ has a ~60-65% recovery rate relative to WT, with a calculated correction factor of ~1.6 for both assays tested (see table 4.5). Correction factors, calculated here, may now be used to determine accurate GDF15 levels in clinical samples when the genotype of the individual is known.

The heterodimer, with one H202D protomer, behaves exactly like the H202D variant homodimer by ELISA. From intact MS we know the GDF15 HZ contains 50% heterodimer, 25% WT homodimer and 25% H202D homodimer. This means that three quarters of the dimers in the HZ sample contain at least one copy of the H202D variant. Using the recovery values for the H202D variant, for three quarters of the GDF15 HZ sample and the recovery values of the WT for the remaining quarter we can calculate a predicted % recovery of ~102-104 for both assays tested. These calculated values match the experimental values of the GDF15 HZ as shown in table 4.5. This confirms the heterodimer concentration is underestimated in the same way as the H202D homodimer.

GDF15 Variant ELISA

Sample	R&D % Recovery	MSD* % Recovery
GDF15 WT homodimer	166.40	167.40
GDF15 H202D	81.70	83.00
GDF15 HZ	101.30	109.40
Calculated Correction Factors		
H202D correction (WT/H202D)	2.04	2.02
HZ correction (WT/HZ)	1.64	1.53

Table 4.5. Recovery values from ELISA for mature GDF15 variants. Average recovery values for mature GDF15 samples as measured by ELISA, presented at two decimal places. Assays performed using commercially available kits from R&D systems (R&D Quantikine ELISA) and MSD (*using R&D DuoSet reagents.) Mature forms of GDF15 tested are wild type (WT), the H202D variant and HZ (heterozygous). GDF15 HZ contains 25% WT homodimer, 25% H202D homodimer and 50% WT/H202D heterodimer as confirmed by MS. Samples tested in a dilution series from 74 to 2000 pg/ml. Calculated correction factors for H202D and HZ samples indicated. ELISA performed by collaborators in the Professor Sir Stephen O'Rahilly Research Group.

4.9 The GDF15 Prodomain: Structure and Interactions

In this section we will examine the structure and interactions of the GDF15 prodomain. The structure of the mature domain of GDF15 is known, however the prodomain structure remains unknown[62–65]. The GDF15 prodomain will now be analysed to determine secondary structure using CD and the interaction between the prodomain and mature domain will be measured by ITC. Secondly, pro-GDF15 is enriched at the extracellular matrix (ECM), unlike mature GDF15[53,54]. This raises the possibility of the prodomain interacting with ECM components such as saccharides, as known for other TGF-β members[57]. Putative interactions with saccharides will be assessed using EMSA based methods.

4.10 GDF15 Prodomain-Saccharide Interactions

Now we will investigate the GDF15 prodomain interaction with saccharides, as we suspect it will interact with sugars in the ECM, due to enrichment of pro-GDF15 at the ECM as mentioned previously[53,54]. More specifically, we suspect an interaction with heparan sulfate (HS), a component of the ECM and highly negatively charged. As mentioned in the purification, the prodomain has many positively charged residues – this could facilitate an interaction. Furthermore, HS binding is common in the wider TGF-β family[57].

For preliminary experiments, we chose to use an electrophoretic mobility shift assay (EMSA). Using this assay visible material, in this case a fluorescent oligosaccharide, is run on a non-denaturing/native gel of agarose or acrylamide. Then to test whether the protein of interest binds to the oligosaccharide it can be incubated and ran on the same gel. If bound, the sample becomes larger and so will travel more slowly though the gel- resulting in a gel shift[133]. To

visualise gel shifts we used DP10-AMAC, a fluorescent tagged heparan sulfate precursor with ten sugar moieties (DP: degree of polymerisation) linked to 2-aminoacridone dye (AMAC). From EMSA analysis we can see a gel shift under the presence of the prodomain (see figure 4.16). The shifted band is diffuse, this may be because of the prominent degradation product present in purified GDF15 prodomain samples bind differently to the probe. Future work could confirm the prodomain-saccharide interaction using alternate techniques – such as ITC and bio-layer interferometry (BLI) using immobilised sugars. Additionally, crystal screening was performed using the same construct with and without sugar additives. Crystal hits were identified but unfortunately, they did not diffract (see figure 4.16 C).

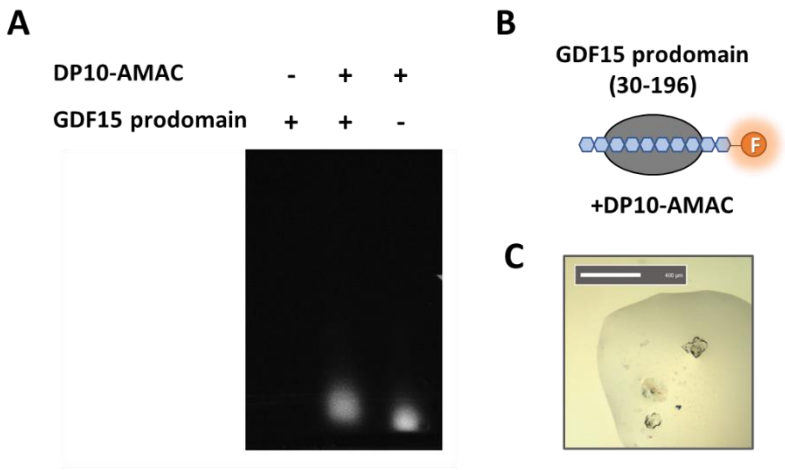


Figure 4.16. GDF15 prodomain HS interaction via EMSA. A) Electrophoretic mobility shift assay (EMSA) performed on the GDF15 prodomain and the fluorescently tagged heparan sulfate precursor DP10-AMAC. B) Schematic of GDF15 prodomain and DP10-AMAC, with the prodomain in grey, sugars in blue and fluorophore of DP10-AMAC in orange. C) Crystallisation screening example hit, with the GDF15 prodomain. Protein at 8 mg/ml in 25 mM CHES pH 9.0, 0.5 M NaCl mixed with screening condition: 0.2 M Lithium sulfate, 0.1 M Sodium acetate 4.5, 50 % w/v PEG 400 at a 1:1 ratio.

4.11 Structure of the GDF15 prodomain

The GDF15 prodomain shows predominantly unstructured spectra by CD (see figure 4.17). We then sought to determine if any secondary structural changes could be observed in the presence of heparin, a putative binding partner of the prodomain as seen by EMSA (see figure

4.16). As certain proteins are known to adopt a structure when bound to specific binding partners. However, other proteins retain their intrinsically disordered structures upon interaction and often described as 'fuzzy complexes' (reviewed by Sharma *et al*., 2015 and Kasahara *et al*., 2019[134,135]). No such change to secondary structure could be seen by CD with the spectra with and without LMWH being largely similar (see figure 4.17). Some changes are seen at lower wavelengths, this may be due to aggregation or phase separation (see figure 4.17) and is discussed in the following section.

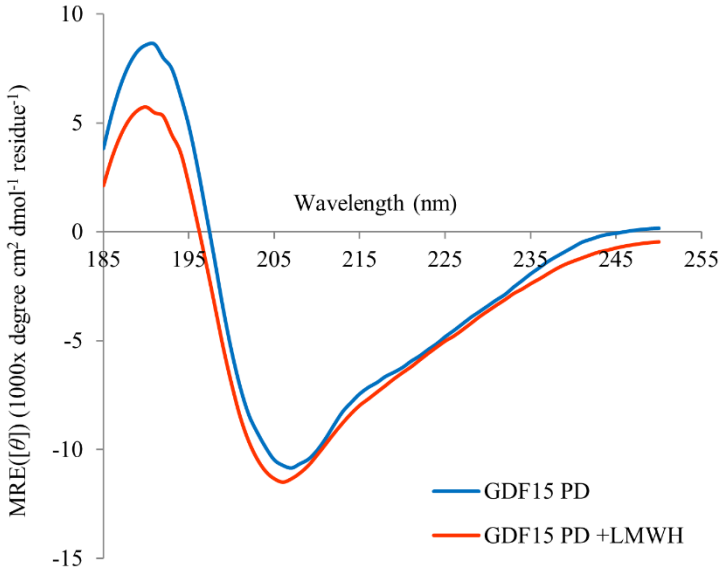


Figure 4.17. Circular dichroism analysis of the GDF15 prodomain and LMWH. A) Far UV circular dichroism (CD) spectra of the GDF15 prodomain (residues 30-196) shown in blue, GDF15 prodomain +0.2 mg/ml low-molecular-weight heparin (LMWH) shown in orange. CD data recorded between 250 and 185 nm in 1 nm increments, in mean residue ellipticity, MRE ([θ]) (1000× degree cm^2 dmol^{-1} residue^{-1}).

