“Life finds a way”
- Dr. Ian Malcolm (1993)
Declaration of Originality

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specified in the text. None of this work has been submitted in whole or in part for consideration for any other degree or qualification at the University of Cambridge, or any other University. This dissertation does not exceed the prescribed word limit.
Acknowledgements

The work presented in this thesis was made possible through the hard work and contributions of many people.

Marko, thank you for all your support and guidance over these last four years, it has been a real pleasure working as part of your group and I’m extremely grateful for the opportunity. Your genuine passion and commitment to solid, meaningful science is a serious source of inspiration for me, and many others. From your personal interest and involvement in each and every project, to your ability to calmly devise a strategy around any seemingly impossible barrier, one would be hard pressed to find a better supervisor. Most importantly, its not all always about the serious stuff and I deeply respect your appreciation for a frothy beverage at the end of the week. You’ll come around to the ‘hipster shit’ one of these days, and I owe you plenty.

Xuelu, I can’t think of a more reliable, hard working and conscientious person - the perfect scientist! I’ve learnt a great deal from you, and am forever in your debt for all of your patience and assistance over the years.

Gerhard, thanks for your patient attempts at explaining all things crystallography, and the many hours you spent refining shitty low resolution structures on my behalf. I appreciate all of your hard work and wish you all the best for the future.

Katharina and Beata - the twin pillars of the MH lab. Nothing works without you, and we mortals live in constant fear of those days when you’re away. Thanks for all of your help and friendly chats over the years, it is hugely appreciated.

To everyone else in the (extended) MH group, past and present: Tomasso, Max, Mathieu, Migle, Piotr, Paul, Aleksei, Marghe, Theo, Matt, Joe, Monique, Luana, Andrew and Michal (honorary) as well as our colleagues next door in the LP lab, you have all contributed in more ways than I could begin to describe here. Thanks for your advice, help, friendship and company, in and out of the lab.

I am also grateful for the various facilities and expertise available with in the Department of Biochemistry, including the Biophysics, Crystallography, and Protein Analysis facilities. Furthermore, continual access to beam time at the Diamond Light Source, European Synchrotron Radiation Facility (ESRF) and Soleil Synchrotron, made the work in this thesis
possible. For ease of access to these world-class resources I am exceptionally grateful. Cheers to Graham and the IT crowd for all technical support, Chris & Tom in photography, and to Clova, Janet and Tim for providing caffeine, sustenance and occasional abuse.

I would like to acknowledge our collaborators in the labs of Professor Thomas Thompson and Professor Timothy Springer across the Atlantic for their willingness to share data and ideas, and for coordinating co-publication of our highly complimentary research. Additionally, many thanks to Hongjia Zhu, Dr. Naomi Robertson and Professor David Spring in the Department of Chemistry, and Dr. Tan Yaw Sing (Bioinformatics Institute, A*STAR, Singapore) for design and synthesis of anti-myostatin peptides.

I would like to thank the Herchel Smith Fund who generously bankrolled this work and allowed me to live a life of relative luxury in Cambridge, as well as providing the opportunity to visit our counterparts in Boston. Many thanks to the managers who oversee the Fund and to Herchel Smith himself who was, by all accounts, a top bloke.

To my friends, both at home in NZ, here in the UK and elsewhere abroad - thanks for keeping me honest, even if that does involve continually slagging me off as a ‘dole-bludging nerdburger who has never worked an honest day in his life’. It’ll pay off one of these days, just you wait.

Mum, Dad and Henry - thanks for getting me here in the first place. Left to my own devices who knows where I might have wound up. It goes without saying that I’m forever grateful for the opportunities and support you’ve given me. This past year has been an unbelievably tough time for all involved, but despite the circumstances, it has been great being together. Onwards and upwards.

And finally, Ivy. Thank you for your love, support, encouragement and eye-rolling disapproval as required. You’ve kept me sane these past two years and I’d be lost without you.
From left: Andrew Oliveira, Aleksei Lulla, Marko Hyvönen, Paul Brear, Natalia Stepurko, Tom Cotton Xuelu Wang, Joe McLoughlin, Teodors Pantelejevs, Matt Ratcliff, Monique Liddar, Katy Davis

Absent: Katharina Ravn, Beata Blaszczyk, Luana Ferrara
Publications

An abridged version of the work presented in this doctoral study has been published, subject to peer-review, in the following manuscript:


Additionally, our collaborators in the group of Prof. Thomas Thompson at the University of Cincinnati have included results from this study in the following manuscript:


Special note: The above papers were simultaneously submitted as preprints to the BioRxiv preprint server and then published back-to-back with the following manuscript from the group of Prof. Timothy Springer, Harvard University:


I would like to thank all authors for their collegial and collaborative spirit, which allowed the co-publication of these highly complementary studies.
Summary

Myostatin is a secreted growth factor of the transforming growth-factor β (TGFβ) superfamily, and a powerful negative regulator of muscle mass in vertebrates. As such, there is considerable interest in developing pharmacological agents which inhibit myostatin signalling in order to stimulate muscle growth in the context of pathological muscle wasting. Like other TGFβ family proteins, myostatin is biosynthesised as an inactive (latent) precursor protein which requires proteolytic processing to liberate the mature bioactive growth factor. To examine the molecular basis of pro-myostatin latency and the mechanism by which it is activated in the extracellular space, I have determined the crystal structure of unprocessed human pro-myostatin and studied the properties of the protein at various stages of activation.

Crystallographic analysis of pro-myostatin reveals a unique domain-swapped dimeric structure, with an open V-shaped conformation distinct from the prototypical family member, TGFβ1. Following cleavage of the prodomains by furin, pro-myostatin persists as a stable non-covalent complex which is resistant to the natural inhibitor follistatin and exhibits significantly weaker bioactivity than the mature growth factor. A number of distinct structural features combine to stabilise the interaction between pro and mature domains and in doing so confer latency to the pro-complex. This facilitates a controlled, step-wise process of activation in the extracellular space and contributes to a complex network of regulatory control.

The results presented here provide a structural basis for understanding the effect of natural polymorphisms on myostatin function and a starting point for structure-guided development of next generation myostatin inhibitors. As a proof-of-concept, I present preliminary data on prodomain derived stapled peptides as inhibitors of myostatin signalling.
## Contents

Poignant quote .................................................. I

Declaration ....................................................... II

Acknowledgements .............................................. III

Publications ..................................................... VI

Summary ......................................................... VII

## Introduction

1.1 Cell signalling ................................................ 15

1.2 The growth-factor cystine-knot (GFCK) superfamily ............. 16

1.3 The transforming growth factor-β (TGFβ) superfamily ............. 18

1.3.1 Evolution of TGFβ superfamily proteins .................... 20

1.3.2 TGFβ superfamily signalling ............................... 21

1.3.3 Receptor activation and signal transduction .................... 21

1.3.4 Intracellular signal propagation .............................. 26

1.3.5 Structure of TGFβ superfamily ligands ..................... 27

1.3.6 Biosynthesis and activation .................................. 29

1.3.7 Role of the prodomain in the TGFβ superfamily .............. 30

1.3.8 Extracellular activation of latent pro-complexes ............... 33

1.3.9 Structural biology of TGFβ superfamily precursor proteins .... 35

1.4 Myostatin – negative regulator of muscle mass ................... 39

1.4.1 Pro-myostatin activation ..................................... 40

1.4.2 Myostatin signalling pathway ................................ 42

1.4.3 Myostatin regulation in the extracellular space ............... 44

1.4.4 Genetic targets of myostatin signalling and biological outcomes ... 46

1.4.5 Myostatin as a therapeutic target ............................ 49
## Production of recombinant human pro-myostatin

1. **Expression construct design**
2. **Protein expression and refolding**
3. **Purification of recombinant human pro-myostatin**
4. **Production of HRV-3C (PreScission) cleavable pro-myostatin**
5. **Analysis of the HRV-3C cleaved pro-myostatin complex**
6. **Conclusion**

## Functional characterisation of pro-myostatin

1. **Isolation of the mature myostatin growth factor**
2. **Quantification of myostatin bioactivity**
   - Maturation myostatin induces SMAD2/3 signalling with high potency
   - Myostatin signalling is inhibited by follistatin
   - The pro-myostatin complex is weakly bioactive
3. **The pro-myostatin complex is resistant to follistatin**
4. **Conclusion**

## Crystallisation & structure determination

1. **Initial pro-myostatin crystallisation trials**
2. **Disorder prediction and N-terminal truncation constructs**
3. **Low resolution structure determination**
   - Phasing by molecular replacement
   - Phasing by single wavelength anomalous dispersion (SAD)
   - Low resolution structure of pro-myostatin
4.4 Protein engineering in pursuit of improved diffraction ........................................ 91
4.4.1 $\beta$4-5 loop deletion ......................................................................................... 92
4.4.2 Surface entropy reduction .................................................................................. 92
4.5 High resolution structure determination and refinement ......................................... 94
4.6 Conclusion ........................................................................................................... 96

Structural analysis of human pro-myostatin .............................................................. 98

5.1 Pro-myostatin shows inter-protomer flexibility ....................................................... 100
5.2 The prodomain $\alpha$1 helix displaces the mature GF wrist helix ............................. 103
5.3 Pro-myostatin has an unexpected open-armed conformation ............................... 104
5.4 Pro-myostatin is a domain-swapped dimer ............................................................ 107
5.5 The prodomain forearm is a key latency conferring element ............................... 109
5.5.1 Interactions of the $\alpha$1 helix ........................................................................ 111
5.5.2 Interactions of the latency lasso and $\alpha$2 helix ............................................. 113
5.6 The prodomain arm stabilises the GF/forearm interaction .................................. 115
5.7 Conclusion ........................................................................................................... 118

Polymorphisms & targeted mutagenesis ..................................................................... 120

6.1 Bioactivity of selected pro-myostatin variants ..................................................... 122
6.2 Inhibition of myostatin by recombinant prodomain variants ............................... 125
6.3 Conclusion ........................................................................................................... 127

Discussion .................................................................................................................. 130

7.1 Proposed model of pro-myostatin biosynthesis and activation ............................ 131
7.2 Extracellular activation and biodistribution ......................................................... 135
7.3 Targeting extracellular activation of pro-myostatin ............................................ 137
7.4 Prodomain derived inhibitors of myostatin signalling ....................................... 138
List of Figures

1. The GFCK superfamily ........................................... 17
2. Cystine-knot motifs in the GFCK superfamily .................. 18
3. Cladogram of TGFβ superfamily members ....................... 19
4. TGFβ receptor ectodomain structures ........................... 23
5. TGFβ superfamily signalling overview .......................... 25
7. Alignment of representative TGFβ superfamily growth factors ... 28
8. Biosynthesis and maturation of TGFβ superfamily growth factors . . . 30
9. Crystal structures of TGFβ superfamily precursors .............. 36
10. Sequence alignment of selected TGFβ superfamily precursors .... 38
11. Sequence alignment of selected TGFβ superfamily precursors - continued ... 39
12. Hyper-muscular phenotypes in myostatin-null animals .......... 40
13. Generalised schematic of myostatin biosynthesis and activation .... 41
14. Myostatin-follistatin interactions ................................ 45
15. Overview of myogenesis .......................................... 47
16. Pro-myostatin expression construct design ....................... 56
17. Pro-myostatin construct expression tests ........................ 57
18. *In vitro* disulfide exchange reaction mechanisms .............. 58
19. SDS-PAGE analysis of small scale refolding trials .............. 60
20. Pro-myostatin refolding time course ................................ 61
22. Engineered HRV-3C cleavable pro-myostatin construct design .... 63
23. Purification of unprocessed pHAT2-MSTN-3C ..................... 64
24. HRV-3C processing of pHAT2-MSTN-3C .......................... 65
51 Prodomain forearm interaction surfaces . . . . . . . . . . . . . . . . . . . . . 115
52 Arm domain interactions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117
53 Human pro-myostatin missense mutations . . . . . . . . . . . . . . . . . . . 121
54 Amino acid residues selected for mutagenesis studies . . . . . . . . . . . . 123
55 Bioactivity of pro-myostatin variants . . . . . . . . . . . . . . . . . . . . . . . 124
56 Inhibition of myostatin signalling by recombinant prodomain variants . . 126
57 Model of pro-myostatin biosynthesis and activation . . . . . . . . . . . . . . 132
58 Latency conferring features of the myostatin prodomain . . . . . . . . . . . 133
59 Immunofluorescence of mouse tibialis anterior muscle cross-section . . . 136
60 Stapled peptides as inhibitors of myostatin signalling . . . . . . . . . . . . 139
61 Double-click stapling on diazido i7-48-55 peptide . . . . . . . . . . . . . . 155
# List of Tables

1. *In vitro* refolding buffer screen. ........................................... 59  
2. Mature myostatin bioactivity in HEK293T cells .......................... 72  
3. Published signalling parameters for mature myostatin .................. 73  
4. Inhibition of mature myostatin signalling by follistatin-288 ............... 74  
5. Bioactivity of myostatin precursor forms in HEK293T cells .......... 75  
6. Crystallographic data collection, processing and refinement statistics . . 97  
7. Experimentally determined SAXS parameters .................................. 106  
8. Inhibition of myostatin signalling by recombinant prodomain variants ....... 127  
9. Inhibition of mature myostatin signalling with stapled peptides ........... 139
Abbreviations

AA ............ Amino acid
ACN ............ Acetonitrile
ActRIIA/B ...... Activin receptor type II A/B
ALK ............ Activin receptor-like kinase
bHLH .......... Basic helix-loop-helix (transcription factors)
BLI ............ Biolayer interferometry
BMP ............ Bone morphogenetic protein
BMP1/TLD ...... Bone morphogenetic protein 1/tolloid-like protease
BSA ............ Bovine serum albumin
CHO cells ...... Chinese Hamster Ovary cells
CI ............. Confidence interval
CV ............. Column volume
CySH .......... Cysteine (reduced)
CySS .......... Cystine (oxidised)
$D_{\text{max}}$ .......... Maximum particle dimension
DMD ............ Duchenne Muscular Dystrophy
DMEM .......... Dulbecco’s Modified Eagle Media
DTT .......... Dithiothreitol
EC$_{50}$ ........ Half maximal effective concentration
ECM .......... Extracellular matrix
EDTA ........... ethylenediaminetetraacetic acid
EM .......... Electron microscopy
ER .......... Endoplasmic reticulum
FSD .......... Follistatin domain
FST288 ....... Follistatin 288 (288 amino acid isoform)
FSTL .......... Follistatin-like
GASP1/2 ...... Growth and differentiation factor-associated serum proteins 1 /2
GDF .......... Growth and differentiation factor
GF .......... Growth factor
GFCK .......... Growth-factor cystine knot
GST ................. Glutathione-S-transferase
HEK293 .............. Human embryonic kidney 293 cells
HEPES ............... 4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid
HRV-3C protease...... Human rhinovirus 3C protease
HSPG .................. Heparan sulfate proteoglycan
IC_{50} ................. Half maximal inhibitory concentration
I.M.A.G.E .............. Integrated Molecular Analysis of Genomes and their Expression Consortium
IPTG ................. Isopropyl β-D-1-thiogalactopyranoside
K_{d} .................. Equilibrium dissociation constant
kDa  .................. Kilodalton
LAP ................... Latency-associated peptide
LTBP .................. Latent TGFβ binding protein
MALDI ................. Matrix-assisted laser desorption/ionisation
MALS .................. Multi-angle light scattering
MAPK .................. Mitogen-activated protein kinase
MBP ................... Maltose binding protein
MND .................. Myonuclear domain
MR ..................... Molecular replacement
mTLL2 ................ Mammalian tolloid-like 2
ND .................... N-terminal domain
PBS ................... Phosphate buffered saline
PCR ................... Polymerase chain reaction
PDB ................... Protein Data Bank
PEG ................... Polyethylene glycol
PPS ................... 3-(1-Pyridinio)-1-propanesulfonate
R_{g} .................. Radius of gyration
RMSD .................. Root-mean-square deviation
RPC ................... Reversed phase chromatography
SAD .................. Single wavelength anomalous dispersion
SAXS .................. Small-angle X-ray scattering
SD .................. Standard deviation
SDS-PAGE ........Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SEC .............Size-exclusion chromatography
SEM .............Standard error
SeMet ............Selenomethionine
sERP ............Surface entropy reduction prediction
TCEP .............Tris(2-carboxyethyl)phosphine
TEV protease.....Tobacco etch virus protease
TFA .............Trifluoroacetic acid
TGFβ ............Transforming growth-factor β
TGN .............Trans-Golgi network
Introduction
1.1 Cell signalling

Cell signalling is a broad term encompassing the processes involved in the transmission of ‘information’ from one cell to another (or within the same cell) resulting in a biological response. These processes can be loosely categorized based on the range over which the signalling occurs, and the nature of the signalling effectors, which may include: ions, gases, small molecules, macromolecules and even electrical impulses.

*Intracrine* signalling describes the process by which a signal generated in a cell, remains within the confines of that cell, and elicits a response there. *Autocrine* signalling on the other hand requires first the secretion of a signalling molecule, which then acts back upon the signalling cell from the exterior.

If a signalling molecule secreted from one cell elicits a response in an immediately proximal cell, the signalling is considered *juxtacrine*. A classic example of juxtacrine signalling is the Notch/Delta system where ligand and receptor are both immobilised on the cell surface, requiring cell-cell contact for activation of signalling. A more abstract example of juxtacrine signalling involves the propagation of electrical action potentials across gap junctions, between neighbouring cardiac cells. Signalling molecules which are secreted into the extracellular space and act on target cells within the local environment are considered *paracrine* signals.

Finally, the long range signals which travel via the circulatory system to act on distant target cells are defined as *endocrine*. Insulin is a textbook endocrine hormone, being synthesised and secreted by pancreatic cells, and acting to regulate glucose metabolism in distant tissues and organs.

Cell signalling systems can also be grouped according to the mechanisms employed to transfer the signal from the target cell surface, across the membrane, to the cellular interior. Three classes of signalling receptors account for almost all signal transduction mechanisms; membrane ion channels, enzyme-linked transmembrane receptors and G-protein coupled transmembrane receptors. Activation (or in some cases deactivation) of these receptors relays the extracellular signal inside the cell, and in many cases triggers a biochemical cascade which allows intracellular amplification of the signal. The MAPK/ERK pathway provides a good example of biochemical signal amplification, in which the small GTPase Ras activates a number of distinct kinases (Raf, MEK, MAPK) each of which have their own downstream
targets [4]. This protein kinase cascade allows rapid amplification of the upstream signal and promotes a robust cellular response.

Careful control of exquisitely tuned signalling systems underpins the success of complex life. Signalling processes control all aspects of biology including embryonic patterning and development, tissue homeostasis and repair, growth and metabolism, immunity, reproduction, sensory perception, movement and so on. It is therefore no surprise that the breakdown of cell signalling integrity is often responsible for systemic pathologies such as cancer, autoimmune disease and diabetes.

1.2 The growth-factor cystine-knot (GFCK) superfamily

One broad type of cell signalling molecules are the so-called ‘growth factors’, a name sometimes used interchangeably (and perhaps incorrectly) with the term ‘cytokine’. These signalling proteins were historically associated with cell proliferation and growth, but the same name is now also used to describe many which can also have a negative effect on cell growth.

There are many familiar names which fall under the growth factor umbrella, including insulin and the insulin-like growth factors (IGFs), epidermal growth factors (EGFs), fibroblast growth factors (FGFs), vascular endothelial growth factors (VEGFs), colony stimulating factors and so on. The definition is loose and is often extended to include further small protein/peptide ligands such as the interleukins and interferons.

One of the largest and most diverse group of growth factors, and the focus of interest here, is the growth-factor cystine-knot (GFCK) superfamily (Fig. 1). This can be further divided into six sub-families; the TGF\(\beta\) superfamily, the DAN family of bone morphogenetic protein (BMP) antagonists, the glycoprotein hormone family (GPH), bursicon hormones (invertebrates only), the platelet-derived growth factor (PDGF) family (including the VEGF sub-family) and the neurotrophic factors [5, 6]. In addition, the multi-domain CCN proteins (CCN1-CCN6) contain a C-terminal cystine-knot domain which is structurally related to the aforementioned GFCKs [7].

Despite poor conservation in terms of sequence, and a plethora of biological functions, the 50+ protein ligands of the GFCK superfamily share a highly conserved common structural feature, the cystine-knot.
Figure 1: The GFCK superfamily. A. Representative protein folds from the GFCK superfamily (PDB accession codes are shown in parentheses for each structure). B. Cladogram showing inferred relationship of representative GFCK superfamily members. Generated from multiple sequence alignment of GFCK superfamily growth factors (neighbour-joining method, without distance corrections, ClustalOmega) and visualised with FigTree v1.4.3. Sub-groups within the superfamily are coloured accordingly and the total number of human growth factors within each subfamily is shown in parentheses. Sequences used for alignment are from human, except for those of the bursicon hormones which are from *Drosophila melanogaster*.

The cystine-knot is a unique structural motif, in which two disulfide bonds and their connecting peptide segments form a loop, which is then threaded by a third disulfide bond, effectively tying the ‘knot’ (Fig. 2) [5]. This arrangement covalently links four sections of polypeptide which are distal in sequence, and thus forms an exceptionally stable protein fold. This complex arrangement of disulfides was first described by McDonald *et al* (1991), who determined the crystal structure of nerve growth factor (NGF) [8]. The structure of many cystine-knot containing growth factors have been determined since and in almost all cases the sequence pattern of disulfide bond formation is conserved (Fig. 2) [9]. The domains which house the cystine-knot are also structurally conserved within the GFCK superfamily, despite poor sequence homology between (and even within) the subfamilies [5]. The GFCK domains are typically characterised by two long β-‘ribbons’, each consisting of two anti-parallel β-strands, and the central cystine-knot which pulls these strands together (Fig. 1).
Invariably these domains dimerise, and exert their biological functions as either homo- or heterodimers [6].

![Diagram of cystine-knot motifs](image)

**Figure 2:** Representative cystine-knot motifs from GFCK subfamilies, coloured by heteroatom (nitrogen blue, oxygen red, sulfur yellow, carbon white or green). In the cystine-knot motif, two disulfide bonds, and their connecting peptide segments (shown in white) form a loop through which a third disulfide passes (shown in green). The six cysteine residues involved in the cystine-knot are denoted I-VI, according to the order in which they appear in the sequence. The disulfide bond forming cysteine pairs (I+IV, II+V, III+VI) are almost entirely conserved across the GFCK superfamily. In most sub-families, the surrounding ring consists of eight residues (of which four are cysteines), however the cystine-knot of NGF and related neurotrophic factors contains a 14-membered ring.

### 1.3 The transforming growth factor-\(\beta\) (TGF\(\beta\)) superfamily

In the late 1970s it was shown that virally transformed cells secreted a ‘factor’ which allowed normal cells to grow in soft agar, a property traditionally reserved for malignant cells [10]. This so-called sarcoma growth factor (SGF) was subsequently shown to consist of multiple protein components which could be separated using chromatographic methods [10]. One such component of the SGF was named transforming growth factor-\(\alpha\) (TGF\(\alpha\)), which is now known to belong to the epidermal growth factor (EGF) family [11]. A separate fraction of SGF was shown to contain an activity which induced the growth of normal rat kidney (NRK) fibroblasts [11]. This activity was termed TGF\(\beta\), and before long, the responsible 25 kDa protein was successfully purified from human platelets and other tissues [12–14].

It became apparent that the physiological role of TGF\(\beta\) was far more complex than initially appreciated, with the effect on cell growth being highly dependent on the context in which TGF\(\beta\) was supplied. In contrast to the original observations, TGF\(\beta\) was also shown to
have a potent inhibitory effect on cell growth under certain conditions [15, 16]. This bifunctionality of TGFβ signalling was shown to result from the synergism with, or antagonism of additional growth factors (e.g PDGF, EGF) and that cellular responses to TGFβ were dependent on the interplay between multiple growth factors within a given context [16]. Furthermore TGFβ was shown to promote wound healing in normal cells and tissues, providing a physiological role for the growth factor outside the context of malignant transformation [17]. These discoveries triggered a landslide of TGFβ research which unearthed the TGFβ superfamily, the largest and most diverse family of cystine-knot growth factors.

The TGFβ superfamily represents a collection of truly multifunctional growth factors which control many aspects of development and tissue homeostasis, in both vertebrates and inver-
tebrates. In mammals, most of the 30+ genes encoding TGFβ superfamily members can be assigned to one of the following phylogenetic subgroups: the TGFβ isoforms, the bone morphogenetic proteins (BMPs), the activins/inhibins, growth and differentiation factors (GDFs), and the glial-derived neurotrophic factors (GDNFs) (Fig. 3) [6, 18].

There are a number of exceptions to these classifications, including anti-müllerian hormone (AMH) and nodal. While they are phylogenetically assigned to specific sub-groups, these proteins are functionally distinct from other family members. The Leftys, which function as antagonists of nodal signalling, are also unique within the TGFβ family. While Lefty 1 and 2 are 95% identical in their mature domains, they share no more than 25% pairwise identity with the remaining superfamily members. Finally, inhibin α heterodimerises with inhibin β chains, giving rise to antagonists which function to suppress the synthesis and secretion of follicle-stimulating hormone (FSH). The inhibin α chain is distantly related in sequence to the inhibin β chains, and is named according to its functional rather than phylogenetic relationship with these proteins.

1.3.1 Evolution of TGFβ superfamily proteins

Cystine-knot containing domains have enjoyed gainful employment as components of cell signalling pathways for more than 500 million years, appearing in the genomes of early pre-bilaterian metazoans (Ctenophora, Cnidaria, Porifera, Placozoa) [19]. As mentioned earlier, cystine-knot growth factors are found in at least six cell signalling families; the TGFβ superfamily, the DAN family of BMP antagonists, the glycoprotein hormone family (GPH), the bursicon hormones (invertebrates only), the platelet-derived growth factor (PDGF) family and the neutrophins (Fig. 1). Sequences representing all but the last of these six families are encoded in pre-bilaterian metazoans, where they presumably evolved alongside other signalling networks including the Wnt/Frizzled and Delta/Notch systems [19, 20].

Within the TGFβ superfamily, the BMPs and their associated DAN family antagonists are thought to have evolved earliest, leading up to the appearance of bilateral symmetry in metazoans (>550 million years ago)[19, 21]. Morphogenetic gradients of primitive BMPs and antagonists (i.e chordin) likely played a role in the early specification of bilateral left-right and dorso-ventral symmetry, and organogenesis. The first TGFβ proper isoform did

---

1It is important to clarify the distinction between ‘TGFβ superfamily’ and ‘TGFβ isoforms’, the latter referring specifically to the subgroup of three true TGFβ growth factors (TGFβ1, TGFβ2 and TGFβ3).
not appear on the scene until after the divergence of deuterostomes (the ancestors of modern day vertebrates) in the lead up to the ‘Cambrian explosion’ (c.a. 530 million years ago) [22]. Because the protostomes diverged prior to TGFβ evolution, their modern day descendants, namely the arthropods (including *Drosophila*) lack TGFβ homologs, but have several BMP homologs (in *Drosophila*: DPP, Screw, GBB) [22]. The activin clade, including inhibins, GDF11 and myostatin (GDF8) also have arthropod homologs, suggesting they arose before the split between deuterostomes and protostomes [22]. Following a prolific evolutionary history of gene duplication and re-purposing, the TGFβ superfamily now represents the most divergent and populous of the GFCK sub-families [6].

1.3.2 TGFβ superfamily signalling

TGFβ superfamily growth factors underpin many biological processes, from the earliest stages of development, to tissue homeostasis and repair in the fully developed organism [18, 23, 24]. While these growth factors are thought to function primarily as autocrine and paracrine signalling molecules, acting within the vicinity of their production, some family members are capable of long range endocrine signalling via the circulation. Some TGFβ superfamily members (nodal, activin, TGFβ1) are known to establish ‘morphogenetic gradients’, decreasing in effective concentration away from their source [25–27]. These gradients allow dose-dependent activation of target cells throughout the morphogenetic field, giving rise to spatial patterns of gene expression. This is rarely a simple case of radial diffusion away from the source, as interaction of morphogens with binding partners (both cell surface, and soluble) can dramatically change the shape of the morphogenetic gradient. Furthermore, cells at each position along the gradient must be capable of ‘detecting’ changes in ligand concentration, and translating these into proportionate cellular responses. The precise mechanisms by which cells read and respond to differential ligand concentrations remains the topic of much debate, and are reviewed elsewhere [23, 28–31].

1.3.3 Receptor activation and signal transduction

TGFβ superfamily signalling occurs via a common set of pathways, utilising a relatively small number of cell surface receptors compared to the number of ligands. The structurally homologous receptors through which ligands of the TGFβ superfamily signal (excluding the
more distantly related GDNF ligands), can be classed as either type I or type II TGFβ receptors [32–34]. Both types contain an extracellular ligand binding domain (ectodomain), single transmembrane helix and an intracellular kinase domain (which is constitutively active in the type II receptor). Specific combinations of homodimeric type I and II receptors interact with their cognate growth factors, forming heterotetrameric receptor complexes capable of signal transduction [33, 35].

The extracellular growth factor binding domains of both type I and II receptors are small (70-110 aa) and contain structurally related ‘three-finger toxin’ folds (Fig. 4) [36]. These folds share two structurally conserved small anti-parallel β-sheets (β-strands 1-2, and 3-5 respectively), and an extended loop between β-strands 4 and 5. Like the ligands to which they bind, the type I and II receptor ectodomains are disulfide-rich, containing up to six disulfide bonds, four of which are conserved across both receptor types. A single transmembrane spanning helix links the extracellular ligand binding domain to the intracellular kinase domain. Unlike other transmembrane receptor kinases, which are almost always tyrosine kinases, both type I and II TGFβ receptors are serine/threonine kinases. This unique kinase activity allows the TGFβ receptors to activate intracellular signal pathways distinct to those activated by tyrosine kinase receptors.

To date, there are 12 known TGFβ superfamily receptors (seven type I and five type II), compared with more than 30 known ligands [6]. Furthermore, certain TGFβ ligands have been shown to form heterodimers and so the number of possible ligands is even higher than the number of genes encoding them [37–39]. This suggests a considerable level of degeneracy in receptor utilisation, and thus poses a problem. How can ligand-specific responses arise from activation of the same receptors? Some of the answer lies in the fact that receptor hetero-oligomerisation (and the contribution of additional co-receptors) increases compositional receptor diversity over the individual receptors themselves. In some cases however, different ligands are capable of activating the same combinations of receptors [33].

Differential expression (both spatial and temporal) of ligands would partly resolve this degeneracy, as appears to be the case for the closely related ligands myostatin and GDF11 which utilise a common set of receptors, but are primarily expressed in different tissues. Myostatin expression is mostly restricted to developing and adult skeletal tissue, while GDF11 is more broadly expressed in developing embryonic tissues including the kidneys, spleen and brain.
Figure 4: Structures of representative type I and II TGFβ receptor ectodomains. Receptor ectodomain structures are coloured in rainbow, from N (blue) to C (red) terminus. Disulfide bonds are shown as sticks, with sulfur atoms in yellow. Type I and II receptor ectodomains consist of a three-finger toxin fold, with structurally conserved β-strands (1-5) and an extended β4-5 loop which typically forms part of the ligand binding interface.

[40, 41]. When the expression patterns of competitive ligands do overlap, differences in affinity for type I and II receptors provides an additional source of specificity at the receptor level [42].

The utilisation of non-canonical signalling receptors also contributes to the diversity of possible TGFβ superfamily signalling outcomes. Anti-müllerian hormone (AMH) is unique within the TGFβ superfamily in that while it signals through a common and promiscuous type I receptor (ALK3), it utilises its own unique type II receptor, AMHRII [43]. Given all AMH signalling proceeds non-redundantly via AMHRII, mutations that affect the function of this receptor are typically deleterious, and linked to developmental disorders of the male reproductive tract [43].
Excitingly, new receptor interactions are still being unearthed for ligands of the TGFβ superfamily. Recently, GDF15 (also known as macrophage inhibitory cytokine, MIC) was shown to be functionally more akin to the GDNF family ligands than to the growth and differentiation factors (GDFs) from which it takes its name [44]. GDF15 signalling is independent of the aforementioned TGFβ receptors and proceeds instead via the GDNF family receptor GFRAL and the tyrosine kinase RET, both of which are expressed in the central nervous system [44].

In the case of the TGFβ isoforms and activin-like ligands, the type II receptor is generally regarded as the high affinity receptor which presumably binds the ligand first, and subsequently recruits the low affinity type I receptor in a stepwise, cooperative manner [45]. The crystal structure of TGFβ3 in complex with type I and II receptor ectodomains shows that the ligand and high affinity type II receptor form a composite docking site for the low affinity type I receptor [45]. BMP receptor complex assembly appears to occur via a different mechanism, with the type I receptor providing the high affinity anchor. Unlike the TGFβ receptors described above, there is no evidence of interaction or appreciable affinity between the ectodomains of BMP receptor subtypes [46, 47].

In many cases, additional co-receptors are required to liberate the full signalling potential of the receptor complex. β-glycan, a membrane anchored proteoglycan, binds to all three TGFβ isoforms and enhances signalling by increasing availability of the ligand at the cell membrane [48]. Endoglin (alternatively called CD105) is another co-receptor for the TGFβ isoforms and acts to potentiate their signalling by interacting with type I and II receptors. Endoglin additionally binds to BMP9 on the surface of endothelial cells and traps the ligand in a mode compatible with type I receptor binding [49]. Similarly, Cripto is a known co-receptor for the ligand Nodal, and repulsive guidance molecule (RGM) family proteins act as auxillary components of BMP receptor assemblies [50, 51].

