4 Modelling dystroglycanopathy in zebrafish embryos by knockdown of *B3GALNT2* and *GMPPB*

4.1 Introduction

4.1.1 The phenotypic spectrum of dystroglycanopathy

Dystroglycanopathy is a subtype of congenital muscular dystrophy (CMD) that is characterised by hypoglycosylation of the α-dystroglycan (α-DG) protein. Dystroglycanopathy is currently clinically classified according to phenotypic severity. Walker-Warburg Syndrome (WWS) is the most severe type of dystroglycanopathy. Affected individuals have such severe muscular dystrophy that they have essentially no muscle tone from birth. There is usually profound intellectual disability (ID), structural brain abnormalities (such as cobblestone lissencephaly, hydrocephalus, and cerebellar malformations), and eye involvement (such as retinal malformations, micropthalmia, and blindness) (216). WWS is usually not compatible with life beyond one year of age.

The slightly less severe subtype is muscle-eye-brain disease (MEB). The phenotype is essentially similar, although patients might live for a few years, and develop some limited communication skills and motor control (217). Online Mendelian Inheritance in Man (OMIM) has recently collectively termed WWS and MEB as muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies) type A1 (MDDGA1; MIM 236670).

In the intermediate range of the spectrum of phenotypic severity, dystroglycanopathy patients have moderate to severe CMD manifesting in infancy or childhood. They may or may not have ID and central nervous system (CNS) involvement (218), and eye abnormalities are rare. This form is called muscular dystrophy-dystroglycanopathy (congenital with mental retardation), type B1 (MDDGB1; MIM 613155).

The mildest form of dystroglycanopathy is limb girdle muscular dystrophy (LGMD). LGMD is characterised by relatively mild muscular dystrophy with later onset. There

may or may not be ID or mild brain abnormalities (219). This form is known as muscular dystrophy-dystroglycanopathy (limb-girdle), type C1 (MDDGC1; MIM 609308) to distinguish it from LGMD patients who do not have hypoglycosylation of α-DG.

4.1.2 Dystroglycan structure, function and glycosylation

Both α-DG and β-dystroglycan (β-DG) are members of the dystrophin-glycoprotein complex (DGC), and they are encoded by a single gene, *DAG1* (MIM 128239). The dystroglycan polypeptide is post-translationally cleaved into the α and β subunits (220). β-DG spans the membrane of the sarcolemma. Intracellularly, it associates with dystrophin to transduce force from the myocyte protein machinery such as filamentous actin. Extracellularly, β-DG remains non-covalently bound to α-DG. α-DG in turn binds to extracellular matrix (ECM) proteins including laminin 2, agrin, perlecan, neurexin and pikachurin (216, 220-222).

α-DG is extensively glycosylated (Figure 4-1). There are five types of protein glycosylation: N-glycosylation, O-mannosylation, C-mannosylation, phospho-serine glycosylation and Glycosylphosphatidylinositol (GPI) anchor formation. α-DG undergoes O-mannosylation, and N-glycosylation, with O-mannosylation being most functionally important (223). First *O*-Mannose glycan residues are added to serine and threonine residues of the core α-DG protein, in the endoplasmic reticulum (ER) (224). The O-mannose next undergoes extension and branching to create complex glycan chains, the exact structures of most of which are not yet well characterised (223). α-DG migrates to the Golgi apparatus where N-acetylgalactosamine (GalNAc) initiated glycans are also added (225).

Binding of α-DG to its ligands is absolutely dependent upon appropriate glycosylation (226). Therefore, pathogenic variants in genes encoding enzymes involved in glycosylation can cause the hypoglycosylation of α-DG, which results in defective ligand binding, and the degeneration of muscle structure characteristic of dystroglycanopathy (227). The IIH6 antibody recognises O-mannosyl glycans, and can therefore be used experimentally to assess the level of glycosylation of α-DG in tissues.

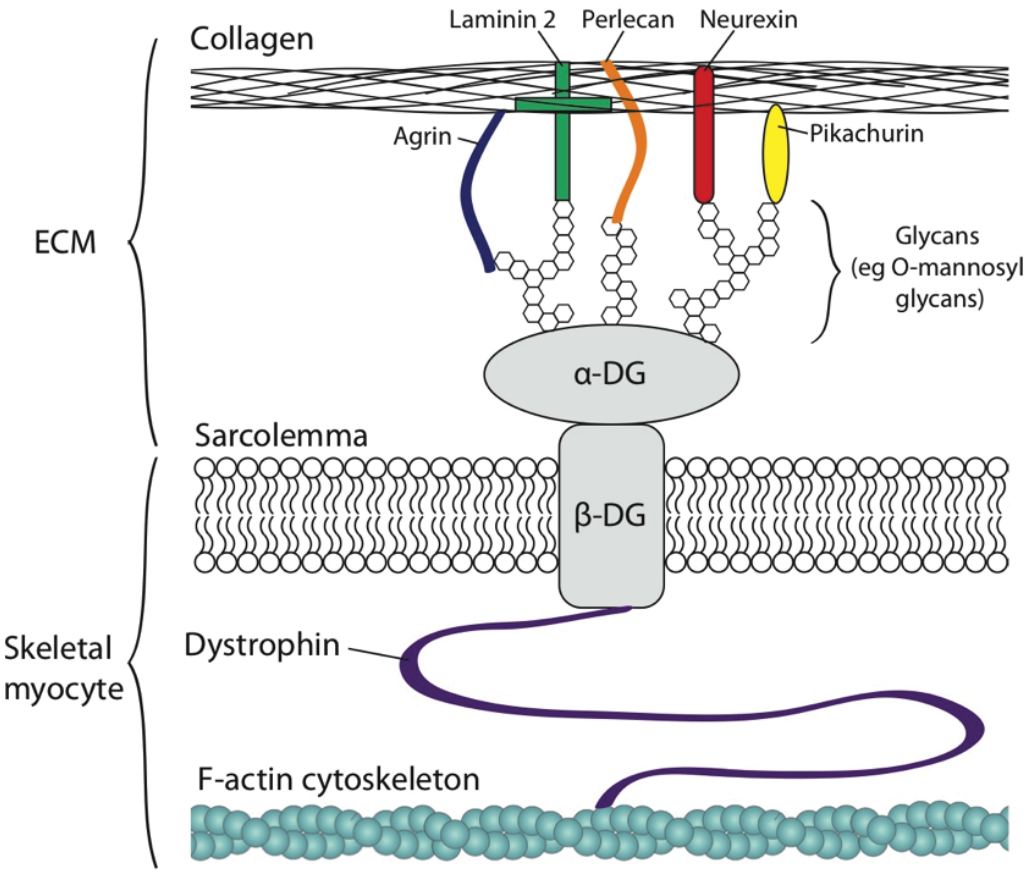

Figure 4 - 1: Model of α-DG interactions.
α-DG is non-covalently bound to the transmembrane protein β-DG, which connects to the sarcolemma cytoskeleton via dystrophin. α-DG binds to several components of the ECM via post-translationally added glycan chains, most importantly O-mannose glycans. Pathogenic variants in genes that encode proteins involved in this glycosylation cause impaired binding and loss of integrity of muscle tissue.

4.1.3 Known dystroglycanopathy-associated genes

Many proteins are involved in the glycosylation of dystroglycan, so it comes as no surprise that dystroglycanopathy is genetically as well as phenotypically heterogeneous. To date, not including the two genes described in this chapter, sixteen genes have been associated with dystroglycanopathy. First, there are rare reports of primary dystroglycanopathy, that is, dystroglycanopathy caused by pathogenic variants in the *DAG1* gene itself. The other fifteen genes are associated with secondary dystroglycanopathy. That is, pathogenic variants result in aberrant function of α-DG because of defects in the post-translational modifications of α-DG, not defects to the core structure of dystroglycan itself. These genes have autosomal recessive inheritance, and they can be classified into four further groups according to the function

of the protein they encode. These groups are: proteins involved in synthesis of dolichol phosphate mannose (Dol-P-Man), proteins involved in O-mannosylation, glycosyltransferases, and proteins with currently unknown function.

Primary dystroglycanopathy (DAG1)

The first report of a possibly pathogenic variant in *DAG1* itself was made in 2010 (228). The patient had learning difficulties, white matter abnormalities, elevated creatine kinase (CK), dyspraxia and facial hypotonia but otherwise no muscle dysfunction. The authors describe these symptoms as a subset of the classical dystroglycanopathy phenotype. The patient had a *de novo*, 2 Megabase (Mb) deletion that overlapped *DAG1*. However, the mutation was heterozygous, whereas variants that cause secondary dystroglycanopathy are recessive. While interesting, because of this, and the atypical spectrum of phenotypes, this case is not regarded as clearly being a case of primary dystroglycanopathy.

A year later, homozygous missense variants in *DAG1* were identified in a patient with LGMD, cognitive impairment, and hypoglycosylation of α-DG (229). This was a more typical dystroglycanopathy phenotype and genotype. A mouse model with the patient's variant had similar abnormalities. Despite this, some scientists in the dystroglycanopathy research community were skeptical that this variant was the sole cause of the patient's phenotype, because it was a single patient and the variants were missense. This skepticism was somewhat allayed by two subsequent studies. In one, simulations suggest that docking between α-DG and its ligands was weakened by the variant in that patient (15). In the other, a homozygous missense variant in *DAG1* was found in two siblings who had severe MEB with macrocephaly and white matter disease (230). The variant, which is in a conserved section of the part of *DAG1* that encodes β-DG, is thought to disrupt the structure of the protein.

Genes encoding proteins involved in synthesis of Dol-P-Man (DPM1, DPM2, DPM3, and DOLK)

DPM synthase is an enzyme which catalyses the formation of Dol-P-Man, a mannosyl donor. In mammals, DPM synthase is a complex consisting of three subunits: DPM1, DPM2 and DPM3 (231). DPM1 is the primary catalytic subunit. While DPM2 may have

some enzymatic activity, its primary function is to stabilise DPM3, which in turn allows DPM1 to be stably expressed at the ER membrane (232).

Defects in the genes encoding the components of DPM synthase can cause wide ranging phenotypic effects. For example, pathogenic variants in *DPM1* can cause congenital disorder of glycosylation type Ie (CDG-Ie). These patients have intellectual and motor disability caused by defects in N-linked glycosylation of proteins (233). In 2009, pathogenic variants in *DPM3* that cause dystroglycanopathy were found (234). These patients had reduced O-mannosylation of α-DG. Pathogenic variants in *DPM2* and *DPM1* have since also been found in dystroglycanopathy patients (235, 236). This demonstrates that the two phenotypically distinct conditions congenital disorder of glycosylation (CDG) and dystroglycanopathy can have the same molecular cause. This is because Dol-P-Man is required for both N-glycosylation and O-mannosylation. It is unclear why different variants in this complex cause the two different phenotypes.

Similarly, DOLK catalyses the formation of dolichol monophosphate, which is a precursor of Dol-P-Man. Variants in *DOLK* have been reported in patients with a phenotype that overlaps dystroglycanopathy and CDG, with defective N-glycosylation and reduced O-mannosylation (237).

Genes encoding proteins involved in O-mannosylation (POMGnT1, POMT1, and POMT2)

POMGnT1 catalyses the transfer of N-acetylglucosamine (GlcNAc) to O-mannose, extending *O*-mannosyl glycan chains (217). POMT1 and POMT2 each have O-mannosyltransferase activity, but this depends on them interacting physically and functionally (238, 239). However, the functions of POMT1 and POMT2 are not interchangeable (240).

In 2001, pathogenic variants in *POMGnT1* were first found to cause MEB in six dystroglycanopathy patients (217). Since then, the phenotypic spectrum of patients with *POMGnT1* variants has been expanded, as some very mildly affected patients have been identified, as well as severely affected patients (241). Different classes of pathogenic *POMGnT1* variants have been implicated, including a duplication in the promoter and an intragenic deletion (242, 243). Patients with pathogenic *POMGnT1* variants have hypoglycosylated α-DG that has reduced ability to bind to its ligands (226).

In 2002, pathogenic *POMT1* variants were found to cause WWS with hypoglycosylation of α-DG in five consanguineous families (216). They can also cause a milder LGMD phenotype, and in some cases cardiac defects (244, 245). In 2005, *POMT2* variants were also implicated in dystroglycanopathy (221). A variety of pathogenic *POMT2* variants can cause disease including intronic deletions and substitutions (246). Pathogenic variants in *POMT1* or *POMT2* that are found in dystroglycanopathy patients can ameliorate catalytic activity (247).

