Downstream processing of a plant-made recombinant human therapeutic enzyme

by

Owen Mazac Pierce

B.Sc. (Hon.), Simon Fraser University, 2009

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Biology
Faculty of Science

SIMON FRASER UNIVERSITY

Fall 2014

Approval

Name: **Owen Mazac Pierce**

Degree: **Master of Science**

Title: ***Downstream processing of a plant-made recombinant human therapeutic enzyme***

Examining Committee: **Chair:** Dr. David Green
Associate Professor

Dr. Allison Kermode
Senior Supervisor
Professor

Dr. Sherryl Bisgrove
Supervisor
Associate Professor

Dr. David Vocadlo
Supervisor
Professor

Dr. Margo Moore
Internal Examiner
Professor
Department of Biology

Date Defended/Approved: July 25, 2014

Partial Copyright Licence

Abstract

Mucopolysaccharidosis I is a lysosomal storage disease that is characterized by a deficiency of the lysosomal hydrolase alpha-L-iduronidase. Enzyme replacement therapy with purified, recombinant IDUA is one treatment option available for individuals that inherit this disease. The goal of my thesis was to engineer a transgenic line of *Arabidopsis thaliana* that could express, in the seed tissue, human iduronidase with a therapeutically acceptable *N*-glycan profile. Human iduronidase was expressed at 5.7% of total soluble protein (TSP) in the *Arabidopsis cgl* mutant line, which is deficient in *N*-acetylglucosamine transferase I, and at 1.5% of TSP in the *Arabidopsis* Golgi-Mannosidase I (GM-I) knockout line. The iduronidase purified from both lines was able to hydrolyze the fluorescent iduronide compound, 4-methylumbelliferyliduronide (4-MUI), at a rate comparable to that of iduronidase produced in Chinese Hamster Ovary cells, commercially available as *Aldurazyme.* Both plant-derived forms of iduronidase possessed primarily high-mannose *N*-glycans (at 95% for *cgl*-iduronidase and 100% for GM1-iduronidase); however, the dominant glycoform in the *N*-glycan profile of *cgl*-iduronidase was $Man_5GlcNAc_2$, whereas most of the *N*-glycans on GM1-iduronidase contained six to eight mannose sugars. Additionally, *cgl*-iduronidase contained a small percentage of complex glycans (and thus potentially immunogenic xylose and/or fucose sugars), which were not found on GM1-iduronidase. Surprisingly GM1-iduronidase did not contain an *N*-glycan at Asn336, which is normally present on the native human iduronidase and on the recombinant iduronidase of *cgl Arabidopsis* seeds and Chinese hamster ovary cells. Of importance to creating a product that is of therapeutic value, another goal was to determine whether the plant-produced iduronidase was amenable to *in vitro* phosphorylation to create the mannose-6-phosphate tag necessary for sequestration and lysosomal delivery in human cells. Both plant-derived forms of the enzymes were able to be phosphorylated *in vitro* by the UDP-GlcNAc:lysosomal enzyme (GlcNAc)-1-phosphotransferase at a rate similar to that of *Aldurazyme*.

Keywords: lysosomal storage diseases; enzyme replacement therapy; plant-produced pharmaceuticals; alpha-L-iduronidase; molecular pharming; UDP-N-acetylglucosamine-lysosomal-enzyme *N*-acetylglucosaminephosphotransferase

Dedication

To my parents.

Acknowledgements

I would like to express my gratitude to all of those who assisted me during my studies at Simon Fraser University. Most importantly, I would like to acknowledge my senior supervisor, Dr. Allison Kermode, for her academic guidance and her support for my research.

I am appreciative for the technical instruction and assistance offered by Dr. Xu He, Dr. Tiehan Zhao, Dr. Ying Zeng, and Reza Babajani, of the Kermode Lab, who taught me the different techniques required for recombinant protein production in transgenic plants. I would like to thank Dr. Tracey Gloster for teaching me how to perform the enzyme kinetics measurements. Thanks to my fellow lab members, Ryan Thomas, Dhani Khalidisan, and Dr. Joost Luecker for their assistance, and for making my studies a more enjoyable experience.

During my studies, I was fortunate enough to visit the labs of Dr. Stuart Kornfeld at Washington University in St. Louis, and Dr. Bill Sly at St. Louis University. Dr. Kornfeld and his post-doctoral fellow, Dr. Yi Qian, instructed me on how to perform the phosphotransferase and uncovering enzyme activity assays. Dr. Kornfeld kindly provided aliquots of the phosphotransferase and uncovering enzymes that were used for the work covered in this thesis.

Finally, I would like to thank Dr. Sherryl Bisgrove and Dr. David Vocadlo, for serving on my thesis committee and providing their advice, as well as Dr. Margo Moore and Dr. David Green for forming the remainder of my examining committee.

Table of Contents

Approval ii
Partial Copyright Licence iii
Abstract iv
Dedication v
Acknowledgements vi
Table of Contents vii
List of Tables x
List of Figures xi

Chapter 1. Introduction 1
1.1. The Lysosomal Storage Diseases (LSDs) 2
1.1.1. Cellular basis of LSDs 3
1.1.2. Addition of the phosphate tag – UDP-*N*-acetylglucosamine:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase 4
1.1.3. Exposing the phosphate tag – *N*-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase: the "uncovering" enzyme 8
1.1.4. Structure, enzyme kinetics, and glycan configuration of β-glucuronidase 10
1.2. The treatment of LSDs 11
1.2.1. Enzyme replacement therapy for the treatment of LSDs 12
1.2.2. Overcoming the therapeutic obstacles associated with ERT 12
1.2.3. Other therapies for the treatment of LSDs 16
1.3. Plants as Bioreactors for Recombinant Protein Production 17
1.3.1. Humanizing *N*-glycans in plants 19
1.3.2. Removing unwanted plant *N*-glycan processing 20
1.3.3. Adding human *N*-Glycan processing enzymes to plants 22
1.3.4. Improving glycan status and yield of plant-produced proteins through seed expression and manipulation of protein targeting 23
Retaining proteins in the ER 23
Expressing proteins in seeds 24
1.4. The Present Study 25
1.5. References 26
1.6. Tables and Figures 36

Chapter 2. Characterization and downstream mannose phosphorylation of human recombinant alpha-L-iduronidase produced in *Arabidopsis* complex glycan-deficient (cgl) seeds 39
2.1. Abstract 39
2.2. Introduction 40
2.3. Results 43
2.3.1. Arabidopsis *cgl* seeds accumulate high levels of active α-L-iduronidase 43

2.3.2. *Cgl*-derived IDUA contains predominantly high-mannose-type *N*-glycans 44
2.3.3. *Cgl*-IDUA exhibits kinetic parameters that are comparable with the commercial enzyme 45
2.3.4. *Cgl*-IDUA is amenable to sequential *in vitro* processing to add the M6P recognition marker 45
2.4. Discussion 46
2.4.1. High levels of active α-L-iduronidase (IDUA) are produced in Arabidopsis *cgl* seeds 46
2.4.2. The maturation of the *N*-glycans of recombinant IDUA is avoided in Arabidopsis *cgl* seeds 48
2.4.3. *Cgl*-IDUA can be processed *in vitro* to generate the M6P recognition marker required for lysosomal targeting in human cells 50
2.5. Experimental Procedures 52
2.5.1. Protein extraction and α-L-iduronidase activity assays on transgenic *cgl* lines 52
2.5.2. Affinity purification of *cgl*-derived α-L-iduronidase 53
2.5.3. *N*-glycan profiles of *cgl*-derived IDUA 53
2.5.4. Anti-horseradish peroxidase affinity chromatography for removal of small amounts of α-L-iduronidase containing plant complex/hybrid *N*-glycans 53
2.5.5. Determination of kinetics of *cgl*-derived IDUA 54
2.5.6. Modification of *cgl*-IDUA by a recombinant soluble $\alpha_2\beta_2$ GlcNAc-1-phosphotransferase 54
2.5.7. Removal of GlcNAc by the uncovering enzyme 55
2.6. Acknowledgements 55
2.7. References 55
2.8. Tables and Figures 60

Chapter 3. Purification, characterization, and phosphorylation of α-L-iduronidase produced in the *Arabidopsis thaliana* Golgi Mannosidase I (GM1) mutant. 70
3.1. Abstract 70
3.2. Introduction 71
3.3. Results 74
3.3.1. Seeds of the *Arabidopsis* GM1 mutant express active α-L-iduronidase 74
3.3.2. GM1-Iduronidase contains a more optimal *N*-glycan profile as compared to *cgl*-iduronidase: higher mannose chain length and no detectable complex *N*-glycan sugars 74
3.3.3. The GM1-iduronidase can be processed *in vitro* by the enzymes of the M6P biosynthetic pathway 75
3.4. Discussion 76
3.4.1. The *Arabidopsis* GM1 mutant line produces α-L-iduronidase (IDUA) with a different *N*-glycan profile than that of *cgl*-IDUA. 76
3.4.2. The GM1 mutant yields recombinant IDUA that is amenable to *in vitro* phosphorylation and may have promise for generating an *in planta* phosphorylation system 78

3.5. Materials and Methods 79
3.5.1. Generation of Transgenic Lines 79
3.5.2. Purification of recombinant α-L-iduronidase 79
3.5.3. Endoglycosidase H digestion 80
3.5.4. Activity assays and determination of kinetic parameters of recombinant IDUA 80
3.5.5. GlcNAc-Phosphate Transfer Assay 81
3.5.6. *N*-glycan Profiling 81
3.6. References 81
3.7. Tables and Figures 84

Chapter 4. Conclusions and Future Directions 95
4.1. References 96

List of Tables

Table 2.1. Accumulation of IDUA in transgenic Arabidopsis seed lines 60

Table 2.2. N-glycans identified in *cgl*-IDUA 61

Table 2.3. N-glycans identified in *cgl*-IDUA after anti-HRP affinity chromatography 62

Table 3.1. Buffers used in this study. 84

Table 3.2. Summary of purification of α-L-iduronidase. 85

Table 3.3. Comparison of kinetic parameters of 4-MUI cleavage by α-L-iduronidase purified from *cgl* and GM1 *Arabidopsis* lines. 86

Table 3.4. A comparison of the kinetic parameters of GlcNAc phosphate transfer *in vitro* from UDP-GlcNAc to several forms of purified, recombinant iduronidase, when catalyzed by soluble phosphotransferase. 87

Table 3.5. The complete *N*-glycan profile of GM1-IDUA. 88

Table 3.6. Site-specific *N*-glycan profiles of GM1-IDUA. 89

List of Figures

Figure 1.1. Addition of mannose-6-phosphate to lysosomal enzymes 37

Figure 1.2. *N*-glycan maturation in Eukaryotic organisms. 38

Figure 2.1. α-L-Iduronidase (IDUA) produced in Arabidopsis *cgl* seeds..... 63

Figure 2.2. *N*-glycans of *cgl*-α-L-iduronidase (IDUA). 64

Figure 2.3. High-mannose type *N*-glycan at site Asn 372 on recombinant *cgl*-IDUA protein (glycopeptide 369–383). 65

Figure 2.4. *N*-glycans on *cgl*-IDUA after anti-HRP affinity chromatography identified by MALDI-TOF MS. 66

Figure 2.5. Michaelis-Menten plot for *cgl*-IDUA 67

Figure 2.6. Phosphorylation of *cgl*-IDUA by the GlcNAc-1-phosphotransferase ($\alpha_2\beta_2$). 68

Figure 2.7. Release of [^{3}H] GlcNAc from [^{3}H] GlcNAc-P ('phosphorylated')-*cgl*-IDUA by the soluble uncovering enzyme. 69

Figure 3.1. Analysis of the purification of IDUA from transgenic *Arabidopsis* GM1-knockout seeds by SDS-PAGE. 90

Figure 3.2. A comparison of the kinetics of 4-MUI cleavage catalyzed by *cgl*-IDUA and GM1-IDUA. 91

Figure 3.3. Endoglycosidase-H digestion of *cgl*-IDUA and GM1-IDUA 92

Figure 3.4. Kinetics of *in vitro* phosphorylation of GM1-IDUA by the soluble phosphotransferase. 93

Figure 3.5. PNGase-F digestion of GM1-IDUA. 94

Chapter 1. Introduction

Therapeutic proteins are vital for the treatment of many serious illnesses and genetic diseases. Recombinant protein production is a large and steadily growing part of the pharmaceutical industry that generates billions of dollars in revenue each year. The lysosomal storage diseases are a class of genetic disease amenable to treatment by recombinant protein therapy. Plants offer the promise of a host that is both free of human pathogens and capable of producing large amounts of protein at minimal cost. Recently, the United States Food and Drug Administration (FDA) approved Taliglucerase alfa, the first therapeutic lysosomal enzyme produced in a plant cell culture.

This thesis is organized in two sections providing a synopsis of the most current information on both the treatment of the lysosomal storage disease, and the use of plants as production platforms for the production of therapeutic proteins, with the focus on lysosomal enzymes. My thesis centers around the downstream (*in vitro)* processing of a plant-made human lysosomal enzyme, so that it is competent for human cell uptake and lysosomal delivery. Because the ultimate goal is the treatment of the disease, the section on human biology is placed first, containing an overview of the lysosomal storage diseases and detailed discussion of the faulty cellular and biochemical mechanisms, followed by a list of the available treatment options. The second section summarizes the current trends in recombinant protein production in plants, especially those studies devoted to *N*-glycan manipulation for high-quality therapeutics as well as expression of protein in the plant seed, highlighting those discoveries and technologies relevant to the production of lysosomal storage enzymes.

1.1. The Lysosomal Storage Diseases (LSDs)

The lysosomal storage diseases (LSDs) are a class of over 70 disorders that are caused by mutations in genes encoding enzymes that reside in the lysosome, or in proteins important to lysosomal function or lysosomal protein trafficking (Vellodi, 2005). The resulting metabolic aberration generally causes the lysosome to accumulate the substrate of the affected enzyme, leading to a range of disease symptoms and in the most severe forms, death in childhood. Lysosomal storage diseases are very rare in the general population; individual diseases are often seen in 1:100,000 births. Collectively, however, they occur in roughly 1:5000 births, and in some populations the frequency of an individual disease can be much higher. An overview of the diseases caused by deficiencies of three lysosomal enzymes that have been expressed in *Arabidopsis* is presented in Table 1.1. The LSDs can be categorized by the type of substrate that is accumulated in the cells and tissues of the body of an afflicted individual. The mucopolysaccharidoses (MPSs) result from storage of glycosaminoglycans (GAGs), nitrogenous polysaccharides (formerly known as mucopolysaccharides) that are typically found in the extracellular matrix and connective tissues. Deficiencies in metabolism leading to glycolipid or sphingolipid accumulation constitute another class of LSD, including Tay-Sachs disease (Kingdon and Russell, 1896) and Gaucher disease (Gaucher, 1882). Other LSDs result from errors in glycoprotein glycan metabolism, such as the deficiency in α-D-mannosidase, α-mannosidosis (Ockerman, 1969).

While there are many specific symptoms associated with individual LSDs, there are also several commonalities. The first trend is that the same disease can occur at a range of severities that is sometimes dependent on the amount of residual enzyme activity in the patient. Point mutations outside the active site can destabilize the enzyme without totally abrogating enzyme activity, leading to a less severe phenotype that presents itself later in life. On the other hand, inheritance of a mutation such as a frameshift that completely destroys a large part of the enzyme is often fatal in childhood. As a corollary, the type of mutation impacts the available treatment options for the patient (discussed in a later section). Secondly, the LSDs tend to share a common set of symptoms that include hepato- and splenomegaly (caused partly by the burden of storage material), joint stiffness and skeletal and facial deformity, and in some cases,

central nervous system (CNS) disease which results in progressive mental retardation. Even in severe cases, the symptoms are often not immediately evident at birth, but begin to manifest themselves early in childhood. The therapeutic outcome and natural history of LSDs can be vastly improved if treatment begins as soon as possible. Thus, there is the need for advanced screening procedures in newborns (Marsden and Levy, 2010), particularly for diseases where therapeutics are available.

1.1.1. Cellular basis of LSDs

Before any discourse on treatment options for the LSDs, one must gain a clear perspective of the cellular biology and biochemistry of the lysosomal enzymes. In the mammalian cell, the majority of lysosomal hydrolases reach the lysosome *via* the endomembrane protein transport pathway and mannose-6-phosphate receptor (M6PR) mediated sorting (Dahms et al., 2006). Proteins bearing *N*-linked high-mannose oligosaccharide chains that contain one or two mannose-6-phosphate (M6P) moieties are recognized by this receptor in the *trans*-Golgi network and are shuttled away in vesicles that eventually fuse with the lysosome. Elaboration of the M6P tag on the glycan occurs in two enzymatic steps in the Golgi complex (Fig 1.1) using phosphate donated from uridine diphosphate-*N*-acetylglucosamine (UDP-GlcNAc). The first enzyme, UDP-GlcNAc:lysosomal enzyme GlcNAc-phosphotransferase, adds GlcNAc-phosphate in phosphodiester linkage to the 6-OH of one or two mannose units on the glycan (Reitman and Kornfeld, 1981). The second enzyme, *N*-acetylglucosamine-1-phosphodiester α-*N*-acetylglucosaminidase hydrolyzes the bond between the anomeric carbon of the GlcNAc and the oxygen of the phosphate group to liberate the GlcNAc and expose the phosphate (Varki et al., 1983). There are two different forms of the M6PR named for their *in vitro* M6P binding properties. The 15-domain cation-independent mannose-6-phosphate receptor (CI-M6PR) contains multiple M6P and M6P-GlcNAc binding domains and is also present at the extracellular face of the plasma membrane, allowing lysosomal enzymes that have escaped into the extracellular medium to be internalized into endosomes and sent to the lysosome. The cation-dependent M6PR, on the other hand, only has a single M6P binding domain and exists as a dimer *in vivo*.

The formation of the lysosomal targeting signal is discussed at length because the final section of this thesis details a method for the replication of this metabolic pathway in plants for the humanization of plant produced lysosomal enzymes. The lysosomal enzyme β-glucuronidase is also reported on because of its significant role in the elucidation of the M6P-biosynthesis pathway; its enzymatic properties have been extensively characterized, and the mouse model of β-glucuronidase deficiency that mimics the human MPS VII disease is used routinely for LSD treatment research.

1.1.2. Addition of the phosphate tag – UDP-*N*-acetylglucosamine:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase

The first step of M6P-tag formation, the transfer of a GlcNAc phosphodiester from UDP-GlcNAc to the high mannose *N*-glycan, is catalyzed by UDP-*N*-acetylglucosamine:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase (henceforth referred to as "the phosphotransferase" within this thesis). Prior to the discovery of this enzyme, it was known that *N*-glycans containing M6P acted as a lysosomal targeting signal, and that a small portion of *N*-glycans on β-glucuronidase contained a terminal GlcNAc bonded to mannose by a phosphate diester. When various glycoproteins and high-mannose oligosaccharides were treated with UDP-GlcNAc and endosomal membrane preparations, only lysosomal proteins would receive GlcNAc-phosphate, indicating that the phosphotransferase not only mediated the first step of M6P addition on lysosomal proteins, but that it could sort out the lysosomal proteins from all of the other secretory proteins (Reitman and Kornfeld 1981; Waheed et al., 1982). The first homogenous preparation of the phosphotransferase came from lactating bovine udder, where a monoclonal antibody was used to obtain 488,000-fold purification (Bao et al., 1996).

The GlcNAc-phosphotransferase is a *cis*-Golgi resident 540-kDa heterohexamer composed of two copies each of 115-kDa α-, 51-kDa β-, and 56-kDa γ-subunits. The α- and β- subunits are each anchored in the *cis*-Golgi by a single transmembrane α-helix. Disulfide bonds exist between the two α-subunits as well as the two β-subunits, and this pair of disulfide-linked homodimers interacts *in vivo* to form the catalytic complex. Both of these subunits are encoded by a single mRNA, transcribed from the gene GNPTAB,

and the transcript product is cleaved between D929 and K930 by the Site-1-Protease (Marschner et al, 2011). The α-subunit sequence has regions homologous to the notch receptor, the transcriptional repressor DMAP binding domain, and the bacterial capsule biosynthesis protein, the latter region likely containing the phosphotransferase active site (Braulke et al., 2008).

The γ-subunit protein is encoded by a separate gene, GNPTG. It is not membrane bound and not strictly necessary for catalysis, but associates with the αβ-complex to increase phosphotransferase activity. It contains a mannose-6-phosphate receptor (MRH) homology domain, a protein domain whose function is to bind high-mannose-type *N*-glycans (Munro et al., 2001). The k_{cat} of the purified αβ-complex towards a typical lysosomal glycoprotein increases roughly three-fold in the presence of the γ-subunit, although K_m is not affected (Qian et al., 2010). A peptide signal or protein-protein interaction that retains the phosphotransferase in the Golgi has yet to be identified.

The functions of the three phosphotransferase subunits have been partially elucidated by the pathologies of mucolipidoses (ML) II, IIIA, and IIIC, the first two resulting from mutations in GNPTAB and the latter from mutations in GNPTG. Extracts of fibroblasts from patients of ML II or IIIA (both severe LSDs), as well as the corresponding mouse models (Lee et al., 2007), are devoid of, or are severely lacking, phosphotransferase activity. The milder phenotype of MLIIIC correlates to fibroblast extracts that have reduced activity towards lysosomal hydrolase glycans, but identical activity towards the synthetic acceptor α-methylmannoside. Those clinical studies have been corroborated by biochemical observations (Qian et al., 2010). A substrate photobinding assay has shown that the α-subunit contains the active site, although the β-subunit is also required for catalysis. While the γ-subunit of the phosphotransferase increases the catalytic efficiency of the phosphotransferase towards some lysosomal proteins (e.g., cathepsin D, uteroferrin, and NPC2), the phosphotransferase lacking the γ-subunit is active, and in the case of a minority of lysosomal proteins (e.g., β-glucuronidase) the absence of the γ-subunit does not appear to reduce phosphorylation (Qian et al., 2010). Thus, the γ-subunit appears to play some modulatory role that enhances the catalysis performed by the α- and β-subunits.

The exact mechanism by which the γ-subunit is involved in catalysis is very much a mystery due to the lack of a crystal structure for the phosphotransferase, although the γ-subunit appears to modulate catalytic activity by interacting with the *N*-glycans of the substrate protein. One hypothesis is that it acts in a manner similar to the β-subunit of glucosidase II (Qian et al., 2010). Glucosidase II is the heterodimeric ER-resident α-1,3 glucosidase that removes the outermost two glucose residues of the high-mannose *N*-linked glycan on the newly-synthesized protein, the first step of *N*-linked glycan trimming. It is known that the role of the glucosidase II β-subunit is to present the high mannose glycan to the catalytic α-subunit. (Watanabe et al., 2009). Like the glucosidase II β-subunit, the phosphotransferase γ-subunit contains a MRH domain. The presence of this domain implicates a role for the γ-subunit in binding the high-mannose glycan, and experimental evidence corroborates this implication. Without the glucosidase II β-subunit, the α-subunit is not active towards terminal glucose residues on high-mannose glycans, but it retains activity toward the synthetic substrate, p-nitrophenyl α-D-glucopyranoside (Watanabe et al., 2009). Likewise, in the absence of the phosphotransferase γ-subunit, phosphotransferase activity towards α-methylmannoside is unchanged, while activity towards glycans on lysosomal proteins is less efficient (Qian et al., 2010).

The phosphotransferase γ-subunit has at least two specific functions. It increases the ratio of bis- to mono-phosphorylated glycans, and it increases the overall level of phosphorylation for some lysosomal enzymes. The exact molecular mechanisms are unknown, however. The γ-subunit may enhance binding time between the substrate and the catalytic core resulting in a better chance for bisphosphorylation to occur (as opposed to monophosphorylation), it might aid catalysis by directing *N*-glycans towards the αβ-complex catalytic site, or it could modulate the tertiary structure of the αβ-complex itself. The latter has been postulated to be the mechanism by which the glucosidase II β-subunit operates (Stigliano et al., 2009).

It does not appear that there is a defined region of protrusions and ridges on the surface of the lysosomal enzyme that would serve as a recognition marker for the phosphotransferase αβ-complex; rather, specific lysine residues present on the lysosomal enzyme are critical for recognition by the phosphotransferase. The exact,

structural basis of these lysines has been extensively explored in the lysosomal hydrolases cathepsin D (Steet et al., 2005) and cathespin L (Warner et al., 2002), as well as DNase I (Nishikawa et al., 1999), a non-lysosomal hydrolase that acquires a small amount of M6P. Phosphotransferase recognition is determined almost exclusively by the spacing between two lysines and the target *N*-glycan. Structural data combined with lysine-to-alanine mutagenesis scanning the length of the substrate protein reveals that pairs of lysines, 34 Å apart, present in DNase I and both cathepsins are the most crucial signals (Warner et al., 2002). Chemical acetylation of all lysine and mutation of specific lysine residues on cathepsin D or L strongly reduces or eliminates M6P acquisition but has no effect on the activity nor on the thermal denaturation profile of the enzyme. These results indicate that the loss of phosphotransferase recognition caused by lysine alteration is just due to the signalling properties of lysine, and not a drastic change in tertiary structure that would destroy a large phosphotransferase recognition patch present on the substrate protein. Conversely, when novel *N*-glycan sites spanning the length of the target protein are introduced, those additional glycans that acquire the M6P are tightly clustered around a point 27 Å from the one lysine out of the signaling pair. Since this point represents the phosphotransferase active site, the active site must always be fixed relative to the position of the lysines, which logically implies that the lysine pair defines phosphotransferase orientation. The DNA methyltransferase-associated protein (DMAP) interaction domain was shown to be the region of the phosphotransferase that binds lysosomal acid hydrolases (Qian et al, 2013).

A beta-hairpin loop has been proposed as the structural determinant found on those lysosomal proteins that receive M6P which allows them to recognize the phosphotransferase. Such a loop has been identified in the crystal structures of cathepsin D (Baldwin et al., 1993) and β-glucuronidase (Hassan et al, 2013). When this lysosomal targeting loop of cathepsin D and a critical lysine residue (Lys 203) are inserted into pepsinogen, a secretory protein that is normally not phosphorylated and that bears 45% sequence identity to cathepsin D, M6P is detected on the *N*-glycans of the chimeric protein (Baranski et al., 1990). It was later shown that the amino acid residue (Lys267), present in the targeting loop of cathepsin, mediates an increase in phosphorylation of the chimeric pepsinogen,although the other residues in the targeting loop aid in increasing the amount of phosphorylation as well (Steet et al., 2005). Beta-

glucuronidase has a lysine in its lysosomal targeting loop. Therefore, it appears that the combination of a critical lysine residue and a specific hairpin loop are the key factors that mediate recognition of lysosomal enzymes that receive M6P by the phosphotransferase. The evolution of the M6P targeting system also suggests that a simple recognition marker is likely to be present (Cuozzo et al., 1998). The emergence of the M6P targeting system early in the vertebrate lineage would require a corresponding phosphorylation signal. The evolution of large surface patch would require a massive, concerted change in the structure of so many lysosomal hydrolases, whereas the point mutation of a pair of lysines would be much more likely to occur. For example, when an *N*-glycosylation site was added to Green Fluorescent Protein (GFP), a protein with 9% lysine content, the *N*-glycans of this mutant GFP were phosphorylated at 18% of cathepsin L levels (Warner et al., 2002). Such a hasty modification to a non-lysosomal protein from an organism with no M6P-targeting pathway should not result in any phosphorylation if a complex surface patch is the primary requirement for phosphotransferase recognition.