Formation of a competent ligand:receptor complex allows the constitutively active type II receptor kinase to phosphorylate the co-localised type I receptor at a conserved regulatory glycine-serine rich ‘GS’ motif within its cytoplasmic domain (Fig. 5). In the absence of ligand, the type I TGFβ receptor ALK5 is inhibited by the peptidyl-prolyl cis-trans isomerase FKBP12 [52]. This interaction is thought to prevent ‘leaky’ activation of signalling which may arise as the result of ligand-independent type I and II receptor collisions in the
Figure 5: Generalised overview of canonical TGFβ/SMAD signalling pathway. Growth factors bind high-affinity receptor ectodomains (grey) in the extracellular space. Subsequent recruitment of the low affinity receptor (blue/purple) completes the heteromeric receptor assembly, displacing inhibitory FKBP12, and allowing transphosphorylation of the cytoplasmic type I receptor kinase by the constitutively active type II receptor kinase. Phosphorylation of the type I receptor allows binding and phosphorylation of R-SMADs (red). Phosphorylated R-SMADs dimerise, and recruit the common SMAD4 (green) before translocation to the nucleus. SMAD complexes modulate the expression of responsive genetic elements via interaction with DNA binding transcription factors and/or coactivators and repressors.
membrane. Ligand-induced assembly of the receptor complex displaces FKBP12, exposing the GS motif phosphorylation sites to the type II kinase (Fig. 5). GS phosphorylation activates the type I kinase, allowing it to recruit and phosphorylate the intracellular effectors of signalling, thus initiating intracellular signal propagation [33, 35, 53].

1.3.4 Intracellular signal propagation

TGFβ superfamily signals are propagated from the cell membrane to the nucleus, via a number of possible intracellular routes. The most well documented of these is the SMAD pathway (Fig. 5). Intracellular signalling begins with phosphorylation of receptor-activated SMADs (R-SMADS) by the type I receptor kinase [35]. In vertebrates, there are 8 distinct SMADs (SMAD1-8), and which of these are phosphorylated is dependent on the composition of the activated ligand:receptor complex. TGFβ isoforms and activin related ligands typically activate R-SMADS 2 and 3, whereas BMP receptor complexes phosphorylate R-SMADS 1, 5 and 8 [54, 55].

The phosphorylated R-SMADs dimerise, and subsequently recruit the common SMAD-4. The activated R-SMAD/SMAD-4 complex is then capable of translocation to the nucleus, where it acts to modulate transcription of target genes either directly as a transcription factor itself, or indirectly via association with DNA binding transcription factors, co-activators and repressors [32, 56]. The remaining SMADs (6 and 7) are inhibitory, and prevent R-SMAD activation by competing for type I receptor kinase binding and/or disrupting the R-SMAD/SMAD-4 interaction. This linear view of intracellular SMAD signalling is highly simplified, and in reality there are additional interactions and levels of regulation at each stage of the process. Furthermore, the combination of possible receptor and SMAD complexes allows a diverse range of context-dependent signalling outcomes [32, 56].

Outside of the canonical SMAD signalling pathway, TGFβ receptors have been shown to activate SMAD-independent signalling pathways, including the ERK, JNK and p38 arms of the mitogen-activated protein kinase (MAPK) signalling pathway, and Rho-like GTPases in certain cell lines. These TGFβ receptor induced pathways may function to modify SMAD signalling, or may result in responses entirely independent of SMAD driven transcription events. The cross-talk between canonical and non-canonical pathways provides a mechanism of fine-tuning TGFβ beta superfamily signalling outcomes [32, 56–58].
1.3.5 Structure of TGFβ superfamily ligands

To date, the structures of 13 TGFβ superfamily growth factors have been determined and deposited in the Protein Data Bank (PDB). The bioactive ligands of the TGFβ superfamily are structurally homologous disulfide-linked dimers, with highly conserved patterns of cysteines, and extended antiparallel β-strands (Fig. 6, 7) [6, 36]. The protomers of this family are described as being analogous in structure to a cupped human hand, with the cystine-knot representing the palm, and the anti-parallel β-strands the fingers. The convex outer surface of the fingers is often referred to as the knuckle region. Additionally, most ligands contain a short ‘wrist’ helix between the first and second β-fingers (Lefty1 and 2 lack the conserved helical sequence). Upon dimerisation, the protomers associate palm to palm in such a way that the wrist helix from one protomer sits within the cupped fingers of the other (Fig. 6D).

Figure 6: Structure of TGFβ superfamily growth factors. A. Structure of the TGFβ1 protomer (PDB ID: 3KFD) coloured by rainbow from N-terminus (blue) to C-terminus (red). B. Same depiction as A, but recoloured to highlight position of the cystine-knot. C. Overlay of four representative structures of TGFβ superfamily protomers highlighting structural similarity (PDB IDs in parentheses). D. Structure of the disulfide-linked homodimer of TGFβ1. The second chain of the dimer is coloured grey.
As a general rule, the type II receptor epitopes are located on the convex ‘knuckle’ or toward the fingertips of each protomer, meaning the type II receptor ectodomain interacts with a single protomer only [6, 59]. Conversely, the type I receptor binding site is a composite site, composed of the wrist helix and concave surface of the fingers [45]. In this case, the type I receptor ectodomain contacts both protomers of the dimeric growth factor. These proteins are highly hydrophobic in nature which are dominated by apolar residues. This makes them notoriously prone to aggregation under physiological buffer conditions, and harsh acidic conditions and/or organic solvents are often required to maintain solubility of the purified mature proteins in vitro.

Figure 7: Clustal X alignment of representative TGFβ superfamily growth factor sequences (mature domains only). Alignments begin from the first residue following the furin cleavage site and are coloured according to sequence conservation from least conserved (white) to most conserved (blue). Cysteines involved in cysteine-knot formation are numbered (I-VI) and the secondary structural features are annotated below the alignment, based on the structure and sequence of TGFβ1.
1.3.6 Biosynthesis and activation

The mature, bioactive ligands of the TGFβ superfamily are small (100-120 amino acids (aa), \(\approx 12\) kDa per protomer), however these proteins are synthesised as much larger precursors which require processing to produce the active form \([6, 60]\). The nascent precursor protein contains an N-terminal signal peptide followed by a prodomain (also called pro-peptide) and finally the C-terminal mature growth factor (GF)(Fig. 8). The prodomain is typically larger (200-300 aa) than the C-terminal GF, and considerably less conserved in sequence (\(\approx 15\%\) pairwise identity across the superfamily, compared to >30% for the GF) (Fig. 10). The prodomain of AMH is exceptionally large (433 aa), while the Lefty isoforms and members of the GDNF sub-family have uncharacteristically small pro-regions (<80 aa)[6]. During translation, the nascent precursor is translocated to the endoplasmic reticulum (ER), during which the signal peptide is cleaved and the precursor protomers dimerise via disulfide linkage in the mature GF domains. The dimeric precursor is then transported through the golgi compartments for post-translational modification prior to secretion from the cell.

TGFβ superfamily precursors are inactive, and minimally, require proteolytic cleavage of the N-terminal prodomain to liberate signalling (Fig. 8). Proteases of the pro-protein convertase subtilisin/kexin (PCSK) family, including furin (PCSK3), are responsible for prodomain cleavage in the TGFβ superfamily \([61]\). These calcium-dependent serine endoproteases recognise and cleave dibasic RXXR motifs in many proprotein substrates, facilitating maturation of these targets as they transit through the secretory pathway \([62]\). Prodomain processing occurs predominantly in the trans-golgi network (TGN) but can also occur extracellularly by secreted pro-protein convertases, as is the case for nodal and myostatin \([63, 64]\).

For some members of the TGFβ superfamily, furin cleavage of the prodomain alone is not sufficient for activation, and further steps are required to liberate signalling. In many instances, the furin cleaved prodomains have significant affinity for their mature GFs, and persist as stable non-covalent pro-complexes even after secretion (Fig. 8) \([60, 65, 66]\). It is becoming increasingly apparent that the prodomains play a diversity of roles in post-translational regulation of TGFβ superfamily growth factors, as reflected by their highly divergent sequences.
**Introduction**

Growth factors of the TGFβ superfamily are biosynthesised as large, inactive precursors which require additional processing steps to liberate the bioactive ligand. Following translocation to the ER, nascent polypeptides dimerise via disulfide linkage in the mature growth factor domain. As the dimeric precursor traffics through the secretory pathway, furin and related pro-protein convertases cleave the N-terminal prodomains. In many cases the prodomains remain associated with the dimeric growth factor, which is secreted from the cell as a non-covalent complex. The secreted pro-complex may spontaneously dissociate to release the growth factor, or additional activating steps may be required.

### 1.3.7 Role of the prodomain in the TGFβ superfamily

Proteolytic processing of latent precursor proteins is a common mechanism for the activation of bioactive molecules, and is well documented in many biological systems. Protein hormones (i.e. insulin), proteases of the digestive system, and proenzymes (zymogens) of the blood coagulation cascade are examples of inactive pro-forms which are activated by controlled proteolysis of auto-inhibitory sequences. Premature activation of these molecules before arrival at their intended sites of action would have potentially disastrous consequences. For example, pepsin of the digestive system is biosynthesised as the inactive zymogen pepsinogen. This prevents the enzyme from digesting the cells of the stomach wall which produce it, allowing it to be safely released to the acidic environment of the stomach lumen where it undergoes an auto-proteolytic activation.

The secreted growth factors of the TGFβ superfamily are no exception, and require proteolytic activation by cleavage of the N-terminal prodomain as described in the previous

---

**Figure 8:** Biosynthesis and maturation of TGFβ superfamily growth factors. Growth factors of the TGFβ superfamily are biosynthesised as large, inactive precursors which require additional processing steps to liberate the bioactive ligand. Following translocation to the ER, nascent polypeptides dimerise via disulfide linkage in the mature growth factor domain. As the dimeric precursor traffics through the secretory pathway, furin and related pro-protein convertases cleave the N-terminal prodomains. In many cases the prodomains remain associated with the dimeric growth factor, which is secreted from the cell as a non-covalent complex. The secreted pro-complex may spontaneously dissociate to release the growth factor, or additional activating steps may be required.
section. Originally, the prodomain was thought to function primarily as a chaperone, facilitating folding, dimerisation and solubilisation of the growth factor during the early stages of biosynthesis [67]. Some GFCK proteins have considerably smaller prodomains, or lack them altogether, which suggests that the prodomain is not necessarily required for cystine-knot growth factor biosynthesis. Instead, the sheer diversity of prodomain sequences across the TGFβ superfamily points to a more complex biological function. We now know that the prodomains of TGFβ superfamily members are important modulators of signalling in many cases [6, 60, 65, 66, 68].

In 1990, Gray and Mason showed that the prodomains of TGFβ1 and activin A were required for successful dimerisation and secretion of the mature ligands [67]. Expression of the mature growth factors with a heterologous prodomain (from parathyroid hormone) failed to yield secreted dimers, while co-expression of the native prodomain in ‘trans’ led to the accumulation of bioactive dimers in the medium. This provided evidence for active involvement of the prodomain in maturation and export of TGFβ superfamily growth factors. Several years prior it had been shown that TGFβ was secreted from cells as an inactive complex with its prodomain, subsequently referred to as the latency associated peptide (LAP) [69, 70].

Since then, non-covalent interactions between pro and mature domains have been described for many family members [60, 65]. The relative affinities of cleaved prodomains for their mature GFs differ markedly across the superfamily. Some pro-complexes dissociate readily following furin proteolysis thus liberating their bioactivity, while other prodomains remain bound with sufficiently high affinity as to prevent the ligand from engaging its receptor.

The various TGFβ superfamily members can be grouped according to the extent to which the prodomain influences growth factor signalling. The BMP2 prodomain has very low affinity for its mature GF, and is known to dissociate following furin cleavage [68, 71]. BMPs 4, 5, 7, 9, 10 and GDF5 are known to form stable non-covalent complexes, the pro and mature domains of which co-migrate during velocity sedimentation in sucrose gradients [68]. Similarly, the recombinantly produced activin A pro-complex remains intact throughout purification, and has been thoroughly characterised in vitro using biophysical and biochemical methods [72]. However, at the picomolar concentrations where mature activin A is known to be biologically active, the prodomain is mostly dissociated and thus confers very little, if
any inhibitory effect on activin A signalling in vivo. The aforementioned BMPs/GDFs have prodomain affinities in a similar range, and likely spontaneously dissociate at low concentrations. As a result these prodomains do not appear to have an inhibitory effect in vivo and can be considered the ‘active’ pro-complexes [60, 73–75].

At the other end of the spectrum the prodomains of myostatin, GDF11 and the TGFβ isoforms are known to form exceptionally stable non-covalent complexes with their mature growth factors, rendering these ligands inactive until further activating processes release the prodomains [69, 76, 77]. These are the so-called ‘latent’ pro-complexes, and they represent an additional paradigm of signal regulation within the TGFβ superfamily. Interestingly, the inhibitory myostatin prodomain has an affinity for its mature GF comparable to those of several bioactive pro-complexes, including BMP4, BMP5 and activin A [60, 72]. This suggests that pro-complex bioactivity (or lack thereof) is not simply determined by the prodomain binding affinity, and that a more complex molecular basis for latency may exist.

The molecular basis for prodomain inhibition of growth factor signalling is an ongoing field of research, with considerable interest in developing prodomain derived biologics as specific inhibitors of growth factor signalling. In 2004, Jiang et al identified the inhibitory region of the myostatin prodomain [78]. Using truncated GST-prodomain fusions, an N-terminal region of the prodomain (residues 42-115) was shown to contain the structural elements required for inhibition of myostatin signalling. The C-terminal portion of the prodomain on the other hand did not inhibit signalling, but was shown to contribute to the stability of the prodomain. Subsequently, an N-terminal sequence corresponding to a predicted aliphatic α1 helix was shown, by mutagenesis, to be essential for stability of the latent pro-TGFβ1 complex [79]. The latent pro-TGFβ1 complex is additionally stabilised by covalent dimerisation of the prodomains [79].

Despite poor overall sequence homology, the inhibitory N-terminal prodomain sequences identified for myostatin and TGFβ1 are generally conserved across the TGFβ superfamily [6]. This suggests a common mode of association of the prodomain with the mature growth factor, and perhaps subtle sequence variations account for the differences in relative affinity and inhibitory potential of the various prodomains.
1.3.8 Extracellular activation of latent pro-complexes

For those TGF\(\beta\) family members secreted as latent pro-complexes, secondary activation processes are required to liberate the mature GF from the inhibitory prodomains. The mechanisms of prodomain release are specific to the individual family members, and represent important regulatory checkpoints in signalling. In some cases, additional proteolysis within the N-terminus of the prodomain is sufficient to trigger dissociation. This is known to be the case for myostatin and its close relative GDF11, both of which contain a recognition site for metalloproteases of the bone morphogenetic protein 1/tolloid (BMP1/TLD) family in the prodomain N-terminus (Arg98, Asp99 in myostatin) [77, 80]. Extracellular cleavage of the prodomain by BMP1/TLD destabilises the pro-myostatin complex, and releases the bioactive ligand [80]. The temporal and spatial expression patterns of the activating proteases may contribute to the regulation of growth factor signalling throughout development and into adulthood [64, 81].

The mechanism of latent TGF\(\beta\) activation is considerably more complex, and relies on the interaction of the latent complex with a number of extracellular proteins. Pro-TGF\(\beta\)1 is secreted from cells as a large latent complex with extracellular ‘milieu’ proteins, including latent TGF\(\beta\) binding proteins (LTBPs) and the leucine rich repeat proteins LRRC32 (GARP) and LRRC33 [82–84]. These proteins function to immobilise latent pro-TGF\(\beta\)1 in the extracellular matrix (ECM), or on the cell membrane in the case of LRRC32 and LRRC33.

During the early stages of biosynthesis, the latent pro-TGF\(\beta\) complex becomes tethered to these proteins covalently, via coupling to cysteines in the very N-terminal ‘protein association region’ of the prodomain. Once immobilised in the extracellular environment, the latent complex may undergo an integrin-dependent activation process. The prodomains of TGF\(\beta\)1 and 3 contain integrin (\(\alpha v\beta 6\) and \(\alpha v\beta 8\)) binding motifs in a region distant to the LTBP/GARP interaction site, as was discovered following determination of the pro-TGF\(\beta\)1 complex crystal structure [82]. Integrin binding tethers the complex to the actin cytoskeleton of an adjacent cell (or the same cell), and mechanical force generated by cytoskeletal movement creates traction across the complex, which is resisted by the covalent linkage to the ECM on the opposite side. Sufficient force across the complex compromises conformational integrity, and allows release of the mature GF from the prodomain [82, 85]. This may
provide a mechanism for the controlled, localised release of the growth factor following cellular distortion at sites of injury, consistent with the known role of TGFβ in wound healing and fibrosis [86].

Over-activation of TGFβ due to loss of retention in the ECM is responsible for the symptoms of Marfan syndrome, a hereditary disorder of the connective tissue. The mutations in fibrillin-1 associated with this disease disrupt the matrix sequestration of LTBP, and latent TGFβ, leading to excessive TGFβ signalling [87].

TGFβ isoforms are not the only growth factors stored as precursor forms in the ECM. Anderson et al (2008) demonstrated sequestration of unprocessed pro-myostatin in the ECM by LTBP3, the primary LTBP expressed in skeletal muscle [64]. Although the prodomains of myostatin and GDF11 contain cysteines analogous to those used by TGFβ for coupling to LTBP1, the interaction with LTBP3 is not dependent on covalent linkage. LTBP1 mediated regulation of signalling has also been described for activin A, the N-terminal association region of which is largely conserved with myostatin/GDF11 [88]. Pro-myostatin and pro-activin make additional interactions with heparan sulfate proteoglycans (HSPGs) including perlecan, which is immobilised in the ECM through interaction with fibrillins [60, 89]. Sequences within the prodomains mediate interactions with the charge-dense heparan sulfate side chains of these ECM proteins. Other pro-TGFβ superfamily members are capable of docking directly onto extracellular fibrillin, which appears to contain promiscuous high affinity binding sites for many growth factors [68].

Extracellular targeting and sequestration regulates the bioavailability of TGFβ superfamily ligands, and gives rise to locally concentrated pools of growth factors which may be activated as required. Interaction of pro-complexes with ECM proteins may also enhance the functional affinity of prodomains for their growth factors, by stabilising inhibitory conformations for example. Non-covalent cross-linking of the activin prodomains by heparan sulfate chains is speculated to maintain the otherwise active pro-complex in a latent state while immobilised in the ECM [72]. In combination with LTBP interaction, this would also allow for a mechanism of activation during injury, analogous to the integrin dependent mechanism employed by TGFβ isoforms.

Given the complex involvement of ECM milieu molecules in TGFβ superfamily storage and signalling, we must take care when making biological assumptions based on in vitro
observations of the isolated growth factors. It is likely that these isolated species exist only transiently in nature.

1.3.9 Structural biology of TGFβ superfamily precursor proteins

The prodomains of the TGFβ superfamily are considerably less conserved in sequence than their respective mature domains, perhaps reflecting divergence in the functional roles of these domains. There are however a number of structural features that appear to be conserved throughout.

The crystal structure of a latent pro-TGFβ1 complex was published in 2011 and provided the first glimpse at a TGFβ superfamily prodomain [82](Fig. 9). Furthermore, this structure provided an elegant molecular explanation for latency of the prodomain bound complex and a model for integrin-dependent activation. Many of the structural features described for the TGFβ1 prodomain are conserved within the sequences of other family members and as a result, the prodomains can be described in general terms to a certain extent. Crystal structures of pro-BMP9 and pro-activin A complexes have since been determined, expanding our understanding of prodomain structure and function [72, 90].

For ease of description, the prodomain can be divided into structurally distinct N- and C-terminal portions. The N-terminal part of the prodomain is characterised by the α-helices which grasp the mature dimeric growth factor (Fig. 9). Together, these ligand binding elements are considered the prodomain ‘forearm’. The C-terminal portion of the prodomain consists of a structurally conserved globular core, otherwise known as the ‘arm’ or ‘shoulder’ domain (Fig. 9).

The N-terminal forearm contains an unstructured ‘protein association’ region, followed by two alpha helices, and a flexible linker to the C-terminal ‘arm’ domain. The first (α1) and second (α2) helices are linked by a ‘latency lasso’ of variable length and together these elements wrap around one protomer of the mature growth factor, occluding both receptor binding sites (Fig. 9). The aliphatic sequence signatures for the α1 and α2 helices are largely conserved throughout the superfamily, suggesting a common mode of interaction between the prodomains and GFs (Fig. 10). Interestingly, the α1 helix and latency lasso are not observed in the pro-BMP9 complex structure, and are assumed to be disordered (Fig. 9). Instead, a unique α-helical extension from the arm domain occupies a similar site on the
Figure 9: Crystal structures of TGFβ superfamily pro-complexes. In each structure the mature growth factor is coloured orange, and one of the associated prodomains is coloured blue, green or pink. The second prodomain in each structure is coloured grey. While the three structures show marked topological differences, the structures of individual domains are largely conserved. In each case, the prodomain can be divided into an N-terminal ‘forearm’ which grasps the mature GF protomer and a globular C-terminal ‘arm’ domain (circled with dotted line). In the structures of pro-TGFβ1 and pro-activin A, the α1 helix of the forearm represents a key interface between the pro and mature domains. In the structure of pro-BMP9, the α1 helix is not visible in the density and is presumably disordered.

mature GF to that of the α1 helix in pro-TGFβ1 and pro-activin A [90].

The C-terminal arm domain consists of a conserved β-stranded fold with two anti-parallel β-sheets, a short α-helix and a considerable hydrophobic core. The arm domain of pro-TGFβ1 contains a unique β8/9 extension which facilitates disulfide dimerisation of the prodomains at a so-called ‘bow-tie’ motif (Fig. 9) [82]. These prodomain cysteines are unique to the TGFβ isoforms, however this does not preclude the potential for non-covalent dimerisation
of the prodomains in other family members. In all published structures so far, the $\beta_1$ strand of the arm domain hydrogen bonds to some extent with the $\beta_7$ strand of the mature GF, creating an extended $\beta$-sheet which spans the mature growth factor and prodomain.

While the structurally characterised prodomains display overall similarity in tertiary structure, the relative orientation of pro and mature domains differs markedly between the three known pro-structures. The prodomains of pro-TGF$\beta_1$ adopt an overall ‘closed’ or ‘crossed-arm’ conformation, as a result of dimerisation at the bowtie and considerable twisting at the $\beta$-sheet interface of mature and prodomains. This conformation creates a doughnut like shape, in which the elements of the prodomain completely encircle the mature dimeric growth factor and form a highly stable latent complex.

The structures of pro-activin A and pro-BMP9 complexes on the other hand reveal ‘open-armed’ conformations in which there is no contact between individual prodomain arms (Fig. 9) [72, 90]. This lack of prodomain interaction presumably decreases their functional affinity relative to that of pro-TGF$\beta$ which gains a significant avidity advantage through covalent dimerisation of the prodomains. The reduced stability of the open-armed complexes provides an explanation for the observed bioactivity of the pro-activin A and pro-BMP9 complexes, which signal with potency comparable to that of their respective mature ligands [72, 90]. Dissociation of these open-armed complexes at low concentrations, or active displacement by receptors competing for the prodomain bound sites, would allow signalling to proceed unimpeded by the prodomain.

From these three structures we can rationalise that the pro-complexes of the TGF$\beta$ superfamily fall into one of at least two subgroups: the open-armed active pro-complexes, or the closed-arm latent complexes. The existence of additional types is possible, and it has even been suggested that these pro-forms may be able to interconvert between open and closed arm conformations depending on interactions with additional binding partners [6, 90].

Despite structural homology at the domain level, the topological diversity and low sequence identity of the prodomains makes structural predictions and homology modelling of other family members very difficult. Structural analysis of additional family members is required if we are to fully understand the structure-function relationships that underpin growth factor regulation in the TGF$\beta$ superfamily.
Figure 10: Structure-guided sequence alignment of selected human TGFβ superfamily precursors. Sequences aligned with Tcoffee-Expresso server using PDB structures 3RJR, 5HLY and 5YCG for structural alignment (inhibin βA is the protomeric subunit of activin A). Alignment was visualised with Jalview and coloured by clustal scheme. Fully conserved residues are indicated with an asterisk. Secondary structural elements are annotated above the sequences for prodomain (blue) and mature (red), based on the structure and sequence of pro-TGFβ1. Alignment continued on next page.
1.4 Myostatin – negative regulator of muscle mass

In 1997, McPherron and colleagues published the discovery of a new TGFβ superfamily member [91]. Using degenerate polymerase chain reaction (PCR) with primers complementary to conserved growth factor sequences, the authors identified a new sequence from mouse genomic DNA, which they designated GDF8. The encoded murine protein was shown to contain all the hallmarks of a TGFβ superfamily ligand, including N-terminal prodomain, furin processing site, and C-terminal domain with characteristic pattern of cysteines required for cystine-knot formation.

Expression of GDF8 mRNA was restricted to the myotome (muscle progenitor) compartment of somites in developing mouse embryos, and almost exclusively in the skeletal muscle of adult mice [91]. In order to determine the biological role of the newly discovered GDF8, transgenic mice were raised with a targeted disruption to the GDF8 gene. Remarkably, homozygous mutants were on average 30% larger than wild-type littermates. This increased size was accounted for by a dramatic increase in skeletal muscle mass, as a result of increased number (hyperplasia) and size (hypertrophy) of muscle fibres [91]. The authors concluded...
by suggesting that GDF8 functions specifically as a powerful negative regulator of skeletal muscle growth, and on that basis, renamed the new growth factor myostatin.

Figure 12: Examples of myostatin-null animals with ‘double-muscled’ phenotypes. A. Belgian-blue bull with loss-of-function myostatin mutation (taken from www.strangeanimals.info/2010/12/belgian-blue-super-cow.html). B. Hyper-muscular myostatin-null whippet (taken from www.europuppy.com/blog). C. Double muscled upper-limb of myostatin-null mouse (lower-right) compared to a wild-type littermate (upper-right)[91].

Shortly after the discovery and preliminary characterisation of myostatin, deleterious mutations were identified within the coding sequence of bovine myostatin in Belgian-blue and Piedmontese breeds [92]. These breeds are characterised by higher levels of muscle than other cattle types and display similar phenotypic characteristics as were seen in myostatin-null mice (Fig. 12). In 2004, a report in the New England Journal of Medicine detailed the identification of a human child with a loss-of-function myostatin mutation, and associated hyper-muscular phenotype similar to those observed in mice and cattle [93]. These findings suggested that the negative regulatory effect of myostatin on muscle was conserved between mammalian species, and highlighted myostatin as a potential target of interest for both agricultural and medical applications.

1.4.1 Pro-myostatin activation

Myostatin is secreted from cells predominantly as the unprocessed precursor (pro-myostatin). The prodomains are then cleaved in the extracellular space by furin-like proteases to form a stable complex consisting of the mature dimeric GF and two non-covalently associated prodomains.

---

2The term ‘pro-myostatin’ refers specifically to the unprocessed precursor form, while ‘pro-myostatin complex’ refers to the non-covalent complex of the mature growth factor and associated prodomains.
prodromes (Fig. 13) [64, 76]. The pro-myostatin complex was shown to circulate as an inactive (latent) form that could be activated by acid treatment, which presumably causes dissociation of the inhibitory prodromes [78, 94]. The bioactive mature ligand could be isolated from the latent complex by reversed-phase chromatography, and was shown to bind the activin type II receptors ActRIIa and ActRIIb [76, 95].

Figure 13: Generalised schematic of myostatin biosynthesis and activation in muscle tissue. Following dimerisation in the endoplasmic reticulum (ER), the pro-myostatin precursor transits through the trans-golgi network (TGN) wherein some fraction may undergo intracellular processing by furin (dashed lines). The majority of pro-myostatin however is secreted as the unprocessed precursor, which becomes immobilised in the ECM. Extracellular cleavage of the prodromes by furin yields a latent non-covalent pro-complex. Additional proteolysis of the prodomain by BMP1/TLD family proteases is required to disrupt the pro-complex and release the bioactive ligand.

The inhibitory capacity of the myostatin prodomain was reinforced by its ability to block mature myostatin from binding ActRIIb in vitro [76]. Furthermore, transgenic mice overexpressing the free prodomain had increased skeletal muscle, suggesting inhibition of the mature ligand in vivo [76, 96]. In addition to inhibition by the prodomain, myostatin signalling could also be effectively suppressed by the known activin antagonist follistatin and a dominant negative form of the ActRIIb receptor [76].

Although it had been shown that the latent pro-myostatin complex could be ‘activated’ by
exposure to harsh acidic conditions, a physiological mechanism for activation was still missing. During characterisation of myostatin secretion, the appearance of a recurring cleavage product in the conditioned medium of Chinese hamster ovary (CHO) cells expressing pro-myostatin suggested secondary proteolysis of the prodomain was occurring [80]. Proteases of the BMP1/TLD family of metalloproteases were identified as the most likely candidates for pro-myostatin complex activation in vivo, cleaving at a site in the prodomain’s N-terminal forearm (Arg98/Asp99). An engineered form of the prodomain with a cleavage resistant BMP1/TLD site resulted in significantly more muscle growth than the wild type prodomain, when injected into adult mice [80]. These findings supported the hypothesis that proteases of the BMP1/TLD family were responsible for pro-myostatin complex activation in vivo (Fig. 13). Specifically, mammalian tolloid-like 2 (mTLL2) of the BMP1/TLD family is expressed predominantly in developing muscle tissue, and thus has been suggested as the putative protease of physiological relevance [80, 97]. The molecular details of pro-myostatin activation remain unclear, including the way in which the non-covalently bound prodomains confer latency to the mature ligand. Access to a high resolution structure of pro-myostatin will be essential for understanding these interactions.

1.4.2 Myostatin signalling pathway

Twenty years on from the discovery of myostatin, the details of its signalling pathway are now reasonably well understood. Myostatin signals preferentially through the high affinity activin type II receptor ActRIIib, and to a lesser extent ActRIIa [76, 95]. Assembly of the complete receptor complex proceeds via recruitment of either the type I activin receptor ALK4, or the type I TGFβ receptor ALK5. The type I receptors bind to myostatin with significantly lower affinity than ActRIIib, and detection of type I receptor binding is dependent on pre-formation of a type II:ligand complex [42, 95]. This suggests a cooperative, sequential mechanism of receptor binding, as is established for TGFβ3 [45]. The structural basis for ActRIIib and ALK4/5 binding cooperativity is unclear, and as of yet there are no available structures of myostatin in complex with its receptors.

The mode of ActRIIib binding to myostatin is likely conserved with that of activin A, which binds the type II receptor via an epitope on the convex ‘knuckles’ of the mature growth factor [98–100]. Unlike activin A however, myostatin is also capable of signalling through
the TGFβ like type I receptor, ALK5. Despite overall structural similarity, the ‘pre-helix region’ of mature myostatin is poorly conserved between activin A and myostatin, and is thought to account for the difference in type I receptor specificity between these ligands. Indeed, swapping the myostatin pre-helix sequence into activin A was shown to be sufficient to permit non-canonical ALK5 signalling \[100\].

GDF11, the closest relative of myostatin (90% identical in the mature domain), also signals through ActRIIb and ALK4/5, but can additionally utilise the type I receptor ALK7 \[101\]. GDF11 was recently shown to induce signalling with higher potency than myostatin in several cell lines, due to subtle structural variations within the type I receptor binding site \[42\]. The ability of myostatin and GDF11 to signal through receptors from both TGFβ and activin subclasses make these ligands unique within the TGFβ superfamily.

Myostatin and GDF11 signalling induces phosphorylation of R-SMADs 2/3 by the activated ALK4/5/7 type I receptor and once activated, these SMADs translocate to the nucleus for modulation of responsive genetic elements. The apparent functional overlap of myostatin and GDF11, and the difficulty in disentangling their effects \textit{in vivo} has led to considerable controversy in the field as is reviewed elsewhere \[102\].

In addition to canonical SMAD-dependent signalling, myostatin is also capable of exerting its muscle-suppressive effect via the p38 MAP kinase (MAPK) pathway \[103\]. This non-canonical pathway has also been implicated in the signalling of other TGFβ superfamily growth factors \[56\]. Activation of p38 is thought to be dependent on upstream activation of the MAP kinase kinase (M KK) TAK1. Loss of myostatin induced p38 activation by expression of a kinase-dead type I receptor (ALK5) suggests that the type I receptor is the upstream branch point for SMAD-dependent and independent pathways \[103\]. Myostatin induced p38 signalling has been shown to operate both independently of the SMAD pathway, and cooperatively with it. This cooperation is likely mediated in-part by the p38 activated transcription factor ATF2, which interacts directly with SMADS 3/4, allowing fine-tuning of the transcriptional response \[103\]. Cross-talk between signalling pathways downstream of the activated type I receptor allows a diversity of context dependent transcriptional outcomes.
1.4.3 Myostatin regulation in the extracellular space

Following secretion from the cell, there is a constant jostle for possession of myostatin by various extracellular proteins. The secreted pro-myostatin precursor is sequestered in the ECM through prodomain mediated interactions with LTBP3 and HSPGs until proteolytic processing by furin and BMP1/TLD proteases releases the mature ligand [60, 64]. This ECM bound precursor is now known to constitute the major storage form of myostatin in skeletal muscle, where it resides predominantly within the interstitial space between muscle fibres [64, 104]. Myostatin is also detectable in serum, and is thought to circulate in the blood primarily as the non-covalent latent pro-complex [64].

Once separated from its inhibitory prodomains, the mature ligand becomes a target for soluble antagonists within the extracellular space. Follistatin is a secreted multi-domain protein, and a broad antagonist of TGF\(\beta\) superfamily ligands. In addition to myostatin, follistatin also binds to and inhibits several of the BMPs, activins, inhibin and GDF11 [6]. Follistatin and other extracellular antagonists are important regulators of TGF\(\beta\) superfamily signalling, and are known to play vital roles in establishing the morphogenetic gradients required for embryogenesis [105]. Follistatin-null mice show widespread birth defects in many tissues and organs, and die shortly after birth [105]. These observations suggest a broad role for follistatin in developmental regulation, likely manifest through modulation of multiple TGF\(\beta\) superfamily growth factors. As mentioned previously, muscle-specific over-expression of follistatin in transgenic mice leads to hyper-muscular phenotypes associated with suppressed myostatin signalling [76].