Genes encoding glycosyltransferases (LARGE, B3GNT1, GTDC2, and POMK)

LARGE is a ubiquitously expressed glycosyltransferase that interacts in the Golgi apparatus with α-DG domains, including the N-terminal domain and the mucin-like domain. LARGE is required for glycosylation of these domains (248). Specifically, LARGE catalyses the addition of repeating units of xylose and glucaronic acid to α-DG, which is required for α-DG to bind to its ligands including laminin (249). Pathogenic variants in *LARGE* can cause dystroglycanopathy (250). Many different *LARGE* variants have been implicated, including intragenic rearrangements (243). Some patients' variants perturb the interaction of LARGE with α-DG (251).

B3GNT1 is an enzyme that interacts with LARGE, and catalyses polymerization of GlcNAc residues (252, 253). B3GNT1 is necessary for α-DG glycosylation (222). In 2013, pathogenic variants in *B3GNT1* were found to cause dystroglycanopathy (16). This study also showed that in human cells, wildtype B3GNT1 increases α-DG glycosylation, but the variant form of B3GNT1 does not.

In 2012, a combination of homozygosity analysis and WES in consanguineous families with WWS was used to find pathogenic variants in *GTDC2* (254). GTDC2 catalyses the addition of GlcNAc epitopes to O-mannosylated α-DG in the ER (255). Pathogenic variants reduce the catalytic activity of GTDC2. Pathogenic variants in *POMK*, a glycosyltransferase also known as *SGK196*, can also cause dystroglycanopathy (256, 257). POMK, along with GTDC2, is involved in synthesising an O-mannosyl trisaccharide structure on α-DG, without which α-DG cannot bind to laminin (258).

Genes encoding proteins with unknown function (FKTN, FKRP, ISPD, and TMEM5)

FKTN was the first identified dystroglycanopathy-associated gene. There is a particular subtype of dystroglycanopathy known as Fukuyama-type CMD (FCMD) which is

relatively common (~1/10,000 births) in Japan. The phenotype consists of hypotonia and muscle weakness first manifesting in infancy, motor delay, ID, seizures in around half of patients, and malformations of the eye and brain (259). In 1998, a 3 kilobase (kb) retrotransposal insertion of a tandem repeat in the 3' untranslated region (UTR) of a novel gene they called *fukutin* or *FKTN*, was identified as the cause of disease in nearly 90% of FCMD cases (260). It is now thought to be a founder mutation that arose around 2000 years ago (261). The mutation results in inappropriate splicing of mRNA (262). Since this discovery, many additional *FKTN* variants that can cause dystroglycanopathy have been identified, both inside and outside the Japanese population (263, 264). Pathogenic *FKTN* variants have also been shown to cause a wider range of severity of phenotypes than just classical FCMD, from LGMD to WWS (265-267). α-DG is hypoglycosylated in the muscle of FCMD patients, and therefore less able to bind to its ligands including laminin, neurexin or agrin (226, 268).

In 2001, pathogenic variants in the *FKTN* related gene *FKRP* were found to cause severe CMD known then as MDC1C (269). This was closely followed by the finding that *FKRP* is also associated with a milder LGMD phenotype with reduced glycosylation of α-DG (270, 271). Pathogenic variants in *FKRP* can also cause a more severe phenotype with cardiac and CNS involvement (272). Both FKTN and FKRP are required for α-DG glycosylation, and they are predicted to be glycosyltransferases. However, despite years of research, their precise functions remain elusive.

Nevertheless, insights can be gained by examining tissue expression patterns and cellular localisation. *FKTN* is expressed prenatally in the brain, particularly in glial cells and astrocytes. This expression is reduced in FCMD cases (273, 274). The subcellular localisation of wildtype FKTN is the Golgi apparatus (275). Some CMD patients' pathogenic variants cause FKTN to be retained in the ER, probably because it is misfolded (276). Interestingly, FKTN interacts with POMGnT1, possibly modulating its activity (277).

In contrast, the subcellular localisation of FKRP has proved controversial. One study found that wildtype FKRP localises to the rough ER, and is therefore likely to be involved in an earlier stage of α-DG glycosylation pathway than FKTN (275). Two other studies contradicted this however. One found FKRP colocalised with the DGC in the sarcolemma (278), whereas the other concluded that wildtype FKRP is located in the Golgi apparatus, and that some pathogenic variants cause FKRP to be retained and

degraded in the ER (279). A further study could not replicate this finding that wildtype and variant FKRP are differently localised (280).

Pathogenic variants in the uncharacterised gene *ISPD* can disrupt O-mannosylation of α-DG, causing dystroglycanopathy (281, 282). As for many dystroglycanopathy-associated genes, the phenotypic spectrum is variable (283). In cultured patients' fibroblasts, wildtype ISPD can restore glycosylation of α-DG (281). Pathogenic variants in *ISPD* and another uncharacterised gene *TMEM5* have been found in fetuses with cobblestone lissencephaly (284). This brain malformation often occurs in patients with dystroglycanopathy. *TMEM5*, which has a predicted glycosyltransferase domain, was confirmed as a dystroglycanopathy-associated gene in 2013 (256).

4.1.4 Frequency of variants, and genotype-phenotype correlations

Several groups have studied the frequency of pathogenic variants in the different dystroglycanopathy-associated genes. Pathogenic variants in *POMT1*, *POMT2*, *POMGnT1*, *FKRP*, *FKTN* and *LARGE* collectively account for up to 50% of cases of dystroglycanopathy, depending on ethnicity, and how the cohort has been screened (219, 285-287). These studies largely agree that *POMT1* and *POMT2* are the most frequently implicated genes, followed by *POMGnT1* and *FKRP*. The most recent of these studies was published in 2009. It is likely that pathogenic variants in all of the currently known dystroglycanopathy-associated genes, including those identified since 2009, will account for more than 50% of cases.

One might assume that pathogenic variants in different dystroglycanopathy-associated genes might cause slightly different phenotypes. While it often initially appears that pathogenic variants in one gene cause a particular subtype of dystroglycanopathy (e.g. *FKTN* variants cause FCMD, variants in *POMGnT1* cause MEB, and variants in *POMT2* cause WWS (217, 221, 260)) further study of each of these cases invariably reveals that the phenotypic spectrum is in fact broad and overlapping (267, 288, 289). In general, there appears to be remarkably little correlation between genotype and phenotype, although some groups have noted that pathogenic *POMT1* and *POMT2* variants tend to cause CNS involvement more frequently than variants in other genes, pathogenic *POMGnT1* variants often cause cerebellar cysts, and loss of function variants may cause more severe phenotypes than missense variants (219, 287, 290,

291). The general lack or correlation between genotype and phenotype make it virtually impossible to predict the causative gene from the phenotype alone.

Even more surprisingly, there appears to be little correlation between the severity of phenotype and the extent of hypoglycosylation of α-DG. For example, one study found that while there was some evidence of such a correlation for a few genes, in other cases, patients could have very mild phenotypes despite a profound hypoglycosylation of α-DG (292). One hypothesis that might explain this observation is that some dystroglycanopathy-associated genes may be involved in the glycosylation of other target proteins in addition to α-DG, and this could contribute to the severity of the phenotype. In another study, the severity of the phenotype in patients with pathogenic *POMGnT1* variants does not correlate with the levels of POMGnT1 protein (293).

4.1.5 Zebrafish models of genetic disease

In vivo models have for decades been vital tools in the study of human diseases. Mice represent the classic laboratory model for human diseases. However other organisms, including zebrafish, are increasing in popularity as disease models for many reasons. Zebrafish are vertebrates with a high homology to humans; 71.4% of human protein-coding genes have at least one direct orthologue in the zebrafish (294). This figure increases to 82% when only looking at genes associated with human disease. An important advantages of zebrafish over mice is that because the embryos develop *ex utero*, it is possible in some cases to use zebrafish to study the effect of knocking out a gene that is embryonic lethal in a placental mammal such as a mouse (one example of this is *dag1*) (295, 296). From a practical point of view, zebrafish are relatively cheap and easy to maintain, they have high fecundity, the embryos develop quickly, and they are amenable to various forms of genetic manipulation.

Zebrafish embryos are particularly good models with which to study muscle diseases, including dystroglycanopathy. In part, this is because a high proportion of each zebrafish embryo is muscle tissue, and the embryos are optically transparent, allowing easy visualization of developing muscle tissue using simple microscopy techniques. Conveniently, the muscle tissue develops quickly; the embryo can swim by 24 hours post fertilisation (hpf), and the muscle is fully differentiated by 48 hpf (295). Furthermore, despite some differences between human and zebrafish muscle tissue at the structural level, they are highly orthologous at the molecular level.

Disadvantages of zebrafish compared to mice include the fact that obviously they are more evolutionarily diverged from humans, and there are some structures, such as the lungs, hair, and teeth, which they do not have, while some organs are fundamentally different, such as their hearts which only have two chambers. However, it is possible to recapitulate genetic cardiac abnormalities in zebrafish hearts despite this profound difference between species (297). Perhaps surprisingly, there are even some early hints that complicated social behaviours in humans have some parallels in zebrafish (298).

There are three main methods of making a zebrafish model of disease: random mutagenesis, site-specific nucleases, and transient inhibition of gene expression. Random mutagenesis involves exposing zebrafish to a mutagen such as *N*-ethyl-*N*-nitrosourea (ENU), which generates random mutations in the genome that are propagated to the next generation. The aim of the zebrafish mutation project is to use this method to generate a zebrafish knockout for every protein-coding gene in the zebrafish genome (299). Random mutagenesis is an efficient and high-throughput method of generating mutants, but identifying all the mutations, and linking them to phenotypes, is challenging. The random nature of the mutations generated also limits the utility of this approach.

Site-specific nucleases are enzymes that consist of two parts: a DNA recognition motif to target the enzyme to a particular genomic locus, and a nuclease to induce double-strand breaks at that locus. Examples include zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPRs) (300). These enable very specific mutations to be introduced into the genome. ZFNs were the first to be developed, but they were not widely used because they were difficult, expensive and time-consuming to make. TALENs and CRISPRs are more promising tools, which can be made in individual laboratories relatively cheaply and easily, and in a high-throughout manner. Problems with site-specific nucleases include possible off-target effects, and the fact that the time from designing the system to having a mutant fish ready to phenotype takes at least several months (301).

The most commonly used tool with which to transiently inhibit expression of a zebrafish gene is morpholino oligonucleotides (MOs). These short, synthetic oligonucleotides are designed to specifically target the mRNA of a gene of interest, and prevent it from being properly spliced or translated. The main advantage of MOs over random

mutagenesis or site-specific nucleases is that they are very quick; one can begin to examine the phenotype of a knockdown organism generated using a MO (known as a morphant) within a few days of identifying a gene of interest. However, there are several disadvantages. MOs only transiently inhibit expression of a target gene. Also, they can be toxic, and cause non-specific phenotypes (302, 303). It can be challenging to confirm that expression of the target gene has been inhibited, and it can be even more challenging to confirm that the expression of no other off-target genes have been inhibited. There are methods to minimise the impact of these problems, so it is important to take them into account when designing experiments and interpreting results.

4.1.6 Animal models of dystroglycanopathy

The zebrafish orthologues of *LARGE*, *POMT1*, *POMT2*, *POMGnT1*, *FKTN*, and *FKRP* were identified in 2008 (304). All of these genes were expressed throughout early development, including in tissues relevant to dystroglycanopathy including the CNS, eye and muscle. Importantly, the authors also found that the antibody IIH6, which recognises glycosylated α-DG in humans, also works in zebrafish. Since then, in zebrafish embryos, dystroglycanopathy-associated genes have often been studied using MOs to knock down a candidate gene. For example, knockdown of *fkrp* results in changes to somites, muscle fibres, neuronal structures and eye shape, hypoglycosylation of α-DG and reduced laminin binding, all of which recapitulates the patients' phenotypes (17, 305). Similar results have been found for orthologues of many other dystroglycanopathy-associated genes including *POMT1*, *POMT2*, *B3GNT1*, *GTDC2*, *ISPD*, and *DAG1* (16, 254, 282, 306, 307).

As well as providing supportive evidence of causality, modelling a candidate dystroglycanopathy-associated gene in a model organism often leads to insights into the molecular pathology of disease that would not be possible on human patients. For example, examination of various tissues of the *Fktn* knockout mouse revealed that some of the structural tissue defects might be cause by disruption of the basal lamina, caused in turn by disruption of α-DG-ligand binding (308). In another important example, mass spectrometry of O-linked glycans in brain samples from mouse knockouts for *Pomgnt1*, *Large*, and *Dag1* yielded evidence that *Pomgnt1* may glycosylate other brain proteins in addition to α-DG for glycosylation. This finding could

have important implications for understanding of pathology in patients (309). Experiments on zebrafish models of dystroglycanopathy have similarly revealed many insights into the disease pathology, such as FKTN and FKRP may have role in protein secretion (310).