4.12 GDF15 Prodomain Phase Separation

During testing of LMWH addition the GDF15 prodomain was observed to become turbid. This turbidity could be due to precipitation (formation of colloidal aggregates) or potentially liquid-liquid phase separation. A similar phenomenon has been observed for histone tails and DNA binding, creating two phases from a single solution. The histone tails are unstructured and highly positively charged (like the GDF15 prodomain) and DNA is an extended and negatively

charged polymer (like HS)[136]. In this section we investigate phase separation by a turbidity assay. In this assay absorbance (or rather, scattering) at 340 nm is measured as the ligand is titrated in – in this case we will use LMWH. Larger particles and condensates will have an increased absorbance at 340 nm, allowing us to measure the appearance of any aggregates or phase separation. This method has also been used to study phase separation of histone tails, which undergo a reversable phase separation with increasing concentration of DNA[136].

From the turbidity assay, we can see the GDF15 prodomain undergoes a similar pattern to the phase separation noted for the histone tails. Upon initial addition of LMWH to GDF15 prodomain the light scattering increases, however this is reversible with increasing concentrations of LMWH (see figure 4.18). We also observe a difference in the behaviour of the prodomain depending on the ionic strength of the solution, as expected for an ionic interaction (see figure 4.18). Finally, phase separation may explain LMWH binding but no change in structure as seen by CD – histone tails, described previously, behave in a similar way with DNA and can retain their disordered structure upon binding[136]. Further work to confirm phase separation could include microscopy to identify phase-separated droplets upon LMWH addition to the prodomain.

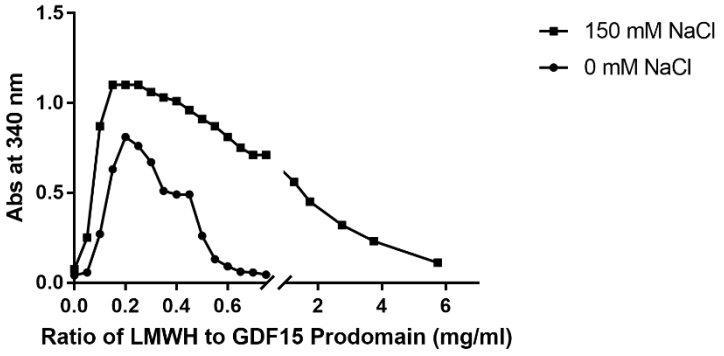


Figure 4.18. GDF15 prodomain - LMWH turbidity assay. Turbidity assay performed on the GDF15 prodomain with a titration of low molecular weight heparin (LMWH). Measurement of absorbance at 340 nm is plotted against increasing concentrations of LMWH in either 0 or 150 mM NaCl as indicated.

4.13 Prodomain-mature domain interactions

Using ITC binding between the prodomain and mature GDF15 was not detected, and no significant heat changes are observed when injecting prodomain into mature domain (see figure 4.19). This matches what was seen with NRTN and its prodomain in chapter 3. Unfortunately, suitable buffer conditions for HDX-MS experiments using pro-GDF15 could not be identified and so this was not performed. Future work could confirm the absence of an interaction between GDF15 domains using alternative techniques – such as bio-layer interferometry (BLI). As mentioned with NRTN in Chapter 3, aggregation may also prevent a genuine interaction as measured by ITC.

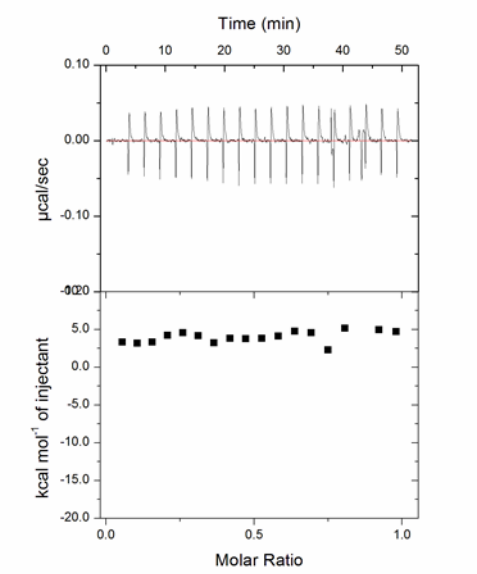


Figure 4.19. Prodomain and mature GDF15 interaction via ITC. Thermogram and isotherm traces generated by isothermal titration calorimetry (ITC) for the titration of the GDF15 prodomain into the mature GDF15.

Conclusions

This chapter describes the production and characterisation of GDF15 variants and pro-forms, through recombinant protein expression and purification, cell-signalling assays, enzyme cleavage and biophysical analysis.

Firstly, we describe the production of human GDF15 constructs; the GDF15 prodomain, mature GDF15, pro-GDF15 and a TEV-cleavable form - pro-TEV-GDF15. All mature domain containing proteins were successfully produced using bacterial expression systems, refolded, and purified to homogeneity and confirmed to be disulfide-linked dimers. Purified proteins were deemed to be relatively pure being free from major contaminants and degradation products, as analysed by SDS-PAGE. In contrast, the GDF15 prodomain was expressed as a soluble protein and purified without refolding. The GDF15 prodomain co-purifies with a degradation product and remains in the final sample. However, the full-length protein is still present. Purified GDF15 samples may now be used to examine the bioactivity, interactions, and structure of the protein.

Secondly, we examined we examined the bioactivity of mature and pro-forms of GDF15 Additionally, we determined the impact of cleavage on cell signalling using a TEV cleavable construct. Bioactivity was measured using a cell-based luciferase reporter gene assay. Refolded and purified mature GDF15 was found to be equivalent to control mature GDF15 (PeproTech), demonstrating the protein to be correctly folded and bioactive with comparable EC_{50} values to the control. The prodomain was found to be inhibitory to cell signalling, as mature GDF15 is approximately 10-fold more active than pro-GDF15.

Addition of the TEV cleavage site to pro-GDF15 was determined to partially reduce bioactivity but enabled the production of cleaved protein for testing. Reduction in bioactivity may be due to furin enzyme present in the bioassay. TEV cleavage of pro-TEV-GDF15 results in a semi-cleaved form, with only one of the two prodomains cleaved from the dimer. The semi-cleaved pro-GDF15 has an increased signalling activity compared to non-cleaved protein. However, activity is still less than mature GDF15 and so once cleaved the GDF15 prodomain has a reduced impact on GDF15 signalling.

Next, we examined the furin cleavage of pro-GDF15. As expected, furin can cleave pro-GDF15 *in vitro*. Furin produces the same cleavage products as seen with TEV cleavage, as described previously, resulting in a semi-cleaved form with one prodomain remaining uncleaved. We also demonstrate furin cleavage can be abolished by mutation of the furin site, preventing the cleavage between the pro and mature domains for pro-TEV-GDF15. Unlike, pro-NRTN no secondary furin site could be identified.

This chapter then demonstrates the H202D GDF15 variant has no impact on RET signalling and bioactivity, as determined by reporter gene assay. Calculated EC_{50} values for the GDF15 WT and H202D variant are equivalent at ~0.4 nM. We also examined the quantification of GDF15 variants by ELISA. To facilitate the analysis a heterozygous mix (GDF15 HZ), containing a heterodimeric species (WT/H202D) was produced to high purity by refolding. GDF15 HZ, WT and H202D mature proteins were then analysed by ELISA using widely used and commercially available kits. ELISA measurements demonstrate GDF15 H202D concentration is underestimated by 50%, and GDF15 HZ sample ~40% as hypothesised. From these results correction factors can be used to determine accurate GDF15 levels in clinical samples when the genotype of the individual is known.

Lastly, we examined the secondary structure and interactions of the GDF15 prodomain. Binding experiments using ITC display no interaction between the prodomain and mature GDF15. However, initial EMSA analysis suggests an interaction between the prodomain and the fluorescently tagged heparan sulfate precursor - DP10-AMAC. Analysis by CD analysis reveals the unstructured nature of the GDF15 prodomain. Furthermore, the structure of the prodomain is not altered by addition of LMWH. Furthermore, using a turbidity assay, the GDF15 prodomain appears to undergo phase separation upon the addition of LMWH. An unstructured prodomain with no interaction with the mature domain was also observed with NRTN constructs, suggesting a common theme for the GDNF family.