1.1.3. Exposing the phosphate tag – *N*-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase: the "uncovering" enzyme

The enzyme responsible for exposing M6P through removal of the blocking GlcNAc residue added by the phosphotransferase is *N*-acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase, or the "uncovering enzyme" (UCE; gene symbol NAGPA). It was first isolated in microsomal preparations of the rat liver (Varki & Kornfeld, 1981) after the observation that intermediately processed M6P glycans contained a terminal α-GlcNAc linked to the mannosylphosphate (Tabas and Kornfeld, 1980). Although initially believed to be a phosphodiesterase, it was shown that the UCE cleaves the C-O bond and not the O-P bond, and therefore it is now correctly classified as a glycosidase (Varki et al., 1983). Lysosomes also contain an *N*-acetylglucosaminidase required for GAG catabolism, but the two enzymes can be differentiated by their substrate requirements (the UCE requires a phosphodiester substrate, while the lysosomal enzyme does not). Improvement in recombinant protein

purification as well as UCE detection assays has led to the development of a quick immunopurification scheme (Mullis et al., 1994).

The UCE is a *trans*-Golgi network (TGN) resident type I membrane protein 515 amino acids in length, bearing a lumenal catalytic domain, a single transmembrane helix, and a 27 amino acid cytoplasmic tail (Do et al., 2002). It is synthesized as a proenzyme that requires activation by the *trans*-Golgi resident endoprotease, furin. The mature protein of 65 kDa bears sialylated, endoglycosidase-H (endo-H) resistant *N*-glycans and exists as a dimer of two disulfide-linked subunits. There is little structural data available on the UCE, other than the location of multiple disulfide bonds, and that there is evidence for a soluble form of the enzyme that lacks its membrane anchor and which exists in trace amounts in the bloodstream (Lee and Pierce, 1995). Enzyme kinetics, substrate specificity, optimum pH, as well as a variety of inorganic and organic inhibitors, have all been characterized (Mullis and Kornfeld, 1994). The enzyme is capable of cleaving GlcNAc moieties present on all high mannose glycans isolated from mammalian cells containing six to nine mannose residues, but it is the most specific for those containing α-(1→2)-linked terminal mannose bearing the GlcNAc. This may be significant in the cellular context, as the M6PR exhibits the lowest K_D towards M6P on terminal α-(1→2)-linked mannose residues. The UCE also demonstrates a catalytic preference for GlcNAc moieties on the α-(1→3)-linked arm of the *N*-glycan, followed by those on the middle arm. The biological ramifications of this differential arm specificity are not entirely clear, as most phosphorylated high-mannose glycans found in the cell can bind the M6PR, regardless upon which arm the phosphate lies. Furthermore, the kinetic parameters are not affected whether or not the glycan is attached to protein, peptide, or has been liberated using EndoH, so this arm specificity may just be the result of steric hindrance caused by the glycan itself.

Many lines of evidence, including fluorescence studies, *N*-glycan status, brefeldin A treatment, and density centrifugation experiments have localized the UCE to the TGN, where it constantly moves to and from the plasma membrane in a clathrin-dependent manner (Lee et al., 2002). A tetrapeptide motif present at the cytoplasmic tail (YHPL) agrees with the consensus motif YXXφ (φ is a bulky, hydrophobic residue) that is known to bind adaptor complex AP-2, a receptor that recruits cargo into clathrin coated

vesicles. A puzzling result is obtained after the truncation of the cytoplasmic tail immediately upstream of this internalization motif: only 15% of UCE activity is detected at the cell surface (as opposed to 100%), implying a TGN exit signal is also removed by the truncation. This led to the discovery of two distinct internalization signals; the strongest is immediately downstream of the YPHL AP-2 binding signal, and a weaker, pentapeptide motif is located at the extreme C-terminus of the enzyme (Nair et al., 2005).

1.1.4. Structure, enzyme kinetics, and glycan configuration of β-glucuronidase

The enzyme β-D-glucuronohydrolase (β-glucuronidase; gene symbol GUSB) is a lysosomal acid hydrolase that has been identified in human, mouse, rat, dog, cat, and is likely present in most vertebrate species. The enzyme liberates terminal β-D-glucuronic acid moieties from GAGs in the lysosome, and its absence in humans results in Sly syndrome (MPS VII). The enzyme itself is translated as a 651 amino acid, 75 kDa preproprotein that is first cleaved of its signal peptide after translocation to the ER, and then of an 18-amino acid C-terminal propeptide (Islam et al, 1993). Tetramerization is required for catalytic activity of the mature, glycosylated protein. Human β-glucuronidase has homology to *Escherichia coli* β-D-galactosidase, a well-characterized retaining glycosidase. This sequence and mechanistic similarity, as well as the β-glucuronidase crystal structure, helped to identify Glu540, Glu451, and Tyr504 as the likely β-glucuronidase active site residues (Islam et al., 1999). Trapping of the intermediate using a fluoro-sugar analog confirmed that Glu540 is the catalytic nucleophile donor responsible for cleavage of the glycosidic bond (Wong et al., 1998). Experiments using site-directed mutagenesis (Islam et al., 1999), along with the crystal structure of β-glucuronidase (Jain et al., 1996), show that Glu451 serves to deprotonate the incoming water molecule that hydrolyzes the enzyme-substrate intermediate. The role of Tyr504 is important, but unknown, as catalytic efficiency decreases 830-fold in its absence.

Human β-glucuronidase has *N*-linked glycosylation sites that harbor high-mannose *N*-glycans with the M6P recognition marker. In addition to lysosomal targeting,

N-glycosylation is required for proper folding (Shipley et al., 1993). Site directed mutagenesis of the *N*-glycan recognition site, or the addition of tunicamycin in the growth medium results in a loss of β-glucuronidase activity that is proportional to the number of nonfunctional *N*-glycosylation sites. The loss in activity is likely due to a failure of non-glycosylated β-glucuronidase to form tetramers. In contrast, removal of all the oligosaccharides via *N*-glycosidase F digestion from active, wild-type β-glucuronidase preparations does not result in a loss of enzymatic activity. The crystal structure reveals that the glycosylation sites are all solvent accessible and do not mediate intrasubunit binding; therefore, these are required during β-glucuronidase folding and tetramer assembly but are not involved in catalysis.

The *N*-glycans of β-glucuronidase at Asn272, Asn420, and Asn631 are phosphorylated (containing 72%, 21%, and 7% of the total M6P label, respectively), while that at Asn173 is not phosphorylated at all. This uneven distribution is due to the varying location of the glycans relative to the position of the phosphotransferase recognition signal. Phosphorylation levels are also dependent on the presence of the C-terminal propeptide. A mutant lacking the propeptide had a lower amount of phosphorylated glycans, and an increased proportion of those phosphorylated glycans retained the blocking GlcNAc residue (Islam et al, 1993). In addition, this mutant achieved only half-maximal catalytic rate. Specific aspects of M6P-elaboration on another lysosomal enzyme, α-L-iduronidase, are noted in chapters 2 and 3.

1.2. The treatment of LSDs

The only known cure for LSDs is gene therapy which is the permanent transformation of the patient's genome with a functional copy of the defective gene (Beck, 2010). Animal studies using retroviral vectors are promising, but are considered unsafe for humans at this time. Available treatments are those that increase the pool of available enzyme (enzyme replacement therapy, enzyme enhancement therapy, hematopoeitic stem cell transplantation), and those that reduce the amount of stored substrate (substrate reduction therapy) (Beck 2010). Below, I provide additional details on enzyme replacement therapy.

1.2.1. Enzyme replacement therapy for the treatment of LSDs

Enzyme replacement therapy (ERT) is the compensation of an enzyme deficiency using intravenously supplied purified enzyme (Grabowski and Hopkin, 2003). The therapy takes advantage of the naturally occurring receptor-mediated endocytosis of lysosomal enzymes via the M6PR. The first LSD to be treated by ERT was type I Gaucher disease using human placental glucocerebrosidase (Barton et al., 1991). This enzyme is taken up by macrophages and other cells of monocytic origin that possess a mannose receptor. Other lysosomal enzymes that possess a M6P tag are recognized and endocytosed by the M6PR. For example, recombinant α-galactosidase A is internalized by the M6PR present on fibroblasts (Marschesan et al., 2012) and kidney endothelial cells (Prabakaran et al., 2012). Enzyme replacement therapy has improved significantly over the past two decades, and the treatment is now available for seven of the LSDs. Glucocerebrosidase is the first lysosomal storage enzyme to be produced in a plant host and is marketed by Pfizer (reviewed in Grabowski et al., 2014).

1.2.2. Overcoming the therapeutic obstacles associated with ERT

A common side effect of therapy involving intravenously supplied recombinant proteins is an adverse immune reaction by the patient, especially if the afflicted individual has large deletion mutations such that the therapeutic protein would present a foreign epitope. The immune system response can be violent and immediately threatening to the patient's life although this effect can be mitigated by the use of antihistamines and corticosteroids. In addition, the subsequent production of antibodies can mediate premature clearance of the enzyme from the serum, a complication observed in the treatment of Pompe disease and Fabry disease by ERT. In many other cases, however, an initially high antibody titre eventually subsides as the patient develops tolerance to the drug over the course of the treatment.

A more formidable limitation of ERT is the obstruction of M6PR-mediated uptake of most intravenously supplied lysosomal enzymes in the central nervous system (CNS) by the blood-brain barrier (BBB). This makes the therapy useless for treating the neuropathies caused by the most severe forms of LSDs, as witnessed during the treatment of severe MPS I disease and types II and III Gaucher disease. Hematopoietic

stem cell transplantation (HSCT; reviewed in Lund, 2013) is often effective in treating the neurological facets of the LSD (because monocytes can migrate through the BBB into the brain), but the transplant is a more difficult procedure than ERT and is further complicated by the possibility of donor-host incompatibility (Boelens et al., 2007). To date, HSCT has been somewhat successful for the management of severe MPS I (Boelens et al., 2013), and more recently, mucolipidosis type II (Lund et al., 2014).

To address the obstacle of delivering enzyme to the CNS, researchers have attempted modified ERT regimes in animal hosts. Up to two weeks of age, the BBB in mice is sufficiently permeable that intravenously supplied enzyme can be taken up by the M6PR in the brain tissue. After this developmental stage, however, the amount of enzyme that reaches the brain after this time period is drastically lower (Urayama et al., 2004). This cellular phenomenon is reflected in studies of mouse models of LSDs where disease symptoms are delayed if the ERT is begun immediately after birth (Gliddon and Hopwood, 2004; Urayama et al., 2008). Permeabilization of the BBB by adrenergic agonists such as epinephrine can increase the amount of enzyme that reaches the neurons (Urayama et al., 2007; Borges et al., 1994), but this approach is not likely to be suitable for regular, lifelong treatment. On the other hand, maintaining an exceptionally high serum concentration of therapeutic enzyme in LSD mice results in the reduction of lysosomal storage products from the CNS tissues, in addition to the reduction in storage normally observed in the visceral organs when normal concentrations of drug are given. This entry to the neuronal tissues might occur by fluid-phase uptake by the brain endothelial cells themselves, or by transit through extracellular fluid in areas of the brain where the BBB is incomplete. The mannose receptor present on macrophages and other cells of mononuclear origin is mainly responsible for the removal of active lysosomal enzyme supplied by ERT. Mice that are null mutants for, or have been given inhibitors of, the mannose receptor retain intravenously supplied enzyme for periods longer than those with a functional mannose receptor (Sly et al., 2006). A treatment regime that minimizes exposure of the intravenously supplied enzyme to the mannose receptor would therefore raise the serum concentration of the enzyme allowing entry to the brain tissues.

Several approaches along this line using the MPS VII mouse model have been successful. The first approach was the administration of a higher-than-normal concentration of β-glucuronidase over the lifespan of the mouse, the rationale being that saturation of the mannose receptor would allow the persistence of an elevated enzyme concentration (Vogler et al., 2004). Both the increased concentration of enzyme and frequency of administration were associated with a lower GAG content in the parietal neocortical neurons, the meninges and brain perivascular cells, and the glia. A similar study using the α-mannosidosis mouse model showed that not only was a higher concentration of intravenously supplied α-mannosidase successful in reducing lysosomal storage in the brain, but that the presence of the M6P tag was largely unimportant (Blanz et al, 2008). A second strategy was the oxidative cleavage of high-mannose oligosaccharides present on β-glucuronidase by treatment with sodium metaperiodate, so that the glycan chains could no longer be recognized by either the mannose or mannose-6-phosphate receptors (Grubb et al., 2008). The half-life of this chemically modified β-glucuronidase in the bloodstream was orders of magnitude above the unmodified enzyme. The modified enzyme was far superior in reducing lysosomal storage in the brain, but not as effective in doing so for the spleen and liver, tissues rich in cells containing the mannose receptor. The authors speculate that either fluid-phase pinocytosis or receptor-mediated endocytosis *via* an uncharacterized receptor on the membrane of the blood-brain barrier was responsible for uptake of the modified β-glucuronidase. This outcome suggests that a treatment capable of reaching all tissues should contain an enzyme population with a range of glycan configurations.

In addition to *N*-glycan modification, the fusion of a short, acidic peptide to the therapeutic enzyme is capable of increasing delivery to the brain. The AAA-peptide tag containing either six or eight aspartate residues prepended to β-glucuronidase (Montaño et al., 2008), and the E6-peptide tag containing a glutamate hexamer prepended to *N*-acetylgalactosamine-6-sulfate sulfatase (E6-GALNS) (Tomatsu et al., 2010), were responsible for increased enzyme serum half-life and better substrate reduction in the brain tissue in the mouse models of MPS VII and MPS IVA, respectively. The tagged enzymes were also directed in increased amounts to the bone, and in the case of E6-GALNS, the heart. Both of these tissues are more difficult to target by ERT than visceral organs like the spleen and liver. In this light, the acidic peptide may serve a dual

purpose. The long half-life of the enzyme in the serum may allow the heart and bone to be targeted in a manner similar to that described above for the brain. The negative charge of the peptide might also be drawn to a positively-charged component of the target cell, such as unidentified transmembrane receptors, or the hydroxyapatite present in membrane-bound vesicles of osteoblasts.

One intriguing, and very modular approach to crossing the BBB is the attachment of an RNA or DNA aptamer specific for a brain tissue marker, to the lysosomal enzyme. An aptamer is an oligonucleotide strand that adopts a three-dimensional configuration that strongly binds a target molecule, in this case a transmembrane receptor. The aptamer sequence is selected by an iterative process known as SELEX (systematic evolution of ligands by exponential enrichment) in which a pool of oligonucleotides is enriched to bind the target molecule. In one study, RNA and DNA aptamers specific for the extracellular domain of mouse transferrin receptor were attached to α-L-iduronidase (Chen et al., 2008). Transferrin is endocytosed into the brain endothelial cells by its receptor, and this aptamer was able to bind the receptor and deliver the aptamer-iduronidase conjugate to the lysosomes. Although the test was not performed on an animal model of MPS I, it demonstrates yet another glycan-independent method for targeting lysosomal enzymes to the brain. One drawback of this method, as well as the method involving the acidic peptide tag, is the potential for an immunogenic reaction caused by the presentation a foreign epitope to the patient's immune system.

Another major hurdle in providing ERT is the astronomical cost of recombinant protein production. This is partly due to the expense of producing large amounts of the recombinant protein in mammalian cell cultures and in some cases, because of the need for *in vitro* protein manipulation ("downstream processing"). The financial burden is further compounded by the frequency of hospital visits. A study from 2006 determined that the mean expense for patients in England and Wales receiving ERT for type I Gaucher disease is £86,000 (roughly 200,000 CAD) per year (Connock et al., 2006). Depending on the type and severity of LSD and the age of the patient, costs can be over double this amount for an individual. A Canadian MPS VI patient is receiving ERT at an annual cost of 1,000,000 CAD.

Toward addressing the large cost of ERT therapeutics, researchers are seeking other production platforms including plant-based systems (see Section 1.3 below). Protalix Biotherapeutics has developed the first lysosomal enzyme for ERT using plant cells as the production host. Uplyso (taliglucerase alfa) is a modified version of human glucocerebrosidase produced in transgenic carrot cell-suspension cultures (Avezier et al., 2009). The human signal peptide is replaced by that of *Arabidopsis thaliana* basic endochitinase. Appended to the C-terminus is the 7-amino-acid tobacco chitinase A protein storage vacuole targeting sequence (DLLVDTM) which may serve to reduce exposure to proteases and increase the uncovering of terminal mannose residues. Due to mannose-receptor-mediated uptake of glucocerebrosidase in Gaucher's disease ERT, the carrot-produced enzyme does not need costly post-purification *in vitro* N-glycan manipulation in order to expose the mannose core of the *N*-glycans, as 90% of the glycans are mannose-terminated. In contrast, CHO-cell-produced Cerezyme requires the sequential catalysis by sialidase, β-galactosidase and β-N-acetylglucosaminidase to expose the high-mannose core. The activity of Uplyso is nearly identical to that of Cerezyme. Clinical trials have shown that the enzyme is well-tolerated by patients and effective at treating Gaucher disease. The product has recently completed phase III clinical studies and has been approved for use by the United States FDA.

1.2.3. Other therapies for the treatment of LSDs

The use of small molecules that can reach CNS tissues is a necessary supplement to many severe forms of LSDs. Substrate reduction therapy (SRT) is the administration of small molecules that inhibit the enzymes responsible for production of the substrate that accumulates in the lysosomes of afflicted individuals. For example, Misglustat, an imino sugar used for SRT in the treatment of Gaucher disease, inhibits glucosylceramide synthase, thereby lowering the total pool of glucocerebroside.

On the other hand, if the LSD results from a point mutation that causes lysosomal enzyme misfolding, pharmacological chaperones can assist in rescuing residual endogenous enzyme production (Parenti, 2009). This treatment is known as enzyme enhancement therapy. To date, most of these chaperones are small molecules that mimic the lysosomal enzyme substrates and bind the enzyme active site, promoting

a favourable enzyme conformation and therefore giving the enzyme a chance to fold properly. This treatment is, therefore, not applicable to those patients with large deletions or mutations in the active site. Iminosugars, such as those based on 1-deoxynojirimycin have been successful in reducing lysosomal storage in cell culture models of Fabry disease (Benjamin et al., 2009), and in clinical trials involving patients with Fabry disease (Giugliani et al., 2013).Using plants in the treatment of LSDs

1.3. Plants as Bioreactors for Recombinant Protein Production

The number of treatments requiring recombinant human proteins is steadily rising each year. To meet this growing demand, the pharmaceutical industry is challenged to develop more robust methods of production that can yield high volumes of complex therapeutic proteins (Faye and Gomord, 2010). Existing platforms have their merits, but many suffer from drawbacks that impair the expression of therapeutic proteins. For example, bacteria can be easily genetically altered and grown very rapidly, but their inability to perform common eukaryotic post-translational modifications to proteins makes them suitable for only the simplest human peptides, such as insulin and human growth hormone. Yeasts offer similar advantages as bacteria in a eukaryotic host, but they are still sufficiently evolutionarily distant from humans that the differences in post-translational processing, notably glycan hypermannosylation, are problematic. It is noteworthy however that recent advances in genetic engineering of the yeast *Pichia pastoris* have allowed for the production correctly glycosylated monoclonal antibodies (Zha, 2013) On the other end of the spectrum, mammalian cell cultures can produce proteins with *N*-glycan profiles suitable for therapeutic administration, but difficulties in large-scale production, potential contamination by viruses or prions, and the overall cost of production make this platform unrealistically expensive for many proteins.

Plant hosts offer the advantage of inexpensive, large-scale production while maintaining the ability to be engineered to perform correct, human-like protein modifications (Sourrouille et al., 2009). They are economically advantageous because large amounts of biomass can be accumulated with a minimum of land, labour, and specialized equipment, and production can be easily scaled up or down depending on

demand. Plants and plant cell cultures are not susceptible to harmful contamination by mammalian viruses or prions, making them safer than mammalian cell cultures in this regard. There are numerous containment strategies to prevent the spread of transgenes into the wild, making plants an environmentally sound host as well.

The physiological traits of plants make them ideal hosts for therapeutic protein production. At the cellular level, plants perform many of the same post-translational modifications as mammals, including *N*-glycosylation, disulfide bond formation, proteolytic cleavage, and oligomer assembly, as exemplified in the assembly of plant-produced antibodies. Many plants can tolerate extensive genetic manipulation, such as the modification of glycosylation pathways (Castilho et al., 2013), or the expression of a substantial amount of foreign protein (De Jaeger et al., 2002). At the physiological level, the diversity of species within the plant kingdom has led to the discovery of numerous species for recombinant protein production, each offering relative advantages and disadvantages. For example, grains can store large amounts of protein in the protective environment of the seed, simplifying protein transport and long-term storage. On the other hand, species with rapid doubling times, such as duckweeds and algae, may be more suitable hosts for proteins that need to be produced more quickly, such as antigens for seasonal vaccines.

Like any industrial process, the two primary goals of expressing human proteins in plants are to maximize product quality and yield. The quality is gauged by the correct post-translational modification of the polypeptide chain(s) to a bioactive form that is tolerated by humans. Of these post-translational modifications, *N*-linked glycosylation has received the most attention by researchers. Protein yield is often measured as the percentage of total soluble protein (%TSP). A review of strategies to humanize *N*-linked glycans through genetic engineering of the plant biosynthetic pathway, and to improve yield by targeted expression at the cellular and subcellular level is presented here. Subsequently, the challenges related to expressing the human phosphotransferase in *Arabidopsis* are discussed.

1.3.1. Humanizing *N*-glycans in plants

Among the many post-translational protein modifications performed by eukaryotes, protein glycosylation is one of the most complex, involving dozens of enzymatic steps and demonstrating variation between the phylogenetic kingdoms (Gomord et al., 2010, Fig 1.2). Carbohydrate attachment to proteins affects a number of chemical and physiological processes and properties, including protein folding, protein thermodynamic stability, cellular recognition, protein-receptor binding, subcellular protein targeting, and catalytic function (Skropeta, 2009). Protein glycosylation is an extremely common modification among human proteins used for therapeutic purposes; therefore, engineering plant lines to produce mammalian glycoforms is an important focus in plant biotechnology research. Many therapeutic proteins harbour *N*-glycans, including lysosomal enzymes.

The secretory pathway is the site of *N*-linked glycosylation, the initial steps of which are shared between plants and animals. A precursor glycan, constructed by the stepwise addition of mannose and GlcNAc units to a dolichol phosphate moiety, is transferred in the ER lumen to the asparagine of the acceptor protein at sites containing Asn-X-Ser/Thr. In the ER lumen, the newly added *N*-glycan is required for proper protein folding. The terminal glucose and mannose residues interact with calnexin, calreticulin, and ER glycosyltransferases and glucosidases in a timing circuit that exports the nascent polypeptide to the endoplasmic reticulum associated degradation (ERAD) pathway in the case of aberrant protein folding (Yoshida and Tanaka, 2010). In the ER lumen and *cis*-Golgi, the sequence of glucose and mannose trimming reactions, as well as the addition of two GlcNAc moieties, is conserved in both plants and animals. In the medial- and *trans*-Golgi, however, plants add sugars in patterns that are vastly different from those found in the *N*-glycans of mammalian proteins. Aside from topical application, these plant complex glycans render the recombinant protein unfit for human therapeutic use. Plant glycans, in particular those bearing β-(1→2)-xylose and α-(1→3)-fucose linked to the core of the *N*-glycan (Bencurova et al., 2004), can cause immune reactions in most patients that are, at worst, life-threatening, and at best, prematurely remove the therapeutic from the bloodstream. The lack of mammalian sugar residues may also impact therapeutic efficacy. Neither sialic acid (whose existence in plants is

debatable) nor β1,4-galactose, a substrate requirement for the human-like sialic acid, are found in plant *N*-glycans. Sialic acid is necessary for increasing glycoprotein half-life in the circulatory system (e.g., for monoclonal antibodies). Those *N*-glycans bearing terminal galactose or mannose residues are rapidly cleared from the bloodstream via the hepatic asialoglycoprotein receptor or macrophage mannose receptor, respectively. On the other hand, rapid clearance of the therapeutic could be of benefit for reducing potential immune responses.

The established techniques of generating humanized *N*-glycans are described below, with a focus on those important for the phosphorylation of lysosomal enzymes in plants. Engineering a plant line that generates phosphorylated high-mannose *N*-glycans would be of immense value for the production of recombinant, human lysosomal enzymes. Such an endeavour would require the use of a mutant background that minimizes *N*-glycan maturation in combination with the addition of the phosphotransferase and the uncovering enzyme, which are the enzymes responsible for the addition of M6P in humans.

1.3.2. Removing unwanted plant *N*-glycan processing

The use of mutant plant lines deficient in one or more *N*-glycan processing enzymes constitutes a strategy to mitigate complex glycan formation on recombinant proteins. Unlike mammals, many of the glycan processing enzymes in the plant Golgi are completely dispensible for the plant's development under normal growth conditions. The simplest application of this strategy is to use a mutant without the undesirable enzymatic activity. *Arabidopsis* mutants devoid of α-(1→3)-fucosyltranferase and β-(1→2)-fucosyltransferase activity (Strasser et al., 2004), or β-(1→3)-galactosyltransferase activity (Strasser et al., 2007) produce *N*-glycans lacking the respective immunogenic sugars. Given the multitude of enzymes that produce undesirable moieties on *N*-glycans, however, this approach might be unreasonable for extensive modulation of the plant glycan pathway, such as generating only high-mannose glycoforms. Since the initial steps of the mammalian and plant *N*-glycan trimming pathways are identical, a more sophisticated strategy is to choose a mutant interrupted in the region of the pathway conserved between the two classes of

organisms. The *Arabidopsis cgl* mutant (deficient in complex glycan formation) lacks *N*-acetylglucosaminetransferase-I (GNT-I), which adds the first GlcNAc moiety on the high-mannose glycan (von Schaewen et al., 1993). In the absence of this GlcNAc, glycosyltransferases that mediate complex glycan formation downstream of GNTI are not as effective, and this results in a higher proportion of high-mannose glycoforms. This mutant was utilized to produce active human α-L-iduronidase containing primarily high mannose *N*-glycans (~50% $Man_5GlcNac_2$, ~25% $Man_{6-9}GlcNAc_2$) with less than 5% of *N*-glycans in the complex form (Downing et al., 2006; He et al. 2012b, 2013; Chapter 2).

With respect to elaboration of the M6P on lysosomal enzymes, only $Man_{6-8}GlcNAc_2$ *N*-glycans serve as a substrate for the lysosomal enzyme phosphotransferase; $Man_5GlcNAc_2$ *N*-glycans are not recognized by this enzyme. Although the *Arabidopsis cgl* mutant produces a significant portion of $Man_{6-8}GlcNAc_2$ glycans, it is still possible to interrupt the *N*-glycan trimming pathway at an earlier stage. Golgi α-mannosidase I (GMI), which is responsible for trimming of high-mannose glycans to $Man_5GlcNAc_2$, is dispensable under normal growth conditions. An *Arabidopsis* double mutant lacking two redundant, paralogous GMI genes exhibits *N*-glycan structures comprising mostly (>80%) of $Man_{7-8}GlcNAc_2$,no detectable complex *N*-glycan formation, with the remainder of the *N*-glycans bearing five or nine mannose rings (Kajiura et al., 2010). The mutant was phenotypically normal under standard growth conditions. The mannosidases that act immediately upstream of GM1 are the ER-localized class I α-mannosidases which are strictly required for protein quality control. *Arabidopsis* mutants lacking class I α-mannoisdases display severely reduced root growth and would likely not be good hosts for protein production (Liebminger et al., 2009). Hence, GM1 is the earliest point at which the *N*-glycan processing pathway can be interrupted, and the higher mannose glycoforms make the GM1 mutant a more attractive host than the *cgl* mutant.