Animal models can also lead to important developments in therapeutic strategies. For example, several groups have shown that overexpression of *Large* in mouse models can improve a dystroglycanopathy phenotype, even when the mouse has a pathogenic variant in another gene such as *Pomgnt1*, suggesting that overexpression of *LARGE* could be a viable therapeutic strategy for human dystroglycanopathy patients, whether or not their pathogenic variants are in *LARGE* (311, 312).

4.1.7 Aims, context, and colleagues

Some parts of this chapter have been published (313, 314). The parts of these two publications that I have reproduced in this chapter were all my work originally. This section briefly summarises the aspects of these two publications with which I was not directly involved, in order to put my own results into context.

The exomes of five patients with dystroglycanopathy were sequenced as part of the UK10K project, under the rare disease consortium, with the aim of identifying novel causative genes. Dr Sebahattin Cirak identified likely candidate genes for two of the patients. The genes were β-1,3-N-acetylgalactosaminyltransferase 2 (*B3GALNT2* [MIM 610194]) and guanosine diphosphate (GDP) mannose pyrophosphorylase B (*GMPPB* [MIM 615320]).

B3GALNT2 encodes a glycosyltransferase that is involved in the synthesis of GlcNac-β1,3GalNac, which is one of the glycans on α-DG, and is required for laminin binding (227, 315, 316). GMPPB catalyses the conversion of mannose-1-phosphate and GTP into GDP-mannose (317). GDP-mannose is required in four glycosylation pathways (Figure 4-2), including O-mannosylation of membrane and secretory glycoproteins, such as α-DG. Pathogenic variants in other members of this pathway can cause dystroglycanopathy (234, 235).

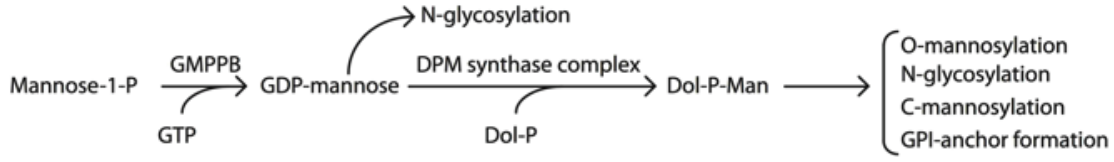

Figure 4 - 2: The function of GMPPB in glycosylation pathways.
This figure has been published in a modified form (313).

Through an international collaboration lead by Professor Francesco Muntoni and Dr A. Reghan Foley, six further individuals (two of whom were siblings) with dystroglycanopathy and *B3GALNT2* variants were identified, and seven further individuals with dystroglycanopathy and *GMPPB* variants were identified. All of the 15 patients had homozygous or compound heterozygous variants. The variants included missense, nonsense and frameshift changes.

All seven patients with *B3GALNT2* variants were severely affected. They had muscle dysfunction, cognitive impairment, and gross abnormalities of the brain with early presentation. Three of the patients also had epilepsy, and five had ophthalmologic abnormalities. Most reached no major motor milestones. In contrast, the phenotype of the eight patients with *GMPPB* variants was milder, and the presentation later. Most patients had hypotonia and poor muscle control, but they could walk. Most had intellectual delay, but mild, and structural brain abnormalities were only detected in three patients.

Dr Elizabeth Stevens and Dr Silvia Torelli confirmed hypoglycosylation of α-DG in most of the 15 patients using multiple methods, including immunoblot of both muscle protein lysate, and flow cytometry of fibroblasts, using the IIH6 antibody. They also transfected myoblasts with recombinant wildtype B3GALNT2, and showed that it localises to the ER, and that some of the variants identified in patients cause B3GALNT2 to mislocalise to the cytoplasm. In contrast, recombinant wildtype GMPPB localises to the cytoplasm, and some of the variants identified in patients result in formation of protein aggregates.

I carried out the work described in the rest of this chapter, under the overall supervision of Dr Derek Stemple and Dr Matthew Hurles, with direct, day-to-day supervision and assistance from Dr Yung-Yao Lin. Our aim was to knockdown orthologues of *B3GALNT2* and *GMPPB* in zebrafish embryos in order to recapitulate the patients' phenotypes to provide further evidence of pathogenicity, and to gain further insight into the pathology of the disease.

4.2 Materials and methods

4.2.1 Sequencing of clones

I obtained clones of *b3galnt2* (BC095777) and *gmppb* (BC078357.1) from Source BioScience (Nottingham, UK). I Sanger sequenced each insert using standard protocols, and confirmed that they did not contain any variants different from the human reference sequence. These sequences were used for all alignments and reagent design.

4.2.2 Reverse transcription polymerase chain reaction

To estimate the expression levels of zebrafish genes, I extracted RNA from 20–30 embryos with an RNeasy kit (QIAGEN, Crawley, UK), followed by reverse transcription with SuperScript III (Life Technologies). Polymerase chain reaction (PCR) was done with RedTaq DNA polymerase kit (Sigma-Aldrich, Dorset, UK). I amplified fragments of *b3galnt2* and *gmppb*, using *actb1* as a positive control. I visualised results using standard agarose gel electrophoresis. Primer sequences are listed in Table 4-1.

Gene	Primer function	Direction	Primer sequence (5'-3')
b3galnt2	Expression analysis	Forward	actcagagctccgcgatg
b3galnt2	Expression analysis	Reverse	cagagcagagatccctcaaa
gmppb	Expression analysis & MO-flanking	Forward	tacagcagcaggtcaatcgt
gmppb	Expression analysis	Reverse	acaacaatggtgccctctct
gmppb	MO-flanking	Reverse	gttctgcccaatcactgctg

Table 4 - 1: Primers used to analyse *b3galnt2* and *gmppb* expression in early zebrafish development, and splicing disruption in *gmppb* morphants.
The sequences of the *actb1* primers have been previously described (282). This table and legend have been published in a modified form (313).

4.2.3 Design and injection of morpholino oligonucleotides

For *b3galnt2*, I obtained one translation blocking (TB) MO and one splice blocking (SB) MO from Gene Tools, LLC (Philomath, OR, USA). For *gmppb*, I obtained one TB MO and three SB MOs. For MO sequences and the predicted effect of each SB MO, please

see Table 4-2. All SB MOs were predicted to result in a frameshift. I also confirmed that there were no known variants in the MO binding site, and that the MO binding site was predicted to be specific.

Gene	MO type	Sequence (5'-3')	Predicted Effect
b3galnt2	TB	CGCCGCCGCTGCACTTCT**CAT**GGAC	NA
b3galnt2	SB	GGTCTGTctgtcaaggagaaataaa	Skipping of exon 2
gmppb	TB	CACCGACAAGAATCAGAGCTTT**CAT**	NA
gmppb	SB	gaaagactgccgtcagttacCTTGA	Retention of intron 2
gmppb	SB	GGACCAGctgaaaacagaaacagat	Skipping of exon 5
gmppb	SB	acagtgttcaaatcctttacCTTGC	Retention of intron 7

Table 4 - 2: Morpholino oligonucleotide sequences and predicted effects.
MO = morpholino oligonucleotide; TB = translation blocking; SB = splice blocking. In the sequences of TB MOs, the letters in bold indicate the start codon. In the sequences of SB MOs, the letters in lower case indicate the portion of the MO predicted to bind in the intron, and the letters in upper case indicate the portion of the MO predicted to bind in the exon. Sequences of *p53* and *dag1* MOs have been described (296, 302).

I injected MOs into 1- to 4- cell-stage Tuebingen Long Fin zebrafish embryos, which I reared as previously described (318). Unless otherwise stated, for *b3galnt2* I injected the TB MO at 4 ng along with a *p53* TB MO at 2 ng, and for *gmppb* I used the second SB MO, which is predicted to result in skipping of the fifth coding exon, at 3 ng dose, coinjected with *p53* MO 6 ng. To compare the eye diameter of different groups of embryos, I used unpaired two-tailed t tests.

4.2.4 Generation of green fluorescent protein-tagged RNA

I generated zebrafish expression plasmids (pCS2fl) containing cDNA of the gene of interest and a C-terminal green fluorescent protein (GFP) tag by Gateway Cloning Technology (Life Technologies, Paisley, UK) according to the manufacturer's instructions. I made mRNA for injection with mMessage mMachine SP6 kit (Ambion, for Life Technologies).

4.2.5 Immunofluorescence staining

I performed immunofluorescence staining on 48 hpf whole-mount embryos as previously described (296). For B3galnt2, I used primary antibodies against laminin (L-9393, Sigma-Aldrich) and β-DG (monoclonal, NCL-b-DG, Leica Microsystems, Milton Keynes, UK). For Gmppb I used primary antibodies against filamentous actin with the use of Alexa-Fluor-594-conjugated phalloidin (Life Technologies), β-DG, laminin, and IIH6.

4.2.6 Evans blue dye assay

For the Evans blue dye (EBD) assay, I immobilised 48 hpf live embryos in 1% low-melting-point agarose (Sigma-Aldrich) containing 0.016% tricaine (Sigma-Aldrich), and I injected a solution of 0.1% EBD (Sigma-Aldrich) into the pericardium. Two hours later, I examined them by confocal microscopy. To assess the significance of the results of the EBD assay, I used unpaired two-tailed t tests.

4.2.7 Immunoblotting

I performed microsome preparation and immunoblot analysis of zebrafish proteins as previously described (282), and quantified results using ImageJ software (319).

4.3 Results

4.3.1 *B3GALNT2* and *GMPPB* are conserved with their zebrafish orthologues

B3GALNT2 and *GMPPB* each have a single orthologue in the zebrafish. Zebrafish B3galnt2 (ENSDARP00000067823) is 53% identical to human B3GALNT2. Interestingly, the amino acid sequence of the galactosyltransferase domain alone is 68.5% identical between the two species (Table 4-3 and Figure 4-3). Zebrafish Gmppb (ENSDARP00000022618) is 81.4% identical to human GMPPB (Table 4-4 and Figure 4-4).

Species	Protein name	Ensembl identifier	% identity with human B3GALNT2 (whole protein)	% identity with human B3GALNT2 (galactosyl-transferase domain)
Pan troglodytes	B3GALNT2	ENSPTRT00000003896	99.6	100
Mus musculus	B3GALNT2	ENSMUST00000099747	88.4	93.3
Danio rerio	B3galnt2	Clone BC095777	54.6	68.5
Drosophila melanogaster	Beta-1,3-galactosyltransferase II	FBtr0088728	21.2	32.2
Caenorhabditis elegans	Beta-1,3-galactosyltransferase sqv-2	Y110A2AL_14_2	25.1	28.8

Table 4 - 3: Percentage identity of B3GALNT2 orthologues of five diverse eukaryotic species with human B3GALNT2.
Orthologues identified by BLAST alignment of the *Homo sapiens* B3GALNT2 sequence (ENSP00000355559) against the genomes of the species shown.

Species	Protein name	Ensembl identifier	% identity with human GMPPB
Pan troglodytes	GMPPB	ENSPTRP00000025773	99.7
Mus musculus	GMPPB	ENSMUSP00000107914	98.1
Danio rerio	Gmppb	ENSDARP00000022618	81.4
Drosophila melanogaster	CG1129	FBpp0078511	70.2
Caenorhabditis elegans	TAG-335	C42C1.5	63.8

Table 4 - 4: Percentage identity of GMPPB orthologues of five diverse eukaryotic species with human GMPPB.
Orthologues identified by BLAST alignment of the *Homo sapiens* GMPPB sequence (ENSP00000418565) against the genomes of the species shown. This table and legend have been published (313).