Chapter 5 - Discussion

In this PhD project I have studied the previously uncharacterised pro-domains of NRTN and GDF15. This has been achieved by development of efficient methods for the production of pure and homogenous pro-NRTN and pro-GDF15 and their isolated pro-domains. I also produced engineered constructs for both growth factors, with the native furin site replaced with a TEV cleavage site to enable controlled cleavage *in vitro* to produce cleaved pro-forms to study the effect of proteolytic processing of the precursor growth factors. With these purified proteins I then could characterise the role of the prodomains in cell signalling, their structure and interactions using cell-based bioactivity assays and biophysical methods for both GDNF family members. Characterisation of the GDF15 and NRTN prodomains improves our understanding of the role in cell signalling and highlights the diversity of structure and function for prodomains in the in the TGF-β superfamily.

5.1 Model of pro-NRTN and pro-GDF15 Inhibition

From the bioactivity assays for NRTN and GDF15, a moderate level of inhibition in bioactivity is seen for the pro-forms. I also determined that the pro-forms remain active and can induce signalling (Chapters 3 and 4). Furthermore, we know the prodomains of both growth factors to be largely unstructured by CD analysis (chapters 3 and 4). An unstructured N-terminal prodomain attached to mature NRTN or GDF15 may reduce the formation of an active RET signalling complex by steric hinderance. This would explain a reduction in bioactivity seen for both pro-NRTN and pro-GDF15 (see figure 5.1).

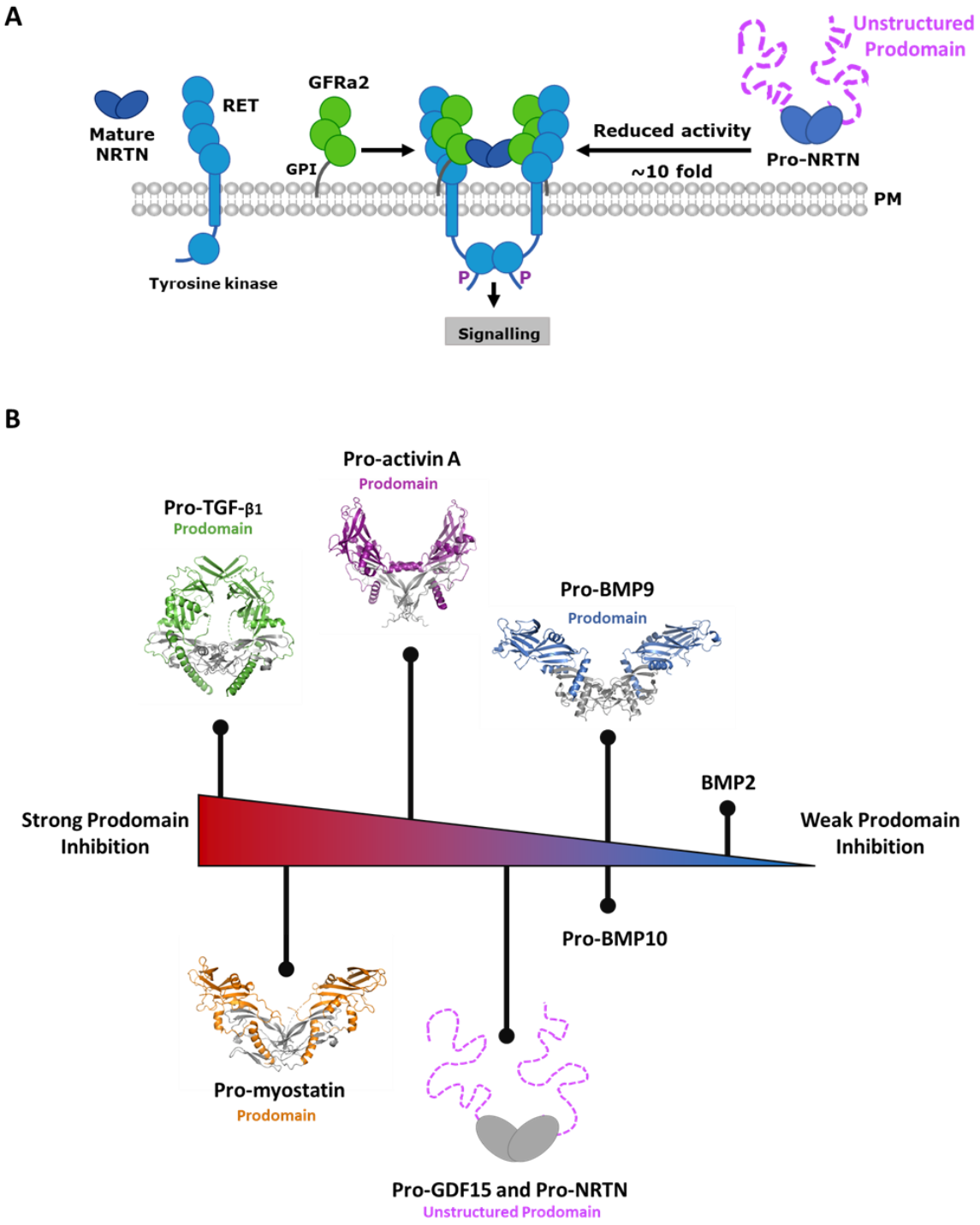


Figure 5.1. Model of pro-NRTN/GDF15 signalling and comparison with the TGF-β superfamily. A) Pro-NRTN, consists of an unstructured prodomain in pink and mature domain in dark blue. Formation of the RET signalling complex can occur though mature or pro-NRTN binding, although pro-NRTN displays approximately a 10-fold reduction in bioactivity. RET receptor shown in light blue, and co-receptor GFRα2 shown in green, phosphorylation of intracellular tyrosine kinas domains represented by P. B) Representative schematic of pro-NRTN/GDF15 signalling and prodomain inhibition in comparison with the wider TGF-β superfamily (for structure references see Introduction figure 1.4).

From cryo-EM structures of the RET extracellular signalling complex we see can that the N-terminus of the mature growth factors, GDF15 and NRTN, are not involved in receptor binding interfaces[73,77]. Additionally, space is available in the formed signalling complex that could accommodate an N-terminal prodomain for both GDF15 and NRTN (See figure 5.2). This would explain how pro-GDF15 and pro-NRTN can remain active. Similarly, the crystal structure of the pro-BMP9-ALK1 complex demonstrates the retention of the prodomain even with receptor ALK1 binding (Introduction, see figure 1.4)[47].

Another possibility is that alternative pro-form interactions compete with receptors for mature domain binding, resulting in a reduction in signalling activity. In the case of GDF15, this may be ECM components such as HS (as discussed in chapter 4). Regulation of cell signalling by ECM interactions is well known for TGF-β family members such as TGF-β1 and HS binding has also been proposed to influence NRTN signalling[16,58].