Besides screening for glycosylation mutants, plant species can be genetically engineered to reduce complex glycan formation. The use of RNA-interference (RNAi) against xylosyl- and fucosyltransferases has been successful in reducing complex glycan formation on antibodies (Strasser et al., 2008), although the degree of silencing varies among species. Biolex Therapeutics has used this approach in the duckweed

Lemna minor, where RNAi is very efficient, to completely eliminate immunogenic sugars for the production of an monoclonal antibody used in the treatment of Hodgkin's lymphoma, which was evaluated in phase I trials (Cox et al., 2006). The moss *Physcomitrella patens* is also attractive because the high frequency of naturally occuring homologous recombination can be exploited for the knockout of glycosyltransferases. The advantage of using recombination over RNAi is not only the guaranteed, complete elimination of unwanted glycosyltransferase activity, but the optional, simultaneous knock-in of a human glycosyltransferase in the middle of the disrupted gene (Huether et al., 2005).

1.3.3. Adding human *N*-Glycan processing enzymes to plants

In addition to removal of glycosyltransferases that add immunogenic sugar moieties, there is a need for the augmentation of the plant biosynthetic pathway to further humanize recombinant glycoproteins. There are several key challenges to reconfiguring the plant *N*-glycan processing pathway in this manner. First, the glycosyltransferase must be inserted appropriately within the “assembly line” of Golgi-localized *N*-glycan processing enzymes. In some cases, targeting signals are conserved across kingdoms and wild-type mammalian enzymes are targetted to the proper location in plants (Wee et al., 1998). Nevertheless, much effort has been expended towards elucidating the localization signals present on these predominantly Type II single pass membrane proteins. Except for glucosidases that act in the early ER, intra-Golgi localization information is almost always present in the transmembrane domain and cytoplasmic tail of the glycosidase or glycosyltransferase, and not in the catalytic domain; likewise, the catalytic domain does not require a specific membrane anchor. Harnessing this modular design afforded by nature, as well as a growing collection of intra-Golgi targeting motifs, researchers have a rich palette for designing novel glycosyltransferases by combining human catalytic and plant targeting domains (Saint-Jore Dupas et al., 2006). This approach has been used for the production of monoclonal antibodies with terminal β1,4-galactose residues (Frey et al., 2009).

A second obstacle is the availability of nucleotide sugar donors in the Golgi. This is of particular importance to sialic acid addition, for example, because plants do not

produce cytidine monophosphate-*N*-acetylneuraminic acid (CMP-Neu5Ac), the sialic acid sugar donor. To generate a plant line expressing sialic acid, Castilho et al. (2008) inserted genes encoding three enzymes of the mammalian biosynthetic pathway that convert UDP-GlcNAc (an available metabolite in the plant Golgi apparatus) to CMP-Neu5Ac under control of the cauliflower mosaic virus 35S promoter and expressed them in *Arabidopsis.* This study was succesful in generating CMP-NeuNAc, demonstrating that significant manipulation of the plant biosynthetic pathway is possible. It also highlights several challenges that researchers face when modifying plant biosynthetic pathways. The coordinated regulation of several genes at once requires attention to promoter choice, as post-transcriptional gene silencing could have resulted in lower CMP-NeuNAc levels.

1.3.4. Improving glycan status and yield of plant-produced proteins through seed expression and manipulation of protein targeting

Proteins that require *N*-glycosylation must be targeted to the secretory pathway; however, within the secretory pathway, additional sorting signals that are added to the protein of interest can have a profound impact on glycoprotein quality and yield. If only a signal peptide is present, the protein will be directed to the apoplast and exposed to glycan processing enzymes in the plant Golgi apparatus. On the other hand, the protein can be retained in the ER by the addition of the C-terminal ER retention signal. Other subcellular locations, such as the cytosol and chloroplasts may improve yields for proteins lacking *N*-glycans. Likewise, targeted expression to certain plant organs through the use of tissue-specific promoters is another strategy to maximize yield.

Retaining proteins in the ER

One approach to circumvent the formation of complex glycans on a recombinant glycoprotein is the addition of the ER retention signal. The C-terminal tetrapeptide motif, KDEL, found on resident ER proteins, binds the ER-localized KDEL receptor so that recombinant proteins are retained in the ER and are not exposed to the xylosyl- and fucosyltransferases present in the later Golgi stacks, resulting in a high-mannose glycan. The technique has been successfully applied to a number of recombinant plant hosts

without the loss of protein activity or a negative impact on protein yield. The glycans of human α-L-iduronidase expressed in *Brassica napus* and *Nicotiana tabaccum* seeds contained roughly 75% less xylose and 50% less fucose when a C-terminal SEKDEL peptide was added, demonstrating some utility of this technique in the production of lysosomal storage enzymes (Galpin et al., 2010), though a more recent study shows this strategy to be relatively ineffective (He et al, 2012a). One other potential problem with the addition of the ER retention signal is the creation of an immunogenic epitope (Petrucelli et al., 2006).

A second issue is that ER retention is not perfect, as KDEL-tagged proteins may still escape the ER and acquire complex glycans in the Golgi apparatus (Fujiyama et al., 2009), possibly due to saturation of the KDEL receptor, or by limited exposure or proteolytic cleavage of the KDEL peptide (Abranches et al., 2008). A perplexing, yet frequently observed outcome of KDEL peptide addition, is targeting to the vacuole in leaf cells and to the protein storage vacuole in the cells of seeds, even in cases where the recombinant protein is of mammalian origin. In the case of immunoglobulins, this may be due to a cryptic sorting determinant present in the heavy chain (Hadlington et al., 2003). On the other hand, recombinant phytase bearing the KDEL peptide was partially sorted to the vacuole of *Medicago trunculata,* even when the same protein lacking the KDEL peptide was totally secreted to the apoplast (Abranches et al., 2008). This phenomenon may also be tissue specific, as one recombinant antibody expressed in tobacco was partially sorted to the vacuole in seed tissue, but completely secreted in leaf tissue (Petrucelli et al., 2006). To summarize, the KDEL peptide is a useful tool in glycan modelling, but it is only effective for certain proteins; the KDEL peptide often does not completely retain the target protein in the ER, and this aberrant targeting is responsible for complex glycan formation and unwanted proteolysis.

Expressing proteins in seeds

As the only protein source for the plant embryo, seeds are naturally geared towards high levels of protein expression (Boothe et al., 2010). They contain very little water and are far richer in protein than vegetative tissue. Seeds serve as hardy, long-term storage containers for recombinant proteins, which facilitates transport to the processing plant and does not require the protein to be purified immediately after

harvest. Recombinant hirudin has been stored for three years in maize seeds, and human proinsulin for seven years in soybean (Cunha et al., 2010). Likewise, recombinant α-L-iduronidase and glucocerebrosidase are stable for long time periods in mature, dry *Arabidopsis* seeds when kept refrigerated (Downing et al., 2007). In addition to their tremendous capacity to serve as stable repositories of recombinant proteins during storage in the dry state, seeds protect the recombinant protein better than leaf tissue during sterilization procedures required for therapeutic protein purification. Using endogenous regulatory sequences that control the expression of leguminous seed storage proteins such as arcelin, phaseolin, and cruciferin, high levels of recombinant protein expression have been achieved.

Seed-based production has some disadvantages as well. One is the time required to transform and screen stable transformants, which can take months. In contrast, transient expression trials can be completed in as little as one week in leaf tissue (Vezina et al., 2009). Furthermore, seeds can sort secretory proteins to the protein storage vacuole (PSV) in an unreliable manner. Although PSV targeting can be a useful technique for the stable accumulation of recombinant proteins, the PSV also contains proteases that may cause unwanted degradation of the target protein (Vezina et al, 2009).

1.4. The Present Study

The goal of my research was the development of a plant-based system for recombinant protein production. Specifically, I sought to optimize the production of recombinant human α-L-iduronidase (IDUA) in *Arabidopsis thaliana* seeds. In order to be therapeutically efficacious, the human IDUA enzyme requires mannose-6-phosphate (M6P) to mediate its uptake by human cells and concomitant delivery to the lysosome. There were two specific objectives:

(1) Improvement of the N-glycan profile of IDUA so that the plant-made product was non-immunogenic and amenable to subsequent phosphorylation and,

(2) To conduct in vitro phosphorylation assays to elaborate the M6P tag on the plant-made IDUA. This was accomplished through the use of a soluble form of the phosphotransferase that was obtained from a research collaborator.

1.5. References

Abranches R, Arcalis E, Marcel S, Altmann F, Ribeiro-Pedro M, Rodriguez J, Stoger E. (2008). Functional specialization of Medicago truncatula leaves and seeds does not affect the subcellular localization of a recombinant protein. *Planta.* **227(3):**649-58.

Aviezer D, Brill-Almon E, Shaaltiel Y, Hashmueli S, Bartfeld D, Mizrachi S, Liberman Y, Freeman A, Zimran A, Galun E. (2009). A plant-derived recombinant human glucocerebrosidase enzyme--a preclinical and phase I investigation. *PloS One.* **4(3):**e4792.

Baldwin ET, Bhat TN, Gulnik S, Hosur MV, Sowder RC 2nd, et al. (1993). Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc Natl Acad Sci USA*. **90:**6796-6800.Bao M, Booth JL, Elmendorf BJ, Canfield WM. (1996). Bovine UDP-N-acetylglucosamine:lysosomal-enzyme *N*-acetylglucosamine-1-phosphotransferase. *J Biol Chem.* **271(49):**31437-45.

Barton NW, Brady RO, Dambrosia JM, Di Bisceglie AM, Doppelt SH, Hill SC, Mankin HJ, Murray GJ, Parker RI, Argoff CE, RPG, Yu KT. (1991) Replacement therapy for inherited enzyme deficiency—macrophage-targeted glucocerebrosidase for Gaucher's disease. *N. Engl. J. Med.* **324**: 1464-70.

Beck, M. (2010). Therapy for lysosomal storage disorders. *IUBMB Life,* **62:**33-40.

Benjamin ER, Flanagan JJ, Schilling A, Chang HH, Agarwal L, Katz E, Wu X, Pine C, Wustman B, Desnick RJ, Lockhart DJ, Valenzano KJ. (2009). The pharmacological chaperone 1-deoxygalactonojirimycin increases alpha-galactosidase A levels in Fabry patient cell lines. *J Inherit Metab Dis.* **32(3):**424-40.

Bent, A. (2006). *Arabidopsis thaliana* Floral Dip Transformation Method. In *Methods in Molecular Biology: Agrobacterium Protocols.* Wang, K Ed.; Humana Press: New York, 2006; Vol. 343., pp 87-103.

Betenbaugh MJ, Tomiya N, Narang S, Hsu JT, Lee YC. (2004). Biosynthesis of human-type *N*-glycans in heterologous systems. *Curr Opin Struct Biol.* **14(5):**601-6.

Blanz J, Stroobants S, Lüllmann-Rauch R, Morelle W, Lüdemann M, D'Hooge R, Reuterwall H, Michalski JC, Fogh J, Andersson C, Saftig P. (2008). Reversal of peripheral and central neural storage and ataxia after recombinant enzyme replacement therapy in alpha-mannosidosis mice. *Hum Mol Genet.* **17(22):**3437-45.

Boelens JJ, Wynn RF, O'Meara A, Veys P, Bertrand Y, Souillet G, Wraith JE, Fischer A, Cavazzana-Calvo M, Sykora KW, Sedlacek P, Rovelli A, Uiterwaal CS, Wulffraat N. (2007). Outcomes of hematopoietic stem cell transplantation for Hurler's syndrome in Europe: a risk factor analysis for graft failure. *Bone Marrow Transplant.* **40(3):**225-33.

Boelens JJ, Aldenhoven M, Purtill D, Ruggeri A, Defor T, Wynn R, Wraith E, Cavazzana-Calvo M, Rovelli A, Fischer A, Tolar J, Prasad VK, Escolar M, Gluckman E, O'Meara A, Orchard PJ, Veys P, Eapen M, Kurtzberg J, Rocha V; Eurocord; Inborn Errors Working Party of European Blood and Marrow Transplant group; Duke University Blood and Marrow Transplantation Program; Centre for International Blood and Marrow Research. (2013). Outcomes of transplantation using various hematopoietic cell sources in children with Hurler syndrome after myeloablative conditioning. *Blood.* **121(19):**3981-7.

Borges N, Shi F, Azevedo I, Audus KL. (1994). Changes in brain microvessel endothelial cell monolayer permeability induced by adrenergic drugs. *Eur J Pharmacol.* **269:**243-248.

Braulke T, Pohl S, Storch S. (2008). Molecular analysis of the GlcNac-1-phosphotransferase. J Inherit Metab Dis. **31(2):**253-7.

Castilho A, Pabst M, Leonard R, Veit C, Altmann F, Mach L, Glössl J, Strasser R, Steinkellner H. (2008). Construction of a functional CMP-sialic acid biosynthesis pathway in Arabidopsis. *Plant Physiol.* **147(1):**331-9.

Castilho A, Neumann L, Gattinger P, Strasser R, Vorauer-Uhl K, Sterovsky T, Altmann F, Steinkellner H. (2013). Generation of biologically active multi-sialylated recombinant human EPOFc in plants. *PLoS One.* **8(1):**e54836.Chen CH, Dellamaggiore KR, Ouellette CP, Sedano CD, Lizadjohry M, Chernis GA, Gonzales M, Baltasar FE, Fan AL, Myerowitz R, Neufeld EF. (2008). Aptamer-based endocytosis of a lysosomal enzyme. *Proc Natl Acad Sci*. **105(41):**15908-13.

Connock M, Burls A, Frew E, Fry-Smith A, Juarez-Garcia A, McCabe C, Wailoo A, Abrams K, Cooper N, Sutton A, O'Hagan A, Moore D. (2006). The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review. *Health Technol Assess.* **10(24):**iii-iv, ix-136.

Cox KM, Sterling JD, Regan JT, Gasdaska JR, Frantz KK, Peele CG, Black A, Passmore D, Moldovan-Loomis C, Srinivasan M, Cuison S, Cardarelli PM, Dickey LF. (2006). Glycan optimization of a human monoclonal antibody in the aquatic plant Lemna minor. *Nat. Biotechnol.* **24(12):**1591-7.

Cunha NB, Araújo AC, Leite A, Murad AM, Vianna GR, Rech EL. (2010). Correct targeting of proinsulin in protein storage vacuoles of transgenic soybean seeds. *Genet Mol Res.* **9(2):**1163-70

Cuozzo JW, Tao K, Cygler M, Mort JS, Sahagian GG. (1998). Lysine-Based Structure Responsible for Selective Mannose Phosphorylation of Cathepsin D and Cathepsin L Defines a Common Structural Motif for Lysosomal Enzyme Targetting. *J Biol Chem.* **273:**21067-76.

Dahms NM, Olson LJ, Kim JJ. (2008). Strategies for carbohydrate recognition by the mannose 6-phosphate receptors. *Glycobiology.* **18(9):**664-78.

Datla RS, Hammerlindl JK, Panchuk B, Pelcher LE, Keller W. (1992). Modified binary plant transformation vectors with the wild-type gene encoding NPTII. *Gene.* **122(2):**383-4.

de Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD. (2006). E unum pluribus: multiple proteins from a self-processing polyprotein. *Trends Biotechnol.* **24(2):**68-75.

De Jaeger G, Scheffer S, Jacobs A, Zambre M, Zobell O, Goossens A, Depicker A, Angenon G. (2002). Boosting heterologous protein production in transgenic dicotyledonous seeds using Phaseolus vulgaris regulatory sequences. *Nat Biotechnol.* **20(12):**1265-8.

Do H, Lee WS, Ghosh P, Hollowell T, Canfield W, Kornfeld S. (2002). Human mannose 6-phosphate-uncovering enzyme is synthesized as a proenzyme that is activated by the endoprotease furin. *J Biol Chem.* **277(33):**29737-44.

Downing WL, Galpin JD, Clemens S, Lauzon SM, Samuels AL, Pidkowich MS, Clarke LA, Kermode AR. (2006). Synthesis of enzymatically active human alpha-L-iduronidase in Arabidopsis *cgl* (complex glycan-deficient) seeds. *Plant Biotechnol J.* **4(2):**169-81.

Downing WL, Hu X, Kermode AR. (2007). Post-transcriptional factors are important for high-level expression of the human α-L-iduronidase gene in *Arabidopsis cgl* (*complex-glycan-deficient*) seeds. *Plant Sci.* **172(2):**327-34.

Fischer HD, Gonzalez-Noriega A, Sly WS. (1980). Beta-g lucuronidase binding to human fibroblast membrane receptors. *J Biol Chem.* **255(11):**5069-74.

Frey AD, Karg SR, Kallio PT. (2009). Expression of rat β(1,4)-*N*-acetylglucosaminyltransferase III in *Nicotiana tabacum* remodels the plant-specific *N*-glycosylation. *Plant Biotechnol J.* **7:**33–48.

Fujiyama K, Misaki R, Sakai Y, Omasa T, Seki T. (2009). Change in glycosylation pattern with extension of endoplasmic reticulum retention signal sequence of mouse antibody produced by suspension-cultured tobacco BY2 cells. *J. Biosci. Bioeng.* **107(2):**165-72.

Galpin JD, Clemens S, Kermode AR. (2010). The carboxy-terminal ER-retention motif, SEKDEL, influences the *N*-linked glycosylation of recombinant human α-l-iduronidase but has little effect on enzyme activity in seeds of *Brassica napus* and *Nicotiana tabacum*. *Plant Sci.* **178(5):**440-7.

Gaucher PCE. De l'epithelioma primitif de la rate, hypertrophie idiopathique de la rate sans leucemie. [academic thesis]. Paris, France; 1882.

Giugliani R, Waldek S, Germain DP, Nicholls K, Bichet DG, Simosky JK, Bragat AC, Castelli JP, Benjamin ER, Boudes PF. (2013). A Phase 2 study of migalastat hydrochloride in females with Fabry disease: selection of population, safety and pharmacodynamic effects. *Mol Genet Metab.* **109(1):**86-92.Gomord V, Fitchette A.-C, Menu-Bouaouiche L, Saint-Jore-Dupas C, Plasson C, Michaud D, Faye L. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol J.* **8:**564-87.

Grabowski GA, Hopkin RJ. (2003). Enzyme therapy for lysosomal storage disease: principles, practice, and prospects. *Annu Rev Genomics Hum Genet.* **4:**403-36.

Grabowski GA, Golembo M, Shaaltiel Y. (2014). Taliglucerase alfa: an enzyme replacement therapy using plant cell expression technology. *Mol Genet Metab.* **112(1):**1-8.

Grubb JH, Vogler C, Levy B, Galvin N, Tan Y, Sly WS. (2008). Chemically modified β-glucuronidase crosses blood–brain barrier and clears neuronal storage in murine mucopolysaccharidosis VII.

Hadlington JL, Santoro A, Nuttall J, Denecke J, Ma JK, Vitale A, Frigerio L. (2003). The C-terminal Extension of a Hybrid Immunoglobulin A/G Heavy Chain Is Responsible for Its Golgi-mediated Sorting to the Vacuole. *Mol Biol Cell.* **14(6):**2592-2602.

Halpin C, Cooke SE, Barakate A, El Amrani A, Ryan MD. (1999). Self-processing 2A-polyproteins--a system for co-ordinate expression of multiple proteins in transgenic plants.

Hassan I, Waheed A, Grubb JH, Klei HE, Korolev S, Sly WS. (2013). High Resolution Crystal Structure of Human β-Glucuronidase Reveals Structural Basis of Lysosome Targeting. *PLoS One.* **8(11):**e79687.

He X, Haselhorst T, von Itzstein M, Kolarich D, Packer NH, Kermode AR. (2012). Influence of an ER-retention signal on the *N*-glycosylation of recombinant human α-L-iduronidase generated in seeds of Arabidopsis. *Plant Mol Biol.* 79(1-2):157-69.

He X, Galpin JD, Tropak MB, Mahuran D, Haselhorst T, von Itzstein M, Kolarich D, Packer NH, Miao Y, Jiang L, Grabowski GA, Clarke LA, Kermode AR. (2012). Production of active human glucocerebrosidase in seeds of Arabidopsis thaliana complex-glycan-deficient (cgl) plants. *Glycobiology*. 22(4):492-503.

He X, Pierce O, Haselhorst T, von Itzstein M, Kolarich D, Packer NH, Gloster TM, Vocadlo DJ, Qian Y, Brooks D, Kermode AR. (2013). Characterization and downstream mannose phosphorylation of human recombinant α-L-iduronidase produced in Arabidopsis complex glycan-deficient (cgl) seeds. *Plant Biotechnol J*. 11(9):1034-43.

Huether CM, Lienhart O, Baur A, Stemmer C, Gorr G, Reski R, Decker EL. (2005). Glyco-engineering of moss lacking plant-specific sugar residues. *Plant Biol (Stuttg).* **7(3):**292-9.

Islam MR, Grubb JH, Sly WS. (1993). C-terminal processing of human beta-glucuronidase. The propeptide is required for full expression of catalytic activity, intracellular retention, and proper phosphorylation. *J Biol Chem.* **268(30):**22627-33.

Islam MR, Tomatsu S, Shah GN, Grubb JH, Jain S, Sly WS. (1999). Active site residues of human beta-glucuronidase. Evidence for Glu(540) as the nucleophile and Glu(451) as the acid-base residue. *J Biol Chem.* **274(33):**23451-5.

Jain S, Drendel WB, Chen ZW, Mathews FS, Sly WS, Grubb JH. (1996). Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. *Nat Struct Biol.* **3(4):**375-81.

Jmoudiak M, Futerman AH.(2005). Gaucher disease: pathological mechanisms and modern management. *Br J Haematol.* **129(2):**178-88.

Kajiura H, Koiwa H, Nakazawa Y, Okazawa A, Kobayashi A, Seki T, Fujiyama K. (2010). Two Arabidopsis thaliana Golgi alpha-mannosidase I enzymes are responsible for plant *N*-glycan maturation. *Glycobiology.* **20(2):**235-47.

Kingdon EC, Russell JSR. (1896). Infantile cerebral degeneration with symmetrical changes in the macula. *Proc Royal Med Chirurgical Soc.* 934- 38.

Kim JJ, Olson LJ, Dahms NM (2009). Carbohydrate recognition by the mannose-6-phosphate receptors. *Curr Opin Struct Biol.* **19(5):**534-42.

Kudo M, Canfield WM. (2006). Structural requirements for efficient processing and activation of recombinant human UDP-N-acetylglucosamine:lysosomal-enzyme-N-acetylglucosamine-1-phosphotransferase. *J Biol Chem.* **281(17):**11761-8.

Lee JK, Pierce M. (1995). Purification and characterization of human serum *N*-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase. *Arch Biochem Biophys.* **319(2):**413-25.

Lee WS, Payne BJ, Gelfman CM, Vogel P, Kornfeld S. (2007). Murine UDP-GlcNAc:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase lacking the gamma-subunit retains substantial activity toward acid hydrolases. *J Biol Chem.* **282(37):**27198-203.

Lee WS, Rohrer J, Kornfeld R, Kornfeld S. (2002). Multiple signals regulate trafficking of the mannose 6-phosphate-uncovering enzyme. *J Biol Chem.* **277(5):**3544-51.

Liebminger E, Hüttner S, Vavra U, Fischl R, Schoberer J, Grass J, Blaukopf C, Seifert GJ, Altmann F, Mach L, Strasser R. (2009). Class I alpha-mannosidases are required for *N*-glycan processing and root development in Arabidopsis thaliana. **21(12):**3850-67.

Loos A, Steinkellner H. (2012). IgG-Fc glycoengineering in non-mammalian expression hosts. *Arch Biochem Biophys.* **526(2):**167-73.

Lund TC. (2013). Hematopoetic stem cell transplant for lysosomal storage diseases. *Pediatr Endocrinol Rev.* **11(Suppl 1):**91-8.

Lund TC, Cathey SS, Miller WP, Eapen M, Andreansky M, Dvorak CC, Davis JH, Dalal JD, Devine SM, Eames GM, Ferguson WS, Giller RH, He W, Kurtzberg J, Krance R, Katsanis E, Lewis VA, Sahdev I, Orchard PJ. (2014). Outcomes after Hematopoietic Stem Cell Transplantation for Children with I-Cell Disease. *Biol Blood Marrow Transplant.* 10 July 2014. [Epub ahead of print].

Marchesan D, Cox TM, Deegan PB. (2012). Lysosomal delivery of therapeutic enzymes in cell models of Fabry disease. *J Inherit Metab Dis.* **35(6):**1107-17.

Marsden D, Levy H. (2010). Newborn screening of lysosomal storage disorders. *Clin Chem.* **56(7):**1071-9.

Marschner K, Kollman K, Schweizer M, Braulke T, Pohl S. (2011). A key enzyme in the biogenesis of lysosomes is a protease that regulates cholesterol metabolism. *Science.* **333(6038):**87-90.

Montaño AM, Oikawa H, Tomatsu S, Nishioka T, Vogler C, Gutierrez MA, Oguma T, Tan Y, Grubb JH, Dung VC, Ohashi A, Miyamoto K, Orii T, Yoneda Y, Sly WS. (2008). Acidic amino acid tag enhances response to enzyme replacement in mucopolysaccharidosis type VII mice. *Mol Genet Metab.* **94(2):**178-89.

Munro S. (2001). The MRH domain suggests a shared ancestry for the mannose-6-phosphate recptors and other N-glycan-recognising proteins. *Curr Biol.* **11(13):**R499-R501.Mullis KG, Huynh M, Kornfeld RH. (1994). Purification and kinetic parameters of bovine liver *N*-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase. *J Biol Chem.* **269(3):**1718-26.

Mullis KG, Kornfeld RH. (1994). Characterization and immunolocalization of bovine *N*-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase. *J Biol Chem.* **269(3):**1727-33.

Nair P, Schaub BE, Huang K, Chen X, Murphy RF, Griffith JM, Geuze HJ, Rohrer J. (2005). Characterization of the TGN exit signal of the human mannose 6-phosphate uncovering enzyme. *J Cell Sci.* **118(Pt 13):**2949-56.

Nishikawa A, Nanda A, Gregory W, Frenz J, Kornfeld S. (1999). Identification of amino acids that modulate mannose phosphorylation of mouse DNase I, a secretory glycoprotein. *J Biol Chem.* 274(27):19309-15.

Ockerman PA. (1969). Mannosidosis: isolation of oligosaccharide storage material from brain. *J Pediatr.* **75:**360–365.

Parenti G. (2009). Treating lysosomal storage diseases with pharmacological chaperones: from concept to clinics. *EMBO Mol Med.* **1(5):**268-79.

Petruccelli S, Otegui MS, Lareu F, Tran Dinh O, Fitchette AC, Circosta A, Rumbo M, Bardor M, Carcamo R, Gomord V, Beachy RN. (2006). A KDEL-tagged monoclonal antibody is efficiently retained in the endoplasmic reticulum in leaves, but is both partially secreted and sorted to protein storage vacuoles in seeds. *Plant Biotechnol J.* **4(5):**511-27.