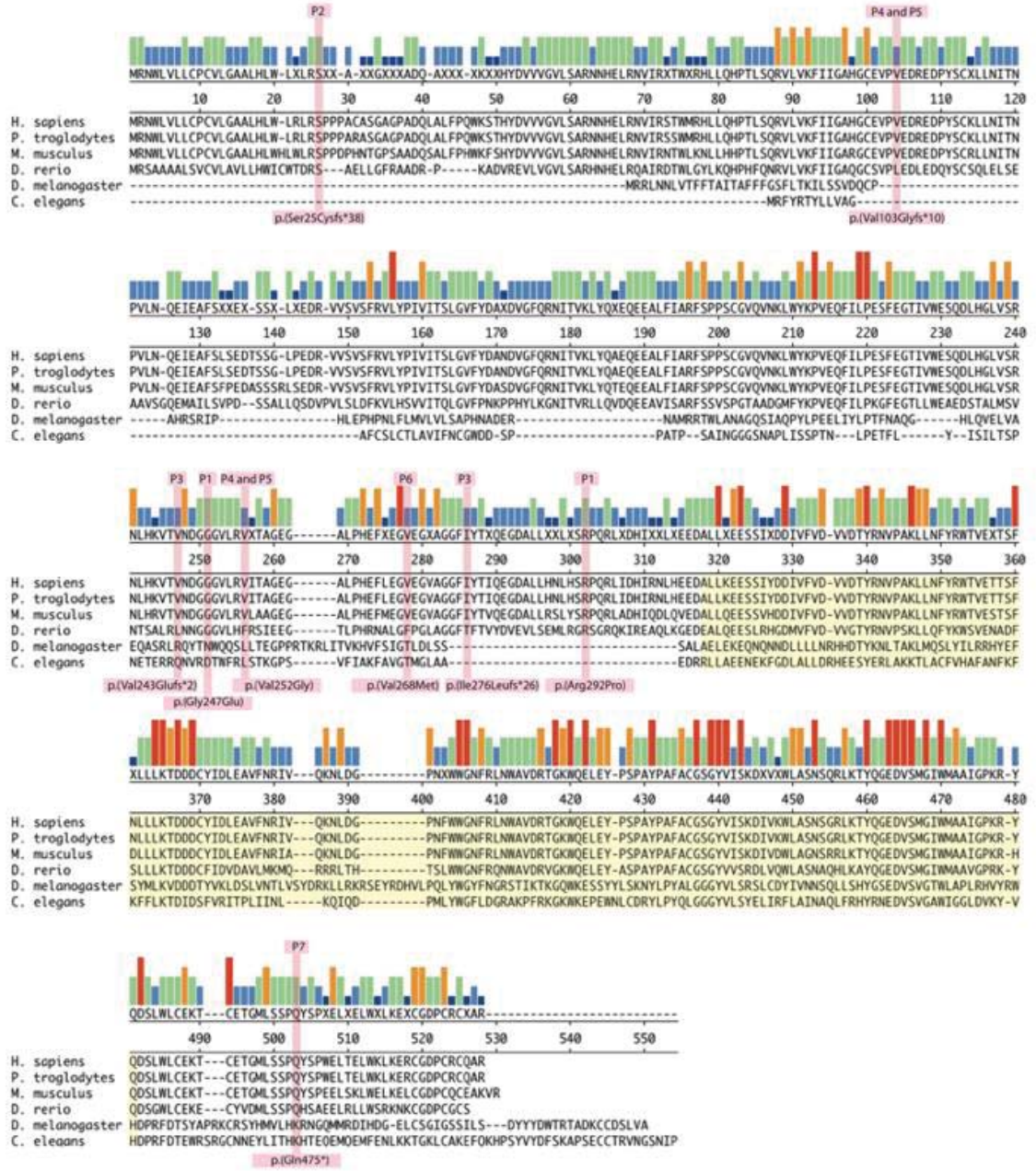

Figure 4 - 3: Protein alignment showing conservation of B3GALNT2.
Sequences are *Homo sapiens* (ENSP00000355559), *Pan troglodytes* (ENSPTRT00000003896), *Mus musculus* (ENSMUST00000099747), *Danio rerio* (clone BC095777), *Drosophila melanogaster* (FBtr0088728), and *Caenorhabditis elegans* (Y110A2AL_14_2). The height and colour of the bars indicates the degree of conservation of each amino acid residue, for example a red bar shows that a residue is conserved across all six species. The residues altered in the muscular dystrophy cases are highlighted in pink. The galactosyltransferase domain is highlighted in yellow.

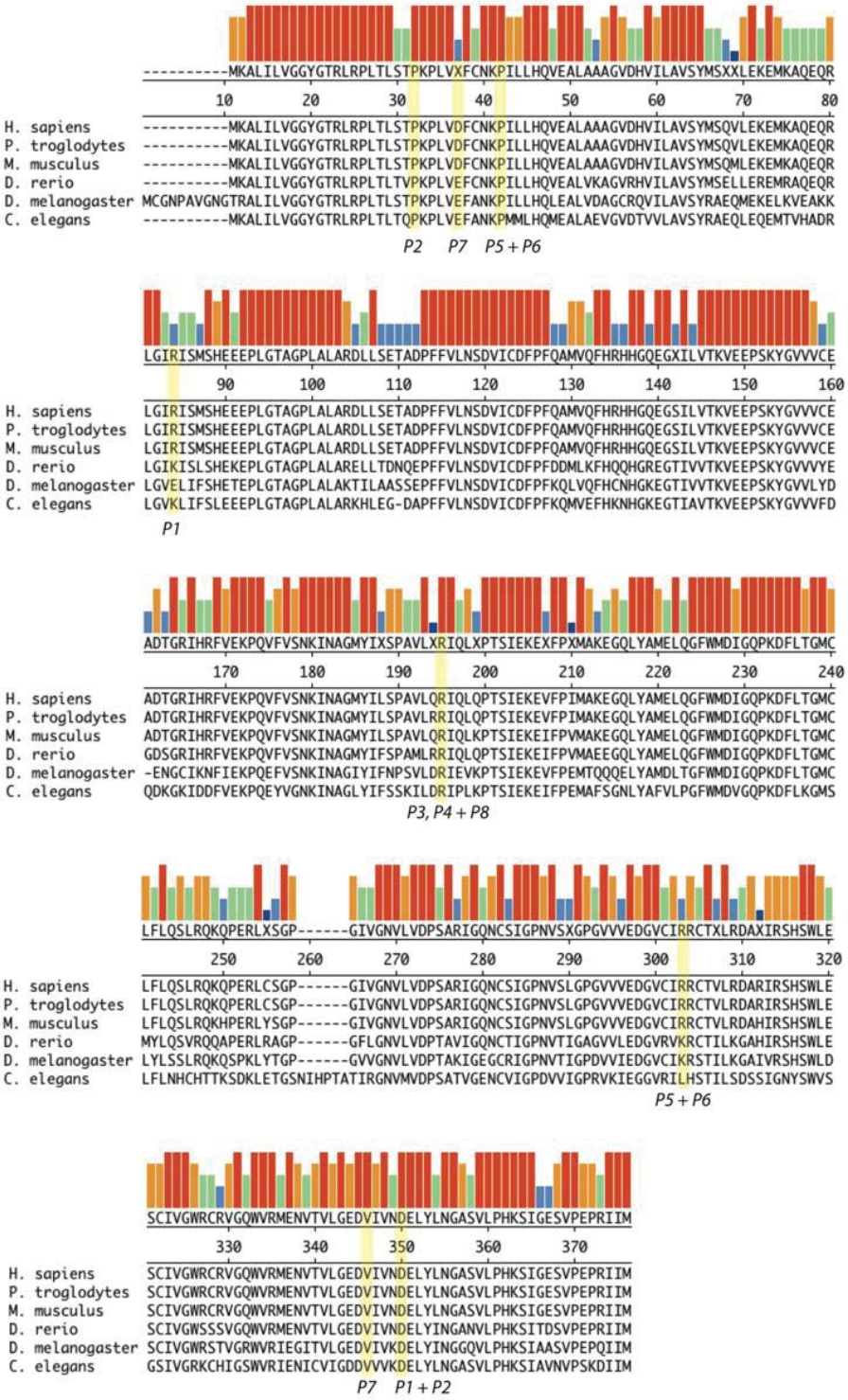

Figure 4 - 4: Protein alignment showing conservation of GMPPB.
Sequences are *Homo sapiens* (ENSP00000418565), *Pan troglodytes* (ENSPTRP00000025773), *Mus musculus* (ENSMUSP00000107914), *Danio rerio* (ENSDARP00000022618), *Drosophila melanogaster* (FBpp0078511), and *Caenorhabditis elegans* (C42C1.5). The height and colour of the bars indicates the degree of conservation of each amino acid residue. The residues altered in the muscular dystrophy cases are highlighted in yellow. This figure and legend have been published (313).

4.3.2 Expression of *b3galnt2* and *gmppb* throughout early zebrafish development

To investigate temporal expression patterns in zebrafish development I reverse transcribed RNA extracted from zebrafish embryos at various stages of development, and amplified a fragment of each gene of interest. Both *b3galnt2* and *gmppb* are clearly expressed at each of the stages tested (Figure 4-5).

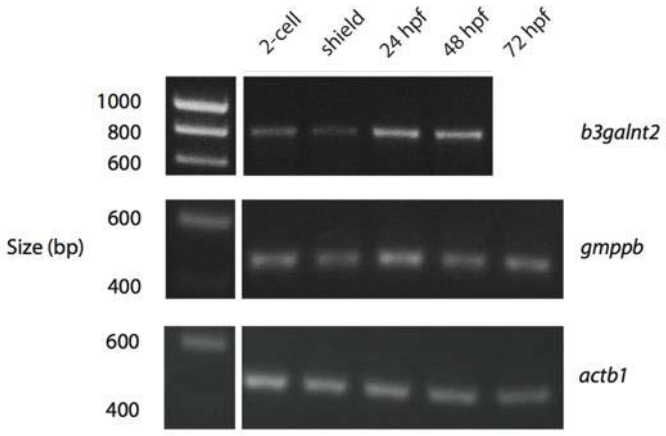

Figure 4 - 5: Reverse transcription PCR shows expression of *b3galnt2* and *gmppb* throughout early zebrafish development.
I performed reverse transcription PCR (RT-PCR) on RNA extracted from wildtype zebrafish embryos at the five developmental stages indicated. Expression of *b3galnt2* and *gmppb* was detected at each stage, using PCR amplification of cDNA fragments. An *actb1* fragment indicates near-equivalent amounts of input cDNA. This figure and legend have been published in a modified form (313, 314).

This is consistent with published *in-situ* hybridisation data suggesting that *b3galnt2* is ubiquitously expressed at early stages and becomes more anteriorly localised by 60 hpf (320), and that at early developmental stages (1-4 somites to 10-13 somites) *gmppb* is expressed primarily in the notochord, periderm and polster, but by prim-15 stage, expression is primarily localised to the head (321).

4.3.3 Finding the optimal morpholino oligonucleotide and dose for *b3galnt2* and *gmppb*

I injected a range of quantities of each MO into zebrafish embryos, and examined their phenotype, to determine the optimal MO and dose for each gene of interest (data not shown). For *b3galnt2* I found that both MOs resulted in similar phenotypes, suggesting they do specifically knock down *b3galnt2*. However, even at very low doses (2.5 ng) the SB MO had such a severe effect on muscle development that normal structures including fibres and myosepta were not discernable, whereas the TB MO had a clearer and more moderate phenotype.

A proportion of MOs cause non-specific *p53* upregulation. This results in phenotypes that can be mistakenly attributed to knockdown of the gene of interest (302, 303). I tested whether this would occur with the *b3galnt2* TB MO, by injecting 4 ng of the TB MO with or without a *p53* TB MO at 2 ng, and comparing the resulting phenotypes. At 24 hpf there was some neurodegeneration in the brains of the embryos without the *p53* MO, which was not present where the *p53* MO had been coinjected (Figure 4-6). Therefore, unless otherwise stated, for *b3galnt2* the TB MO was injected at 4 ng along with a *p53* TB MO at 2 ng.

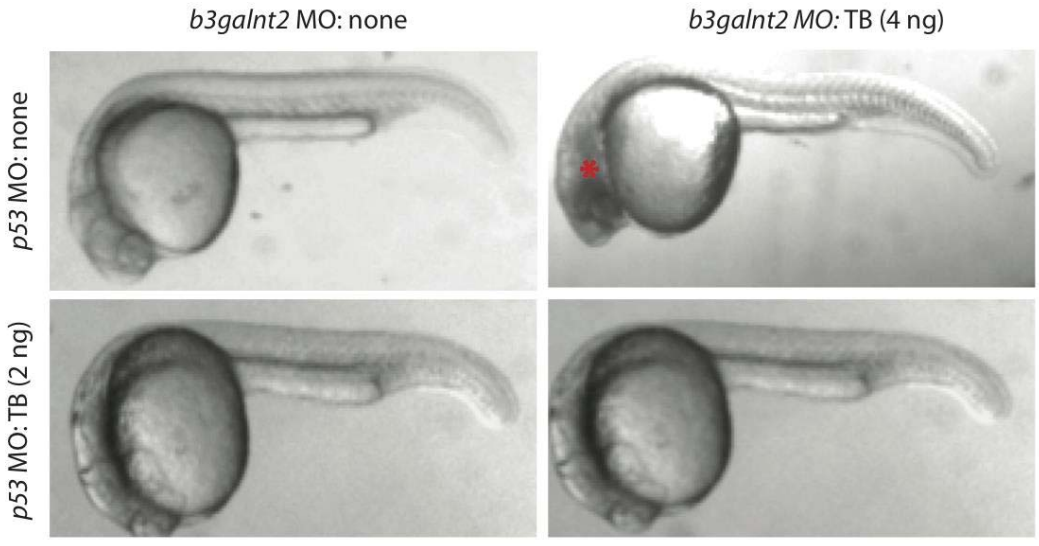

Figure 4 - 6: Coinjection of *p53* MO rescues the neurodegeneration induced by *b3galnt2* MO.
Injection of the *b3galnt2* TB MO (4 ng) on its own induces neurodegeneration, which makes the brain appear cloudy and dark (red asterisk). This is rescued then *p53* TB MO (2 ng) is coinjected, indicating that neurodegeneration is a non-specific phenotype (302, 303).