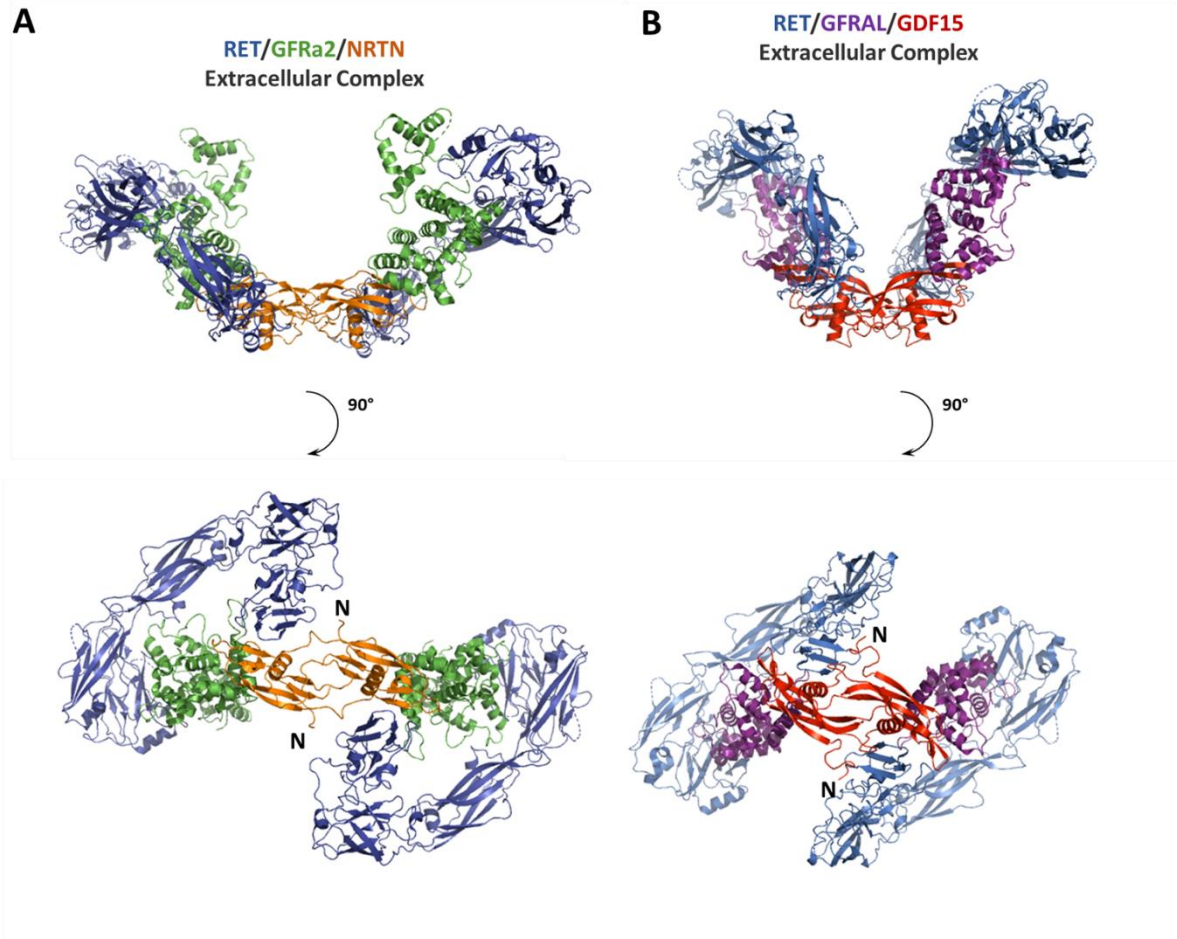


Figure 5.2 Cryo-EM Structures of NRTN and GDF15 Extracellular Signalling Complexes. A) The structure of the NRTN/GFRα2/RET receptor signalling complex, with the RET receptor shown in blue, the co-receptor GFRα2 shown in green and mature NRTN ligand shown in orange. B) The structure of the GDF15/GFRAL/RET receptor signalling complex, with the RET receptor shown in blue, the co-receptor GFRAL shown in purple and mature GDF15 ligand shown in red. Free N-termini of GDF15 and NRTN indicated (PDB codes: 6Q2J, 6Q2O)[73].

Steric hindrance by the unstructured prodomains may not be sequence specific. To explore this, the prodomains of each growth factor could be replaced by an unstructured and similarly charged peptide. If inhibition levels are equivalent to WT proteins, this would provide evidence to support the hypothesis that inhibition is driven through steric hindrance, reducing the efficiency of receptor complex formation and activation. To further characterise inhibition of the prodomains for both growth factors a truncation series of the pro-forms could be produced to identify the minimum inhibitory length of the prodomain. Free prodomain could also be added to cell-based assays for both GDF15 and NRTN to examine inhibition (if any) and whether or not the prodomains alone have any RET signalling activity.

Additionally, it cannot be ruled out cell-signalling assay activity could also be due to bacterial contaminants, such as lipopolysaccharides (LPS). For GDF15 assays the positive control was mammalian derived mature GDF15 and so would be less likely to contain such contaminants like LPS. For NRTN assays the positive control was mature NRTN that was produced from bacteria by AstraZeneca. However, this NRTN sample was also produced for *in vivo* animal studies and so was required to contain low LPS levels – this can be achieved via endotoxin removal procedures. In both cases the positive control has equivalent to the activity of mature domains produced in this work. Therefore, variable LPS levels, due to different expression and purification methods, are unlikely to have significantly influenced the cell-signalling assays used here. Further work could test GDF15 in the NRTN bioassay and vice versa to confirm no signalling activity. Furthermore, as RET signalling activates the MAPK cascade downstream signalling activation could also be confirmed such as detection of phosphorylated ERK and AKT via western blotting. This has been used previously to study GDNF proteins e.g., GDF15[64].

5.2 NRTN, GDF15 and the TGF-β Superfamily

Prodomain signalling inhibition in the TGF-β superfamily exists on a spectrum (see figure 5.1 B). From strong inhibitors to weaker ones. For example, pro-activin A and pro-myostatin

signalling is completely inhibited by the uncleaved prodomain[32,33] whereas, pro-BMP9 and pro-BMP10 display no significant inhibition[47]. Mature and pro-forms of BMP9 and BMP10 show no differences in potency via cell bioactivity assays and have a similar affinity for their ALK1 receptor by surface plasmon resonance (SPR). Furthermore, the prodomain of BMP9 can remain bound post Alk1 binding[47]. NRTN and GDF15 sit between these examples, displaying mild inhibition from their prodomains (chapters 3 and 4).

As with other family members, cleavage restores activity. Post cleavage, using the TEV cleavable constructs, we can see full activity is restored to NRTN and is equivalent to mature NRTN (Chapter 3). Similarly, activin A displays a similar profile following cleavage *in vitro*[33]. Other family members, such as myostatin only display a partial increase in bioactivity following cleavage[32,45]. Partial activity is also restored post-cleavage for GDF15; however, cleavage is not complete, and a semi-cleaved form is produced – this explains why full activity is not restored (chapter 4).

NRTN, unlike GDF15 has an additional furin site located within the prodomain (Chapters 3 and 4). Cleavage by furin within the NRTN prodomain may impact activity, for example reduce the moderate inhibition of NRTN signalling by the prodomain but this remains untested It is possible that furin may cleave pro-NRTN and pro-TEV-NRTN in the cell-based bioassays at the NRTN prodomain additional furin site impacting the activity observed. It is also hard to determine if this is genuine site, as cleavage does not seem to occur unless the furin site between mature and prodomains of NRTN is mutated (Chapter 3). An additional furin site would reduce the specificity of furin cleavage for NRTN. TGF-β prodomains typically remain associated to mature domains post cleavage and this is often referred to as the pro-complex[26] - as is the case for activin A, myostatin and others[32,33]. In contrast, GDF15 and NRTN display no evidence of interaction between their pro and mature domains, even in the uncleaved form, from ITC and HDX-MS analysis. In addition, the disordered nature of the NRTN and GDF15 prodomains are also atypical for the TGF-β superfamily[26] (see Chapters 3 and 4). However, not all TGF-β proteins form stable pro-complexes, the BMP2 prodomain is known to dissociate following furin cleavage[41,48].

5.3 Beyond the TGF-β Superfamily

GDNF family members, pro-GDF15 and pro-NRTN, both display reduced bioactivity, an unstructured prodomain and no detectable interaction between mature and prodomains. This is unlike typical TGF-β superfamily members, with structured prodomains that remain bound to the mature domains post-cleavage[8]. The similar properties of the prodomains for both GDF15 and NRTN, identified in this work, suggest a common theme for the GDNF family and their atypically short prodomains.

Beyond the TGF-β superfamily similarities between GDF15, NRTN and other growth factors can be seen. Growth factors with similar properties include nerve growth factor (NGF) and oncostatin M (OSM). Human OSM has a short and basic C-terminal propeptide of 31 amino acids that can be cleaved from the mature domain[137]. Originally identified as an inhibitor of tumour growth, this growth factor is now known to be an inflammatory cytokine[138]. Pro-OSM was also found to be able to bind its receptors and induce signalling, but with reduced activity. A similar change in bioactivity was observed between OSM and pro-OSM as seen with mature and pro-forms of the GDNF family members examined here – approximately 5 to 10-fold in growth inhibition assays[137].

