Molecular modeling and *in vitro* mutagenesis to analyse a mannosylphosphohydrolase of biotherapeutical importance

Celine Everaert

Master's dissertation submitted to obtain the degree of
Master of Biochemistry and Biotechnology
Major Bioinformatics and Systems Biology
Academic year 2011-2012

Promoter: Prof. Dr. Nico Callewaert
Scientific supervisor: Katrien Claes
Department of Biochemistry and Microbiology
Laboratory for Protein Biochemistry & Biomolecular Engineering
VIB - Department for Molecular Biomedical Research

Copromoter: Dr. Ewald Pauwels
Department Physics and Astronomy
Center for Molecular Modeling

“This document and the information in it are provided in confidence, for the sole purpose of evaluating Celine Everaert for her master dissertation, and may not be disclosed to any third party or used for any other purpose without the explicit written permission of Nico Callewaert ”

Preface

When I started at the university, I was a young adolescent, in search of new knowledge. Ever since I was a little, I was asking a lot of questions. How does this work? Why does it work this way? Sometimes my parents or teachers could give me answers, but ever so often they didn't know the answer either. In my fifth and sixth year of highschool we had some courses about cell function and genetics. That is where my interest in life science was formed. This was the basis for my choice to study biochemistry and biotechnology. Now, five years later, I think I can say that I gained a lot of knowledge, new friendships, adulthood and wisdom of life. I have learned how to find answers myself; on scientific questions, especially in this thesis.

Therefore I would like to thank some people. First of all the persons who coached me during the last year: I thank Prof. Nico Callewaert for the opportunity to do my thesis in his lab, for his constructive feedback and advice. My gratitude also goes to Ewald Pauwels for the modelling support, specifically his rapid mail answers up until midnight (sometimes even funny and/or geeky). Many thanks also to Katrien Claes for being my guide through the lab, helping me with the experimental part, and for the chats in between. Also, I am very grateful to the other lab members, especially Petra and Charlotte (the LSD-experts), for giving me sometimes small but always very useful hints and for the nice work atmosphere I could experience.

I am also very lucky to receive a lot of support of my family. This year has been emotionally very hard, due to the loss of a close friend and my grandmother, with whom I was inseparable. The support of my family was very welcome. Mam, thanks for the chats on our way to Oma, your infinite patience and support during my study. Papa thanks for the healthy meals, discussion about the meaning of life and your unsalted opinion about genetic modification. Emil, Titi and Artuur it is a pleasure to be your big sister. I would also like thank my friends; the girls from university (Ans, Annelynn, Charlot, Dagmar, Daphne, Karen & Ruth), as we all had some struggle moments to complain about but also many other moments where we had a lot of fun (eg. the ladies nights), but also my friends musicians (too many to mention); making and listening to music is the ideal way to relax for me, and all the others.

Table of Contents

List of figures

List of abbreviations

AC-LL	acid-cluster dileucine
Ala351	Ala351Leu mutant
Ala631	Ala631Leu mutant
AOX	alcohol oxidase I
CCV	clatherin coated vesicles
CD-M6PR	cation-dependent mannose-6-phosphate receptor
CcMan5	*Cellulosimicrobium cellulans* mannosidase 5
CHO	Chinese hamster ovary
CI-MGPR	cation-independent mannose-6-phosphate receptor
CNS	central nervous system
DSA-FACE	DNA sequencer aided fluophore-assisted carbohydrate electrophoresis
ER	endoplasmatic reticulum
ERT	enzyme replacement therapy
GAG	glycosaminoglycan
GH	glycosyl hydrolase
GlcNAc	N-acetylglucosamine
GRAS	generally recognized as safe
IGF-II	insulin-like growth factor 2
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria Broth
LSD	lysosomal storage disorder
M6PR	mannose-6-phosphate receptor
Man-Pi-6-Man	mannose 6-phosphate-mannose
Pi-6-Man	mannose-6-phosphate
MD	molecular dynamics
MM	molecular mechanics
MPS	mucopolysacharidosis
NMR	nuclear magnetic resonance

OD	optical density
PCT	pharmacological chaperone therapy
QM/MM	quantum mechanics/molecular mechanics
QM	quantum mechanics
rhGAA	recombinant human acid alpha-glucosidase
RMSD	root mean square displacement
RMSF	root mean square fluctuation
TGN	trans-Golgi network
UDP	uridine diphosphate

Abstract English

Lysosomal storage disorders (LSDs) are a group of more than 50 inherited monogenic disorders. LSDs are mostly due to loss-of-function mutations in genes encoding lysosomal hydrolases resulting in impaired substrate degradation. An increasing part of the LSDs can successfully be treated with enzyme replacement therapy (ERT). A major lysosomal targeting pathway is based on modification of enzymes with mannose-6-phosphate (Pi-6-Man) moieties. The critical factor in therapeutic success of ERT is that the enzymes need to be modified with a high density of Pi-6-Man modifications on their N-glycans. This glycotope is recognized by the Pi-6-Man receptor (M6PR) at the surface of human cells, leading to endocytosis of the receptor:enzyme complex. The enzyme is then delivered to the lysosomes, complementing the enzymatic defect.

Nowadays ERT enzymes are usually produced in mammalian cells but this makes the therapy very expensive and not very efficacious because of the low levels of Pi-6-Man modifications. Yeast cells could be a great alternative, as they produce a similar glycotope on their cell wall mannoproteins, mannose-6-Pi-mannose (Man-Pi-6-Man). To produce enzymes for ERT in yeast we need a specific enzyme, *Cellulosimicrobium* cellulans mannosidase 5 (CcMan5). CcMan5 uncaps the mannose from Man-Pi-6-Man that is present twice on the N-glycans at the enzymes for ERT. A higher demannosylation rate and especially the rapid demannosylation of both Man-Pi-6-Man modifications, is desired in the production process of the enzymes. The N-glycan possesses two different regio-isomers with the Man-Pi-6-Man modification. The aim of the project is to model the interaction between CcMan5 and the different phosphorylated glycan regio-isomers. This will yield insight into how the protein can be engineered so that more complete uncapping can be obtained faster. The project consists of an experimental and computational part. In the computational part we will simulate possible mutations and dock the Pi-6-Man N-glycans on it. In parallel, these mutations were be made in the lab and enzymatically evaluated. The produced mutants were the Ala351Leu and Ala631Leu point mutants and the Y615-P642 deletion mutant.

The deletion mutant did not express in the periplasmatic fraction, probably due to folding problems. The Ala351Leu point mutant is presumably less active then the wild type and Ala631Leu mutant. The docking results showed that it is possible that the two regio-isomers bind the active site in an opposite manner. The free bindings energy of the best results of one of the regio-isomers is 1kJ/mol higher than the other. This could be an explanation for the favourable uncapping of just one mannose from the Man-Pi-6-Man's on enzymes for ERT. Still, more experiments are needed.

Abstract Nederlands

Lysosomale opslag ziekten (LSDs) zijn een groep van meer dan 50 overerfbare, monogenische aandoeningen. LSDs zijn meestal het gevolg van gebrekkige substraat degradatie door verliesmutaties in genen coderend voor lysosomale hydrolases. Een toenemend aantal LSDs kunnen succesvol behandeld worden met enzym vervanging therapie (ERT). De kritische factor voor succes is dat de enzymen moeten gemodificeerd zijn met een hoge densiteit aan mannose-6-fosfaat (Pi-6-Man) modificaties op hun N-glycanen. Dit glycotoop wordt herkend door de Pi-6-Man receptor (M6PR), aanwezig op het oppervlak van humane cellen, resulterend in endocytose van het receptor:enzym complex. Het enzym wordt dan aan het lysosoom geleverd waar het enzymatisch defect erdoor wordt gecomplementeerd.

Momenteel worden ERT enzymen voornamelijk aangemaakt in zoogdiercellen maar dit maakt de therapie zeer duur en niet efficiënt door de lage graad aan Pi-6-Man modificaties. Gistcellen kunnen een goed alternatief zijn, aangezien ze gelijkaardige glycotopen produceren op hun mannoproteïnes in de celwand. Om de enzymen voor ERT in gist te produceren is er een specifiek enzym nodig, *Cellulosimicrobium* cellulans mannosidase 5 (CcMan5). CcMan5 verwijdert het eindstandig mannose van mannose-6-fosfaat-mannose (Man-Pi-6-Man) dat tweemaal aanwezig is op de N-glycanen. Een hogere demannosylatie snelheid en in het bijzonder de snelle demannosylatie van beide Man-Pi-6-Man modificaties, is noodzakelijk in het productie proces van de enzymen. Het N-glycaan bevat twee regio-isomeren met de Man-Pi-6-Man modificatie. Het doel van het project is om de interactie tussen CcMan5 en de verschillende gefosforyleerde regio-isomeren van het glycaan te modellen. Dit zal inzicht geven in hoe het proteïne kan aangepast worden zodat complete verwijdering van het eindstandig mannose snel bereikt kan worden. Het project bestaat uit een experimenteel en computationeel deel. In het computationeel deel zullen we mogelijke mutaties simuleren en hierop het Pi-6-Man N-glycaan docken. Gelijktijdig zullen we deze mutaties aanmaken in het labo en enzymatisch evalueren. De geproduceerde mutanten waren de Ala351Leu en Ala631Leu puntmutanten en de Y615-P642 deletiemutant.

De deletiemutant werd niet aangemaakt in de periplasmatische fractie, waarschijnlijk door problemen met de opvouwing. De Ala351Leu puntmutant is wellicht minder actief dan het wildtype en de Ala631Leu mutant. De docking resultaten tonen dat het mogelijk is dat de twee regio-isomeren in een tegenovergestelde richting de actieve site binden. De vrije bindingsenergie van de beste resultaten van één van de regio-isomeren is 1kJ/mol hoger dan het ander. Dit kan verklaren waarom vaker slechts een mannose van de Man-Pi-6-Man's op de enzymen voor ERT wordt verwijdert. Er zijn echter nog meer experimenten nodig om een sluitende conclusie te trekken.

I Introduction

I.1 Lysosomal Storage Disorders

Lysosomal storage disorders (LSDs) are a large group of more than 50 inherited monogenic disorders, including Fabry, Gaucher and Pompe disease. Every form of LSD is individually rare, although their combined prevalence is estimated to be 1 in 8000 (Schultz et al., 2011). Most of the mutations are autosomal recessively inherited, but some are X-chromosome linked (eg. Fabry and Hunter disease) (Bach et al., 1973; Germain et al., 2010).

I.1.1 Normal function of the Lysosome

In 1955 De Duve et al. described the lysosome as an acidic organelle (pH 4.5), which differs substantially from the surrounding cytosol (De Duve et al., 1955). Lysosomes are found exclusively in animal cells. Plants and yeast have a vacuole instead (Samaj et al., 2005). Cells have three possible ways to take up material, depending on the size of the molecules. (Figure 1). Soluble macromolecules are taken by *(I)* endocytosis in the cell, where they form part of an endosome that matures towards a lysosome. *(II)* Autophagy is the process that degrades the cells-own organelles and bulk cytoplasm. Autophagy is a major action during starvation to maintain the cellular homeostasis. It has also an important role in inflammation and programmed cell death. Whole cells including bacteria or other large, insoluble particles are taken up by *(III)* phagocytosis (Holtzman, 1989). The molecules that must be degraded are transported to the lysosomes. The lysosomes vary in shape and size (0.2-20μm) and usually there are hundreds in a cell. Primary lysosomes are roughly spherical and do not contain particulate or membrane debris. Secondary lysosomes are larger and irregularly shaped. They result from a fusion of primary lysosomes with late endosomes and other membrane bound organelles and vesicles (Lodish, 2008).

The main function of lysosomes is degradation of components that have become obsolete for the cell or organism. They are able to break down cellular waste but also macromolecules like lipids, proteins, carbohydrates and others. The resulting compounds are returned to the cytoplasm, where they are used as building blocks for new macromolecules. For this breakdown, at least 50 hydrolases (Journet et al., 2002) and seven integral membrane proteins (Eskelinen et al., 2003) are present in the acidic environment of the lysosome (Lodish, 2008). The pH of 4-6 is crucial since all present enzymes work most efficiently under these conditions. Hence, no degradation of the macromolecules will appear in the cytosol after leakage of these enzymes. Also, this low pH is important for cargo release, maturation of transport vesicles, autophagy, neurotransmitter loading, intracellular trafficking and mannose-6-Phosphate receptor (M6PR) recycling. The acidic environment is maintained by two types of transport proteins, ATP stimulated V-type proton pumps and a chloride ion channel.

Lysosomal enzymes are synthesized in the endoplasmic reticulum (ER) and transported to the Golgi apparatus, where they undergo post-translational modifications. An important one is the attachment of mannose-6-phosphate groups (Pi-6-Man) to some of the oligosaccharide side chains. This serves as a transport signal that is recognized by a specific

receptor, the M6PR. Small transport vesicles, containing the receptor-bound enzymes, are pitched off from the Golgi and fuses with the lysosomes.

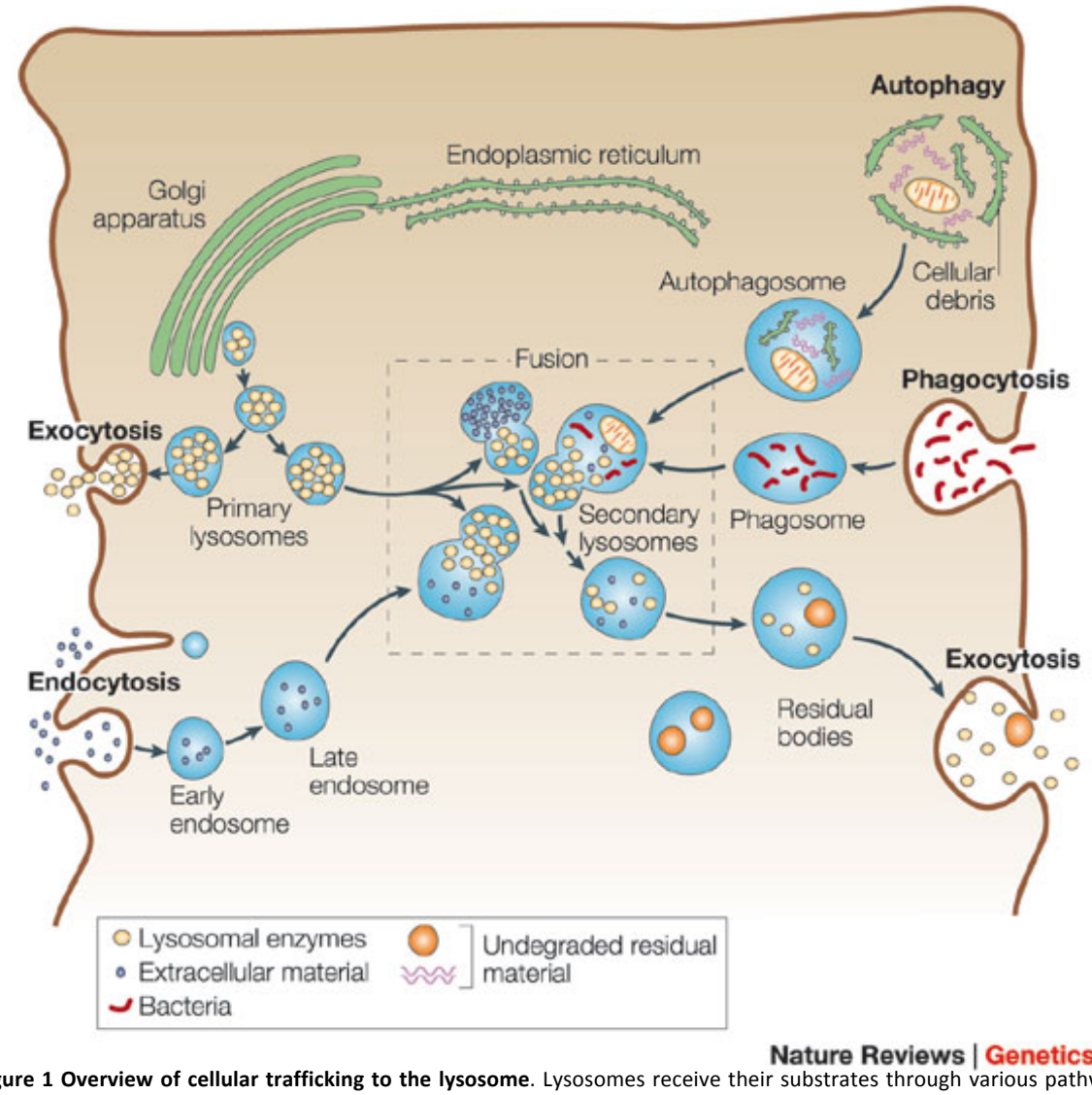


Figure 1 Overview of cellular trafficking to the lysosome. Lysosomes receive their substrates through various pathways, including endocytosis, phagocytosis and autophagy, depending on the size. Lysosomal proteins reach the primary lysosome directly, via the trans-Golgi network, or indirectly by the plasma membrane. After fusion with the endosomes, phagosomes or autophagosomes, the primary lysosomes mature into secondary lysosomes, which are larger. Degradation of macromolecules starts and the residuals are removed by exocytosis. (Desnick and Schuchman, 2002)

I.1.2 Pathophysiology of LSD

LSD's can be caused by problems with proteins involved in vesicular traffic (Blanz et al., 2010; Reczek et al., 2007), the biogenesis of lysosomes, pH regulation (Lee et al., 2010) or the sorting mechanism. However, LSDs are mostly due to loss-of-function mutations in genes encoding lysosomal hydrolases, resulting in impaired substrate degradation (Futerman and van Meer, 2004; Parkinson-Lawrence et al., 2010). Mutations for one type of LSD are normally monogenic, but in different patients, the mutated locations and resulting amino acids could vary (Kroos et al., 2008). For a dozen of genes involved in lysosomal function, LSDs are known. Probably there are also LSDs for the other genes, but they are not yet discovered. Some of them will have such a huge impact that they are lethal in early stages.

Symptoms and onset of most LSDs vary, which can be explained by the mutated gene and the difference in organs affected (Futerman and van Meer, 2004; Kroos et al., 2008; Schultz et al., 2011). The fact that the disease is very heterogeneous and most newborns appear normal; makes it hard to diagnose LSDs just based on symptoms. Nevertheless, early detection is desired so that one can start with treatment before the onset of the irreversible pathology (Nakamura et al., 2011; Staretz-Chacham et al., 2009). Also, no obvious genotype-phenotype correlation has been found. So, the clinical course of the disease cannot be predicted based on mutational analysis, meaning that individuals with the same mutation could be asymptomatic whether others suffer from a severe form (Futerman and van Meer, 2004).

The accumulation of undegraded substrates leads to more and larger lysosomes causing cell and tissue dysfunction and clinical manifestations. The extensive range of symptoms indicates that many secondary biochemical and cellular pathways must be activated, causing tissue pathology, altered gene expression and the activation of tertiary biochemical pathways (Futerman and van Meer, 2004).

Thus, different explanations for the impact of lysosomal damage on tissues can be given:

- An explanation for the pathology would be leakage of some lysosomal components such as hydrolases and metabolites. However, the presence of lysosomal components in the cytosol is normal so that this claim is refused. Under various physiological and pharmacological conditions, e.g. apoptosis, lysosomal leakage is a normal process (Ferri and Kroemer, 2001). Thereby, some integral membrane proteins are able to pump metabolites out of lysosomes into the cytosol (Eskelinen et al., 2003).
- Also, accumulation of undegraded metabolites can block intracellular transport due to an overload (Marks and Pagano, 2002).
- Defective intracellular signalling could cause symptoms, e.g. changes in production of ceramide, a lipid second messenger (Hannun and Obeid, 2002; Li et al., 2002).
- Mutations could result indirectly in changes of secondary metabolites, having an impact on tissue function (Hollak et al., 1994).
- The number of gene-expression changes is overwhelming. These changes in gene expression were determined by gene profiling, mainly using microarrays. The change in expression and the changed genes depend on the type of mutation, stage of disease progression and the tissue (Brooks et al., 2003).