Popovic T, Puizdar V, Brzin J. (2002). A novel subtilase from common bean leaves. *FEBS Lett.* **530(1-3):**163-8.

Prabakaran T, Nielsen R, Satchell SC, Mathieson PW, Feldt-Rasmussen U, Sørensen SS, Christensen EI. (2012). Mannose-6-phosphate receptor and sortilin mediated endocytosis of α-galactosidase A in kidney endothelial cells. (2012). *PLoS One.* **7(6):**e39975.Qian Y, Lee I, Lee WS, Qian M, Kudo M, Canfield WM, Lobel P, Kornfeld S. (2010). Functions of the alpha, beta, and gamma subunits of UDP-GlcNAc:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase. *J Biol Chem.* **285(5):**3360-70.

Qian Y, Flanagan-Steet H, van Meel E, Steet R, Kornfeld SA. (2013). The DMAP interaction domain of UDP-GlcNAc:lysosomal enzyme *N*-acetylglucosamine -1-phosphotransferase is a substrate recognition module. *Proc Nat Acad Sci USA.* **110(25):**10246-51.

Reitman ML, Kornfeld S. (1981). Lysosomal enzyme targeting. *N*-Acetylglucosaminylphosphotransferase selectively phosphorylates native lysosomal enzymes. *J Biol Chem.* **256(23)**:11977-80.

Ruuska SA, Ohlrogge JB. (2001). Protocol for small-scale RNA isolation and transcriptional profiling of developing Arabidopsis seeds. *Biotechniques.* **31(4):**752, 754, 756-8.

Saint-Jore-Dupas C, Nebenführ A, Boulaflous A, Follet-Gueye ML, Plasson C, Hawes C, Driouich A, Faye L, Gomord V. (2006). Plant *N*-glycan processing enzymes employ different targeting mechanisms for their spatial arrangement along the secretory pathway. *Plant Cell*. **18(11):**3182-200.

Scott HS, Bunge S, Gal A, Clarke LA, Morris CP, Hopwood JJ. (1995). Molecular genetics of mucopolysaccharidosis type I: diagnostic, clinical, and biological implications. *Hum Mutat.* **6(4):**288-302.

Shipley JM, Grubb JH, Sly WS. (1993). The role of glycosylation and phosphorylation in the expression of active human beta-glucuronidase. *J Biol Chem.* **268(16):**12193-8.

Skropeta, D. (2009). The effect of individual *N*-glycans on enzyme activity. *Bioorg Med Chem.* **17(7):**2645-53.

Sly WS, Vogler C, Grubb JH, Levy B, Galvin N, Tan Y, Nishioka T, Tomatsu S. (2006). Enzyme therapy in mannose receptor-null mucopolysaccharidosis VII mice defines roles for the mannose-6-phosphate and mannose receptors. *Proc Nat Acad Sci.* **103(41):**15172-15177.

Sourrouille C, Marhsall B, Lienard D, Faye L. (2009). From neanderthal to nanobiotech: from plant potions to pharming with plant factories. In *Methods in Molecular Biology: Recombinant Proteins in Plants.* Faye L, Gomord V Eds.; Humana Press: New York, 2009; Vol 483., pp1-23.

Steet R, Lee WS, Kornfeld S. (2005). Identification of the minimal lysosomal enzyme recognition domain in cathepsin D. *J Biol Chem.* **280(39):**33318-23.

Stigliano ID, Caramelo JJ, Labriola CA, Parodi AJ, D'Alessio C. (2009). Stigliano IDGlucosidase II beta subunit modulates *N*-glycan trimming in fission yeasts and mammals., Caramelo JJ, Labriola CA, Parodi AJ, D'Alessio C. *Mol Biol Cell.* **20(17):**3974-84.

Strasser R, Altmann F, Mach L, Glössl J, Steinkellner H. (2004). Generation of Arabidopsis thaliana plants with complex *N*-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. *FEBS Lett.* **561(1-3):**132-6.

Strasser R, Bondili JS, Vavra U, Schoberer J, Svoboda B, Glössl J, Léonard R, Stadlmann J, Altmann F, Steinkellner H, Mach L. (2007). A unique beta1,3-galactosyltransferase is indispensable for the biosynthesis of *N*-glycans containing Lewis a structures in Arabidopsis thaliana. *Plant Cell.* **19(7):**2278-92.

Strasser R, Stadlmann J, Schähs M, Stiegler G, Quendler H, Mach L, Glössl J, Weterings K, Pabst M, Steinkellner H. (2008). Generation of glyco-engineered Nicotiana benthamiana for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. *Plant Biotechnol J.* **6(4):**392-402.

Tomatsu S, Montaño AM, Dung VC, Grubb JH, Sly WS. (2009). Mutations and polymorphisms in GUSB gene in mucopolysaccharidosis VII (Sly Syndrome). *Hum Mutat.* **30(4):**511-9.

Tomatsu S, Montaño AM, Dung VC, Ohashi A, Oikawa H, Oguma T, Orii T, Barrera L, Sly WS. (2010). Enhancement of drug delivery: enzyme-replacement therapy for murine Morquio A syndrome. *Mol Ther*. **18(6):**1094-102.

Urayama A, Grubb JH, Banks WA, Sly WS. (2007). Epinephrine enhances lysosomal enzyme delivery across the blood–brain barrier by up-regulation of the mannose 6-phosphate receptor. *Proc Nat Acad Sci.* **104(31):**12873-12878.

Urayama A, Grubb JH, Sly WS, Banks WA. (2004). Developmentally regulated mannose 6-phosphate receptor-mediated transport of a lysosomal enzyme across the blood-brain barrier. *Proc Nat Acad Sci.* **101(34):**12658-12663.

Urayama A, Grubb JH, Sly WS, Banks WA. (2008). Mannose 6-phosphate receptor-mediated transport of sulfamidase across the blood-brain barrier in the newborn mouse. *Mol Ther.* **16(7):**1261-6.

Varki A, Sherman W, Kornfeld S. (1983). Demonstration of the enzymatic mechanisms of α-N-acetyl-d-glucosamine-1-phosphodiester N-acetylglucosaminidase (formerly called α-N-acetylglucosaminylphosphodiesterase) and lysosomal α-N-acetylglucosaminidase. *Arch Biochem Biophys.* 222(1):145-9.

Vellodi A. (2005). Lysosomal storage disorders. Br J Haematol. **128(4)**:413-31.

Vogler C, Levy B, Grubb JH, Galvin N, Tan Y, Kakkis E, Pavloff N, Sly WS. (2004). Overcoming the blood-brain barrier with high-dose enzyme replacement therapy in murine mucopolysaccharidosis VII. *Proc Natl Acad Sci.* **102(41):**14777-82.

von Schaewen A, Sturm A, O'Neill J, Chrispeels MJ. (1993). Isolation of a mutant Arabidopsis plant that lacks *N*-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex *N*-linked glycans. *Plant Physiol.* **102(4):**1109-18.

Warner JB, Thalhauser C, Tao K, Sahagian GG. (2002). Role of *N*-linked oligosaccharide flexibility in mannose phosphorylation of lysosomal enzyme cathepsin L. *J Biol Chem*. **277(44):**41897-905.

Watanabe T, Totani K, Matsuo I, Maruyama J, Kitamoto K, Ito Y. (2009). Genetic analysis of glucosidase II beta-subunit in trimming of high-mannose-type glycans. *Glycobiology.* **19(8):**834-40.

Wei Y, Yen TY, Cai J, Trent JO, Pierce WM, Young WW Jr. (2005). Structural features of the lysosomal hydrolase mannose 6-phosphate uncovering enzyme. *Glycoconj J.* **22(1-2):**13-9.

Wong AW, He S, Grubb JH, Sly WS, Withers SG. (1998). Identification of Glu-540 as the catalytic nucleophile of human beta-glucuronidase using electrospray mass spectrometry. *J Biol Chem.* **273(51):**34057-62.

Yoshida Y, Tanaka K. (2010). Lectin-like ERAD players in ER and cytosol. *Biochim Biophys Acta*. **1800(2):**172-80.

Zha D. (2013). Glycoengineered *Pichia*-based expression of monoclonal antibodies. *Methods Mol Biol.* **988:**31-43.

1.6. Tables and Figures

Table 1. Overview of the biochemical and clinical features of three targeted LSDs

Disease	Gaucher Disease[1]	Hurler Syndrome / MPS I[2]	Sly Syndrome / MPS VII[3]
Defective gene	• GBA (glucocerebrosidase)	• IDUA (α-L-iduronidase)	• GUSB (β-D-glucuronidase)
Gene function	• Cleaves β-linked glucose from glucosylceramide	• Cleaves α-L-iduronic acid moieties from GAGs	• Cleaves β-D-glucuronic acid moieties from GAGs
Storage product	• Glucosylceramide, glucosylsphingosine	• Heparan sulfate, dermatan sulfate	• Heparan sulfate, dermatan sulfate, chondroitin sulfate
Year discovered	• 1882	• 1919	• 1969
Incidence (general pop)	• ~ 1:60,000; (Ashkenazi Jews: 1:1000)	• ~1:80,000	• <1:250,000
General features of the disease	• Macrophage activation leads to increased production of interleukins 1β and 6, also increased tumour necrosis factor- α • Hepatosplenomegaly	• Hepatosplenomegaly, corneal clouding, joint stiffness	• Short stature, hepatosplenomegaly , mild to severe mental retardation, craniofacial dysmorphism
Clinical spectrum of disease	• Type I: adult onset non-neuronopathic • Type II: acute neuronopathic, death within 3 yrs • Type III: chronic neuronopathic, death in 3rd/4th decade	• Hurler: Severe mental retardation and skeletal malformation, death in early childhood. • Schie: Late-onset, normal intelligence normal lifespan • Hurler-Schie: normal intelligence, skeletal pathology	• Proposed clinical spectrum: • Early severe lethal form, • Intermediate form, • Mild form
ERT Available	• Cerezyme, Uplyso	• Aldurazyme	• None
Other Treaments	• Substrate reduction therapy (Misglustat)	• Hematopoetic stem cell therapy	• Surgical relief of deformities

[1] Jmoudiak and Futerman (2005). [2] Scott et al. (1995). [3] Tomatsu et al. (2009).

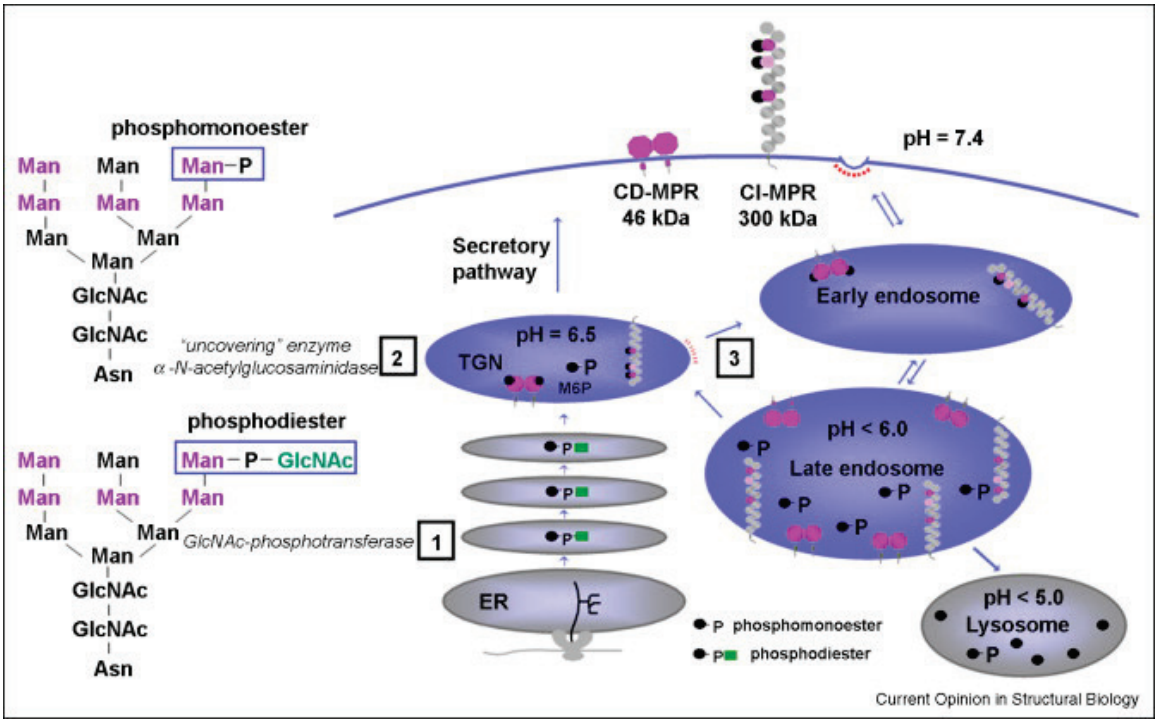


Figure 1.1. Addition of mannose-6-phosphate to lysosomal enzymes
Trafficking of lysosomal enzymes through the secretory pathway to the lysosome is depicted. The mannose-6-phosphate tag is added to the *N*-glycans of lysosomal enzymes by the action of the GlcNAc-phosphotransferase (step 1) and the uncovering enzyme (step 2) to one of the pink mannose residues as depicted on the left hand side of the figure. Recognition of the M6P marker by the cation-dependant or the cation-independent MP6R (step 3) allows the lysosomal enzyme to be trafficked to the lysosome through the endosomal network. The lysosomal enzyme is represented as a black circle. Reprinted from Kim et al. (2009) with permission.

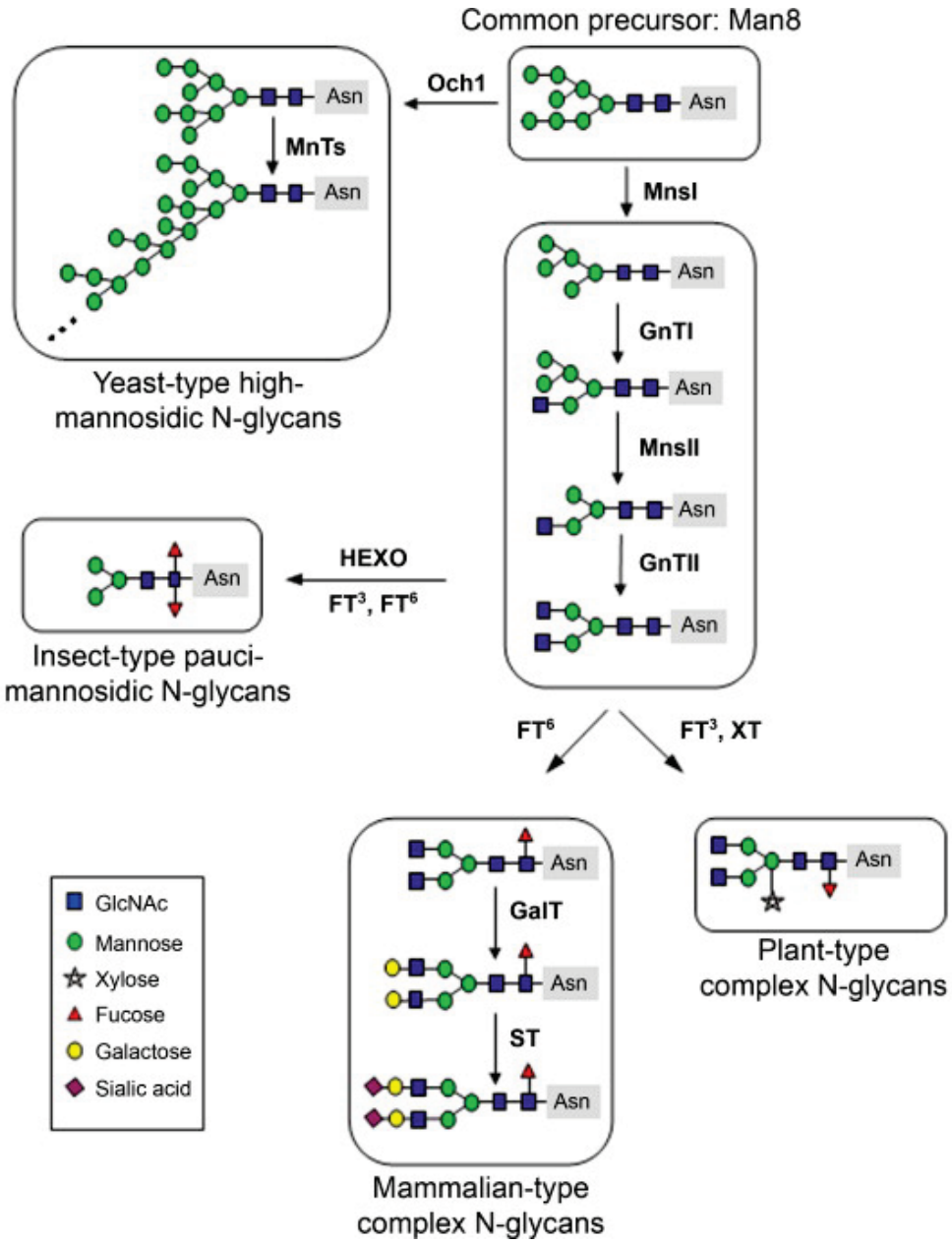


Figure 1.2. *N*-glycan maturation in Eukaryotic organisms.
Schematic presentation of selected components of the N-glycosylation pathways in humans, yeast, insect cells and plants. The common ER-resident oligosaccharide precursor Man8 acts as starting point for further modifications along the Golgi apparatus. Och1: α1,6-mannosyltransferase; MnTs: mannosyltransferases; Mns: mannosidase; GnT: N-acetylglucosaminyltransferase; GalT: α1,4-galactosyltransferase; ST: α2,6-sialyltransferase; HEXO: hexosaminidase (N-acetylglucosaminidase); XT: β 1,2-xylosyltransferase; FT: core fucosyltransferase; Fucose can be transferred in α1,3-linkage (plant typical) and α1,6-linkage (mammalian typical). Interestingly, both forms are synthesized in insect cells. Oligosaccharide abbreviations according to www.proglycan.com. Figure and caption from Loos and Steinkellner (2012). Reprinted under a Creative Commons licence.

Chapter 2. Characterization and downstream mannose phosphorylation of human recombinant alpha-L-iduronidase produced in *Arabidopsis* complex glycan-deficient (cgl) seeds.

2.1. Abstract

Mucopolysaccharidosis (MPS) I is a lysosomal storage disease caused by a deficiency of α-L-iduronidase (IDUA) (EC 3.2.1.76); enzyme replacement therapy is the conventional treatment for this genetic disease. *Arabidopsis cgl* mutants are characterized by a deficiency of the activity of *N*-acetylglucosaminyl transferase I (EC 2.4.1.101), the first enzyme in the pathway of hybrid and complex *N*-glycan biosynthesis. To develop a seed-based platform for the production of recombinant IDUA for potential treatment of MPS I, *cgl* mutant seeds were generated to express human IDUA at high yields and to avoid maturation of the *N*-linked glycans on the recombinant human enzyme. Enzyme kinetic data showed that *cgl*-IDUA has similar enzymatic properties to the commercial recombinant IDUA derived from cultured Chinese hamster ovary (CHO) cells (Aldurazyme™). The *N*-glycan profile showed that *cgl*-derived IDUA contained predominantly high-mannose-type *N*-glycans (94.5%), and the residual complex/hybrid *N*-glycan-containing enzyme was efficiently removed by an additional affinity chromatography step. Furthermore, purified *cgl*-IDUA was amenable to sequential *in vitro* processing by soluble recombinant forms of the two enzymes that mediate the addition of the mannose-6-phosphate (M6P) tag in mammalian cells—UDP-GlcNAc:lysosomal enzyme *N*−acetylglucosamine (GlcNAc)−1−phosphotransferase—and GlcNAc−1−phosphodiester α−*N*−acetylglucosaminidase (the 'uncovering enzyme'). *Arabidopsis* seeds provide an alternative system for producing recombinant lysosomal enzymes for enzyme replacement therapy; the purified enzymes can be subjected to downstream processing to create the M6P, a recognition marker essential for efficient receptor-mediated uptake into lysosomes of human cells.

2.2. Introduction

Lysosomal storage diseases (LSDs) collectively represent over 70 inherited metabolic disorders. The underlying etiology of many LSDs involves a deficiency of a single acid hydrolase. Enzyme replacement therapy (ERT) is the conventional treatment. While some enzyme replacement therapeutics are in the 'pipeline' and are under different phases of clinical evaluation (van Gelder *et al*., 2012), to date, these therapeutics have been registered for only a few of the LSDs: Gaucher disease, Fabry disease, Pompe disease and mucopolysaccharidoses (MPS I, II, and VI). ERT is a process that takes advantage of plasma membrane receptor uptake to deliver recombinant enzyme into a patient's cells following intravenous infusion of the replacement enzyme (Neufeld, 2011). Mucopolysaccharidosis I (MPS I) disease is an LSD characterized by the deficiency of α-L-iduronidase (IDUA), an enzyme involved in the stepwise degradation of the glycosaminoglycans heparan sulphate and dermatan sulphate. Because of the widespread distribution of these glycosaminoglycans in tissues and organs, severely affected humans with no residual α-L-iduronidase typically die in early childhood due to profound skeletal, cardiac and neurological disturbances (Clarke, 2008). The current approved ERT for MPS I (*Aldurazyme*™ or *Laronidase*), based on the recombinant IDUA from Chinese hamster ovary (CHO) cells, is prohibitively expensive, costing approximately US$450 000 per year for an average 12-year-old child. As costs are calculated on the basis of the weight of an individual, treatments for teenagers and adults are even higher.

Plant-based systems offer an alternative protein production platform that could reduce the costs associated with these therapies. An enzyme therapeutic for Gaucher disease generated by a plant platform (carrot suspension cell cultures) was approved by the US Food and Drug Administration in 2012 (Maxmen, 2012; Shaaltiel *et al*., 2007) and represents the first plant-based recombinant therapeutic approved for parenteral administration. Some of the advantages of plant expression systems include the following: relatively low production costs due to an ability to rapidly accumulate biomass and a lack of susceptibility to contamination by human pathogens as can occur in mammalian expression systems, such as CHO cells or human fibroblasts. Seeds, as natural depots for stable accumulation of proteins, offer additional advantages as a 'host'

platform: recombinant proteins are usually stable in the mature seeds placed in conventional (cool, dry) storage conditions, and the seeds remain viable for several years. Thus, seeds serve as a stable repository of the therapeutic protein providing the convenient option of temporally separating the production of the recombinant protein from its purification and downstream processing (Kermode, 2012).

Nonetheless, the use of plants as platforms for the production of recombinant lysosomal enzymes suitable for ERT poses at least three technical challenges. First, sufficient yields of the recombinant enzyme must be obtained to render economic viability to the production platform. Second, a nonimmunogenic and active recombinant protein must be generated. While the initial steps of protein *N*-glycosylation and subsequent *N*-glycan trimming that are essential for proper protein folding are common to plant and animal cells, differences occur in *N*-glycan maturation as proteins transit through the distal compartments of the Golgi complex (*medial*-, *trans*- and *trans*-Golgi networks). In these compartments, enzymes convert the high-mannose-type *N*-glycans of proteins to complex *N*-glycans by a series of sequential reactions that rely on the accessibility of the glycan chain(s) to the Golgi processing machinery (Gomord *et al*., 2010; Kermode, 1996). In mammalian cells, core α1,6 fucose residues and terminal sialic acid residues are commonly added, whereas bisecting β1,2 xylose and core α1,3 fucose residues are assembled onto the trimmed *N*-glycans of plant-synthesized proteins. The presence of β1,2 xylose or α1,3 fucose residues on plant-produced pharmaceuticals can render the product immunogenic, which is a concern for any therapeutic destined for parenteral administration (reviewed in Gomord *et al*., 2010).

Third, there are challenges associated with generating a therapeutically efficacious product; the parenterally administered recombinant enzyme must have suitable targeting signals for endocytosis into patient cells and for intracellular delivery to the lysosome. For most lysosomal enzymes (e.g. α-L-iduronidase and several others), this requires the cellular recognition marker, mannose-6-phosphate (M6P), on the replacement protein.

In mammalian cells, Golgi-specific enzymatic processing mediates M6P elaboration, a process involving the modification of select terminal mannose residues of

some of the *N*-glycans of lysosomal hydrolases. Thus, lysosomal hydrolases typically possess various *N*-glycans, namely those of the high-mannose-, complex- or M6P-terminated types. The M6P recognition motif normally mediates the targeting of soluble lysosomal enzymes from the *trans*-Golgi network to the lysosome, but can also mediate the internalization of inappropriately secreted lysosomal enzymes via the so-called "salvage" pathway. Enzyme replacement therapy for LSDs largely exploits this endogenous salvage (uptake) pathway of human cells, which is effected by cell surface endocytic receptors, such as the M6P receptors (reviewed in Gary-Bobo *et al*., 2007). "Retrieval," or sequestration, of an extracellular (e.g. recombinant) lysosomal enzyme due to recognition of its M6P marker can result in delivery to the lysosome by receptor-mediated endocytosis via early and late endosomes, enabling the correction of an enzyme deficiency. Thus, a key challenge lies in the fact that plant cells do not have the endogenous machinery to create the M6P tag on a recombinant protein; a viable option is to create this on the purified protein *in vitro*, that is, as part of downstream processing.

In this study, we address the three challenges associated with plant-based platforms for the production of recombinant lysosomal enzymes destined for ERT, focussing on human IDUA. Through selection and inbreeding, we identified an *Arabidopsis thaliana* (Arabidopsis) seed line that significantly exceeds our previously reported yields of recombinant human IDUA by at least threefold (Downing *et al*., 2006, 2007); in this line, IDUA represented 5.7% of the total soluble seed protein (TSP), and the yield was stable through T3 and T4 generations. This Arabidopsis seed line represents a mutant background (the *complex glycan-deficient* or *cgl1* C5 mutant), in which *N*-glycan maturation is avoided due to a deficiency of the Golgi enzyme *N*-acetylglucosamine transferase I (GnT I); importantly, we show by graphitized carbon liquid chromatography–tandem mass spectrometry that high-mannose-type *N*-glycans accounted for 94.5% of the *N*-glycans on the *cgl*-IDUA. Kinetic parameters of the *cgl*-derived IDUA were comparable with those of Aldurazyme™ (the commercial CHO cell–derived IDUA); we further demonstrate that the plant recombinant enzyme was amenable to *in vitro* downstream processing using two recombinant soluble human enzymes: GlcNAc-1-phosphotransferase and the 'uncovering enzyme', to generate the M6P recognition motif.

Our results show that Arabidopsis *cgl* seeds provide a viable production system for recombinant human IDUA, generating enzyme that is potentially suitable for treating patients with MPS I.

This research is published in Plant Biotechnology Journal (He et al., 2013), for which the Post-Doctoral Fellow, Dr. Xu He (Kermode lab) was first author. Specifically my contributions were: (1) establishing the kinetic parameters of the cgl-IDUA in comparison to the CHO-cell-derived IDUA (Aldurazyme^{TM}) and (2) contributing to the downstream processing of purified *cgl*-IDUA to elaborate the M6P tag.