Crystal structures of both myostatin and activin A in complex with follistatin provide significant insight into the molecular basis of inhibition [100, 106–108]. Both full length follistatin (FST288), and the shorter paralog which lacks the C-terminal FSD3 domain (FSTL3), form high affinity interactions with mature myostatin through a conserved binding mode (Fig. 14). The N-terminal domains (ND) of both paralogs occupy the putative type I receptor site and form contacts with both myostatin protomers, while the FSD1 and FSD2 domains mask the putative type II receptor site on the convex face of a single mature protomer. In this mode, two molecules of follistatin wrap around the dimeric ligand, effectively blocking all four putative receptor binding sites (Fig. 14). Additionally, full length follistatin has a high affinity for heparin which facilitates targeting of its bound ligand to the cell surface for
Figure 14: Crystal structures of myostatin in complex with follistatin paralogs FSTL3 and FST288. The mature myostatin dimer is shown as orange/pale orange surface representation, with location of putative receptor binding sites indicated in the left most structure. Follistatin chains are shown as ribbon representations, one chain of which is coloured by rainbow from N (blue) to C (red) terminus. The second follistatin chain is coloured grey. The binding mode is conserved between the two paralogs, with ND domains and FSD1/FSD2 domains interacting with the putative type I and II receptor sites, respectively. FST288 has an additional C-terminal FSD3 domain that is not present in FSTL3.

endocytosis and degradation [109, 110].

In addition to follistatin, the growth and differentiation factor-associated serum proteins 1 and 2 (GASP-1 and GASP-2) are known to specifically antagonise myostatin and GDF11 in the extracellular space [111, 112]. These multi-domain proteins contain a follistatin domain which is involved in growth factor recognition, however the additional domains are required for formation of a high affinity complex [111]. Furthermore, pro-myostatin is known to bind to the ECM associated proteoglycans decorin and perlecan which function to modulate its activity in the extracellular space [60, 113].

The aforementioned secreted proteins regulate myostatin signalling either by directly blocking receptor epitopes, or by immobilising the latent growth factor in the ECM. Together, these interactions contribute to a complex network of regulatory control.
1.4.4 Genetic targets of myostatin signalling and biological outcomes

The physiological effects of myostatin signalling on muscle growth and homeostasis are well described, however the cellular processes which underpin these biological outcomes remain the topic of investigation. In addition to the summary given below, the biology of myostatin and myogenesis is reviewed comprehensively elsewhere [114–116].

During skeletal muscle development, embryonic myoblasts proliferate before undergoing terminal differentiation and fusing to form multinucleate myocytes (myofibres) (Fig. 15). A subset of progenitor cells remain undifferentiated and reside within the developed muscle as quiescent ‘satellite cells’. These multipotent cells, characterised by expression of Pax7, represent the primary muscle stem cell niche in adults and become activated following injury to facilitate muscle regeneration and maintain tissue homeostasis. The aforementioned myogenic processes are controlled by the expression and repression of muscle-specific transcription factors including myogenin, myf5, MyoD and MRF4 from the basic helix-loop-helix (bHLH) family of transcription factors. Upstream of the myogenic bHLH factors, the paired homeobox transcription factors Pax3 and Pax7 are genetic controllers of early embryonic myogenesis [116, 117].

Myostatin signalling, via the SMAD2/3 pathway, is known to exert its suppressive effect on muscle growth by modulating expression patterns of myogenic transcription factors and in doing so suppresses the proliferation and terminal differentiation of myoblasts. In vitro, myoblasts exposed to myostatin accumulate in G0/G1 phase as a result of upregulation of the cyclin-dependent kinase (Cdk) inhibitor P21 and concomitant decrease in Cdk2 levels [114, 118, 119]. In addition to suppression of proliferation, myostatin signalling also blocks myoblast differentiation as evidenced by the inability of cultured myoblasts to form myofibers in the presence of myostatin [114, 119, 120]. Myostatin-induced down-regulation of MyoD, myogenin and Myf5 expression is thought to prevent myoblasts from successfully exiting the cell-cycle, thus preventing the terminal differentiation of myoblasts required for muscle fibre formation. Reciprocally, knockdown of endogenous myostatin using antisense RNA promotes the proliferation and differentiation of C2C12 myoblasts, presumably as a result of liberated myogenin expression [121].

Myostatin signalling in postnatal muscle is thought to target the resident satellite cells, suppressing their activation and self-renewal. Again, the Cdk inhibitor P21 is implicated
Embryonic progenitors → Proliferating myoblasts → Myocytes → Multi-nucleate myotubes

Quiescent satellite cells → Activation following injury → Self-renewal of undifferentiated state

Figure 15: Simplified schematic of myogenesis, with associated key myogenic transcription factors shown for each stage of development. Embryonic muscle development involves the early determination of myoblasts from embryonic progenitors, which then differentiate and fuse to form multi-nucleate myotubes and subsequently muscle fibres. The transcription factors Pax3 and Pax7 control the early stages of myogenesis, while the basic helix-loop-helix factors Myf5, MyoD, myogenin and Mrf4 control differentiation of myoblasts into mature muscle fibres. A subset of undifferentiated progenitors reside within the developed muscle as quiescent Pax7 positive satellite cells. These cells retain a level of stemness which allows them to differentiate into mature myoblasts as required for muscle regeneration following injury.

In myostatin-induced satellite cell suppression, blocking the G1 to S phase transition, and maintaining their quiescence [122]. In line with this model, myostatin-null mice show higher proportions of activated satellite cells over wild-type, while the addition of exogenous myostatin causes a decrease in markers of satellite cell activation in cell culture [122, 123]. Furthermore, myostatin-null mice display enhanced muscle regeneration following injury, with improved inflammatory responses and reduced fibrosis when compared to wild-type mice [123, 124]. Together these data point towards a clear role for myostatin as a negative regulator of satellite cell activation in mature skeletal muscle. Speculatively, this negative regulation may be important for maintaining quiescence of the satellite cell pool and thus the regenerative potential of muscle tissue over time.

Inhibition of myostatin signalling by genetic ablation or expression of antagonistic molecules leads to dramatic muscle growth in mice and recapitulates the phenotypes of naturally occurring myostatin-null cattle, sheep, dogs and even humans. On a cellular level, increased
muscle mass occurs as a result of increased size (hypertrophy) or number (hyperplasia) of myofibres, or a combination of the two. In myostatin-deficient cattle breeds, hyperplasia seems to account for their increased muscle mass and presumably results from uninhibited myoblast proliferation during early development [114, 125]. Conversely, analyses of myostatin deficient mice show varying levels of hyperplasia and hypertrophy depending on the experimental model [76, 91, 96, 114, 126].

In contrast, partial myostatin suppression seems to result primarily in fibre hypertrophy. Transgenic mice expressing antagonists of myostatin signalling (including the prodomain, and dominant negative forms of the growth factor) display marked increase in fibre size, but not fibre number [96, 126]. It is possible that the expression of these myostatin antagonists occurs too late in development to significantly affect the myoblast proliferative phase and thus the number of mature myofibres remain unchanged [114]. In addition to controlling cellular proliferation and differentiation, there is also significant evidence to suggest myostatin negatively regulates myofibrillar protein synthesis [127, 128]. In the absence of myostatin, disinhibited protein synthesis appears to be the primary contributor to muscle fibre hypertrophy, at least in-part via suppression of the mammalian-target of rapamycin (mTor)/Akt signalling pathway [115, 129].

Although myostatin has traditionally been considered the primary regulator of muscle mass, other TGFβ superfamily ligands have been implicated in myogenic control. Administration of an ActRIIb-Fc fusion to myostatin-null mice resulted in hypertrophy to an even higher extent than observed for untreated myostatin-null mice [130]. This suggested that more than one ligand, signalling through the ActRIIb pathway, was acting as a negative regulator of muscle mass. Indeed, simultaneous inhibition of activin A and myostatin using specific neutralising antibodies recapitulated the hypertrophic phenotypes seen for pan-TGFβ superfamily inhibition [131]. Activin A is now emerging as a second negative regulator of muscle mass and was shown to exert an even more notable effect on musculature in primates than myostatin [131]. Additionally, GDF11 has been shown to negatively effect muscle growth via the same downstream signalling pathway (and target genes) as myostatin [41]. While these findings are unsurprising given GDF11 and myostatin are 90% identical in their mature domains, previous studies have suggested the contrary is true and that age-related decrease of GDF11 is responsible for muscle degeneration in mice [132–134].
In addition to the well-established roles of myostatin in myogenesis, low levels of expression are also detectable in adipose tissue [91]. Intriguingly, myostatin-null mice have considerably less adipose tissue than their wild-type counterparts and are less susceptible to high-fat diet induced obesity and insulin resistance [135, 136]. Furthermore, myostatin mRNA and protein levels are elevated in the skeletal muscle of obese and insulin resistant human patients, and tend to decrease in line with weight loss [137–139]. The effect of myostatin on adipogenesis and insulin resistance seem to be related to changes in the metabolic activity of skeletal muscle tissue which is an important regulator of glucose metabolism [140–142]. Suppression of myostatin signalling in wild-type and diabetic mice leads to improved glucose uptake and utilisation by muscle cells [143, 144]. As a result, pharmacological targeting of myostatin may hold potential for the treatment of metabolic diseases, including diabetes and obesity.

1.4.5 Myostatin as a therapeutic target

Since its discovery as a key negative regulator of muscle in 1997, myostatin has gained considerable interest as a potential therapeutic target. By blocking myostatin signalling, one could promote skeletal muscle growth which would be desirable in the context of muscle-wasting disorders including muscular dystrophy, cancer-related cachexia, and sarcopenia (age-related muscle loss). The real question is, would increasing muscle mass without correcting the underlying pathology improve clinical outcomes for sufferers of these diseases?

Pre-clinical studies to assess the potential of myostatin as a therapeutic target have typically relied on the use of animal models of muscular dystrophy. The most frequently employed is the mdx mouse, which due to a loss-of-function dystrophin mutation, exhibits muscle degeneration somewhat comparable to that experienced by patients with Duchenne muscular dystrophy (DMD)[145]. This model is far from perfect however, and there is considerable debate concerning its worth as an experimental surrogate for human DMD [146, 147]. Inhibition of myostatin signalling in mdx mice using neutralising antibodies, recombinant prodomain, or soluble ActRIIb receptor ligand traps, consistently improves overall muscle mass and absolute strength as a result of fibre hypertrophy [143, 148–151]. The effects on overall pathophysiology and specific muscle strength however are much less consistent. While some studies have reported improved specific force and decreased rates of muscle
degeneration, others fail to demonstrate improvement in functional metrics [143, 148–152]. Furthermore, in a mouse model of laminin-deficient muscular dystrophy, myostatin knockout entirely failed to alleviate pathological outcomes, despite increased muscle mass [153].

Taken together, these studies suggest that myostatin blockade can compensate for muscle degeneration by increasing muscle mass in some cases, but fails to correct the underlying pathological weaknesses which arise from mutations in unrelated proteins. The use of myostatin therapy as an adjunct to disease modifying treatments may provide a more effective means of combatting muscle wasting disorders. Indeed, simultaneous knockdown of ActRIIb and restoration of dystrophin function using exon-skipping methods was shown to significantly improve functional scores in mdx mice, over and above the use of either treatment alone [154].

A number of myostatin targeting therapies have progressed to clinical trials, to gauge their safety and efficacy in treating muscular atrophic disorders [155]. Clinical results have been disappointing overall, with several candidates failing to meet primary endpoints in phase II trials (Bimagrumab by Novartis, MYO-029 by Wyeth, PINTA-745 by Atara, PF-06252616 by Pfizer). Furthermore, the ActRIIb-Fc fusion ACE-031 (Acceleron) was withdrawn from phase II trials due to safety concerns.

The general lack of success so far is likely related to the inability of myostatin blockade to correct underlying pathophysiology, but also may stem from an inherent lack of specificity of these molecules. The aforementioned drug candidates are either antibodies or ligand-traps which target mature myostatin, and may cross-react with structurally related yet functionally distinct TGF\(\beta\) superfamily growth factors. This is particularly true for receptor derived ligand-traps which exhibit a natural promiscuity for multiple ligands. In some cases, simultaneous inhibition of multiple ligands may be desirable, for example dual myostatin/activin A suppression has been shown to be more effective over either ligand alone in promoting muscle growth in monkeys [131]. Unintentional inhibition of the wrong ligands however may have severe consequences. ACE-031 (Acceleron) trials were cancelled due to bleeding complications that possibly stemmed from unanticipated BMP9 blockade [155]. Undeterred, Acceleron are now actively pursuing multi-ligand blockade with their latest development ACE-083, a locally-acting biologic therapeutic based on follistatin which in principle would inhibit both myostatin and activin A in skeletal muscle. Following promising results from
phase I clinical trials, ACE-083 has progressed to phase II in patients with facioscapulohumeral muscular dystrophy and Charcot-Marie-Tooth disease [156].

An emerging strategy for pharmacological blockade of myostatin is to target the latent precursor forms rather than the mature bioactive ligand. One could effectively suppress myostatin signalling by stabilising the latent prodomain bound complex, or inhibiting proteolytic processing of the precursor. This would confer a significant specificity advantage as the prodromains of TGF\(\beta\) superfamily growth factors are less conserved than the bioactive mature domains. Furthermore, uncleaved pro-myostatin represents the primary storage form of myostatin in skeletal muscle and likely persists for longer than the mature ligand which is present only transiently following activation [64]. The Cambridge MA based biotech company Scholar Rock is pursuing such an approach with the development of SRK-015, a monoclonal antibody which blocks proteolytic activation of pro-myostatin with high specificity [104]. Clinical assessment of SRK-015 for the treatment of spinal muscular atrophy (SMA) is currently underway (http://www.scholarrock.com/pipeline/srk-015-for-sma/intro accessed June, 2018).

Understanding the prodomain interactions which confer latency to the pro-myostatin complex will be essential for advancing these efforts. To this end, I set out to solve the crystal structure of the pro-myostatin precursor and analyse the molecular details of its latency and activation.
1.5 Aims and scope of research

The primary objective of this study was to increase our understanding of pro-myostatin latency and activation. An enhanced appreciation of the molecular details which underpin these processes is not only desirable from an inquisitive academic viewpoint, but may also assist with the development of next generation myostatin targeting therapeutics. The core aims of this doctoral thesis are as follows:

1) Develop a protocol for the production of recombinant human pro-myostatin

A ready supply of the pro-myostatin precursor will be required for downstream structural and functional analysis. The resulting protein must be bioactive, structurally homogenous and available in high yield. To this end, I will employ a bacterial expression system and in vitro refolding strategy to produce the recombinant protein in its native dimeric form. A number of expression constructs and refolding buffers will be screened in order to identify the best conditions for production of a high quality surrogate of native pro-myostatin.

In addition to wild-type pro-myostatin, I will generate an engineered construct in which the native furin protease cleavage site is replaced with a HRV-3C ‘PreScission’ protease site. This should allow us to efficiently cleave the prodomain in vitro, with a cost-effective protease that is readily available in our lab.

2) Functional characterisation of myostatin precursor forms

Once the proteins required for this study have been produced, the bioactivity of the uncleaved precursor, and the HRV-3C cleaved pro-myostatin complex will be assessed using mammalian cell-based luciferase reporter assays. The extent to which the pro-complex persists after prodomain cleavage, and the degree of latency conferred by the prodomain will also be investigated.

Furthermore, the mature growth factor will be purified from the pro-myostatin complex using reversed-phase chromatography, and its bioactivity assayed. This will serve to validate the quality of our bacterially expressed/refolded protein. The bioactivity values can be compared to published activity data for mature myostatin expressed in mammalian systems.

I will also seek to characterise the interactions of the pro-myostatin complex with the known antagonist follistatin, and other binding partners if possible. While follistatin is known to bind the free growth factor, its ability to displace the prodomain or form a ternary complex
with the GF and prodromains is unknown. Biophysical methods will be used to address these questions.

3) Pro-myostatin structure determination

To investigate the molecular mechanism by which pro-myostatin remains latent I aim to crystallise pro-myostatin (the uncleaved precursor and/or the cleaved complex) and determine its molecular structure using X-ray crystallography. Pro-TGFβ family members are notoriously difficult to crystallise due to the inherent flexibility of these proteins and thus it is envisaged that considerable protein engineering will be required to encourage crystallisation. Access to a high resolution structure of pro-myostatin will grant us invaluable insight into the interactions between pro and mature domains, and should expose the molecular basis for latency of the pro-complex. Additionally, structural information will allow us to rationalise the effect of naturally occurring myostatin polymorphisms on the function of the protein and may provide a novel basis for the development of next generation myostatin inhibitors.
Production of recombinant human pro-myostatin
The initial objective of this study was to develop a robust protocol for the production of recombinant human pro-myostatin, for structural and functional analysis. Typically, disulfide rich TGFβ superfamily growth factors are produced in eukaryotic expression systems which are capable of facilitating correct disulfide bond formation and post-translational modification. In this instance however the need for large quantities of structurally homogenous protein encouraged us to utilise a bacterial expression system. This system provides a number of additional advantages over mammalian protein expression. Firstly, because bacteria do not express furin protease we can be sure that all expressed protein will be of the unprocessed form. A bacterial system also permits rapid generation and expression of variant constructs and easily permits introduction of labelled amino acids (i.e selenomethionine for phasing X-ray data).

The pro-myostatin sequence contains a single N-linked glycan site (Asn71-Ile72-Ser73) in the putative latency lasso region. By using a bacterial expression system we accept that our recombinant pro-myostatin will lack native glycosylation and that this may affect the protein bioactivity. A previous study, however, has shown that the myostatin prodomain retains its native inhibitory properties even in the absence of native glycosylation patterns [78].

The cytoplasm of *Escherichia coli* (*E.coli*) is a reducing environment which does not favour the formation of disulfide bonds and as a result, disulfide-rich recombinant proteins are often produced as insoluble aggregates (inclusion bodies). For this reason, we developed a method to solubilise and refold the protein to its native form. This chapter describes the expression, refolding and purification of recombinant human pro-myostatin.

### 2.1 Expression construct design

The DNA sequence encoding full length human pro-myostatin, including the entire N-terminal prodomain and C-terminal mature domain but lacking the signal peptide (residues 19-375, Uniprot 014793), was amplified by polymerase chain reaction (PCR) from I.M.A.G.E. consortium cDNA (clone 30915216) with primers incorporating 5’ and 3’ restriction sites for cloning into pHAT/pBAT vectors (Fig. 16A).

Three expression constructs were made for the initial round of expression testing; pBAT4-MSTN.001 (untagged), pHAT2-MSTN.001 (N-term 6xHis-tag with short linker) and pHAT2-
MSTN.002 (N-term 6xHis-tag with 11 residue linker) as shown in figure 16B.

Figure 16: First generation pro-myostatin construct design. A. pBAT4 and pHAT2 expression vectors employed in this study. For vector sequences and maps see http://hyvonen.bioc.cam.ac.uk/pBAT-vectors. B. First generation pro-myostatin constructs used for expression and refolding trials.

### 2.2 Protein expression and refolding

Sequence verified pBAT or pHAT2 constructs (Appendix 1) were transformed into chemically-competent E.coli BL21 (DE3) cells also bearing a vector encoding tRNA gene for rare arginine codons AGA/AGG (pUBS520)[157]. Small scale expression tests showed good overexpression of all three constructs and, as expected, the protein was present exclusively in the insoluble pellet under all expression conditions (Fig. 17).

There are several strains of E.coli which have been modified to promote soluble expression
of disulfide-rich proteins. These include strains with mutations in the reductive thioredoxin (Trx) and glutathione reductase (Gor) pathways, as well as those which co-express disulfide isomerases and sulfhydryl oxidases to facilitate disulfide exchange reactions for formation of native disulfide bond patterns [158–161]. Additionally, direction of translated polypeptides to the oxidising bacterial periplasm is a well established approach for production of disulfide bonded eukaryotic proteins [162].

These engineered strains have been shown to be effective in many cases for producing soluble disulfide-rich proteins in E.coli, including full length antibodies [163]. Our lab however has had limited success with these strains when trying to express TGFβ superfamily proteins which contain the structurally important and topologically complex cystine-knot. In most instances the proteins are present exclusively in the insoluble pellet when expression is attempted in these modified strains (data not shown).

*In vitro* refolding methods are commonly employed for the production of disulfide-rich proteins which are unable to be expressed as soluble protein in E.coli. A number of *in vitro* refolding strategies exist, including either rapid dilution or dialysis of the solubilised (but still denatured) protein into non-denaturing buffers which allow the protein to refold to its native form. Alternatively refolding can be done while the denatured protein is immobilised...
on an affinity-matrix, by slowly changing the composition of the buffer to reduce the concentration of the denaturant. In this study I employed the rapid-dilution approach as this method is well established in the Hyvönen lab and has been shown previously to work well for proteins of the TGFβ superfamily \[72, 164\].

The correct folding of proteins containing structural disulfide bonds is dependent on a series of disulfide shuffling reactions which allow the native arrangement of disulfides to form (Fig. 18). These reactions are rate-limiting in terms of protein folding kinetics, but are accelerated by enzyme catalysis \textit{in vivo} \[165\]. By supplementing our refolding buffer with a thiol-containing redox pair, we can promote thiol-disulfide exchange \textit{in vitro}. Assuming disulfide bond rearrangement is under thermodynamic control, the most stable (and therefore likely the native) arrangement of disulfide bonds is expected to accumulate over time.

![Diagram of thiol-disulfide exchange reactions](image)

**Figure 18:** \textit{In vitro} thiol-disulfide exchange reactions. Disulfide shuffling reactions are critical for the correct folding of disulfide-containing proteins. A. Under alkaline conditions the deprotonated thiolate anion predominates and is a potent nucleophile. B. In this \textit{intermolecular} step, a free protein-thiolate attacks the sulfur atom of a reagent disulfide bond, displacing one thiolate and generating a mixed-disulfide species. C. A second protein-thiolate attacks the mixed-disulfide bond (this time \textit{intramolecularly}), displacing the reagent thiolate and forming a new protein disulfide bond.

As TGFβ superfamily proteins are known to form functional disulfide-linked dimers, this provides a simple metric for identifying potential refolding hits. The accumulation of covalent dimers can be observed by non-reducing SDS-PAGE and implies (but does not prove)
formation of the native protein conformation.

Expression of first-generation constructs was scaled up to one litre culture volume and the inclusion bodies prepared as per Material and Methods section 8.2. Solubilised protein at a stock concentration of 1 mg/mL was then rapidly diluted 1:10 into a custom refolding buffer screen (Table 1), supplemented with a cysteine/cystine (CySH/CySS) redox pair to facilitate thiol-disulfide exchange reactions.

Table 1: In vitro refolding buffer screen.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Buffer</th>
<th>pH</th>
<th>Additive 1</th>
<th>Additive 2</th>
<th>Additive 3</th>
<th>Additive 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1M Na Phosphate</td>
<td>7.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.1M Tris</td>
<td>8.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.1M Tris</td>
<td>8.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.1M Tris</td>
<td>9.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.1M CHES</td>
<td>9.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.1M CAPS</td>
<td>10.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.1M Na Phosphate</td>
<td>7.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.1M Tris</td>
<td>8.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.1M Tris</td>
<td>8.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.1M Tris</td>
<td>9.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.1M CHES</td>
<td>9.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.1M CAPS</td>
<td>10.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.1M Na Phosphate</td>
<td>7.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td>1 M PPS</td>
</tr>
<tr>
<td>14</td>
<td>0.1M Tris</td>
<td>8.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td>1 M PPS</td>
</tr>
<tr>
<td>15</td>
<td>0.1M Tris</td>
<td>8.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td>1 M PPS</td>
</tr>
<tr>
<td>16</td>
<td>0.1M Tris</td>
<td>9.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td>1 M PPS</td>
</tr>
<tr>
<td>17</td>
<td>0.1M CHES</td>
<td>9.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td>1 M PPS</td>
</tr>
<tr>
<td>18</td>
<td>0.1M CAPS</td>
<td>10.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td></td>
<td>1 M PPS</td>
</tr>
<tr>
<td>19</td>
<td>0.1M Na Phosphate</td>
<td>7.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td>1 M PPS</td>
</tr>
<tr>
<td>20</td>
<td>0.1M Tris</td>
<td>8.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td>1 M PPS</td>
</tr>
<tr>
<td>21</td>
<td>0.1M Tris</td>
<td>8.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td>1 M PPS</td>
</tr>
<tr>
<td>22</td>
<td>0.1M Tris</td>
<td>9.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td>1 M PPS</td>
</tr>
<tr>
<td>23</td>
<td>0.1M CHES</td>
<td>9.5</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td>1 M PPS</td>
</tr>
<tr>
<td>24</td>
<td>0.1M CAPS</td>
<td>10.0</td>
<td>2/0.2 mM CySH/CySS</td>
<td>0.5 mM EDTA</td>
<td>0.5 M NaCl</td>
<td>1 M PPS</td>
</tr>
</tbody>
</table>

SDS-PAGE analysis of refolding trials was completed after 7 days (Fig. 19). Bands corresponding to putative disulfide-linked dimers were present in several conditions for all three constructs. Interestingly, the bands corresponding to dimeric pro-myostatin ran on SDS-PAGE with an apparent molecular weight of >100 kDa, much higher than expected for this protein of 85 kDa.

Refolding condition #16 (50 mM Tris pH 9.0, 1 M PPS, 0.5 mM EDTA, 2 mM cysteine, 0.2 mM cystine) was chosen for further analysis as this condition appeared to yield a good

---

3-(1-Pyridinio)-1-propanesulfonate (PPS) is a zwitterionic non-detergent sulfobetaine, a class of compound used to aid protein solubility during refolding, and thought to outcompete the abortive hydrophobic interactions that drive aggregation [166].
Figure 19: Non-reducing SDS-PAGE analysis of small scale refolding trials after 7 days. The three constructs of pro-myostatin were screened against 24 refolding buffers to identify conditions which facilitate formation of the native disulfide-linked dimer. Bands corresponding to the dimeric species (arrowheads) were observed for all three constructs in many of the tested conditions. Condition 16 was selected for subsequent large scale refolding and purification.
production of dimeric species. Furthermore, this condition does not contain salt, making it amenable to ion-exchange chromatography for the initial purification step. Preliminary refolding trials showed that of the three initial constructs, pHAT2-MSTN.002 (with 6xHis-tag and long linker) appeared to refold with higher efficiency than the short linker and untagged constructs (Fig. 19). This construct yielded the highest proportion of dimeric species and as such was taken forward for large scale refolding and purification.

![SDS-PAGE analysis of pHAT2-MSTN.002 refolding](image)

Figure 20: Non-reducing SDS-PAGE analysis of pHAT2-MSTN.002 refolding (condition #16) over the course of 7 days, showing accumulation of disulfide-linked dimer.

SDS-PAGE analysis of the pro-myostatin refolding products under reducing conditions led to disappearance of the putative dimer band and appearance of a product with an apparent molecular weight half that of the dimeric species, confirming the presence of a disulfide-linked dimer under non-reducing conditions (not shown). Refolding samples from condition #16 taken at intervals over the course of 7 days show that dimerisation of pHAT2-MSTN.002 is a slow process, and dimeric species are not visible by SDS-PAGE until at least 8 hours after addition of protein to the refolding buffer (Fig. 20). Dimeric products continue to accumulate up until day 7, with a concomitant decrease in monomeric species.

### 2.3 Purification of recombinant human pro-myostatin

One litre batches of pro-myostatin (pHAT2-MSTN.002) refolding solution were loaded onto a Source 15Q anion exchange column (GE Healthcare). This step provides a means of capturing and concentrating the protein from the refolding solution, and simultaneously removing the unwanted buffer components. The bound protein was eluted with increasing
NaCl concentration, and typically eluted as multiple broad peaks (Fig. 21A). Regardless, the disulfide-linked dimer was greatly enriched over contaminating species in the peak eluting at ≈ 300 mM NaCl (Fig. 21A).

Figure 21: Purification of unprocessed pHAT2-MSTN.002. A. Chromatogram of anion-exchange capture of pHAT2-MSTN.002 from refolding solution, and non-reduced SDS-PAGE analysis of peak fractions showing enriched disulfide linked dimer (pink arrowhead). B. Pooled fractions from A were further purified by size-exclusion chromatography. The fractions pooled at each stage are indicated by pink fraction numbers.

Unlike the mature dimeric growth factors of the TGFβ superfamily which are extremely hydrophobic and typically cannot be purified using standard chromatographic methods, the pro-forms have improved solubility and tend to behave as normal globular proteins. As such I was able to further purify the captured protein by size-exclusion chromatography (SEC), in which the protein eluted as a single symmetrical peak (Fig. 21B). The peak contained majority disulfide-linked species, but also traces of contaminating lower molecular weight species.

Mass spectrometry of the purified product confirmed the expected molecular mass (data not shown). There was some batch-to-batch variability in yield and purity, but typically one could expect to obtain 2-5 mg of purified protein per litre of refolding solution.
It should be noted that while this protein construct contains an N-terminal 6xHis-tag, it was not used for purification. Affinity purification using the 6xHis-tag is not ideal in this system as it cannot separate the desired dimeric species from the unwanted misfolded and aggregated species which also contain a 6xHis-tag. Instead, the tag was retained because of its apparent beneficial effect on refolding efficiency, and because it provides a useful handle for detection and immobilisation as required for downstream applications.

### 2.4 Production of HRV-3C (PreScission) cleavable pro-myostatin

In nature, activation of pro-myostatin first requires proteolytic cleavage of the prodomain by furin-like pro-protein convertases. In my hands, efficiency of furin cleavage in vitro was relatively poor and the recombinant furin required for this purpose is expensive. For these reasons we chose to engineer a variant of pro-myostatin in which the native furin site is replaced with an HRV-3C (PreScission) protease site (Fig. 22). This should allow efficient proteolysis of the prodomain in vitro, using a protease we produce routinely in-house. This approach had previously been pioneered by Dr Xuelu Wang in our lab, for pro-activin A studies [72].

![Figure 22: Engineered HRV-3C cleavable pro-myostatin. The native furin motif (RSRR) has been substituted for an HRV-3C protease site.](image)

Throughout the course of construct design and refolding trials, it became apparent that alteration of the furin cleavage site sequence drastically reduced the refolding efficiency (data not shown). Initially I hypothesised that replacing the basic residues of the furin motif with mostly hydrophobic amino acids was the source of this problem. Alanine scanning of the furin site residues, however, showed that loss of basic residues was not sufficient to disrupt refolding (data not shown). Eventually it was realised that altering the length of the linker between the prodomain and mature domain was likely responsible for perturbed refolding. By substituting the furin site in such a way as to retain the wild-type linker length, I was able to generate an HRV-3C cleavable construct (pHAT2-MSTN-3C) which refolded under...
the same conditions as our wild-type pro-myostatin construct.

Curiously, pHAT2-MSTN-3C did not bind efficiently to the Source 15Q anion exchange resin under the original purification conditions. Fortunately, switching to an alternate anion exchange resin (HiTrap Q, GE Healthcare) allowed successful capture from the refolding solution (Fig. 23A). In most instances, traces of putative monomeric species co-eluted with the disulfide-linked dimer. It is possible that a fraction of the protein forms non-covalent dimers which have similar physio-chemical properties to the native disulfide-linked dimer and thus are not well separated by chromatographic methods. As for pHAT2-MSTN.002, the uncleaved protein could be purified further by size-exclusion chromatography (Fig. 23B).

In order to generate the non-covalent pro-complex, purified pHAT2-MSTN-3C was incubated overnight with HRV-3C protease. HRV-3C is a cysteine protease and is typically used under reducing conditions which maintains reactivity of the catalytic thiol group. The reducing agent present in the protease storage buffer was removed by buffer exchange prior
to incubation with pro-myostatin, which contains many structural disulfide bonds that are susceptible to reduction. Even in the absence of reducing agent, HRV-3C protease retained sufficient activity and the majority of the starting material was processed following incubation at 4°C overnight.

Following HRV-3C treatment, in addition to SDS-PAGE bands corresponding to the expected 31 kDa monomeric prodomain and 24 kDa dimeric mature growth factor (which runs as a diffuse, blurry band), there was another cleavage product which ran as a 50-60 kDa product (Fig. 24). This may represent a partially cleaved product, which would arise if only a single prodomain is cleaved from the dimeric growth factor (55 kDa). A partially cleaved ‘semi-latent’ product has been previously reported when furin was used to process pro-myostatin in vitro [167]. The presence of this intermediate cleavage product was stubborn in our preparation, and could not be removed even after the addition of a molar excess of HRV-3C protease.

Furthermore, cleavage of the pro-myostatin precursor revealed an additional high molecular weight contaminant, initially concealed by the uncleaved product. This contaminating species could not be separated chromatographically.

Figure 24: SDS-PAGE analysis of unprocessed and HRV-3C processed pro-myostatin. A. Purified unprocessed pro-myostatin. B. Purified HRV-3C processed pro-myostatin complex. The pro-myostatin complex preparation contains a number of contaminating species including a putative partially-cleaved product, and a high MW species. Under non-reducing conditions, the mature GF dimer migrates as a diffuse, blurry band. Each species is depicted in cartoon representation with mature GF in orange and prodomain in grey.
2.5 Analysis of the HRV-3C cleaved pro-myostatin complex

Size-exclusion chromatography of HRV-3C cleaved pro-myostatin demonstrated formation of a stable non-covalent complex between the cleaved prodromains and mature growth factor. These components co-eluted as a single symmetrical peak with the same retention volume as the uncleaved pro-myostatin dimer (Fig. 25). The aforementioned partially cleaved species also eluted at the same volume, suggesting non-covalent interaction of a single free prodomain, with the singly cleaved precursor variant.

To further validate complex formation, size-exclusion chromatography with multi-angle light scattering (SEC-MALS) was employed to measure the mass of both unprocessed and HRV-3C cleaved pro-myostatin. MALS is a technique used to measure the absolute molar mass and size of macromolecules in solution. This method relies on accurate measurement of the intensity and angular dependence of laser light scattered by the sample. The overall intensity of scattered light is proportional to the concentration and molar mass of the macromolecule, while the angular dependence of scattering (anisotropy) is related to the particle size (radius of gyration, \( R_g \)). When coupled to a size-exclusion chromatography system, absolute measurements of molar mass and particle size can be made continuously as size-separated fractions elute from the column. This prevents contaminating aggregates from masking the signal of the smaller particles of interest. Furthermore, by measuring the molar mass and particle size across an eluting peak, the polydispersity/homogeneity of the sample can be determined.