LSDs are classified based on the nature of the accumulating substrate or on characterization of the defective enzyme. Here we will discuss a selection of LSDs for which enzyme replacement therapy (ERT) is available and by this they are relevant for our project.

I.1.2.1 Pompe Disease

Pompe disease is an inherited metabolic myopathy, characterized by the deficiency of lysosomal α-glucosidase (Hers, 1963). This leads to storage of glycogen in the lysosome. The prevalence differs from one in 40,000 in the Dutch population (Ausems et al., 1999) to one in 146,000 in the Australian one (Meikle et al., 1999). Pompe disease presents a spectrum of features in which symptoms can manifest at any age with median age of onset around 2

months (Engel et al., 1973). The major manifestations are skeletal muscle and cardiac weakness (Figure 2). However, the α-glucosidase gene is needed in all tissues, so manifestations are found in all organ systems. The severity depends on the glycogen production of the system (van der Ploeg and Reuser, 2008). Symptoms are feeding difficulties, respiratory infections, hypotonia and very few movements due to muscle weakness. The mean age of death is 6.0-8.7 months.

Figure 2 Patient with severe form of Pompe disease, a alpha-glucosidase defect, with typical symptoms as feeding difficulties and head lag due to muscle weakness. The life expectance is around 7 months.

In patients with less progressive forms, the onset of symptoms ranges from infancy to adulthood and clear diagnosis based on symptoms cannot be made (Wokke et al., 1995). The dominant symptom is again skeletal muscle dysfunction, resulting in immobility and respiratory problems. The heart is sporadically affected (Talsma et al., 2002). The life expectancy of an adult with Pompe disease is about 40-70 years. The causes of death are pulmonary complications or aneurysm due to accumulation of glycogen in vascular smooth-muscle cells (Matsuoka et al., 1988).

As mentioned above, changes in skeletal muscle are predominant in both classic and late onset Pompe disease. Small lysosomes expand during disease progression and they interfere with the fiber structure of the muscle. This results in atrophy and reduced performance per unit of muscle mass (Hesselink et al., 2002). A low caloric intake might enhance the loss of muscle mass and function, as ageing and immobilisation do.

Diagnosis is mostly based on residual activity of α-glucosidase, which can be detected using a specific and sensitive artificial substrate, 4-methylumbelliferyl-α-D-glucopyranoside. The residual activity is a good indication for the severity. Infants with the classical Pompe disease have less than 1% residual activity, while children and adults with the less severe form, have no more than 30% of average normal activity (Mehler and DiMauro, 1977). Full correction of the enzyme is not needed, since carriers with 50% residual activity are not affected. Furthermore, several other assays, based on antibodies (Umapathysivam et al., 2001), mRNA (Martiniuk et al., 1986) or other artificial substrates (Jack et al., 2006), can be used to screen newborns. Also genotyping is possible as the gene is well characterized. The α-glucosidase gene is localized on chromosome 17q25.2-q25.3 and contains 19 exons (Martiniuk et al., 1986). Both alleles need to harbour a pathogenic mutation before the phenotype develops. The mutations affect the splicing process, mRNA stability or biosynthesis. There are several studies that indicate the role of secondary genetic or non-genetic factors. In some patients with complete null mutations, first symptoms manifest in childhood whereas others remained presymptomatic until late adulthood (Kroos et al., 2007).

I.1.2.2 Fabry Disease

Fabry disease is a pathology induced by lysosomal glycosphingolipid and globotriaosylceramide deposition caused by absence or deficient activity of lysosomal exoglycohydrolase α-galactosidase A (Kint, 1970). This orphan disease has a high prevalence in Italy (Spada et al., 2006) and Taiwan (Hwu et al., 2009). Fabry disease has an X-chromosome linked inheritance pattern but heterozygotes could also develop some symptoms (MacDermot et al., 2001a, b). The primary disease process starts in infancy or even during fetal stage (Vedder et al., 2006) but, in contrast to other LSDs, the patients remain asymptomatic during first years of life (Hopkin et al., 2008). One of the first symptoms is chronic or episodic pain due to the damage of small nerves of the peripheral somatic and autonomic nerve system (Dutsch et al., 2002). From adolescence on, cardiac and cerebrovascular abnormalities may be present. The main visible symptom is angiokeratomas, which are red, tiny, painless elevations of the skin. During aging, damage to vital organ systems develops, resulting in a limited life expectancy by renal diseases, cardiovascular and cerebrovascular complications (Schiffmann et al., 2009).

Diagnosis could be based on an enzymatic assay using for example 4-methylumbelliferyl-alpha-D-galactopyranoside as artificial substrate. This can demonstrate the deficient activity of α-galactosidase in plasma or leukocytes (Mayes et al., 1981). Plasma globotriaosylceramide levels are also used but this is more time consuming (Vedder et al., 2007). Affected females could have a normal activity although they carry the mutation, therefore their status is determined by genotyping (Germain et al., 2010). The location of the gene is known, chromosome X 100.65-100.66 Mb. This and the fact that the gene has a small size make precise characterization of the mutation possible. In Italy and Taiwan newborn screens are performed, as the prevalence is much higher than in the rest of the world.

I.1.2.3 Mucopolysaccharidoses

Mucopolysaccharidoses (MPS) are a group of rare genetic disorders of glycosaminoglycan (GAG) catabolism (Muenzer, 1986). Each MPS disorder stands for a mutation in a different gene involved in GAG degradation. MPS is characterized by accumulation of partially degraded GAG in the lysosomes and elevation of GAG in the urine, blood (Tomatsu et al., 2005a; Tomatsu et al., 2005b) and cerebral spinal fluid (Federico et al., 1981). Most MPSs are autosomal recessive, except MPS II, which is an X-linked recessive disorder. The severity depends on the kind of mutation and varies from stillbirth to undiagnosed for several years (Vijay and Wraith, 2005). MPS patients could have both somatic and cognitive involvement. Skeletal changes result in loss of joint range of motion, restricted mobility and growth slowing or arrest, resulting in a short stature (Figure 3). Other symptoms include vision loss, hearing loss, decreased pulmonary function and cardiac disease. Mental retardation can be profound in patients with severe disease. As with the other diseases, early and accurate diagnosis is very important to start with treatment as soon as possible and prevent thereby irreversible damage. Measurement of the urinary GAG levels is a useful screening test for MPS disorders but false-negative results are very common (Mahalingam et al., 2004). In contrast, enzyme activity assays are considered more accurate. Gene sequencing can follow biochemical diagnosis to identify the mutations present.

Figure 3 Child with mucopolysaccharidosis type IV, a defect in the GAG synthesis, with typical skeletal changes resulting in growth slowing. (http://www.gfmer.ch/genetic_diseases_v2/gendis_detail_list.php?cat3=619)

I.1.2.4 Gaucher Disease

Gaucher disease, the most common LSD, refers to a cluster of autosomal recessive disorders resulting from mutations in the gene encoding glucocerebrosidase. This is an enzyme that normally cleaves the glucose residue from ceramide. Glucocerebroside is formed form the catabolism of glycolipids, derived from cell membranes of senescent leukocytes and erythrocytes (Brady et al., 1965). The enzyme defect results in a glucocerebroside accumulation in phagocytes but also in cells of the central nervous system (CNS). This results in activation of macrophages and consequent secretion of cytokines such as interleukin 1, interleukin 6 and tumour necrosis factor (Robbins et al., 2010).

Figure 4 Patient with Gaucher Disease, defective glucocerebrosidase, with typical enlargement of the liver and spleen (http://geneticpeople.com/?tag=gaucher-disease)

Gaucher disease may occur at any age and in any population, but it is overrepresented in Ashkenazi Jewisch (Beutler et al., 1993) and Northern Sweden population. Gaucher diseases are classified in three subtypes. Type I, the most common one (90%), is associated with subtotal deficiency of lysosomal acid β-glucocerebrosidase. Infiltrating macrophages in liver, spleen and bone marrow is typical. This

leads often to bleeding, anemia and massive enlargement of the liver and spleen (Figure 4). Macrophages in the bone marrow are leading to necrosis, resulting in impaired function of large joints (Mikosch and Hughes, 2010). Neurologic manifestations, resembling Parkinsonism, take place in middle life (Bultron et al., 2010). Type II, an acute neurological illness, is a very rare disorder causing death in the first five years of live. Type III, less than 10% of Gaucher disease patients, is a chronic neuropathic form. Typical symptoms are epilepsy and nerve deafness.

Gaucher disease may be identified by tissue biopsy of bone marrow or liver, but abnormal macrophages are also observed in some other haematological conditions. For a specific diagnosis, one must measure the acid β-glucosidase activity in leukocytes (Cox, 2010)

I.2 Treatment of LSD

Because of the complexity of LSD phenotypes and pathophysiology, therapeutic interventions are challenging. Consequently, most symptoms of the LSD-patients are managed with a conventional multidisciplinary approach (Germain, 2010). To address the underlying defect ERT, chaperones, substrate reduction therapy and bone marrow transplantation have been introduced. But each of these treatments has their limitations and unsolved issues.

I.2.1 Bone marrow transplantations

This aggressive therapy involves replacement of the patient's stem cells by multipotent hematopoietic stem cells out of the bone marrow of the donor. These donor cells contain sufficient amounts of enzyme to correct the deficiency. Before transplantation the natural immune system of the patient must be destroyed by chemotherapy, so that the donor cells survive. Since the first bone marrow transplantation, hundreds of patients with LSD received donor cells. The best results are obtained for MPS type IV and children under the age of 2 years. They show a significant gain in cognitive function but the skeleton does not respond as well. Unfortunately bone marrow transplantation is related with a high mortality rate and have variable success rates (Hobbs et al., 1981). That is the reason why they only use it for life-threatening diseases (Robbins et al., 2010).

I.2.2 Substrate reduction therapy

The concept of this principle is to reduce the amount of storage material instead of enhancing the degrading enzymes. To reduce storage material, the enzymes that synthesize the compound are inhibited. The reduced biosynthesis should offset the catabolic effect resulting in a balance. Therefore substrate reduction therapy is only possible when there is residual activity. N-butyldeoxynojirimycin, an imino sugar, has the ability to inhibit the enzyme that synthesizes glucosylceramide, the storage compound in Gaucher disease (Platt et al., 1994). This is a good alternative for persons with mild manifestations for whom ERT is not an option. The small molecules are able to cross the blood-brain barrier so that it could be used for CNS pathologies. As for Gaucher disease, substrate reduction therapies for MPSs and Niemann-Pick were developed successfully.

I.2.3 Molecular chaperone therapy

Missense mutations often result in the inability to fold properly leading to toxic aggregates or increased degradation. The ER has a quality control system (ERAD) based on chaperones

that facilitates correct folding and sensor molecules that tag improperly folded proteins for proteolysis (Ellgaard and Helenius, 2003). The use of pharmacological chaperone therapy (PCT) is based on the concept that they may assist the folding of mutated enzymes, which can still be catalytically active. LSDs are an excellent candidate for PCT, as they only need a minor increase in enzyme activity. PCTs are already developed for Gaucher disease (Yu et al., 2007), Pompe disease (Okumiya et al., 2007) and different types of MPS. These drugs can be ingested orally but long-term effects are not known. The cost depends on the used small molecule. With the second-generation compounds one tries to find less expensive ones. The small molecules could penetrate cell membranes so that they are widely distributed in tissues and organs. They have potential for treatment of diseases with CNS involvement as they can cross the blood-brain barrier (Valenzano et al., 2011).

I.2.4 Gene Therapy

Gene therapy is the delivery of a wild-type copy of the defective gene into the recipient cells to restore enzyme activity. Gene therapy is very promising for LSDs, as they are monogenic diseases and a small amount of enzyme can have a profound impact. Gene therapy could overcome systemic and neurological defects. Not all cells need to be genetically modified, as the M6PR makes cross-correction possible. Diffusion of lysosomal enzymes in the cerebrospinal fluid may enhance distribution in the CNS. This therapy can be accomplished with *ex vivo* and *in vivo* strategies using different viral vectors as gamma-retroviruses, lentiviruses, adeno-associated viruses and others. *In vivo* refers to injection of the vector directly into a tissue (intraparenchymal), the CNS ventricular system (intrathecal) or into the circulation (systemic). *Ex vivo* therapy is based on the isolation of stem, progenitor or differentiated cells of patients or a normal donor. After transduction and selection, the cells are transplanted in the patient. The success rate depends on the proper combination of vector, transgene product, dosing and delivery route. The biggest hurdles are oncogenicity and immunogenicity (Gritti, 2011).

I.2.5 Enzyme Replacement Therapy

In most inherited diseases there are no specific pathways available to deliver the enzyme in the desired intracellular compartment, which makes ERT impossible. In contrast, some LSDs, like Fabry, Pompe and MPS, can be treated by ERT thanks to the M6PR pathway (Beck, 2002; Beutler et al., 1991; Giugliani et al., 2011; Lee et al., 2003; Weinreb et al., 2002). Lysosomal targeting of both endogenously synthesised and exogenously supplied enzymes happens via the M6PR (Braulke and Bonifacino, 2009). For binding to this receptor, Pi-6-Man are crucial (Zhu et al., 2004). Therefore, lysosomal enzymes are post-translationally modified with N-glycans containing these Pi-6-Man groups. Despite the success, a major problem derives from enzyme distribution. The recombinant enzymes depend on the M6PR pathways for delivery to the lysosomes (Wraith, 2006). They are also unable to cross the blood-brain barrier so that ERT is not successful for neurologic disorders. To cross the blood-brain barrier, systems are designed based on production of fusion proteins. These systems are still in progress. There are also problems with ERT efficacy when, due to the pathophysiology, cell abnormalities occur, like impaired receptor trafficking (Cardone et al., 2008). ERT may cause an immune response but only a few cases are described (Koeberl and Kishnani, 2009).

Enzymes for ERT are usually produced in milk of transgenic mice (Bijvoet et al., 1996; Bijvoet et al., 1998) and rabbits (Bijvoet et al., 1999), in insect cells, the Chinese Hamster Ovary

(CHO) cell line (VanHove et al., 1996) or in human cells by over-expression of the native gene (Corchero et al., 2011; Lachmann, 2011). But mammalian cells have some drawbacks such as high costs to cultivate and they have a high potential for contamination by viruses (Demain and Vaishnav, 2009). Furthermore, they have a poor secretion of enzymes, which have evolved to be transported to the lysosomes. Producing the enzymes for ERT in mammalian cells is complicated because enzymes with high levels of Pi-6-Man are targeted to the own lysosomes and only those with low levels of lysosome-targeting modification are secreted. Consequently, the lysosomal targeting efficacy of these enzymes is rather limited, requiring high doses. All of this, in combination with the limited patient number, results in an extremely high cost of €150,000 – 250,000 per patient per year, a challenge for society. The authorities stimulate the pharmaceutical industry with extra incentives to produce ERTs for other LSDs or to develop cheaper therapies as they are covered by the orphan drug regulation (Lachmann, 2011).

I.2.5.1 Pompe disease

In the 1960s, Pompe was the first LSD treated with ERT using α-glucosidase preparations from *Aspergillus niger* and human placenta. The first trials were without benefits for almost all patients, as the dosing was too low (Hug and Schubert, 1967). Afterwards α-glucosidase was produced in milk of transgenic rabbits (Bijvoet et al., 1999) and in CHO cells (Myozyme®) (VanHove et al., 1996). When both enzyme preparations were given in high doses of 10 mg/kg, the lysosomal glycogen storage in the heart, skeletal muscle, smooth muscle and in several other organs was diminished but the brain was unaffected. The longest survivors of the severe form in the first trial were 8 years old, while untreated infants do not survive beyond 1 year of age. But not all patients respond as well to the therapy. The younger the patients are at the start of the ERT, the better the results (Cho et al., 2012; van der Ploeg and Reuser, 2008). Patients with the less severe form stabilized their pathology after long-term treatment (Winkel et al., 2004).

I.2.5.2 Fabry disease

After the isolation and sequencing of the cDNA encoding human α-galactosidase A and the availability of a mouse model and a CHO cell production system, one was able to produce the recombinant enzyme. There are two commercial therapies available. Agalsidase beta (Fabrazyme®) is the engineered form of α-galactosidase A produced in the CHO cell line, while agalsidase alpha (Replagal®) was produced in human skin fibroblasts. Patients need an intravenous injection every 2 weeks. The doses are different for both products, 0.2 mg/kg for agalisidase alpha and 1.0 mg/kg for agalisidase beta. Both products result in normalized globotriaosylceramide levels in various organs, with numerous symptomatic benefits. Guidelines for these ERTs vary from country to country. The two distinct proteins are approved for use in Europe (Ramaswami, 2011).

I.2.5.3 Mucopolysacharidoses

The first approved ERT for MPS was laronidase (Aldurazyme®) for MPSI. It is available since 2003 in the USA and Europe. Patients could walk for an increased distance, decreased liver volume and decreased urinary GAG levels were observed (Clarke et al., 2009). For MPS II idursulfase (Elaprase®) is available since 2006 and for MPS VI galsufase (Naglazyme®) since 2005. As MPS I and MPS II involve CNS, ERT does not prevent cognitive decline. Most of the

patients developed neutralizing antibodies, as a normal immune reaction against foreign proteins. But immune tolerance might enhance treatment efficacy, as the neutralizing antibodies does not induces break down of the enzyme anymore.

I.2.5.4 Gaucher disease

Gaucher is an attractive target for ERT as it lacks CNS involvement, however the uptake mechanisme is not based on the M6P pathway. The defective enzyme for Gaucher is acid β-glucocerebrosidase. In the beginning it was purified from human placenta to treat some patients but it had not the desired outcome. It was shown that the enzyme formed complex carbohydrate chains with mostly terminal sialic acid residues. Sequential removal of the sialic acid and galactose residues resulted in a better uptake. This mannose-terminated human placental glucocerebrosidase was later known as alglucerase (Ceredase). After 6 months treatment, the spleen size was substantially reduced and red blood cell and platelets counts were improved (Furbish et al., 1981). As thousands of placentas were required for the treatment of a single patient, new production systems were designed. Imiglucerase (Cerezyme) is human glucocerebrosidase engineerd in CHO cell line. More than 5,000 patients are treated worldwide with auspicious effects; increase in hemoglobin concentration and in white blood cell and platelets counts, improvements in bone mineralization density and a decrease of biomarkers (Weinreb et al., 2002). Two new versions are available, velaglucerase-alpha produced in an immortalized human fibrosarcoma cell line (Zimran et al., 2010) and taliglucerase-alpha produced in plant cells (Shaaltiel et al., 2007).

I.2.5.5 Mannose-6-phosphate function and recognition

The formation of Pi-6-Man is catalysed by the action of two enzymes. First a N-acetylglucosaminyl-1-phosphotransferase, which transfers a N-acetylglucosamine-1-phosphate (GlcNAc-1-phosphate) from uridinediphosphate-GlcNAc (UDGlcNAc) to C6 hydroxyl groups of mannose, generating phosphodiester forms (Lazzarino and Gabel, 1989). The GlcNac-1-phosphotransferase is a hexameric complex of tree subunits which are encoded by two genes. The α/β subunits recognize a common conformation on the lysosomal enzymes (Lee et al., 2007). Although the role of the γ subunit is still unclear, one suggests that it could have a possible role in efficiency. The second enzyme is N-acetylglucosamine-1-phosphodiester α N-acetyl glucosaminidase, localised in the trans-Golgi network, which hydrolyses the phosphodiester. This results in exposed Pi-6-Man residues (Rohrer and Kornfeld, 2001).