2.3. Results

2.3.1. Arabidopsis *cgl* seeds accumulate high levels of active α-L-iduronidase

Previously, we reported that using the promoter and other regulatory sequences of the *Phaseolus vulgaris arcelin 5-I* gene, three transgenic *Arabidopsis cgl* lines with exceptional activities were identified; these possessed IDUA activities of 820 ± 63, 745 ± 75 and 423 ± 19 units/mg TSP (Downing *et al*., 2007; lines A5.5, A4.7 and A6.3, respectively). The highest yielding line accumulated IDUA to approximately 18 μg IDUA/mg TSP (Downing *et al*., 2007). Here, we further improved the yield by selfing of plants and by further selection. T2 seeds from the three lines were germinated on selection media (25 mg/L kanamycin in half-strength Murashige and Skoog [MS] medium). Ten to 15 transgenic seedlings from each line were grown to maturity and selfed to obtain T3 seeds. Protein was extracted from each seed stock, and the expression levels were determined by activity assays and Western blot analyses. Ten transgenic seedlings from each of the highest expressers were selfed to generate T4 seeds. Interestingly, the highest expressing line of T3 seeds accumulated IDUA at 11.2% TSP (Table 2.1 and Figure 2.1a, lane 2), but the expression level of the subsequent generation (T4) decreased significantly (Table 2.1). Although we have not determined the zygosity of the transgene, the highest yielding line (with relatively stable trans-generation expression) accumulated IDUA at 5.7% TSP (Table 2.1).

2.3.2. *Cgl*-derived IDUA contains predominantly high-mannose-type *N*-glycans

There are six consensus sites for *N*-linked glycosylation on α-L-iduronidase (Asn-X-Ser/Thr); all sites are utilized in human cells (Kakkis, 2005). In CHO cells hosting production of the recombinant protein, the enzyme is secreted, and all six glycosylation sites are used, but the *N*-linked glycans themselves display high intrasite heterogeneity (Zhao *et al*., 1997). Some of the *N*-linked glycans of the mature enzyme remain in a high-mannose form (Asn 372; Asn 415 is mixed high mannose and complex); at least two of the sites are modified to complex forms (Asn 110 and Asn 190); two carry M6P tags for receptor-mediated uptake/lysosome delivery (Asn 336 and Asn 451) (Zhao *et al*., 1997).

In the Arabidopsis *cgl* mutant, the *N*-glycan status of the recombinant protein appears to depend on both the target protein and the plant tissues that 'host' its expression (see Discussion). As shown in Figure 2.2 and Table 2.2, 94.5% of the *N*-glycan structures detected in the purified *cgl*-IDUA (Figure 2.1b) by carbon LC MS/MS were of the oligomannosidic type containing 1–7 hexose residues in addition to the pentasaccharide *N*-glycan core. Among these structures, a Man5 structure (i.e. $Man_5GlcNAc_2$) was the most abundant (50.6%). Complex and hybrid structures containing core α 1,3 fucose or β 1,2 linked xylose accounted for 5.5% of the total structures detected (Table 2.2). One glycopeptide could be identified from the *cgl*-IDUA samples, indicating incomplete digestion with PNGase A. Nevertheless, this allowed the identification of high-mannose-type *N*-glycosylation at site Asn 372 (Figure 2.3).

The purified *cgl*-IDUA sample was subjected to further affinity chromatography with anti-horseradish peroxidase (HRP) antibodies to determine whether the small amounts of recombinant enzyme containing plant complex and hybrid *N*-glycans (approximately 5.5%) could be efficiently removed. As shown in Figure 2.4 and Table 2.3, Man5 accounted for 93.8% of the *N*-glycans, while Man6–Man8 accounted for 6.2% of the *N*-glycans. No plant complex or hybrid *N*-glycans were detected. Thus, the anti-HRP chromatography step gave rise to a single dominant *N*-glycan structure of IDUA.

2.3.3. *Cgl*-IDUA exhibits kinetic parameters that are comparable with the commercial enzyme

IDUA was purified to homogeneity from T4 seeds (line A4.7) by sequential use of Con A-sepharose chromatography and anti-IDUA affinity chromatography as described in He et al. (2012c) (Fig. 2.1b). Using 4-MUI as substrate, the *cgl*-IDUA exhibited a K_m of 44 μM, and a V_{max} of 5.6 μM min^{-1} mg^{-1} (Fig. 2.5). The values are comparable to those of IDUA produced in CHO cell cultures (Aldurazyme™), for which the K_m is 24 μM and the V_{max} is 3.9 μM min^{-1} mg^{-1} (He *et al*., 2012c).

2.3.4. *Cgl*-IDUA is amenable to sequential *in vitro* processing to add the M6P recognition marker

Most lysosomal enzymes including α-L-iduronidase require a M6P tag for efficient uptake and lysosomal delivery in human cells. In mammalian cells, the M6P tag is generated by the sequential action of two enzymes: UDP-GlcNAc:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase (also known as GlcNAc-1-phosphotransferase) and *N*-acetylglucosamine 1-phosphodiester α-*N*-acetylglucosaminidase (the uncovering enzyme). The endogenous mammalian GlcNAc-1-phosphotransferase is a heterohexamer comprised of three subunits ($\alpha_2\beta_2\gamma_2$) and is localized within the *cis*-Golgi (Bao *et al*., 1996). GlcNAc-1-phosphotransferase catalyses the transfer of GlcNAc-P from UDP-GlcNAc to C6 of select mannoses of the lysosomal enzyme's high-mannose-type oligosaccharides. The α/β subunits contain the catalytic function of the phosphotransferase and further are responsible for the specificity of the reaction by recognition of a conformation-dependent determinant of the lysosomal enzyme (Kudo *et al*., 2005; Lee *et al*., 2007; Qian *et al*., 2010). The human uncovering enzyme is synthesized as a proenzyme that is activated by the endoprotease furin in the *trans*-Golgi network (Do *et al*., 2002). This enzyme excises the terminal GlcNAc, exposing the M6P recognition marker on lysosomal hydrolases. Both the GlcNAc-1-phosphotransferase (as a $\alpha_2\beta_2$ heterotetramer) and the uncovering enzyme have been generated as soluble forms (Do *et al*., 2002; Kudo and Canfield, 2006) and so can be used to conduct the *in vitro* phosphorylation of purified recombinant lysosomal enzymes.

We first investigated whether the *N*-glycan terminal mannose residues of the *cgl*-IDUA could be modified *in vitro* using recombinant soluble GlcNAc-1-phosphotransferase ($\alpha_2\beta_2$). Equal amounts of the phosphotransferase (1.5 µg) were incubated with increasing concentrations of *cgl*-IDUA in the presence of 75 µm UDP-[^{3}H]GlcNAc for 1 h, and transfer of [^{3}H]GlcNAc-P to *cgl*-IDUA was measured. Figure 6 illustrates the activity of the GlcNAc-1-phosphotransferase towards *cgl*-IDUA based on the data of one representative experiment. Kinetic parameters were determined using the double-reciprocal plot from 3 experiments. *Cgl*-IDUA exhibited a K_m of 4.3 µm and a k_{cat} of 0.10 min^{-1} (V_{max} of 21.7 pmol/h/µg).

Next we examined the efficiency by which the uncovering enzyme cleaved GlcNAc from the phosphorylated *cgl*-IDUA by monitoring the concentrations of free [^{3}H]GlcNAc over a time course. The amount of liberated [^{3}H]-GlcNAc increased with time (Fig. 2.7), suggesting that the covering GlcNAc of the phosphorylated IDUA can be readily removed by the uncovering enzyme to expose the M6P tag.

2.4. Discussion

In this study, we sought to address the technical challenges of plant-based production of recombinant enzymes as replacement therapeutics for patients with lysosomal storage disease (LSD). To recap, these challenges pertain to the development of economically viable production systems that generate non-immunogenic and functional therapeutics (see Introduction).

2.4.1. High levels of active α-L-iduronidase (IDUA) are produced in Arabidopsis *cgl* seeds

Arabidopsis seeds have attracted considerable attention as a platform to produce recombinant proteins. In this regard, Arabidopsis has several advantages: its life cycle is short, it can be readily grown in a controlled environment, and compared with several other seed-based systems, it is not a food crop, so the regulatory issues surrounding its use as a source of recombinant proteins are reduced considerably. In addition, exceptionally high levels of functional recombinant proteins can be accumulated in

Arabidopsis seeds. For example, a single-chain variable fragment (scFv) antibody is accumulated to 74 µg/mg of TSP in homozygous Arabidopsis seeds (De Jaeger *et al*., 2002); likewise, the maximal accumulation of antihepatitis A virus antibodies reaches 9.8 µg/mg in dry seeds (Loos *et al*., 2011). In our study, we obtained accumulation levels of IDUA that significantly exceeded our previous reports (Downing *et al*., 2006, 2007), reaching 57 µg/mg TSP (corresponding to 8.0 µg/mg dry seeds) in T4 transgenic Arabidopsis seeds. At this level of expression, one milligram of purified IDUA can be obtained from less than 1 g of seeds using the chromatographic methods of the present study.

One might be able to apply the lessons learned from Arabidopsis to production of recombinant proteins in a larger seed system, possibly in seeds of a cereal or another dicot. However, efforts to completely knock-out *GnT I* expression in hosts other than Arabidopsis, for example in dicots such as potato and tobacco, lead to a reduction in complex *N*-glycans on endogenous plant proteins (Wenderoth and von Schaewen, 2000), but the extent of silencing would not adequately control the *N*-glycan maturation of a recombinant therapeutic protein. In addition, the yield of IDUA in the stable high-expressing *cgl* Arabidopsis line (A4.7) of the present study is approximately 47 times higher than that of transgenic maize seeds (1.2 µg/mg TSP) (He *et al*., 2012c). The amount of IDUA accumulation in *cgl* seeds also exceeds other seed-based systems; it is about 25–30 times higher than the IDUA levels accumulated by seeds of transgenic tobacco (2.2 µg/mg TSP) and *Brassica napus* (1.9 µg/mg TSP) using the same (arcelin gene) regulatory sequences to drive the human gene expression (Galpin *et al*., 2010).

The development of automated sowing and harvesting of pilled Arabidopsis seeds for bulk production of biomass (Loos *et al*., 2011) represents another advantage for the use of Arabidopsis as a host production system. With the expression level of *cgl*-IDUA at 7–9 µg/mg dry seeds (line A 4.7), the estimated annual production yields of IDUA per square metre (sqm) would range from 1.2 g to 1.6 g (based on 200 mg seeds per plant; 300 plants per sqm; 3 generations per year). A single patient of 40 kg (upper estimate for body weight of 12-year-old patient) (Pastores *et al*., 2007) would annually require approximately 1.2 g of purified IDUA (based on weekly infusions of 0.58 mg/kg of body weight as recommended for Aldurazyme;

http://www.aldurazyme.com/global/az_us_home.asp). With an estimated worldwide incidence of MPS I as 1 in 100 000 live births (Muenzer *et al*., 2009) (or potentially approximately 70 000 patients globally), the generation of sufficient *cgl*-IDUA to meet the potential annual market needs would require approximately 140 000 sqm of growth area (approximately 34.6 acres), taking into account the purification yields. Yet this calculation is based on many assumptions, and the true feasibility of using Arabidopsis seeds to produce enzyme therapeutics for LSD patients will rely upon numerous factors, not the least of which is the associated costs and efficiency of downstream processing (see subsequent Discussion).

2.4.2. The maturation of the *N*-glycans of recombinant IDUA is avoided in Arabidopsis *cgl* seeds

Aside from achieving sufficient yields of a therapeutic recombinant protein, plant-specific *N*-glycan maturation limits the use of plant-based systems as factories to produce recombinant pharmaceutical proteins; the presence of xylose and fucose residues on 'matured' *N*-glycans of the protein may render the product immunogenic if administered parenterally (see Introduction). The human IDUA enzyme has six consensus signals for *N*-linked glycosylation in the ER; these are all utilized in human cells in which the enzyme is targeted to the lysosome via the Golgi complex. In CHO cells hosting production of the recombinant protein, the enzyme is secreted, and again all six sites are used, but the *N*-linked glycans are heterogeneous (Zhao *et al*., 1997) (see Results). To reduce or eliminate plant-specific *N*-glycan maturation, various strategies have been implemented such as a targeted localization of recombinant proteins within the ER or ER-derived compartments, down-regulating glycosyltransferases (e.g. *N*-acetylglucosamine transferase I, α1,3 fucosyltransferases and β1,2 xylosyltransferase) by RNAi technology or by the use of mutants, and expressing human glycosyltransferases (e.g. β1,4-galactosyltransferase) (reviewed in Gomord *et al*., 2010; Kermode, 2012; He *et al*., 2012c). Towards developing various plant-based platforms for the production of recombinant lysosomal enzymes for use in ERT, we have used some of the strategies applied to recombinant human IDUA and/or glucocerebrosidase. Two strategies for ER retention of recombinant IDUA have included KDEL tagging (He *et al*., 2012b) and exploitation of a unique mRNA-targeting

mechanism in maize seeds, a process which effectively localizes recombinant IDUA to ER-derived protein bodies (He *et al*., 2012c). Only this latter strategy is effective in controlling the *N*-glycan maturation of IDUA and further has the advantage of yielding a native enzyme with no additional (i.e. foreign) amino acid sequences (He *et al*., 2012c). For example, the KDEL-tagged IDUA of Arabidopsis seeds contains 88% complex/hybrid *N*-glycans (He *et al*., 2012b), while the mRNA-mediated targeting of IDUA to ER-derived protein bodies in maize seeds generates only 8.2% hybrid/complex *N*-glycans (He *et al*., 2012c).

As shown in the present study (and in He *et al*., 2012a for glucocerebrosidase), use of seeds of the Arabidopsis *cgl* mutant is also a viable means to control the *N*-glycan maturation status of recombinant lysosomal enzymes. The *cgl* mutant is defective in the gene encoding the Golgi enzyme *N*-acetylglucosamine transferase I (GnT I) (EC 2.4.1.101) (von Schaewen *et al*., 1993; Strasser *et al*., 2005). GnT I transfers a GlcNAc residue from UDP-GlcNAc to the acceptor substrate $Man_5GlcNAc_2$ to produce $GlcNAcMan_5GlcNAc_2$, which is essential for the subsequent action of α1,3 fucosyltransferases and β1,2 xylosyltransferase to add their respective (plant-specific) sugar residues. The deficiency of GnT I activity in the *cgl* mutant leads to a vast reduction in complex and hybrid *N*-glycans on the plant (endogenous) proteins (von Schaewen *et al*., 1993), although the effectiveness of this strategy for manipulating the *N*-glycan status of recombinant proteins appears to be protein specific (Kermode, 2012) and is considerably more effective for recombinant human IDUA (this study) than for recombinant human glucocerebrosidase (He *et al*., 2012a).

Earlier reports of the *cgl* mutant showed an absence of complex *N*-glycans on endogenous glycoproteins of plant leaf extracts or callus extracts, with the predominant type of *N*-glycan being Man5 ($Man_5GlcNAc_2$) (von Schaewen *et al*., 1993; Strasser *et al*., 2004, 2005). Later, Frank *et al*. (2008) demonstrated the conditional nature of this particular *cgl* mutant (*cgl1* C5), for which a point mutation leading to a destabilized GnT I enzyme accounts for the phenotype (Strasser *et al*., 2005). For example, tunicamycin treatment relieves the 'folding block' for the mutant GnT I protein, permitting its transit to the Golgi complex. More recently, our results of the *N*-glycan profiles for *cgl1 C5* mutant-derived recombinant lysosomal enzymes (this study and He *et al*., 2012a,b) suggest that

there is a partial restoration of GnT I activity during seed development in this particular mutant. Despite this, we show here that *cgl*-derived IDUA contained predominantly high-mannose-type *N*-glycans (94.5%), with a Man5 structure accounting for 50.6%. The small amount of recombinant IDUA containing complex/hybrid *N*-glycans (5.5%) was efficiently removed by an additional purification step—an affinity chromatography using conjugated antihorseradish peroxidase (HRP) antibodies. Interestingly, this led to a single dominant *N*-glycan of IDUA (Man5: 93.8%) without detectable plant-specific complex/hybrid *N*-glycans (i.e. those containing xylose/fucose residues). Thus, from the point of view of controlling *N*-glycan maturation of IDUA, the strategy of hosting expression in the *cgl1* seed background is very efficient, and the enzyme is stable as a secreted protein.

Incomplete digestion of *cgl*-IDUA by PNGase A allowed the structural analysis of an *N*-glycan at one consensus *N*-glycosylation site (Asn 372); this site possessed a high-mannose-type *N*-glycan (Man 6–9). The high-mannose-type *N*-glycan at Asn 372 is consistently generated on recombinant IDUA regardless of the host production system; it occurs in both Arabidopsis seeds (this study and He *et al*., 2012b for IDUA-KDEL) and in mammalian (CHO) cell cultures (Zhao *et al*., 1997). This suggests that within the folded mature IDUA protein, the *N*-glycan at this site is protected from 'maturation' by its inaccessibility to the modifying enzymes of the Golgi complex in both plant and mammalian host cells (i.e. the α1,3 fucosyltransferases and β1,2 xylosyltransferase in the former, and the β1,4 galactosyltransferase and sialyltransferase enzymes in the latter).

2.4.3. *Cgl*-IDUA can be processed *in vitro* to generate the M6P recognition marker required for lysosomal targeting in human cells

Another challenge associated with plant-based systems for production of LSD enzyme therapeutics is the lack of the endogenous machinery in plant cells required to generate the M6P recognition marker; this tag is ultimately of importance for efficacy of most LSD enzyme therapeutics as their uptake and subsequent lysosomal targeting in human cells occur largely via M6P-receptor-mediated transport systems. The formation of the M6P recognition marker in mammalian cells is catalysed by the sequential action

of two enzymes: GlcNAc-1-phosphotransferase and the 'uncovering' enzyme. Of the two mammalian/human enzymes involved in creating the M6P tag, the first—the GlcNAc-1-phosphotransferase—is essential for lysosomal trafficking of most lysosomal enzymes as mutations in the gene can cause mucolipidosis II, mucolipidosis IIIA or mucolipidosis IIIC (Kudo *et al*., 2006). No lethal disease has been directly linked with a deficiency of the uncovering enzyme perhaps because the cation-independent M6P receptors can also recognize the disaccharide Man-P-GlcNAc (Olson *et al*., 2010), although a link between mutations in the gene encoding the uncovering enzyme and persistent stuttering has been reported (Lee *et al*., 2011). CHO cell–produced IDUA typically possesses two M6P tags (at Asn336 and Asn451), which are primarily of the biphosphorylated variety: $P_2Man_7GlcNAc_2$ (Zhao *et al*., 1997).

In the present study, we tested the efficacy of an *in vitro* process to achieve the addition of M6P on *cgl*-IDUA. Kinetic parameters of the soluble GlcNAc-1-phosphotransferase with *cgl*-IDUA as a substrate were determined (K_m of 4.3 μm and a k_{cat} of 0.10 min^{-1}). These values are comparable with those of CHO-IDUA treated with alkaline phosphatase (He *et al*., 2012c). The *cgl*-IDUA was amenable to the addition of GlcNAc-P by the soluble recombinant phosphotransferase; through monitoring of the release of [^{3}H]-GlcNAc, we further showed that the covering GlcNAc of the phosphorylated IDUA was readily removed by the uncovering enzyme to expose the M6P tag.

Despite this advance in the downstream processing of a plant-made recombinant lysosomal enzyme, the efficiency of the first step of the *in vitro* process (that mediated by the soluble GlcNAc-1-phosphotransferase) clearly needs to be improved. This could be achieved either by modifying the *in vitro* process itself, or, more readily, by modifying the characteristics of the plant recombinant lysosomal enzyme, such as optimizing the number of mannose residues that represent the predominant species of high-mannose-type *N*-glycans. The phosphotransferases of rat liver and simple eukaryotes (*Acanthamoeba castellani*, and *Dictyostelium discoideum*) seem to require *N*-glycans containing Man6 and higher, partly because they contain the required α1,2-linked mannose residues (Couso *et al*., 1986; Ketcham and Kornfeld, 1992). No phosphorylation of *N*-oligosaccharides comprised of Man5 is detected, at least on the

target hydrolase that was studied – uteroferrin (Couso *et al*., 1986; Ketcham and Kornfeld, 1992). The k_{cat} value of the phosphotransferase for the plant recombinant IDUA as substrate is about 18 times less than that for the mammalian lysosomal protease cathepsin D (Qian *et al*., 2010). The lower k_{cat} value may suggest that, as compared to cathepsin D, the *N*-glycans on the *cgl*-IDUA are less efficiently phosphorylated by the soluble phosphotransferase. Yet the K_m value suggests that the phosphotransferase binds very effectively to the plant IDUA, and part of the efficiency of the phosphotransferase is indeed reliant upon its affinity for the target hydrolase (Qian *et al*., 2010). The extent of phosphorylation of a glycoprotein by the phosphotransferase is influenced by the position of the *N*-glycans relative to the binding site for the phosphotransferase; this modifying enzyme functions best on target lysosomal hydrolases with Man6–Man8 *N*-glycans (Varki and Kornfeld, 1980). This factor may indeed have influenced the extent of GlcNAc-P transfer in the present case as the majority of the *N*-glycans of *cgl*-IDUA are oligomannosidic containing Man5 (Man5 structures account for 50.6%), although approximately 26% are Man 6–8. Perhaps also notable, despite the *N*-glycan at Asn372 of *cgl*-IDUA being of favourable mannose number, this site on IDUA appears to be protected from accessibility to the phosphotransferase, at least *in situ* (Zhao *et al*., 1997).

In summary, we have addressed some of the technical challenges of plant-based production of recombinant enzymes for treatment of lysosomal storage diseases. We are now conducting work to improve the efficiency of the *in vitro* phosphorylation of recombinant IDUA in part through the manipulation of the length of the high-mannose-type *N*-glycans on this human enzyme and by other means.

2.5. Experimental Procedures

2.5.1. Protein extraction and α-L-iduronidase activity assays on transgenic *cgl* lines

Procedures for extraction of total soluble protein from seeds and analyses of α-L-iduronidase (IDUA) activities were as described in the study of Downing *et al*. (2006). Protein amounts were determined using the Bio-Rad DC protein assay kit (Bio-Rad

Laboratories, Mississauga, Ontario, Canada) using bovine serum albumin as the standard. The IDUA activities of *cgl* seed extracts were determined at 37 °C and pH 4.5 using 2 mm 4-methylumbelliferyl-α-L-iduronide (4-MUI) as the substrate. Reactions were performed in a total volume of 15 µL, in 0.1 M dimethylglutarate buffer, pH 4.5, 2 mm sodium metabisulfite and 0.7% bovine serum albumin. Reactions were stopped by the addition of 0.1 m glycine buffer, pH 10.7. Fluorescence of the reaction product, 4-methylumbelliferone (4-MU), was determined (λ_{ex} = 365 nm, λ_{em} = 460 nm). The activities were expressed as units/mg TSP, where one unit is defined as 1 nmol 4-MU/min.

2.5.2. Affinity purification of *cgl*-derived α-L-iduronidase

Affinity purification of *cgl*-IDUA using concanavalin A-Sepharose (ConA-Sepharose) chromatography and affinity chromatography with a monoclonal antibody against IDUA was carried out essentially as described previously (He *et al*., 2012c).

2.5.3. *N*-glycan profiles of *cgl*-derived IDUA

Purified IDUA (approximately 4 µg) was resolved by 10% SDS-PAGE, and the IDUA protein bands recovered from the gel. *N*-glycans were released from tryptic peptides obtained after in-gel digestion, and graphitized carbon liquid chromatography MS/MS (carbon LC MS/MS) was used for *N*-glycan analysis as outlined in He *et al*. (2012c).

2.5.4. Anti-horseradish peroxidase affinity chromatography for removal of small amounts of α-L-iduronidase containing plant complex/hybrid *N*-glycans

Approximately 5% of the IDUA derived from *cgl* seeds contained matured (complex or hybrid) *N*-glycans, that is, *N*-glycans with xylose and/or fucose. This small fraction of the IDUA was removed by passing the purified IDUA (a single band on SDS-PAGE) through an anti-horseradish peroxidase affinity column (recycling overnight at 4 °C) according to the method described by He *et al*. (2012c). The column specifically binds to xylose and/or fucose residues (Wilson *et al*., 1998), allowing for the removal of

any IDUA containing these sugars from the sample. The resultant IDUA was resolved on 10% SDS-PAGE, and its *N*-glycan composition was determined by MALDI-TOF MS analyses as described in He *et al*. (2012b).

2.5.5. Determination of kinetics of *cgl*-derived IDUA

The Michaelis–Menten kinetic parameters of 4-MUI cleavage by *cgl*-IDUA were determined using IDUA enzyme purified by ConA-Sepharose and anti-IDUA affinity chromatography. The fluorometric activity assay outlined above was performed using a total volume of 100 µL. The substrate (4-MUI) was used at concentrations of 1 to 800 µm to ensure the coverage of a sufficient range encompassing the K_m. Reactions were started upon the addition of enzyme (0.035 µg per assay) and allowed to proceed for 10 min, over which time it had been determined that the rate of the reaction remained linear. Reactions were stopped by the addition of 0.1 m glycine buffer, pH 10.7. The amount of 4-MU produced at each substrate concentration was determined by measuring the fluorescence and converting the value into a concentration using a standard curve that was prepared using 4-MU concentrations between 0.1 and 125 µm. Rates of catalysis were determined by dividing the reaction time and concentration of the enzyme and fit to a Michaelis–Menten curve using GRAFIT. All measurements were made in triplicate.

2.5.6. Modification of *cgl*-IDUA by a recombinant soluble $\alpha_2\beta_2$ GlcNAc-1-phosphotransferase

For *in vitro* phosphorylation of *cgl*-IDUA that had been purified by sequential ConA-Sepharose and anti-IDUA affinity chromatography, the recombinant $\alpha_2\beta_2$ GlcNAc-1-phosphotransferase (1.5 µg) was added to the reaction mixtures containing various concentrations of *cgl*-derived IDUA in 50 mm Tris–HCl, pH 7.4, 10 mm $MgCl_2$, 10 mm $MnCl_2$, 75 µm UDP-[^{3}H]GlcNAc (1 µCi) and 2 mg/mL bovine serum albumin in a final volume of 50 µL. The assay was carried out as described by Qian *et al*. (2010). Apparent K_m and k_{cat} values were generated from double-reciprocal plots using a least-square approximation for the best fit line. The values were the average of three separate determinations.

2.5.7. Removal of GlcNAc by the uncovering enzyme

The uncovering enzyme activity assay was performed according to Mullis and Ketcham (1992) and Kornfeld *et al*. (1998). Twenty µmoles of IDUA was phosphorylated by the soluble phosphotransferase overnight at 37°C under the conditions described above. Phosphorylated *cgl*-IDUA was passed through Sephadex-G-25, and the void volume was collected. The fraction was adjusted to 600 µL with a buffer containing Tris–Maleate pH 6.7, Triton X-100 0.5% and 500 ng uncovering enzyme. The reaction mix was incubated at 37 °C, and at each time point, 30 µL was taken for analysis. After stopping the reaction, the mixture was passed through a 2-mL CM-cellulose chromatography column pre-equilibrated with 50 mm dimethylglutarate (pH 6.0) and eluted with 2 mL of the same buffer. The concentration of [^{3}H]-GlcNAc of the eluted fraction was determined by liquid scintillation counting.

2.6. Acknowledgements

We thank Dr. John Hopwood for providing the anti-α-L-iduronidase antibodies and Dr. Stuart Kornfeld for providing the recombinant GlcNAc-1-phosphotransferase ($\alpha_2\beta_2$) and uncovering enzyme. This work was funded by the Michael Smith Foundation for Health Research (ARK), the Australian Research Council (MvI), and a Wellcome Trust Research Career Development Fellow Award (TMG).