Samples of the uncleaved pro-myostatin precursor and the HRV-3C cleaved pro-myostatin complex were analysed by SEC-MALS, and as expected, they eluted as symmetrical peaks with almost identical retention volumes (Fig. 25). The molar mass of unprocessed pro-myostatin was calculated as 84.5 kDa (cf. expected 85.4 kDa) confirming the expected dimeric state under native conditions. The HRV-3C cleaved pro-myostatin complex had a calculated molar mass of 83.4 kDa, indicating stable complex formation between the dimeric mature growth factor and the prodromains following proteolysis. These results show that the prodromains of our HRV-3C cleavable pro-myostatin variant can be cleaved \textit{in vitro} and that they remain associated with the dimeric growth factor as a stable non-covalent complex following cleavage.
2.6 Conclusion

This chapter describes the development of a protocol for the production of recombinant human pro-myostatin. Using an \textit{E. coli} expression system and \textit{in vitro} refolding strategy, constructs of both wild-type pro-myostatin and an HRV-3C cleavable engineered variant were successfully expressed and refolded to yield the desired disulfide-linked dimeric proteins. Interestingly, the non-native plasmid derived residues at the protein N-terminus (6xHis-tag and linker) appear to contribute to the refolding efficiency of these constructs. The refolded proteins could be purified in milligram quantities using combinations of anion exchange and size-exclusion chromatography, and displayed good solubility and stability under near physiological buffer conditions. The HRV-3C cleavable variant was successfully processed by HRV-3C protease to yield a stable non-covalent complex of the mature growth factor and its associated prodomains. Formation of the non-covalent pro-myostatin complex was confirmed by SEC-MALS, with the measured molecular mass consistent with that expected for two prodomains and the dimeric mature growth factor.
Functional characterisation of pro-myostatin
Having established a protocol for the production of pro-myostatin, I then sought to evaluate the bioactivity of the protein in its different forms. Analysis of the bioactivity would allow me to 1) assess the quality of the refolded protein and 2) investigate the extent to which the prodomain confers latency to the non-covalent pro-myostatin complex. While there is already data demonstrating that the furin cleaved pro-myostatin complex is latent (inactive) until secondary cleavage of the prodomains by BMP1/TLD family proteases, I deemed it necessary to confirm this with my bacterially expressed protein. This is particularly important given the native furin site has been substituted for an HRV-3C protease site in the engineered pro-myostatin variant.

3.1 Isolation of the mature myostatin growth factor

As a critical quality control step, I sought to isolate the mature myostatin GF and compare its signalling potency to published bioactivity values. Comparison of bioactivity with myostatin produced in mammalian expression systems would allow validation of my bacterially expressed and refolded product.

In order to test the bioactivity of the mature GF it was first purified from the HRV-3C processed complex produced in the previous chapter. Because the mature ligand is heavily disulfide linked and therefore structurally very stable, it is possible to purify it away from its associated prodomains under harsh conditions using reversed-phase chromatography (RPC) without compromising the structural integrity of the ligand. Unlike most chromatographic methods which use a hydrophilic stationary phase, RPC utilises a hydrophobic stationary phase typically consisting of hydrocarbon chains covalently linked to a silica matrix. This allows hydrophobic molecules to adsorb to the stationary phase, which can then be eluted with increasing concentrations of an organic solvent (i.e. acetonitrile, ACN) which reduces the polarity of the mobile phase. ACN is a protein denaturant at high concentrations and so it is expected the HRV-3C cleaved complex will dissociate during RPC, allowing separation of the prodomains and the mature growth factor. While the mature growth factor is highly stable, the prodomains are structurally much less robust and so will likely be denatured following their passage through the RPC column. Once separated, fractions containing the mature ligand can be dried to remove volatile ACN and then resuspended immediately before use. This method is routinely used for purification of mature TFG/β superfamily
ligands, and can be done without compromising their bioactivity [72, 76, 168].

Batches of the purified HRV-3C processed pro-myostatin complex were loaded onto an ACE C8 300 4.6x250 mm RPC column and eluted using a gradient of 10-90% ACN. The mature growth factor eluted first as a broad peak, followed by the prodomain peak (Fig. 26). Fractions containing the mature GF (25 kDa) were pooled and dried to powder by centrifugal evaporation for storage at -80°C.

![Figure 26: Isolation of mature myostatin GF from the HRV-3C processed complex. A. Chromatogram showing separation of mature GF and prodomains by reversed-phase chromatography. B. Non-reducing and reducing SDS-PAGE analysis of purified mature GF.](image)

### 3.2 Quantification of myostatin bioactivity

To measure the bioactivity of purified mature myostatin, a cell-based dual luciferase assay using cultured human embryonic kidney (HEK293) cells was employed. The specific strain used in our lab, HEK293T, is a derivative of HEK293 which stably expresses the large T antigen from Simian vacuolating virus 40 (SV40), which allows competent replication of vectors containing the SV40 origin of replication [169].

In this assay, HEK293T cells are cultured to confluence in 96-well plates and then simultaneously transfected with two vectors; pGL3-CAGA and pRL-SV40. The first of these encodes the firefly luciferase reporter gene, under control of a SMAD2/3 (and therefore myostatin) responsive CAGA element containing promoter [170]. In this way, myostatin signalling is directly linked to the expression of firefly luciferase. The second plasmid, pRL-SV40, encodes
Sea Pansy (*Renilla reniformis*) luciferase which is constitutively expressed within the cell following transfection. The expression level of *Renilla* luciferase is independent of myostatin signalling and provides an internal control for normalisation of the firefly luciferase activity.

Twenty-four hours post-transfection, protein is added to the cells at a number of different concentrations and incubated overnight allowing expression of firefly luciferase. The cells are then lysed and the firefly luciferase substrate added to the lysate. Luciferase present in the sample catalyses the conversion of the substrate to a luminescent product which can be quantified with a luminometer. Providing the substrate is not limiting, the luminescence of the product is proportional to the amount of luciferase present in the sample. After the firefly luciferase measurement is made, the reaction is quenched and the *Renilla* luciferase substrate is then added to the same sample, again yielding a quantifiable luminescent product. Instead of using the raw firefly luciferase luminescence values, the ratio of firefly/*Renilla* luminescence is taken for each sample, which allows normalisation for any variation in the measurement not attributable to myostatin signalling (cell number/viability etc). The full method is detailed in Material and Methods section 8.8.

### 3.2.1 Mature myostatin induces SMAD2/3 signalling with high potency

Purified mature myostatin was serially titrated (1 pM to 50 nM) into cell culture medium, and incubated with cultured HEK293T cells bearing the luciferase reporter genes described above. Following 24 hours exposure to varying concentrations of mature myostatin, cells were lysed and the luciferase activity quantified. Each protein concentration was tested in triplicate and furthermore the entire assay was run six times independently over the course of this project using different batches of purified mature myostatin, and luciferase substrates from different sources.

Plotting the data obtained from these experiments yielded a clearly sigmoidal dose-response with defined lower and upper maxima. The data from each assay were fit to a non-linear regression dose-response model (Fig. 27A) in order to determine the half maximal effective concentration (EC$_{50}$) of the protein, that is, the concentration at which half of the maximal activity is achieved. From the calculated dose-response curves, an EC$_{50}$ value for each assay was calculated independently, with these values ranging from 0.06 – 0.12 nM (Table 2). Global-fitting analysis of all datasets to a single model gave a global EC$_{50}$ value of 0.09 ±
0.01 nM (Fig. 27B).

Figure 27: Mature myostatin signalling in HEK293T cells. A. Six independent bioactivity datasets individually fit to dose-response model for EC$_{50}$ calculation. B. Global fit analysis of all datasets. Data shown are the means of triplicate measurements ± SEM.

Table 2: Mature myostatin bioactivity in HEK293T cells

<table>
<thead>
<tr>
<th>Assay no.</th>
<th>EC$_{50}$ (nM)</th>
<th>EC$_{50}$ (nM) 95 % CI*</th>
<th>Log [EC$_{50}$] (nM) ± SEM**</th>
<th>Hill slope</th>
<th>Hill slope 95% CI*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.08</td>
<td>0.05 to 0.13</td>
<td>-1.08 ± 0.10</td>
<td>1.05</td>
<td>0.65 to 1.72</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>0.06 to 0.08</td>
<td>-1.19 ± 0.03</td>
<td>1.41</td>
<td>1.17 to 1.70</td>
</tr>
<tr>
<td>3</td>
<td>0.12</td>
<td>0.09 to 0.16</td>
<td>-0.92 ± 0.06</td>
<td>1.71</td>
<td>1.16 to 2.60</td>
</tr>
<tr>
<td>4</td>
<td>0.12</td>
<td>0.11 to 0.14</td>
<td>-0.91 ± 0.02</td>
<td>1.41</td>
<td>1.24 to 1.60</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>0.09 to 0.12</td>
<td>-0.99 ± 0.02</td>
<td>1.72</td>
<td>1.40 to 2.20</td>
</tr>
<tr>
<td>6</td>
<td>0.08</td>
<td>0.07 to 0.09</td>
<td>-1.11 ± 0.03</td>
<td>1.55</td>
<td>1.21 to 2.02</td>
</tr>
<tr>
<td>Global fit</td>
<td>0.09</td>
<td>0.08 to 0.10</td>
<td>-1.04 ± 0.02</td>
<td>1.36</td>
<td>1.19 to 1.55</td>
</tr>
</tbody>
</table>

* CI Confidence interval  
** SEM Standard error of the mean

The EC$_{50}$ values determined here for refolded and isolated mature myostatin are very similar to that reported by Le et al 2018 (0.08 nM) who, using a similar assay, tested the bioactivity of commercially available mature myostatin (R&D systems) produced in a mammalian expression system (Table 3)[3]. Furthermore, our protein appears to signal with even higher potency than that documented previously in the literature, including the values reported by Walker et al 2017 who tested the bioactivity of mature myostatin obtained from a number of commercial retailers [42, 171].
Table 3: Published signalling parameters for mature myostatin

<table>
<thead>
<tr>
<th>Assay type</th>
<th>EC₅₀ (nM)</th>
<th>EC₅₀ (nM) 95 % CI*</th>
<th>Protein source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luciferase (HEK293)</td>
<td>0.48</td>
<td>0.46 to 0.50</td>
<td>Acceleron Pharma</td>
<td>Walker et al 2017 [42]</td>
</tr>
<tr>
<td>Luciferase (HEK293)</td>
<td>0.39</td>
<td>N/S</td>
<td>R&amp;D systems</td>
<td>Walker et al 2017 [42]</td>
</tr>
<tr>
<td>Luciferase (HEK293)</td>
<td>0.66</td>
<td>N/S</td>
<td>Peprotech</td>
<td>Walker et al 2017 [42]</td>
</tr>
<tr>
<td>Luciferase (HEK293)</td>
<td>0.72</td>
<td>N/S</td>
<td>N/S</td>
<td>Walker et al 2018 [2]</td>
</tr>
<tr>
<td>Luciferase (HEK293T)</td>
<td>0.078</td>
<td>N/S</td>
<td>R&amp;D systems</td>
<td>Le et al 2018 [3]</td>
</tr>
<tr>
<td>Luciferase (A204)</td>
<td>1.37</td>
<td>N/S</td>
<td>R&amp;D systems</td>
<td>Latres et al 2015 [171]</td>
</tr>
<tr>
<td>Luciferase (HEP2G)</td>
<td>5.4</td>
<td>4.90 to 6.10</td>
<td>N/S</td>
<td>Walker et al 2017 [42]</td>
</tr>
<tr>
<td>FSH stimulation (LβT2)</td>
<td>0.08</td>
<td>0.07 to 0.10</td>
<td>N/S</td>
<td>Walker et al 2017 [42]</td>
</tr>
</tbody>
</table>

* CI Confidence interval
N/S - Not shown

3.2.2 Myostatin signalling is inhibited by follistatin

To further validate the bioactivity of our recombinant mature myostatin, I tested the ability of the natural antagonist follistatin to inhibit myostatin signalling. Antagonism of myostatin by follistatin is well documented in the literature, and the structural basis for inhibition has been studied in painstaking detail [100, 107].

Using the aforementioned dual-luciferase assay, the ability of follistatin-288 (the 288 amino acid paralog) to inhibit mature myostatin signalling in HEK293T cells was assessed. Increasing concentrations of recombinant follistatin were titrated with mature myostatin (0.25 nM), and the myostatin:follistatin mixtures then applied to cells. As expected, follistatin inhibited myostatin signalling in a dose-dependent manner (Fig. 28). Non-linear regression fitting of a dose-response inhibition model (variable slope, 4 parameters) showed a good fit to the data (R-squared: 0.98), and an IC₅₀ value of 0.8 nM (Table 4).

The potent bioactivity of mature myostatin and its susceptibility to inhibition by follistatin provides good evidence that the recombinant protein produced in this study is correctly folded, and represents a good surrogate for the native protein. In section 3.3, I extend my investigation of follistatin, by examining its interactions with the pro-myostatin complex.
3.2.3 The pro-myostatin complex is weakly bioactive

It is well established that following furin cleavage of pro-myostatin, the dimeric growth factor and cleaved prodomains remain non-covalently associated as an inactive complex. This ‘latent’ complex is unable to signal until BMP1/TLD cleavage triggers dissociation of the prodomains and exposes the receptor epitopes [76, 78, 80].

Having validated our refolded protein by testing bioactivity of the mature ligand, I then attempted to measure the bioactivity (if any) of the precursor forms using the same dual-luciferase assay as described in the previous paragraphs. It should be noted from the outset that the HRV-3C cleaved complex preparation contained some contaminating species and therefore the reported concentrations (and resultant signalling parameters) are approximate only.

Table 4: Inhibition of mature myostatin signalling by follistatin-288

<table>
<thead>
<tr>
<th></th>
<th>IC₅₀ (nM)</th>
<th>IC₅₀ (nM) 95% CI*</th>
<th>Log [IC₅₀] (nM) ± SEM**</th>
<th>Hill slope</th>
<th>Hill slope 95% CI*</th>
</tr>
</thead>
<tbody>
<tr>
<td>FST2-288</td>
<td>0.80</td>
<td>0.73 to 0.88</td>
<td>-0.10 ± 0.02</td>
<td>-4.92</td>
<td>-8.28 to -3.70</td>
</tr>
</tbody>
</table>

* CI Confidence interval  
** SEM Standard error of the mean
Figure 29: Bioactivity of myostatin precursor forms in HEK293T cells. Unprocessed pro-myostatin (black circles) shows no signalling activity at the highest concentrations of protein tested. HRV-3C processed pro-myostatin complex (blue lines, 3 independent assays) shows low level bioactivity compared with the isolated mature growth factor (green lines, 3 independent assays). Data shown are the means of triplicate measurements ± SEM.

Table 5: Bioactivity of myostatin precursor forms in HEK293T cells

<table>
<thead>
<tr>
<th>Assay no.</th>
<th>Assay no.</th>
<th>EC50 (nM)</th>
<th>EC50 (nM) 95 % CI*</th>
<th>Log [EC50] (nM) ± SEM**</th>
<th>Hill slope</th>
<th>Hill slope 95% CI*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mature GF</td>
<td>1</td>
<td>0.10</td>
<td>0.09 to 0.12</td>
<td>-0.99 ± 0.02</td>
<td>1.72</td>
<td>1.36 to 2.20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.12</td>
<td>0.09 to 0.16</td>
<td>-0.92 ± 0.06</td>
<td>1.71</td>
<td>1.16 to 2.60</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.12</td>
<td>0.11 to 0.14</td>
<td>-0.91 ± 0.02</td>
<td>1.41</td>
<td>1.24 to 1.60</td>
</tr>
<tr>
<td>Pro-myostatin complex</td>
<td>1</td>
<td>17.4</td>
<td>11.0 to 32.9</td>
<td>1.24 ± 0.10</td>
<td>0.61</td>
<td>0.47 to 0.78</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12.6</td>
<td>8.53 to 21.3</td>
<td>1.10 ± 0.09</td>
<td>0.87</td>
<td>0.61 to 1.23</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5.92</td>
<td>4.80 to 7.28</td>
<td>0.77 ± 0.04</td>
<td>1.23</td>
<td>0.96 to 1.66</td>
</tr>
<tr>
<td>Uncleaved pro-myostatin</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* CI confidence interval
** SEM Standard error of the mean

As expected, unprocessed pro-myostatin showed no signalling activity, even at the highest concentrations tested (Fig. 29). This is consistent with the existing notion that minimally, cleavage of the furin site is necessary to liberate the signalling activity of pro-TGFβ superfamily proteins. Unexpectedly, the HRV-3C processed complex was not entirely latent, but showed some signalling activity albeit with a much lower potency than that of the purified mature ligand (Fig. 29). As it was very difficult to obtain reproducible data points at high concentrations of the pro-complex, the maximal response could not be determined accurately and so the calculated EC50 values should be used as an indicator only (Table
5). Nonetheless, it can be concluded that the pro-myostatin complex has considerably lower activity than the mature ligand.

Low-level activity of the pro-myostatin complex, traditionally thought to circulate in serum as a totally inactive complex, has been observed previously by Szlama et al who argued the unexpected activity is due to partial dissociation of the prodromains at the very low concentrations used in bioassays [167]. This is plausible given the EC\textsubscript{50} of mature myostatin (0.09 nM) is much lower than the dissociation constant (\(K_d\)) of the pro-complex (8 nM)[68]. It is also possible that a fraction of the pro-complex may undergo activation by proteolysis following prolonged exposure to cells in this assay. Lower levels of residual activity may be observed if the experiment would be run over a shorter time frame.

This situation is quite different to what has been seen previously for other pro-complexes. Under similar assay conditions the pro-TGF\(\beta\)1 complex is stable and shows little measurable activity [82]. On the other hand, the activity of pro-activin A and pro-BMP9 complexes is equivalent to that of their isolated mature growth factors. This suggests that their prodromains are entirely dissociated and thus exert a negligible inhibitory effect at the picomolar concentrations where the mature growth factors are active [72, 90]. While we do observe low-level activity for our pro-myostatin complex in cellular assays, it is likely that the physiological complex is further stabilised by additional interactions \textit{in vivo}. Cross-linking of the prodromains by ECM bound interactors may provide a mechanism for increasing functional affinity of the complex, as has been proposed for pro-activin A [72].

These results confirm that the non-covalently associated myostatin prodromains confer significant latency to the mature growth factor, to a lesser extent than pro-TGF\(\beta\)1, but greater than for pro-activin A and pro-BMP9. In nature, secondary cleavage of the prodomain by BMP1/TLD family proteases is required to trigger dissociation of the prodromains and liberate the full signalling potential of the ligand.

### 3.2.4 Acid activation of the latent pro-myostatin complex

Before identification of the BMP1/TLD family proteases responsible for activation of the pro-myostatin complex it was shown that the latent complex in serum could be activated by acidification of the sample [94]. Subjecting the complex to extremely low pH presumably causes dissociation of the prodromains and release of the bioactive growth factor. Curiously,
attempts by the Thompson lab (Uni. of Cincinnati) to purify the mature ligand from the acid-treated complex using a follistatin affinity column were unsuccessful despite the acid-treated sample displaying significant bioactivity [2]. The authors suggested that perhaps the prodomains remain bound to the mature ligand, but in a ‘triggered state’ with reduced latency. Here I attempted to test the acid activation of my HRV-3C processed complex and search for evidence of this ‘triggered-state’.

Figure 30: Analysis of acid-treated pro-myostatin complex. A. Bioactivity of acid treated pro-myostatin complex measured by dual-luciferase reporter assay. The acid-treated complex (red) displays moderately increased signalling activity over the untreated complex (blue), however this increased activity is lost over time (green), or following size-exclusion chromatography (black). For reference, the EC50 value for mature myostatin is indicated by the vertical dashed line. B. Size-exclusion chromatography of the untreated, or acid-treated precursor and pro-myostatin complex.

Samples of the HRV-3C cleaved pro-myostatin complex were acidified with 1M HCl for 30 min, before neutralisation with concentrated Tris base and NaOH. The protein appeared stable throughout this process, and did not show any evidence of precipitation following centrifugation. Size-exclusion chromatography of the acid treated and untreated complexes showed no significant change in elution profile between the two samples, suggesting the complex remains largely associated following acidification (Fig. 30B). The acidified complex was then tested alongside the untreated complex in the dual-luciferase bioactivity assay. The acid-treated complex gave only a modest (2-3 fold) increase in signalling activity over the untreated complex, still considerably lower than that of the mature growth factor (Fig. 30A).

It is possible that a small proportion of the pro-complex, undetectable by SEC, does dis-
sociate during acid treatment and that this contaminating growth factor is responsible for the modest increase in activity. In order to test this I re-attempted the bioactivity assay using acidified pro-complex which had been purified by SEC after acidification to remove any contaminating free ligand from the acidified complex preparation. The purified material no-longer showed an increase in activation over the untreated sample (Fig. 30A). This suggested that either the additional activity yielded by acid-activation is the result of contaminating free ligand which can be removed by SEC, or alternatively, the acid-activated state is short lived and reverts to a less active state over time. In fact, by incubating the acid-activated complex for a period of time equivalent to that required for SEC experiments, a loss of acid-activation was observed (Fig. 30A).

Together, these results indicate that acid treatment of the complex does increase bioactivity however the mechanism of activation remains unclear. Whether the additional activity is attributable to contaminating free ligand or an alternate pro-complex conformation is unknown, however the activation appears to be a reversible process. The original pro-myostatin acid activation studies were done by acid-treating whole serum samples from mice, not the isolated protein itself [94]. It should be considered that in serum the pro-myostatin complex may interact with additional binding partners which function to stabilise the latent complex. Activation by acidification of serum may therefore represent liberation of the pro-complex from a third party interaction rather than changing the nature of the pro-complex.

3.3 The pro-myostatin complex is resistant to follistatin

Inhibition of myostatin signalling by follistatin is well documented in the literature. Two molecules of follistatin bind the dimeric mature GF with high affinity, blocking both type I and II receptor epitopes [100, 107]. The natural affinity of follistatin for heparin sulfate then directs the follistatin:myostatin complex to the cell surface where it can be internalised and degraded [109, 110].

The interactions of follistatin with the precursor forms of myostatin however are unknown. It was previously shown in our lab, using biolayer interferometry (BLI), that dissociation of the pro-activin A complex was enhanced in the presence of follistatin [72]. This is consistent with displacement of the prodomains by follistatin as it competes for common binding sites.
Using the same experimental approach, we sought to determine if follistatin could similarly displace mature myostatin from its bound prodomains.

Biolayer interferometry (BLI) is an optical analytical technique used to quantify biomolecular interactions by monitoring the optical thickness of biomolecules immobilised on the tip of a ‘biosensor’. White light projected down a glass-fibre biosensor is reflected from two surfaces, a layer of macromolecule immobilised on the sensor tip and a reference layer, giving rise to a specific pattern of interference of the reflected light. When the biosensor is dipped into a solution containing a binding partner (analyte), the binding event between the immobilised molecule and the analyte in solution increases the optical thickness of the biological layer (while the reference layer remains constant). This causes a wavelength shift and changes the pattern of interference which can be measured in real time. With this information, parameters including the rates of dissociation/association, and binding affinity can be derived for biomolecular interactions.

The HRV-3C processed pro-complex was captured on the biosensor tip via the prodomain 6xHis-tag, which binds with high affinity to immobilised anti penta-His antibodies. By dipping the sensors into a simple buffer solution, we can monitor spontaneous dissociation of the mature growth factor over time as the thickness of the biological layer decreases. Furthermore, we can supplement the dissociation buffer with follistatin, to determine if it is capable of enhancing dissociation rates.

Using this setup, slow spontaneous dissociation of the mature growth factor from the pro-complex was observed, consistent with the low levels of bioactivity seen for the HRV-3C cleaved pro-complex (Fig. 31A). Interestingly, no additional dissociation of the complex was observed when the buffer was supplemented with 500 nM follistatin-288. This is in stark contrast to the pro-activin A complex which showed significantly enhanced dissociation in the presence of follistatin (Fig. 31B).

As expected, uncleaved pro-myostatin showed little or no dissociation from the biosensor, with or without follistatin (Fig. 31C). This is a good indicator that the 6xHis-tag/antibody immobilisation strategy is robust, and that background dissociation of the prodomain does not contribute to the signal we observe. It was entirely unexpected however, that when uncleaved pro-activin A was exposed to follistatin, we actually observed an increase in response, suggesting follistatin is capable of forming a ternary complex with activin A even
Figure 31: Dissociation of pro-myostatin and pro-activin A complexes as monitored by biolayer interferometry. A. Pro-myostatin complex shows slow dissociation in buffer (indicated by decrease in response), which is not enhanced by the addition of follistatin-288. B. Pro-activin A complex is susceptible to rapid dissociation by follistatin-288. C. Uncleaved pro-myostatin shows no dissociation. D. Uncleaved pro-activin A is capable of forming a ternary complex with follistatin-288 (as indicated by increased response), even prior to cleavage of the prodomains by furin.

These results show that unlike pro-activin A, the processed pro-myostatin complex is extremely stable, and resistant to dissociation by follistatin and that myostatin only becomes susceptible to follistatin inhibition following dissociation of the prodomains from the mature GF.

3.4 Conclusion

In this chapter we examined the functional properties of the precursor forms of pro-myostatin, and the mature GF isolated from the HRV-3C processed pro-complex. The isolated GF dis-
played a potent signalling response in HEK293T cells and was susceptible to antagonism by the natural inhibitor follistatin. Unprocessed pro-myostatin was shown to be entirely inactive at all concentrations tested, however the HRV-3C processed pro-complex showed low-level bioactivity, orders of magnitude less potent than the isolated GF. This activity observed for the ‘latent’ pro-myostatin complex is likely the result of partial activation/dissociation of the complex during the course of the bioactivity assays. The pro-complex could be further activated by acid treatment in vitro, however this was not sufficient to liberate the full signalling potential of the mature GF, which is likely dependent on secondary cleavage of the prodomain by BMP1/TLD proteases. Finally, unlike pro-activin A, mature myostatin cannot be displaced from its prodomains by follistatin, suggesting enhanced stability of the pro-myostatin complex.

In order to understand the molecular determinants of latency and the enhanced stability of pro-myostatin, I set out to determine its structure by X-ray crystallography.
Crystallisation & structure determination
It is well established that following furin cleavage, the non-covalently associated prodomains maintain myostatin in a (near) inactive, or ‘latent’ state by blocking receptor binding [76, 78, 80]. Concordant with these reports, in the previous chapter we demonstrated that our HRV-3C cleaved pro-myostatin complex is weakly bioactive, exhibiting as little as 1% of the signalling activity of the free ligand.

The molecular mechanism by which the myostatin prodomain confers latency to the growth factor has been the topic of considerable interest in the field, with prodomain derived therapeutics an attractive prospect for selective inhibition of myostatin signalling [148, 155, 172]. The inhibitory region of the myostatin prodomain has been mapped to the N-terminal region, analogous to the helical ‘forearm’ of pro-TGFβ1 [78]. Based on sequence homology and structural predictions, it is assumed that the prodomain α1 helix of myostatin interacts with the growth factor type I receptor binding site in a similar mode to that observed for pro-TGFβ1 and pro-activin A [79, 82, 172, 173]. Interactions of the α1 helix alone however are unlikely to confer the latency we observe for the pro-complex, given the α1 sequence is also highly conserved in non-latent family members.

Unlike pro-TGFβ1, pro-myostatin lacks the cysteines responsible for covalent cross-linking of the prodomains, which is known to be important for the latency of the pro-TGFβ1 complex [82]. In the absence of this stabilising feature, the question arises as to what drives the latency of the pro-myostatin complex on a molecular level. To answer this question I sought to determine the structure of pro-myostatin using X-ray crystallography, and analyse the structural determinants of its latency at atomic level resolution.

4.1 Initial pro-myostatin crystallisation trials

Initial crystallisation trials were focussed on the first generation wild-type pro-myostatin construct pHAT2-MSTN.002 (Fig. 16) as this was the construct we could obtain in highest purity and yield. The purified protein was tested for crystallisation in a number of commercial 96-well sparse-matrix screens using the sitting-drop vapour diffusion method (Fig. 32). These trials yielded very small protein crystals (Fig. 36, panel A) in a number of conditions, which diffracted to ≈ 8Å when exposed to synchrotron radiation, and were indexed in the cubic space group I23 (autoPROC). Crystallisation could be optimised by addition of commercial small molecule additives (Silver Bullets, Hampton Research), which yielded
slightly larger crystals (Fig. 36, panel B). Diffraction quality however was unimproved.

Figure 32: Crystallisation by vapour diffusion. A. Diagram showing process of crystallisation by vapour diffusion under ideal conditions. An initially undersaturated protein solution becomes more concentrated as water diffuses from the drop, until saturation levels permit nucleation. Following nucleation, crystal growth (without additional nucleation) occurs in the metastable zone until an equilibrium is established between protein in solution and crystallised protein. B. In the sitting-drop vapour diffusion method a mixture of protein and reservoir solution is dispensed adjacent to a volume of pure reservoir solution in a sealed well. As water diffuses from the drop to the reservoir, the protein and precipitant concentration in the drop gradually increases, promoting controlled nucleation and crystal growth.

4.2 Disorder prediction and N-terminal truncation constructs

Protein flexibility and structural disorder is undesirable in terms of crystallisation as these elements introduce conformational heterogeneity which prevents molecules from forming the well-ordered and stable packing interactions that define the crystal lattice. In order to identify regions of potential local disorder within the pro-myostatin sequence I used the DISOPRED3 server (University College London) which predicts secondary structural elements and structural disorder based on the primary amino acid sequence and existing structural information [174]. The results showed that the pro-myostatin sequence (not including the signal peptide) contains a stretch of 18 putatively disordered/conformationally flexible residues in the very N-terminus prior to the start of the α1 helix (Fig. 33). Furthermore, the 6xHis-tag and linker present in our pHAT2-MSTN.002 expression construct contribute an additional unstructured stretch of 20 amino acids.
Cryrstallisation trials using the untagged construct (pBAT4-MSTN.001) yielded slightly larger crystals of the same cubic form as previous hits (Fig. 36, panel C), but again diffraction quality remained unchanged. To further address the issue of protein flexibility I made a number of new constructs, incrementally truncating the predicted flexible N-terminus at ∆Gly37, ∆Ser50 and ∆Asn71 (Fig. 34). Truncation of the protein N-terminus significantly impacted the refolding of these constructs, with the ∆Ser50 and ∆Asn71 truncations yielding no discernible dimeric species, presumably due to partial or complete deletion of the α1 helix (data not shown). The ∆Gly37 construct did give some dimeric product in refolding trials, however refolding appeared to be considerably less efficient than full length constructs and the purified product was heavily contaminated with misfolded species. Nonetheless crystals of the ∆Gly37 construct were obtained and while still cubic in morphology, were significantly smaller than those seen previously (Fig. 36, panel D).

To get around the problem of poorly refolding truncated constructs, I designed a construct in which a tobacco etch virus (TEV) protease site was introduced into the N-terminus of our best refolding construct pHAT2-MSTN.002. By integrating the TEV site into the sequence immediately N-terminal to the desired truncation point, it was possible to refold and purify the full length protein as previously and then remove the flexible N-terminus after purification by proteolysis with TEV.

Figure 33: DISOPRED3 (UCL-CS) structural disorder prediction profile for human pro-myostatin. Confidence scores above 0.5 indicate structural disorder and provide a guide for construct modification to minimise conformational flexibility. Truncation at G37 removes the predicted disordered N-terminus. [http://bioinf.cs.ucl.ac.uk/psipred/](http://bioinf.cs.ucl.ac.uk/psipred/).
The wild-type pro-myostatin sequence was modified considerably over the course of this study to reduce conformational flexibility and facilitate crystallisation. Using this strategy I was able to obtain large quantities of highly pure pro-myostatin, truncated at Thr43 (pHAT2-MSTNΔT43)(Fig. 35A). Truncation at this point removes the 6xHis-tag, linker and the entire predicted disordered N-terminus, but should leave the prodomain α1 helix intact. This protein yielded significantly larger crystals than any of the aforementioned constructs, in a number of different crystallisation conditions (Fig. 36, panels E-G). These crystals were again indexed in cubic space group I23, with unit cell parameters $a = 197 \, \text{Å}$, $b = 197 \, \text{Å}$, $c = 197 \, \text{Å}$, $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 90^\circ$ (autoPROC). Despite their greatly increased size and beautiful overall morphology these crystals diffracted weakly, giving 4-5 Å resolution at best (Fig. 35C).

I then attempted to optimise these crystals using conventional (and less conventional) methods including: crystallisation condition and cryo-protectant screening, seeding (macro, micro and cross-matrix), small molecule additive screening, variable temperature, co-crystallisation with heparin sulfates, chemical cross-linking, lysine methylation, room-temperature diffraction and crystal dehydration using the HC1 dehumidifier at Diamond Light Source (DLS). The attempted optimisation of these crystals was exhaustive, and the failed attempts are
Figure 35: pHAT2-MSTNΔT43 crystallisation and diffraction. A. SDS-PAGE analysis of purified pHAT2-MSTNΔ43 following cleavage of the N-terminus with TEV. B. Crystal of pHAT2-MSTNΔ43 mounted in cryo-loop for diffraction analysis. C. Representative low resolution (>5Å) diffraction pattern of cubic pHAT2-MSTNΔT43, with detector edge set to 4Å.

too many to describe here. All in all, over 500 crystals were tested for diffraction and almost 50 low resolution datasets were collected. It became clear that this cubic crystal form was inherently low resolution, presumably as the result of high solvent content and/or poor crystal packing. Nonetheless, undeterred and in possession of an abundance of low resolution diffraction data, we attempted to solve the structure of pro-myostatin.

4.3 Low resolution structure determination

During an X-ray diffraction experiment we measure the intensity of waves scattered by ‘planes’ of atoms within the crystal lattice. If waves diffracted from different planes within the lattice satisfy Bragg’s law⁴ and thereby interfere constructively, diffraction peaks can be measured in the resulting diffraction pattern. From the intensity of diffraction peaks, the amplitudes of scattered waves can be derived (square-root of the intensity), however we are unable to determine the relative phase relationship of these waves. Without the phase information we are unable to calculate the electron density at a given point in the unit cell of the crystal. This is called the phase problem, and solving it requires some knowledge of the atomic coordinates of the structure.