Two types of Pi-6-Man specific receptors recognize the Pi-6-Man residues on lysosomal proteins, both are members of the P-type lectins (Dahms and Hancock, 2002). A 46 kDa or cation dependent Pi-6-Man Receptor (CD-M6PR) and a 300 kDa or cation independent receptor (CI-M6PR) were identified. The CD-M6PR shows enhanced ligand binding in the presence of divalent cations. Their common task is the delivery of newly synthesized acid hydrolases from the trans-Golgi network to endosomes. The CI-M6PR has several other functions in physiological processes. It has a role in insulin-like growth factor 2 (IGF-II) degradation, activation of transforming growth factor-β1 and as tumour suppressor. As the CI-MPR and CD-MPR binds at different pH, the CI-MPR is responsible for the uptake of extracellular enzymes.

The repetitive structure of the luminal domain of the CI-M6PR shows sequence similarity with the luminal domain of the CD-M6PR. The same essential mutations (Hancock et al., 2002) and the cysteine distribution could indicate that they have similar disulfide-bonding and tertiary structures. Two high affinity sites are found within the CI-M6PR whereas only one in the CD-M6PR but it exists primarily as a dimer (Figure 6) (Ghosh et al., 2003). The X-ray crystallographic structure of CD-M6PR is known in unligated state and in complex with Pi-6-Man (Roberts et al., 1998). Each monomer consists of an N-terminal α-helix and nine anti-parallel β-strands forming two β-sheets. A flattened β-barrel structure is formed due to the hydrophobic interactions between the β-sheets (Figure 5). The six cysteine residues form three intramolecular disulfide bonds that are essential for a correct conformation. The terminal Pi-6-Man residue is buried in the receptor, as the carbohydrate-recognition domain lies relatively deep. There is a great structural difference between the Pi-6-Man bounded and unbounded state, which indicates movement between the two states (Olson et al., 2002).

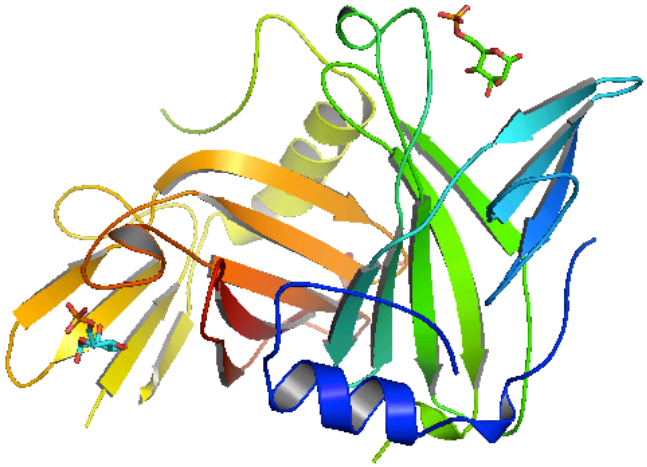


Figure 5 Structure of CD-M6PR in Pi-6-Man bounded state. Each monomer consists of a β-barrel (nine anti-parallel β-strands forming two β-sheets) and an N-terminal α-helix.

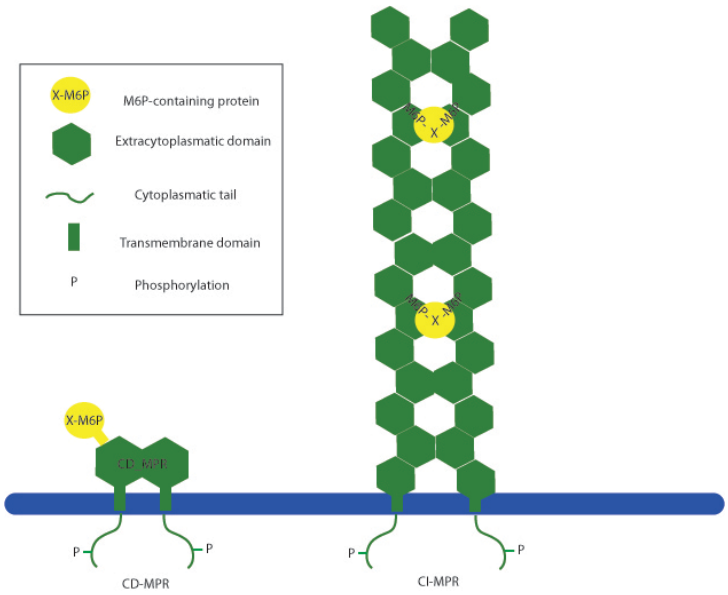


Figure 6 Cation dependent and independent Mannose-6-phosphate receptor. Left: CD-MP6R existing out of two monomers and bound to a protein with the Pi-6-Man N-glycan. Right: CI-MPR existing out of 15 repeating structures. Interaction of two monomers is guided by proteins wit Pi-6-Man N-glycan. The cytoplasmatic tail of both receptors contains sorting signals.

The M6PRs have in their cytoplasmatic tails different sorting signals for lysosomal transport. The synthesis of the Pi-6-Man-recognition signal is complete in the late-Golgi compartments. This allows that the lysosomal hydrolases bind to M6PR and exit the trans-Golgi network (TGN) in clathrin-coated vesicles (CCV) (Campbell and Rome, 1983). Efficient sorting depends on an acid-cluster-dileucine (AC-LL) near the carboxyl-terminus of the cytoplasmatic tails of the receptors (Johnson and Kornfeld, 1992). This AC-LL motif could bind to the CCVs by multiple binding sites (Honing et al., 1997). After budding of the CCVs to the early endosome, the clathrin is released. The M6PR

containing vesicles are then transported to the cell periphery along microtubule tracks. M6PRs are recycled to the Golgi when they delivered the protein in the endosome. There is an indication that the dileucine motif is essential for recycling (Tikkanen et al., 2000). The CI-MPR could bind a diverse group of M6P-containing ligand and transport them to the cell surface.

I.2.5.6 Developed solutions for problems with ERT

First, solutions for better distribution were explored. Genzyme patented in 2003 an *in vitro* approach based on two different enzymes to get a better distribution toward the lysosomes. GlcNAc-phosphotransferase transfers acetylglucosamine-1-phospate from UDP-GlcNAc to the 6 position of a 1,2-linked mannose. The second is a phosphodiesterase that removes the N-glucosamine after the previous reaction. An alternative approach is carbonyl-coupled chemistry to generate an oxime bond using an aminooxy-derived glycan (Figure 7). First the hydrazine reactive group was converted to an aminooxy reactive group, followed by deprotection. The purified aminooxy ligand was then coupled to periodate-oxidized recombinant human acid α-glucosidase (rhGAA). This resulted in changed pharmacokinetics and dynamics profiles and a higher efficiency. Another approach to enhance distribution is based on a heterodimer of two proteins. The first domain comprised a cell-penetrating peptide, which facilitated active transport across a biological membrane and into the lysosome, and the second domain comprised a lysosomal enzyme (Urbanelli et al., 2011; Zhu et al., 2009).

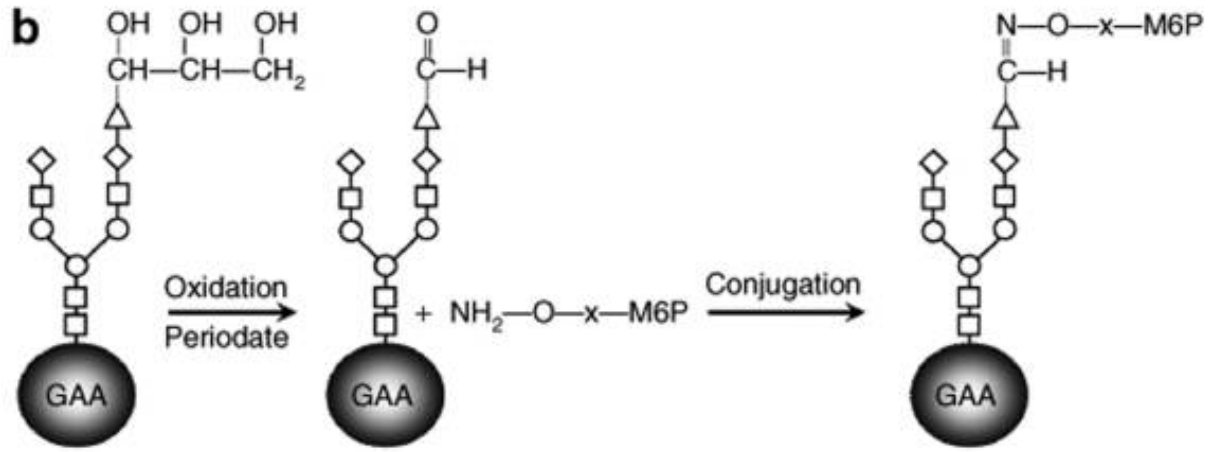


Figure 7 Conjugation of synthetic glycan to rhGAA. The sialic acids on the enzyme were oxidized with periodate before reacting with the reactive group (aminooxy) on the synthetic glycan to generate oxime-neo-rhGAA. Symbols in diagram: open circle, mannose; open square, *N*-acetylglucosamine; open diamond, galactose; open triangle, sialic acid. GAA, acid α-glucosidase; M6P, mannose 6-phosphate; rhGAA, recombinant human acid α-glucosidase. (Zhu et al., 2009)

Second, to reduce costs expression in plants is explored. Plants have some advantages over prokaryotic systems, namely the correct processing of eukaryotic gene products. A major problem is that plant cells do not contain a lysosome but a vacuole without M6PRs. They succeeded to produce active recombinant enzymes in tobacco leaf cells and seed storage tissue (Fogher, 2003).

I.3 Enzyme Replacement Therapy in yeast

A better production platform for ERT could be yeast because of its higher yields and its cost effectiveness (Demain and Vaishnav, 2009). Production of the enzymes in yeast could reduce the cost dramatically so that all patients over the world could be treated with the same budget as used to treat 10% or so of the patients in the developed world, as is the situation now. However, there are some problems that must be solved. Nico Callewaert's group did already a big effort to produce the enzymes in *Pichia pastoris* and *Yarrowia lipolytica*, both generally recognized as safe (GRAS) organisms. They made strains that produce multi-mg quantities α-galactosidase for the treatment of Fabry disease and of α-glucosidase for the treatment of Pompe disease in cooperation with Oxyrane. This was achieved, for α-galactosidase, by co-overexpression of HAC1 transcription factor (manuscript submitted for publication).

I.3.1 Advantages and disadvantages of yeast

Yeast is eukaryotic, which implicates that it is able to perform post-translational modifications such as disulfide bonds and eukaryotic N-glycosylation. This is essential for correct folding of many proteins so that there is an overall success rate of correct folding for more than 60% for eukaryotic proteins. The generation of an expression strain is possible in two weeks, which is partially due to the fast generation time of two hours. Yeast does not secrete much endogenous proteins in the medium so that purification of secreted proteins is easy. But there is often proteolytic activity present in the medium. This can be prevented by using protease-deficient strains or a medium pH far from the optimal pH for protease activity. In *Pichia pastoris* the tightly regulated Alcohol Oxidase I (AOXI) promoter often used for recombinant protein expression as it produces 30% of the cellular mRNA rate when it is switched on (Callewaert, 2011; Cregg et al., 1989; Yurimoto et al., 2011). Heterogeneous glycosylation was another problem but this was solved by the use of glyco-engineered strains.

I.3.2 N-glycan mimic in Yeast

Yeasts do not use Pi-6-Man based targeting for lysosomal protein sorting but they produce quite similar Man-Pi-6-Man on their secreted and cell wall-associated proteins. So a strain was developed, which overproduces this Man-Pi-6-Man glycotype. To mimic the human glycosylation as best as possible, the *OCH1* locus was inactivated. This resulted in a *Pichia pastoris* and a *Yarrowia lipolytica* strain with $Man_8GlcNAc_2$ N-glycans (De Pourcq et al., 2012; Vervecken et al., 2004). It has been shown that a higher phosphorylation rate could be obtained by overexpression of the MNN4 protein in *Yarrowia lipolytica* and PNO1 in *Pichia pastoris* (Akeboshi et al., 2009; Odani et al., 1996). Overexpression of *MNN4* in the strains induced at least one Man-Pi-6-Man glycotype in more than 80% of the protein-linked N-glycans (Figure 8). Such high levels are needed for efficient ERT. The obtained N-glycans have an endstanding mannose that prevents binding with the M6PR, these need to be removed.

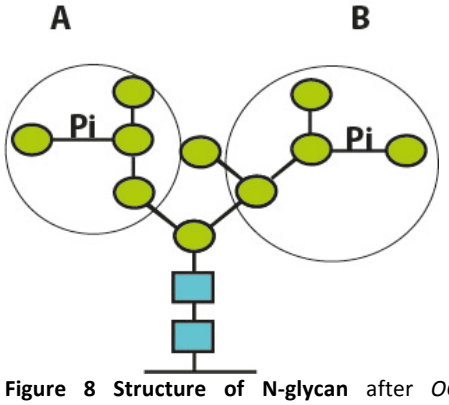


Figure 8 Structure of N-glycan after *OCH1* inactivation and *MNN4* overexpression. A and B are regio-isomers. Green: mannose, blue: GlcNac

I.4 CcMan5

Arthrobacter luteus (currently classified as *C. cellulans)* enzyme cocktails are used for DNA transformation of spheroplasting yeast cells because it is able to degrade cell walls. The group of Nico Callewaert realised that this cocktail should contain an enzyme that is able to remove the capping mannose from phosphorylated N-glycans, as these are abundant in yeast cell walls. They also realised that the enzyme should be similar to α-mannosidases. These belong to one of five glycosyl hydrolase (GH) families. After sequencing the *C. cellulans* genome and a BLAST search with representatives from the GH families, five GH92 genes are found. These five proteins were expressed, after codon optimization, in *E. coli*. *C. cellulans* mannosidase *5* (CcMan5) was found to posses no α-mannosidase activity but only mannosylphosphohydrolase activity. This makes it very specific to uncap the endstanding mannose from Man-Pi-6-Man moieties. CcMan4 had α-1,2-mannosidase activity on both neutral and Man-Pi-6-Man substituted N-glycan branches. The combination of CcMan4 and CcMan5 can generate yeast N-glycans with terminal Pi-6-Man and results in much higher densities of the lysosome targeting Pi-6-Man glycotype than on current therapeutic enzymes.

As in the other GH92 members, CcMan5 catalysis involves Ca^{2+}, which assists in distorting the mannoside away from its ground state towards the transition state (Zhu et al., 2010). The N-terminal domain houses the GH92-homologous catalytic activity and the C-terminal domain houses three carbohydrate-binding modules. The X-ray structure of the CcMan5 catalytic domain (1-771) has been determined. The protein consists of two domains, an N-terminal β-sandwich domain (8-271) and a C-terminal (αα)6 barrel domain (291-771), connected by an α–linker (272-290) (Figure 10). The conserved Ca^{2+} ion is placed on the interface between both domains and octahedrally coordinated by the carbonyl oxygen of Asn588, Glu589 and Asp662. Treatment of *Yarrowia* and *Pichia*-produced lysosomal enzymes with CcMan5 and the α-1,2-mannosidase CcMan4 (also cloned from *C. cellulans*) resulted for α-glucosidase in higher Pi-6-Man levels compared to Myozyme®. The uptake by human Pompe fibroblast was 18(+/-7) times more efficient than with Myozyme® (Callewaert et al. manuscript submitted for publication).

The uncapping of the Man-Pi-6-Man to Pi-6-Man is an essential step in the process of producing N-glycans that could be recognized by the M6PR. The phosphate bound to the anomeric oxygen constitutes a strong leaving group. Tyr626 and the amide of Gly72 are able to donate a H-bond to the non-glycosidic oxygens in the phosphate. CcMan5 is an effective catalyst for this uncapping, but one of the regio-isomers of the phosphorylated N-glycans is a markedly slower substrate (A) than the other regio-isomer (B) (Figure 8 and Figure 9). One of the problems in manufacturing these enzymes in *Yarrowia lipolytica* is that this yeast also secretes endogenous phosphatases, which can remove the phosphate moiety uncovered by CcMan5. Therefore, it is important that CcMan5 can operate more rapidly also on its less-preferred substrate, to minimize the side reaction due to dephosphorylation. Preliminary docking of both phosphorylated N-glycan branch structures indicates that both regio-isomers have a different binding mode on the enzyme. The faster-converted isomer is more complementary to the catalytic site than the more slowly converted one.

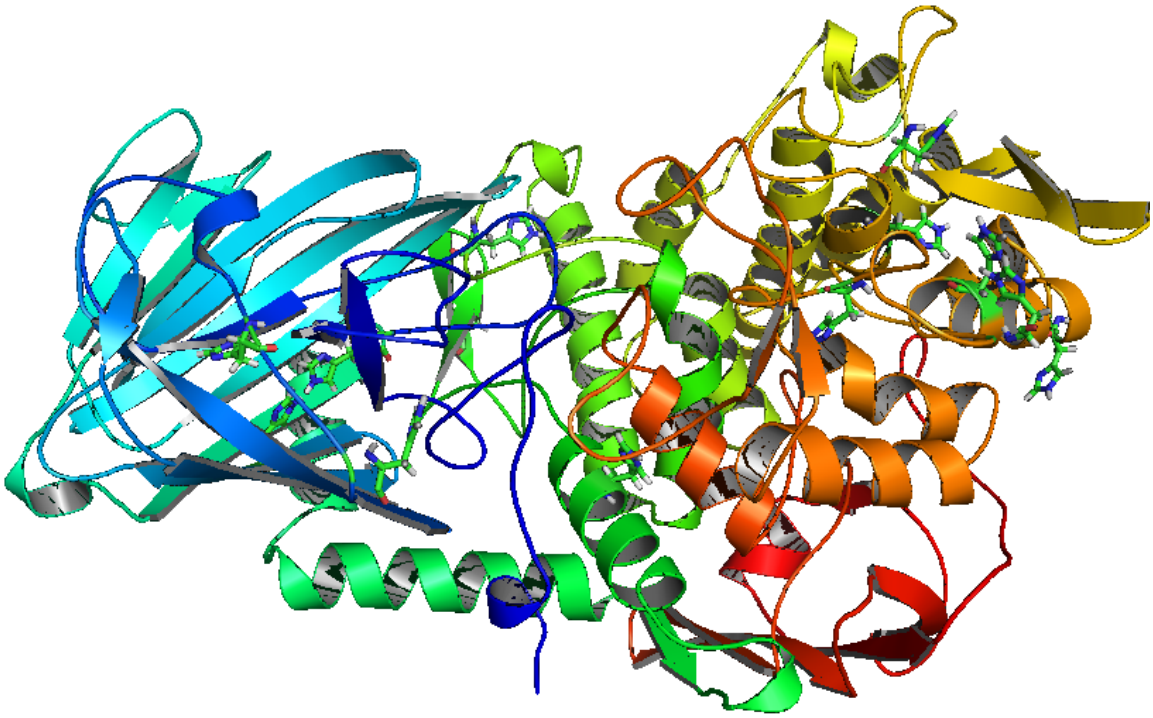


Figure 10 Structure of CcMan5 experimentally determined by X-ray. N-term β-sandwich domain (8-271) and a C-terminal (αα)6 barrel domain (291-771), α–linker (272-290).