For *gmppb*, the TB MO and the first SB MO gave no discernable phenotype, even at very high doses (12.5 ng). The third SB MO also gave no phenotype. The second SB MO, which is predicted to result in skipping of the fifth coding exon**,** gave a consistent and relevant phenotype at reasonable doses, without excessive lethality. Therefore, unless otherwise stated, the second SB MO was used at 3 ng dose, coinjected with *p53* MO 6 ng.

4.3.4 Morpholino oligonucleotides reduce the expression of *b3galnt2* and *gmppb*

To confirm the efficacy and specificity of the *b3galnt2* MO I cloned *b3galnt2* into a GFP expression vector, made RNA from this, and injected it into zebrafish embryos (25 pg). At 24 hpf I showed using confocal microscopy that these embryos express the wildtype recombinant GFP-tagged *b3galnt2* RNA. This was suppressed when coinjected with *b3galnt2* MO (Figure 4-7), indicating that the *b3galnt2* MO effectively knocks down *b3galnt2*.

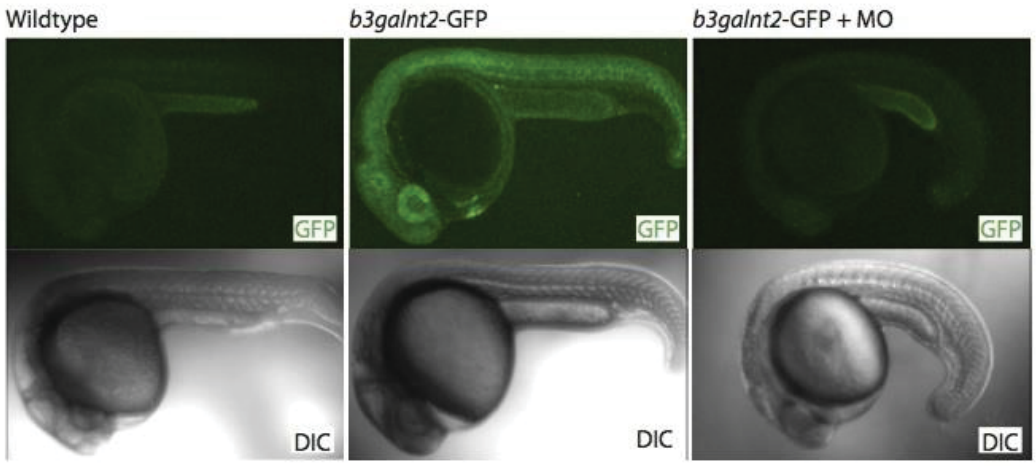

Figure 4 - 7: The *b3galnt2* MO inhibits expression of recombinant GFP-tagged *b3galnt2* RNA.
Embryos express wildtype recombinant GFP-tagged *b3galnt2* RNA (25 pg). This is suppressed when coinjected with the MO (*b3galnt2* TB 4 ng coinjected with *p53* TB 2 ng). Photographs were taken at 24 hpf. This figure and legend have been published (314).

To test the efficacy of the *gmppb* MO, I extracted RNA from wildtype and MO-injected embryos and performed reverse transcription PCR (RT-PCR) with primers flanking the MO binding site. I found that the morphants had a lower abundance of gmppb cDNA than wildtype, whereas the abundance of cDNA of a housekeeping gene (*actb1*) was approximately equal in the two groups, indicating that the *gmppb* MO effectively knocks down *gmppb* (Figure 4-8).

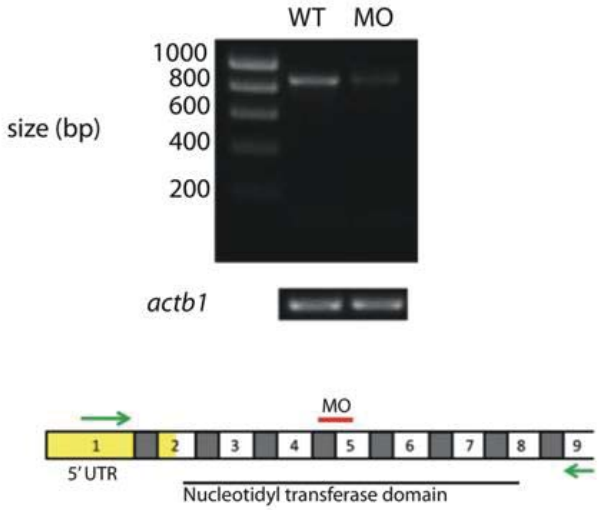

Figure 4 - 8: The *gmppb* splice blocking MO disrupts RNA splicing.
I extracted RNA from 48 hpf wildtype zebrafish embryos (WT) alongside embryos injected with the *gmppb* SB MO (2.5 ng, 5 ng and 7.5 ng pooled). I performed RT-PCR and PCR amplified a ~900 bp fragment of *gmppb* cDNA using primers (indicated by green arrows on schematic diagram) that bind either side of the MO binding site (indicated by red line). A clear reduction in band intensity is seen in the MO embryos, indicating that the MO disrupts correct mRNA splicing. An *actb1* fragment indicates near-equivalent amounts of input cDNA. This figure and legend have been published (313).

4.3.5 *b3galnt2* morphants have gross morphological defects including hydrocephalus and impaired motility

I examined the gross morphological phenotype of *b3galnt2* knockdown embryos at 48 hpf using light microscopy, and compared it to wildtype embryos. Consistently observed phenotypes were hydrocephalus, curvature of the tail, severely impaired motility, mild retinal degeneration, growth restriction, pericardial effusion, enlarged and dysmorphic yolk, delayed or failed hatching, and mild hypopigmentation (Figure 4-9A).

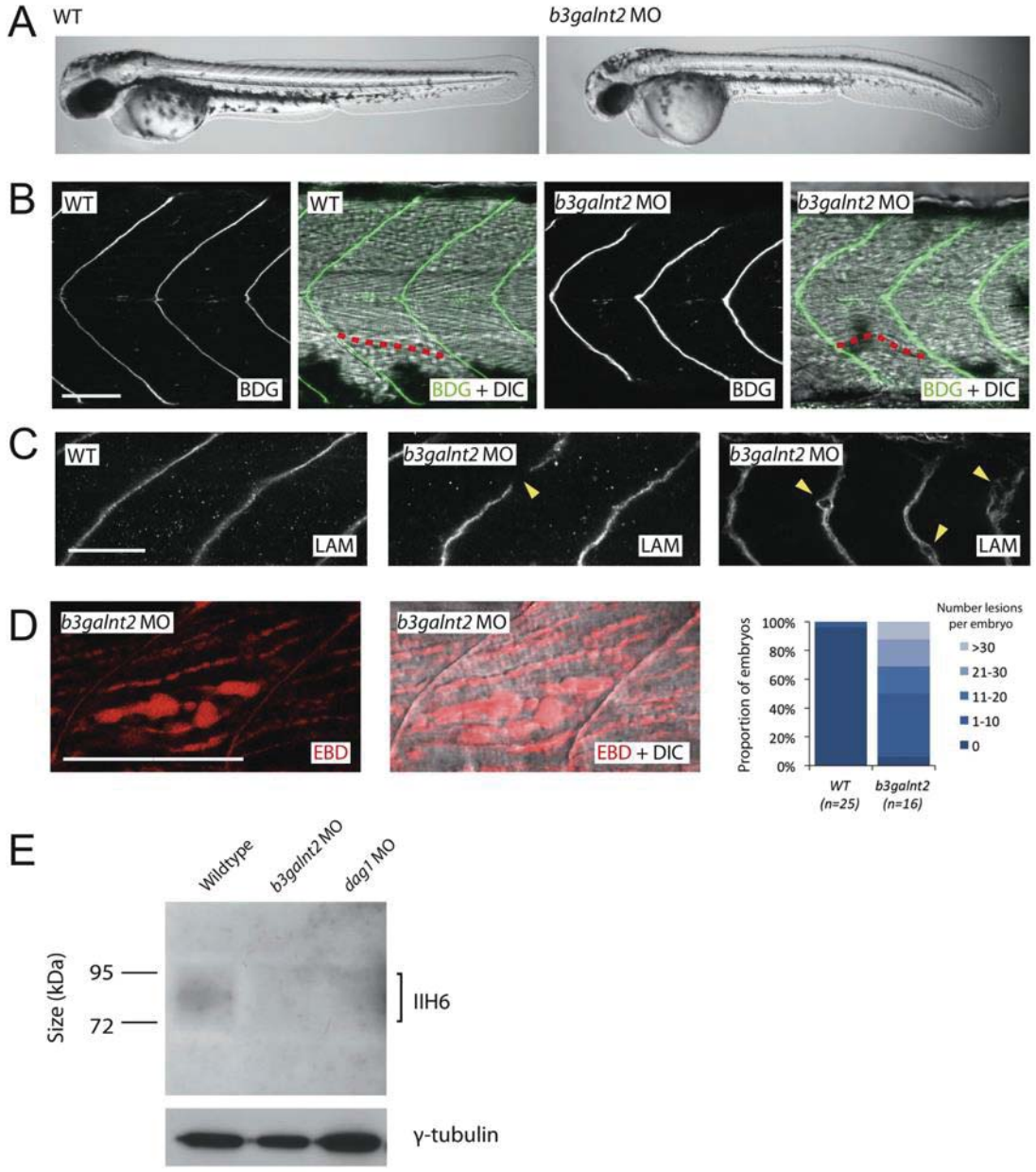

Figure 4 - 9: *b3galnt2* knockdown zebrafish embryos have muscle defects and hypoglycosylated α-DG at 48 hpf.
(A) Whole-mount pictures of live embryos show gross morphological defects. **(B)** Immunofluorescence staining by an antibody against β-DG and differential interference contrast (DIC) microscopy showed that the muscle fibres are disordered. One sample fibre is highlighted in red. **(C)** Immunofluorescence staining by an antibody against laminin (LAM) shows gaps in the myosepta of knockdown embryos and degeneration of the ECM. **(D)** Evans blue dye assay (EBD) highlights frequent lesions between muscle fibres in *b3galnt2* morphants, which are very rarely seen in wildtype embryos. **(A–D)** *b3galnt2* MO: *b3galnt2* TB MO (4 ng) coinjected with *p53* TB MO (2 ng). Scale bars represent 50 mm. **(E)** Immunoblotting by isolated microsome protein from 48 hpf embryos and IIH6 antibody showed a reduction in glycosylated α-DG in the *b3galnt2* knockdown embryos. *b3galnt2* MO: *b3galnt2* TB MO (5 ng) coinjected with *p53* TB MO (2.5 ng). *dag1* MO: *dag1* TB MO (5 ng). This figure and legend have been published (314).

The occurrence of hydrocephalus is particularly interesting as hydrocephalus was observed in three of the muscular dystrophy patients. Hydrocephalus is also caused by knocking down other dystroglycanopathy-associated genes such as *fktn*, *fkrp*, and *ispd* in zebrafish embryos (282). I counted the number of embryos that had hydrocephalus in a group of morphants and a group of wildtype embryos. 72% of *b3galnt2* morphants had some degree of hydrocephalus compared to 0% of wild type (Table 4-5). This is a highly significant difference ($p = 2.2 \times 10^{-16}$, Fisher's exact test).

	Wildtype	**Morphant**
Without hydrocephalus	72 (100%)	11 (28%)
With hydrocephalus	0 (0%)	28 (72%)
Total	72	39

Table 4 - 5: b3galnt2 morphants are significantly more likely to have hydrocephalus than wildtype embryos.
Morphant = *b3galnt2* TB 4ng + *p53* 2ng. Classified according to appearance under light microscope at 48 hpf. Fisher exact test; $p = 2.2 \times 10^{-16}$.

4.3.6 *b3galnt2* morphants have muscle defects including gaps in the myosepta and lesions between fibres

To characterise the muscle phenotype, I performed immunofluorescence staining. Compared to the chevron-shaped somite boundaries flanking straight muscle fibres in wildtype embryos, *b3galnt2* morphants consistently showed slightly U-shaped somites and disordered muscle fibres (Figure 4-9B). Laminin staining revealed occasional gaps in the myosepta (the connective tissues where the muscle fibres anchor), suggesting disruption of the ECM (Figure 4-9C).