2.7. References

Bao, M., Booth, J.L., Elmendorf, B.J. and Canfield, W.M. (1996) Bovine UDP-N-acetylglucosamine:lysosomal-enzyme *N*-acetylglucosamine-1-phosphotransferase. I. Purification and subunit structure. *J. Biol. Chem.* 271, 31437–31445.

Clarke, L.A. (2008) The mucopolysaccharidoses: a success of molecular medicine. Expert. Rev. Molec. Med. 10, e1.

Couso, R., Lang, L., Roberts, R.M. and Kornfeld, S. (1986) Phosphorylation of the oligosaccharide of uteroferrin by UDP-GlcNAc:glycoprotein *N*-acetylglucosamine-1-phosphotransferases from rat liver, *Acanthamoeba castellani*, and *Dictyostelium discoideum* requires alpha 1, 2-linked mannose residues. *J. Biol. Chem.* 261, 6326–6331.

De Jaeger, G., Scheffer, S., Jacobs, A., Zambre, M., Zobell, O., Goossens, A., Depicker, A. and Angenon, G. (2002) Boosting heterologous protein production in transgenic dicotyledonous seeds using Phaseolus vulgaris regulatory sequences. *Nat. Biotechnol.* 20, 1265–1268.

Do, H., Lee, W.S., Ghosh, P., Hollowell, T., Canfeld, W. and Kornfeld, S. (2002) Human mannose 6-phosphate-uncovering enzyme is synthesized as a proenzyme that is activated by the endoprotease furin. *J. Biol. Chem.* 277, 29737–29744.

Downing, W.L., Galpin, J.D., Clemens, S., Lauzon, S.M., Samuels, A.L., Pidkowich, M.S., Clarke, L.A. and Kermode, A.R. (2006) Synthesis of enzymatically active human α-L-iduronidase in *Arabidopsis cgl* (complex glycan-deficient) seeds. *Plant Biotechnol. J.* 4, 169–181.

Downing, W.L., Hu, X. and Kermode, A.R. (2007) Post-transcriptional factors are important for high-level expression of the human α-L-iduronidase gene in *Arabidopsis cgl* (complex glycan-deficient) seeds. *Plant Sci.* 172, 327–334.

Frank, J., Kaulfürst-Soboll, H., Rips, S., Koiwa, H. and von Schaewen, A. (2008) Comparative analyses of *Arabidopsis* complex glycan 1 mutants and genetic interaction with staurosporin and temperature sensitive3a. *Plant Physiol.* 148, 1354–1367.

Galpin, J.D., Clemens, S. and Kermode, A.R. (2010) The carboxy-terminal ER-retention motif, SEKDEL, influences the *N*-linked glycosylation of recombinant human α-L-iduronidase but has little effect on enzyme activity in seeds of *Brassica napus* and *Nicotiana tabacum*. *Plant Sci.* 178, 440–447.

Gary-Bobo, M., Nirdé, P., Jeanjean, A., Morère, A. and Garcia, M. (2007) Mannose 6-phosphate receptor targeting and its applications in human diseases. *Curr. Medicinal Chem.* 14, 2945–2953.

van Gelder, C.M., Vollebregt, A.A., Plug, I., van der Ploeg, A.T. and Reuser, A.J. (2012) Treatment options for lysosomal storage disorders: developing insights. *Expert Opin. Pharmacother.* 13, 2281–2299.

Gomord, V., Fitchette, A.C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D. and Faye, L. (2010) Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J.* 8, 564–587.

He, X., Galpin, J.D., Tropak, M.B., Mahuran, D., Haselhorst, T., von Itzstein, M., Kolarich, D., Packer, N.H., Miao, Y.S., Jiang, L., Grabowski, G.A., Clarke, L.A. and Kermode, A.R. (2012a) Production of active human glucocerebrosidase in seeds of *Arabidopsis thaliana* complex-glycan-deficient (*cgl*) plants. *Glycobiology*, 22, 492–503.

He, X., Haselhorst, T., von Itzstein, M., Kolarich, D., Packer, N.H. and Kermode, A.R. (2012b) Influence of an ER-retention signal on the *N*-glycosylation of recombinant human α-L-iduronidase generated in seeds of *Arabidopsis*. *Plant Mol. Biol.* 79, 157–169.

He, X., Haselhorst, T., von Itzstein, M., Kolarich, D., Packer, N.H., Gloster, T., Vocadlo, D., Clarke, L.A., Qian, Y. and Kermode, A.R. (2012c) Production of α-L-iduronidase in maize for the potential treatment of a lysosomal storage disease. *Nat. Commun.* 3, 1062.

He X, Pierce O, Haselhorst T, von Itzstein M, Kolarich D, Packer NH, Gloster TM, Vocadlo DJ, Qian Y, Brooks D, Kermode AR. (2013). Characterization and downstream mannose phosphorylation of human recombinant α-L-iduronidase produced in *Arabidopsis* complex glycan-deficient (cgl) seeds. *Plant Biotechnol J.* **11(9):**1034-43.

Kakkis, E. (2005) α-L-Iduronidase: the development of aldurazyme (Larconidase). In Directory of Therapeutic Enzymes (McGrath, B.M. and Walsh, G., eds), pp. 239–260. New York: CRC Press, Taylor & Francis Group.

Kermode, A.R. (1996) Mechanisms of intracellular protein transport and targeting. *Crit. Rev. Plant Sci.* 15, 285–423.

Kermode, A.R. (2012) Seed expression systems for molecular farming. In Molecular Farming in Plants: Recent Advances and Future Prospects (Wang, A. and Ma, S., eds), pp. 89–123. New York: Springer Science + Business Media.

Ketcham, C.M. and Kornfeld, S. (1992) Characterization of UDP-N-acetylglucosamine:glycoprotein *N*-acetylglucosamine-1-phosphotransferase from *Acanthamoeba castellanii. J. Biol. Chem.* 267, 11654–11659.

Kornfeld, R., Bao, M., Brewer, K., Noll, C. and Canfield, W.M. (1998) Purification and multimeric structure of bovine *N*-acetylglucosamine-1-phosphodiester α-N-acetylglucosominidase*. J. Biol. Chem.* 273, 23203–23210.

Kudo, M. and Canfield, W.M. (2006) Structural requirements for efficient processing and activation of recombinant human UDP-N-acetylglucosamine:lysosomal-enzyme-N-acetylglucosamine-1-phosphotransferase. *J. Biol. Chem.* 281, 11761–11768.

Kudo, M., Bao, M., D'Souza, A., Ying, F., Pan, H., Roe, B.A. and Canfield, W.M. (2005) The α- and β- subunits of the human UDP-N-acetylglucosamine:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase are encoded by a single cDNA. *J. Biol. Chem.* 280, 36141–36149.

Kudo, M., Brem, M. and Canfield, W.M. (2006) Mucolipidosis II (I-Cell Disease) and mucolipidosis IIIA (Classical Pseudo-Hurler Polydystrophy) are caused by mutations in the GlcNAc-phosphotransferase α/β-subunits precursor gene. *Am. J. Hum. Genet.* 78, 451–463.

Lee, W.S., Payne, B.J., Gelfman, C.M., Vogel, P. and Kornfeld, S. (2007) Murine UDP-GlcNAc:lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase lacking the gamma-subunit retains substantial activity toward acid hydrolases. *J. Biol. Chem.* 282, 27198–27203.

Lee, W.S., Kang, C.S., Drayna, D. and Kornfeld, S. (2011) Analysis of mannose 6-phosphate uncovering enzyme mutations associated with persistent stuttering. *J. Biol. Chem.* 286, 39786–39793.

Loos, A., Droogenbroeck, B.V., Hillmer, S., Grass, J., Kunert, R., Cao, J., Robinson, D.G., Depicker, A. and Steinkellner, H. (2011) Production of monoclonal antibodies with a controlled *N*-glycosylation pattern in seeds of *Arabidopsis thaliana. Plant Biotechnol. J.* 9, 179–192.

Maxmen, A. (2012) Drug-making plant blooms. *Nature*, 485, 160.

Muenzer, J., Wraith, J.E. and Clarke, L.A. (2009) International consensus panel on management and treatment of mucopolysaccharidosis I. Mucopolysaccharidosis I: management and treatment guidelines. *Pediatrics*, 123, 19–29.

Mullis, K.G. and Ketcham, C.M. (1992) The synthesis of substrates and two assays for the detection of *N*-Acetylglucosamine-1-phosphodiester α-N-acetylglucosaminidase (uncovering enzyme). *Anal. Biochem.* 205, 200–207.

Neufeld, E.F. (2011) From serendipity to therapy. *Annu. Rev. Biochem.* 80, 1–15.

Olson, L.J., Peterson, F.C., Castonguay, A., Bohnsack, R.N., Kudo, M., Gotschall, R.R., Canfield, W.M., Vollman, B.F. and Dahms, N.M. (2010) Structural basis for recognition of phosphodiester-containing lysosomal enzymes by the cation-independent mannose 6-phosphate receptor. *Proc. Natl Acad. Sci. USA* 107, 12493–12498.

Pastores, G.M., Arn, P., Beck, M., Clarke, J.T., Guffon, N., Kaplan, P., Muenzer, J., Norato, D.Y., Shapiro, E., Thomas, J., Viskochil, D. and Wraith, J.E. (2007) The MPS I registry: design, methodology, and early findings of a global disease registry for monitoring patients with Mucopolysaccharidosis Type I. *Mol. Genet. Metab.* 91, 37–47.

Qian, Y., Lee, I., Lee, W.S., Qian, M., Kudo, M., Canfield, W.M., Lobel, P. and Kornfeld, S. (2010) Functions of the α, β, and γ subunits of UDP-GlcNAc:Lysosomal enzyme *N*-acetylglucosamine-1-phosphotransferase. *J. Biol. Chem.* 285, 3360–3370.

von Schaewen, A., Sturm, A., O'Neill, J. and Chrispeels, M.J. (1993) Isolation of a mutant *Arabidopsis* plant that lacks *N*-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-mediated complex *N*-linked glycans. *Plant Physiol.* 102, 1109–1118.

Shaaltiel, Y., Bartfeld, D., Hashmueli, S., Baum, G., Brill-Almon, E., Galili, G., Dym, O., Boldin-Adamsky, S.A., Silman, I., Sussman, J.L., Futerman, A.H. and Aviezer, D. (2007) Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnol. J.* 5, 579–590.

Strasser, R., Altmann, F., Mach, L., Glössl, J., Stadlmann, J. and Steinkellner, H. (2004) Generation of Arabidopsis thaliana plants with complex *N*-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. *FEBS Lett.* 561, 132–136.

Strasser, R., Stadlmann, J., Svoboda, B., Altmann, F., Glössl, J. and Mach, L. (2005) Molecular basis of *N*-acetylglucosaminyltransferase I deficiency in *Arabidopsis* thaliana plants lacking complex *N*-glycans. *Biochem. J.* 387, 385–391.

Varki, A. and Kornfeld, S. (1980) Structural studies of phosphorylated high mannose-type oligosaccharides. *J. Biol. Chem.* 255, 10847–10858.

Wenderoth, I. and von Schaewen, A. (2000) Isolation and characterization of plant *N*-acetyl glucosaminyltransferase I (GntI) cDNA sequences. Functional analyses in the Arabidopsis cgl mutant and in antisense plants. *Plant Physiol.* 123, 1097–1108.

Wilson, I.B.H., Harthill, J.E., Mullin, N., Ashford, D. and Altmann, F. (1998) Core α 1,3-fucose is a key part of the epitope recognized by antibodies reacting against plant *N*-linked oligosaccharides and is present in a wide variety of plant extracts. *Glycobiol.* 8, 651–661.

Zhao, K.W., Faull, K.F., Kakkis, E.D. and Neufeld, E.F. (1997) Carbohydrate structures of recombinant human α-L-iduronidase secreted by Chinese hamster ovary cells. *J. Biol. Chem.* 272, 22758–22765.

2.8. Tables and Figures

Table 2.1. Accumulation of IDUA in transgenic Arabidopsis seed lines

Line	T2		T3 (±SD)		T4 (±SD)	
	%TSP	(µg/mg)	%TSP	(µg/mg)	%TSP	(µg/mg)
A4.7	1.7	2.7	7.2 ± 0.6	9.8 ± 0.6	5.7 ± 0.4	8.0 ± 0.8
A5.5	1.8	3.0	11.2 ± 2.7	16.3 ± 2.1	2.5 ± 0.1	4.0 ± 0.5
A6.3	0.94	1.6	2.60 ± .02	4.3 ± 0.4	1.6 ± 0.1	2.3 ± 0.2

Table 2.2. N-glycans identified in *cgl*-IDUA

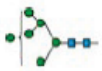
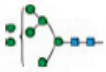
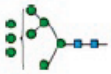
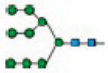
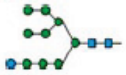
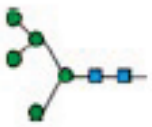
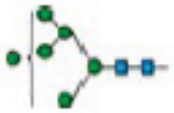
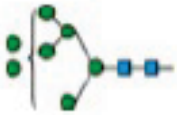
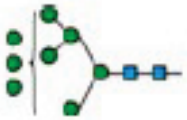
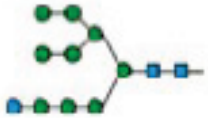
	Abbreviation	Structures	Rel. amount (%)	Hex	HexNAc	Fuc	Xyl
Complex and hybrid type structures	MMX		1.3	3	2		1
	MMXF		3.8	3	2	1	1
	Man5F		0.4	5	2	1	
Oligomannosidic structures	Man4		3.7	4	2		
	Man5		50.6	5	2		
	Man6		3.6	6	2		
	Man7		9.0	7	2		
	Man8		13.8	8	2		
	Man9		11.0	9	2		
	Man9 + Hex		2.8	10	2		

Table 2.3. N-glycans identified in *cgl*-IDUA after anti-HRP affinity chromatography

N-glycan type High-mannose type	Det. m/z $[M + Na]^+$	Structures	IDUA Rel. int. [%]
Man5	1257.507		93.83
Man6	1419.556		1.59
Man7	1581.598		2.57
Man8	1743.722		1.65
Man9	1905.742		0.36

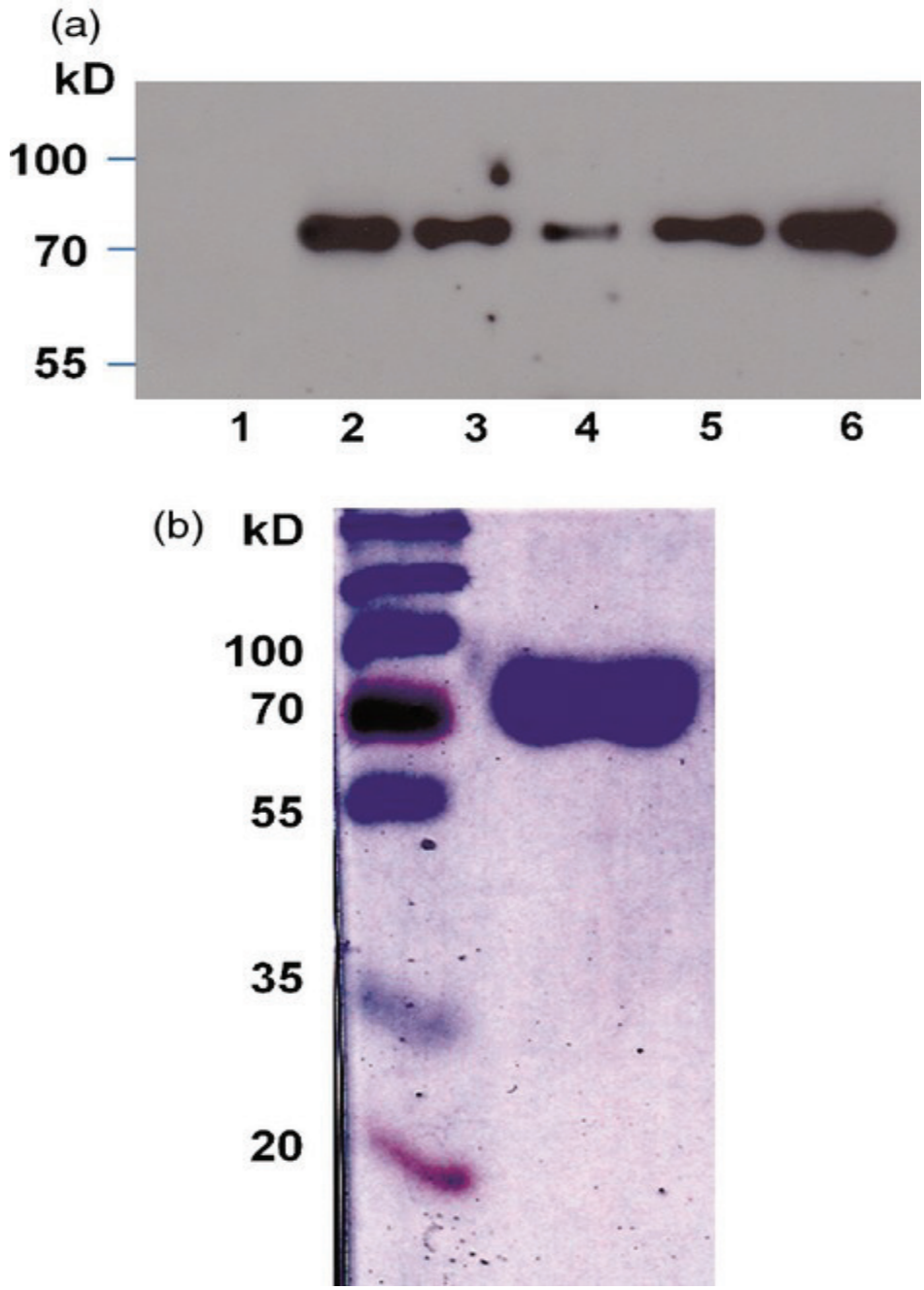


Figure 2.1. α-L-Iduronidase (IDUA) produced in Arabidopsis *cgl* seeds.
(a) Western blot analysis of IDUA produced in T3 seeds of 3 high-expressing lines. Lanes 1–4: 0.4 µg of total soluble protein was loaded from the crude extracts of *cgl* untransformed seeds (lane 1) and from seeds of transformed lines A5.5, A4.7 and A6.3, respectively (lanes 2–4). Lanes 5–6 show 20 ng and 50 ng of purified *cgl*-IDUA (from line A4.7), respectively. (b) Five micrograms of affinity-purified *cgl*-IDUA (line A4.7) was loaded on a 10% gel, and after SDS-PAGE, the gel was stained with Coomassie Brilliant Blue and destained. Molecular weights of the prestained protein markers are indicated on the left.

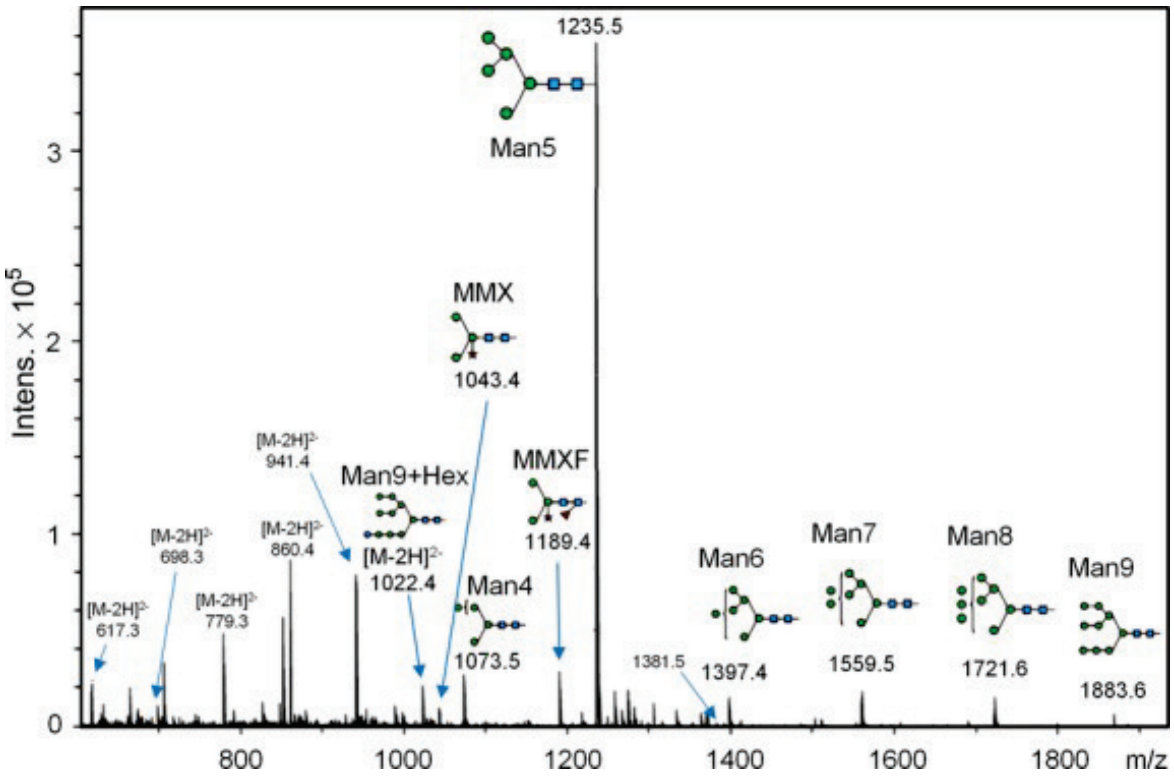


Figure 2.2. *N*-glycans of *cgl*-α-L-iduronidase (IDUA).
Summed mass spectra of *N*-glycans from *cgl*-IDUA. Signals were identified as $[M–H]^-$ signals unless stated otherwise. Oligomannosidic structures comprised more than 90% of the structures identified (see also Table 2).

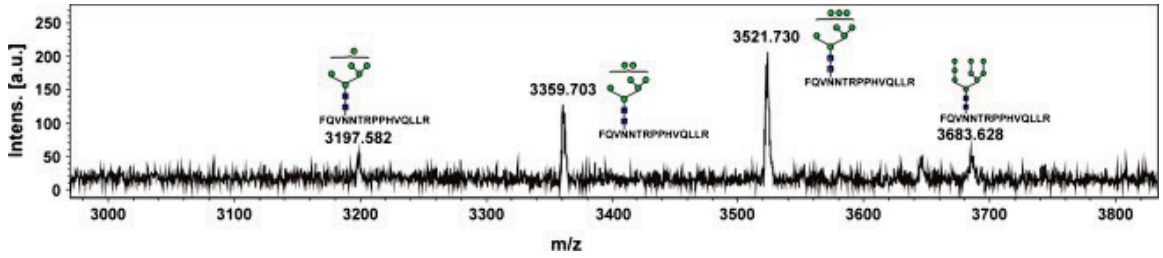


Figure 2.3. High-mannose type *N*-glycan at site Asn 372 on recombinant *cgl*-IDUA protein (glycopeptide 369–383).

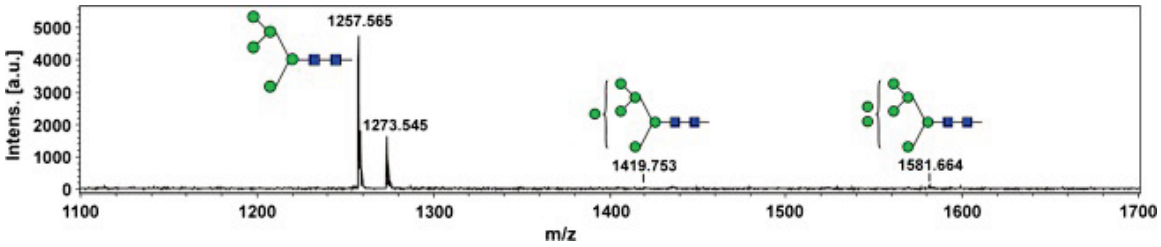
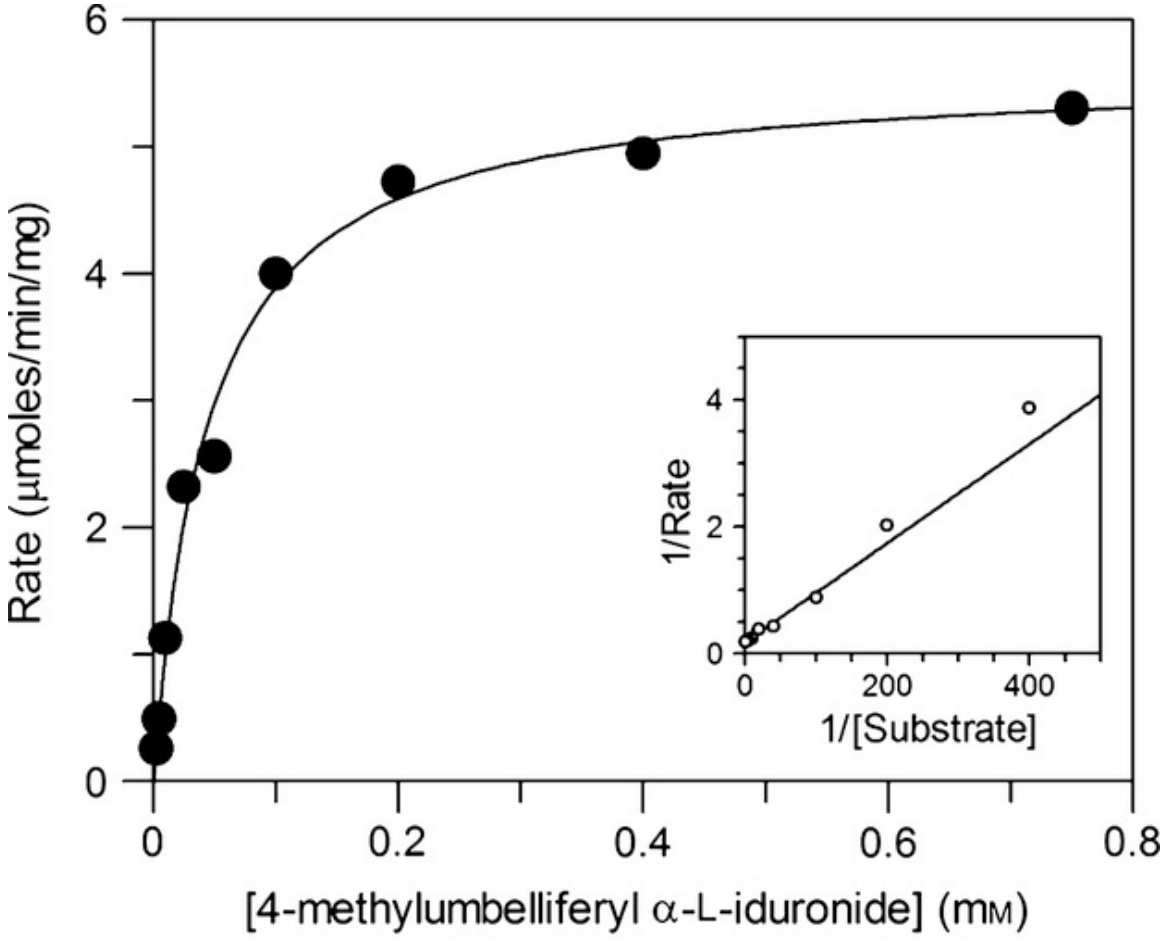


Figure 2.4. *N*-glycans on *cgl*-IDUA after anti-HRP affinity chromatography identified by MALDI-TOF MS.