As well as NRTN and GDF15, other growth factors are also known to have unstructured prodomains, such as NGF. NGF is neurotrophic factor involved in development and maintenance of the nervous system and has a basic N-terminal propeptide of around 100 amino acids. The propeptide structure of NGF has been extensively studied revealing it to be intrinsically unstructured/disordered by nuclear magnetic resonance (NMR). NMR experiments also revealed transient interactions between the pro and mature domains. Additionally, the propeptide has been shown to alter NGF effects as pro-NGF, and is able to signal independently as well[113].

Alternative receptor binding could be potential roles for the prodomains and/or pro-forms for GDF15 and NRTN, this could lead to a biologically distinct outcome. This is the case for both NGF and BDNF with mature proteins signalling via TrkA and B tyrosine kinase receptors, whereas pro-forms promote apoptosis through p75 and sortilin[114,115].

5.4 The GDF15 prodomain, HS and Phase Separation

Approximately one third of TGF-β superfamily members bind heparin or HS oligosaccharides - including TGF-β1, TGF-β2, certain BMPs, GDFs and GDNFs [57]. Binding experiments between the GDF15 prodomain and HS also demonstrate an interaction via EMSA (see Chapter 4). This interaction would also explain the enrichment of prodomain containing forms of GDF15, pro-GDF15 and semi-cleaved pro-GDF15, at the HS rich ECM[53,54]. Localisation of GDF15, via the prodomain would allow storage and release of the growth factor under stressful conditions (reviewed in [86]). Localisation of GDF15 at the secreting cells surface may also be a suitable way to regulate GDF15. As GFRAL, the GDF15 co-receptor, is only expressed in specific regions of the brain stem localisation/storage away from this region would prove an effective regulatory mechanism without the need for a fully inhibitory prodomain [63–65], like TGF-β1 and myostatin[16,32,45]

HS was also shown to induce phase separation for the GDF15 prodomain using a turbidity assay (see Chapter 4) however, this is not conclusive. Microscopy and NMR experiments could validate phase separation. This would be a novel finding and phase separation for HS binding proteins and has not been previously characterised. Analogous examples can be found in histone biology. Experiments into histone-tail interactions with DNA, reveal the disordered histone-tail to undergo phase separation upon DNA binding, this has been further characterised by microscopy work and NMR[136]. Further work could include characterisation of the GDF15 prodomain via NMR, as well as the GDF15-HS interaction. Additionally, CD could be performed on the pro-GDF15 protein to see if the presence of the mature domain induces prodomain folding.

Phase separation may also explain the semi-cleaved products seen upon GDF15 proteolysis by TEV and furin enzymes in chapter 4. Conformational changes post cleavage could lead to a similar phase separation as seen with the prodomain alone, this would exclude access for further cleavage. Alternatively, cleavage at one site may induce conformational changes that occlude the second cleavage site. Further work is needed to characterise phase separation and examine the pro-GDF15 construct to see if it displays the same properties. Semi-cleaved pro-GDF15 has been identified it the ECM[53,54], it is unclear if it is a functionally important state

– or just partially alters activity and localisation in the same way as pro-GDF15 uncleaved. Crystallization of the semi-cleaved GDF15 protein was not attempted.

As mentioned in Chapter 4 the GDF15 prodomain contains a putative N-linked glycosylation site (UniProt - Q99988). This contrasts with the NRTN prodomain that contains no predicted glycosylation sites (UniProt - Q99748). As the GDF15 prodomain was produced via a bacterial expression system this glycosylation would be absent. Glycosylation sites are known to improve the solubility and stability of many proteins including numerous protein based pharmaceuticals[139]. If applied to the GDF15 prodomain, increased solubility from glycosylation may reduce the aggregation/phase separation seen in this work (Chapter 4). Additionally, the formation of the semi-cleaved pro-GDF15 may also be due to aggregation – preventing access for a protease e.g., furin or TEV protease. Glycosylation may alter this result in the same way as well.

5.5 The GDF15 H202D Variant

The H202D variant has no impact on bioactivity for GDF15 signalling as speculated previously[86] (see chapter 4). This contradicts previous research suggesting the variant had differential biological effects in cancer models, however this may be due to poor experimental design and is disputed[86,96]. In contrast, the variant does impact on the quantification of GDF15 by ELISA, using commonly used commercial kits (See chapter 4). The variant likely disrupts part of an antibody binding site, explaining a reduction in detection by ELISA. This would also explain why N-terminal affinity-tags also reduced detection via ELISA. From variant recovery rates, by ELISA, we can derive a correction factor to determine more accurate levels of GDF15 when the genotype is known. This is of particular relevance and in clinical studies when investigating the links between GDF15 and its numerous associated pathologies, such as cachexia (reviewed in Lockhart *et al.,* 2020[86]).

The correction factor for the GDF15 H202D variant will be of use for previously collected data sets. Our collaborators have already collected a significant data set from clinical samples via ELISA with the problematic antibody. Such clinical samples are limited and so replication of experiments with better antibodies is less feasible. Additionally, as the genotype for each sample is already known an appropriate correction factor can be applied retrospectively with

ease. An alternate antibody, without such variant based detection issues, would be preferable for future studies. This would eliminate the need to apply such corrections factors going forwards.

5.6 Conclusions

This thesis highlights the diversity of prodomain structure and function in the TGF-β superfamily and suggests a common theme of unstructured and weakly inhibitory prodomains for the GDNF subfamily. The role of the GDF15 prodomain in localisation and HS binding gives the prodomain a potential biological function, this could be further characterised and investigated. However, the biological role of the NRTN prodomain remains largely unknown. Initially we anticipated a more significant role for the prodomains, as seen with other more "typical" TGF-β superfamily members, instead of the relatively minor effects on signalling demonstrated in the work[8,56]. The presence of the prodomain could be explained by a distinct activity, independent of the mature domain, as seen with NGF[113]. Other functions, such as influencing protein folding and secretion *in vivo*, are also possible - as well as no distinct function. Further work is needed to understand the structure and function of the prodomains for the GDNF subfamily.

References

1. Fujii, T. *et al.* Pheromonal activities of the bombykol isomer, (10E,12E)-10,12-hexadecadien-1-ol, in the pheromone gland of the silkmoth Bombyx mori. *J. Insect Physiol.* **121**, 104018 (2020).

2. Bryan, P. N. Prodomains and protein folding catalysis. *Chem. Rev.* **102**, 4805–16 (2002).

3. Harrison, C. A., Al-Musawi, S. L. & Walton, K. L. Prodomains regulate the synthesis, extracellular localisation and activity of TGF-β superfamily ligands. *Growth Factors* **29**, 174–86 (2011).

4. Yosten, G. L. C. & Kolar, G. R. The Physiology of Proinsulin C-Peptide: Unanswered Questions and a Proposed Model. *Physiology* **30**, 327–332 (2015).

5. Cabral, W. A. *et al.* Mutations near amino end of alpha1(I) collagen cause combined osteogenesis imperfecta/Ehlers-Danlos syndrome by interference with N-propeptide processing. *J. Biol. Chem.* **280**, 19259–69 (2005).

6. Harrison, C. A., Al-Musawi, S. L. & Walton, K. L. Prodomains regulate the synthesis, extracellular localisation and activity of TGF-β superfamily ligands. *Growth Factors* **29**, 174–86 (2011).

7. Wakefield, L. M. & Hill, C. S. Beyond TGFβ: roles of other TGFβ superfamily members in cancer. *Nat. Rev. Cancer* **13**, 328–341 (2013).

8. Hinck, A. P., Mueller, T. D. & Springer, T. A. Structural biology and evolution of the TGF-β family. *Cold Spring Harb. Perspect. Biol.* **8**, a022103 (2016).

9. Joulia-Ekaza, D. & Cabello, G. Myostatin regulation of muscle development: molecular basis, natural mutations, physiopathological aspects. *Exp. Cell Res.* **312**, 2401–2414 (2006).

10. Pangas, S. A. Growth factors in ovarian development. *Semin. Reprod. Med.* **25**, 225–234 (2007).

11. Salazar, V. S., Gamer, L. W. & Rosen, V. BMP signalling in skeletal development, disease and repair. *Nat. Rev. Endocrinol.* **12**, 203–221 (2016).

12. Sánchez, M. P. *et al.* Renal agenesis and the absence of enteric neurons in mice lacking GDNF. *Nature* **382**, 70–73 (1996).