Next, I quantified the severity of muscle damage by the EBD assay. EBD is an azo dye that binds to proteins such as albumin and is transported in the serum. It fluoresces upon protein binding and infiltrates muscle, where it penetrates compromised sarcolemma and accumulates at lesions between muscle fibres (interfibre spaces) (310, 322). Compared to wildtype embryos, *b3galnt2* morphants showed more severely damaged muscle, with increased number of lesions, ranging from less than 10 to more than 30 lesions per embryo (Figure 4-9D).

4.3.7 *b3galnt2* morphants have hypoglycosylated α-dystroglycan

Immunoblot analysis with the IIH6 antibody on protein extracts from wildtype embryos and *b3galnt2* morphants showed a reduction of the IIH6 signal in *b3galnt2* morphants, indicating that knockdown of *b3galnt2* led to reduced functional glycosylation of α-DG (Figure 4-9E). This is consistent with the human data and strongly suggests that this may be the molecular mechanism behind the phenotypes described.

4.3.8 Coinjection with wildtype RNA fails to rescue the *b3galnt2* morphant phenotype

In experiments using MOs, coinjection of wildtype RNA of the targeted gene can reduce the severity of the phenotype, providing supporting evidence that the phenotype is specific to knockdown of the gene of interest (323). To this end, I injected GFP-tagged wildtype human *B3GALNT2* mRNA (100 pg) into the cell of zebrafish embryos at the 1-cell stage, along with the MO. At 24 hpf I confirmed the expression of wildtype recombinant GFP-tagged *B3GALNT2* RNA by fluorescence microscopy. At 48 hpf I phenotyped the embryos by examining gross morphology, performing the EBD assay, and measuring the diameter of the eyes.

Human wildtype *B3GALNT2* RNA did not rescue the phenotype of *b3galnt2* zebrafish morphants. Compared to controls that had only been injected with *b3galnt2* MO, embryos that had been injected with *b3galnt2* MO and GFP-tagged *B3GALNT2* RNA had slightly more severe morphological defects such as curvature (Figure 4-10A). They had severe muscle lesions, with a slightly higher median number of lesions per embryo than controls that had only been injected with *b3galnt2* MO (Figure 4-10B). This difference is not statistically significant ($p=0.2$). They had a reduction in eye diameter compared to wildtype embryos and controls that had only been injected with GFP-tagged *B3GALNT2* RNA, which was not significantly different to that seen in controls that had only been injected with *b3galnt2* MO ($p=0.26$) (Figure 4-10C). Finally, a slightly higher proportion of embryos that had been injected with *b3galnt2* MO and GFP-tagged *B3GALNT2* RNA had hydrocephalus than controls that had only been injected with *b3galnt2* MO (Figure 4-10D). Control embryos that had only been injected with GFP-tagged *B3GALNT2* RNA were indistinguishable from wildtype embryos, except that a small proportion of them had mild hydrocephalus, whereas none of the

uninjected embryos did (Figure 4-10D). Increasing the injected dose of RNA to 200 pg also did not result in phenotypic rescue (data not shown).

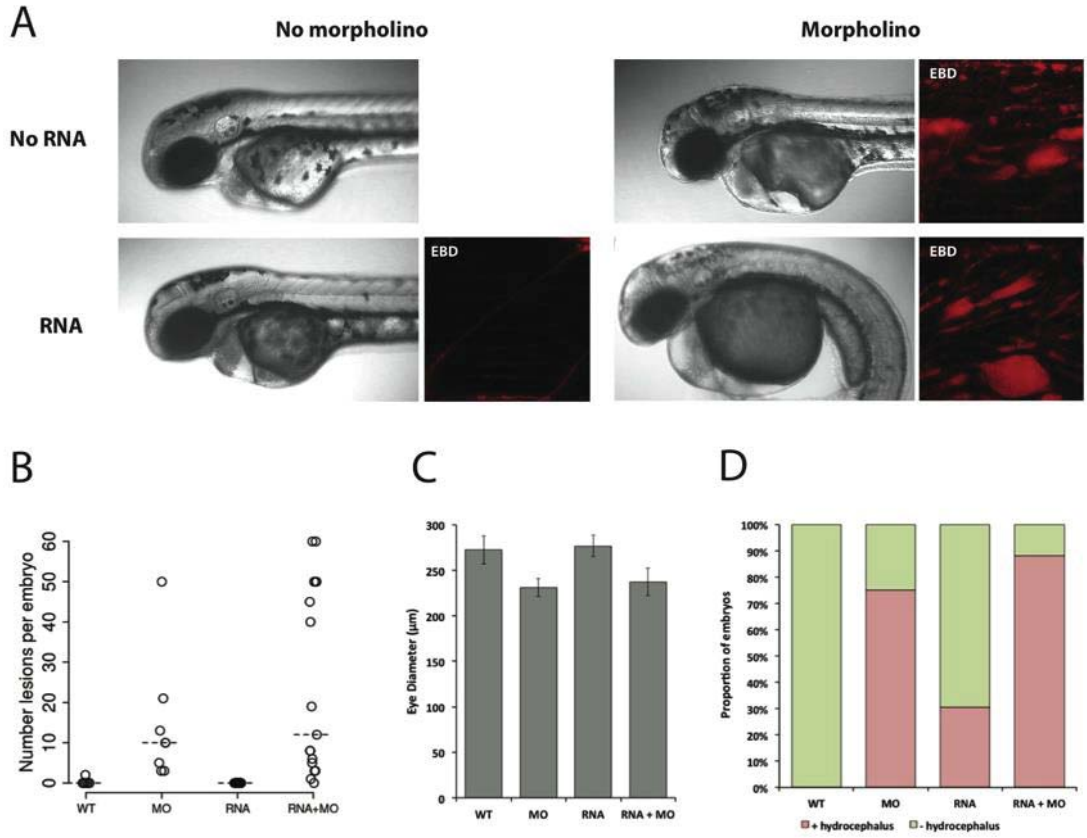

Figure 4 - 10: Coinjection with wildtype human *B3GALNT2* RNA fails to rescue the *b3galnt2* morphant phenotype.
48 hpf zebrafish embryos. Expression of GFP was confirmed at 24 hpf. **(A-D)** MO or Morpholino = *b3galnt2* TB MO (4 ng) coinjected with *p53* TB MO (2 ng). RNA = GFP-tagged wildtype human *B3GALNT2* mRNA (100 pg), WT = wildtype. Sample size for the WT, MO, RNA and RNA+MO groups respectively are 9, 8, 23 and 17 embryos. **(A)** Phenotype of anterior portion of embryos photographed using a confocal microscope. EBD = Evans blue dye assay. The appearance of pericardial oedema may be exaggerated in these images due to the injection of EBD. No photograph of the EBD assay on wildtype embryos was taken in this experiment. **(B)** Number of lesions. Dashed horizontal line indicates median. **(C)** Diameter of eyes. **(D)** Proportion of embryos that have hydrocephalus.

I tested whether the divergence between human and zebrafish orthologues of B3GALNT2 could explain this failure to rescue, by injecting GFP-tagged wildtype zebrafish *b3galnt2* mRNA (200 pg) into zebrafish embryos, along with the *b3galnt2* MO, and performing the EBD assay at 48 hpf. GFP-tagged wildtype zebrafish *b3galnt2* mRNA resulted in no improvement in the gross morphology of the embryos, or the appearance of muscle lesions (Figure 4-11).

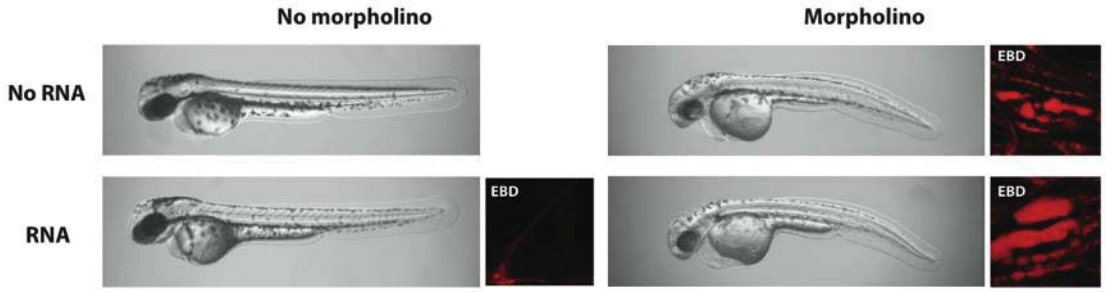

Figure 4 - 11: Coinjection with wildtype zebrafish *b3galnt2* RNA fails to rescue the *b3galnt2* morphant phenotype.
48 hpf zebrafish embryos. Expression of GFP was confirmed at 24 hpf. Morpholino = *b3galnt2* TB MO (4 ng) coinjected with *p53* TB MO (2 ng). RNA = GFP-tagged wildtype zebrafish *b3galnt2* mRNA (200 pg). Phenotype of embryos photographed using light microscope. EBD=Evans blue dye assay. EBD assay was not performed on wildtype embryos in this experiment. Each whole embryo picture is representative of at least 20 embryos, and each EBD picture is representative of at least five embryos.

4.3.9 *gmppb* morphants have gross morphological defects including micropthalmia and impaired motility

I examined the gross morphological phenotype of *gmppb* knockdown embryos at 48 hpf using light microscopy, and compared it to wildtype embryos. Morphologically, *gmppb* morphants were shorter than wildtype uninjected embryos at 48 hpf and often had bent tails. Other phenotypes included hypopigmentation, micropthalmia, hydrocephalus, increased lethality, and reduced motility (Figure 4-12A). The difference in diameter of the eyes of wildtype and *gmppb* morphant embryos was statistically significant, following correction for body length ($p < 1 \times 10^{-7}$; Figure 4-12B). Although none of the dystroglycanopathy cases reported micropthalmia, this is a phenotype that is common in individuals with severe forms of CMD, such as WWS and MEB (216, 221, 254, 281, 282, 314).

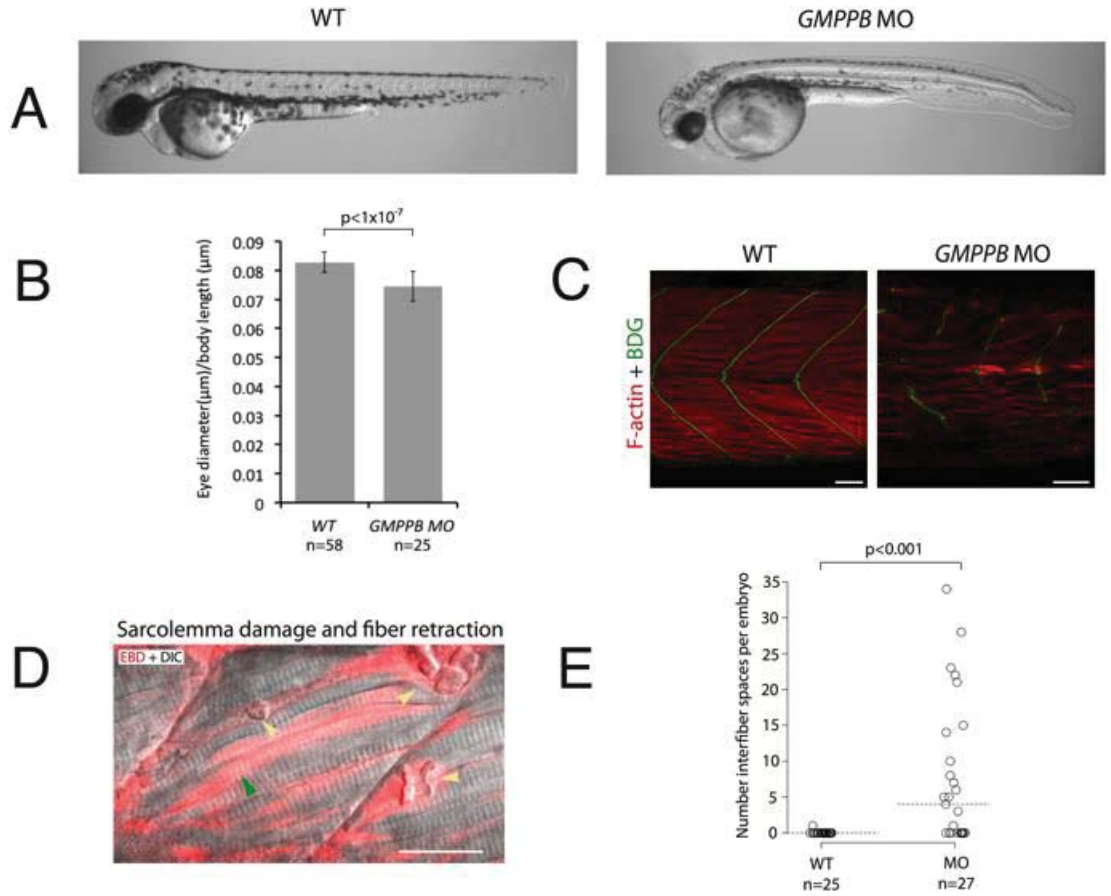

Figure 4 - 12: *gmppb* knockdown zebrafish embryos have morphological defects, damaged muscle, and hypoglycosylated α-DG at 48 hpf.
(A) Bright-field microscopy of live embryos shows morphological defects of the *gmppb*-MO-injected embryos (injected with *gmppb* splice-blocking MO 3 ng + p53 MO 6 ng) as compared to uninjected wildtype embryos. **(B)** I measured the eye diameter of *gmppb* morphants and wildtype embryos, normalised each eye diameter measurement to the embryo's body length, and assessed statistical significance using an unpaired two-tailed t-test. **(C)** Phalloidin staining of filamentous actin (red) and immunostaining with an antibody against β-DG (green). **(D)** Live *gmppb*-MO-injected embryos injected with EBD (red) and imaged by confocal microscopy. Some fibres showed EBD infiltration, indicating damage to the sarcolemma (green arrow), and other fibres detached from the myosepta and retracted (yellow arrow) and thus left a space. The following abbreviation is used: DIC, differential interference contrast. The scale bar represents 25 mm. **(E)** *gmppb* morphants have significantly more interfibre spaces than do wildtype uninjected embryos. The horizontal dotted line shows the median. This figure and legend have been published in a modified form (313).