⁴Bragg’s law (nλ=2dsinθ) defines the conditions required for constructive scattering of X-rays by a crystal lattice. If the pathlength difference (2dsinθ) between two waves is equal to an integer multiple (n) of the wavelength (λ), scattered waves remain in phase and produce measurable intensities at a diffraction angle of θ.
Figure 36: Pro-myostatin crystals. Crystallisation construct and screen details are given below each image.
4.3.1 Phasing by molecular replacement

One way in which the phase information can be deduced is by using the atomic coordinates of a structurally homologous protein, providing the similarity between the unknown and known structure is high enough (typically >25% sequence identity). This method, known as molecular replacement (MR), attempts to match the observed diffraction to the theoretical diffraction back-calculated from the existing structure in many possible orientations within the unit cell. When a good match is found, the related phases are essentially borrowed from the known structure and the electron density maps can then be calculated for the unknown structure.

Following data reduction we attempted to solve the structure of pro-myostatin by MR with various search models derived from the existing structures of pro-TGFβ1, pro-activin A and pro-BMP9. Unfortunately we were unable to obtain a definitive solution using this method, presumably due to structural divergence of pro-myostatin from the search models. It would seem obvious to use the highly conserved mature growth factor from an existing structure as the search model however this also proved unsuccessful. Any change in inter-protomer angle between protomers of the growth factor would render the dimer a useless search model, and the mature protomer alone likely represents too small a proportion of the overall scattering mass.

4.3.2 Phasing by single wavelength anomalous dispersion (SAD)

Having failed to yield a sensible solution by MR, we turned to single-wavelength anomalous dispersion (SAD) phasing to solve the phase problem. This method relies on the incorporation of heavy atoms into the protein, which diffract with measurable differences in intensity when irradiated near the absorption peak of the heavy element. This so-called ‘anomalous scattering’ of the heavy atoms can be used to determine their location (the ‘sub-structure’) within the asymmetric unit. The sub-structure and anomalous differences can then be used to estimate the phases (and subsequently the electron density) for the entire structure. The anomalous signal contribution is very small compared to the native protein diffraction, and thus anomalous differences must be measured with extreme precision and high redundancy to provide sufficient signal to noise.
For experimental phasing we used a seleno-methionine (SeMet) labelled form of our pHAT2-MSTNΔ43 construct. By substituting native methionine for seleno-methionine during protein expression we were able to introduce 7 selenium atoms into each chain of pro-myostatin. SeMet derivatisation had no noticeable impact on the expression, purification or crystallisation of this construct, and we were able to collect a number of datasets at the selenium absorption peak wavelength (0.97970 Å).

Given the high-level symmetry of the cubic space group (I23), it was possible to collect data from these crystals with exceptionally high multiplicity (redundancy). By collecting diffraction data through a full 360° rotation, symmetry related reflections are measured multiple times which greatly improves the signal to noise ratio. This is particularly important when measuring the small anomalous differences required for SAD phasing.

### 4.3.3 Low resolution structure of pro-myostatin

Our highly redundant SAD datasets showed strong anomalous signal which was successfully used to find the heavy-atom substructure and solve the structure of pro-myostatin at 4.2 Å resolution. Unsurprisingly, the crystal contacts between individual molecules in the unit cell were sparse and the entire unit cell was permeated by large solvent-filled channels (crystal solvent content >60% estimated by Matthews coefficient). The protein electron density was extremely ambiguous and only large structural features (β-strands, α-helices) were discernible. Fortunately the quality of electron density maps could be significantly improved by merging data from multiple crystals grown under identical conditions. The best seven datasets were chosen for merging based on assessment of their quality and mutual compatibility with regards to diffraction quality, similarity of unit cell dimensions, resulting $R_{\text{merge}}$ and quality of the anomalous signal (Table 6, data for individual crystals in Appendix 3).

With the improved electron density map we were able to successfully dock the structure of the mature growth factor (from PDB: 3HH2) into the density, trace most of the protein backbone, and identify bulky side-chains. The strong electron-density for selenium atoms also made it possible to assign the protein seleno-methionines (Fig. 37). We were, however, unable to unambiguously assign the sequence of the prodomain, and the registry and connectivity of the chain was unclear in many areas.
Despite the low resolution of the structure, we were able to clearly observe the overall topological arrangement of the mature growth factor and prodomains. Strikingly, pro-myostatin was observed in an unexpected V-shaped, open-armed conformation with no apparent contact between the prodomain arms (Fig. 37). This is in stark contrast to the prototypical latent pro-TGFβ1 which adopts a closed-arm conformation. Our first low resolution glimpse at pro-myostatin showed that it more closely resembles the structures of non-latent pro-activin A and pro-BMP9.

4.4 Protein engineering in pursuit of improved diffraction

While our initial structure provided some insight into the gross topological/conformational features of pro-myostatin, it lacked the resolution to clearly resolve the molecular details of the pro/mature domain interactions. In pursuit of improved diffraction resolution we continued to modify our crystallisation construct. Critically, we now had access to the low resolution structure and knowledge of the crystal packing interactions. Given the inherently poor diffraction of the I23 crystal form, it seemed sensible to try and disrupt the existing crystal contacts and thereby promote crystallisation in an alternative crystal form.
Figure 38: The cubic unit cell of low resolution pro-myostatin crystals contains 24 copies of the dimeric protein. This figure shows the distribution of these molecules within the unit cell (each shown as a different color), as viewed down the three-fold symmetry axis. Inset shows the location of the β4-5 loop within a crystal contact interface. Residues of the β4-5 loop not resolved in the electron density are represented by a dashed line.

### 4.4.1 β4-5 loop deletion

From the low resolution structure, weak/partial density of the prodomain β4-5 loop could be seen extending into a crystal contact with two adjacent molecules (Fig. 38). We rationalised that deletion of this loop sequence would destabilise the crystal contact, and prevent crystallisation in the I23 space group. Furthermore, removal of this flexible surface loop may encourage crystallisation through reduction of surface entropy. The pHAT2-MSTNΔ43β4-5del construct did indeed crystallise in a new form, this time as clusters of small indistinct crystals (Fig. 36, panel H). While the size and morphology of these crystals could be optimised by iterative cycles of micro-seeding (Fig. 36, panel I), diffraction beyond 5-6 Å was never observed.

### 4.4.2 Surface entropy reduction

In a final attempt to improve diffraction resolution, I turned to a surface entropy reduction approach. Reducing the number of highly entropic flexible residues (Arg, Lys, Glu, Gln) present on the protein surface is a commonly employed method used to promote crystallisation of tricky targets. It is argued that by reducing the conformational entropy of the
Cotton, T.R  
Crystallisation & structure determination

Figure 39: Surface-entropy reduction mutagenesis. Predicted entropic residues (red spheres) were mapped onto the low resolution structure of pro-myostatin. Clusters 1 and 3 were considered good candidates for mutagenesis, while cluster 2 occupied a domain interface and so was excluded. Crystallisation constructs were generated in which the residues of clusters 1 and 3, and both clusters combined, were mutated to alanine. Non-reducing SDS-PAGE analysis (right) shows the final purified surface-modified variants used for crystallisation.

protein surface, the entropic penalty is reduced for intermolecular interactions, thus making crystallisation thermodynamically more favourable [175].

We submitted the sequence of pro-myostatin to the UCLA Surface Entropy Reduction prediction (SERp) server (http://services.mbi.ucla.edu/SER/), which identifies clusters of highly entropic surface exposed residues as potential candidates for mutagenesis based on calculated conformational entropy profile, secondary structure and sequence conservation [175]. The server identified three clusters of putative surface-exposed high-entropy amino acids, which we analysed by mapping onto our low resolution crystal structure (Fig. 39). Residues from the first cluster (K217, Q218, E220) occupied solvent exposed positions in the prodomain arm and were considered good candidates for mutagenesis, while those from the second cluster were proximal to a functional domain interface and so were excluded from screening as not to risk disrupting the native protein structure. Residues of the third predicted cluster (G319, K320) were buried within a crystal contact involving the mature GF domain in the original crystal form.

Based on these results, a new set of crystallisation constructs were designed in which the residues of clusters 1 and 3 were mutated to alanine, both individually, and also together as a single construct. These surface modified variants were generated using TEV cleavable pHAT2-MSTNΔ43 as the template to allow for N-terminal truncation as previously described. All three variant proteins could be refolded and purified as per usual (Fig. 39)
Constructs with individually mutated clusters gave no crystal hits, however when both alanine clusters were combined in a single construct (pHAT2-MSTNΔ43SERp), the protein crystallised readily in a number of sparse-matrix conditions (Fig. 36, panels J-L). Unlike the previously observed cubic crystals, this surface engineered variant crystallised in an entirely new form, and excitingly, diffracted to much higher resolution.

4.5 High resolution structure determination and refinement

A high resolution dataset was collected from a single surface-engineered pro-myostatin crystal which crystallised in 10% PEG 6000, 0.1 M HEPES pH 7.0. The resulting dataset was indexed in monoclinic space group C121 with unit cell parameters a = 168.16 Å, b = 36.3 Å, c = 120.45 Å, α = 90°, β = 104.4°, γ = 90° (autoPROC). This dataset was used to re-solve the structure of pro-myostatin at 2.6 Å, using one symmetrical half of the low resolution dimer structure as the input model for phasing by molecular replacement. In this instance MR gave a single, highly scoring solution. The resulting electron density map was vastly improved over the low resolution maps (Fig. 40A), and we were able to successfully build and refine 92% and 81% of the residues into the two protomer chains (final R_work/R_free: 0.215/0.260). The remaining regions were disordered and not resolved in the electron density, with chains A and B missing 27 and 62 residues respectively.

Interestingly, there was a significant difference in the quality of electron density between different parts of the two chains, which is plainly reflected in the B-factors (Fig. 40B). Given the disparity in quality between the protomers, our structural interpretations are based on the analysis of the best resolved parts of both chains. The refined high resolution structure was then used as a template to re-build and refine the low resolution structure with the correct registry (final R_work/R_free: 0.274/0.301). The refined low and high resolution structures, and their associated reflection data were deposited in the protein data bank (PDB) with accession codes 5NTU and 5NXS respectively (final data processing and refinement statistics in Table 6). Many thanks to Dr. Gerhard Fischer for his invaluable guidance and assistance with all aspects of data collection, processing and refinement.
Figure 40: A. Comparison of representative electron density from low and high resolution structures. In both figures the same β-strand (L249-T258) from each structure is viewed from the same orientation and the electron density map is contoured to 1.5 σ. B. B-factor analysis of high resolution pro-myostatin structure (PDB: 5NTU). The protein backbone is rendered by B-factor from lowest (thin blue tubes) to highest (thick red tubes). There is significant difference in the B-factors between individual chains of the dimer. C. Ramachandran plots for refined pro-myostatin structures showing backbone dihedral angles for each amino acid in the structure with respect to energetically favoured/allowed regions (RAMPAGE server, http://mordred.bioc.cam.ac.uk/~rapper/rampage.php).
4.6 Conclusion

Obtaining diffraction quality crystals of human pro-myostatin proved to be an exceptionally challenging task, as seems to be the trend with proteins of the pro-TGFβ superfamily. First generation constructs crystallised exclusively in the same cubic form, and although superficially beautiful, were inherently poorly diffracting due to high solvent content and sparse crystal-contacts. Nonetheless, by combining data from a number of low resolution datasets we were able to solve the structure of pro-myostatin at 4.1 Å using SAD phasing of seleno-methionine derivative crystals.

To achieve greater resolution I employed a surface entropy modification approach in which a number of surface exposed residues were mutated to alanine in an attempt to reduce conformational flexibility at the protein surface and promote crystallisation. Access to the low resolution structure was immensely beneficial for this, as it allowed pre-screening of mutagenesis candidates based on their approximate position within the structure.

A surface-modified variant with five alanine substitutions per chain crystallised readily in an entirely new form to those previously obtained, and had greatly improved diffraction, demonstrating the power of this approach for the crystallisation of otherwise intractable targets. Molecular replacement was used to solve the structure at 2.6 Å, which was subsequently refined and deposited in the PDB alongside the previously determined low resolution structure. The time taken from obtaining the initial crystal hit to depositing the final high resolution structure was nearly two years.

Pro-myostatin is the fourth member of the pro-TGFβ superfamily for which structural information is now available in addition to pro-activin A, pro-TGFβ1 and pro-BMP9. These structures represent a small fraction of the total number of TGFβ superfamily members and only the tip of a potentially very structurally diverse iceberg. In the next chapter I discuss the molecular structure of pro-myostatin, and explore the structural features which underpin the stability and latency of the pro-myostatin complex.
Table 6: Crystallographic data collection, processing and refinement statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>High resolution</th>
<th>Low resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB code:</td>
<td>5NTU</td>
<td>5NXS</td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synchrotron / beamline</td>
<td>DLS / I-03</td>
<td>DLS / I-03</td>
</tr>
<tr>
<td>X-ray wavelength [Å]</td>
<td>0.97625</td>
<td>0.9797</td>
</tr>
<tr>
<td><strong>Data processing</strong> *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C 1 2 1</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Unit cell (a, b, c) [Å]</td>
<td>168.16, 36.30, 120.45</td>
<td>196.83, 196.83, 196.83</td>
</tr>
<tr>
<td>α, β, γ [°]</td>
<td>90.0, 104.4, 90.0</td>
<td>90.0, 90.0, 90.0</td>
</tr>
<tr>
<td>Resolution limits [Å]†</td>
<td>76.26-2.58 (2.62-2.58)</td>
<td>98.41-4.19 (4.27-4.19)</td>
</tr>
<tr>
<td>Number of protomers in ASU</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>No of total / unique reflections</td>
<td>90386 / 22474</td>
<td>1517107 / 9470</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>4.0 (4.2)</td>
<td>160.2 (128.0)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.056 (0.826)</td>
<td>0.141 (2.690)</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.071 (1.037)</td>
<td>0.148 (2.732)</td>
</tr>
<tr>
<td>I/σI</td>
<td>11.7 (1.3)</td>
<td>28.6 (3.0)</td>
</tr>
<tr>
<td>CC1/2</td>
<td>0.997 (0.583)</td>
<td>0.996 (0.923)</td>
</tr>
<tr>
<td>Completeness [%]</td>
<td>97.7 (99.6)</td>
<td>100.0 (100.0)</td>
</tr>
<tr>
<td>Anomalous completeness [%]</td>
<td>100.0 (100.0)</td>
<td></td>
</tr>
<tr>
<td>Anomalous multiplicity</td>
<td>84.0 (66.0)</td>
<td></td>
</tr>
<tr>
<td>Anomalous signal (</td>
<td>DANO</td>
<td>/σ(DANO) )</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rwork/Rfree [%]</td>
<td>0.215 / 0.260</td>
<td>0.274 / 0.301</td>
</tr>
<tr>
<td>No. of unique/free reflections used</td>
<td>22310 / 1118</td>
<td>9460 / 478</td>
</tr>
<tr>
<td>R.m.s deviations:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond lengths [Å]</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>bond angles [°]</td>
<td>0.55</td>
<td>0.53</td>
</tr>
<tr>
<td>Ramachandran analysis: (no. / % of residues)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favoured</td>
<td>535 / 95 %</td>
<td>463 / 89 %</td>
</tr>
<tr>
<td>Allowed</td>
<td>24 / 4 %</td>
<td>49 / 9.5 %</td>
</tr>
<tr>
<td>Outliers</td>
<td>2 / 0.4 %</td>
<td>3 / 0.6 %</td>
</tr>
<tr>
<td>Number of atoms/B-factors:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>4404 / 100.5</td>
<td>3556 / 91.7</td>
</tr>
<tr>
<td>Solvent atoms</td>
<td>30 / 76.0</td>
<td>0 / -</td>
</tr>
<tr>
<td>Heterogen atoms</td>
<td>86 / 90.8</td>
<td>0 / -</td>
</tr>
<tr>
<td>Mean/Wilson B-factor</td>
<td>100.2 / 95.7</td>
<td>91.7 / 225.3 §</td>
</tr>
</tbody>
</table>

* Processing statistics for 5NXS are shown for data merged from 7 crystals
† Data in parenthesis are for the highest resolution shell.
‡ Data in parenthesis are for the low resolution shell (98.41-11.38 Å)
§ The Wilson B-factor is ill-defined due to the low resolution of this structure
Structural analysis of human pro-myostatin
The primary goal of this doctoral study was to elucidate the molecular structure of human pro-myostatin, in order to better understand the structural basis for growth factor latency and activation. In this chapter we present the sought after crystal structure and an in-depth analysis of the unique (and shared) molecular features of pro-myostatin, and their functional implications for signalling and regulation.

Pro-myostatin, as expected, forms a disulfide-linked homodimer in which two protomers are linked covalently through a conserved intermolecular disulfide in the mature GF domain (Fig. 41). The individual chains of pro-myostatin, like other pro-TGFβ superfamily proteins, consists structurally of an N-terminal prodomain and C-terminal mature GF domain.

The dimeric myostatin GF is typical of the TGFβ superfamily, with each GF protomer
consisting of four antiparallel β-strands (‘fingers’), covalently linked to one another by a central cystine-knot motif (Fig. 41). The two GF protomers of the dimeric assembly associate in characteristic palm-to-palm fashion, with equivalent Cys339 residues forming an intermolecular disulfide bond between the two.

The prodomain also contains structural elements conserved in other family members and can be divided into N-terminal ‘forearm’ and globular C-terminal ‘arm’ sub-domains (Fig. 42). The forearm region consists of the α-helices 1 & 2 (and connecting loops) N-terminal of the BMP1/TLD site, which form intimate encircling interactions around the mature GF and provide the primary pro-mature interface. Upstream of the BMP1/TLD site is the globular C-terminal ‘arm’ domain, which consists of two anti-parallel β-sheets (also known as a ‘jelly-roll’ or ‘β-sandwich’ fold) and a small α-helix, and sits atop its associated mature GF protomer.

Figure 42: Structural domains of pro-myostatin coloured by rainbow from N (blue) to C (red) termini. A. The single protomeric chain of the pro-myostatin dimer consists of an N-terminal prodomain and C-terminal mature GF domain. B. The prodomain can be further subdivided into the N-terminal ‘forearm’ helices, and the globular C-terminal ‘arm’. C. The mature GF domain with disulfides of the cystine-knot shown in stick and ball representation.

5.1 Pro-myostatin shows inter-protomer flexibility

While the individual chains of our high and low resolution structures align exceptionally well (Ca RMSD: 0.68 Å, 227 atoms), the dimeric structures cannot be fully aligned due to significant shifts in the ‘inter-protomer angle’, as measured between the central dimerisation disulfide and residue Gln358 at the fingertip of each mature GF (Fig. 43A). The dimer of
the low resolution structure adopts a more closed conformation (89.2°) than that of the high resolution structure (108.5°). This shift in angle arises due to conformational flexibility about the dimerisation disulfide bond which allows the individual protomers to pivot about this central axis.

Figure 43: Inter-protomer flexibility in myostatin crystal structures. A. The individual chains of the low (grey) and high resolution (red) crystal structures of pro-myostatin align well, however the dimer cannot be aligned due to a significant shift in the inter-protomer angle. B. Structures of the mature myostatin dimer aligned to a single protomer show variable inter-protomer angles (in parentheses) as measured between Gln358 of each chain and the central dimerisation disulfide bond. Aligned protomers are coloured grey.

Variability of the inter-protomer angle is also observed in existing structures of mature myostatin, including the apo and follistatin-288 bound growth factors (PDBs 5JI1 & 3HH2 respectively) (Fig. 43B)[42]. These crystal structures appear to have trapped unique inter-
protomer angles, perhaps dependent on the presence or absence of binding partners, or the nature of crystal packing interactions.

The crystal structure of myostatin bound to a neutralising antibody (PDB: 5F3H) displays a dramatic inter-protomer angle, in which fingers of the protomers approach the horizontal plane \(172^\circ\) \[42\]. This exaggerated ‘open’ conformation is also observed in the crystal structure of apo-GDF11 \[42\]. Similarly, activin A also exhibits conformational flexibility, with inter-protomer angles ranging from \(50^\circ\) to \(108^\circ\) in various crystal structures \[72\]. Together, these observations highlight the extreme malleability of the TGF\(\beta\) superfamily growth factors, and the influence of interaction partners (bona fide or crystallographic) on protein conformation.

This conformational flexibility may also explain the difficulty in crystallising pro-TGF\(\beta\) superfamily members and the current dearth of structural information. Restricting inter-protomer dynamics by introducing novel interactions at the dimer interface may represent a broadly applicable strategy for improving the propensity of these proteins towards crystallisation.

The extent of inter-protomer flexibility outside of the context of a crystal structure (i.e. dynamics in solution) and the relevance of this conformational plasticity on the biological function of these growth factors remains to be seen. Greenwald et al (2004) propose that the binding of two membrane tethered type II receptors drives activin A into a ‘wings-spread’ conformation which has an increased affinity for the type I receptor \[99\]. In this way, ligand flexibility plays a direct role in assembly of the receptor complex at the membrane.

One could imagine engineering a growth factor in such a way as to introduce an additional inter-molecular interaction which would restrict inter-protomer dynamics or shift the conformational equilibrium to favour a subset of inter-protomer angles, in order to assess the role of conformation in growth factor signalling. The dynamics of differentially-labelled heterodimeric growth factors could also be interrogated using a single molecule FRET based approach.
5.2 The prodomain $\alpha_1$ helix displaces the mature GF wrist helix

As described in section 1.3.5, the mature GF domains of most TGF$\beta$ superfamily growth factors share a structurally conserved $\alpha_3$ wrist helix between the first and second $\beta$-fingers. Upon dimerisation, this short helix occupies an interfacial position within the palm of the opposing GF protomer and forms part of the presumed type I receptor binding site [100]. Based on the structure of TGF$\beta$3 in complex with its type I receptor (ALK5) ectodomain (PDB: 2PJY), it is reasonable to assume that myostatin shares a conserved binding mode with this common receptor [45]. Curiously, in both of our crystal structures the prodomain $\alpha_1$ helix occupies the hydrophobic groove of the palm and displaces the wrist helix from its expected position, destroying the putative type I receptor site (Fig. 44). The displaced wrist helix and pre-helix sequence form a $\beta$-hairpin which is resolved in the crystal contact of one chain and interfaces with the solvent exposed ‘back’ face of the prodomain $\alpha_1$ helix.

![Image](image-url)

Figure 44: Displacement of the mature GF wrist helix in pro-myostatin. Left. Structure of mature myostatin from myostatin:follistatin complex (PDB: 3HH2, follistatin not shown) with wrist helix present in the type I receptor binding site. Middle. Myostatin GF from the pro-myostatin structure (PDB: 5NTU, prodomain not shown) with wrist helix displaced by the prodomain $\alpha_1$ helix. Right. Structures from A and B superimposed and rotated by 90° to show change in conformation of the wrist helix.

In the absence of a structured wrist helix at the type I receptor site it is unlikely that myostatin would be able to engage its type I receptor and as such the prodomain renders the GF conformationally incompetent for signalling. Interestingly, the N-terminal domain (ND) of follistatin is however able to occupy the type I receptor binding site, and potently inhibit signalling without displacing the wrist helix [100]. The wrist helix is similarly displaced by the prodomain $\alpha_1$ helix in the structures of pro-TGF$\beta$1 and pro-activin A, although the displaced sequence is not visible in these structures. Pro-BMP9 is an unusual exception as
the α1 helix of the prodomain lacks any discernible electron density in the crystal structure and is thought to be disordered, while the wrist helix of the mature GF remains undisturbed in the type I receptor site [90]. Instead, the BMP9 prodomain α5 helix occupies a ‘non-invasive’ position alongside the wrist helix, more reminiscent of the follistatin ND binding mode.

5.3 Pro-myostatin has an unexpected open-armed conformation

Pro-myostatin, like pro-TGFβ1, forms a latent complex following cleavage of the prodomain at the furin site. It was therefore envisaged that pro-myostatin would be structurally comparable to pro-TGFβ1 in its overall topology and perhaps adopt a similar ‘closed’ conformation, albeit lacking the prodomain cross-linking disulfide which is unique to TGFβ isoforms. Pro-myostatin however, was shown to exist in an open-armed, V-shaped conformation in which there is no contact between prodomain arms. This is more similar to that seen for pro-activin A and pro-BMP9, both of which form non-latent pro-complexes (Fig. 45).

Figure 45: Comparison of the structural topology of pro-myostatin with existing pro-TGFβ family structures. Mature GFs are coloured orange in all structures while the prodomains are coloured blue, grey, green or pink.

To assess the relevance of this open-armed conformation in solution and extend these observations to the pro-myostatin complex, small-angle X-ray scattering (SAXS) was employed to examine the overall conformational of pro-myostatin and the HRV-3C processed complex.
SAXS is a biophysical method used to determine low resolution structural information (i.e. particle size and shape) for macromolecules in solution, as well as for interrogating sample polydispersity and conformational transitions. In brief, a sample of macromolecule in solution is irradiated with collimated monochromatic X-rays (typically from a synchrotron source), and the scattered X-ray intensities recorded by a detector. The random orientation of macromolecules in solution gives rise to isotropic scattering, and as such the collected intensities can be radially averaged. The resulting scattering information is equivalent to the scattering of a single particle averaged in all possible orientations. In this case, scattered intensities closest to the beam (the small-angle scattering) are of interest because these intensities represent the scattering contribution of the low resolution features [176].

Solution scattering profiles of unprocessed pro-myostatin and the HRV-3C cleaved complex overlay well, and the estimated radius of gyration ($R_g$) value (both Guinier and real-space derived) is unchanged following cleavage of the prodomain (Fig. 46A). Furthermore, the pairwise distance distribution functions $P(r)$, which describe the distribution of inter-atomic distances within the molecule, overlay well and produce comparable estimates of the maximum particle size $D_{max}$ for both samples (Fig. 46B).

Similarity of the SAXS scattering parameters (Table 7) suggests that pro-myostatin does not undergo a significant conformational rearrangement following cleavage of the prodomain. This observation, however, does not exclude the possibility of local structural changes which are undetectable using this method. Moreover, this data suggests that any conclusions drawn about the overall topology of pro-myostatin, based on the crystal structure, are likely also applicable to the pro-myostatin complex.

In order to further explore the solution structure of pro-myostatin, SAXS data was used to calculate a molecular envelope of both the unprocessed precursor and the HRV-3C cleaved complex. The resulting ab initio envelopes show similar overall shapes, with extended conformations that easily accommodate our crystal structures (Fig. 46C, D). Both envelopes contain additional scattering volume in the space above and between the open prodomain arms, in which a range of additional conformations with smaller inter-protomer angles could be accommodated. This suggests an inherent flexibility of these pro-forms, consistent with the variable inter-protomer angles observed in crystal structures.
Figure 46: Small-angle X-ray scattering (SAXS) analysis of pro-myostatin. A. SAXS scattering profile showing intensity ($I$) vs. scattering angle ($q$) for unprocessed pro-myostatin (black circles) and HRV-3C cleaved pro-myostatin complex (blue circles). B. Pairwise distance distribution functions $P(r)$ calculated by DATGNOM ($q_{max}$=0.2) for unprocessed pro-myostatin (black line) and HRV-3C cleaved pro-myostatin complex (blue line). Both functions smoothly approach zero at their respective values of $D_{max}$. C. Ab initio SAXS envelope (DAMFILT) of unprocessed pro-myostatin, with docked pro-myostatin crystal structure. D. Ab initio SAXS envelope of HRV-3C cleaved pro-myostatin complex, with docked pro-myostatin structure.

Table 7: Experimentally determined SAXS parameters

<table>
<thead>
<tr>
<th></th>
<th>[Protein] (mg/ml)</th>
<th>$I(0)$</th>
<th>$R_g$ (Guinier, Å)</th>
<th>$D_{max}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncleaved pro-myostatin</td>
<td>7.5</td>
<td>0.0827</td>
<td>40</td>
<td>140.6</td>
</tr>
<tr>
<td>Pro-myostatin complex</td>
<td>9</td>
<td>0.0495</td>
<td>38.9</td>
<td>146.6</td>
</tr>
</tbody>
</table>
Further validation of the open-armed, V-shaped topology observed for pro-myostatin came from the lab of Professor Timothy Springer (Harvard Medical School), whose group was simultaneously working on structural analyses of pro-myostatin. While pursuing diffraction quality crystals of pro-myostatin, the Springer group turned to negative-stain electron microscopy (EM) as a means of visualising their mammalian expressed protein. Using this method they were clearly able to resolve the distinctive V-shape of both the unprocessed precursor and the furin cleaved complex, which were essentially indistinguishable from one another [3]. The concordance of these results obtained using orthogonal and complementary structural methods provides validation of the topology of pro-myostatin we observe in our crystal structures.

5.4 Pro-myostatin is a domain-swapped dimer

When building the pro-myostatin crystal structure it became apparent that electron density was missing for the sequence linking the prodomain ‘forearm’ and ‘arm’ domains, between Asp95 and Glu107. This unresolved region contains the BMP1/TLD site (Arg98/Asp99), cleavage of which is required for activation of pro-myostatin in vivo [80]. Because pro-myostatin is homo-dimeric, there are two possible structural connectivities between the forearm and arm domains within the dimeric assembly. The first of these would involve the forearm and arm domain of a single chain interacting with the same mature GF protomer and forming one symmetrical half of the dimer (Fig. 47A). The other would involve the forearm from one side of the symmetrical dimer linked to the arm domain on the other, in a domain-swapped arrangement (Fig. 47B). In this way, the forearm and arm domains from the same chain interact with both mature GF protomers.

Measuring the distances between the last visible residue of the forearm (Asp95) and the first of the arm domain (Glu107) allowed us to conclude that the prodomain most likely adopts a domain-swapped configuration in which the forearm interacts with the mature GF from one chain, while the arm domain crosses the dimeric symmetry axis and interacts with the GF protomer of the other chain. In this arrangement, the distance between Asp95 and Glu107 is 23 Å (2.1 Å per residue), compared with 35 Å (3.2 Å per residue) for the alternate connectivity (Fig. 47). The latter arrangement, while possible (assuming maximum length of 3.5 Å per residue for an extended β-strand) would require a highly constrained trajectory.
of the missing 11 residues to bridge this gap, which is unlikely given the disordered nature of these amino acids in our crystal structure.

The linker regions containing the furin cleavage sites are also weakly resolved in the electron density. We were however able to build the entire main-chain between the prodomain arm and the mature GF in one of the two protomers. In doing so, this revealed an additional domain-swapped arrangement in which the arm domain from chain A interacts with the mature GF of chain B. Together, this domain-swapping gives a criss-crossed structure in which the forearm of chain A interacts with mature GF from the same chain before crossing over to the arm domain which interacts with the mature GF of chain B (Fig. 47).

![Figure 47: Possible prodomain forearm-arm connectivity. In these figures, one protomer of the dimer is coloured in rainbow from N (blue) to C (red) termini, while the second protomer is coloured transparent grey. The missing sequence of the BMP1/TLD site linker is represented by blue circles. A. In this arrangement, the forearm and arm interact with the same mature GF domain, and the 11 missing residues of the BMP1/TLD linker must traverse a distance of 35 Å. B. In our proposed connectivity, the forearm and arm from the same chain adopt a domain-swapped configuration in which only 23 Å separates the last visible residue of the forearm with the first visible residue of the arm.](image)

It is interesting to note that density for the furin site and associated linker sequence is at
least partially visible in our structure of pro-myostatin, while it is entirely missing in the structures of pro-activin A and pro-TGFβ1. In pro-myostatin, 12 residues of the furin site linker span a distance of 34 Å, while in pro-activin A, the unresolved 10 residues need only span 16 Å and therefore may take a more meandering route. This suggests that the furin site is more structurally constrained in pro-myostatin and thus potentially less accessible to the protease when compared to the disordered, flexible furin sites of the aforementioned family members.

Pro-activin A and pro-TGFβ1 are known to form similar domain-swapped configurations as observed here for pro-myostatin, and this is thought to have important implications for the biosynthesis of TGFβ superfamily growth factors [72, 177]. Firstly, the interaction of the helical forearm with the mature GF protomer from the same chain allows stabilisation and solubilisation of the mature GF during biosynthesis and secretion, and maintains the protomer in a dimerisation-compatible configuration. Secondly, domain-swapping may help facilitate heterodimerisation of growth factors when two or more family members are co-expressed in the same cell. Heterodimeric TGFβ superfamily members are known to exist as biologically relevant entities, often with bioactivities comparable to or even more potent than that of their respective homodimers [37–39]. Differential affinity of the prodomain arm for its own mature GF versus that of another family member may help drive such heterodimerisation.

### 5.5 The prodomain forearm is a key latency conferring element

The prodomain forearm provides the primary molecular interface between the pro and mature domains of pro-myostatin and is known to be a key driver of latency for the pro-myostatin complex [78]. Structurally, the N-terminal forearm consists of α-helices 1 and 2, and the flexible loop which connects the two (commonly referred to as the ‘latency lasso’). Together, these elements encircle the mature GF protomer and form key interactions on both its concave and convex surfaces. This helix-loop-helix motif and its mode of interaction with the mature GF is structurally similar to that first described for pro-TGFβ1 and subsequently for pro-activin A however the pro-activin A forearm contains an additional β-hairpin at its C-terminus (Fig. 48).

Conserved helical signatures corresponding to the forearm α1 and α2 helices are also widely
present in sequences of other pro-TGFβ family members for which structural information is currently unavailable [6]. This suggests that an evolutionarily conserved mode of pro-mature interaction may exist within the superfamily despite low overall sequence homology between prodomains.

Figure 48: A. Comparison of prodomain forearms motifs from pro-myostatin, pro-TGFβ1 and pro-activin A crystal structures. Prodomain forearms are shown as blue ribbons while the associated mature GF proomers are shown as orange surface representations. The second mature GF proterom is shown as transparent ribbon. B. ClustalX alignment of prodomain forearm sequences. The alignment begins at the first amino acid following the predicted signal peptide, with residues coloured by percentage identity, from least conserved (white) to most conserved (dark blue). Secondary structure annotation shown below the alignment is based on the structure and sequence of pro-myostatin. Residues of the pro-myostatin α1 helix which interact directly with the mature GF are indicted by arrowheads.

The core inhibitory region of the myostatin prodomain was initially mapped by Jiang et al (2004) who showed that residues 42-115 of the prodomain provided the necessary elements to inhibit myostatin signalling in vitro [78]. This region spans the entire forearm from the N-terminus of the α1 helix to the beginning of the arm domain. The α1 helix was subse-
quently shown to be critical for stability of the latent pro-TGFβ1 complex, supporting the notion that the forearm is a key latency conferring feature of latent pro-TGFβ superfamily complexes [79].