Glycan cartoons are according to the recommendations of the consortium of functional glycomics (www.functionalglycomics.org).



Parameter	Value	Std. Error
Vmax	5.5959	0.1765
Km	0.0438	0.0053

Figure 2.5. Michaelis-Menten plot for *cgl*-IDUA
Assays were performed in triplicate. The inset on the right is the Lineweaver–Burk plot.

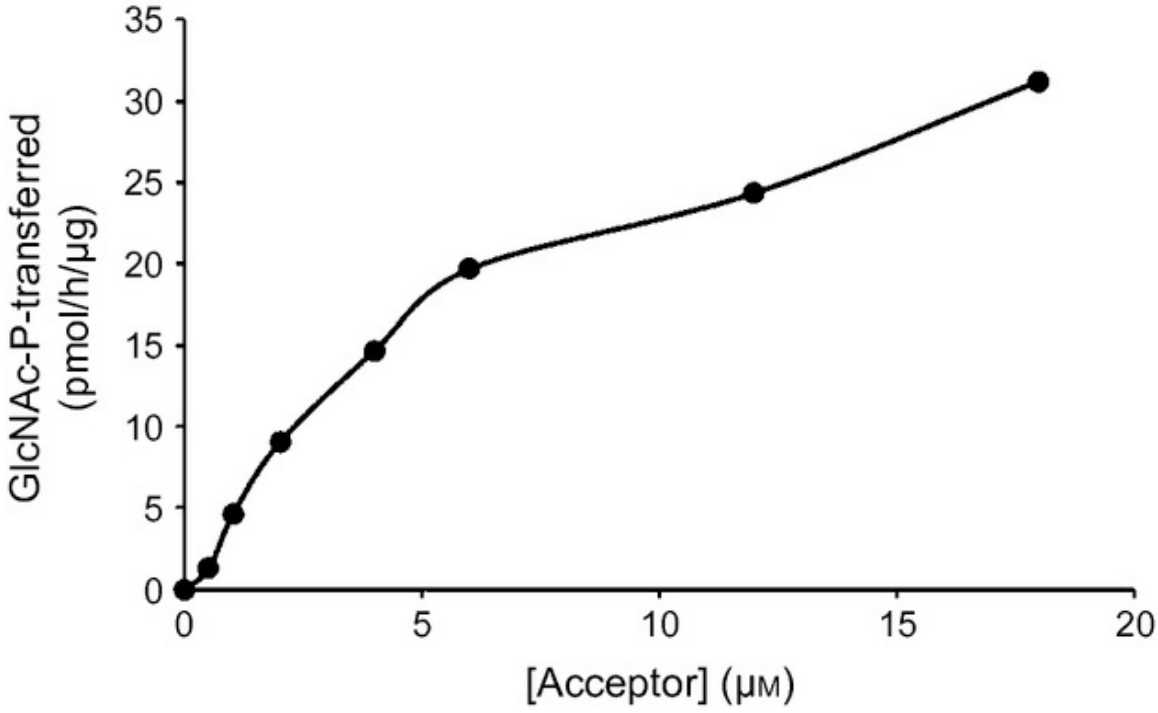


Figure 2.6. Phosphorylation of *cgl*-IDUA by the GlcNAc-1-phosphotransferase ($\alpha_2\beta_2$).

The activity of the GlcNAc-1-phosphotransferase was expressed as pmoles of [^{3}H]GlcNAc-P transferred per h/µg of the GlcNAc-1-phosphotransferase. The graph is based on data from one typical experiment.

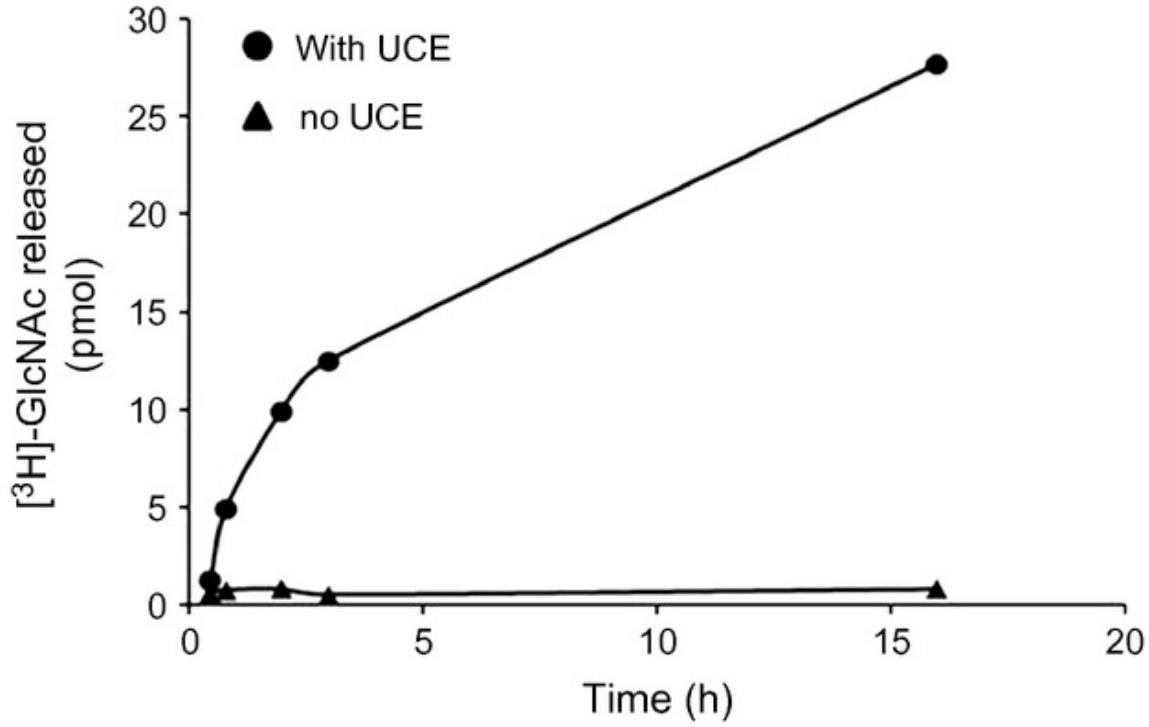


Figure 2.7. Release of [^{3}H] GlcNAc from [^{3}H] GlcNAc-P ('phosphorylated')-*cgl*-IDUA by the soluble uncovering enzyme.

[3H] GlcNAc-P-cgl-IDUA was incubated with or without 500 ng of soluble recombinant uncovering enzyme. At the times indicated, a 30-µL aliquot was removed, and the amount of free [3H] GlcNAc was determined. Each point is the average of two determinations.

Chapter 3. Purification, characterization, and phosphorylation of α-L-iduronidase produced in the *Arabidopsis thaliana* Golgi Mannosidase I (GM1) mutant.

3.1. Abstract

Mucopolysaccharidosis I is a heritable lysosomal storage disease that is caused by a deficiency of the lysosomal enzyme, α-L-iduronidase (IDUA). Enzyme replacement therapy that utilizes purified, recombinant IDUA is the most successful treatment for the disease. Previously, IDUA was produced in seeds of the *Arabidopsis cgl* mutant (which is deficient in the GlcNAc Transferase I). A minor fraction of this *cgl*-IDUA (~6%) possesses matured *N*-glycans, and thus potentially immunogenic xylose and fucose residues. Here, we show that IDUA produced in the seed tissues of the *Arabidopsis* Golgi Mannosidase I mutant (GM1-IDUA) line harbors a more favorable *N*-glycan profile compared to IDUA produced in the *cgl* line. All of the *N*-glycans on GM1-IDUA were of the high-mannose form (i.e. xylose and fucose are not detected), with 92% of the *N*-glycans bearing six to eight mannose rings. The recombinant IDUA of GM1 seeds was produced at 1.5% of total soluble protein, and the purified protein exhibited kinetics identical to those of *cgl*-IDUA. The GM1-IDUA was also amenable to phosphorylation *in vitro* by a soluble form of the UDP-GlcNAc:lysosomal enzyme GlcNAc-phosphotransferase at a rate comparable to IDUA produced in Chinese Hamster Ovary cells and in the *cgl* mutant line. The results of this study show that the *Arabidopsis* Golgi Mannosidase I mutant can serve as a suitable host for the production of recombinant lysosomal enzymes for the treatment of lysosomal storage diseases.

3.2. Introduction

Lysosomal Storage Disesases (LSDs) comprise a class of over 70 heritable metabolic disorders. The molecular basis for these conditions usually stems from a defect in a particular lysosomal hydrolase which results in the accumulation of undegraded or partially degraded substrate(s). Mucopolysaccharidosis I (MPS I), an LSD caused by the deficiency in the lysosomal enzyme α-L-iduronidase (IDUA), leads to pathology as a result of both primary and secondary consequences of the abnormal lysosomal accumulation of glycosaminoglycans (Bach et al., 1972; see Chapter 2). This disrupts the proper functioning of multiple organ systems including the liver, spleen, and heart, and of the skeletal and central nervous systems. For patients with no residual α-L-iduronidase activity, the disease is often fatal in the first decade of life. Mucopolysaccharidosis I is one of the few LSDs that can be treated by enzyme replacement therapy (ERT), which is the parenteral administration of the purified, recombinant enzyme *via* weekly infusions (Kakkis et al., 2001). This therapy exploits mechanisms of extracellular receptors present on patient cells that mediate the internalization of the circulating recombinant protein through endocytosis (see Chapter 2). *Aldurazyme*^{TM}, a United States Food and Drug Administration (FDA)-approved therapeutic for MPS I treatment, is a recombinant form of IDUA produced in Chinese Hamster Ovary (CHO) cells that is effective in the treatment of MPS I (Jameson et al., 2013). Treatment regimens can cost over $450,000 per year, stimulating the demand for a more inexpensive source of the enzyme.

Over the past two decades, there has been an increasing focus on the use of plants as expression platforms for recombinant, therapeutic proteins. Plants offer several key advantages that are attractive for this purpose, including the capacity to quickly produce large amounts of protein, and the amenability towards extensive engineering of the *N*-glycan maturation pathway for the production of proteins with a homogenous and therapeutically acceptable *N*-glycan profile. Through the use of seed-specific regulatory sequences, the recombinant protein can be directed to accumulate in the seed, thereby allowing the purification of the protein to be decoupled from biomass harvest. In 2012, Elelyso, a recombinant form of glucocerebrosidase produced in carrot cell culture for the treatment of Gaucher Disease, was the first plant-produced

recombinant enzyme for ERT to be approved by the United States FDA (reviewed in Grabowski et al., 2014).

As the majority of therapeutic proteins on the market are glycoproteins, a lingering challenge in the development of plant-based systems for recombinant protein production is the "humanization" of the protein-N-linked oligosaccharides synthesized by the plant host. In particular, and relevant to the production of lysosomal enzymes, the engineering of *N*-glycan status for the purpose of preventing plant-specific *N*-glycan maturation has received the most attention. While the initial steps of *N*-glycosylation and *N*-glycan trimming that occur in the endoplasmic reticulum and *cis*-Golgi are common to plants and animals, the divergence in the later stages of *N*-glycan maturation between these two groups of organisms is sufficient to render proteins produced in plants either therapeutically less effective or potentially immunogenic in human patients. The two key challenges are the elimination of immunogenic β-(1→2)-xylose residues and α-(1→3)-fucose residues linked to the *N*-glycan core, and the addition of the mannose-6-phosphate (M6P) carbohydrate recognition tag, the latter required for receptor-mediated uptake of the recombinant enzyme into the cells of the patient (see Chapter 2).

We utilized the seeds of the *Arabidopsis cgl* mutant as a host for recombinant human IDUA production (He et al, 2013). This mutant is deficient in *N*-acetylglucosamine transferase I, the Golgi-resident glycosyltransferase which is the first enzyme to modify the $Man_5GlcNAc_2$ core after the *N*-glycan has been processed by mannosidases and glucosidases in the ER and *cis*-Golgi. While this form of IDUA was expressed at an appreciable level (5.7% of total soluble protein), *N*-glycan profiling experiments revealed that a minor fraction of the protein (roughly 5% of the purified IDUA) harbored *N*-glycans that contained xylose or α-(1→3)-linked fucose. Additionally, half of the *N*-linked glycans were of the $Man_5GlcNAc_2$ glycoform, which is not optimal for downstream processing of the purified protein to create the M6P tag.

In respect to these concerns, the *Arabidopsis* Golgi Mannosidase I (GM1) mutant is an attractive alternative to the *cgl* mutant line. Golgi mannosidase I is responsible for trimming the terminal mannose residues from $Man_8GlcNAc_2$ *N*-glycans to yield the intermediate $Man_5GlcNAc_2$ glycoform. The reports to date further indicate that xylose

and fucose residues are not added to endogenous or recombinant glycoproteins isolated from this mutant, and the vast majority of *N*-glycans remain in the $Man_{6-8}GlcNAc_2$ glycoform (Kajiura et al., 2010). In addition to the lowered potential of creating an immunogenic product, the glycoforms are predicted to be better substrates for the downstream addition of mannose-6-phosphate to the *N*-glycan. Notably, the phosphotransferase responsible for the first step of M6P elaboration does not accept the $Man_5GlcNAc_2$ glycoform as a substrate.

In this study, we compared IDUA purified from seeds of the *Arabidopsis* GM1 mutant to that from *Arabidopsis cgl* mutant seeds and from CHO cells (Aldurazyme™). We show that the GM1 mutant generates IDUA with a therapeutically advantageous *N*-glycan profile as far as an absence of xylose and fucose are concerned. Despite the increased abundance of Man_{6-8} *N*-glycans over Man_5 *N*-glycans, there was no appreciable increase in the efficiency of the in vitro phosphorylation process. A possible cause was the surprising lack of an *N*-glycan at site Asn336 on the GM1-IDUA, which is normally present on the native human iduronidase and on the recombinant iduronidase of *cgl Arabidopsis* seeds and Chinese hamster ovary cells. For CHO-cell derived IDUA (Aldurazyme™), the two M6P tags for receptor-mediated uptake/lysosome delivery are located at Asn336 and Asn451 (Zhao *et al*., 1997). Surprisingly GM1-iduronidase did not contain an *N*-glycan at Asn336, which is normally present on the native human iduronidase and on the recombinant iduronidase of *cgl Arabidopsis* seeds and Chinese hamster ovary cells.

My contributions to this research were the transformation and selection of the T1 generation of *Arabidopsis thaliana* GM1 transformants, the purification of GM1-IDUA from the *Arabidopsis* seed, the measurement of the enzymatic activity of GM1-IDUA, and the measurement of the rate of phosphorylation of GM1-IDUA by the soluble phosphotransferase. Dr. Xu He performed the line selection of the T2 and T3 generations, and Dr. Hiroyuki Kajiura and Dr. Kazuhito Fujiyama of Osaka University performed the *N*-glycan profiling experiments.

3.3. Results

3.3.1. Seeds of the *Arabidopsis* GM1 mutant express active α-L-iduronidase

Previously, the *Arabidopsis cgl* mutant was used as a transgenic host for the expression of recombinant iduronidase (IDUA). The *Arabidopsis* GM1 mutant was transformed by a binary vector that harbors the IDUA gene under the control of the Arcelin regulatory sequences, which had also been used to transform the *Arabidopsis cgl* mutant. After screening seedlings on kanamycin for several generations, a T3-generation line that expressed IDUA in the mature dry seeds at 1.5% of TSP was identified.

In order to assess the kinetic properties of the recombinant enzyme, IDUA was purified from the seeds of the high-expressing GM1 line using concanavalin A and monoclonal antibody affinity chromatography as described previously (He et al., 2012). The kinetic profile of 4-MUI cleavage by the purified GM1-IDUA was nearly identical to that of *cgl*-IDUA and to the commercial (CHO-cell-derived) enzyme *Aldurazyme*^{TM} (Table 3.3, Fig. 3.3).

3.3.2. GM1-Iduronidase contains a more optimal *N*-glycan profile as compared to *cgl*-iduronidase: higher mannose chain length and no detectable complex *N*-glycan sugars

The IDUA synthesized in the *Arabidopsis cgl* line bears *N*-glycans primarily of the $Man_5GlcNAc_2$ configuration, and a small proportion of the protein possesses complex *N*-glycans. In order to ameliorate the *N*-glycan configuration for downstream phosphorylation by the phosphotransferase, which generally accepts only a high mannose *N*-glycan with six to eight mannose rings as a substrate, the *Arabidopsis* GM1 mutant was chosen as a suitable transgenic host. Seeds of the GM1 mutant generated IDUA with *N*-glycans primarily of the higher mannose residue length, and there was no detectable complex *N*-glycans on the recombinant human enzyme in contrast to the IDUA of *cgl* mutant seeds (Tables 3.5, 3.6).

Two methods were used to assess the *N*-glycan status of GM1-iduronidase. First, digestion by endoglycosidase H (Endo H), which cleaves high-mannose *N*-glycans from the aspargine residue to which they are bound, indicated that none of the *N*-glycan pool had undergone extensive processing past the high-mannose form. To more precisely determine the *N*-glycan status of GM1-iduronidase, liquid chromatography-tandem mass spectroscopy was used to generate both site-specific and total *N*-glycan profiles. Human iduronidase bears six *N*-glycans, two of which are phosphorylated *in vivo* in mammalian cells (Asn336 and Asn451; see Chapter 2). The *N*-glycans of GM1-iduronidase occupied five of these sites; however, Asn336, which normally carries a phosphorylated *N*-glycan when synthesized in human cells, was not detected on GM1-IDUA (Fig 3.5; Table 3.6). The overwhelming majority of the *N*-glycans present on GM1-iduronidase (91.8%) contained six to eight mannose rings, while the remainder of the *N*-glycans contained five or nine mannose rings, or more than seven mannose rings and a terminal glucose ring. Importantly, no *N*-glycans that contained the potentially immunogenic xylose or fucose residues indicative of complex *N*-glycan formation were detected (Table 3.6).

3.3.3. The GM1-iduronidase can be processed *in vitro* by the enzymes of the M6P biosynthetic pathway

In mammalian cells, the addition of phosphate to the 6-OH of a terminal mannose ring occurs in two steps. First, GlcNAc-phosphate is transferred to the *N*-glycan from a donor UDP-GlcNAc molecule, in a reaction catalyzed by the phosphotransferase. Next, the resulting phosphodiester bond is hydrolyzed by the uncovering enzyme to remove the GlcNAc from the *N*-glycan, thereby yielding the M6P tag required for sequestration of the recombinant lysosomal hydrolase by human cells.

The efficiency of the first step, was assessed by an *in vitro* assay that measures the transfer of tritiated GlcNAc-phosphate to the *N*-glycan of purified, recombinant GM1 iduronidase (Fig 3.4; methods noted in Chapter 2). In this assay, the transfer of GlcNAc-phosphate is catalyzed by a soluble form αβ-phosphotransferase that lacks both transmembrane domains. As depicted in Table 3.3, the catalytic efficiency (V_{max}/K_m) of the transfer of GlcNAc-phosphate to GM1-IDUA is comparable of that to CHO-

iduronidase (Aldurazyme™ treated with alkaline phosphatase) and to *cgl-IDUA*, Representative of an improved affinity of the phosphotransferase for the GM1-IDUA, the K_m was an order of magnitude smaller.

3.4. Discussion

3.4.1. The *Arabidopsis* GM1 mutant line produces α-L-iduronidase (IDUA) with a different *N*-glycan profile than that of *cgl*-IDUA.

In this study, we sought to develop a line of *Arabidopsis thaliana* that could express, in the seed tissues, α-L-iduronidase (IDUA) with a more preferable *N*-glycan profile compared to that produced by seeds of the *Arabidopsis cgl* mutant. Seeds of the *cgl* mutant produce active IDUA at a significant percentage of total soluble protein, with the best line providing stable expression across generations and accumulating the protein at 5.7% TSP. However, the dominant *N*-glycan configuration was $Man_5GlcNAc_2$, which does not appear to be recognized by the phosphotransferase (He et al., 2013). Additionally, a small percentage of the *cgl-IDUA* harbours complex *N*-glycans, which possess xylose or fucose sugar residues and thus a potentially immunogenic product. In contrast, the present study shows that the GM1-IDUA contains *N*-glycans primarily of the $Man_{6-8}GlcNAc_2$ configuration; no complex *N*-glycans are detected. Both of these improvements demonstrate that the GM1 line may represent a preferable choice as a plant production platform for human lysosomal enzymes. On the other hand, *N*-linked glycans were found at five sites on GM1-IDUA, whereas IDUA produced in CHO cells bears six *N*-glycans. We are presently conducting *N*-glycan analyses to determine whether there are six N-glycans on IDUA of the *cgl* line of *Arabidopsis* The *N*-glycan at Asn336 was not detected on GM1-IDUA. This is one of two *N*-glycans of CHO-cell-derived IDUA that receives mannose-6-phosphate (Zhao et al., 1997). Accordingly, the lack of an *N*-glycan at this site could have a negative impact on the rate of uptake (or absolute uptake) by human cells of a phosphorylated form of GM1-IDUA. Importantly, the absence of the *N*-glycan at Asn336 did not appear to impact the catalytic efficiency of GM-1 towards the synthetic, fluorogenic substrate 4-MUI (Table 3.3; Fig. 3.2). This observation corroborates a recent study that demonstrated that the *N*-glycan at Asn 372 is involved in the enzymatic activity of IDUA (Bie et al., 2013; Maita et al., 2013). In the

Maita et al. (2013) study It was shown that the fraction of residual enzymatic activity of IDUA that had been incubated with endoglycosidase H over a time-course correlated very strongly to the fraction of deglycosylation at Asn 372, and not at Asn 336 (or any other *N*-linked glycosylation site). These results are also in accordance with the crystal structure of IDUA (Bie et al., 2013), which reveals that mannose 7 of the *N*-glycan at Asn 372 makes non-covalent interactions with idopyranosyluronic acid in the active site of the enzyme, whereas Asn 336, which is located on the opposite face of the enzyme, makes no such contacts with the substrate. In addition to enzyme activity, the different *N*-glycan profiles of *cgl*- and GM1-IDUA might modulate the thermostability of the enzyme, although this was not tested.

The reduction in *N*-glycan site occupancy of recombinant proteins has been observed in mammalian, yeast, and fungal expression platforms, likely resulting from the fact that the oligosaccharyltransferase (OST) complex exhibits a preference in substrate that varies slightly between the Eukaryotic kingdoms. The exact identity of the tripeptide sequon (i.e., Asn-X-Ser/Thr) that serves as the site of *N*-linked protein glycosylation (Kasturi et al., 1995), and the amino acid residues flanking this sequence (Ben-Dor et al., 2004), can influence the occupancy of the *N*-linked glycosylation site. Several strategies to counteract this reduction in *N*-glycan site occupancy include altering the growth conditions (e.g. within the culture medium for CHO cells; Gawlitzek et al., 2009), expressing foreign OST subunits that have altered substrate specificities (Choi et al., 2012), and engineering the tripeptide sequon of the recombinant protein. As GM1-IDUA exhibits normal enzyme activity and possesses at least one *N*-glycan that is capable of being phosphorylated, it is not yet clear if it is necessary to remedy the absence of the *N*-glycan at Asn336. Nevertheless, it may be possible to subtly modify the amino acid sequence surrounding Asn336 to make it recognizable by the plant OST complex. In this regard, studying the consensus sequences of the glycopeptides obtained from the secretory proteins of *Arabidopsis* (Song et al., 2013) might aid in the development of a chimeric IDUA for expression in the GM1 *Arabidopsis* background. However, the effect of modifying amino acids surrounding the Asn-X-Ser/Thr site may well lead to problems in protein folding.

3.4.2. The GM1 mutant yields recombinant IDUA that is amenable to *in vitro* phosphorylation and may have promise for generating an *in planta* phosphorylation system

Many intravenously supplied, recombinant lysosomal enzymes (such as IDUA for MPS I disease, and β-D-glucuronidase for MPS VII disease) utilize M6P to enter the target cells by receptor mediated uptake. Therefore, this challenge must be addressed if plants are to be used as a lysosomal enzyme production platform. The elaboration of high-mannose *N*-glycans of lysosomal enzymes with M6P has been performed *in vitro*. One method is the enzymatic transformation of a high-mannose glycan by stepwise reaction with the purified phosphotransferase and uncovering enzyme (Canfield and Kudo, 2005; He et al., 2013). This requires the purification of the two additional modifying enzymes, as well as performing the enzymatic reactions themselves. In the present study, we found that the GM1-IDUA was amenable to the *in vitro* phosphorylation process, and showed tenacity (i.e. a lower K_m) for binding to the soluble phosphotransferase. It is presently difficult to assess the degree to which the lack of an *N*-glycan at Asn336 reduced the efficiency of the *in vitro* phosphorylation process. Future experiments involving the comparison of the phosphorylated forms of *cgl*- and GM1-IDUA in terms of their *N*-glycan profiles and their uptake by cultured MPS I fibroblasts will help to determine which mutant line would be most useful for the production of a recombinant form of iduronidase. Alternative methods to achieve the necessary downstream processing of plant-made recombinant lysosomal enzymes may be considered. One is the conjugation of synthetic M6P-containing oligosaccharides to chemically modified glycans on the purified lysosomal enzyme (Zhu, 2009). In this method, the synthetic, bisphosphorylated oligomannopyranose is derivatized to incorporate, at its reducing end, a linker molecule with a free hydrazine or amino-oxy group. The purified lysosomal enzyme is then treated with sodium metaperiodate to oxidize vicinal hydroxyls to aldehydes which are attacked by the nucleophilic linker moiety. Regardless, both *in vitro* methods introduce an additional step in lysosomal enzyme purification which inevitably equates to a lower yield and increased costs. On the other hand, the construction of an *Arabidopsis* line that expresses the α- and β-subunits of the phosphotransferase and the uncovering enzyme in the seed tissues could be valuable for the production of properly phosphorylated lysosomal enzymes that do not require any further manipulation once they are purified.

3.5. Materials and Methods

3.5.1. Generation of Transgenic Lines

A gene expression construct consisting of the coding sequence of the human IDUA gene and the promoter and terminator sequences of the Arcelin 5-I gene from *Phaseolis vulgaris* was placed inside the T-DNA boundary regions of the binary vector, pRD-400 (Downing et al., 2006). This binary vector was transformed into *Agrobacterium tumefaciens* GV3101. The resulting strain of *Agrobacterium* was used to transform a line of *Arabidopsis thaliana* bearing a knockout mutation in the Golgi-Mannosidase 1 (GM1) gene by the floral dip procedure. The first generation of transformants (T1 generation) was selected for by germinating the seeds from the plants of the T0 generation on plates containing kanamycin (25 mg/mL). The steady-state level of IDUA in soluble protein extracts from seeds of the T1 generation plants was assessed by western blot, using a polyclonal antibody specific for IDUA as described in Downing et al. (2006). The fluorometric activity assay was used to measure the rate of 4-methylumbelliferyl iduronide (4-MUI) cleavage (Downing et al., 2006). The lines with the highest levels of IDUA activity were propagated for two more generations through germination on kanamycin plates.