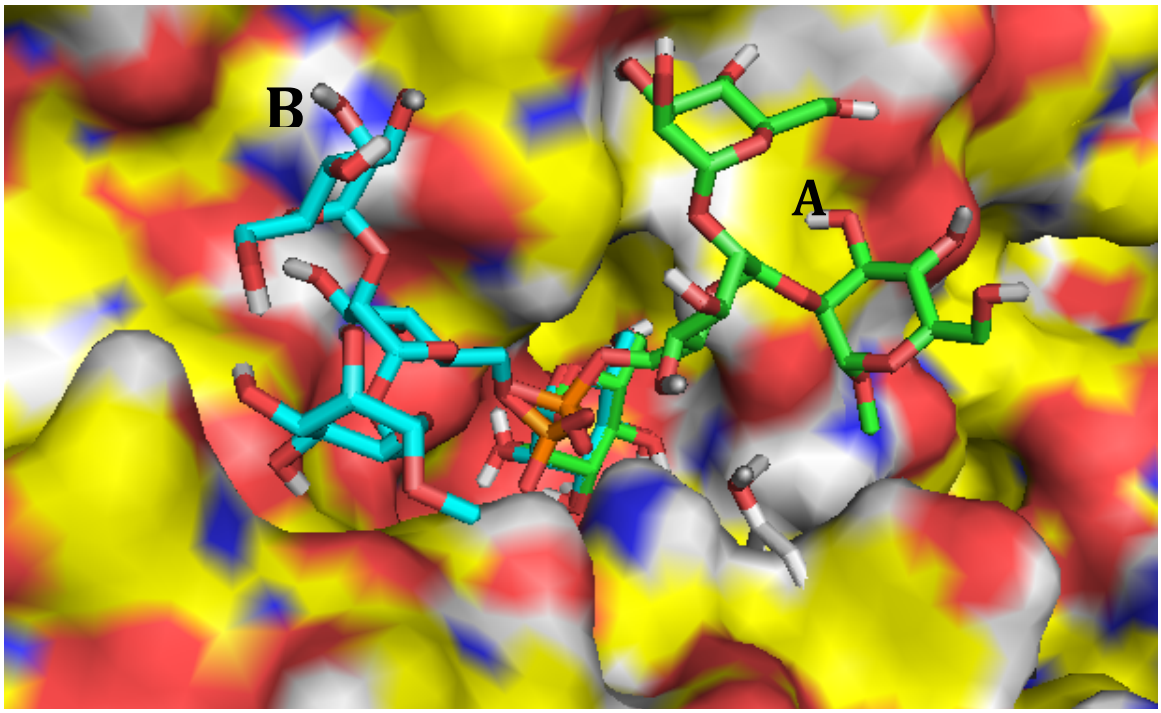


Figure 9 Docking of two parts of the N-glycan both regio-isomers, slower (A,green) and faster (B,blue), have a different orientation at the surface.

I.5 Molecular Modeling

Knowledge of the 3D structures of proteins provides insight into their functions, reaction pathways and interactions. The structures are usually determined by experimental methods, such as X-ray crystallography, high-resolution electron microscopy or nuclear magnetic resonance (NMR) spectroscopy. These methods are expensive, time consuming and not always applicable. Complementary to these, information on the structure and dynamics of proteins can be determined by molecular modeling, encompassing a variety of theoretical and computational methods to model the behaviour of molecules.

Molecular modeling is used for the study of small molecules as well as large biomolecules. Each of these techniques usually relies, in one way or another, on a description of the energy of the molecule as a function of its coordinates. The nature of this description depends on the size of the molecule and the purpose of the simulation. At the top end, all atoms and electrons in the molecular system are described via the Schrödinger equation, in principle allowing the prediction of all properties in an *ab initio* fashion. This is done in quantum mechanics simulations, which come in various flavours depending on the approximations adopted: Hartree-Fock simulations, Density Functional Theory calculations, Post-Hartree Fock, … However, these *ab initio* methods are extremely expensive from a computational point of view, requiring lots of integral calculations and effectively limiting the size of the molecular system to about 10^2 atoms. In semi-empirical methods, this limit is extended to about 10^3-10^4 atoms by using a database for recurring integrals in the calculations. To enable simulations on macromolecules such as proteins or DNA, which are much larger, it is necessary to introduce even bigger approximations: electrons are no longer explicitly described in a Schrödinger equation, but instead an empirical parameterized model is used in which their effect is implicitly incorporated in the atomic description, often relying on well-known expressions of classical mechanics. For instance, the many-body electrostatic interaction between nuclei and electrons in a molecule, requiring the construction and evaluation of an electronic wavefunction, is replaced by a much simpler expression of the Coulomb interaction between atomic point charges. These approximations are a lot faster and sometimes even better describe reality because of the empirical parameterisation (Kukol, 2008). As such, the computational limit goes up to easily 10^8 atoms.

However, to accomplish this, these methods sacrifice the universal applicability of quantum mechanics: each molecular system needs to be described by a specific (potential) energy function, suitable only for that type of molecular system. The functional form of this energy function in terms of nuclear coordinates is called a *force field*. It consists of equations used to calculate (potential) energy and forces from nuclear coordinates only. These equations are usually subdivided in bonded interactions (bond length, angle, dihedral and improper terms) and nonbonded interactions (Coulomb and Van der Waals interactions). A typical example of such a functional form is the TIP3P force field for water (Jorgensen et al., 1983).

$$E = \sum_{b}^{bonds} K_b(b - b_0)^2 + \sum_{\theta}^{angles} K_\theta(\theta - \theta_0)^2 + \sum_{i,j}^{nuclei} \left\{ \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right\}$$

As they are tailor-made, different force fields cannot be combined unless they are specifically suited to one another. Additionally, chemical reactions cannot be described by (classical) force fields, since they involve bond formation or breaking. The application of a non-reactive force field, however, requires the definition of a constant topology, explicitly outlining the bonds present in a molecule.

In principle, molecular modeling methods using force fields could even predict the 3D structure of proteins based on the sequence alone, potentially closing the gap between the numbers of released structures and annotated protein sequences. However, these methods require vast computational resources to perform this treat (Lindorff-Larsen et al., 2011) and are therefore often abandoned in favour of more pragmatic, database-driven approaches such as Rosetta (Leaver-Fay et al., 2011) or iTasser (Roy et al., 2010). Still, several other methods in the molecular modeling panoply can yield extraordinary insight in the structure, function and dynamics of a protein, particularly when an initial or homologous 3D structure of the protein is already known. In this work specifically, molecular dynamics and docking simulations are used to get information on the flexibility, interaction and relative movement of a protein and its ligand. Insight in the reaction pathways of a protein can be obtained with the aid of quantum mechanics/molecular mechanics (QM/MM) simulations.

I.5.1 Molecular Dynamics Simulations

Molecular Dynamics (MD) is a very powerful toolbox in modern molecular modeling. It enables us to follow and understand structure and dynamics with detail. Other than in energy minimization, where the potential energy of a molecule is locally minimized by varying its nuclear coordinate, kinetic energy is added to the equation in MD. This allows the molecule to move at a certain temperature and surpass barriers in the conformational landscape. As such, MD can be used to find a global minimum on the potential energy surface. This structure – or rather ensemble of structures – represents the true conformation of a molecule that would be observed experimentally. In addition, this technique gives us a keen insight on the conformational dynamics of a molecule, allowing weak or flexible regions to be identified.

In MD, essentially, Newton's equations of motion are iteratively solved numerically over finite timesteps. Given an initial set of atomic coordinates and velocities (related to the instantaneous temperature through Boltzmann's equation), the algorithm starts by calculating the potential energy and corresponding atomic forces from an energy function, such as a force field, a quantum mechanics method or even empirical scoring function. Assuming the acceleration due to these forces (Newton's second equation) remains constant for just a brief timestep, the nuclei can now be propagated over that timestep, yielding new coordinates for the nuclei and associated new velocities. Starting from that point, the atomic forces can again be calculated, after which further propagation is possible, and so on. Usually, Verlet integration (Frenkel and Smit, 2002) is used to propagate Newton's equation of motion, as outlined from the equations below:

$$\underbrace{x_n \qquad v_n}_{\downarrow}$$

$$v_{n+1/2} = v_n + \frac{\Delta t}{2}\frac{F_n(x_n)}{M}$$

$$x_{n+1} = x_n + \Delta t \cdot v_{n+1/2}$$

$$v_{n+1} = v_{n+1/2} + \frac{\Delta t}{2}\frac{F_{n+1}(x_{n+1})}{M}$$

$$\underbrace{\downarrow \qquad \downarrow}_{x_{n+1} \qquad v_{n+1}}$$

The choice of the timestep in this algorithm must be sane and cannot be too large, otherwise the assumption that the atomic acceleration remains constant will break down. For most molecular systems, a timestep of 1 fs is commonly used. This is 10 orders of magnitude smaller than the period of the fastest molecular motion: bond stretch involving hydrogens. In order to simulate molecular motion with a longer period, MD simulations must be conducted for a multitude of timesteps. If we assume a protein to fold within 1 s, this would require 1.000.000.000.000.000 (one quadrillion) MD timesteps. Given that it usually takes more than one thousand of a second in real-time on an averaged-size computer to simulate just one timestep of 1 fs for a typical protein system, one can easily see that an MD simulation of one protein folding event quickly becomes computationally daunting (about 30.000 years). And obviously more than one folding event needs to be simulated in order to have any statistical meaning. Still, the advent of powerful supercomputers and advances in molecular modeling theory (Shaw et al., 2008) have brought protein folding simulations within reach (for very few researchers though).

I.5.2 Docking

Docking is the simulation of the protein and the ligand in solution that predicts the preferred orientation of a ligand on a receptor, given their representative initial structures. Docking methods are often used in drug design to perform virtual screens on large libraries of compounds. It gives insight in how the ligand inhibits the target, which is of great importance for lead optimization. It is a faster and cheaper screening method than *in vitro* screening. Usually docking programs consist of a scoring function and a docking algorithm. The scoring function determines the most optimized model and ranks candidate dockings, while the docking algorithm determines structures by a conformational search.

The scoring function tries to quickly solve the problem of estimating the free binding energy of a protein-ligand complex, which is actually an extraordinarily complex task. Scoring functions are therefore often based on empirical or statistical observations, although force fields can also be used.

Search algorithms can be divided into two main categories: systematic and stochastic. Stochastic methods rely on randomness. They are more suited for higher-dimensional problems, such as flexible ligand docking. Systematic methods have less degrees of freedom

and are often used for rigid docking. Although usually all degrees of torsional freedom in the ligand are explored, this is computationally intractable for the receptor. The protein receptor could be rigid or some side-chains are allowed to be flexible. The first one is less accurate but computationally less intense. Flexibility is a very important aspect of molecular recognition. A receptor may adopt different conformations in the unbound and bounded states. To accelerate the scoring calculation, some docking methods use a grid map. This limits the search space on the protein surface to a defined region by the user.

II Aims

At the moment enzymes for ERT are produced in mammalian cells but the lab of prof. Nico Callewaert wants to produce these enzymes in yeast, as it could reach higher yields and it is more cost effective. But the in yeast produced enzymes have N-glycans distinct from mammalian N-glycans, so that the distribution to the lysosomes is blocked. CcMan5 has an important role in the production process of enzymes for ERT, as it modifies the N-glycans on the produced enzymes. CcMan5 activity results in uncapping of the Man-Pi-6-Man to Pi-6-Man, so that these are recognized by the M6PR. More double uncapping of the N-glycan would lead to a better M6PR recognition on the lysosome surface and improved uptake of the enzymes by the patients so that lower quantities of therapeutics are needed. Besides the higher conversion rate, the incubation time of the yeast cells with CcMan5, after purification, could be lower so that phosphatases have less time to degrade the uncapped Pi-6-Man on the enzymes used for ERT.

In this master project we want to engineer the CcMan5 enzyme so that both Man-Pi-6-Man N-glycan branches are uncapped faster and that the overall conversion rate is higher. This all could have a huge impact on the ERT efficiency but could reduce the ERT production cost as well. With the mutations we want to enhance the binding of both branches of the N-glycan on the enzyme surface by creating more space, as well as enhancing the activity of CcMan5 to the difficult-to-uncap branch. The mutations should give us more information about how CcMan5 works and how the N-glycan binds the protein. We will adopt combined experimental-theoretical approach to achieve this.

Molecular modeling will be used to understand how the two different N-glycan branches preferentially bind to the CcMan5 protein. It will also be used to explore different possible mutations that would result in enough space and binding stability for both branches. We will try to simulate these mutations by using modeling techniques, giving us the possibility to analyze many mutations simultaneously, which makes it cost and time effective. For each mutation we will perform energy minimization and molecular dynamics. These techniques should give us an indication for the stability of the proteins, but could also denote which residues are flexible. If the mutations are stable, docking will be performed. The energy-based docking score is the factor to rank the mutations and compare them with the native protein.

Some of these mutants will also be produced in the wet lab. We will first generate the mutations in the expression plasmid using site directed mutagenesis. After which expression of the mutated enzymes will be checked in *E. coli*. In a final step we will purify the different CcMan5 enzymes and test their activity on MNN4 glycans using DSA-FACE (DNA Sequencer-Aided Fluorophore-Assisted Carbohydrate Electrophoresis). Based on the activity of dilution series of the enzymes, we can conclude whether the mutants are more or less active than the native protein.

The obtained results can then be used to develop an improved version of the CcMan5 enzyme, which will greatly enhance the efficacy of the produced enzymes for ERT in our yeast.

Contribution of third parties

Construction pLSAHCcMan5domain plasmid: Petra Tiels

Sequencing: VIB core facilities

DSA-FACE analysis: Petra Tiels

Purification: Katrien Claes

III Results

The project consists of an experimental and a computational part. Both parts were performed simultaneously. The experimental part was executed at the lab of Prof. Nico Callewaert. The computational part was in co-operation with Dr. Ewald Pauwels of the Center for Molecular Modelling.

III.1 Experimental part

III.1.1 Defining mutants

For this project, we are working on CcMan5, which possesses mannosylphosphohydrolase activity. This makes it very specific to uncap the endstanding mannose from Man-Pi-6-Man moieties. If we speak about CcMan5, we mean the active domain, which is functional per se. Because the computational and experimental part was initiated at the same time, we started to make some mutations based on the structure of the protein, previous mutations, a preliminary docking study and an available alignment of the GH92 family. There were already some previous mutations available, which gave insights in the reaction mechanism. These allowed us to define the extended binding site of the substrate but they all resulted in an enzyme with less conversion efficiency. Based on the structure we will start by making two points mutations, i.e. Ala351Leu (Ala351) and Ala631Leu (Ala631). The first one is lying in the active site, near the calcium. The Ala631 is part of the conserved sequence between the CcMan5-like proteins. With this mutation we will try to discriminate whether the N-glycan binding site extends to this area of the protein, as there are doubts about how the N-glycan binds. We visualised Ala351 and Ala631 with PyMOL (Figure 11 and Figure 12). We will mutate both alanines to a leucine, which has more volume and this should disrupt binding of the N-glycan to this part of the protein.

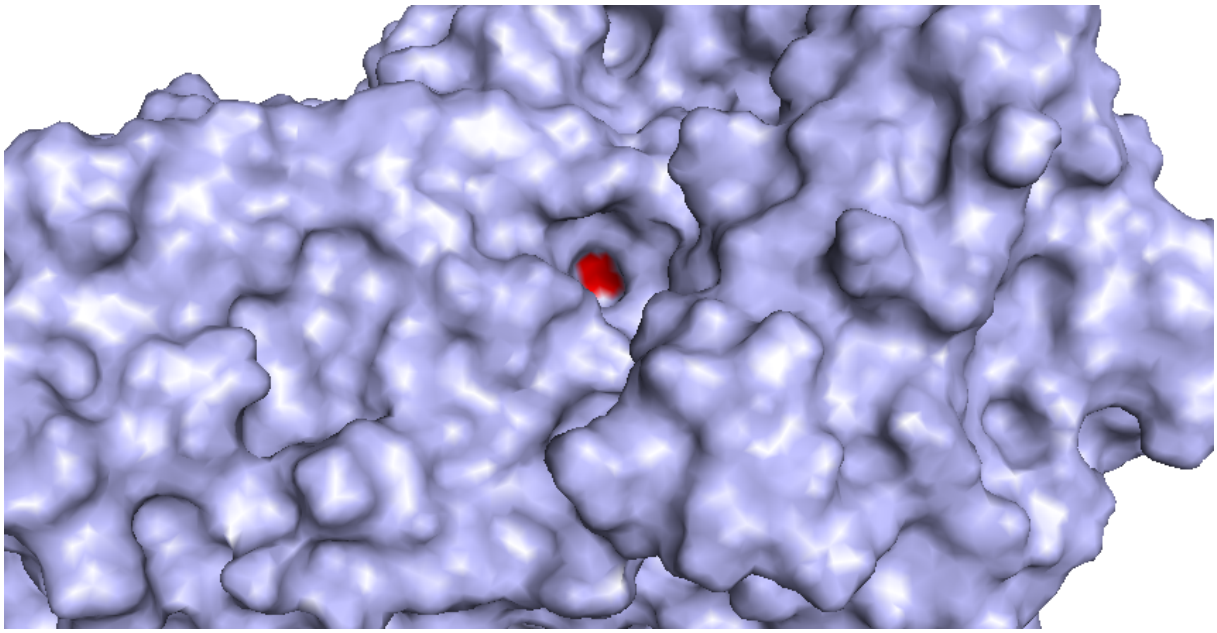


Figure 11 Surface visualisation of Ala351 position blue=CcMan5 protein; red= Ala351

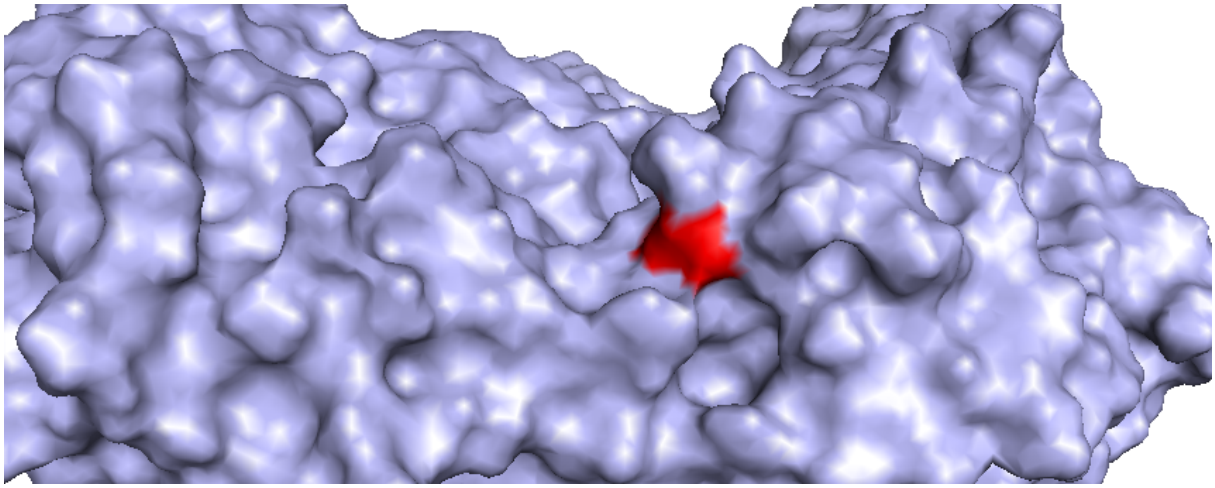


Figure 12 Surface visualisation of Ala631 position blue=CcMan5 protein; red= Ala631

The GH92 family members have a largely conserved fold and catalytic site, but differ in loop structures. These differences in the loops likely help to determine the substrate specificity. Based on the alignment of the GH92 family (Figure 14) we saw that there were differences in the loop length between amino acid 615 and 642. When we looked to the dockings, it seemed that the slow converting part of the N-glycan results in a steric clash with this loop region of the protein. We concluded to make two deletion mutants, a long (Y615-P642) and a short (S634-P642) one (Figure 13).