4.3.10 *gmppb* morphants have muscle defects including disordered fibres, incomplete myosepta, and interfibre spaces

To characterise muscle defects in *gmppb* knockdown zebrafish embryos, I used phalloidin to label filamentous actin, along with immunostaining with antibodies against β-DG (which localises to the myosepta, the connective tissue to which muscle fibres anchor). I observed that the muscle fibres in *gmppb* morphants were sparse and disordered. Furthermore, fibres were frequently observed to span two somites, indicating damage or incomplete development of the myosepta (Figure 4-12C).

To further explore the muscle phenotypes in *gmppb* morphants, I injected EBD into the pericardium of embryos at 48 hpf. Compared with uninjected control embryos, *gmppb* morphants had significantly more EBD accumulation within interfibre spaces ($p < 0.001$; Figure 4-12D-E). In addition, EBD infiltrated both retracted and some intact muscle fibres in *gmppb*-knockdown embryos, suggesting that sarcolemma integrity was compromised prior to muscle fibre breakdown.

4.3.11 *gmppb* morphants have hypoglycosylated α-dystroglycan

Next, I investigated whether the laminin-binding glycan on α-DG is reduced in *gmppb* morphants. To do this, I performed immunoblots with the IIH6 antibody on membrane proteins enriched from wildtype embryos and *gmppb* morphants, as well as *dag1* morphants as a negative control. After normalisation to γ-tubulin loading control, *gmppb* morphants showed a slight but clear reduction in IIH6 levels (71% of that of the wildtype embryos) and *dag1* morphants showed a strong reduction in IIH6 (15% of that of the wild-type embryos) (Figure 4-13A). To confirm this finding, I performed double immunostaining with the IIH6 antibody and an antibody against laminins. In wildtype embryos, laminin and glycosylated α-DG colocalised at the myosepta. In *gmppb* morphants, the IIH6 staining was severely reduced, and laminin staining revealed widened myosepta, indicating a reduction in glycosylation of a-DG associated with abnormal basement-membrane structure (Figure 4-13B).

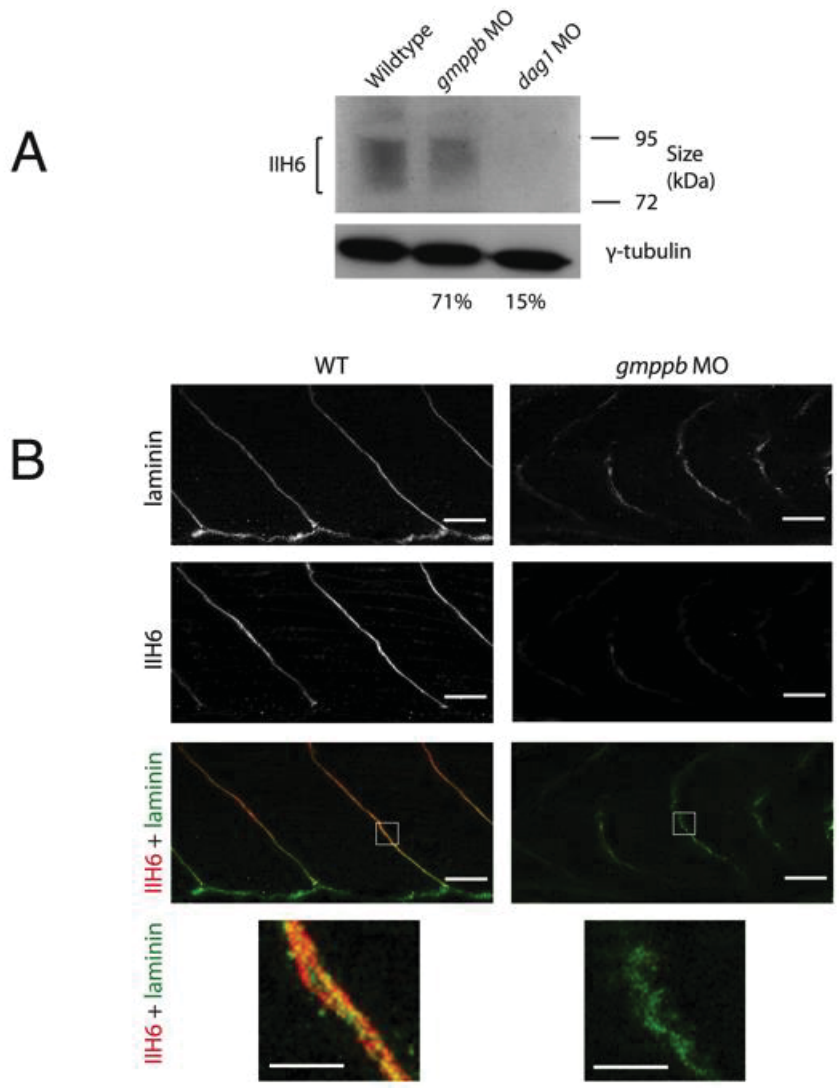

Figure 4 - 13: α-DG of *gmppb* morphants is hypoglycosylated relative to wildtype embryos.
(A) An immunoblot shows a reduction in α-DG glycosylation. Percentage figures indicate the intensity of morphant bands relative to that of the wildtype and are adjusted for the γ-tubulin loading control. ''*gmppb* MO'' indicates embryos injected with *gmppb* MO 3 ng + p53 MO 6 ng, and ''*dag1* MO'' indicates embryos injected with *dag1* TB MO (5 ng). **(B)** Antibodies used are against laminins, (polyclonal antibody L9393, Sigma-Aldrich, Dorset, UK), and glycosylated α-DG (monoclonal antibody IIH6, from Professor Kevin Campbell). Staining of laminins revealed abnormal myosepta structure, while the fluorescent intensity of IIH6 epitope was significantly reduced in *gmppb* morphants. White square indicates magnified area. Scale bar = 25 μm except for magnified area where scale bar = 6 μm. This figure and legend have been published in a modified form (313).

4.4 Discussion

4.4.1 Summary

Zebrafish are an appropriate model with which to study many human diseases, and have been particularly fruitful in the study of muscular dystrophy such as dystroglycanopathy. Here I have shown that zebrafish are an appropriate model with which to study the two dystroglycanopathy-associated genes *B3GALNT2* and *GMPPB*, because of moderate or high levels of conservation respectively, and appropriate temporal expression. I designed MOs to knock down the orthologues of both genes, and I optimised and validated the MOs. I performed extensive characterisation of the phenotypes of my two models. In each case, I demonstrate that aspects of the patients' phenotypes have been recapitulated in the zebrafish embryos at three levels: gross morphology, muscle structure and molecular level (hypoglycosylation of α-DG). This contributes to the body of data, which also includes the clinical and cellular data generated by colleagues as described in section 4.1.7, that supports the conclusion that pathogenic variants in *B3GALNT2* or *GMPPB* can cause dystroglycanopathy.

4.4.2 Phenotypic rescue

One method of confirming that knockdown of a gene of interest specifically causes a morphant phenotype, is to coinject wildtype RNA of the targeted gene and demonstrate reduction in the severity of the phenotype (323). The RNA can be from the zebrafish gene, or, if there is sufficient homology between zebrafish and humans, RNA from the human orthologue can be used. This letter method has the advantage that, if rescue is achieved, one can next demonstrate failure to rescue with a construct containing the patient's variant, providing compelling evidence of pathogenicity of that variant (254, 324).

In this study, I have demonstrated that neither zebrafish *b3galnt2* RNA nor human *B3GALNT2* RNA rescues the phenotype of *b3galnt2* morphant zebrafish embryos. This failure to rescue in no way suggests that the phenotypes I have detailed are *not* specific to knockdown of *b3galnt2*. While expression of endogenous genes is subject to exquisitely specific and complex spatial-temporal regulation, rescue experiments

induce ubiquitous, unregulated, high levels of expression, which can render rescue ineffective, and adversely affect embryonic development (325, 326). Furthermore, GFP tags in some cases alter the behaviour of recombinant proteins (327). Therefore, it is to be anticipated that injection of wildtype RNA does not always rescue a morphant phenotype. My results highlight the importance of appropriate spatial-temporal expression of *b3galnt2*.

4.4.3 The function of B3GALNT2

Very recently, a new patient with pathogenic *B3GALNT2* variants has been identified (328). This girl had dystroglycanopathy, but with a milder phenotype than the patients we described, expanding the phenotypic spectrum of this group of patients.

In addition to its role in muscle integrity, α-DG acts as a receptor by which some pathogens including Lassa virus, enter cells (329). When α-DG is hypoglycosylated, these pathogens cannot enter. A recent study elegantly exploited this fact to screen for genes that may cause dystroglycanopathy (256). The authors used mutagenised, haploid HAP1 cell lines, and identified those that were resistant to virus entry. They next identified the genes that were mutated in these cell lines, concluding that these were likely to be involved in α-DG glycosylation, and were therefore good candidates for dystroglycanopathy. Some known genes were identified, including *LARGE*, *ISPD* and *DAG1*, as were several novel genes including *B3GALNT2*. This further emphasises the importance of *B3GALNT2* in the glycosylation of α-DG, and the pathology of dystroglycanopathy.

The mannose residues initially added to the mucin-like domain of α-DG are extensively extended and branched. For example, a β1,4 linked GlcNAc groups can be added to a mannose residue by an unknown GlcNAc transferase (227). After this, a novel β1,3-N-acetylgalactosaminyltransferase acts on this structure to complete the trisaccharide GalNAcβ1-3-GlcNAc-β1,4-Man (315). This trisaccharide is required for ligand binding of α-DG. We speculated that this novel β1,3-N-acetylgalactosaminyltransferase was B3GALNT2 (314). This trisaccharide is phosphorylated and further extended by LARGE (249).

An important subsequent study used various biochemical techniques including mass spectrometry to demonstrate that GTDC2 is the unknown GlcNAc transferase mentioned above, and furthermore confirmed our speculation as to the precise function

of B3GALNT2 (258). Additionally, the authors found that POMK phosphorylates the 6-position of the mannose of this trisaccharide, which is required for the activity of LARGE (Figure 4-14). Thus, the pieces of the dystroglycanopathy puzzle are beginning to fall into place.

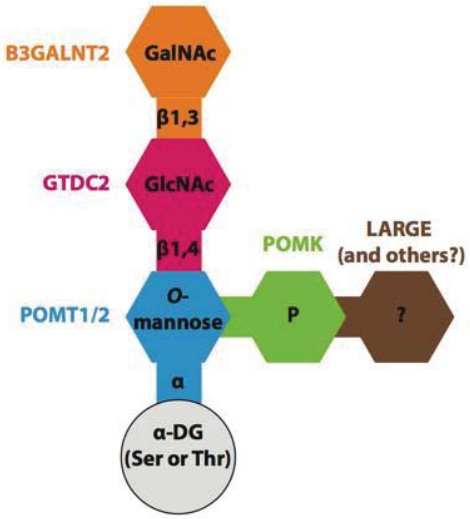

Figure 4 - 14: B3GALNT2 catalyses the synthesis of the trisaccharide GalNAcβ1-3-GlcNAc-β1,4-Man, which is required for laminin binding.
The enzyme that catalyses the addition of each monosaccharide is shown in the same colour as the monosaccharide. The precise mechanism of action of LARGE is not yet fully understood, and it may require the activity of other enzymes (258).