5.5.1 Interactions of the α1 helix

Analysis of our pro-myostatin crystal structure shows that the prodomain α1 helix is helical from Arg45 to Leu64 and occupies the hydrophobic groove on the concave face of the mature GF protomer with a similar binding mode to the corresponding helices in pro-TGFβ1 and pro-activin A (Fig. 48). By binding to the mature GF in this way the α1 helix displaces the wrist helix and destroys the type I receptor site.

The α1 helix is amphipathic in nature, with its hydrophobic face buried in the GF interface and polar residues oriented towards the solvent. Of the seven aliphatic residues in the myostatin α1 helix, six are highly conserved within the superfamily (Ile53, Ile56, Ile60, Leu61, Leu64, Leu66), the latter four of which are fully conserved in GDF11, activin A, and TGFβ isoforms (Fig. 48B). The only hydrophobic residue unique to the α1 helix of myostatin is Ile58, which projects outwards away from the mature GF interface towards the displaced wrist helix sequence (Fig. 49).

Figure 49: Interactions of the prodomain forearm with the mature GF. The prodomain forearm (grey ribbon) forms key interactions with the mature GF domain (orange ribbon). Residues of the forearm involved in the GF interface are shown as sticks, coloured blue for polar/charged residues and orange for hydrophobic residues. The N-terminal half of the α1 helix is dominated by polar residues, while the C-terminal half is buried deeper within the GF interface and is rich in aliphatic, hydrophobic residues.

The importance of the conserved α1 hydrophobic residues was highlighted by Asari et al who
showed by alanine scanning that substitution of aliphatic residues diminishes the inhibitory capacity of synthetic α1 derived peptides [178]. Asari et al also highlighted Trp44 at the very N-terminus of the α1 helix as an important residue for effective inhibition of myostatin signalling. This is surprising given this residue is poorly resolved in our structure and appears to extend into solvent away from the mature GF (Fig. 49). Similarly, Ile58 was identified as a critical residue for inhibition of myostatin by α1 derived peptides. Again, this is somewhat intriguing given Ile58 does not form part of the mature GF interface in the crystal structure but instead appears to interface with aliphatic residues on the displaced wrist helix β-hairpin structure (Fig. 49). If this interaction persists under physiological conditions, it could be imagined that the β-hairpin helps to stabilise the α1:GF interaction.

While the aforementioned hydrophobics are located predominantly in the C-terminal half of the α1 helix, the N-terminal half interacts with the mature GF via polar contacts (Fig. 49). Arg45 of the α1 helix (conserved only pro-GDF11) hydrogen bonds with the main-chain carbonyls of Glu274 and Ser276 located on the N-terminal extension of the mature GF, adjacent to the furin cleavage site (Fig. 50A). Similarly, Lys49 of the α1 helix makes a salt-bridge with the carboxyl group of the Glu274 side chain, also on the N-terminal extension. Arg52 and Lys63 make additional polar contacts with residues of the mature GF fingers. It has been suggested that the cationic residues of the prodomain α1 helix are required for the non-covalent interaction of pro-TGFβ1 with latent TGFβ binding proteins (LTBPs) during biosynthesis and deposition in the extracellular matrix [79]. These non-covalent interactions may promote subsequent disulfide linkage of the TGFβ prodomain (through Cys33) to the LTBP, a process required for integrin-dependent activation of TGFβ signalling. Myostatin is also known to interact non-covalently with LTBP3, and speculatively, the aforementioned charged residues of the α1 helix may mediate such an interaction.

Arg65 at the very C-terminus of the α1 helix forms a particularly well defined stacking interaction with the side chains of Tyr111 and His112 from the BMP1/TLD site linker of the opposing protomer (Fig. 50B). In this interaction, the charged guanidinium group stacks with the histidine imidazole group in planar fashion. The aromatic ring of Tyr111 then sits perpendicular to the plane of the histidine side chain. This interaction is equivalent to the ‘fastener’ motif described for pro-TGFβ1, which was shown to be important for latency of the pro-complex [82]. The equivalent α1 arginine, and histidine-tyrosine pair are also conserved in the sequence of pro-GDF11 suggesting the fastener may be a general feature
of latent pro-TGFβ superfamily members.

There has been considerable interest in the development of α1 helix mimicking peptides as potent and specific inhibitors of myostatin, for the treatment of muscular atrophic disorders. Takayama et al put considerable effort into this avenue of research, having originally developed a range of synthetic peptidic inhibitors of myostatin signalling by truncating the mouse pro-myostatin α1 helix at various points [172]. The most potent of these (Trp44-Leu66) was shown to bind mature myostatin with a $K_d$ of 30 nM, and yielded increased muscle mass when injected into mdx mice. Further elaboration of these peptides by mutagenesis and incorporation of non-natural chemical moities at the N-terminus led to modestly improved potency of inhibition [173, 178].

It is unlikely however that the wild-type α1 helix alone is sufficient to explain the latency of the pro-myostatin complex in nature. The EC$_{50}$ measured for mature myostatin (0.09 nM) is more than 300-fold lower than the $K_d$ of the highest affinity α1 derived helix (30 nM) described to date, and thus at the concentrations required for bioactivity very little of the α1 helix would remain associated, let alone outcompete the type I receptor. This is consistent with the fact that the pro-activin A complex is non-latent despite having an α1 helix in which many of the GF interacting residues are conserved. It seems likely that further structural elements are required to increase the functional affinity of the prodomain and confer latency to the pro-myostatin complex.

5.5.2 Interactions of the latency lasso and α2 helix

Immediately C-terminal to the α1 helix is a loop which wraps around between the tips of the mature GF fingers. This loop, originally termed the ‘latency lasso’ for pro-TGFβ1, links the α1 helix on the concave face of the mature GF with the α2 helix on the convex side. Interestingly, the pro-myostatin latency lasso contains a five residue insert (also conserved in pro-GDF11) which creates a small ‘lasso helix’ not observed in other pro-structures (Fig. 48). The α2 helix sits atop the convex face of the mature GF protomer and occupies the putative type II receptor binding site, the location of which can be inferred from existing structures of activin A in complex with ActRIIB (Fig. 51) [98, 99]. This interaction is primary hydrophobic in nature and additionally stabilised by a hydrogen bond between Tyr94 of the α2 and the backbone carbonyl of Asn349 (Fig. 50C).
Cotton, T.R  
Structural analysis of human pro-myostatin

Figure 50: Key interactions of the prodomain forearm. A. N-terminal polar interactions of the α1 helix (grey ribbon) with the mature GF (orange ribbon and sticks). B. Fastener residue interactions. Arg65 of the α1 helix (grey) forms a stacking interaction with His112 and Tyr111 from the BMP1/TLD site linker of the opposite prodomain (blue). This fastener motif, and associated acidic residues (Asp109, Asp110) comes within proximity of the basic furin recognition motif (red). C. Hydrogen bonding between Tyr94 and Asn349 stabilises the interaction of the forearm α2 helix (grey) on the convex face of the mature GF.
Figure 51: A. Surface representation of one mature myostatin GF protomer (white) and the interacting prodomain forearm (grey ribbon). The putative ActRIIB receptor binding site (green) has been mapped onto the surface of the mature GF, based on the structure of activin A in complex with ActRIIB (PDB: 1S4Y). B. Mature myostatin GF coloured by surface hydrophobicity, ranging from most polar (blue) to most hydrophobic (red). The \( \alpha_2 \) helix masks the primary hydrophobic surface on the convex face of the mature GF, which also overlays with the putative ActRIIB receptor epitope as shown in A. C. Same as in B, but the structure has been rotated 180° to show interaction of the \( \alpha_1 \) helix with the hydrophobic concave face of the mature GF.

Together, the elements of the prodomain forearm constitute the primary interface between the pro and mature domains, and function to occlude both putative type I and II receptor binding sites. In doing so, the forearm masks the major hydrophobic surfaces of the mature growth factor, making the pro-form considerably more soluble than the free ligand (Fig. 51). Consistent with these observations, the prodomain forearm of GDF11 was shown to remain associated with the mature GF following proteolytic activation and significantly improved solubility of the growth factor [179].

5.6 The prodomain arm stabilises the GF/forearm interaction

While the prodomain forearm constitutes the primary interface between the prodomain and mature GF domain, its interaction alone is unlikely to fully explain the latency of the pro-myostatin complex. The additional interaction of the globular prodomain arm with the mature GF and forearm however may provide sufficient stabilisation to prevent dissociation of the prodomains following furin cleavage, thus maintaining growth factor latency.

The globular arm of the myostatin prodomain consists of a ‘\( \beta \)-sandwich’ fold, with two antiparallel \( \beta \)-sheets and a small \( \alpha \)-helix. While this fold is structurally conserved with other TGF\( \beta \) family members, it lacks the \( \beta 8/9 \) hairpin ‘bow-tie’ motif which facilitates disulfide dimerisation of the pro-TGF\( \beta 1 \) prodomains. Despite structural homology of the
arm domain, its orientation relative to the mature GF is unique to myostatin (Fig. 45). Because the prodomains are not covalently linked via the arm domain, as is the case for pro-TGFβ1, pro-myostatin is able to adopt an open-armed conformation in which the arm domain sits parallel to the mature GF, giving rise to a considerably enhanced interface with the mature GF and its associated forearm.

This unique arm orientation has a number of implications for the pro-myostatin complex. Firstly, it gives rise to an extended anti-parallel β-sheet interface between β1 of the arm domain and β7 of the mature GF (Fig. 52A). This interface is stabilised by eight inter-strand hydrogen bonds, compared with five for pro-activin A and only three for pro-TGFβ1. Further β-sheet hydrogen bonding is not possible for pro-TGFβ1 given the conformational restriction imposed by the covalently linked prodomains. Given pro-GDF11 is known to adopt a similar open-armed conformation, and has considerable sequence similarity to pro-myostatin at this interface, it is likely to form a similarly extensive hydrogen bond network across the GF:arm β-sheet interface [179].

In addition to the aforementioned hydrogen bonding network, the prodomain arm further stabilises the pro-complex by forming a considerable interaction surface with the prodomain forearm on either side of the mature GF (Fig. 52B). Within this interface, the β6/β7 loop of the prodomain arm partially extends over the latency lasso, further stabilising the arm:forearm interface. On the opposite side of the mature GF, the fastener motif connects the α1 helix to the prodomain arm, completing the ring of elements which encapsulate the mature GF fingers.

The buried interface between the arm domain and the mature GF/forearm in pro-myostatin (2175 Å²) is significantly larger than that of pro-TGFβ1 (1423 Å²) and pro-activin A (1696 Å²) and likely contributes to the enhanced stability of the complex, preventing dissociation of the prodomains after furin cleavage (Fig. 52B). In this way, the arm domain functions not only to improve solubility and stability of pro-myostatin, but also to lock the inhibitory helices of the forearm in place. The reduced arm interface of pro-activin A would render the furin processed complex more prone to dissociation, consistent with the results of our BLI experiments (Fig. 31) and the bioactivity of the pro-activin A complex [72]. Conversely, the pro-TGFβ1 complex is capable of compensating for a reduced arm domain interface through disulfide dimerisation of the prodomains, which would increase their functional affinity.
Figure 52: Arm domain interactions. A. Hydrogen-bonding at the interface of mature GF (orange) and arm domain (grey, blue or green) β-sheet interface. The pro-myostatin interface is stabilised by a more extensive hydrogen bonding network than pro-TGFβ1 and pro-activin A. B. Analysis of the arm domain interface with the mature GF/forearm. In the rotated view (bottom) the arm domain has been removed to expose the interaction surface of the mature GF and associated forearm (all residues within 4 Å of the arm domain are coloured red).
5.7 Conclusion

Analysis of the pro-myostatin crystal structure has illuminated a number of key structural characteristics which contribute to the latency of the pro-myostatin complex. Unexpectedly, pro-myostatin was found to adopt an open-armed, V-shaped topology reminiscent of non-latent pro-TGF\(\beta\) family members. This domain-architecture was confirmed in solution using SAXS and shown to have considerable conformational flexibility about the dimerisation disulfide. Furthermore, the pro-myostatin dimer forms a domain-swapped arrangement in which criss-cross connectivity provides stabilising tethers between the individual domains of the complex which may have implications for growth factor biosynthesis and activation.

The conserved forearm helices of the prodomain form critical latency conferring interactions with the hydrophobic surfaces of mature GF, disrupting the wrist helix of the putative type I receptor site, and occluding the type II receptor site. The globular arm domain forms a stabilising interface with the forearm and mature GF and has a larger buried surface than observed for other family members. This enhanced interface is likely to impede dissociation of the prodromains following furin cleavage, allowing the pro-complex to persist in a stable, latent form which requires further proteolysis by BMP1/TLD for full activation.

In conclusion, we show here that pro-myostatin achieves latency through a different mechanism to the canonical pro-TGF\(\beta\)1 complex and that analysis of overall prodomain topology (open versus closed) is not necessarily sufficient to predict whether or not a pro-complex will be latent. Pro-myostatin (and pro-GDF11 by homology) represents a new paradigm of structure-function relationship within the pro-TGF\(\beta\) superfamily, in which latency is associated with an open-armed conformation. Unlike pro-TGF\(\beta\)1 which utilises covalent linkage of the prodromains to form a high affinity pro-complex, myostatin relies on an enhanced non-covalent pro-mature interface for growth factor latency.
Polymorphisms & targeted mutagenesis
Loss-of-function mutations in the gene encoding myostatin are well known for their ability to yield hyper-muscular phenotypes in animals [92]. The gene encoding myostatin in double-muscled Belgian-blue cattle features an 11 base pair deletion which results in a severely truncated and non-functional form of the protein [92]. In other double-muscled breeds individual missense mutations are sufficient to confer hyper-muscular phenotypes, as is the case for Limousine and Piedmontese cattle which carry the myostatin mutations F94L and C313Y respectively [180, 181].

While most mutations described for myostatin result in a non-functional protein and increased musculature, it could be imagined that a mutation which destabilises the inhibitory prodomain without disrupting the mature growth factor would have the opposite effect. To date, 129 naturally occurring missense mutations have been documented in the sequences of human pro-myostatin (Ensembl genome assembly GRCh38.p10 accessed 05.06.17, Appendix 4). Of these mutations, 87 involve substitution of amino acids in the prodomain (Fig. 53).

We analysed the location and interactions of these residues in our pro-myostatin structure and ranked them in order of likeliness to affect the bioactivity of the pro-myostatin complex (Appendix 4).

Figure 53: Distribution of described missense variants in human pro-myostatin. Polymorphisms shown in red were selected for analysis in this study. Mutations responsible for hypertrophy in cattle are shown in blue. In Belgian-blue cattle an 11bp deletion frameshift results in a truncated product lacking the mature GF (D273∆), while the F94L and C313Y missense mutations are associated with double muscling in Limousine and Piedmontese breeds respectively.
6.1 Bioactivity of selected pro-myostatin variants

In order to extend our investigation of the molecular determinants of pro-myostatin latency, we designed a number of pro-myostatin variants in which we introduced naturally occurring structural polymorphisms, or made targeted substitutions of residues which we had identified as potentially important for latency of the pro-complex (Fig. 54).

Natural polymorphisms involving residues of the fastener motif (R65C, Y112H, H112R) were chosen for analysis, as these residues are presumed to form key latency conferring interactions in the pro-myostatin complex. Non-conserved mutation of the equivalent residues in pro-TGFβ1 has been shown to release latency of the pro-complex [82]. A non-natural R65A variant was also selected in order to abolish all side-chain functionality. In addition to fastener mutations, the natural variant A84G was chosen as this residue sits at the interface of the latency lasso and arm domain (Fig. 54). Trp203 from the prodomain arm also forms part of the forearm/arm interface, forming a hydrogen bond between the Trp203 indole nitrogen and the backbone of Lys83 (Fig. 54). Trp203 was substituted for Ala, His and Phe, in an attempt to disable the interaction without compromising the hydrophobic core of the arm domain. Finally, the known human polymorphic variant K153R, which has been associated with muscle and obesity related phenotypes, was chosen for analysis [182, 183]. Lys153 occupies a seemingly non-functional site in the prodomain arm (Fig. 54), but has been suggested to enhance furin processing of the precursor [184].

While choosing candidates for mutagenesis we decided to exclude those that had previously been, or were currently being investigated elsewhere. During the course of this study we became aware that the group of Prof. Thomas Thompson (University of Cincinnati) was undertaking an extensive mutagenesis study of pro-myostatin, also with the goal of uncovering the molecular requirements for latency. The Thompson group kindly disclosed the identity of their mutagenic targets, which they had selected based on analysis of the conserved forearm elements of the pro-TGFβ1 structure. These targets focussed predominantly on conserved hydrophobic residues of the α1 helix. Furthermore, the Thompson lab agreed to carry out mutagenesis experiments on our behalf, using the candidates we had selected.

The aforementioned natural polymorphisms, or structure-guided mutations were introduced into constructs of pro-myostatin for transient transfection in HEK293-(CAGA)$_{12}$ luciferase reporter cells by our collaborators in the Thompson lab. In this experimental setup, cells
stably transfected with a myostatin responsive (CAGA₁₂) firefly luciferase reporter gene were transiently transfected with plasmids bearing the various pro-myostatin variants (Materials and Methods section 8.9). Additionally, the expression plasmid also contained a gene encoding Renilla luciferase within an internal ribosome entry site (IRES), allowing constitutive co-expression of Renilla luciferase for normalisation of the firefly luciferase signal. Cells were simultaneously transfected with expression constructs of furin and human tolloid-like 2 (hTLL-2) proteases. Using this assay system we were able to analyse the effect of mutations on the relative level of myostatin signalling activity compared to the wild-type protein.

When cells were transfected with only low levels of hTLL-2, secreted pro-myostatin variants (including the wild-type) showed minimal signalling activity, indicating that cleavage of the BMP1/TLD site is a minimal requirement for activity of these variants (Fig. 55A). Increasing the amount of co-transfected hTLL-2 from 5 ng to 25 ng yielded a considerable increase in signalling response, with a number of variants deviating from the wild-type (Fig. 55B).

Substitution of Arg65 from the fastener motif to alanine led to a ≈1.5 fold increase in signalling activity over the wild-type, however the R65C variant displayed reduced activity
Figure 55: Bioactivity of pro-myostatin variants in cell-based luciferase reporter assay. A. Raw luciferase signalling responses for pro-myostatin variants with 5ng hTLL-2 transfection. B. Raw luciferase signalling responses for pro-myostatin variants with 25ng hTLL-2 transfection. C. Luciferase signalling responses for pro-myostatin variants with 5ng hTLL-2 transfection, normalised to the wild-type response. D. Luciferase signalling responses for pro-myostatin variants with 25ng hTLL-2 transfection, normalised to the wild-type response. E. Ratio of the normalised signalling responses in presence of 25 or 5 ng hTLL-2 transfection. Data shown are means ± SEM. All data shown in this figure was produced by our collaborators in the group of Prof. Thomas Thompson at the University of Cincinnati.
This loss of activity may be attributable to impaired secretion and bioavailability of the protein following introduction of a reactive unpaired cysteine. The naturally occurring fastener polymorphisms Y111H and H112R also significantly enhanced pro-myostatin activation, by $\approx 2.5$ and $\approx 1.8$-fold respectively (Fig. 55). These findings were in good agreement with those of the Thompson lab, in which disruption of the fastener by Y111A and H112A mutations also led to increased signalling activity [2]. This provides good evidence for a role of the fastener motif in maintaining latency of the pro-myostatin complex. Disruption of these stacking interactions appear to reduce latency of the furin cleaved complex, leading to elevated signalling responses.

There are a number of possible mechanistic explanations for the observed increased activity of the fastener variants. Disruption of the fastener interaction may increase dissociation rates of the inhibitory prodomain fragments following proteolysis at the furin and BMP1/TLD sites. Alternatively, specific polymorphisms may actually enhance proteolytic processing by increasing the accessibility and lability of the cleavage sites. The pro-myostatin fastener motif appears to hold the negatively charged BMP1/TLD site linker within proximity of the positively charged furin site, and in doing so may impose a restriction on the conformational flexibility required for efficient proteolysis at these sites (Fig. 50B). Interestingly, the fastener variants show proportionally higher activity than other variants with increased levels of hTLL-2 co-expression, supporting a hypothesis of enhanced proteolysis (Fig. 55E).

Unlike the fastener mutations, the natural K153R variant gave only a modest increase in myostatin signalling over the wild-type, while the conservative A84G variant had no noticeable effect on bioactivity (Fig. 55). Mutations of Trp203 gave considerably reduced signalling activity, suggesting loss of the tryptophan moiety is disruptive to protein folding, secretion and/or stability. This is reasonable given the location of Trp203 within a tightly packed hydrophobic interface. It should be noted that secreted protein levels were not quantified for the different variants tested here and it remains possible that altered expression accounts for differences in bioactivity.

6.2 Inhibition of myostatin by recombinant prodomain variants

To further probe the effect of the aforementioned mutations on pro-myostatin signalling, our collaborators in Cincinnati introduced the same mutations into an *E. coli* expression
The prodomain variants could be successfully expressed as soluble maltose-binding protein (MBP) fusion proteins and purified to homogeneity (Fig. 56). Mutations of Trp203 and the R65C variant gave very poorly soluble protein, consistent with the reduced endogenous signalling activity observed for these variants. As such, these variants were excluded from further analysis.

We then tested the ability of the purified prodomain variants to inhibit myostatin signalling in trans. Mature myostatin (0.25 nM) was titrated with increasing concentrations of the prodomain variants and then applied to SMAD2/3 responsive HEK293T cells. In this experimental setup the wild-type recombinant prodomain inhibited mature myostatin with an IC$_{50}$ of 0.9 nM (Fig. 56). Unexpectedly, none of the prodomain variants showed any obvious deviation in inhibitory capacity from the wild-type prodomain (Fig. 56, Table 8). This suggests that these mutations have no meaningful effect on latency of the pro-complex when it is reconstituted from its individual components in trans.

![Figure 56: Inhibition of myostatin signalling by recombinant prodomain variants. A. SDS-PAGE analysis of purified MBP-prodomain fusions. B. Inhibition of myostatin signalling by recombinant wild-type and variant prodomains in HEK293T dual-luciferase reporter assay. Increasing concentrations of prodomain variants were titrated into 0.25 nM mature myostatin. Data shown are means of triplicate measurements, normalised to 100% activity. Error bars show SD.](image-url)
Table 8: Inhibition of myostatin signalling by recombinant prodomain variants

<table>
<thead>
<tr>
<th>Variant</th>
<th>IC50 (nM)</th>
<th>IC50 (nM) 95 % CI*</th>
<th>Log [IC50] (nM) ± SEM**</th>
<th>Hill slope</th>
<th>Hill slope 95% CI*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>0.9</td>
<td>0.83 to 1.08</td>
<td>-0.02 ± 0.03</td>
<td>-2.01</td>
<td>-2.8 to -1.53</td>
</tr>
<tr>
<td>R65A</td>
<td>1.6</td>
<td>1.42 to 1.90</td>
<td>0.22 ± 0.03</td>
<td>-2.24</td>
<td>-2.93 to -1.74</td>
</tr>
<tr>
<td>A84G</td>
<td>1.3</td>
<td>1.18 to 1.54</td>
<td>0.13 ± 0.03</td>
<td>-1.99</td>
<td>-2.55 to -1.58</td>
</tr>
<tr>
<td>Y111H</td>
<td>1.1</td>
<td>0.88 to 1.14</td>
<td>0.001 ± 0.03</td>
<td>-1.86</td>
<td>-2.42 to -1.45</td>
</tr>
<tr>
<td>H112R</td>
<td>1.1</td>
<td>0.95 to 1.28</td>
<td>0.04 ± 0.03</td>
<td>-1.93</td>
<td>-2.60 to -1.46</td>
</tr>
<tr>
<td>K153R</td>
<td>1.2</td>
<td>1.01 to 1.42</td>
<td>0.08 ± 0.04</td>
<td>-1.95</td>
<td>-2.78 to -1.41</td>
</tr>
</tbody>
</table>

* CI Confidence interval
** SEM Standard error of the mean

It is possible that the reconstituted pro-complexes fail to reach the same conformational equilibrium as the endogenously synthesised proteins on the timescale of these experiments, masking the effect of any polymorphisms on inhibition when the prodomain is supplied in trans. Alternatively, if the primary mechanism by which these mutations affect signalling is via altering efficiency of proteolysis, this may not be resolved in the context of prodomain antagonism in trans.

Ideally we would have expressed and purified the endogenously biosynthesised pro-myostatin variants in order to analyse the effect of these mutations in a more physiological context, with proper quantification of protein concentration and controlled proteolytic cleavage at furin and BMP1/TLD sites. This would allow us to disentangle the effect of mutations on proteolysis of the precursor versus stability/dissociation of the pro-complex. Unfortunately this extended analysis proved infeasible given the time constraints of the project.

### 6.3 Conclusion

Having identified a number of potential latency conferring interactions from structural analysis of pro-myostatin, we sought to probe their functional relevance by introducing targeted mutations at these sites, or recapitulating naturally occurring missense mutations involving the residues of interest. The stacked resides of the fastener motif were of most interest as the equivalent residues had been identified as important for latency in related pro-TGFβ1.

Introduction of natural missense mutations at the fastener motif (Y111H, H112R) led to considerably increased signalling responses when these variants were expressed in SMAD2/3 responsive HEK293-CAGA12 luciferase reporter cells. The enhanced bioactivity of these
variants appears to be dependent on cleavage at the BMP1/TLD site, suggesting that these mutations do not liberate the pro-myostatin complex from latency altogether, but may increase the efficiency of proteolysis. This is plausible given the apparent conformational constraint imposed on the BMP1/TLD and furin sites by the fastener. Furthermore, disruption of the fastener may promote dissociation of the prodomain fragments following proteolysis.

The apparent reduction of latency, and increased bioactivity observed for fastener mutations is consistent with that previously reported for pro-TGF\(\beta\)1, and with the results of the Thompson lab pro-myostatin mutagenesis study. In addition to fastener mutations, the Thompson study identified the conserved hydrophobic residues I53 and I56 of the \(\alpha1\) helix as important for latency of pro-myostatin [2]. Mutation of these residues to alanine significantly increased bioactivity of the pro-complex, and furthermore, an I56A variant was bioactive even in the absence of BMP1/TLD site cleavage. These findings show that destabilisation of the inhibitory \(\alpha1\) interface is sufficient to liberate the growth factor from prodomain inhibition.

Curiously, recombinant MBP-prodomain fusions carrying the aforementioned mutations did not recapitulate the same pattern of effects seen for the endogenously produced variants when tested for inhibitory capacity in \textit{trans}. Y111H and H112R variant prodomains inhibited mature myostatin to the same degree as the wild-type protein, suggesting these mutations do not affect latency of the pro-complex when reconstituted in \textit{trans}. The Thompson lab similarly tested inhibition of myostatin by purified prodomain variants, and while their I53A variant was a significantly less potent antagonist, the inhibitory capacity of the remaining variants was not consistent with the pattern of effects observed in endogenous mutagenesis experiments [2].

Further investigation is required to fully explain the effect of these various mutations on the bioactivity of pro-myostatin, and to extend this analysis to include additional naturally occurring polymorphisms. It is possible that human polymorphisms that disrupt the pro-myostatin complex and liberate signalling (i.e. those affecting the fastener) may result in muscle-related phenotypes, however most data relating to polymorphic variation is acquired from anonymised studies and as such we are unable to correlate mutations with phenotypic information.
Discussion
The development and maintenance of multicellular life is dependent on strictly controlled signalling processes between and within cells. Loss of signal regulation leads to a breakdown of biological integrity, which often results in pathological outcomes. Perhaps the best known product of aberrant cell-signalling is cancer, a heterogenous and diverse spectrum of disease unified by a loss of ability to regulate cellular proliferation.

One way in which multicellular organisms achieve regulation of cell-signalling is by producing and storing signalling molecules as inactive precursors, which can be activated in a controlled manner following stimulation. Storage of an extracellular pool of latent or semi-latent growth factors additionally allows rapid signalling in response to stimuli without the need to first biosynthesise, process and secrete the proteins. Analogous to an armed missile in a silo, the secreted growth factor needs only the launch codes to deploy, rather than being built from scratch as and when required. In the TGFβ superfamily, the prodomains play a diversity of roles in extracellular regulation and storage, as reflected by their highly divergent amino acid sequences. This same divergence makes it particularly difficult to make accurate structural and functional predictions about these proteins.

As a strong negative regulator of muscle mass in animals, over-activation of myostatin is associated with pathological muscle atrophy. For this reason it is expected to be subject to strict regulatory measures and a controlled process of activation. In this doctoral study I sought to structurally and functionally characterise pro-myostatin, the major storage form of myostatin in skeletal muscle, with the goal of improving our understanding of its regulation and activation in the extracellular space.

### 7.1 Proposed model of pro-myostatin biosynthesis and activation

Based on the results presented in this work and elsewhere, we propose a general model for the biosynthesis and activation of myostatin as follows. The domain-swapped architecture observed for the pro-myostatin dimer suggests a stable protomeric structure could fold independently following translation, with the amphipathic forearm helices solubilising the hydrophobic GF domain (Fig. 57A, B). The exposed globular arm domain may then drive non-covalent dimerisation with an identical chain, followed by disulfide linkage of the mature GFs (Fig. 57C). The unprocessed pro-myostatin dimer is then secreted from the cell and immobilised in the ECM of muscle tissue through prodomain mediated interactions.
Figure 57: Model of pro-myostatin biosynthesis and activation. Stages of pro-myostatin biosynthesis (A-C), structural features which contribute to pro-myostatin latency (D), sequential proteolysis of furin and BMP1/TLD sites (E, F) and dissociation of prodomain fragments (G, H) to release the bioactive growth factor (I). Adapted from Cotton et al (2018), figure 8 [1].
The pro-myostatin precursor contains a number of structural features which collectively contribute to latency in the extracellular space and facilitate a controlled process of activation (Fig. 57D, 58). The prodomain α1 helix, with its conserved pattern of aliphatic residues, binds to the mature GF in such a way that displacement of the wrist helix destroys the type I receptor binding site. The α1 helix is further stabilised in place by the fastener motif which links the α1 helix to the β1 strand of the arm domain and is conserved in other latent pro-TGFβ superfamily members (Fig. 58). The pro-myostatin arm domain which sits atop the mature GF and associated forearm, forms a considerably larger interaction surface than observed for pro-TGFβ1 and pro-activin A. This extensive non-covalent interface stabilises the inhibitory grip of the forearm and prevents dissociation of the prodomains following furin cleavage.

Pro-myostatin is activated in the extracellular space by sequential proteolysis of the furin and BMP1/TLD sites (Fig. 57E-F). The furin site in pro-myostatin appears to be structurally constrained, perhaps slowing the rate of furin proteolysis with comparison to other family members that have more solvent accessible and labile sites. Prior to furin cleavage, the furin site and flanking elements may partially obscure the BMP1/TLD site or restrict conformation flexibility required for efficient proteolysis. While it is possible that extra-
cellular BMP1/TLD cleavage could occur before furin processing, as it stands there is no data to show that BMP1/TLD cleavage alone is sufficient for activation. The requirement of sequential proteolysis for activation is the topic of ongoing investigation in the Hyvönen lab.

Cleavage of the furin site does not result in a significant structural rearrangement, based on SAXS analysis of unprocessed and processed forms. Concordantly, furin cleavage of pro-myostatin does not alter the rate of hydrogen/deuterium (H/D) exchange, suggesting structural stability is unaffected by furin processing [3]. Furin cleavage alone is insufficient to fully activate pro-myostatin as the complex is stabilised by extended non-covalent interactions which prevent dissociation of the bioactive ligand. The prodomains form sufficiently strong non-covalent interactions to prevent release of the bioactive ligand and are even capable of outcompeting the high affinity antagonist follistatin. The domain-swapped architecture of the myostatin prodomains may confer additional functional affinity to the furin processed complex through a ‘pseudo-avidity’ effect. Prior to BMP1/TLD cleavage the prodomain arm is tethered to the forearm on the opposite side of the dimer, and so even if it were to temporarily dissociate from its interface with the mature GF, it would remain localised through linkage to the forearm which would promote re-binding to the GF.

Subsequent proteolysis at the BMP1/TLD site removes the final tether securing the arm domain which is then free to dissociate (Fig. 57G). Loss of the stabilising arm domain would reduce the affinity of the forearm, which dissociates last and releases the mature ligand (Fig. 57H). This model is consistent with the results of Le et al who observed increased rates of H/D exchange at the pro-mature interfaces and concentration dependent dissociation of prodomain fragments following cleavage of the BMP1/TLD site [3]. Sequential dissociation of the prodomain elements is additionally supported by the findings of Pepinsky et al (2017) who show the forearm fragment of pro-GDF11 can remain bound to the mature GF in the absence of the stabilising arm domain, without affecting ligand bioactivity [179].

Following dissociation of the forearm helices, the wrist helix is then free to spring back into place and re-establish the type I receptor site (Fig. 57I). The free mature GF is now competent for engagement of its receptors at the cell surface, but simultaneously becomes susceptible to inhibition by soluble antagonists in the extracellular space. Given the competition for myostatin by its prodomains, antagonists and receptors, it is likely that the
free ligand exists only transiently in nature. This would explain why the free ligand only represents a small proportion of the total myostatin detectable in both serum and muscle tissue [185].

The latency conferring features described above are mostly conserved in the sequence of pro-GDF11 and furthermore the overall structural topology of pro-GDF11 is similar to that of pro-myostatin [179]. As such our proposed model of biosynthesis and activation may also apply to pro-GDF11. For more distantly related members of the TGFβ superfamily, it is incredibly difficult to infer structure-function relationships based on the existing limited structural information available. For members with particularly divergent prodomain sequences (AMH and GDF15 for example), experimental structural data will be required to fully understand the role of these prodomains in biology.

### 7.2 Extracellular activation and biodistribution

Pro-myostatin exists extracellularly in various states; the unprocessed precursor, latent pro-myostatin complex, antagonist bound and free ligand. Traditional methods of quantification such as ELISA using antibodies raised against the mature ligand are typically not able to discriminate between these various forms and as such detected myostatin levels are not necessarily correlated with ‘active’ myostatin levels. This problem highlights the need for highly specific reagents, which are capable of discriminating between the various forms of the protein.