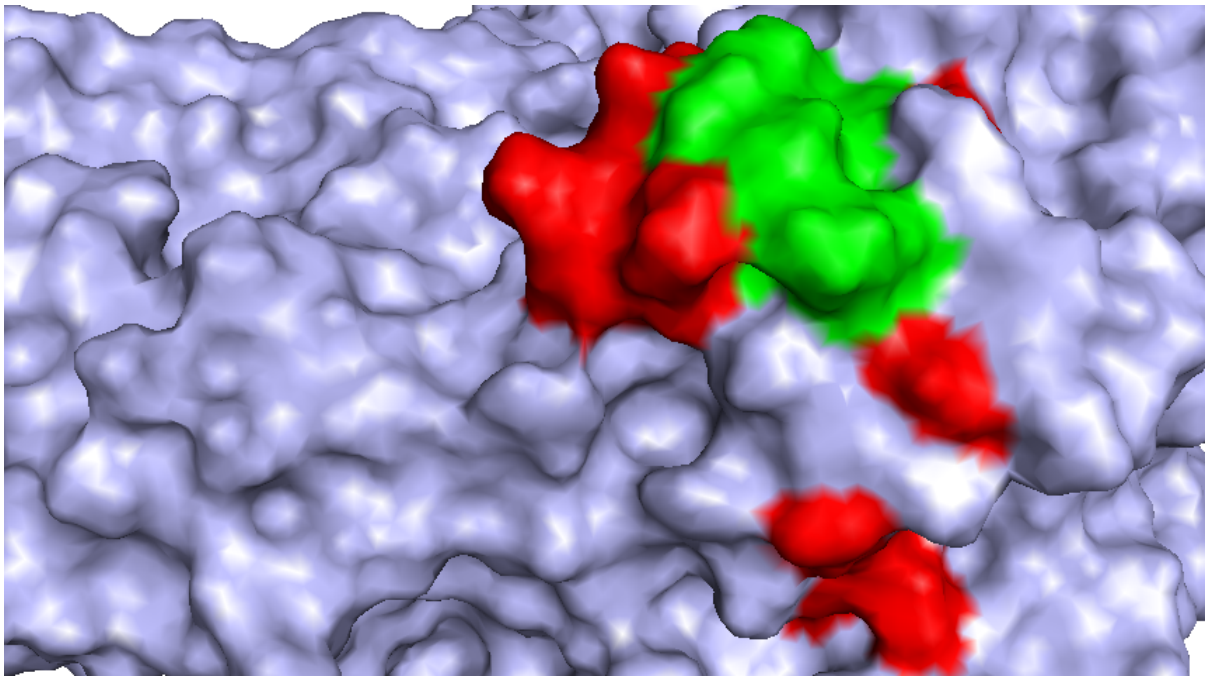


Figure 13 Surface visualisation of deletion mutants blue=CcMan5 protein; green=short deletion of S634 to P642; red + green=long deletion of Y615 to P642

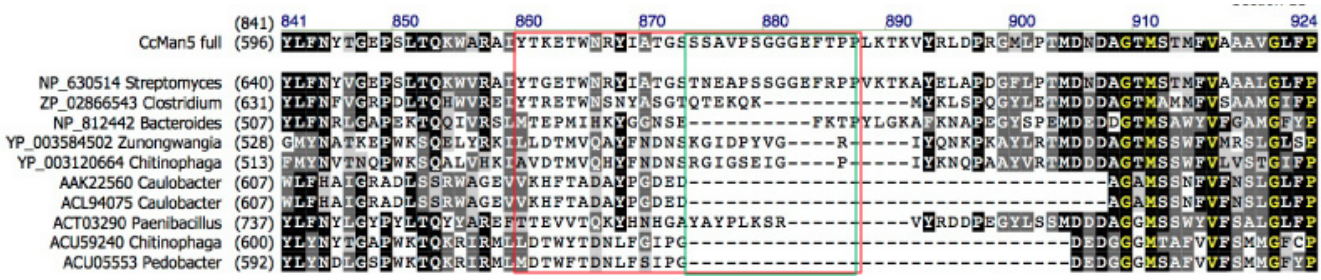


Figure 14 Alignment of the GH92 family, visualisation ot the long (red) and short (green) deletion mutant.

III.1.2 pLSAHCcMan5domain plasmid

A CcMan5domain expression vector (Figure 15) is available for the expression of the CcMan5domain regulated by the pL-promoter of phage lamda. A His-tag is added N-terminally, which can be cleaved of with a caspase as the DEVD amino sequence is available. This tag can be used for antibody detection or purification. There is also a constitutively expressed ampiciline resistance gene available for positive selection. T7 and T3 sites can be used for PCR or sequencing by the respective forward and reverse conventional primers.

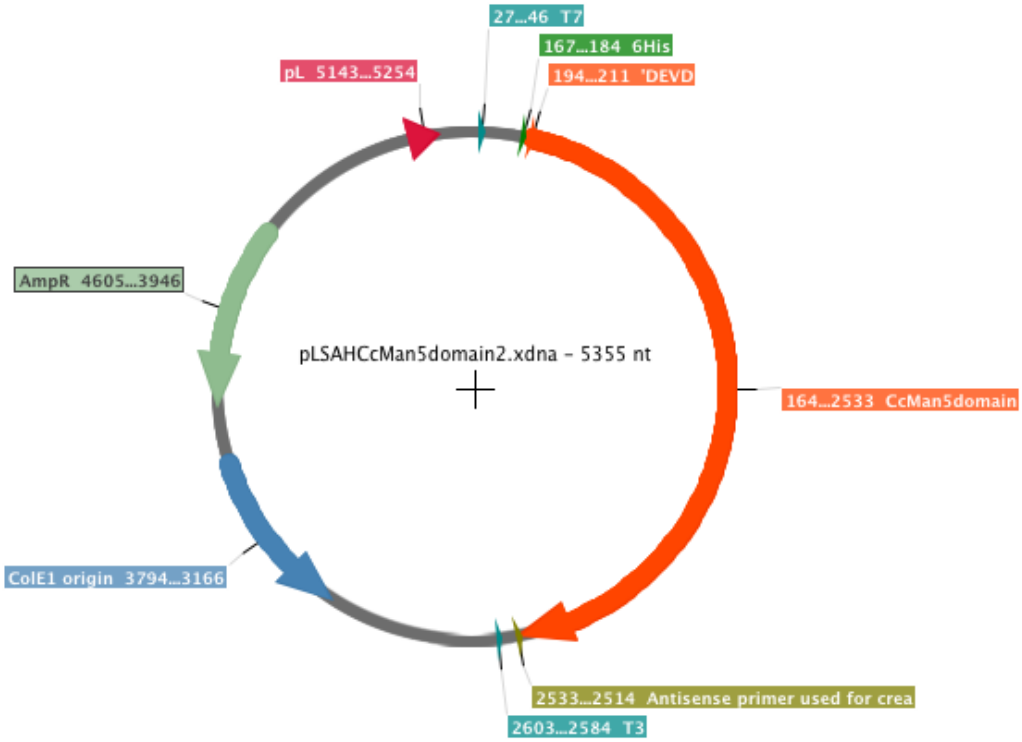


Figure 15 pLSAHCcMan5domain plasmid with features as the N-terminal His-tag, pL-promoter inducible in *E.coli* BL21 with IPTG, ampiciline resistance and ORI for replication in *E. coli*. The plasmid encodes for the CcMan5 domain which possesses the active site. The T7 and T3 sites could be used for PCR or sequencing by respectivily the forward and reverse primers.

III.1.3 Site directed mutagenesis of pLSAHCcMan5 domain plasmid

In this plasmid we started to make the point mutants by using QuikChange site directed mutagenesis (Figure 16). We did this with a one-step and two-step strategy. The one step strategy is the most straightforward one. After PCR with the primers containing the mutation, the original methylated plasmids are destroyed by a DpnI digest. In the two-step strategy we used the combination of the T7 forward primer with a reverse primer with the mutated site and T3 reverse primer with a forward primer with the changed nucleotides for the mutation. After PCR we got small pieces with overlap so that a second PCR reaction resulted in the T7-T3 fragment. After a restriction digest on the T7-T3 fragment and the plasmid, the pieces were ligated resulting in the plasmid with the mutation. To create the deletion mutants we used also two different strategies. The first has primers flanking each side of the deletion and needs ligation afterwards. For the second strategy the primers flank both sides so that there is partially overlap.

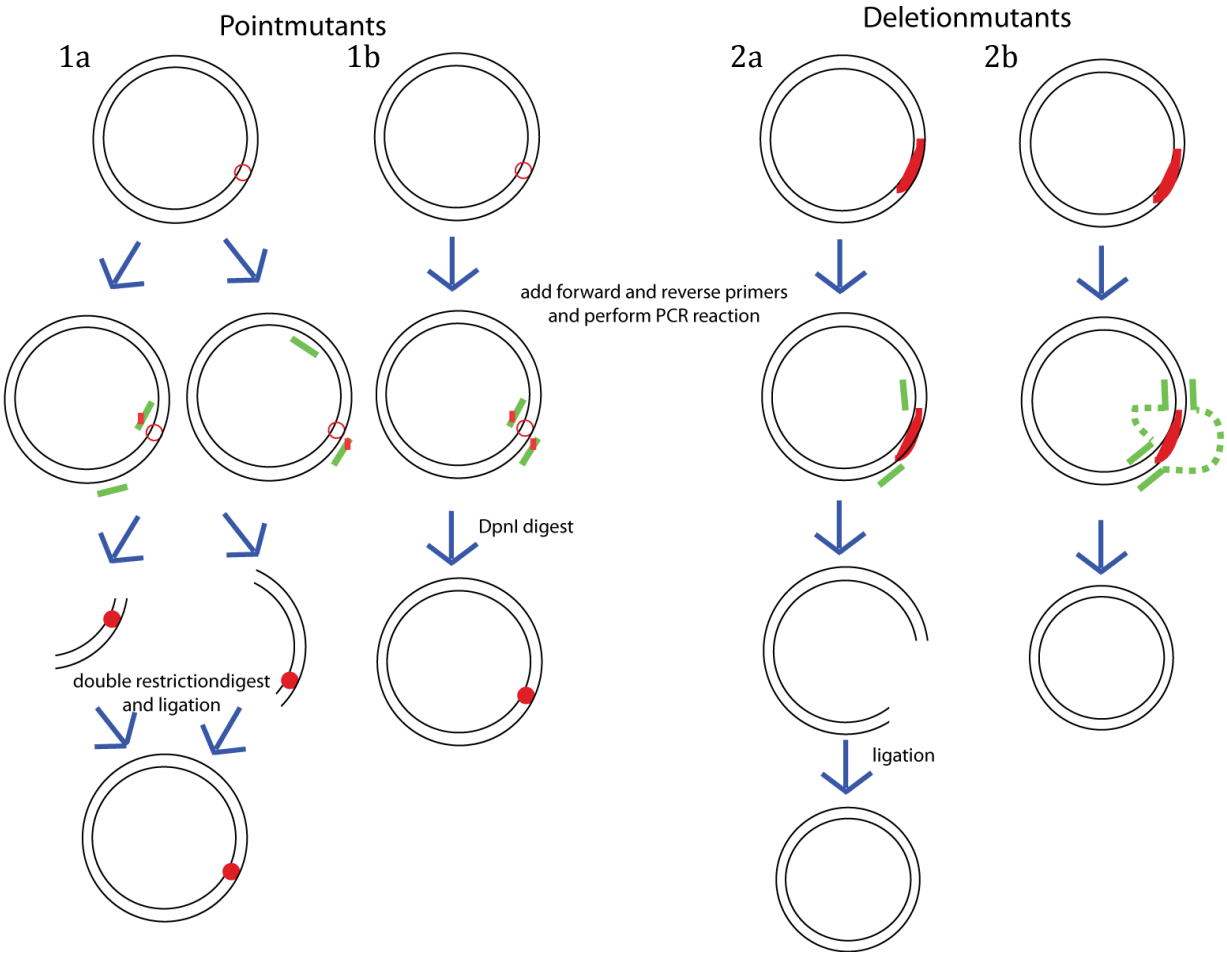
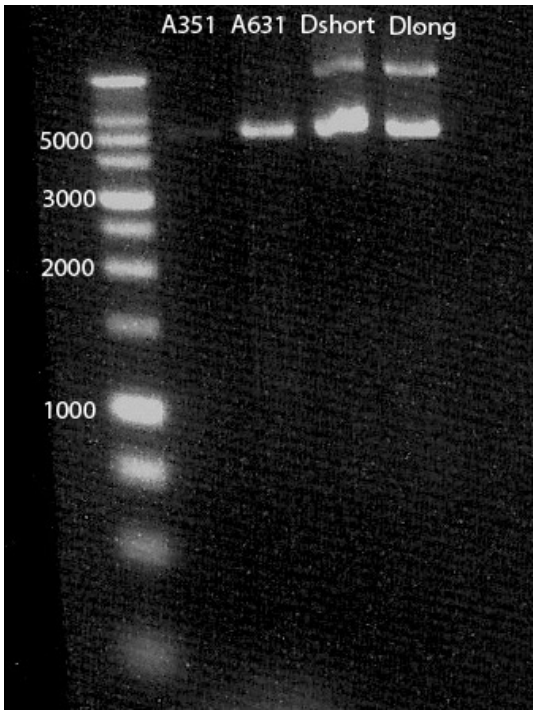


Figure 16 Site directed mutagenesis of point and deletion mutants 1a A two-step reaction to create a point mutant: two small fragments were created by PCR and afterwards a PCR was performed with the outer primers. On the T7-T3 fragment and the plasmid, a restriction digest and ligation reaction was performed, resulting in the plasmid with the mutation. **1b** Quikchange site-directed mutagenesis: a PCR with primers containing the mutation is performed, followed by a Dpn I digest to destroy the original plasmid. **2a** Deletion with flanking primers: each primer flanks one side of the deletion, which create a gap and need ligation **2b** Deletion with overlapping primers: primers flanking both sides of the deletion, so that no gap is created.



By using a high-fidelity polymerase that has 3' to 5' proofreading, we hoped to avoid unwanted mutations in the gene. After testing different PCR protocols, with different temperatures, annealing times and polymerases, we got only one correct Ala631 mutant after sequencing. Although the agarose gel (Figure 17) shows fragments of the desired length, the transformation would not succeed most of the time. The Ala631 mutant wouldn't express so that we recloned a small fragment containing the mutation back into the original vector. This way we could exclude that other regions were mutated resulting in the lack of expression.

Figure 17 Agarose gel for Ala351, Ala631, deletion short and deletion long mutant after site directed mutagenesis based on method 1b for the pointmutants (after DpnI digest) and 2a for the deletion mutants (after T4 ligase reaction).

III.1.4 Synthetic gene synthesis and restriction site engineering

As the site directed mutagenesis didn't work that well, we decided to order the mutated sequences. The amount of unique restriction sites in the CcMan5 gene was low, especially around the region where we wanted the mutation, so we designed a new CcMan5 gene with more restriction sites. This would enable us to make mutants by ordering only a small piece of synthetic DNA. We started with a list of robust restriction enzymes that are available in the lab. For each enzyme we looked after the recognition sequence and if it was a blunt end or sticky end cutter. The blunt end cutters are not as interesting as the sticky ones, because combination of two blunt ends could result in inverse integration of the fragment. Then we searched for similar sequences in the gene. For each possible adaptation we checked if the amino acid stayed the same and if the codon usage was high enough for *E. coli*. This resulted in four new and good restriction sites (Figure 18).

The changed nucleotides are:

- SpeI: 1187-1192: accagc to a^ctagt
  tggtcg tgatc^a
- HindIII: 1497- 1502: aagcat to a^agctt
  ttcgta to ttcga^a
- SmaI: 1573-1578: tccggg to ccc^ggg
  aggccc ggg^ccc
- KpnI: 1820-1825: ggcacc to ggtac^c
  cctcgg c^catgg

We ordered the new synthetic CcMan5 domain gene flanked by the NsiI and BglII restriction sites from GenScript. We also ordered shorter fragments for the Ala351 mutant, between

the SpeI and HindIII restriction sites, and for the long deletion between KpnI and NcoI sites. As the short deletion would have less influence, we decided to test first the long deletion. First we cloned the synthetic gene into our original plasmid, sequenced it and did a control digest with SpeI and HindIII. The deletion mutants were checked by colony PCR with a primer against a fragment of the deletion and the T3 primer. As a positive control another colony PCR was performed with a primer flanking the deletion and the T3 primer. The positive Ala351 point mutant and long deletion mutant were checked by sequencing. After transformation to *E. coli* MC1061λ, minipreps were performed to obtain a larger amount of plasmid.

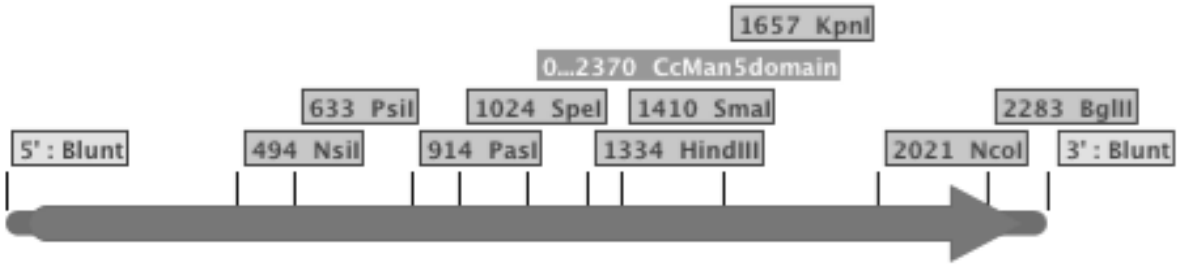


Figure 18 CcMan5 with unique restriction sites. SpeI, HindIII, SmaI and KpnI were newly introduced

III.1.5 Induction of protein expression

As the plasmids with mutations are obtained, the following step is the production of the proteins in *E. coli*. The obtained plasmids were transformed to *E. coli BL21,* which harbours the pICA2 plasmid. This plasmid contains a *LaqIq*-gene, a constitutively expressed repressor for the lac-promoter. The expression of *CcMan5* is induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (patented WO98/48025 (29.10.1998)). IPTG binds the LacIq-repressor, which results in the expression of an antirepressor that binds the cI-monomers. Binding on cI-monomers prevents dimerisation, necessary for repression of the pL-promoter. This results in an active *CcMan5*-promoter and transcription of the *CcMan5*-gene.

As the CcMan5 protein has a N-terminal His-tag, we used this tag to do Western blot. We used mouse anti-His as a primary antibody and goat anti-mouse DyLight 800 as a secondary antibody. Native CcMan5 was loaded as a positive control. Often the induction of our mutants did not succeed, the reason why remains unknown. Finally, we got for each mutant Western Blots with positive result (Figure 19), which means that the CcMan5 mutants could be expressed in *E. coli*.