4.4.4 The function of GMPPB

Since the publication of our study, another group also found pathogenic variants in *GMPPB* by exome sequencing in a single family with dystroglycanopathy, brain abnormalities, and seizures, confirming the importance of *GMPPB* in the glycosylation of α-DG (330).

GDP-man is the substrate required for N-glycosylation, and also for the synthesis of Dol-P-Man, which is in turn required for all forms of glycosylation (Figure 4-2). The synthesis of Dol-P-Man is catalysed by the DPM synthase complex, which consists of DPM1, DPM2, and DPM3. One might therefore expect the phenotype of patients with pathogenic *GMPPB* variants to be more similar to those with pathogenic variants in components of the DPM synthase complex, than those with pathogenic variants in genes encoding more downstream enzymes in the pathway such as *B3GALNT2* or

POMT1. For example, patients with pathogenic variants in components of the DPM synthase complex or *DOLK* can also have defects in N-glycosylation, and an associated phenotype of CDG (233-237). However, surprisingly, patients with pathogenic *GMPPB* variants have neither defective N-glycosylation, nor signs of CDG (313).

There are several possible reasons for this. For example, Dol-P-Man might not be equally important in the N-glycosylation and O-mannosylation pathways. Previous research has shown that two enzymes in the glycosylation pathway compete for a common substrate, and can use this substrate differently, supporting this hypothesis (331). Alternatively, N-glycosylation could occur before O-mannosylation and the amount of Dol-P-Man would therefore be depleted by N-glycosylation before O-mannosylation starts. Also, the possibility of tissue-specific defects in N-glycosylation cannot be excluded.

Experiments in the pig showed that the enzyme GDP-man pyrophosphorylase consists of two subunits (317). GMPPA has GDP-Glc pyrophosphorylase activity, whereas GMPPB catalyses the formation of GDP-mannose from Mannose-1-phosphate and GTP. GMPPB has a high affinity for synthesising GDP-mannose, and a low but detectable affinity for synthesising GDP-Glc. Studies on the affinity of GMPPA to synthesise GDP-man have not been published thus far, but these results suggest some functional overlap which may be relevant to the phenotype of our patients. In humans, GMPPA and GMPPB are 30% identical.

Since the publication of our study, another group found that pathogenic variants in *GMPPA* cause a syndrome of achalasia (constriction of gastric cardia), deficiency of tear secretion, ID, gait abnormalities, neurological defects, and feeding difficulties (332). Interestingly, these patients also do not have a N-glycosylation defect. The amount of GDP-man increases in the cells of these patients, suggesting that GMPPA might negatively regulate GMPPB.

GMPPB orthologues have been knocked down in various species, including *Saccharomyces cerevisiae*, *Aspergillus fumigatus*, *Arabidopsis thaliana*, *Solanum tuberosum*, *Trypanosoma brucei*, and *Leishmania mexicana* (333-339). This caused glycosylation defects and a range of pathogenic phenotypes from defective cell growth to lethality. This severity suggests that complete loss of function of *GMPPB* might be lethal. This hypothesis is supported by the fact that we did not identify any case with two null alleles (313), and also by the fact that *GMPPB* was not identified as a

candidate dystroglycanopathy-associated gene by the lassa virus screen I have described (256).

4.4.5 Zebrafish phenotypes in context

With the exception of *dag1* (307), no stable germ line zebrafish mutants have been generated as models of dystroglycanopathy. Instead, most groups have used MOs. Some studies that use MO knockdown technology in zebrafish embryos to study dystroglycanopathy-associated genes focus more on some phenotypes than others. For example Buysse *et al.* do not report extensively on the gross morphology of their *b3gnt1* model, and Avsar-Ban *et al.* do not report extensively on the muscle structure of their *pomt1* and *pomt2* models (16, 306). Additionally, some studies used coinjection of *p53* MO, and others did not. Studies of *fkrp* and *gtdc2* zebrafish models reported generalised neurodegeneration, manifesting as opaque appearance of neurons in the brain, in models (17, 254). However, this observation is a well-reported non-specific effect of MOs, which is often ameliorated by the coinjection of a *p53* MO (Figure 4-6) (302, 303), which neither of these studies did. Therefore it is conceivable that this observation in these studies, which initially appears to be an interesting phenotype specific to knockdown of these genes, is in fact an artifact.

Despite these differences in methodology and reporting, some interesting insights can be gained from comparing the phenotypes of the *b3galnt2* and *gmppb* models described here to zebrafish models of six other dystroglycanopathy-associated genes (*fkrp*, *b3gnt1*, *gtdc2*, *pomt1*, *pomt2* and *ispd*) (Table 4-6). The following phenotypes are universal, in every model for which they are described: developmental delay, curved tail, impaired motility, U-shaped somites, disordered muscle fibres and hypoglycosylated α-DG. These defining characteristics of zebrafish dystroglycanopathy models are all present in my *b3galnt2* and *gmppb* models. Other phenotypes are more gene-specific. For example, a dysmorphic yolk and hypopigmentation was observed only in models of *b3galnt2*, *gmppb*, and *pomt2*, pericardial effusion is observed only in models of *b3galnt2*, *fkrp*, and *pomt2*, and micropthalmia is observed only in models of *gmppb*, *fkrp*, *gtdc2*, and *ispd*.

As I have discussed, B3GALNT2 and GTDC2 are both involved in synthesising a trisaccharide on α-DG that is essential for ligand binding (258). Given the close proximity of the activity of these proteins in the glycosylation pathway, one might

assume that the phenotypes of the patients and the zebrafish models of these genes might be more similar to one another than they are to the phenotypes of patients or zebrafish models with deficiencies in genes active elsewhere in the glycosylation pathway. This does not appear to be the case (Table 4-6) (254, 314). This may be because of the methodological differences discussed, or it may be that B3GALNT2 and GTDC2 have targets in addition to α-DG, which may be different from one another. The known dystroglycanopathy-associated genes whose function is closest in the glycosylation pathway to *GMPPB* are *DPM1*, *DPM2*, *DPM3*, and *DOLK*. Zebrafish models of these genes have not yet been generated and studied. It would be interesting to see whether the phenotypes of these models would have some of the features more specific to my gmppb model, such as micropthalmia, retracted muscle fibres and damage to sarcolemma.

Gene	Reference	Developmental delay	Hydrocephalus	Microcephaly	Abnormal brain	Abnormal notochord	Impaired motility	Curved tail	Dysmorphic yolk	Hypopigmentation	Pericardial effusion	Micropthalmia	Retinal degeneration	U-shaped somites	Disordered muscle fibres	Gaps in myosepta	Lesions between muscle fibres	Retracted muscle fibres	Damage to sarcolemma	Hypoglycosylated α-DG
b3galnt2	(314)	Y	Y	N	N	X	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	Y
gmppb	(313)	Y	Y	N	N	X	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y
fkrp	(17, 305)	Y	N	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	X	Y	X	Y
b3gnt1	(16)	X	X	X	X	X	X	X	X	X	X	X	X	Y	Y	Y	Y	Y	Y	Y
gtdc2	(254)	Y	Y	Y	Y	X	Y	Y	N	N	N	Y	Y	Y	Y	X	X	X	X	Y
pomt1	(306)	Y	X	N	X	X	X	Y	N	N	N	N	N	Y	X	X	X	X	X	Y
pomt2	(306)	Y	X	N	X	X	X	Y	Y	Y	Y	N	Y	Y	X	X	X	X	X	Y
ispd	(282)	X	Y	N	Y	X	Y	X	X	X	X	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 4 - 6: Phenotypic comparison of zebrafish dystroglycanopathy models.
All models made by MO knockdown. Y = phenotype observed; N = phenotype not observed; X = phenotype not investigated or described.

4.4.6 Technical limitations of this study

There are technical challenges associated with the use of MOs, as discussed in section 4.1.5. In this study, I took various measures to limit the effect of these challenges on my results. For example, I coinjected a *p53* MO along with the MOs for my genes of interest, in order to limit any generalised p53 upregulation, which may have produced non-specific results (302, 303). I also confirmed that the *b3galnt2* and *gmppb* MOs inhibit the expression of *b3galnt2* and *gmppb*. Despite the technical limitations, MOs have hitherto been the method of choice for generating zebrafish models of dystroglycanopathy. The overlap between the phenotypes I observed in my two models, and the phenotypes observed in other dystroglycanopathy models, support the hypothesis that those phenotypes are specific to dystroglycanopathy models, as discussed in section 4.4.5.

Nevertheless, it is likely that in coming years the method of choice for generating zebrafish models of dystroglycanopathy will move towards CRISPRs. CRISPRs are cheap, easy to generate, and allow very specific, targeted mutations to be introduced into the genome (300). While the experiments would take longer than for MOs, this disadvantage would be outweighed by the fact that the results are likely to be much more consistent and specific. Technological advances continue to improve the specificity of CRISPRs (340). The existence of endogenous CRISPRs as an immune-like mechanism is prokaryotes has been known for years (341), however their use as a genome-editing tool was not developed until 2012, and are still in the process of being optimised and becoming widely available (342). Had CRISPRs been available at the time this study was designed (2011), it is probable that we would have elected to use them rather than MOs to generate zebrafish models of *B3GALNT2* and *GMPPB*.

4.4.7 Future research

A further experiment that could be performed is to attempt to rescue the phenotype of the *GMPPB* morphants by coinjection of wildtype human or zebrafish *GMPPB* RNA. Unfortunately, it was not possible to carry out this experiment in the time frame of this project. While a positive result from this experiment would have provided useful supportive evidence that the phenotypes we observed in the *GMPPB* morphants were specific to knockdown of *GMPPB*, my colleagues and I did not feel that this experiment

was essential for the following reasons. First, as discussed in section 4.4.2, coinjection of wildtype RNA does not always result in phenotypic rescue. Second, even without this experiment, we were confident in our conclusion that *GMPPB* variants in the patients were causative of disease, because the phenotype of the *gmppb* morphants was very consistent with those of the patients and those of other zebrafish models of dystroglycanopathy-associated genes, and because of all the other clinical and cellular evidence.

In the wider field of dystroglycanopathy research, many questions remain. Indeed, identification of the genes associated with disease is only the first step. Elucidating the precise role of each gene in the glycosylation of α-DG is vital to understand disease pathology at the molecular level. Here, the collaboration of clinicians, human geneticists, and model organism researchers with biochemists will prove fruitful. For example, Yoshida-Moriguchi *et al.* have shed light on the exact function of several dystroglycanopathy-associated genes, including *B3GALNT2* (258).

Several observations suggest that some dystroglycanopathy-associated genes may be involved in the glycosylation of other target proteins in addition to α-DG. These include the heterogeneity of dystroglycanopathy (both in terms of phenotypic severity and organ systems affected), the lack of correlation between the extent to which α-DG is hypoglycosylated, and the clinical severity of disease (292), and the known promiscuity of some bacterial glycosyltransferases (343-345). Identifying any other target proteins of dystroglycanopathy-associated genes would undoubtedly improve understanding of the disease. Similarly, identification of secondary modifiers may help to explain why some patients are less severely affected than others, and may point towards therapeutic targets.

Therapeutic sugar supplementation can treat patients with some glycosylation defects. The best example of this is CDG caused by pathogenic variants in the gene encoding mannose phosphate isomerase, which catalyses the conversion of fructose-6-phosphate to mannose-6-phosphate. The resulting deficiency in mannose-6-phosphate (which is a precursor to mannose-1-phosphate, the substrate of GMPPB) causes a multisystem phenotype including gastrointestinal and liver disease (346). Orally administrated mannose can improve clinical symptoms of these patients, and the levels of glycosylation of glycoproteins (347). This works because mannose can be converted to mannose-6-phosphate, catalysed by hexokinase. This is normally a minor alternative pathway, but is promoted in the presence of high doses of exogenous mannose.

Mannose supplementation in water improves the phenotype of a zebrafish MO model of CDG-1b (348). This suggests that a similar approach, using supplementation of mannose or GDP-mannose, could be a therapeutic avenue worth exploring in patients with pathogenic *GMPPB* variants. Furthermore, a zebrafish model could be a useful initial tool to test the viability of this idea. It is clear that the zebrafish will continue going from strength to strength as a model of dystroglycanopathy.