It is now established that unprocessed pro-myostatin is the primary species in muscle tissue, and that the majority of this protein is stored extracellularly in the interstitial space between muscle fibres and around peripheral nuclei (Fig. 59) [104]. There is evidence to suggest that while held in the ECM by LTBP5s, pro-myostatin is at least partially resistant to furin cleavage, suggesting the precursor is released from the ECM before it becomes a target for extracellular furin [64]. The mechanism by which LTBP5s retain pro-myostatin in the ECM and the processes which allow its release are thus far unknown and represent an intriguing avenue of future research.
Figure 59: Immunofluorescence of healthy mouse tibialis anterior muscle. Tibialis anterior cross-sections were probed with the antibody GDF8-086 (red) which specifically recognises pro-myostatin and the furin processed pro-complex, anti-laminin (green) and co-stained for nuclei with DAPI (blue). Pro-myostatin/pro-complex is stored exclusively in the extracellular space where it co-localises with laminin (an ECM marker) in the interstitial space between muscle fibres (arrow and arrowhead) and around peripheral myo-nuclei (asterix). Figure taken from Pirruccello-Straub et al 2018 [104].

In serum, the furin processed latent complex is thought to be the most abundant circulating species [64, 104]. Whether this represents protein that was processed by intracellular furin prior to secretion, or soluble furin in the extracellular space remains unknown. Using specific anti-pro/latent myostatin antibodies, researchers at Scholar Rock (Cambridge, MA) showed that extracellular pro-myostatin is increased in muscle tissue following glucocorticoid induced atrophy, while latent pro-myostatin complex in serum is decreased [104]. The authors rationalise this observation by suggesting that enhanced proteolytic processing of the pro-complex in atrophying muscle is responsible for the observed decrease in circulating pro-complex.

The relevance of these observations in other forms of pathological muscle atrophy are as of yet unknown, and changes in myostatin precursor activation and bio-distribution during atrophy may prove to be context dependent.
7.3 Targeting extracellular activation of pro-myostatin

The previously mentioned Scholar Rock study implicates extracellular activation of ECM deposited pro-myostatin as a key regulatory step in myostatin signalling during muscle atrophy [104]. As a result, the blockade of extracellular activation represents a novel and attractive therapeutic strategy for suppression of myostatin signalling in the context of muscle wasting disease. This could be achieved by either inhibiting precursor processing by furin and/or BMP1/TLD family proteases, or by stabilising the latent pro-complex to such a degree that it is unable to release the bioactive GF. This approach has the added advantage of targeting the divergent prodomains over the highly conserved mature GFs, thus improving chances of achieving truly specific inhibition of signalling.

This approach was taken by Pirruccello-Straub et al of Scholar Rock who developed a series of monoclonal antibodies which bind to unprocessed pro-myostatin and the latent pro-myostatin complex with high specificity and affinity ($K_d < 3$ nM) [104]. The lead therapeutic candidate SRK-015 was shown to suppress extracellular activation of pro-myostatin (but not pro-GDF11), by inhibiting proteolysis at the BMP1/TLD site. SRK-015 was able to successfully protect against corticosteroid induced muscle atrophy in mice, and also promoted muscle growth with concomitant increase in absolute muscle strength in healthy mice. Together, these encouraging results demonstrate that targeting extracellular activation of pro-myostatin is a viable concept for myostatin inhibition in vivo. It is possible that this approach could also be extended to other members of the pro-TGFβ superfamily, including the prototypical TGFβ isoforms. In the context of established cancer, TGFβ signalling contributes to immune suppression in the tumor microenvironment [186]. As such, inhibition of TGFβ signalling has been shown to improve infiltration of immune effectors and potentiate the immune response to cancer cells [187]. In combination with modern cancer immunotherapies, TGFβ inhibition has the potential to greatly improve treatment efficacy as demonstrated recently in a mouse model of colon cancer [188].

The crystal structure of pro-myostatin presented in this study will be of value to those who seek to suppress its activation in the extracellular space. This structural insight will allow molecular rationalisation of biological observations and should prove a useful resource for guiding the development of next generation anti-myostatin agents. In addition to suppression of activation by macromolecular binders à la Scholar Rock, stabilisation of latent...
pro-complex protein-protein interfaces with small molecules may offer an alternate approach. The structural insight resulting from this work will be a useful resource for any structure-based drug design efforts targeting pro-myostatin.

### 7.4 Prodomain derived inhibitors of myostatin signalling

The more traditional approach of targeting myostatin signalling through inhibition of the GF:receptor interaction remains a problematic strategy. The high-level structural conservation between mature growth factors of the TGFβ superfamily make it difficult to selectively inhibit specific family members without off-target effects.

One way in which to overcome this lack of target specificity is to adapt the naturally inhibitory prodomains, or parts thereof, as biological therapeutics. The prodomains have evolved to selectively recognise their GF ligands and therefore represent a useful starting point for development as potent and specific inhibitors. Using a recombinant myostatin prodomain (as IgG-Fc fusion), Bogdanovich et al (2005) were able to demonstrate improvement of functional metrics in mdx mice, above what had previously been achieved with myostatin neutralising antibodies [148].

A different approach was taken by Takayama and colleagues who generated a range of inhibitory synthetic peptides based on the α1 helix sequence of the myostatin prodomain [172]. Interestingly, this study identified W43-L66 (referred to as ‘peptide 7’) from the mouse myostatin α1 sequence as the best candidate for inhibition of myostatin signalling. The mouse α1 helix sequence differs from the human sequence by substitution of only two residues (K49R and S50Y), both of which were shown to confer additional inhibitory capacity over the human sequence.

Access to the crystal structure of human pro-myostatin now offers us the possibility to use structure-guided design in order to elaborate on this foundational work, in pursuit of better inhibitors. Linear peptides, like those described above, are inherently poor therapeutic molecules due to their susceptibility to proteolytic degradation and short half-life. Additionally, isolated linear peptides are generally unstructured in solution and the structural reorganisation required for target engagement incurs an entropic penalty that makes binding unfavourable [189]. For these reasons, peptides are often chemically modified in order to improve pharmacological properties and binding affinity. In collaboration with the group of
David Spring in the Department of Chemistry, we synthesised and tested both linear and chemically ‘stapled’ versions of the previously identified peptide 7. Chemical stapling involves macrocyclization of a linear peptide by covalently linking the side chains of two amino acids [189]. By constraining the helical conformation of the peptide, stapling strategies can significantly improve target affinity and peptide stability [190].

![A](image1)

![C](image2)

Figure 60: Prodomain α1 helix mimicking peptides as inhibitors of myostatin signalling. A. Sequences of synthetic peptides tested. Peptide 7 (Takayama et al 2015) is based on the mouse prodomain α1 helix and differs from human sequence by the two residues shown in pink [172]. B. Diagram of α1 helix of myostatin prodomain (PDB: 5NTU) showing the position of bis-triazole staple in i,i+7 configuration. C. Preliminary experiments show both linear and stapled peptides inhibit myostatin signalling in cell-based luciferase reporter assay.

### Table 9: Inhibition of mature myostatin signalling with stapled peptides

<table>
<thead>
<tr>
<th>Peptide</th>
<th>IC50 (nM)</th>
<th>IC50 (nM) 95% CI*</th>
<th>Log [IC50] (nM) ± SEM**</th>
<th>Hill slope</th>
<th>Hill slope 95% CI*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide 7</td>
<td>2171</td>
<td>1692 to 2778</td>
<td>3.337 ± 0.05</td>
<td>-1.49</td>
<td>-2.159 to -1.078</td>
</tr>
<tr>
<td>Peptide 7 stapled</td>
<td>973</td>
<td>827.4 to 1145</td>
<td>2.988 ± 0.03</td>
<td>-1.22</td>
<td>-1.491 to -1.014</td>
</tr>
</tbody>
</table>

* CI: Confidence interval  
** SEM: Standard error of the mean

Both the native peptide 7 sequence and a variant with azide containing non-natural amino acids at positions 48 and 55 were made by solid-phase synthesis in the Department of Chemistry (Fig. 60A). The diazido peptide was subsequently reacted with a dialkynyl compound to form a bis-triazole linkage between the non-natural amino acids using the ‘double-click’ chemistry methodology described in Lau et al 2015 [191]. In this so-called i, i+7 configuration, the linker connects positions separated by two full helical turns (Fig.
The position and configuration of stapling was chosen based on computational analysis of the pro-myostatin structure (courtesy of Dr. Yaw Sing Tan, Bioinformatics Institute, A*STAR, Singapore). Structural analysis suggests that in this configuration the triazole linker itself may form additional interactions with the mature GF and possibly increase target affinity (data not shown).

The purified peptides were tested for inhibition of myostatin signalling in our HEK293T cell-based luciferase assay. At concentrations greater than 10 µM cell toxicity was observed, with a clear change in cell morphology and dose-dependent reduction in constitutive Renilla luciferase expression (data not shown). The assay was optimised accordingly to utilise lower peptide concentrations at which toxicity was not observed.

Preliminary experiments show that both the linear and stapled peptides inhibit myostatin signalling in a dose-dependent manner (Fig. 60C). Linear peptide 7 inhibited signalling with a low micromolar IC\(_{50}\) value comparable to that described by Takayama et al (2017) for similar synthetic peptides (1 - 3.5 µM)[173]. Stapled peptide 7 showed a modest improvement in potency over the linear peptide (Table 9), however further experimentation and careful quantification of peptide concentration is required to validate these findings.

These preliminary results are encouraging, and demonstrate that stapled peptides can be used to inhibit myostatin signalling with equal or higher potency than has previously been described for linear peptides of a similar nature. The envisaged pharmacological benefits of peptide stapling make this an attractive avenue of investigation. Further iterations of structure-guided design will be required to continue the optimisation of these peptides.

7.5 Clinical validity of myostatin as a therapeutic target

Regardless of the pharmacological method by which inhibition of myostatin signalling is achieved, it remains unclear if myostatin represents a clinically meaningful target. Numerous studies to date have investigated the effect of myostatin inhibition on muscle mass and function in both pre-clinical models of muscle-related pathologies and in human clinical trials, with mixed results (reviewed here [155]).

Inhibition of post-natal myostatin signalling consistently promotes muscle growth through increased size (hypertrophy) of muscle fibres. Hypertrophy is often associated with an
increase in absolute strength, however in many cases the mass-corrected ‘specific strength’
is unimproved, or indeed worse i.e strength does not increase proportionally with mass. This
may be the result of increasing the cytoplasmic volume of muscle fibres without increasing
the number of nuclei which govern transcriptional control in the surrounding ‘myonuclear
domains’ (MNDs) [192].

There is considerable debate concerning the role of satellite cell activation and recruitment
in hypertrophy, with some studies suggesting that myostatin-blockade induced hypertrophy
is the result of satellite cell activation [122, 193]. Conflicting studies show that activation
and incorporation of satellite cells into existing myofibers is too low compared to the overall
increase in myofiber size to fully explain hypertrophy [194, 195]. In one study, satellite cell
recruitment was shown to occur only after the onset of hypertrophy in mice treated with
a soluble ActRIIb-Fc ligand trap [195]. This suggests that in some cases hypertrophy can
occur without satellite cell activation, and that myostatin exerts its function directly on
the myofibers. A multiple-phase model of hypertrophy seeks to reconcile these disparate
findings by supposing that enhanced protein synthesis drives an initial increase in muscle
fibre size up to a specific MND size threshold, after which recruitment of satellite cells
is required for additional growth without compromising functionality of the fibre [195].
Therapeutic strategies which promote the incorporation of satellite cells during the early
stages of hypertrophy may provide a means of maintaining nuclei number and improving
the functional output of enlarged muscle in the context of muscle wasting disease. The long
term effects of supra-physiological satellite cell activation remain a concern however as it is
possible that premature depletion of the quiescent satellite cell pool\textsuperscript{5} may exacerbate the
process of muscle degeneration as the disease progresses.

Perhaps the most important consideration to be made when developing myostatin targeting
therapeutics is identification of the appropriate clinical indication. In some cases, simply
increasing muscle mass may be sufficient to restore function or ameliorate symptoms
of disease. In others, without a concomitant improvement in functionality, muscle hyper-
trophy may confer no clinical benefit or worsen the condition. Pre-clinical and clinical
studies have investigated the effect of myostatin inhibition across a wide range of patholo-
gies including muscular dystrophies (Duchenne, Becker, Limb-girdle, facioscapulohumeral,\footnote{The natural decline in satellite cell activity over time and associated loss of regenerative potential, is thought to be responsible for age-related loss of skeletal muscle mass [196].}

141
dysferlin-deficient, laminin-deficient), neuro-muscular conditions (amyotrophic-lateral sclerosis, spinal-muscular atrophy), chronic disuse and age-related muscle decline (sarcopenia), cancer-cachexia, X-linked myotubular myopathy, chronic kidney disease, chronic obstructive pulmonary disorder, sporadic inclusion body myositis and even metabolic diseases including type I insulin-resistant diabetes [155]. The exceptionally mixed results of these studies suggest that myostatin inhibition does not provide a broadly applicable therapeutic strategy for the treatment of muscle-related pathologies and that clinical outcomes are dependent on the underlying disease mechanism.

From studies to date, it seems that myostatin inhibition is unlikely to find use as a stand-alone monotherapy, particularly for indications with a genetic basis in which muscle physiology is fundamentally altered. Myostatin blockade may however provide a clinically effective therapeutic approach for enhancing muscle function and alleviating symptoms in certain acquired indications, and/or as an adjunct to treatments which seek to correct the underlying pathological mechanism. Emerging gene therapy approaches show promise for the treatment of Duchenne muscular dystrophy, with encouraging pre-clinical findings [197], and clinical trials now underway (https://clinicaltrials.gov/ct2/show/NCT03375164 accessed July 2018).

The near exclusive expression of myostatin in skeletal muscle, and the target-specificity achievable by inhibiting extracellular activation of the precursor forms makes myostatin an excellent candidate for safe and selective therapeutic intervention, although efficacy remains a question.
Materials & methods
8.1 Cloning and expression of pro-myostatin constructs

The DNA sequence encoding full length human pro-myostatin (Uniprot 014793, residues 19-375) was amplified from IMAGE clone (30915216) by polymerase chain reaction (PCR) and cloned into pBAT4 or pHAT2 vectors, using traditional restriction enzyme methods. The protein sequence of all sequence verified expression constructs are listed in Appendix 1. All sequence modifications were made using a multi-step extended-overlap PCR protocol using Phusion polymerase according to the manufacturers guidelines. In the first PCR reaction, oligonucleotides containing the modified sequence and a region complementary to the template DNA were used to amplify modified 5’ and 3’ fragments of the template. In the second PCR reaction, these fragments, with overlapping ends, were used to template extension of the full-length construct. The sequences of oligos used for cloning are listed in Appendix 2. Chemically competent DH5α cells were used for all cloning and sub-cloning steps.

For the crystallography construct pHAT2-MSTNΔT43 the sequence Glu36-Cys42 was replaced with the TEV protease recognition sequence (ENLYFQGS), to allow in vitro truncation of the disordered N-terminus. Surface entropy reduction mutations (G319A, K320A, K217A, Q218A, E220A) were subsequently introduced into the pHAT2-MSTNΔT43 construct.

For generation of HRV-3C cleavable pro-myostatin construct (pHAT2-MSTN-3C), the native furin site and surrounding residues (Arg263-Leu270) were substituted with the HRV-3C protease recognition sequence (LEVLFQGP) to permit efficient cleavage of the prodomain in vitro. Retention of the native furin linker length was required for efficient refolding of the HRV-3C substituted construct.

All expression constructs were sequence verified prior to heat-shock transformation into competent BL21(DE3)+pUBS520 cells, which were then cultured overnight at 37 °C on LB agar (+ 100 µg/mL ampicillin & 25 µg/mL kanamycin). Single colonies were cultured in 2 mL LB media at 37 °C until OD_{600} 0.6-1.0 and then induced with 400 µM IPTG. After 3 hours at 37 °C, cells were harvested by centrifugation, lysed using bugbuster (Novagen) and assessed for protein expression by SDS-PAGE. A small amount of the uninduced culture was used to prepare glycerol stocks (8% glycerol) which were stored at −80 °C. All subsequent cultures for protein expression were inoculated from these glycerol stocks.
For large scale expression, cells were grown in 1L 2xYT media until $\text{OD}_{600}$ 0.8-1.0, and induced as above for 3 hours at $37^\circ C$. Cell pellets were harvested by centrifugation at 4000 g for 20 minutes and stored at $-20^\circ C$ until required.

For expression of seleno-methionine (SeMet) labelled protein, a metabolic suppression protocol utilising minimal media was used to minimise endogenous methionine production. Cells were supplemented with seleno-methionine prior to induction of protein expression. This protocol is a slightly modified version of that described by Van Duyne et al 1993 [198].

8.2 Inclusion body preparation and refolding

All pro-myostatin constructs used in this thesis were expressed exclusively as insoluble inclusion bodies in *E.coli*. Cell pellets were defrosted and resuspended in 50 mM Tris pH 8.0, 5 mM EDTA and 10 mM DTT, to a final volume of 30 mL per litre culture cell pellet. Resuspended cells were lysed by 4-5 passages through Emulsiflex C5 homogenizer (with addition of 0.5% Triton-X after passage 2), and then incubated with DNase (100 $\mu$L, 2 mg/mL) at room temperature (RT) for 30 min. The lysate was centrifuged (15000 g, 20 min) and the supernatant discarded. The inclusion body pellet was thoroughly resuspended in wash buffer 1 (50 mM tris pH 8.0, 5 mM EDTA, 10 mM DTT, 0.5% Triton-X) using a handheld homogenizer before centrifugation again (15000 g, 20 min). The inclusion bodies were washed as above twice more using wash buffer 2 (50 mM tris pH 8.0, 5 mM EDTA, 10 mM DTT, 1 M NaCl) and wash buffer 3 (50 mM tris pH 8.0, 5 mM EDTA, 10 mM DTT). A final centrifugation step was used to pellet the washed inclusion bodies.

Prepared inclusion bodies (from 1 L culture pellet) were resuspended in 5 mL TCEP (100 mM pH 7.0), and solubilised by addition of 15 mL solubilisation buffer (50 mM Tris pH 8.0, 8 M Gdn-HCl, 10 mM EDTA, 0.1 M Cystine). The solubilisation mixture was incubated at RT with shaking for 30-40 min, before clarification by centrifugation (15000 g, 20 min). The supernatant was carefully removed and filtered (0.22 $\mu$m, Sartorius) before being buffer exchanged into 6 M urea + 20 mM HCl\textsuperscript{6}. Solubilised and buffer exchanged protein was diluted to 1 mg/mL with the above urea solution, and used for subsequent refolding trials.

\textsuperscript{6}Urea spontaneously degrades to generate cyanate ions in solution, which can lead to carbamylation of protein amines. In order to avoid unwanted carbamylation, all urea solutions are deionised by anion exchange prior to use, and acidified to drive the reaction equilibrium toward formation of urea instead of isocyanic acid.
For small-scale refolding trials, 100 µL of solubilised protein was rapidly mixed (by pipetting) with 900 µL of ice-cold refolding buffer (supplemented with 2 mM cysteine and 0.2 mM cystine) in 96-well deep well block format. A number of other cysteine/cystine ratios were also tested, but no obvious improvement in refolding efficiency was observed. Initially, 24 refolding buffers from our in-house screen were tested (Table 1). Refolding mixtures were incubated at 4°C for 7 days and then analysed by non-reducing SDS-PAGE. For large scale refolding of all pro-myostatin constructs, 100 mL of solubilised protein (1 mg/mL) was rapidly diluted 1:10, with vigorous stirring, into cold refolding buffer (100 mM Tris pH 9.0, 1 M pyridinium propyl sulfobetaine (PPS), 0.5 mM EDTA, 2 mM cysteine and 0.2 mM cystine) which was identified from the above screen.

8.3 Purification of pro-myostatin constructs

All pro-myostatin constructs were purified according to the following general protocol, however minor modifications to the protocol were required in some cases. These changes are detailed where necessary.

One litre batches of refolding solution were filtered (0.65 µm cartridge, Sartopure) and loaded overnight (1 mL/min) onto a 10 mL Source 15Q anion exchange column (GE Healthcare) equilibrated with 50 mM Tris pH 9.0. The column was then washed with five column volumes (CVs) of the equilibration buffer, before elution of the bound protein with a linear gradient of 0-100% elution buffer (50 mM Tris pH 9.0, 1 M NaCl) over 20 column volumes. For pHAT2-MSTN-3C this initial anion exchange capture step was completed using 5 mL HiTrap Q HP (GE Heathcare) column, instead of Source 15Q.

For crystallographic constructs with TEV cleavable N-termini, fractions pooled from anion exchange were buffer exchanged to 20 mM Tris pH 8.0, 150 mM NaCl and incubated with TEV protease (200 µL, 2 mg/mL) overnight at 4°C. TEV protease was also buffer exchanged prior to use, to remove reducing agent present in the storage buffer. The TEV cleavage mixture was incubated (1 hour, 4°C) with PureCube Ni-NTA resin (Cube Biotech, Germany) to remove TEV protease and the cleaved N-terminal fragment. The flow-through containing the N-terminally truncated pro-myostatin construct was collected.

All pro-myostatin constructs were purified further by size-exclusion chromatography (SEC). Pooled fractions from ion-exchange capture were concentrated by centrifugation (10 kDA
Amicon Ultra, MilliporeSigma) and loaded onto HiLoad superdex 200 16/60 gel filtration column (GE Healthcare) equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl. The peak fractions were assessed for purity by reduced and non-reduced SDS-PAGE (12% acrylamide).

To generate the pro-myostatin complex *in vitro*, purified pHAT2-MSTN-3C was incubated with GST-HRV-3C protease (at 4:1 mass ratio) in 50 mM Tris pH 8.0, 150 mM NaCl for up to 3 days at 4°C. The protease was buffer exchanged (PD10 column, GE Healthcare) into the above buffer prior to use, to remove reducing agent present in the storage buffer. The HRV-3C processed pro-myostatin complex was further purified by SEC as above. Co-elution of the prodomains and mature GF was indicative of stable complex formation following processing by HRV-3C. Finally, the pooled fractions from SEC were incubated with PureCube glutathione agarose (Cube Biotech, Germany) to remove the residual GST-tagged protease.

The mature myostatin GF dimer could be purified from the HRV-3C cleaved complex by reversed-phase chromatography (RPC). 1-3 mg batches of the purified pro-myostatin complex were treated with acetonitrile (ACN) and trifluoroacetic acid (TFA) to final concentrations of 10% and 0.1% respectively. This mixture was then loaded onto ACE C8 300 4.6 x 250 mm RPC column, equilibrated with 10% ACN and 0.1% TFA. The column was washed with 5 CVs of the equilibration buffer, and the protein eluted with a linear gradient 0-100% elution buffer (90% ACN, 0.1% TFA) over 20 CVs. Peak fractions were assessed by SDS-PAGE, and those containing the mature GF dimer (24 kDa) were pooled and dried to powder by centrifugal evaporation. Mature myostatin is extremely insoluble in physiological buffers, and so was resuspended in 10 mM HCl prior to use.

All pro-myostatin constructs were shown to have the expected molecular mass by matrix-assisted laser desorption ionisation (MALDI) mass-spectrometry\(^7\) (data not shown). The concentration of purified protein was carefully quantified spectrophotometrically by absorbance at \(A_{280}\) nm using calculated extinction coefficients (ExPASy protparam) with the assumption that all cysteines form disulfides. All protein concentrations were calculated from the average absorbance of triplicate measurements made immediately prior to use.

\(^7\)All mass-spectrometry was performed by the Protein and Nucleic acid Chemistry (PNAC) facility within the Department of Biochemistry, University of Cambridge.
8.4 Expression and purification of MBP-prodomain fusions

A cDNA fragment encoding the wild-type human myostatin prodomain (Uniprot 014793, residues 24-262) was cloned into pET28a vector to give a final construct with N-terminal 6xHis-tag and MBP-fusion. The four cysteines within the prodomain sequence were mutated to serine to improve expression and solubility of the prodomain fusion, and furthermore a number of surface entropy reduction mutation were introduced into MBP to improve stability as per Moon et al (2010) [199]. For prodomain variants, specific mutations were made using QuickChange PCR with PfuUltra II (600670, Agilent technologies) according to manufacturer’s instructions. MBP-prodomain constructs were expressed in Rosetta(DE3)pLacI cells, which were grown in 2xYT media (34 µg/mL chloramphenicol, 25 µg/mL kanamycin) until OD$_{600}$ 0.6-0.8 then overnight at 18°C following induction with 400 µM IPTG. Cell pellets were harvested by centrifugation (4000 g, 20 min) and resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 10 mM imidazole) with the addition of protease inhibitor (cOmplete mini, EDTA-free, Roche). Resuspended cells were lysed with 3-4 passages through Emulsiflex C5 homogenizer and then clarified by centrifugation (15000 g, 20 min). The supernatant containing soluble protein was filtered and incubated with 1 mL PureCube Ni-NTA agarose for 1 hour at 4°C. The resin was collected in a gravity flow column and washed with 5 CVs of wash buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole), before elution of 6xHis-tagged protein with 0.5 mL fractions of elution buffer (wash buffer + 500 mM imidazole).

MBP-prodomain fusions from Ni-NTA capture step were further purified by size-exclusion chromatography on a HiLoad Superdex 200 16/60 column (GE Healthcare), equilibrated with 20 mM Tris pH 8.0, 500 mM NaCl. The peak fractions were assessed for purity by SDS-PAGE (12%) analysis.

8.5 Pro-myostatin crystallisation

Purified pro-myostatin constructs were concentrated to 10 mg/mL and screened for crystallisation in a range of commercial 96-well screens (Molecular Dimensions, Qiagen, Rigaku).

---

8MBP-prodomain construct cloning and expression was done by our collaborators at the University of Cincinnati who had previously developed this protocol. We received the frozen cell pellets, and completed purification of the proteins.
Reagents, USA). Sitting drops with 150 nL reservoir and either 150 nL (1:1) or 300 nL (1:2) protein solution were dispensed using a Mosquito crystallisation robot (TPP Labtech). Crystallisation screens were incubated at 19°C, and periodically imaged by RockImager (Formulatrix). Preliminary crystallisation conditions were optimised further by preparing customised 96-well screens with Dragonfly robotics (TPP Labtech).

Unprocessed human pro-myostatin with N-terminal truncation (pHAT2-MSTN∆T43) crystallised optimally in 0.1 M Na acetate (pH 4.2) with 1 M ammonium phosphate, 6-10 mg/mL protein with a 1:2 ratio of reservoir to protein in the drop. Crystals reached maximum size after 7 days. Selenomethione derivative pHAT2-MSTN∆T43 crystallised under identical conditions to the native protein.

The optimised crystallography construct, with N-terminal truncation and five surface-entropy reduction mutations G319A, K320A, K217A, Q218A, E220A (pHAT2-MSTN∆T43\textsubscript{SERp}) crystallised overnight in 10% PEG 6000, 0.1 M HEPES pH 7.0, 6-10 mg/mL protein and a 1:2 ratio of reservoir to protein in the drop. Crystals were transferred to a drop of reservoir solution containing 26% ethylene glycol for cryo-protection before flash freezing in liquid nitrogen.

8.6 Data collection and structure determination

Throughout the course of this study, diffraction data from pro-myostatin crystals was collected at the following synchrotrons: Diamond Light Source (Oxford, UK), ESRF (Grenoble, France) and Soleil (Gif-sur-Yvette, France). Data collected at all of these facilities greatly contributed to the development of our crystallography system for pro-myostatin, however the data which were ultimately used for determination of both low and high resolution structures was collected at Diamond Light Source on beamline I-03.

For SAD phasing of SeMet derived pHAT2-MSTN∆T43, diffraction data was collected for multiple crystals at the selenium-peak wavelength (0.97970 Å). All datasets were processed individually with autoPROC (XDS, Pointless, Aimless, CCP4 Suite) and the seven best were selected for merging based on quality of diffraction and anomalous signal, similarity of unit cell parameters and the resulting $R_{\text{merge}}$ [200–203]. Multiple datasets were merged by autoPROC/Aimless using a resolution cutoff of 4.19 Å, and then Autosol (Phenix) was used for SAD phasing of the merged dataset [204, 205]. Coot was used to build the atomic
model into electron-density maps, and phenix.refine and autoBUSTER (Global Phasing Ltd, https://www.globalphasing.com) were used for iterative cycles of model refinement against the experimental data.

For determination of the high resolution structure, diffraction data was collected from a single cryo-cooled crystal of surface modified pro-myostatin (pHAT2-MSTN\textDelta T43\text{SERp}) and processed to 2.59 Å with autoPROC. One symmetrical half of the partially refined low resolution structure was used as a search model for molecular replacement of the high resolution dataset with Phaser [206]. Structure building and refinement was completed as above.

Data collection, processing and refinement statistics for both high and low resolution structures are listed in Table 6. Both structures were deposited in the Protein Data Bank (PDB) with accession codes 5NTU (high resolution) and 5NXS (low resolution). All structural figures in this thesis were prepared using UCSF Chimera (Version 1.12).

8.7 SAXS data collection and analysis

Small-angle X-ray scattering data was collected at Soleil Synchrotron (Gif-sur-Yvette, France) on SWING biosaxs beamline using a SEC-coupled SAXS system. Samples of purified pro-myostatin (40 \mu L, 7.5 mg/mL) and HRV-3C processed pro-myostatin complex (20 \mu L, 9 mg/mL) were injected (0.075 mL/min) onto a Superdex 200 Increase 10/300 column (GE Healthcare) equilibrated with 20 mM Tris pH 8.0, 150 mM NaCl. Scattering data was collected as the eluting protein passed through a quartz flow-cell. Sample was maintained at constant temperature of 293 K for the duration of data collection. 250 frames of scattering data were collected (at 12000 eV beamline energy) for each sample, with frame duration of 0.75 s and 0.25 s spacing between frames. Scattered X-rays were detected by an PCCD170170 detector (AVIEX) with a sample-detector distance of 1784 mm, allowing data collection over an angular range (q) of 0.006-0.613 Å.

Soleil Synchrotron in-house software (FOXTROT 3.4.1) was used for initial data processing steps, including the averaging of frames collected across the elution peaks (based on \( R_g \) value), and buffer subtraction. Averaged and buffer corrected data was plot as scattered intensity (\( I \)) against scattering angle (\( q \)), for analysis of the radius of gyration (\( R_g \)) and forward scattering \( I(0) \) based on Guinier approximation using SCATTER 3.0 [207]. Linearity in the Guinier plots at low \( q \) indicates the data was not compromised by the presence of
high MW aggregates. DATGNOM (ATSAS package, EMBL) was used to calculate pairwise-distance distributions \( P(r) \) from which the maximum particle dimension \( (D_{\text{max}}) \) could be estimated [208]. For \( P(r) \) calculation, datasets were truncated to exclude data collected at angles greater than 0.2 and \( P(r) \) distribution functions were shown to approach zero smoothly at \( D_{\text{max}} \), indicative of a reasonable \( P(r) \) solution.

DAMMIN (ATSAS, EMBL) was used for \textit{ab initio} modelling of molecular envelopes of pro-myostatin and the HRV-3C processed complex. Thirty-four independent models were calculated (with assumption of P2 symmetry) for each dataset. This ensemble of models was then averaged with DAMAVER (ATSAS, EMBL) and filtered with DAMFILT (ATSAS, EMBL) to produce the final model. The high resolution pro-myostatin crystal structure was docked manually into the envelope and visualised with UCSF Chimera.

### 8.8 Dual-luciferase reporter assay

The bioactivity of the various forms of myostatin produced in this study was assessed using a dual-luciferase reporter assay in cultured HEK293T cells\(^9\). The general protocol is as follows:

Cells were cultured as per standard protocols in 96-well flat bottom culture plates with Dulbeccos Modified Eagle Medium (DMEM, Life Technologies) + 10% foetal bovine serum (FBS, Life Technologies). 96-well plates were seeded with 8000 cells per well (100 \( \mu \)L media per well) and cultured at 37 °C with 5% CO2 until approximately 80% confluent (typically 24 hours). After 24 hours, 33 ng pGL3-CAGA (with SMAD2/3 responsive firefly luciferase reporter) and 17 ng pRL-SV40 (with constitutively expressed \textit{Renilla} luciferase) plasmids were mixed with 0.2 \( \mu \)L FuGENE HD transfection reagent (Promega) and added to each well for transient transfection of cells. 24 hours following transfection, the medium was aspirated and replaced with 100 \( \mu \)L DMEM (+0.5% FBS) containing an appropriate dilution of the protein being tested for bioactivity. Each concentration point was replicated three times per experiment, and the outermost wells of the plate were filled with media only, to avoid unwanted edge effects arising from increased evaporation in these wells.

For myostatin inhibition assays (with follistatin, MBP-prodomains and synthetic peptides),

---

\(^9\)HEK293T cells were a generous gift from Dr Trevor Littlewood, Department of Biochemistry, Uni. of Cambridge.
serial dilutions of the antagonist were prepared by diluting stock solutions into DMEM (0.5% FBS) containing 0.25 nM mature myostatin GF (a concentrated stock of myostatin GF was first prepared by resuspending lyophilised myostatin in 10 mM HCl). The myostatin:antagonist mixtures were incubated at room temperature for 30 min, before applying to cells as usual. Follistatin-288 used in this study was kindly prepared by Katharina Ravn in the Hyvönen group, according to the protocol of Harrington et al (2006) [107].

Cells were incubated with the proteins of interest overnight, then washed once with PBS (this must be done carefully as HEK293T cells are weakly adherent) and lysed with Passive Lysis Buffer (20 µL per well, Promega) and shaking at room temperature for 30 min. An aliquot of 4 µL cell lysate from each well was transferred to a black flat-bottomed half area 96-well plate for measurement of luciferase activity. Using a PHERAsstar microplate reader (BMG LABTECH, Germany), 15 µL of firefly luciferase substrate (LAR II, Promega) was injected per well, and the resulting luminescence then measured for a duration of 2 s after an initial delay of 4 s. The firefly luminescence reaction was quenched by addition of 15 µL Stop & Glo reagent (Promega) which contains an inhibitor of firefly luciferase, and the substrate for Renilla luciferase. The Renilla luminescence was measured as for firefly luciferase above. Myostatin dependent firefly luciferase activity was normalised against the constitutively expressed Renilla luciferase activity, to adjust for variation in cell number and viability between wells. The ratio of firefly:Renilla luciferase luminescence was plot as a function of agonist (or antagonist) concentration, and fit to non-linear dose-response (variable slope) models for EC\textsubscript{50} and IC\textsubscript{50} determination. GraphPad Prism 7 was used for all dose-response curve fitting and data analysis.