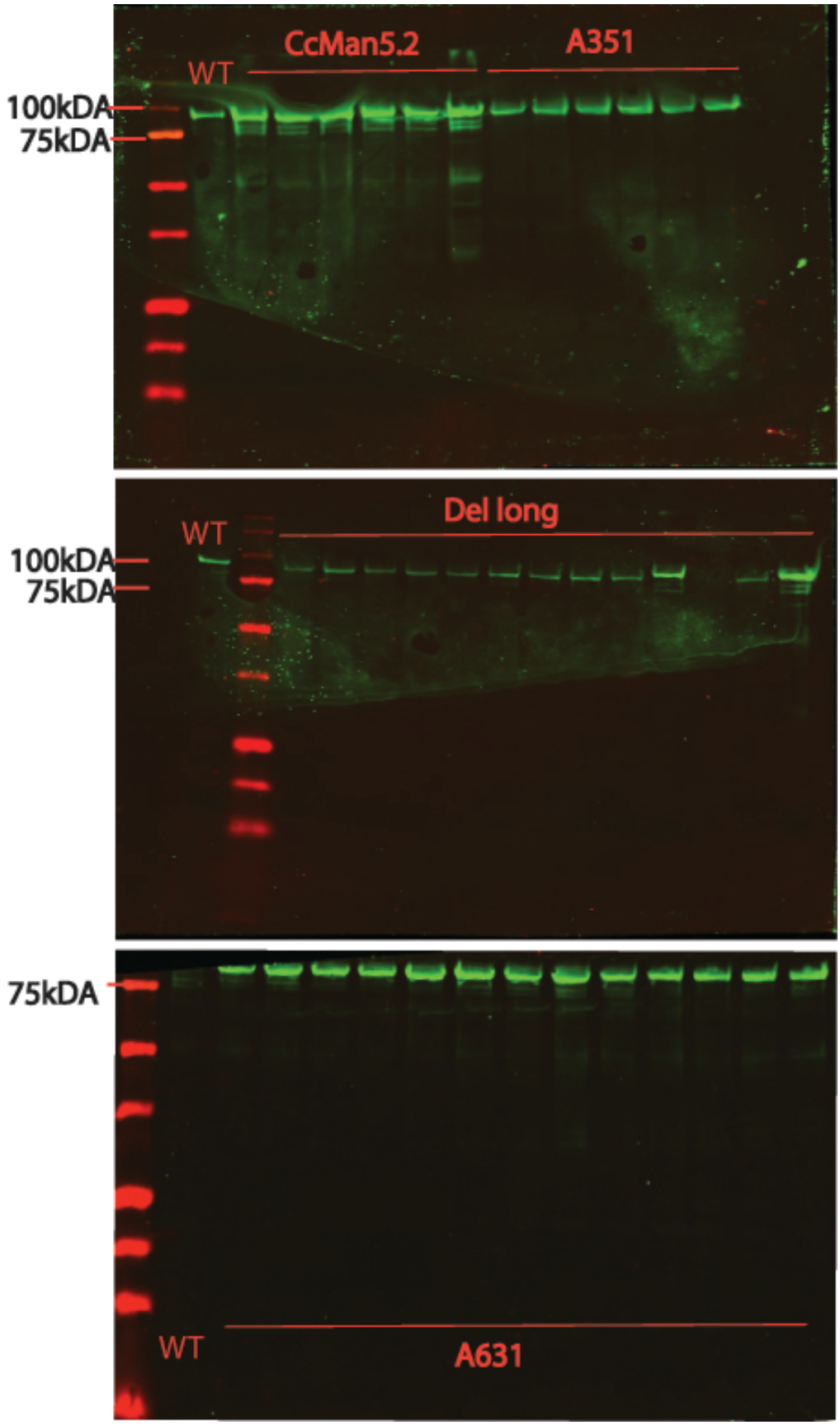


Figure 19 Western Blots after IPTG induction of WT (=Original CcMan5), CcMan5.2 (=restriction sites introduced), Ala351 mutant, long deletion and Ala631 mutant, with in each lane a different clone. The CcMan5 protein has a molecular weight of 83.83 kDa. We used anti-His as a primary antibody and goat anti-mouse DyLight 800 as a secondary antibody

III.1.6 Fraction extraction

For each mutant we inoculated one well-expressing colony in 4 ml LB. Then we isolated the medium, periplasm, water-soluble and insoluble fraction to determine where the CcMan5 protein is located in the cell. The pellets were dissolved in 1.5 ml for the periplasm, 0.4ml for the soluble fraction and 0.45ml for the insoluble fraction so that we loaded respectively 15µl, 2µl and 2.25µl, for the medium we loaded 25µl. We prepared SDS-PAGE gels for Coomassie and Western Blot (Figure 20 and Figure 21).

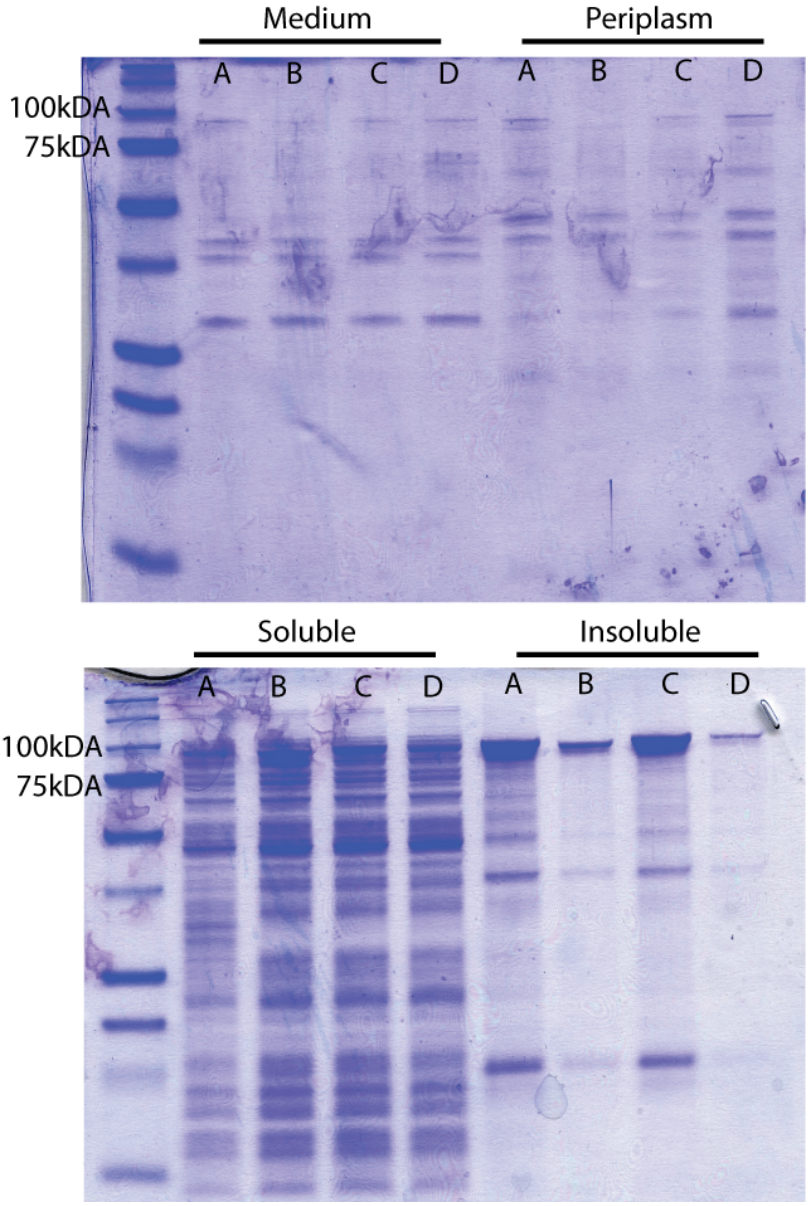


Figure 20 Coomassie gels of different extracted fractions for native CcMan5 (A), long deletion mutant (B), Ala351 mutant (C) and Ala631 mutant (D). For each sample, the extracted fractions are medium, periplams, soluble and insoluble fraction. The CcMan5 protein has a mocular weight of 83.83 kDa.

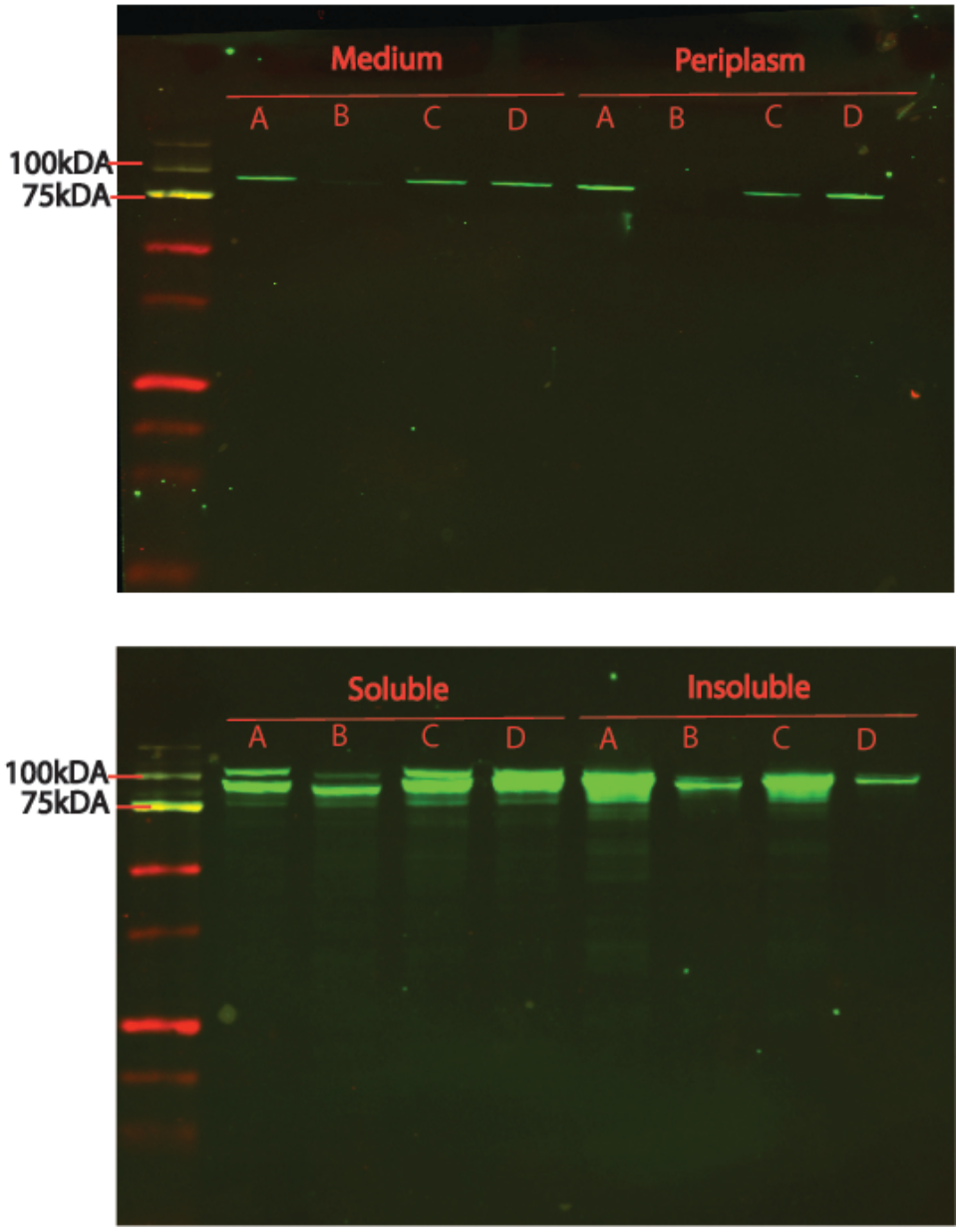


Figure 21 Western blot with mouse anti-penta His and goat anti-mouse Dylight 800 of different extracted fractions for native CcMan5 (A), long deletion mutant (B), Ala351 mutant (C) and Ala631 mutant (D). For each sample, the extracted fractions are medium, periplams, soluble and insoluble fraction. The CcMan5 protein has a mocular weight of 83.83 kDa.

The amount of CcMan5 in the medium and periplasm is substantially lower than in the water-soluble and insoluble fraction. The amount of the deletion mutant in the periplasm is lower than the amount of the native and deletion mutants. The soluble fraction contains more contaminant proteins than the other fractions. This makes that this fraction cannot be used for activity tests without His-tag purification.

III.1.7 Activity assay

For the determination of the CcMan5 activity we used glycans from a modified *Yarrowia lipolytica* strain, which overproduces the phosphorylated substrate glycans. This modification was induced by MNN4 protein overexpression. Different dilutions of the enzyme preparations will be used to digest these glycans. After incubation we perform a DNA-sequencer aided fluorophore assisted carbohydrate electrophoresis (DSA-FACE). DSA-FACE is a very sensitive high-throughput system to analyse N-glycans. It is based on capillary electrophoresis and detection of fluorophore-labelled N-glycans by laser-induced fluorescence. Molecules are separated based on molecular weight and charge (Callewaert et al., 2001; Callewaert et al., 2002). This gives us a profile of the CcMan5 substrates and products that can be analyzed and compared to the profile obtained with the original CcMan5 and the negative control. Based on this we can decide if the mutation resulted in a faster conversion of the less preferred substrate or not.

First we extracted periplasm of the three mutants but also of native CcMan5 before and after restriction site introduction. Then we added the glycans from the MNN4 strain. The reaction was performed in duplo, for one hour at 37°C or overnight. The reactions were stopped by heatshock of 5 min at 100°C. Then the samples were analysed together with the dextran and RNaseB as calibration and MNN4 sugars as negative control. The dextran calibration shows a peak at the glucose units, while the RNaseB shows $Man_xGlcNAc_2$ where x varies from 5 to 9. In our negative control we observe at 4500 a $GlcNAc_2Man_8$, at 3250 and 2500 respectively a $GlcNAc_2Man_8$ with one and two Man-Pi-Man modifications.

If we compare the profiles of the samples that incubated one hour (Figure 22), we can see at first sight that the profiles of CcMan5.1, CcMan5.2 and the Ala631 mutant are very similar. For the CcMan5.1 and 5.2 this was expected, as the amino acid stayed the same and so the activity. From this point on we could use CcMan5.2 instead of the version with less restriction sites. The profiles of these active proteins show a peak around 2500 and 1500, which represents a $GlcNAc_2Man_8$ respectively with one and two uncapped Pi-6-Man. The Ala351 mutant shows the 2500 peak but the 1500 peak is absent. The profile of the long deletion mutant is similar to the MNN4 negative control, which could be explained by the fact that on western blot no protein was detected in the periplasmatic fraction. The patterns for the overnight-incubated samples (Figure 23) are almost the same, only the relative fluorescence was generally higher. Except for the Ala351 mutant, which was fully converted.

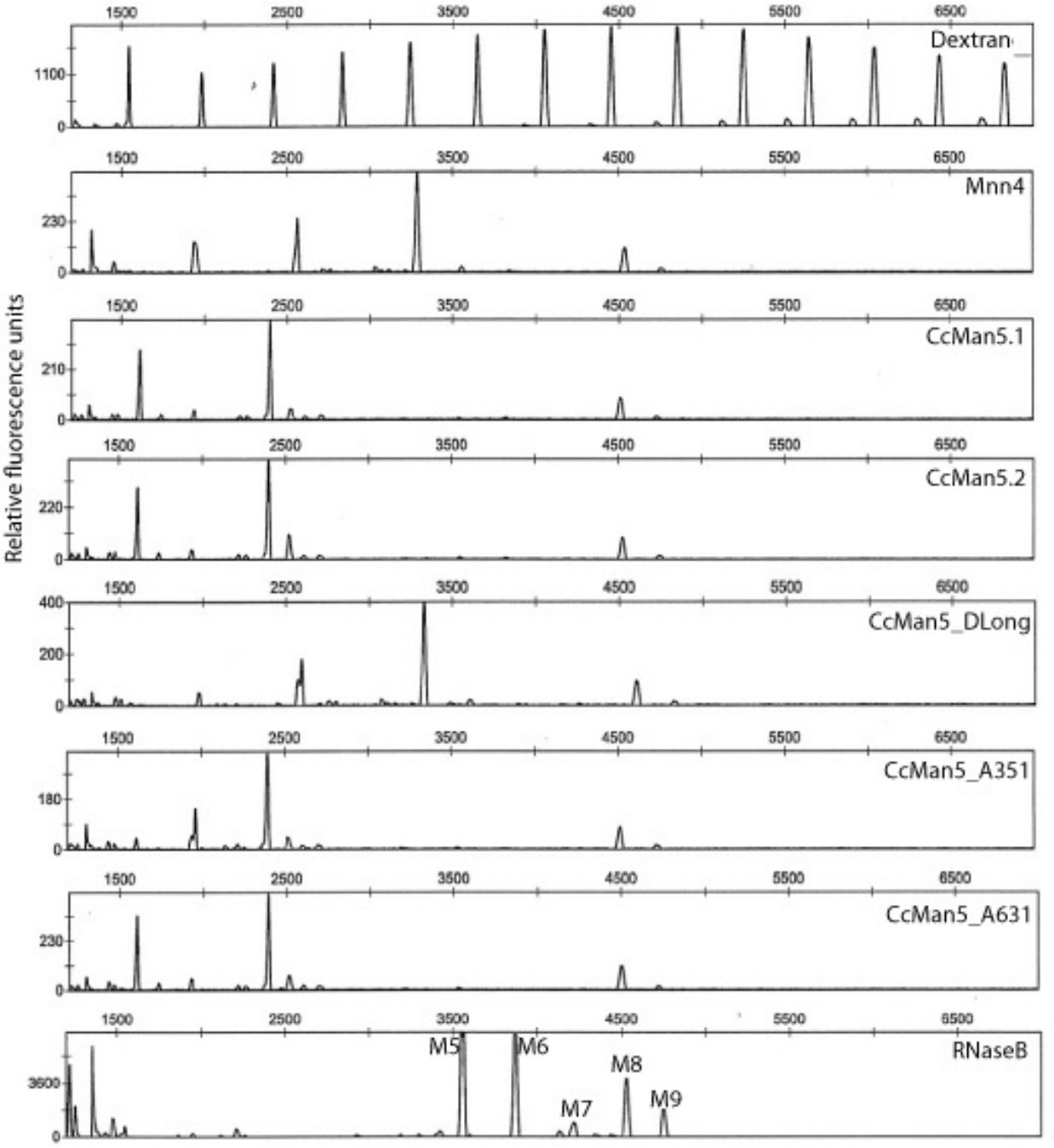


Figure 22 N-glycan profiles of native CcMan5 and mutants of the samoles incubated for 1h. The N-glycan profiles of the CcMan5.1 is compared to the profiles of the mutants and the CcMan5.2 protein after 1h incubation. The dextran calibrations with glucose units is diplayed at the top. At the bottom the results of the referential N-glycans of RNaseB is showed. The second profile represents the untreated N-Glycans from the Mnn4 mutant strain, which is the negative control. CcMan5_DLong: long deletion mutant, CcMan5_A351: Ala351 mutant, CcMan5_A631: Ala631 mutant