13. Moore, M. W. *et al.* Renal and neuronal abnormalities in mice lacking GDNF. *Nature* **382**, 76–79 (1996).

14. Colak, S. & Ten Dijke, P. Targeting TGF-β Signaling in Cancer. *Trends in cancer* **3**, 56–71 (2017).

15. McDonald, N. Q. *et al.* New protein fold revealed by a 2.3-A resolution crystal structure of nerve growth factor. *Nature* **354**, 411–414 (1991).

16. Shi, M. *et al.* Latent TGF-β structure and activation. *Nature* **474**, 343–349 (2011).

17. Vitt, U. A., Hsu, S. Y. & Hsueh, A. J. W. Evolution and Classification of Cystine Knot-Containing Hormones and Related Extracellular Signaling Molecules. *Mol. Endocrinol.* **15**, 681–694 (2001).

18. Holbourn, K. P., Acharya, K. R. & Perbal, B. The CCN family of proteins: structure-function relationships. *Trends Biochem. Sci.* **33**, 461–473 (2008).

19. Guo, J. & Wu, G. The signaling and functions of heterodimeric bone morphogenetic proteins. *Cytokine Growth Factor Rev.* **23**, 61–67 (2012).

20. Mellor, S. L. *et al.* Localization of activin beta(A)-, beta(B)-, and beta(C)-subunits in humanprostate and evidence for formation of new activin heterodimers of beta(C)-subunit. *J. Clin. Endocrinol. Metab.* **85**, 4851–4858 (2000).

21. Israel, D. I. *et al.* Heterodimeric bone morphogenetic proteins show enhanced activity in vitro and in vivo. *Growth Factors* **13**, 291–300 (1996).

22. Little, S. C. & Mullins, M. C. Bone morphogenetic protein heterodimers assemble heteromeric type I receptor complexes to pattern the dorsoventral axis. *Nat. Cell Biol.* **11**, 637–643 (2009).

23. Hata, A. & Chen, Y.-G. TGF-β Signaling from Receptors to Smads. *Cold Spring Harb. Perspect. Biol.* **8**, (2016).

24. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

25. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

26. Sengle, G., Ono, R. N., Sasaki, T. & Sakai, L. Y. Prodomains of transforming growth factor beta (TGFbeta) superfamily members specify different functions: extracellular matrix interactions and growth factor bioavailability. *J. Biol. Chem.* **286**, 5087–5099 (2011).

27. Molloy, S. S., Thomas, L., VanSlyke, J. K., Stenberg, P. E. & Thomas, G. Intracellular trafficking and activation of the furin proprotein convertase: localization to the TGN and recycling from the cell surface. *EMBO J.* **13**, 18–33 (1994).

28. Holyoak, T., Kettner, C. A., Petsko, G. A., Fuller, R. S. & Ringe, D. Structural basis for differences in substrate selectivity in Kex2 and furin protein convertases. *Biochemistry* **43**, 2412–2421 (2004).

29. Thomas, G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat. Rev. Mol. Cell Biol.* **3**, 753–766 (2002).

30. Blanchet, M.-H. *et al.* Cripto recruits Furin and PACE4 and controls Nodal trafficking during proteolytic maturation. *EMBO J.* **27**, 2580–2591 (2008).

31. Anderson, S. B., Goldberg, A. L. & Whitman, M. Identification of a novel pool of extracellular pro-myostatin in skeletal muscle. *J. Biol. Chem.* **283**, 7027–7035 (2008).

32. Cotton, T. R. *et al.* Structure of the human myostatin precursor and determinants of growth factor latency. *EMBO J.* **37**, 367–383 (2018).

33. Wang, X., Fischer, G. & Hyvönen, M. Structure and activation of pro-activin A. *Nat. Commun.* **7**, 12052 (2016).

34. Ge, G., Hopkins, D. R., Ho, W.-B. & Greenspan, D. S. GDF11 forms a bone morphogenetic protein 1-activated latent complex that can modulate nerve growth factor-induced differentiation of PC12 cells. *Mol. Cell. Biol.* **25**, 5846–5858 (2005).

35. Wolfman, N. M. *et al.* Activation of latent myostatin by the BMP-1/tolloid family of metalloproteinases. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15842–15846 (2003).

36. Le, V. Q. *et al.* Tolloid cleavage activates latent GDF8 by priming the pro-complex for dissociation. *EMBO J.* **37**, 384–397 (2018).

37. Gray, A. M. & Mason, A. J. Requirement for activin A and transforming growth factor--beta 1 pro-regions in homodimer assembly. *Science* **247**, 1328–30 (1990).

38. Green, D. Coagulation cascade. *Hemodial. Int.* **10 Suppl 2**, S2-4 (2006).

39. Zipfel, P. F. & Skerka, C. Complement regulators and inhibitory proteins. *Nat. Rev. Immunol.* **9**, 729–740 (2009).

40. Mi, L.-Z. *et al.* Structure of bone morphogenetic protein 9 procomplex. *Proc. Natl. Acad. Sci.* **112**, 201501303 (2015).

41. Sengle, G. *et al.* Targeting of bone morphogenetic protein growth factor complexes to fibrillin. *J. Biol. Chem.* **283**, 13874–13888 (2008).

42. Dong, X. *et al.* Force interacts with macromolecular structure in activation of TGF-β. *Nature* **542**, 55–59 (2017).

43. Liénart, S. *et al.* Structural basis of latent TGF-β1 presentation and activation by GARP on human regulatory T cells. *Science* **362**, 952–956 (2018).

44. Campbell, M. G. *et al.* Cryo-EM Reveals Integrin-Mediated TGF-β Activation without Release from Latent TGF-β. *Cell* **180**, 490-501.e16 (2020).

45. Walker, R. G. *et al.* Molecular characterization of latent GDF8 reveals mechanisms of activation. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E866–E875 (2018).

46. Bidart, M. *et al.* BMP9 is produced by hepatocytes and circulates mainly in an active

mature form complexed to its prodomain. *Cell. Mol. Life Sci.* **69**, 313–324 (2012).

47. Salmon, R. M. *et al.* Molecular basis of ALK1-mediated signalling by BMP9/BMP10 and their prodomain-bound forms. *Nat. Commun.* **11**, 1621 (2020).

48. Israel, D. I., Nove, J., Kerns, K. M., Moutsatsos, I. K. & Kaufman, R. J. Expression and characterization of bone morphogenetic protein-2 in Chinese hamster ovary cells. *Growth Factors* **7**, 139–150 (1992).

49. Wakefield, L. M. *et al.* Recombinant latent transforming growth factor beta 1 has a longer plasma half-life in rats than active transforming growth factor beta 1, and a different tissue distribution. *J. Clin. Invest.* **86**, 1976–1984 (1990).

50. Johnson, K. E. *et al.* Biological activity and in vivo half-life of pro-activin A in male rats. *Mol. Cell. Endocrinol.* **422**, 84–92 (2016).

51. McLennan, I. S. & Pankhurst, M. W. Anti-Müllerian hormone is a gonadal cytokine with two circulating forms and cryptic actions. *J. Endocrinol.* **226**, R45-57 (2015).

52. Picard, J.-Y. & Josso, N. Persistent Müllerian duct syndrome: an update. *Reprod. Fertil. Dev.* **31**, 1240–1245 (2019).

53. Bauskin, A. R. *et al.* The propeptide mediates formation of stromal stores of PROMIC-1: role in determining prostate cancer outcome. *Cancer Res.* **65**, 2330–2336 (2005).

54. Bauskin, A. R. *et al.* The TGF-beta superfamily cytokine MIC-1/GDF15: secretory mechanisms facilitate creation of latent stromal stores. *J. Interf. cytokine Res. Off. J. Int. Soc. Interf. Cytokine Res.* **30**, 389–397 (2010).

55. Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200 (2010).

56. Sengle, G., Ono, R. N., Sasaki, T. & Sakai, L. Y. Prodomains of Transforming Growth Factor β (TGFβ) Superfamily Members Specify Different Functions. *J. Biol. Chem.* **286**, 5087–5099 (2011).

57. Rider, C. C. & Mulloy, B. Heparin, Heparan Sulphate and the TGF-β Cytokine

Superfamily. *Molecules* **22**, 713 (2017).

58. Sandmark, J. *et al.* Structure and biophysical characterization of the human full-length neurturin-GFRa2 complex: A role for heparan sulfate in signaling. *J. Biol. Chem.* **293**, 5492–5508 (2018).

59. Runeberg-Roos, P. *et al.* Developing therapeutically more efficient Neurturin variants for treatment of Parkinson's disease. *Neurobiol. Dis.* **96**, 335–345 (2016).