8.9 Bioactivity of pro-myostatin variants in HEK293 cells

This protocol for analysing the effect of mutations on the bioactivity of pro-myostatin was developed, and executed by our collaborators in the lab of Professor Thomas Thompson at the University of Cincinnati.

For assessment of pro-myostatin variant signalling activity, a HEK293 cell line was established with a stably transfected SMAD-responsive CAGA\textsubscript{12} firefly luciferase reporter gene. These cells were seeded in 96-well poly-D-Lys coated plates (Greiner Bio-one, GmbH Germany) at 20000 cells per well (100 µL media per well), and cultured until 75-85% confluent.
For transfection, 25 ng pSF-CMV-FMDV IRES-Rluc bearing pro-myostatin constructs and Renilla luciferase within an internal ribosome entry site (IRES), 50 ng pcDNA4 encoding furin protease and 5 or 25 ng pcDNA3 5 encoding human tolloid-like 2 (hTLL2) were combined with TransIT-LT1 reagent (Mirus Bio LLC) in OPTI-MEM reduced serum media (Life Technologies). Transfection mixture (25 µL) was added directly to the growth media in each well, and incubated for 6 hours at 37°C with 5% CO2.

Following transfection period, media was exchanged with 100 µL serum free media and incubated for a further 30 hours before lysis with 20 µL Passive Lysis Buffer (Promega) per well (with shaking 800 rpm, room temperature, 20 min). Cell lysates were transferred to a fresh 96-well plate (black) and 40 µL firefly substrate LAR II (Promega) added per well, for measurement of firefly luminescence using a Synergy H1 Hybrid Plate Reader (BioTek). The firefly luciferase reaction was quenched and Renilla luminesce measured following addition of 40 µL of Stop & Glo substrate (Promega). Firefly luciferase activity was normalised against Renilla luciferase activity. Firefly/Renilla luminescence measurements were repeated in triplicate for each pro-myostatin variant, and the entire experiment repeated three times.

8.10 Multi-angle light scattering (MALs)

Size-exclusion chromatography coupled MALS experiments were performed using a Superdex 200 Increase 10/300 column (GE Healthcare) in conjunction with a DAWN HELEOS II light scattering instrument and Optilab T-rEX refractive index detector. Instrumentation was thoroughly pre-equilibrated with running buffer (20 mM Tris pH 8.0, 150 mM NaCl) and calibrated with a sample of bovine serum albumin (BSA, Thermo Scientific). Samples of pro-myostatin and the HRV-3C cleaved complex (100 µL at 1-1.5 mg/mL) were injected onto the SEC column (0.5 ml/min) and analysed by MALS following elution from the column. The resulting experimental data was analysed using ASTRA software (Wyatt Technology).

8.11 Biolayer interferometry (BLI)

Dissociation of pro-myostatin and pro-activin A complexes was assessed by biolayer interferometry (BLI) using a ForteBio Octet RED96 instrument (Pall ForteBio, USA). Unprocessed and HRV-3C processed pro-myostatin and pro-activin A were immobilised, through
N-terminal 6xHis-tags on anti-penta His (HIS1K) biosensors by dipping sensors into solutions of each protein (20 µg/mL) for 90 seconds loading time. Biosensors with immobilised proteins were then dipped into kinetic buffer (PBS, 0.1% BSA, 0.02% Tween-20) with or without additional 500 nM Follistatin-288. Dissociation of material from the biosensor was monitored by the change in biolayer thickness (nm) at the sensor tip, over a period of 900 s. Follistatin-288 used in this study was kindly prepared by Katharina Ravn in the Hyvönen group, according to the protocol of Harrington et al (2006) [107].

8.12 Peptide synthesis

Solid-phase peptide synthesis was carried out by our collaborators in the group of Professor David Spring in the Department of Chemistry, Cambridge. Fmoc protected amino acids were coupled on Rink amide MBHA resin (Merck), either manually or using a CEM Liberty Automated Microwave Peptide Synthesiser. Solution-phase double-click stapling of diazido peptides was completed as per Lau et al 2015, by reaction of a dialkynyl linker under Cu(I) catalysis [191]. Peptides were purified by semi-preparative HPLC.

Figure 61: Reaction scheme for Cu(I) catalysed double-click stapling on diazido i7-48-55 peptide. Figure courtesy of Hongjia Zhu, Department of Chemistry.
References


References


<table>
<thead>
<tr>
<th>Reference</th>
<th>Title</th>
<th>Journal/Volume</th>
<th>Year</th>
<th>Pages</th>
</tr>
</thead>
</table>


Cotton, T.R

References


[207] R. R. *Scatter 3.0, Diamond Light Source, Didcot, UK.*

Appendices
9.1 Appendix 1: Expression construct protein sequences

First generation expression construct sequences:

Vector derived sequences are underlined and modified residues are shown in bold.

**pBAT4-MSTN.001**

<table>
<thead>
<tr>
<th>gpVDLNENSEQKENVEKEGLCNACTWRQNTSSRIAIIQILSKLRLLETAPNISDKVIRQLLPKAPPLRELQIDYDQVRDDSSDGSLLEDDDYHATTETIITMTESDFLMQVGDGPCCFKFSSKIQYNNKVPK VQAQLIIYLRPVEPTTTTFVQILRLIKPMDGTTRYGIRSLKLDMNPDPGTGIWQSIDVKTGLQNLKQVESNLGIEIKALDENGHLAVTFPGPEGDNFLEKVDTPTKRSSRDQFLDCLDHEHSTESRCCRCYPYPLTVDEFGWYIAPKRYKANCYSGCECEFVFLQKYPHTLVLHQAQPGRSAGPCCTPTKMSPINLYFNGKEQIIYGIKIPAMVVDRCGCS</th>
</tr>
</thead>
</table>

**pHAT2-MSTN.001**

<table>
<thead>
<tr>
<th>gpVDLNENSEQKENVEKEGLCNACTWRQNTSSRIAIIQILSKLRLLETAPNISDKVIRQLLPKAPPLRELQIDYDQVRDDSSDGSLLEDDDYHATTETIITMTESDFLMQVGDGPCCFKFSSKIQYNNKVPK VQAQLIIYLRPVEPTTTTFVQILRLIKPMDGTTRYGIRSLKLDMNPDPGTGIWQSIDVKTGLQNLKQVESNLGIEIKALDENGHLAVTFPGPEGDNFLEKVDTPTKRSSRDQFLDCLDHEHSTESRCCRCYPYPLTVDEFGWYIAPKRYKANCYSGCECEFVFLQKYPHTLVLHQAQPGRSAGPCCTPTKMSPINLYFNGKEQIIYGIKIPAMVVDRCGCS</th>
</tr>
</thead>
</table>

**pHAT2-MSTN.002**

<table>
<thead>
<tr>
<th>gpVDLNENSEQKENVEKEGLCNACTWRQNTSSRIAIIQILSKLRLLETAPNISDKVIRQLLPKAPPLRELQIDYDQVRDDSSDGSLLEDDDYHATTETIITMTESDFLMQVGDGPCCFKFSSKIQYNNKVPK VQAQLIIYLRPVEPTTTTFVQILRLIKPMDGTTRYGIRSLKLDMNPDPGTGIWQSIDVKTGLQNLKQVESNLGIEIKALDENGHLAVTFPGPEGDNFLEKVDTPTKRSSRDQFLDCLDHEHSTESRCCRCYPYPLTVDEFGWYIAPKRYKANCYSGCECEFVFLQKYPHTLVLHQAQPGRSAGPCCTPTKMSPINLYFNGKEQIIYGIKIPAMVVDRCGCS</th>
</tr>
</thead>
</table>

**pHAT2-MSTN-3C**

<table>
<thead>
<tr>
<th>gpVDLNENSEQKENVEKEGLCNACTWRQNTSSRIAIIQILSKLRLLETAPNISDKVIRQLLPKAPPLRELQIDYDQVRDDSSDGSLLEDDDYHATTETIITMTESDFLMQVGDGPCCFKFSSKIQYNNKVPK VQAQLIIYLRPVEPTTTTFVQILRLIKPMDGTTRYGIRSLKLDMNPDPGTGIWQSIDVKTGLQNLKQVESNLGIEIKALDENGHLAVTFPGPEGDNFLEKVDTPTKRSSRDQFLDCLDHEHSTESRCCRCYPYPLTVDEFGWYIAPKRYKANCYSGCECEFVFLQKYPHTLVLHQAQPGRSAGPCCTPTKMSPINLYFNGKEQIIYGIKIPAMVVDRCGCS</th>
</tr>
</thead>
</table>

Crystallography expression constructs:

Vector derived sequences are underlined and modified residues are shown in bold.

**pHAT2-MSTNΔG37**

<table>
<thead>
<tr>
<th>gpVDLNENSEQKENVEKEGLCNACTWRQNTSSRIAIIQILSKLRLLETAPNISDKVIRQLLPKAPPLRELQIDYDQVRDDSSDGSLLEDDDYHATTETIITMTESDFLMQVGDGPCCFKFSSKIQYNNKVPK VQAQLIIYLRPVEPTTTTFVQILRLIKPMDGTTRYGIRSLKLDMNPDPGTGIWQSIDVKTGLQNLKQVESNLGIEIKALDENGHLAVTFPGPEGDNFLEKVDTPTKRSSRDQFLDCLDHEHSTESRCCRCYPYPLTVDEFGWYIAPKRYKANCYSGCECEFVFLQKYPHTLVLHQAQPGRSAGPCCTPTKMSPINLYFNGKEQIIYGIKIPAMVVDRCGCS</th>
</tr>
</thead>
</table>
Cotton, T.R
Appendices

KYPHTHLVHQANPRGSAGPCCTPTKMSPINMLYFNGKEQIIYGKIPAMVVDCGCS

pHAT2-MSTNΔS50

MGSSRIEAIKIQILSLRLETAPNISDVIQQLLPPKAPPPLRELIDQYDVRDDSDGSLEDDDHYATTETIT
IMPTESDFLMQVDGBPCCFFKFSKKIQYNKVKQLWILRPVETTTTVQILRLIPLKMDGTRYTGI
RSLLKDLMNPGTGWQSIDVTVQNLWKQESNLGIEIKALDENGHDLATFPGPEDGLNPFLLEVKTDT
PKRSRDRFGLDCEHSTESCRRYPALLEAFGDWIIAPKRYKANYCSGECIFVFLQKYPHTHLVHQAN
PRGSAGPCCTPTKMSPINMLYFNGKEQIIYGKIPAMVVDCGCS

pHAT2-MSTNΔN71

MGISDKVIRQLLPKAPPPLRELIDQYDVRDDSDGSLEDDDHYATTETITIMPTESDFLMQVDGBKCCF
FKFSKKIQYNKVKQLWILRPVETTTTVQILRLIPLKMDGTRYTGIIRSLKLDNPGTGIWQSIDVKT
VLQNLWKQESNLGIEIKALDENGHDLATFPGPEDGLNPFLLEVKTDPKRSRDFGLDCEHSTESC
CRYPLTLVDEAFGDWIIAPKRYKANYCSGECIFVFLQKYPHTHLVHQANPRGSAGPCCTPTKMSPINMLY
FNGKEQIIYGKIPAMVVDCGCS

pHAT2-MSTNΔA43

MSHSHHHSMDIEFVDEGSGPVDLNENSEQKENVEKENLYFQGSTWRQNTKSSIEAIKIQLSLKLET
APNISDVIQQLLPPKAPPPLRELIDQYDVRDDSDGSLEDDDHYATTETITIMPTESDFLMQVDGBKCCF
FKFSKKIQYNKVKQLWILRPVETTTTVQILRLIPLKMDGTRYTGIIRSLKLDNPGTGIWQSIDVKT
VLQNLWKQESNLGIEIKALDENGHDLATFPGPEDGLNPFLLEVKTDPKRSRDFGLDCEHSTESC
CRYPLTLVDEAFGDWIIAPKRYKANYCSGECIFVFLQKYPHTHLVHQANPRGSAGPCCTPTKMSPINMLY
FNGKEQIIYGKIPAMVVDCGCS

pHAT2-MSTNΔA43-B4/5del

MSHSHHHSMDIEFVDEGSGPVDLNENSEQKENVEKENLYFQGSTWRQNTKSSIEAIKIQLSLKLET
APNISDVIQQLLPPKAPPPLRELIDQYDVRDDSDGSLEDDDHYATTETITIMPTESDFLMQVDGBKCCF
FKFSKKIQYNKVKQLWILRPVETTTTVQILRLIPLKMDGTRYTGIIRSLKLDNPGTGIWQSIDVKT
VLQNLWKQESNLGIEIKALDENGHDLATFPGPEDGLNPFLLEVKTDPKRSRDFGLDCEHSTESC
CRYPLTLVDEAFGDWIIAPKRYKANYCSGECIFVFLQKYPHTHLVHQANPRGSAGPCCTPTKMSPINMLY
FNGKEQIIYGKIPAMVVDCGCS

pHAT2-MSTNΔA43-SERp (cluster 1)

MSHSHHHSMDIEFVDEGSGPVDLNENSEQKENVEKENLYFQGSTWRQNTKSSIEAIKIQLSLKLET
APNISDVIQQLLPPKAPPPLRELIDQYDVRDDSDGSLEDDDHYATTETITIMPTESDFLMQVDGBKCCF
FKFSKKIQYNKVKQLWILRPVETTTTVQILRLIPLKMDGTRYTGIIRSLKLDNPGTGIWQSIDVKT
VLQNLWAAPAANLIGIEIKALDENGHDLATFPGPEDGLNPFLLEVKTDPKRSRDFGLDCEHSTESC
CRYPLTLVDEAFGDWIIAPKRYKANYCSGECIFVFLQKYPHTHLVHQANPRGSAGPCCTPTKMSPINMLY
FNGKEQIIYGKIPAMVVDCGCS

pHAT2-MSTNΔA43-SERp (cluster 3)

MSHSHHHSMDIEFVDEGSGPVDLNENSEQKENVEKENLYFQGSTWRQNTKSSIEAIKIQLSLKLET
APNISDVIQQLLPPKAPPPLRELIDQYDVRDDSDGSLEDDDHYATTETITIMPTESDFLMQVDGBKCCF
FKFSKKIQYNKVKQLWILRPVETTTTVQILRLIPLKMDGTRYTGIIRSLKLDNPGTGIWQSIDVKT
VLQNLWKQESNLGIEIKALDENGHDLATFPGPEDGLNPFLLEVKTDPKRSRDFGLDCEHSTESC
CRYPLTLVDEAFGDWIIAPKRYKANYCSGECIFVFLAAPYPHTHLVHQANPRGSAGPCCTPTKMSPINMLY
FNGKEQIIYGKIPAMVVDCGCS

pHAT2-MSTNΔA43-SERp (clusters 1+3)
Appendices

MBP-prodomain fusion sequences:

Vector derived sequences are underlined, MBP sequence in red and modified residues are shown in bold.

**pET28a-MBP-Pro (wild-type)**

MGHHHHHHHLVPRGS

**pET28a-MBP-Pro (R65A)**

MGHHHHHHHLVPRGS

**pET28a-MBP-Pro (R65C)**

MGHHHHHHHLVPRGS

**pET28a-MBP-Pro (A84G)**

MGHHHHHHHLVPRGS

183
The text on the page appears to be a scientific sequence or set of sequences, possibly related to protein engineering or bioinformatics. The sequences include amino acid strings typical of protein descriptions, with references to specific positions in the sequence (e.g., pET28a-MBP-Pro (Y111H), pET28a-MBP-Pro (H112R), pET28a-MBP-Pro (K153R), pET28a-MBP-Pro (W203A)). The sequences are presented in a tabular format with rows and columns, indicating multiple entries or variations of the same protein configuration.

The page number at the bottom of the page is 184, indicating this is part of a larger document or publication.
Cotton, T.R

Appendices

SDFLM&vdGKPKSSFKFFSSKIQYNKVKAQLWILRPVETPTTVQILRLIKPMKDGTGTRYGIRSLKLD
MNPMTGIAQSIDVKTVLQNWLQPFESNLGIEIKALDENGHDIAVTFPGPGEGLNFLEVKTDPK

pET28a-MBP-Pro (W203H)

MGHHHHGLVPRGSMKIEEGKLVIWINGDKGNYGAEGKKEFKDGTGKVTEHPDKEEKFQVAAATGDG
PDIIIFWADRFGQYAQSNLAEITPAAAFQDKLYPFWDVRYNGKLIAYPIAVEALSILDNLNNPK
TWEENPIKELAKGKSAKMKSALFNQLQPEYFTWPILAAADGGYAFKYAAAGYDKIDVGVNAGAKAGLTLFVLDL
IKNNHMADTDYSISIAEAAPNKGETAMINGPWAVSNIDSAVNYGVTLPFTKQGPSKFPVGVLSAGINAA
SPNKEIAGELENYLITDEGIAEAVNDKPLGAVALSYEMLKAPRLOAATMENAKQGEIMPNIQPMSAFW
YAVRTAVINAASGRQTVDAAIAAQTNAAAALFLVLFQGPNSNEQKENVEKGLSNASTWQRQNTKSSRI
EAIIQILSLHELITAPNISKDIVRQQLLPKAPRRELIDLQYDVQRDSDSGSLEDDDDYHATTETIITMPTE
SDFLM&vdGKPKSSFKFFSSKIQYNKVKAQLWILRPVETPTTVQILRLIKPMKDGTGTRYGIRSLKLD
MNPMTGIHQSIDVKTVLQNWLQPFESNLGIEIKALDENGHDIAVTFPGPGEGLNFLEVKTDPK

pET28a-MBP-Pro (W203F)

MGHHHHGLVPRGSMKIEEGKLVIWINGDKGNYGAEGKKEFKDGTGKVTEHPDKEEKFQVAAATGDG
PDIIIFWADRFGQYAQSNLAEITPAAAFQDKLYPFWDVRYNGKLIAYPIAVEALSILDNLNNPK
TWEENPIKELAKGKSAKMKSALFNQLQPEYFTWPILAAADGGYAFKYAAAGYDKIDVGVNAGAKAGLTLFVLDL
IKNNHMADTDYSISIAEAAPNKGETAMINGPWAVSNIDSAVNYGVTLPFTKQGPSKFPVGVLSAGINAA
SPNKEIAGELENYLITDEGIAEAVNDKPLGAVALSYEMLKAPRLOAATMENAKQGEIMPNIQPMSAFW
YAVRTAVINAASGRQTVDAAIAAQTNAAAALFLVLFQGPNSNEQKENVEKGLSNASTWQRQNTKSSRI
EAIIQILSLHELITAPNISKDIVRQQLLPKAPRRELIDLQYDVQRDSDSGSLEDDDDYHATTETIITMPTE
SDFLM&vdGKPKSSFKFFSSKIQYNKVKAQLWILRPVETPTTVQILRLIKPMKDGTGTRYGIRSLKLD
MNPMTGIHQSIDVKTVLQNWLQPFESNLGIEIKALDENGHDIAVTFPGPGEGLNFLEVKTDPK
9.2 Appendix 2: PCR oligonucleotide sequences

Cloning of full length pro-myostatin constructs

FWD (NcoI) 5’ TATATCCATGGGACCAGTAGATCTAAATGAAAAAT 3’
FWD (BamHI) 5’ TATATGGATCCGACCAGTAGATCTAAATGAGAAC 3’
REV (NotI) 5’ TATATAGGGCGGCTTATAGACCCACAGCGGCTACTAC 3’

Mutagenesis of furin cleavage site to HRV-3C site

FWD 5’ CTAGAAGTACTATTTCAAGGACCAGACTGTGATGAGCACTCAACA 3’
REV 5’ TGGTCCTTTGAAATATGACTTTCTAGTTTTGGTTGCTGTCTGTTACCTT 3’

N-terminal truncation at G37

FWD 5’ TATATCCATGGGCTGTGTAATGCATGTACTTGG 3’

N-terminal truncation at S50

FWD 5’ TATATCCATGGGCTCTTCAAGAATAGAAGCCATTAAG 3’

N-terminal truncation at N71

FWD 5’ TATATCCATGGGCAACATCAGCAAGATGTTATAAGA 3’

Introduction of N-terminal TEV cleavage site

FWD 5’ AACCTGTATTTTCAGGGCTCCACTTGGAGACAAAACACTAAA 3’
REV 5’ GCCCTGAAAATACAGGTTCTCTTTTTCCACATTTTCTTT 3’

Deletion of β4/5 loop

FWD 5’ ATTGGCGGCGGCATCCGATCTCTGAAAAGCT 3’
REV 5’ AATGCCGCCGCCGATGAGTCTCAGGATTTG 3’

Surface entropy mutations (K217A, Q218A, E220A)

FWD 5’ CAAAATTGGCTCGCGGCGCCTGCGTCCAACTTAGGC 3’
REV 5’ GCCTAAGTTGGACGCAGGCGCCGCGAGCCAATTTTG 3’

Surface entropy mutations (G319A, K320A)

FWD 5’ GAATTTGTATTTTTAGCGGCGTATCCTCATACTCAT 3’
REV 5’ ATGAGTATGAGGATACGCCGCTAAAAATACAAATTC 3’

Cloning mammalian constructs (pSF-CMV-FMDV-Rluc vector)

FWD (NcoI) 5’ ACGCCATGGATGCAAAAAACTGCAACTCTG 3’
REV (EcoRV) 5’ CGTATGATATCTCATGAGCACCCACAGCGG 3’

Cloning MBP-prodomain fusion for bacterial expression

FWD (BamHI) 5’ ACGGGATCCAATGAGAACAGTGAGCAAAAAGAAAATGTGG 3’
REV (EcoRI) 5’ TATGAATTCTCATTTTGGTGTGTCTGTTACCTTGAC 3’

Solubility improving prodomain cysteine substitutions

C39S+C42S FWD 5’ GGAAAAAGAGGGGCTGCTAATGCATCTACTTTGGAGACAAAAAC 3’
C39S+C42S REV 5’ GTGTTTTGTCTCCAAGTAGATGCATTAGACAGGCCCCTCTTTTCCC 3’
C137S+C138S FWD 5’ GGAATAAACAAATCTAGCTTTTAAATTTAGC 3’
C137S+C138S REV 5’ GCTAAATTTAAAGAGCTAGATTTTGGGTTTTCCC 3’

Prodomain polymorphism mutations
R65A FWD 5’ ATCCTCAGTAAACTTGCTCTGGAAACAGCTCC 3’
R65A REV 5’ AGGAGCTGTTTCCAGACAGAAGTTTACTGAGG 3’
R65C FWD 5’ ATCCTCAGTAAACTTTTGTCTTGAAACAGCTCC 3’
R65C REV 5’ AGGAGCTGTTTCCAGACAAAGTTTACTGAGG 3’
A84G FWD 5’ AACTTTTACCAAAGGTCTCTCCACTCCGG 3’
A84G REV 5’ TCCCCGGAGTGGAGACCTTTGGGTAAAAG 3’
Y111H FWD 5’ GAAGATGACGATCCACCGCTACCAACGG 3’
Y111H REV 5’ TCCGTTGTAGCGTGGTGATICGTCCATCTTCCC 3’
H112R FWD 5’ AGATGACGATTATGAGTACCTAAACGGAAAC 3’
H112R REV 5’ TGTTTCCGTTTGTAGGCTCGATAATCGTCCATCTTCCC 3’
K153R FWD 5’ AATAAAGTAGTAAGGGCCCAACTATGGATATAT 3’
K153R REV 5’ CCATAGTTGGGCCCTTACTACTTATTGTATTG 3’
W203A FWD 5’ AGGCACTGGTATTGCTCAGAGGATTGATG 3’
W203A REV 5’ ATCAATGCTCTGAGCAATCCAGTGGCC 3’
W203F FWD 5’ AGGCACTGGTATTTCGTAGCAGGATTGATG 3’
W203F REV 5’ ATCAATGCTCTGAGCAATCCAGTGGCC 3’
W203H FWD 5’ AGGCACTGGTATTCCACACCAGGATTGATG 3’
W203H REV 5’ ATCAATGCTCTGAGCAATCCAGTGGCC 3’
## 9.3 Appendix 3: Crystallographic data used for merging

Crystallographic data used for multi-crystal merging and phasing. All data were collected in a single session at Diamond Light Source, beamline I-03 at a wavelength of 0.9797 Å.

<table>
<thead>
<tr>
<th>Data processing*</th>
<th>Crystal 1</th>
<th>Crystal 2</th>
<th>Crystal 3</th>
<th>Crystal 4</th>
<th>Crystal 5</th>
<th>Crystal 6</th>
<th>Crystal 7</th>
<th>merged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spacegroup</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Unit cell (a, b, c [Å])</td>
<td>197.28 197.28 197.28</td>
<td>197.01 197.01 197.01</td>
<td>196.39 196.39 196.39</td>
<td>196.42 196.42 196.42</td>
<td>195.80 195.80 195.80</td>
<td>198.46 198.46 198.46</td>
<td>196.70 196.70 196.70</td>
<td>196.83 196.83 196.83</td>
</tr>
<tr>
<td>α,β,γ [°]</td>
<td>90.0 90.0 90.0</td>
<td>90.0 90.0 90.0</td>
<td>90.0 90.0 90.0</td>
<td>90.0 90.0 90.0</td>
<td>90.0 90.0 90.0</td>
<td>90.0 90.0 90.0</td>
<td>90.0 90.0 90.0</td>
<td>90.0 90.0 90.0</td>
</tr>
<tr>
<td>No of total/unique reflections</td>
<td>340350 / 9470</td>
<td>269893 / 8707</td>
<td>147990 / 8373</td>
<td>180483 / 8705</td>
<td>134953 / 8673</td>
<td>77422 / 9148</td>
<td>342136 / 9403</td>
<td>1517107 / 9470</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>35.9 (30.6)</td>
<td>31.0 (26.2)</td>
<td>17.8 (15.0)</td>
<td>20.7 (17.3)</td>
<td>15.6 (13.3)</td>
<td>8.5 (7.3)</td>
<td>36.4 (30.9)</td>
<td>160.2 (128.0)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.122 (1.694)</td>
<td>0.107 (2.933)</td>
<td>0.096 (2.022)</td>
<td>0.105 (1.840)</td>
<td>0.086 (1.576)</td>
<td>0.082 (1.002)</td>
<td>0.105 (1.913)</td>
<td>0.141 (2.690)</td>
</tr>
<tr>
<td>Rfactors</td>
<td>0.124 (1.722)</td>
<td>0.109 (2.990)</td>
<td>0.099 (2.099)</td>
<td>0.107 (1.895)</td>
<td>0.089 (1.639)</td>
<td>0.087 (1.079)</td>
<td>0.107 (2.229)</td>
<td>0.148 (2.732)</td>
</tr>
<tr>
<td>I/sigma</td>
<td>15.4 (1.8)</td>
<td>16.6 (1.3)</td>
<td>15.1 (1.3)</td>
<td>14.3 (1.5)</td>
<td>14.9 (1.4)</td>
<td>12.6 (1.4)</td>
<td>17.9 (1.6)</td>
<td>28.6 (3.0)</td>
</tr>
<tr>
<td>CC1/2</td>
<td>0.997 (0.865)</td>
<td>1.000 (0.702)</td>
<td>0.999 (0.604)</td>
<td>1.000 (0.701)</td>
<td>0.999 (0.722)</td>
<td>0.998 (0.977)</td>
<td>0.999 (0.706)</td>
<td>0.996 (0.923)</td>
</tr>
<tr>
<td>Completeness [%]</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
<td>99.9 (100.0)</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
</tr>
<tr>
<td>Anomalous completeness [%]</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
<td>99.8 (100.0)</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
<td>100.0 (100.0)</td>
</tr>
<tr>
<td>Anomalous multiplicity</td>
<td>18.8 (15.8)</td>
<td>16.2 (13.6)</td>
<td>9.3 (7.7)</td>
<td>10.9 (9.0)</td>
<td>8.2 (6.9)</td>
<td>4.4 (3.8)</td>
<td>19.1 (16.0)</td>
<td>84.0 (66.0)</td>
</tr>
</tbody>
</table>

(*) Values for high resolution shell are given in parenthesis for indication only; all integrated data were used during merging.

(** values in parenthesis are given for the low resolution shell)
## 9.4 Appendix 4: Pro-myostatin polymorphisms

Missense mutations in pro-myostatin from Ensembl database (Ensembl genome assembly GRCh38.p10, accessed on 05.06.2017). Scoring of possible effect on myostatin bioactivity is based on visual analysis of the mutated positions in the structure.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Chain A</th>
<th>Chain B</th>
<th>Domain/ feature</th>
<th>Comment</th>
<th>Scoring key</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2K</td>
<td>0</td>
<td>0</td>
<td>Signal peptide</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>V8I</td>
<td>0</td>
<td>0</td>
<td>Signal peptide</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Y11H</td>
<td>0</td>
<td>0</td>
<td>Signal peptide</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>M14T</td>
<td>0</td>
<td>0</td>
<td>Signal peptide</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>M14I</td>
<td>0</td>
<td>0</td>
<td>Signal peptide</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>I16T</td>
<td>0</td>
<td>0</td>
<td>Signal peptide</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>A18S</td>
<td>0</td>
<td>0</td>
<td>Signal peptide</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>G19I</td>
<td>0</td>
<td>0</td>
<td>Signal peptide</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>E25K</td>
<td>0</td>
<td>0</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>N26H</td>
<td>0</td>
<td>0</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>E28K</td>
<td>0</td>
<td>0</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Q29K</td>
<td>0</td>
<td>0</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>K30T</td>
<td>0</td>
<td>0</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>T43I</td>
<td>0</td>
<td>1</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>W44R</td>
<td>0</td>
<td>2</td>
<td>Forearm</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>R45K</td>
<td>0</td>
<td>1</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>K49E</td>
<td>1</td>
<td>2</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>R52K</td>
<td>3</td>
<td>3</td>
<td>Forearm</td>
<td>Removal of a hydrogen bond</td>
<td>0</td>
</tr>
<tr>
<td>I53K</td>
<td>1</td>
<td>1</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>E54K</td>
<td>1</td>
<td>2</td>
<td>Forearm</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>A55T</td>
<td>1</td>
<td>1</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Q29R</td>
<td>2</td>
<td>2</td>
<td>Forearm</td>
<td>Disruption of the fastener, introduction of a lone cysteine</td>
<td>2</td>
</tr>
<tr>
<td>R65C</td>
<td>3</td>
<td>3</td>
<td>Forearm</td>
<td>Disruption of the fastener</td>
<td>3</td>
</tr>
<tr>
<td>R65H</td>
<td>3</td>
<td>3</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>E67K</td>
<td>1</td>
<td>1</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>K74R</td>
<td>1</td>
<td>1</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>L80V</td>
<td>2</td>
<td>2</td>
<td>Forearm</td>
<td>Removal of an apolar interaction</td>
<td>2</td>
</tr>
<tr>
<td>A84G</td>
<td>3</td>
<td>3</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>R89Q</td>
<td>1</td>
<td>1</td>
<td>Forearm</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>I91T</td>
<td>3</td>
<td>3</td>
<td>Forearm</td>
<td>Polar substitution in an apolar site</td>
<td>3</td>
</tr>
<tr>
<td>D95H</td>
<td>0</td>
<td>1</td>
<td>Forearm</td>
<td>Close to Tolloid recognition site</td>
<td>0</td>
</tr>
<tr>
<td>K88S</td>
<td>0</td>
<td>0</td>
<td>Tolloid linker</td>
<td>Close to Tolloid recognition site</td>
<td>0</td>
</tr>
<tr>
<td>D99G</td>
<td>0</td>
<td>0</td>
<td>Tolloid linker</td>
<td>Mutation of Tolloid recognition site</td>
<td>0</td>
</tr>
<tr>
<td>D103N</td>
<td>0</td>
<td>0</td>
<td>Tolloid linker</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>L106F</td>
<td>1</td>
<td>0</td>
<td>Tolloid linker</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>D109E</td>
<td>1</td>
<td>1</td>
<td>Tolloid linker</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>D110N</td>
<td>1</td>
<td>2</td>
<td>Tolloid linker</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Y111H</td>
<td>3</td>
<td>3</td>
<td>Fastener</td>
<td>Loss of pi-cation interaction</td>
<td>0</td>
</tr>
<tr>
<td>H112R</td>
<td>3</td>
<td>3</td>
<td>Fastener</td>
<td>Loss of stacking in the fastener</td>
<td>0</td>
</tr>
<tr>
<td>T114A</td>
<td>1</td>
<td>1</td>
<td>Shoulder</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>T115M</td>
<td>2</td>
<td>2</td>
<td>Shoulder</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>M121T</td>
<td>3</td>
<td>3</td>
<td>Shoulder</td>
<td>Removal of hydrophobic residue</td>
<td>3</td>
</tr>
<tr>
<td>M129R</td>
<td>0</td>
<td>0</td>
<td>Shoulder</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>V131A</td>
<td>0</td>
<td>0</td>
<td>Shoulder</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Mutation</td>
<td>Frequency</td>
<td>Shoulder</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K136Q</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K136R</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C138F</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S144F</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y148C</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K153R</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A154T</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L156i</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R156i</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E164K</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P166S</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T167A</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q172R</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q172H</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I177L</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I177T</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I177M</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M180T</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M180V</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G183R</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T184I</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R185T</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T187N</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P188A</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R190Q</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R190L</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M196L</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M196V</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G201D</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S205I</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D207H</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T210I</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L212M</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N222S</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I225V</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I225T</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D236V</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A238T</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A238V</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T240N</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F241L</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P242S</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P244A</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L249P</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N250D</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P251L</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K256M</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V257L</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P261S</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R265K</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G269R</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G269D</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D273V</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E274K</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T277A</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R280Q</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R283H</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D289E</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A292G</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D296E</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K305E</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K305N</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N307S</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E312D</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F317L</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V327I</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H328N</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P332L</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P338R</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C340R</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M345L</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I348T</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N354H</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N354D</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q358K</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y361F</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G362E</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P365T</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P365L</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A366V</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R371G</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R371C</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R371L</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G373R</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Growth factor

Disruption of a disulfide bridge