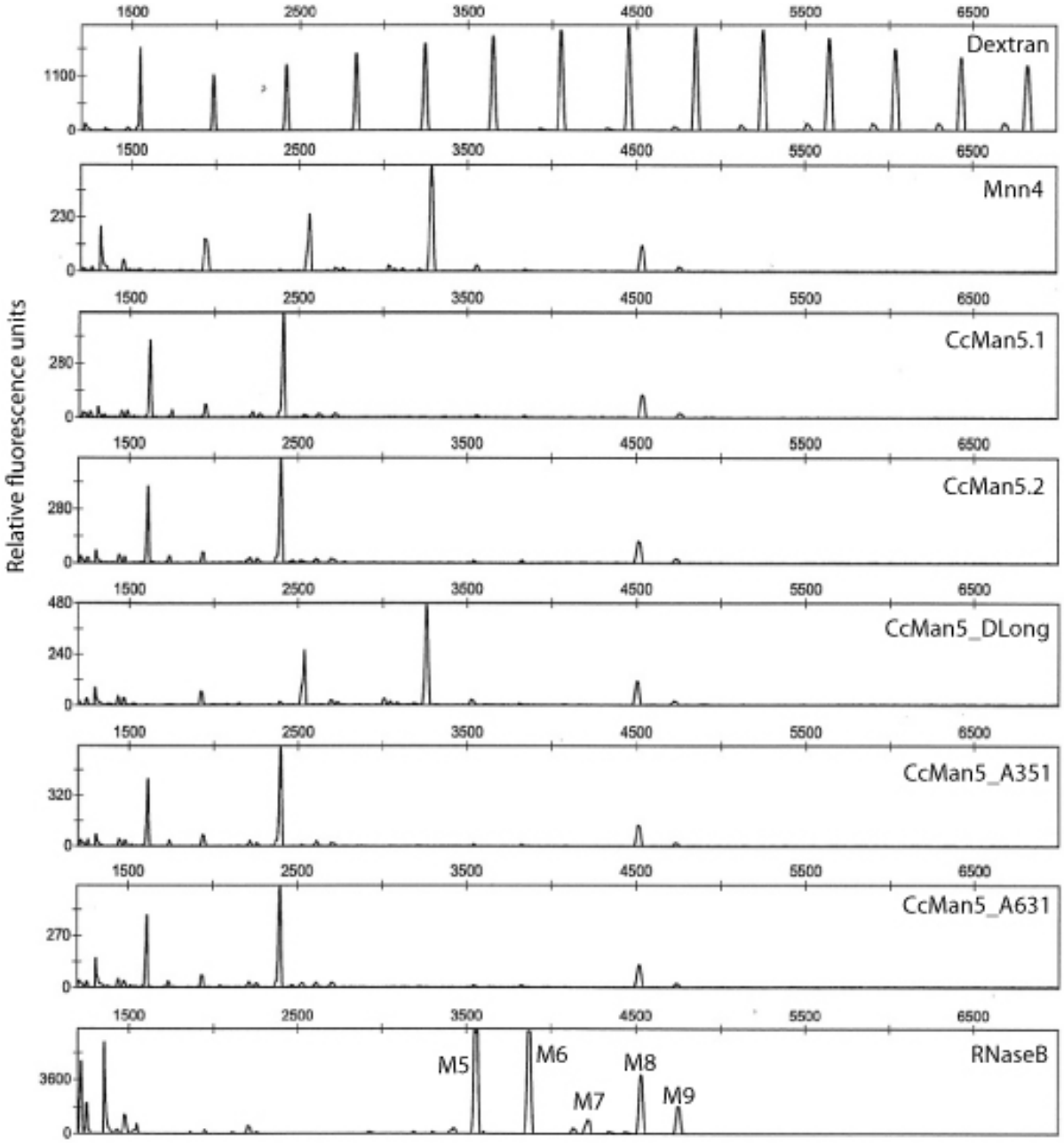


Figure 23 N-glycan profiles of native CcMan5 and mutants of the samples incubated overnight. The N-glycan profiles of the CcMan5.1 is compared to the profiles of the mutants and the CcMan5.2 protein after overnight incubation. The dextran calibrations with glucose units is diplayed at the top. At the bottom the results of the referential N-glycans of RNaseB is showed. The second profile represents the untreated N-Glycans from the Mnn4 mutant strain, which is the negative control. CcMan5_DLong: long deletion mutant, CcMan5_A351: Ala351 mutant, CcMan5_A631: Ala631 mutant

III.1.8 Protein purification

Based on experience with previous CcMan5 mutants, we supposed that the mutated CcMan5 has approximately the same characteristics as the original form. The expression of CcMan5 is most pure in the *E. coli* periplasm so this fraction is used for purification. By using hyperosmotic shock and lysozyme treatment, we can extract the periplasmic fraction. The protein was purified out of the extract by metal ion affinity chromatography (IMAC) based on the N-terminal His-tag. The His-tag binds the Ni^{2+}-Sepharose column. Afterwards the protein was eluted by competition with imidazole. We purified the Ala351, Ala631 and deletion long mutant. First we injected the periplasma using a 50 ml superloop. Then buffer A with 8% buffer B (i.e. 50mM imidazole) was added for 15 column volumes to remove aspecifically bounded proteins. Elution was performed at 200mM imidazole, and for each mutant we saw a small peak immediately after the increase of buffer B. Buffer B was further increased to 100% to ensure complete elution, but it seemed that 200mM of imidazole was enough to obtain the protein. (Graphs in addendum). We tested the different fractions with SDS-PAGE and Coomassie or Western Blot to see respectively, whether the protein was indeed pure and whether it was present in the elution fraction. For the Coomassie (Figure 24) we loaded the wash fraction, purified protein fraction, the pellet and periplasm extract. In Figure 24, you can see that the purification succeeded quite well resulting in a single band on the gel. In the deletion long mutant some breakdown products or aspecfic proteins remain in the purified fraction. Western Blot was performed on different fractions; before purification, the purified protein, the rest fraction after purification, the pellet and the periplasm. But it didn't succeed even after different tries.

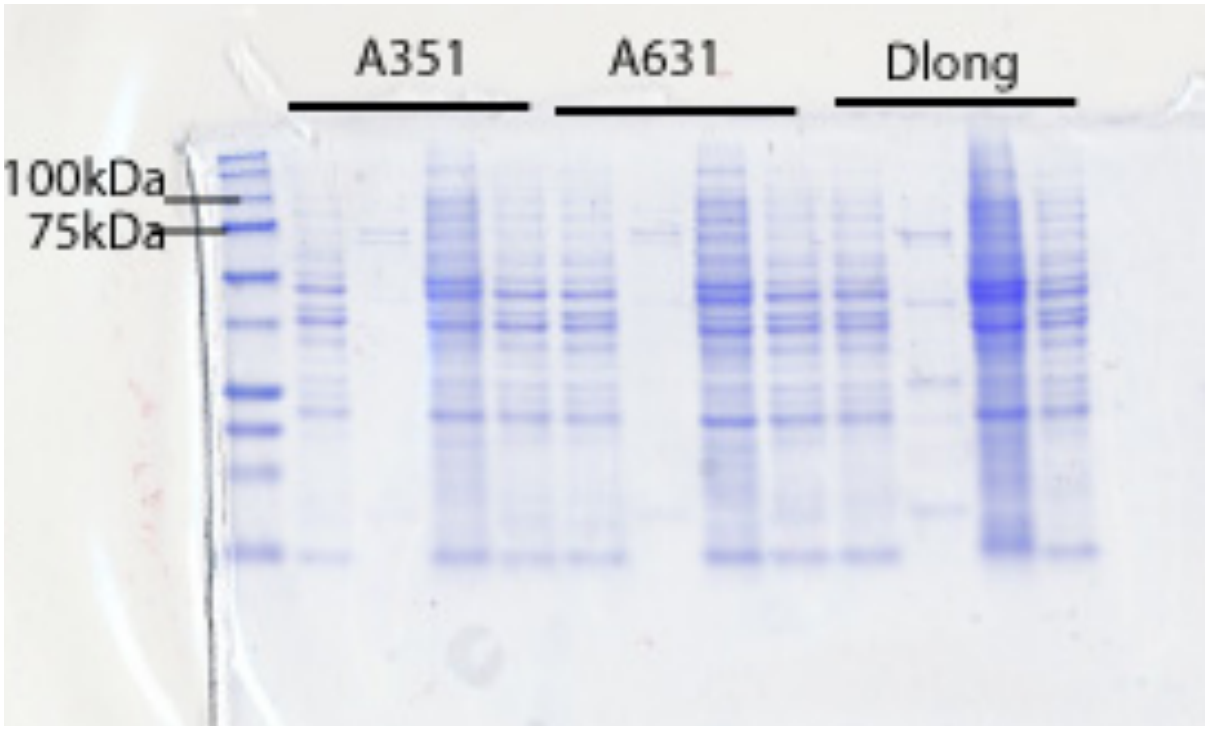


Figure 24 Coomassie to check purification for the Ala 351, Ala 631 and deletion long mutant , where for each respectively the wash fraction, purified protein fraction, pellet and periplasm is loaded.

Enzyme concentration was determined using BCA (Figure 25).

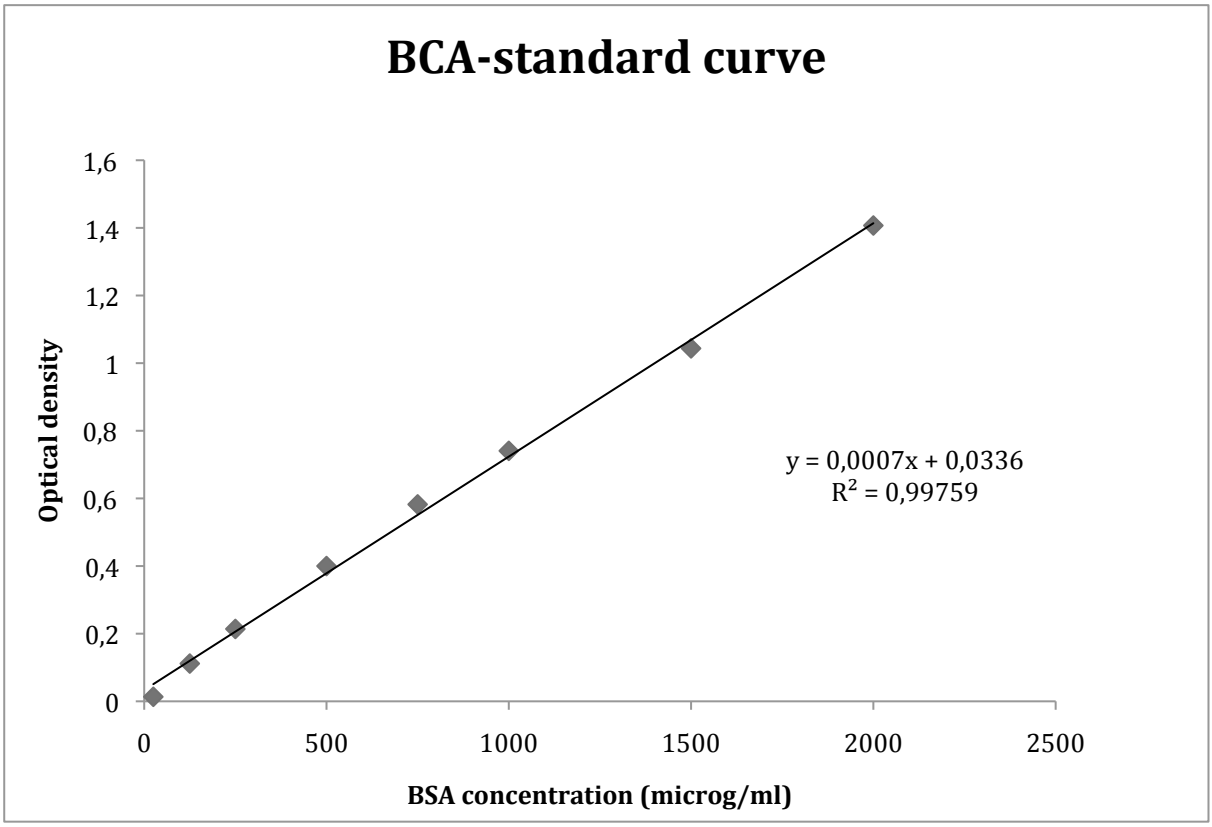


Figure 25 BCA-standard curve BSA concentration (μg/ml) as function of OD

The best fitting linear function had a R-squared value of 0.998 so that we could conclude that this line is a good approximation. Then we calculated the amount of protein out of this function. The OD was calculated for a ½ dilution.

A351: OD_{Sample}-OD_{Blanco}= 0.0128 → beneath lowest standard standard of 25μg/ml

A631: OD_{Sample}-OD_{Blanco}= 0.0133 → beneath lowest standard standard of 25μg/ml

Deletion long: OD_{Sample}-OD_{Blanco}= 0.0828 → 140.67 μg/ml

The concentrations were very low, so upscalling the start quantity is needed. As we saw on previous western blots that the deletion mutation was not expressed in the periplasmatic fraction, it is odd that more protein is in the fraction. There are also contaminating proteins in this fraction. These proteins could explain this higher quantity, as well as the fact that the same column was used different times and the method to prepare the periplasmatic fraction was different.

III.2 Computational part

As described in the aims, we will use molecular modeling to try and identify CcMan5 mutants that optimize the binding of the N-glycan as well as improve the uncapping-reaction of the Man-Pi-6-Man.

III.2.1 Creation of topology file for the native and mutants

To get started we need a PSF file where the topology of our protein is described. This means a description of the used atom types, active bonds, angles, dihedrals and improper torsion angles present in the molecular system. Using the psfpackage in the TK Console of VMD, a visualisation program (Humphrey et al., 1996)**,** we can create the PSF files. Therefore we wrote a short script where we selected the A chain of the homodimer and created the PSF and PDB file for this chain and the calcium ions present. The psfgen package uses a CHARMM topology database file, in our situation we used top_all27_prot_na.rtf. As histidine residues can have different protonation states, we decided to use the HSE state, in which the epsilon nitrogen is always protonated. No disulfide bridges needed to be defined via a patch, since CcMan5 has none. At the end of the script we created also a PSF and PDB file where the protein is solvated. A box of size 10x10x10 angstroms was defined under periodic boundary conditions, containing the protein and sufficient water molecules to warrant a separation between periodic images of at least ten angstroms.

For the deletion we removed amino acids 614 until 642 in the original PDB file, as we did for the short deletion were we removed amino acids 635 until 642. Then we created a PSF and PDB file as described for the native protein. For the point mutation we changed amino acids 631 and 351 from alanine to leucine by adding a mutate patch in the script.

III.2.2 N-glycan description

Producing a PSF and PDB file for the N-glycan asked some more work, as no standard topology was available. CHARMM topologies for carbohydrates, like alpha-D-mannose, are declared in top_all36_carb.rtf. There are also some patches to make the bonds in between but there are non available for the 1,6 alpha bonds and the Man-Pi-6-Man. For these two we created patches by ourselves based on existing one's. Therefore we had to declare enough angles and distances so that the structure could be defined.

1,6-alpha bond between two mannose residues:

```
! axial-axial 1->6 linkage
! GENTBSO1
PRES 16aa          -0.07 ! (i)1->6(i-1) equatorial at C1 and equatorial
at C6
dele atom 1HO6
dele atom 2HO1
dele atom 2O1
ATOM 1C6   CC321    0.00 !
ATOM 1O6   OC301   -0.36 !
ATOM 2C1   CC3162   0.29 !
BOND 1O6 2C1
!    I    J    K    L      R(IK)   T(IKJ)    PHI   T(JKL)   R(KL)
IC   1C5  1C6  1O6  2C1    1.5616  105.37  172.71  113.93   1.3902 !
PSI
IC   1C6  1O6  2C1  2O5    1.4311  113.93   63.49  108.63   1.4110 !
PHI
```

```
IC   2O5  1O6 *2C1  2C2    1.4110  108.63  122.09  110.89   1.5316
IC   2O5  1O6 *2C1  2H1    1.4110  108.63 -121.92  111.32   1.0837
!IC   1O6  2C1  2O5  2C5    1.3902  108.63 -176.53  113.95   1.4709
!IC   1O6  2C1  2C2  2C3    1.3902  110.89 -178.57  113.29   1.4987
```

Man-Pi-6-Man linkage:

```
PRES M6P          -1.00  ! add mannose-phosphate
DELE ATOM 2HO1
DELE ATOM 1HO6
GROUP
ATOM 1C6    CC321     0.03 !
ATOM 1H61   HCA2      0.09 ! axial Hydrogen
ATOM 1H62   HCA2      0.09 ! axial Hydrogen
ATOM 1O6    OC30P    -0.62 !
ATOM P6     PC        1.50 !
ATOM OP62   OC2DP    -0.82 !
ATOM OP63   OC2DP    -0.82 !
ATOM 2O1    OC30P    -0.62 !
ATOM 2C1    CC3162    0.30 !
ATOM 2H1    HCA1      0.09 !
BOND 1O6   P6  OP62 P6  OP63 P6   2O1 P6
IC  1C5   1C6  1O6    P6      1.5323  109.44  -84.54  127.66   1.5968
IC  1C6  1O6   P6   2O1       1.4102  127.66 0.0  102.94   1.5071
!IC  1C6  1O6   P6   2O1       1.4102  127.66 -174.21  102.94   1.5071
IC  1O6  2O1  *P6  OP62       1.5968  102.94  113.03  118.81   1.5471
IC  1O6  2O1  *P6  OP63       1.5968  102.94 -113.60  109.37   1.4811
IC  2C2   2C1  2O1    P6      1.5323  109.44  180.0  127.66   1.5071
IC  2C1   2O1   P6   1O6      1.4102  127.66 -174.21  102.94   1.5968
IC  2O5   2C1  2O1   P6       1.3600  108.92  53.4   107.6     1.5071
```

We also needed a PDB file for the first mannose, as this structure is used as a scaffold to generate the structures of the other carbohydrates in our N-glycan. In the script itself we link step by step each carbohydrate residue, e.g. patch 12aa A:1 A:2 makes an alpha-1,2 bond from residue 2 to residue 1.

III.2.3 Energy minimization

The structure in a PDB file is usually obtained by X-ray crystallography under conditions that hardly resemble the cellular environment, not necessarily yielding the active protein conformation. In addition, no experimentally defined structures were available for the deletion and mutation protein variants. To get a more accurate structure we performed an energy minimization with the NAMD package (Phillips et al., 2005). 5000 steps of steepest descent minimization were performed. The last frame of the resulting trajectory was considered to be the minimized structure, and was later used as a starting point for molecular dynamics and docking simulations. The effect of the minimization was less pronounced for the native protein structure than for the mutated protein variants. In Figure 26 we showed the structure before and after minimization.

The native structure (Figure 26) is almost the same before and after minimization. For the short (Figure 29) and long deletion (Figure 27) the minimization results in a shorter loop. For the small deletion, there are also some local changes, e.g. the β-sheets under the deleted loop. The Ala351 mutant (Figure 30) resulted in a structure were the alanine side chain is turned. The bigger side chain resulted in a structure where a hole near the active site is filled up. On the other hand the Ala631 (Figure 28) mutant resulted in more space, it pushed the loop a bit higher.

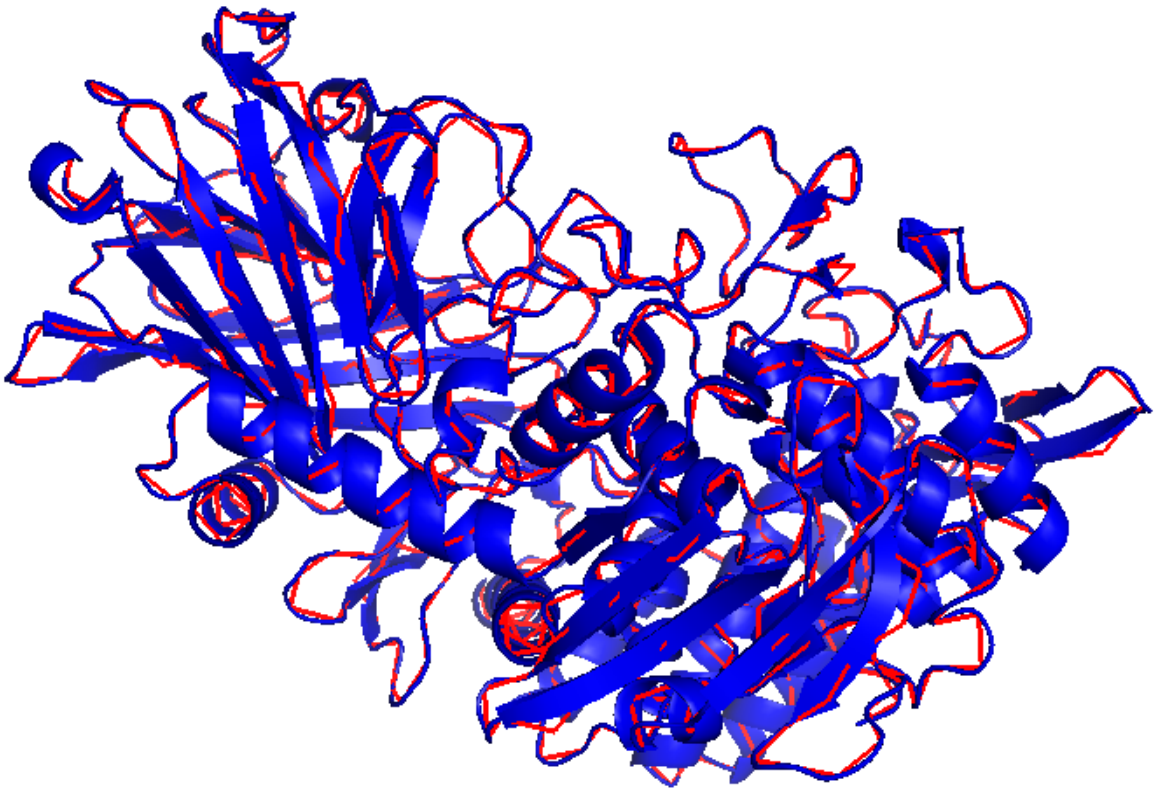


Figure 26 Native CcMan5 X-ray structure before (blue) and after (red) minimization. The minimized structure was obtained after 5000 steps of minimization. The resulting structure is almost the same as X-ray structure.