60. Mochizuki, M. *et al.* Growth factors with enhanced syndecan binding generate tonic signalling and promote tissue healing. *Nat. Biomed. Eng.* **4**, 463–475 (2020).

61. Kotzbauer, P. T. *et al.* Neurturin, a relative of glial-cell-line-derived neurotrophic factor. *Nature* **384**, 467–470 (1996).

62. Yang, L. *et al.* GFRAL is the receptor for GDF15 and is required for the anti-obesity effects of the ligand. *Nat. Med.* **23**, 1158–1166 (2017).

63. Hsu, J.-Y. *et al.* Non-homeostatic body weight regulation through a brainstem-restricted receptor for GDF15. *Nature* **550**, 255–259 (2017).

64. Mullican, S. E. *et al.* GFRAL is the receptor for GDF15 and the ligand promotes weight loss in mice and nonhuman primates. *Nat. Med.* **23**, 1150–1157 (2017).

65. Emmerson, P. J. *et al.* The metabolic effects of GDF15 are mediated by the orphan receptor GFRAL. *Nat. Med.* **23**, 1215–1219 (2017).

66. Wang, X. Structural studies of GDNF family ligands with their receptors—Insights into ligand recognition and activation of receptor tyrosine kinase RET. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 2205–2212 (2013).

67. Parkash, V. *et al.* The structure of the glial cell line-derived neurotrophic factor-coreceptor complex: insights into RET signaling and heparin binding. *J. Biol. Chem.* **283**, 35164–72 (2008).

68. Wang, X., Baloh, R. H., Milbrandt, J. & Garcia, K. C. Structure of artemin complexed with its receptor GFRalpha3: convergent recognition of glial cell line-derived neurotrophic

factors. *Structure* **14**, 1083–92 (2006).

69. Mahato, A. K. & Sidorova, Y. A. RET Receptor Tyrosine Kinase: Role in Neurodegeneration, Obesity, and Cancer. *Int. J. Mol. Sci.* **21**, (2020).

70. Airaksinen, M. S. & Saarma, M. THE GDNF FAMILY: SIGNALLING, BIOLOGICAL FUNCTIONS AND THERAPEUTIC VALUE. *Nat. Rev. Neurosci.* **3**, 383–394 (2002).

71. Paratcha, G., Ledda, F. & Ibáñez, C. F. The neural cell adhesion molecule NCAM is an alternative signaling receptor for GDNF family ligands. *Cell* **113**, 867–879 (2003).

72. Ibáñez, C. F., Paratcha, G. & Ledda, F. RET-independent signaling by GDNF ligands and GFRα receptors. *Cell Tissue Res.* **382**, 71–82 (2020).

73. Li, J. *et al.* Cryo-EM analyses reveal the common mechanism and diversification in the activation of RET by different ligands. *Elife* **8**, (2019).

74. Adams, S. E. *et al.* A two-site flexible clamp mechanism for RET-GDNF-GFRα1 assembly reveals both conformational adaptation and strict geometric spacing. *Structure* (2021). doi:10.1016/j.str.2020.12.012

75. Kjaer, S., Hanrahan, S., Totty, N. & McDonald, N. Q. Mammal-restricted elements predispose human RET to folding impairment by HSCR mutations. *Nat. Struct. Mol. Biol.* **17**, 726–731 (2010).

76. Du, Z. & Lovly, C. M. Mechanisms of receptor tyrosine kinase activation in cancer. *Mol. Cancer* **17**, 58 (2018).

77. Bigalke, J. M. *et al.* Cryo-EM structure of the activated RET signaling complex reveals the importance of its cysteine-rich domain. *Sci. Adv.* **5**, eaau4202 (2019).

78. Tsui, C. C., Gabreski, N. A., Hein, S. J. & Pierchala, B. A. Lipid Rafts Are Physiologic Membrane Microdomains Necessary for the Morphogenic and Developmental Functions of Glial Cell Line-Derived Neurotrophic Factor In Vivo. *J. Neurosci.* **35**, 13233–13243 (2015).

79. Pierchala, B. A., Milbrandt, J. & Johnson, E. M. J. Glial cell line-derived neurotrophic

factor-dependent recruitment of Ret into lipid rafts enhances signaling by partitioning Ret from proteasome-dependent degradation. *J. Neurosci.* **26**, 2777–2787 (2006).

80. Li, A. Y. *et al.* RET fusions in solid tumors. *Cancer Treat. Rev.* **81**, 101911 (2019).

81. Lantieri, F., Griseri, P. & Ceccherini, I. Molecular mechanisms of RET-induced Hirschsprung pathogenesis. *Ann. Med.* **38**, 11–19 (2006).

82. Doray, B. *et al.* Mutation of the RET ligand, neurturin, supports multigenic inheritance in Hirschsprung disease. *Hum. Mol. Genet.* **7**, 1449–52 (1998).

83. Belli, C. *et al.* Progresses Toward Precision Medicine in RET-altered Solid Tumors. *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.* **26**, 6102–6111 (2020).

84. Sariola, H. & Saarma, M. Novel functions and signalling pathways for GDNF. *J. Cell Sci.* **116**, 3855–3862 (2003).

85. Trevaskis, J. L. *et al.* Neurturin and a Glp-1 Analogue Act Synergistically to Alleviate Diabetes in Zucker Diabetic Fatty Rats. *Diabetes* db160916 (2017). doi:10.2337/db16-0916

86. Lockhart, S. M., Saudek, V. & O'Rahilly, S. GDF15: A Hormone Conveying Somatic Distress to the Brain. *Endocr. Rev.* **41**, 610–642 (2020).

87. Lin, L. F., Doherty, D. H., Lile, J. D., Bektesh, S. & Collins, F. GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. *Science* **260**, 1130–1132 (1993).

88. Sampaio, T., Savall, A., Gutierrez, M. Z. & Pinton, S. Neurotrophic factors in Alzheimer's and Parkinson's diseases: implications for pathogenesis and therapy. *Neural Regen. Res.* **12**, 549 (2017).

89. Mahato, A. K. & Sidorova, Y. A. Glial cell line-derived neurotrophic factors (GFLs) and small molecules targeting RET receptor for the treatment of pain and Parkinson's disease. *Cell Tissue Res.* **382**, 147–160 (2020).

90. Oiwa, Y., Yoshimura, R., Nakai, K. & Itakura, T. Dopaminergic neuroprotection and

regeneration by neurturin assessed by using behavioral, biochemical and histochemical measurements in a model of progressive Parkinson's disease. *Brain Res.* **947**, 271–283 (2002).

91. Kordower, J. H. *et al.* Neurodegeneration prevented by lentiviral vector delivery of GDNF in primate models of Parkinson's disease. *Science* **290**, 767–773 (2000).

92. Gash, D. M. *et al.* Functional recovery in parkinsonian monkeys treated with GDNF. *Nature* **380**, 252–255 (1996).

93. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

94. Lerner, L. *et al.* MAP3K11/GDF15 axis is a critical driver of cancer cachexia. *J. Cachexia. Sarcopenia Muscle* **7**, 467–482 (2016).

95. Petry, C. J. *et al.* Associations of vomiting and antiemetic use in pregnancy with levels of circulating GDF15 early in the second trimester: A nested case-control study. *Wellcome open Res.* **3**, 123 (2018).

96. Wang, X., Chrysovergis, K., Bienstock, R. J., Shim, M. & Eling, T. E. The H6D variant of NAG-1/GDF15 inhibits prostate xenograft growth in vivo. *Prostate* **72**, 677–689 (2012).

97. Lindmark, F. *et al.* H6D polymorphism in macrophage-inhibitory cytokine-1 gene associated with prostate cancer. *J. Natl. Cancer Inst.* **96**, 1248–1254 (2004).

98. Hayes, V. M. *et al.* Macrophage inhibitory cytokine-1 H6D polymorphism, prostate cancer risk, and survival. *Cancer Epidemiol. biomarkers Prev. a Publ. Am. Assoc. Cancer Res. cosponsored by Am. Soc. Prev. Oncol.* **15**, 1223–1225 (2006).

99. Chung, H. K. *et al.* GDF15 deficiency exacerbates chronic alcohol- and carbon tetrachloride-induced liver injury. *Sci. Rep.* **7**, 17238 (2017).