3.5.2. Purification of recombinant α-L-iduronidase

Purification of the recombinant IDUA was as outlined in He et al. (2012). Seeds that were harvested from mature, desiccated plants were ground into a powder using a mortar and pestle that had been cooled by liquid nitrogen. Fifty milliliters of protein extraction buffer (Table 3.1) were added to every gram of ground seeds. The mixture was centrifuged at 3000 × *g* for 15 minutes, and the resulting supernatant was filtered through MiraCloth and centrifuged again at 30,000 × *g* for 45 minutes. The supernatant from the second centrifugation was recycled overnight on a 2 mL concanavalin A (ConA) Sepharose column using a peristaltic pump. The ConA column was washed with 50 column volumes of protein extraction buffer, and the bound proteins were thrice eluted in 25 column volumes of ConA elution buffer. Using a Millipore Centrifugal Filter Unit (30 kDa M.W. cutoff), the protein in all three elution fractions was concentrated and pooled,

and the alpha-methyl mannose was removed from solution. This concentrated protein solution was then recycled, overnight, over a column with an monoclonal antibody affinity column specific for IDUA (the monoclonal antibody was kindly supplied by J. Hopwood). The affinity column was washed with 100 column volumes of protein extraction buffer, and thrice eluted in 50 column volumes of IDUA elution buffer. All of the elution fractions were pooled, the protein was concentrated and the buffer was changed to IDUA storage buffer (Table 3.1).

All of the steps of the protein purification were performed at 4 ºC. Over the course of the purification procedure, the total protein concentration was assessed using the BioRad DC Protein Assay, while the IDUA activity was monitored using the fluorometric activity assay. Aliquots from representative fractions over the course of the purification procedure were resolved using SDS-PAGE and visualized by staining with Coomassie Brilliant Blue.

3.5.3. Endoglycosidase H digestion

One thousand units of endoglycosidase H (Endo H; NEB) were incubated with 2 µg each of *cgl*- and GM1-IDUA, according to the manufacturer's instructions. The reaction was resolved by SDS-PAGE and visualized by staining with Coomassie Brilliant Blue.

3.5.4. Activity assays and determination of kinetic parameters of recombinant IDUA

The activity of iduronidase was monitored by cleavage of the synthetic, fluorescent substrate, 4-methylumbelliferyl iduronide (4-MUI). In the standard assay, an aliquot of the seed extract containing GM1-IDUA was added to activity assay buffer (Table 3.1) containing 1 mM 4-MUI. Assays were performed in a volume of 50 µL for 10 min at 37 ºC, at which point 50 µL of glycine stop buffer was added and the fluorescence was read using a Molecular Devices Spectramax M2e microplate reader.

For the determination of the kinetics of iduronidase activity, 4-MUI was added over a concentration range from 2.5 to 1000 μM as described in He et al. (2013) (see Chapter 2).

3.5.5. GlcNAc-Phosphate Transfer Assay

Purified IDUA was added at a range of 0.1 to 2 μM to 50 μL of phosphotransferase reaction buffer (Table 3.1) along with 1 ug of purified, soluble UDP-GlcNAc:lysosomal enzyme phosphotransferase. After a 1-h incubation, the reaction was quenched and the protein was precipitated by adding 1 mL of phosphotransferase stop buffer. The resulting pellet was washed twice with stop buffer, then sonicated, and neutralized by the addition of 1 mL of 50 mM Tris (pH 11). The resulting suspension was added to a scintillation cocktail from which tritium decay was recorded in a scintillation counter.

3.5.6. *N*-glycan Profiling

To determine the number of *N*-glycosylation sites present on GM1-IDUA, an aliquot of the protein was digested over a time-course with PNGase F. Aliquots of the GM1-IDUA were first resolved by SDS-PAGE, then digested with trypsin or chymotrypsin and finally analyzed by nano-LC tandem mass spectrometry.

3.6. References

Bach G, Friedman R, Weissmann B, Neufeld EF. The defect in the Hurler and Scheie syndromes: deficiency of α-L-iduronidase. *Proc Nat Acad Sci USA.* **69(8):**2048-51.

Ben-Dor S, Esterman N, Rubin E, Sharon N. (2004). Biases and complex patterns in the residues flanking protein *N*-glycosylation sites. *Glycobiology*. **14:**95–101.

Bie H, Yin J, He X, Kermode AR, Goddard-Borger ED, Withers SG, James MN. (2013). Insights into mucopolysaccharidosis I from the structure and action of α-L-iduronidase. *Nat. Chem Bio.* **9(11):**739-45.

Canfield WM, Kudo M. (2005). Soluble GlcNAc Phosphotransferase. U.S. Patent 6,905,896, filed 21 Dec 2001 and issued 14 Jun 2005.

Choi BK, Warburton S, Lin H, Patel R, Boldogh I, Meehl M, d'Anjou M, Pon L, Stadheim TA, Sethuraman N. (2012). Improvement of *N*-glycan site occupancy of therapeutic glycoproteins produced in *Pichia pastoris*. *Appl Microbiol Biotechnol.* **95:**671-82.

Downing WL, Galpin JD, Clemens S, Lauzon SM, Samuels AL, Pidkowich MS, Clarke LA, Kermode AR. (2006). Synthesis of enzymatically active human alpha-L-iduronidase in Arabidopsis *cgl* (complex glycan-deficient) seeds. *Plant Biotechnol J.* **4(2):**169-81.

Gawlitzek M, Estacio M, Fürch T, Kiss R. (2009). Identification of cell culture conditions to control N-glycosylation site-occupancy of recombinant glycoproteins expressed in CHO cells. *Biotechnol Bioeng* **103:**1164-75.

Grabowski GA, Golembo M, Shaaltiel Y. (2014). Taliglucerase alfa: an enzyme replacement therapy using plant cell expression technology. *Mol Genet Metab.* **112(1):**1-8.He X, Pierce O, Haselhorst T, von Itzstein M, Kolarich D, Packer NH, Gloster TM, Vocadlo DJ, Qian Y, Brooks D, Kermode AR. (2013). *Plant Biotechnol J.* **11(9):**1034-43.

He X, Haselhorst T, von Itzstein M, Kolarich D, Packer NH, Gloster T, Vocadlo D, Clarke LA, Qian Y, and Kermode AR. (2012). Production of α-L-iduronidase in maize for the potential treatment of a lysosomal storage disease. *Nat. Commun.* **3**, 1062. doi: 10.1038/ncomms2070. Kinal H, Park CM, Berry JO, Koltin Y, Bruenn JA. (1995). Processing and secretion of a virally encoded antifungal toxin in transgenic tobacco plants: evidence for a Kex2p pathway in plants. *Plant Cell.* **7(6):**677-88.

Maita N, Tsukimura T, Taniguchi T, Saito S, Ohno K, Taniguchi H, Sakuraba H. (2013). Human α-L-iduronidase uses its own *N*-glycan as a substrate-binding and catalytic module. *Proc Nat Sci USA.* **110(36):**14628-33.Jameson E, Jones S, Wraith JE. (2013). Enzyme replacement therapy with laronidase (**Aldurazyme**(®)) for treating mucopolysaccharidosis type I. *Cochrane Database Syst Rev.* **11:**CD009354.

Jiang L, Rogers JC. (1999). The role of BP-80 and homologs in sorting proteins to vacuoles. Plant Cell. 11: 2069–2071.

Kakkis ED[1], Muenzer J, Tiller GE, Waber L, Belmont J, Passage M, Izykowski B, Phillips J, Doroshow R, Walot I, Hoft R, Neufeld EF. (2001). Enzyme-replacement therapy in mucopolysaccharidosis I. *N Engl J Med.* **344(3):**182-8.Kasturi L, Eshleman JR, Wunner WH, Shakin-Eshleman SH. (1995). The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence *N*-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. **270(24):**14756-61.

Kudo M, Canfield WM. (2006). Structural requirements for efficient processing and activation of recombinant human UDP-*N*-acetylglucosamine:lysosomal-enzyme-*N*-acetylglucosamine-1-phosphotransferase. *J Biol Chem.* **281(17):**11761-8.

Song W, Mentink RA, Henquet MG, Cordewener JH, van Dijk AD, Bosch D, America AH, van der Krol AR. (2013). *N*-glycan occupancy of *Arabidopsis N*-glycoproteins. *J Proteomics.* **93:**343-55.

Zhao KW, Faull KF, Kakkis ED, Neufeld EF. Carbohydrate structures of recombinant human α-L-iduronidase secreted by Chinese hamster ovary cells. *J. Biol. Chem.* **272:** 22758–22765.

Zhu (2009). Methods for introducing mannose 6-phosphate and other oligosaccharides onto glycoproteins. U.S. Patent 20090022702, filed 24 Sep 2008, issued 22 Jan 2009.

3.7. Tables and Figures

Table 3.1. Buffers used in this study.

Buffer	Composition
Protein extraction buffer	20 mM Tris-HCl (pH 7.4), 500 mM NaCl, 0.5 mM PMSF
ConA wash buffer	20 mM Tris-HCl (pH 7.4), 500 mM NaCl
ConA elution buffer	20 mM Tris-HCl (pH 7.4), 500 mM NaCl, 15% a-methyl-mannose
IDUA affinity elution buffer	50 mM sodium citrate (pH 4.0), 2M NaCl
IDUA storage buffer	20 mM dimethyl glutarate (pH 6.0), 200 mM NaCl, 5% (w/v) glycerol, 5% (w/v) ethanol
IDUA activity assay buffer	100 mM dimethyl glutarate (pH 4.5), 3.5 mg/mL BSA, 1 mM sodium metabisulfite
Phosphotransferase assay buffer	50 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, 10 mM $MnCl_2$, 75 uM UDP-[^{3}H]-GlcNAc (1uCi), 2 mM ATP, 2 mg/mL, 2 mg/mL BSA,

Table 3.2. Summary of purification of α-L-iduronidase.

Purification Stage	Vol (mL)	Total Activity (nmol MU /min)	Total Protein (mg)	Specific Activity (nmol MU/ min/ mg)	Purity (fold)	Yield (%)
Clarified	100	19000	130	150	1.0	100
ConA eluate	20	11000	4.2	2500	17	57
Affinity column eluate	0.60	2000	0.21	9600	66	11

Table 3.3. Comparison of kinetic parameters of 4-MUI cleavage by α-L-iduronidase purified from *cgl* and GM1 *Arabidopsis* lines.

Source	V_{max} [1]	K_m [2]	V_{max} / K_m
GM1	24	37	0.66
cgl	21	37	0.57

[1] Units of umol 4-MU formed / min / mg of IDUA
[2] Units of uM 4-MUI

Table 3.4. A comparison of the kinetic parameters of GlcNAc phosphate transfer *in vitro* from UDP-GlcNAc to several forms of purified, recombinant iduronidase, when catalyzed by soluble phosphotransferase.

Source	V_{max} [1]	K_m [2]	V_{max} / K_m
GM1[3]	7	0.4	20
cgl[4]	21.7	4.3	5.0
CHO (Aldurazyme)[3]	32	7	5

[1] Units of umol pmol [^{3}H]GlcNAc transferred / hour / µg soluble phosphotransferase
[2] Units of µM IDUA
[3] This study
[4] He et al., 2013

Table 3.5. The complete *N*-glycan profile of GM1-IDUA.

Structure		Ratio (%)
High-mannose type structure	M5	1.7
	M6B	5.8
	M7A	7.1
	M7B	13.3
	M8A	65.6
	M9	1.9
Glucose-linked structure	GlcM8A	2.9
	GlcM9	1.7
Total High mannose type		95.4
Total Glucose-linked		4.6

Table 3.6. Site-specific *N*-glycan profiles of GM1-IDUA.

Structure	Asn110	Asn190	Asn336[1]	Asn372	Asn415	Asn451
$Hex_5HexNAc_2$	4.9	3.7	-	4.1	11.3	8.8
$Hex_6HexNAc_2$	9.0	3.5	-	4.3	13.2	15.8
$Hex_7HexNAc_2$	25.9	5.4	-	3.4	11.3	31.0
$Hex_8HexNAc_2$	60.2	62.7	-	61.4	64.2	44.4
$Hex_9HexNAc_2$	-	20.9	-	22.4	-	-
$Hex_{10}HexNAc_2$	-	3.8	-	4.4	-	-

[1]No *N*-glycan was detected on Asn336

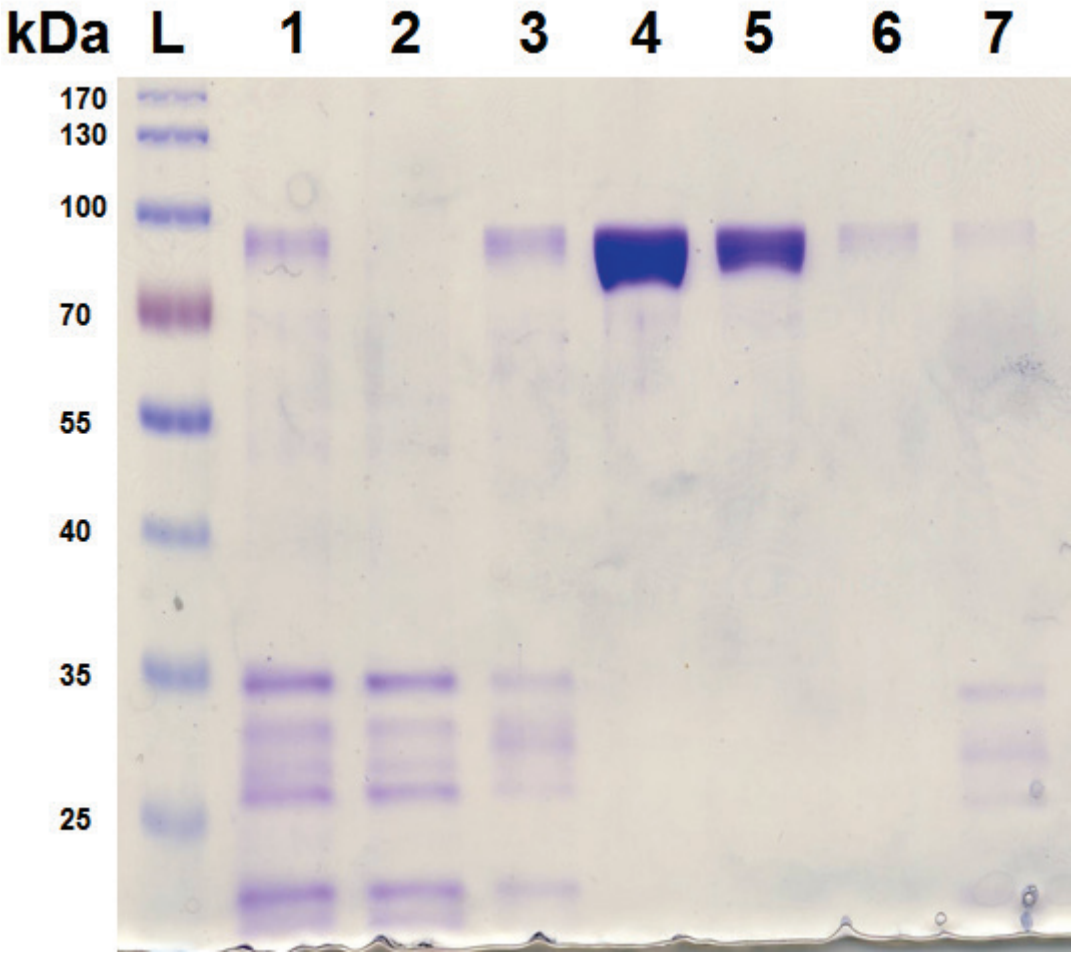


Figure 3.1. Analysis of the purification of IDUA from transgenic *Arabidopsis* GM1-knockout seeds by SDS-PAGE.

Lanes: (L) Ladder; (1) Clarified extract; (2) Concanavalin A (ConA) flow-through fraction ; (3) ConA elution fraction; (4-6) Elution fractions from three separate rounds of anti-IDUA affinity chromatography; (7) flow-through fraction after the third round of anti-IDUA affinity chromatography.

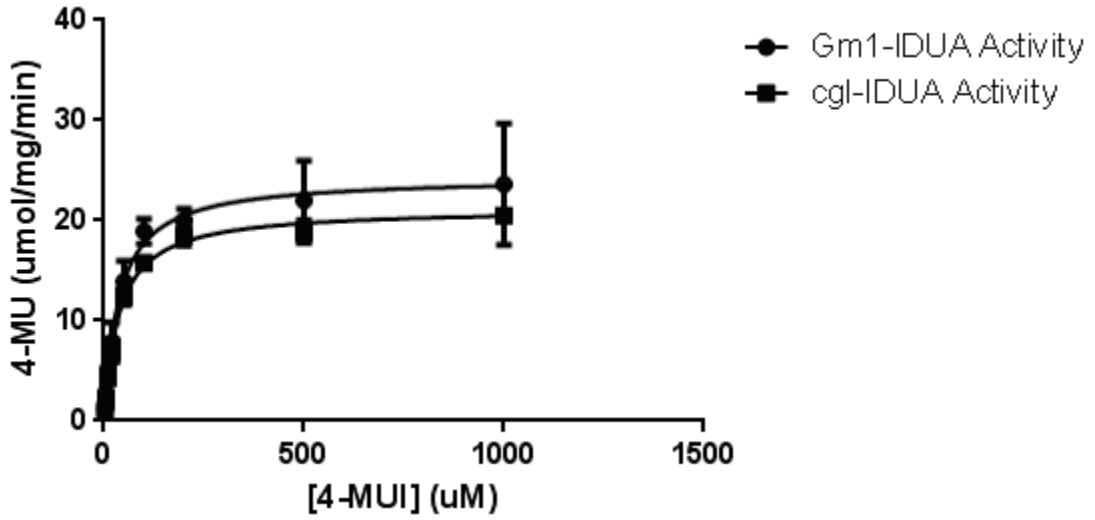


Figure 3.2. A comparison of the kinetics of 4-MUI cleavage catalyzed by *cgl*-IDUA and GM1-IDUA.

A reaction buffer containing 4-MUI over a range of concentrations up to 1000 uM was incubated with iduronidase that had been purified from the *cgl* and GM1 mutant *Arabidopsis* lines. The data are visualized as a Michaelis-Menten plot, while the kinetic parameters are tabulated in Table 3.3. All measurements were performed in triplicate.

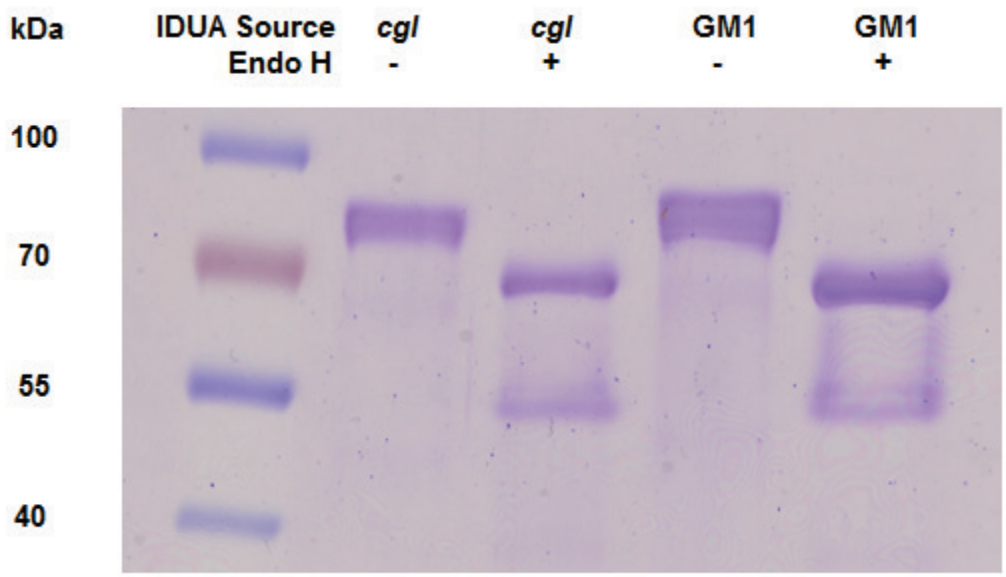


Figure 3.3. Endoglycosidase-H digestion of *cgl*-IDUA and GM1-IDUA
For each reaction, 2 ug of IDUA was digested and resolved by 10% SDS-PAGE.

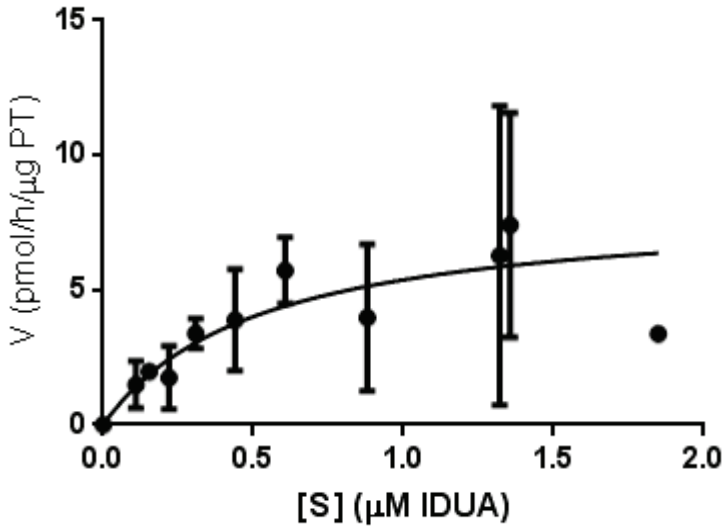


Figure 3.4. Kinetics of *in vitro* phosphorylation of GM1-IDUA by the soluble phosphotransferase.

GM1-IDUA was added over a range of 0 to 2 µM into 50 µL aliquots of the radioactive phosphotransferase assay buffer (Table 3.1), and the reaction was incubated for 1 h at 37 °C. The phosphotransferase activity was measured by pelleting the GM1-IDUA and measuring the radioactivity retained in the pellet. The data are visualized as a Michaelis-Menten plot of pmol of GlcNAc-phosphate transferred per hour per µg of phosphotransferase, against the concentration of GM1-IDUA added to the reaction. The K_m was 0.5 µM of GM1-IDUA, while the V_{max} was 8 pmol GlcNAc-phosphate/h/µg phosphotransferase (see also Table 3.4).

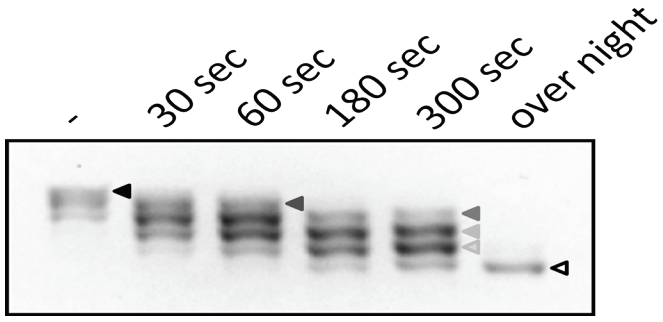


Figure 3.5. PNGase-F digestion of GM1-IDUA.
Aliquots of GM1-IDUA were digested by PNGase-F for increasing durations and resolved by SDS-PAGE.

Chapter 4. Conclusions and Future Directions

This study demonstrated that the seeds of *Arabidopsis thaliana* can be used as an expression platform for the production of recombinant human α-L-iduronidase (IDUA). Two mutant backgrounds of Arabidopsis, *cgl* and Golgi-Mannosidase I, were used to produce IDUA that was catalytically identical to the commericially available form of the enzyme that is produced in Chinese Hamster Ovary cells. Secondly, the plant-produced IDUA harbored predominantly high-mannose *N*-glycans. In particular, the Arabidopsis Golgi Mannosidase I mutant was able to produce human IDUA with no potentially immunogenic xylose or fucose residues, with most *N*-glycans containing six to eight mannose rings. Finally, both forms of the plant-produced iduronidase were also amenable to the addition of the mannose-6-phosphate tag through treatment with a soluble form of the UDP-GlcNAc:lysosomal enzyme phosphotransferase *in vitro.* This research paves the way for future attempts at engineering *Arabidopsis* to produce lysosomal enzymes, including:

1. The expression, purification, and characterization of another lysosomal hydrolase, β-D-glucuronidase (for treatment of MPS VII) in the seeds of *Arabidopsis cgl* and GM1 mutants.

2. The elaboration of mannose-6-phosphate on lysosomal enzymes *in planta* through the expression of the phosphotransferase and the uncovering enzyme in the *Arabidopsis* secretory pathway.

To elaborate on the second goal, the expression of the mammalian (human) phosphotransferase in a plant host is met with the additional challenge of releasing the alpha and beta subunits from the nascent proprotein so that they can associate as a mature protein. In mammalian cells, the phosphotransferase is processed by the site-1-protease in the *cis*-Golgi. *Arabidopsis* contains an ortholog of the mammalian site-1-protease that shares a high percentage of sequence identity, and has been demonstrated to cleave a class of transcription factors called Rapid Alkalinisation Factors (RALFs). Additionally, an expression study in CHO cells has deduced that the mature α- and β-subunits cannot be expressed individually and must be cleaved from

the αβ-proprotein (Kudo and Canfield, 2006). Both transmembrane domains are necessary for correct processing of the wild-type phosphotransferase site if the system is to occur endogenously, although the transmembrane domains are dispensible if the cleavage site for furin, a member of the kexin subfamily of proprotein convertases, is inserted in place of the natural cleave site, resulting in a secreted version of the protein (Kudo and Canfield, 2006). Kexins are subtilases that are located in the eukaroytic secretory pathway and they process prohormones and zymogens at a conserved dibasic recognition site (reviewed in Rockwell et al., 2002). Homologs are present in yeast, fungus, and mammals, and although an ortholog has yet to be formally identified in any plant species, furin-like activity has been identified in the Golgi apparatus of tobacco BY2 cells (Jiang and Rogers, 1999) and the leaves of *Arabidopsis* (Matos et al, 2008), tobacco (Kinal et al., 1995), and the common bean (Popovic et al., 2002). It is unknown whether or not there is furin-like activity in the seeds, or if the site-1-protease was responsible for this activity, although subtilisin-like proteases have been isolated from plants (Tanaka et al., 2001) and identified in computational searches of the *Arabidopsis* genome (Rautengarten et al., 2005). If the wild-type phosphotransferase cleavage site is not processed in the seed tissue of *Arabidopsis*, then replacing the cleavage site for either that found in RALFs or that used in the tobacco BY2 study may allow for phosphotransferase activation. The *in planta* system could then be completed by transgenic expression of the uncovering enzyme and of the target human lysosomal hydrolase.

4.1. References

Kudo M, Canfield WM. (2006). Structural requirements for efficient processing and activation of recombinant human UDP-N-acetylglucosamine:lysosomal-enzyme-N-acetylglucosamine-1-phosphotransferase. J Biol Chem. 281(17):11761-8.

Matos JL, Fiori CS, Silva-Filho MC, Moura DS. (2008). A conserved dibasic site is essential for correct processing of the peptide hormone AtRALF1 in Arabidopsis thaliana. *FEBS Lett.* **582(23-24):**3343-7.

Popovic T, Puzidar V, Brzin J. (2002). A novel subtilase from common bean leaves. *FEBS Lett.* **530(1-3):**163-8.

Rautengarten C, Steinhauser D, Büssis D, Stintzi A, Schaller A, Kopka J, Altmann T. (2005). Inferring hypotheses on functional relationships of genes: Analysis of the Arabidopsis thaliana subtilase gene family. *PLoS Comput Biol.* **1(4):**e40.

Rockwell NC, Krysan DJ, Komiyama T, Fuller RS. (2002). Precursor processing by kex2/furin proteases. *Chem Rev.* **102(12):**4525-48.

Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y. (2001). A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development.* **128(23)**:4681-9.