100. Kempf, T. *et al.* GDF-15 is an inhibitor of leukocyte integrin activation required for survival after myocardial infarction in mice. *Nat. Med.* **17**, 581–588 (2011).

101. Zhang, Y. *et al.* Over-expression of growth differentiation factor 15 (GDF15) preventing

cold ischemia reperfusion (I/R) injury in heart transplantation through Foxo3a signaling. *Oncotarget* **8**, 36531–36544 (2017).

102. Liu, J. *et al.* Renoprotective and Immunomodulatory Effects of GDF15 following AKI Invoked by Ischemia-Reperfusion Injury. *J. Am. Soc. Nephrol.* **31**, 701–715 (2020).

103. Moschovaki-Filippidou, F. *et al.* Growth Differentiation Factor 15 Ameliorates Anti-Glomerular Basement Membrane Glomerulonephritis in Mice. *Int. J. Mol. Sci.* **21**, (2020).

104. Luan, H. H. *et al.* GDF15 Is an Inflammation-Induced Central Mediator of Tissue Tolerance. *Cell* **178**, 1231-1244.e11 (2019).

105. Bootcov, M. R. *et al.* MIC-1, a novel macrophage inhibitory cytokine, is a divergent member of the TGF-beta superfamily. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11514–11519 (1997).

106. Baek, S. J., Kim, K. S., Nixon, J. B., Wilson, L. C. & Eling, T. E. Cyclooxygenase inhibitors regulate the expression of a TGF-beta superfamily member that has proapoptotic and antitumorigenic activities. *Mol. Pharmacol.* **59**, 901–908 (2001).

107. Connolly, E. *et al.* Neurturin regulates the lung-resident macrophage inflammatory response to viral infection. *Life Sci. alliance* **3**, (2020).

108. Coll, A. P. *et al.* GDF15 mediates the effects of metformin on body weight and energy balance. *Nature* **578**, 444–448 (2020).

109. Day, E. A. *et al.* Metformin-induced increases in GDF15 are important for suppressing appetite and promoting weight loss. *Nat. Metab.* **1**, 1202–1208 (2019).

110. Knudsen, L. B. & Lau, J. The Discovery and Development of Liraglutide and Semaglutide. *Front. Endocrinol. (Lausanne).* **10**, 155 (2019).

111. Suter, U., Heymach, J. V & Shooter, E. M. Two conserved domains in the NGF propeptide are necessary and sufficient for the biosynthesis of correctly processed and biologically active NGF. *EMBO J.* **10**, 2395–400 (1991).

112. Uegaki, K. *et al.* BDNF binds its pro-peptide with high affinity and the common val66met polymorphism attenuates the interaction. *Int. J. Mol. Sci.* **18**, (2017).

113. Yan, R. *et al.* The Structure of the Pro-domain of Mouse proNGF in Contact with the NGF Domain. *Structure* **27**, 78-89.e3 (2019).

114. Teng, H. K. *et al.* ProBDNF induces neuronal apoptosis via activation of a receptor complex of p75NTR and sortilin. *J. Neurosci.* **25**, 5455–5463 (2005).

115. Nykjaer, A. *et al.* Sortilin is essential for proNGF-induced neuronal cell death. *Nature* **427**, 843–848 (2004).

116. Peränen, J., Rikkonen, M., Hyvönen, M. & Kääriäinen, L. T7 vectors with modified T7lac promoter for expression of proteins in Escherichia coli. *Anal. Biochem.* **236**, 371–3 (1996).

117. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

118. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389-94 (2015).

119. Vuillard, L., Rabilloud, T. & Goldberg, M. E. Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation. *Eur. J. Biochem.* **256**, 128–135 (1998).

120. Goldberg, M. E., Expert-Bezançon, N., Vuillard, L. & Rabilloud, T. Non-detergent sulphobetaines: a new class of molecules that facilitate in vitro protein renaturation. *Fold. Des.* **1**, 21–27 (1996).

121. Yamaguchi, H. & Miyazaki, M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **4**, 235–251 (2014).

122. Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1760–5 (2009).

123. Hilgarth, R. S. & Sarge, K. D. Detection of sumoylated proteins. *Methods Mol. Biol.* **301**, 329–338 (2005).

124. Tian, S., Huang, Q., Fang, Y. & Wu, J. FurinDB: A database of 20-residue furin cleavage site motifs, substrates and their associated drugs. *Int. J. Mol. Sci.* **12**, 1060–1065 (2011).

125. Hashimoto, T. *et al.* CLAC: a novel Alzheimer amyloid plaque component derived from a transmembrane precursor, CLAC-P/collagen type XXV. *EMBO J.* **21**, 1524–1534 (2002).

126. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119–139 (2005).

127. Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260 (2000).

128. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392–400 (2008).

129. Micsonai, A. *et al.* BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46**, W315–W322 (2018).

130. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**, 595–602 (2019).

131. Oganesyan, I., Lento, C. & Wilson, D. J. Contemporary hydrogen deuterium exchange mass spectrometry. *Methods* **144**, 27–42 (2018).

132. Szláma, G., Trexler, M. & Patthy, L. Latent myostatin has significant activity and this activity is controlled more efficiently by WFIKKN1 than by WFIKKN2. *FEBS J.* **280**, 3822–3839 (2013).

133. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2**, 1849–1861 (2007).

134. Sharma, R., Raduly, Z., Miskei, M. & Fuxreiter, M. Fuzzy complexes: Specific binding without complete folding. *FEBS Lett.* **589**, 2533–2542 (2015).

135. Kasahara, K., Terazawa, H., Takahashi, T. & Higo, J. Studies on Molecular Dynamics of Intrinsically Disordered Proteins and Their Fuzzy Complexes: A Mini-Review. *Comput. Struct. Biotechnol. J.* **17**, 712–720 (2019).

136. Turner, A. L. *et al.* Highly disordered histone H1-DNA model complexes and their condensates. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 11964–11969 (2018).

137. Linsley, P. S., Kallestad, J., Ochs, V. & Neubauer, M. Cleavage of a hydrophilic C-terminal domain increases growth-inhibitory activity of oncostatin M. *Mol. Cell. Biol.* **10**, 1882–1890 (1990).

138. Hermanns, H. M. Oncostatin M and interleukin-31: Cytokines, receptors, signal transduction and physiology. *Cytokine Growth Factor Rev.* **26**, 545–558 (2015).

139. Solá, R. J. & Griebenow, K. Effects of glycosylation on the stability of protein pharmaceuticals. *J. Pharm. Sci.* **98**, 1223–1245 (2009).

Appendix

Primer	Sequence
NRTN prodomain Rev	5’-GATGAAAGCTTTCAACGCGCACGACGACGACGCGG-3’
NRTN + TEV site Fwd	5’-GAAAACCTGTATTTCCAGTCTGCGCGTCTGGGCGCGCGTCCG-3’
NRTN + TEV site Rev	5’-AGACTGGAAATACAGGTTTTCGCCCGCACGACGACGCGGGCC-3’
NRTN C23S Fwd	5’-TCACTCCATGGGCATTTGGATGAGCCGTG-3’
NRTN-WT Fwd	5’-TATATCCATGGGCATTTGGATGTGCCGTGAAGGCCTGCT-3’
Generic HAT2 plasmid Rev	5'-TATATAGTTTAAACCCCTCAAGACCCGTTTAGAGG-3'

Appendix - Table 1. Oligonucleotides used for amplifying different human NRTN fragments. Primer sequences for cloning (Sigma-Aldrich). Fwd indicates forward primer and Rev indicates reverse primers. Wild type (WT).

Primer	Sequence
NcoI_STREPII_mGDF15_Fwd	TATATACCATGGCTTGGAGCCACCCGCAGTTCGAAAAAGGAAGC
NcoI_mGDF15_Fwd	TATATACCATGGCTCGAAATGGAG
HindIII_GDF15_Rev	CTCAAGCTTAAATGCAATGGCAATC
mGDF15_STREPII_Rev_N-Term	CTCCATTTCGAGCCATTGTGCTTCCTTTTTCGAACTG
HindIII_GDF15_fwd	GATTGCCATTGCATTTAAGCTTGAG

Appendix - Table 2. Oligonucleotides used for the cloning of tagged GDF15 constructs. Primer sequences for cloning (Sigma-Aldrich). Fwd indicates forward primer and Rev indicates reverse primers. Wild type (WT). Restriction sites and affinity tags encoded within primer sequences indicated.