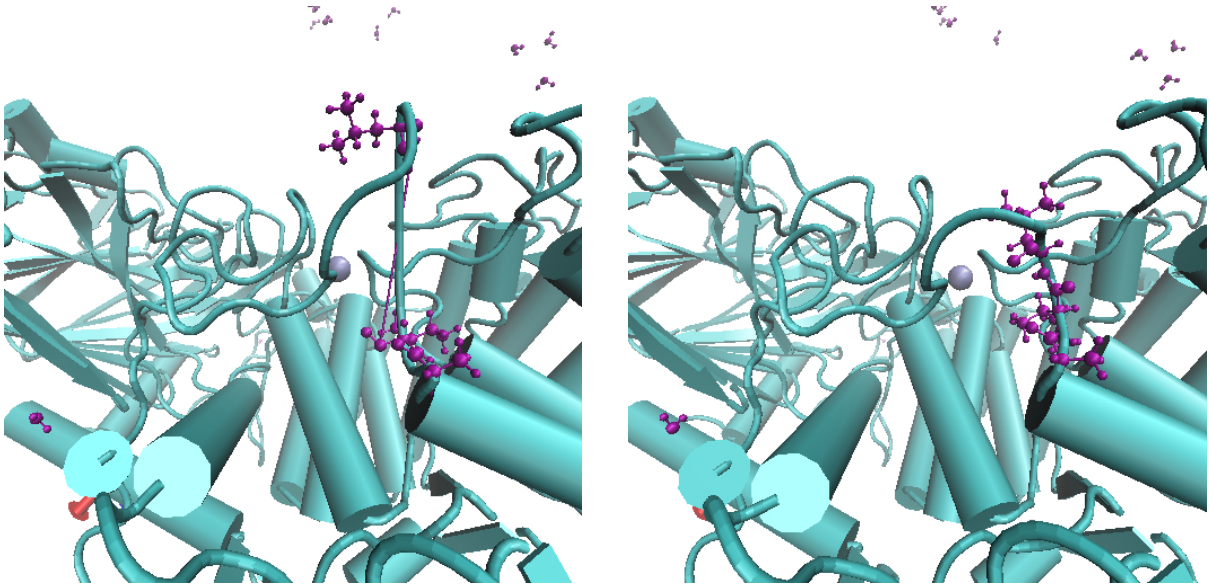


Figure 27 Long deletion mutant (615-642) before (left) and after (right) minimization. This resulted in a shorter loop and almost no effect on the rest of the structure.

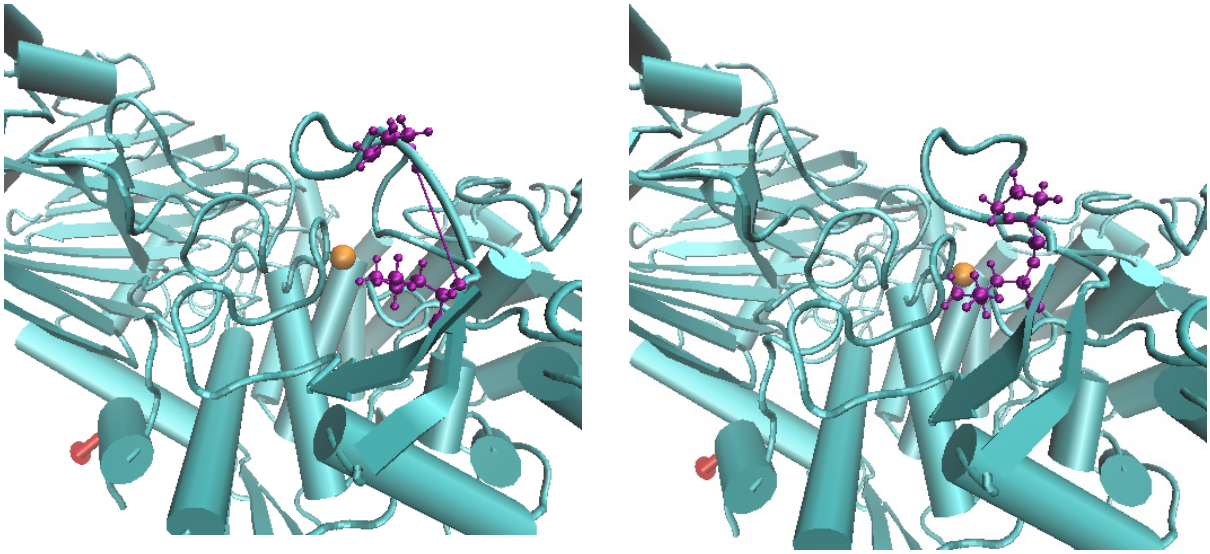


Figure 29 Short deletion mutant (634-642) before (left) and after (right) minimization. This resulted in a shorter loop and the β-sheets beneath are also affected.

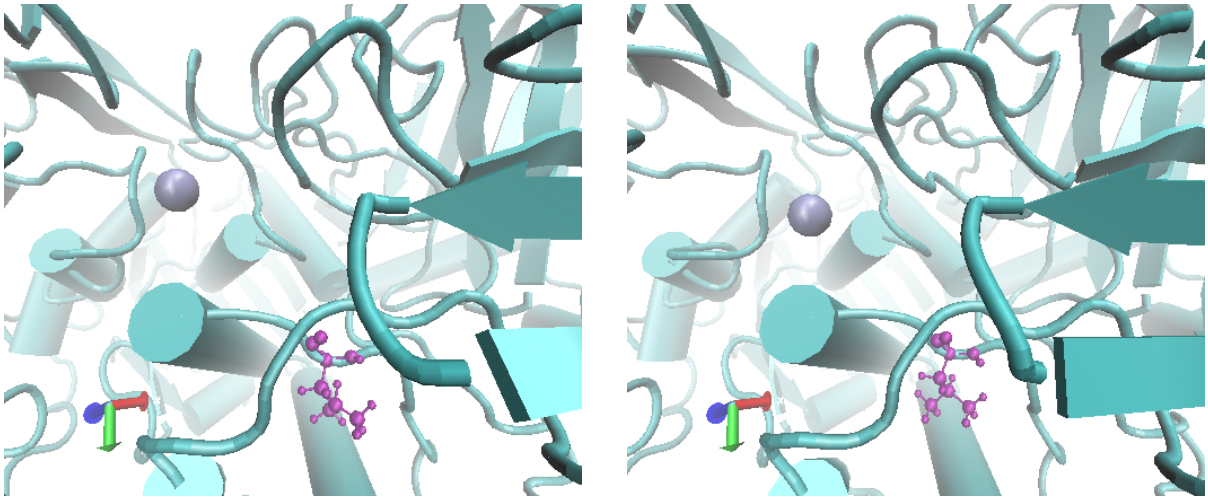


Figure 30 Ala351 mutant before (left) and after (right) minimization. The original alanine was replaced by leucin. The leucine turned around his own axis.

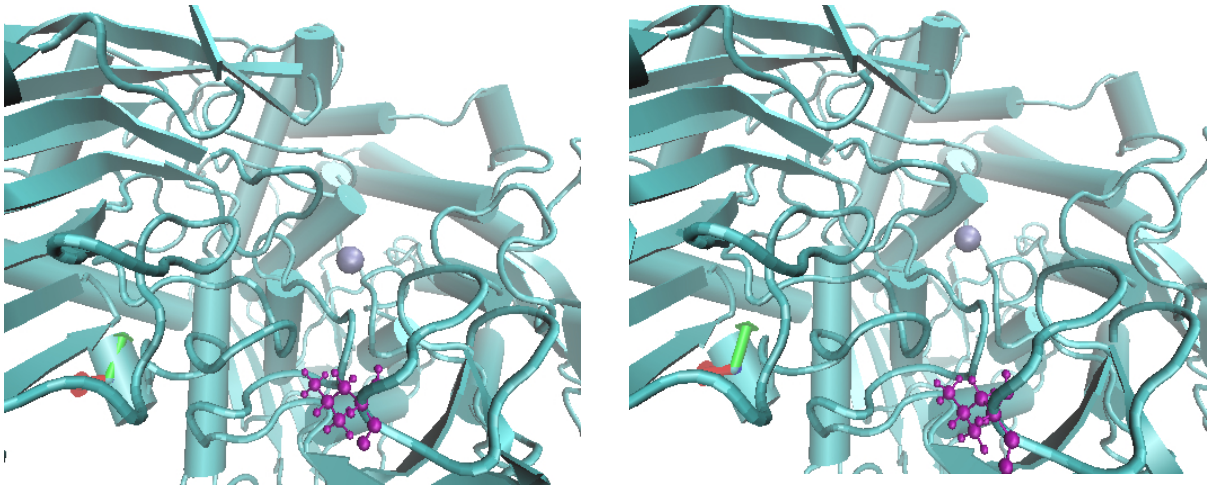


Figure 28 Ala631 mutant before (left) and after (right) minimization. The original alanine is replaced by a leucine. This results in more space, the loop above is pushed a bit upwards.

III.2.4 Molecular dynamics for CcMan5 natives and mutations

The structure in a PDB file is obtained by X-ray crystallography under conditions that hardly resemble the cellular environment. In addition, only static information is obtained. MD simulations allow us to analyze the complete structural and dynamic features of a solvated enzyme (Karplus and Petsko, 1990). MD simulations starting from a mutated CcMan5 form will yield a consensus structure for that mutant, which can then be used in docking simulations. Also, the side-chain flexibility of those groups lining the active site will be examined, allowing us to identify which parts of the protein need to be made flexible in molecular docking simulations.

We started MD for the native CcMan5 protein and four mutants (Ala351, Ala631, deletion long and short) and this during 1ns as this is the period needed to see secondary structure changes. We did this in timesteps of 2fs as we set the H-bonds unchangeable. Simulations were performed in the NVT ensemble, imposing constant volume and temperature. After each MD simulation we calculated with GROMACS the root mean square fluctuation (RMSF) and root mean square displacement (RMSD). The RMSF value gives us an indication about the local flexibility of the chosen atoms. RMSD is the movement of the chosen atoms in function of time. It gives an indication of the general movement in water.

$$RMSD_{S1(t1)-S2(t2)} = \sqrt{\sum_{i=1}^{N}\left(\bar{r}_i^{1}(t_1) - \bar{r}_i^{2}(t_2)\right)^2}$$

$$RMSF = \sqrt{\left(r_n(t)\right)^2 - \left\langle r_n(t)\right\rangle^2}$$

III.2.4.1 Molecular dynamics of native CcMan5

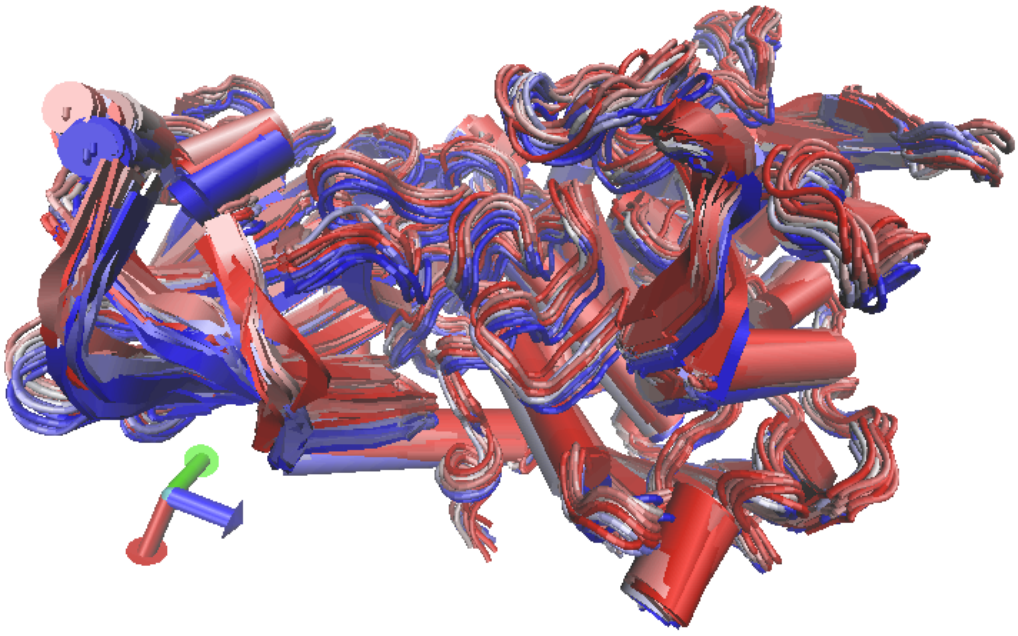


Figure 31 Molecular dynamics side view of native CcMan5. It shows the start position (blue) to the end position (red) in steps of 25 frames. The total simulation time was 1 ns in steps of 2fs with NVT conditions after 5000 steps of minimization.

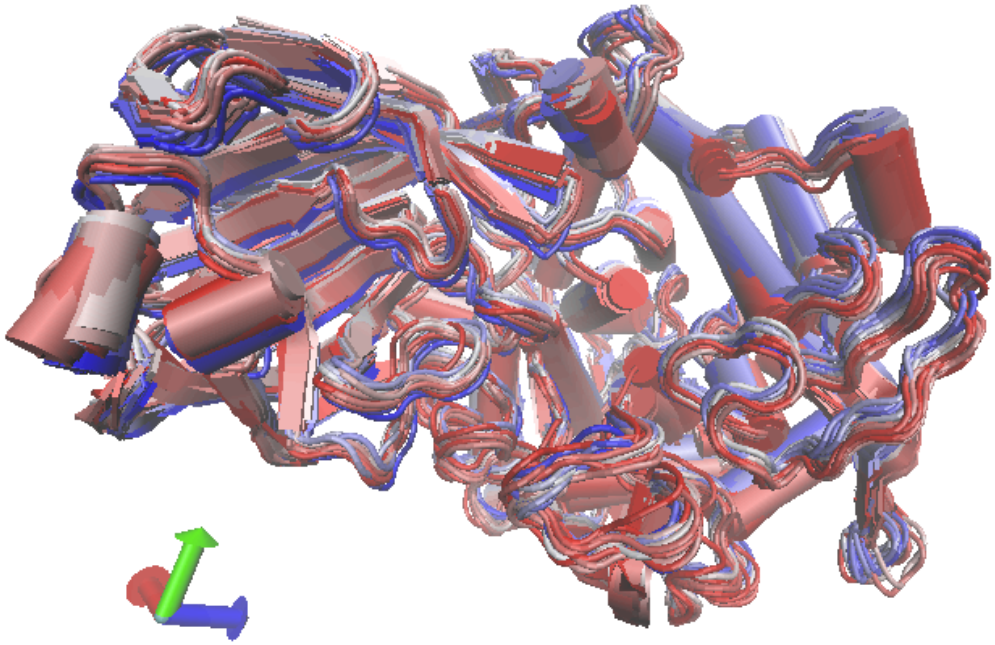


Figure 32 Molecular dynamics top view on active site of native CcMan5. It shows the start position (blue) to the end position (red) in steps of 25 frames. The total simulation time was 1 ns in steps of 2fs with NVT conditions.

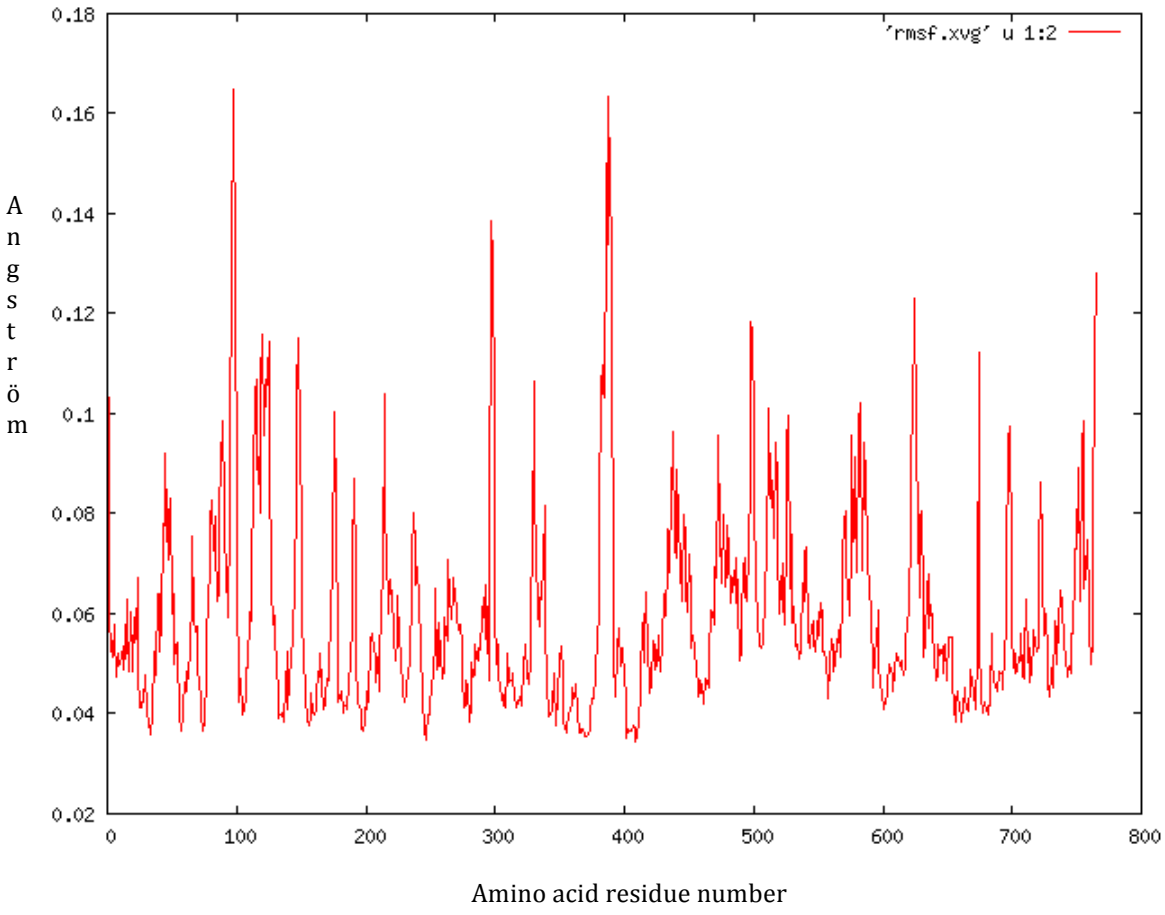


Figure 33 RMSF calculation after molecular dynamics on native CcMan5. The x-axis shows the amino acid residue number, while the y-axis shows the variation in angström during the simulation.

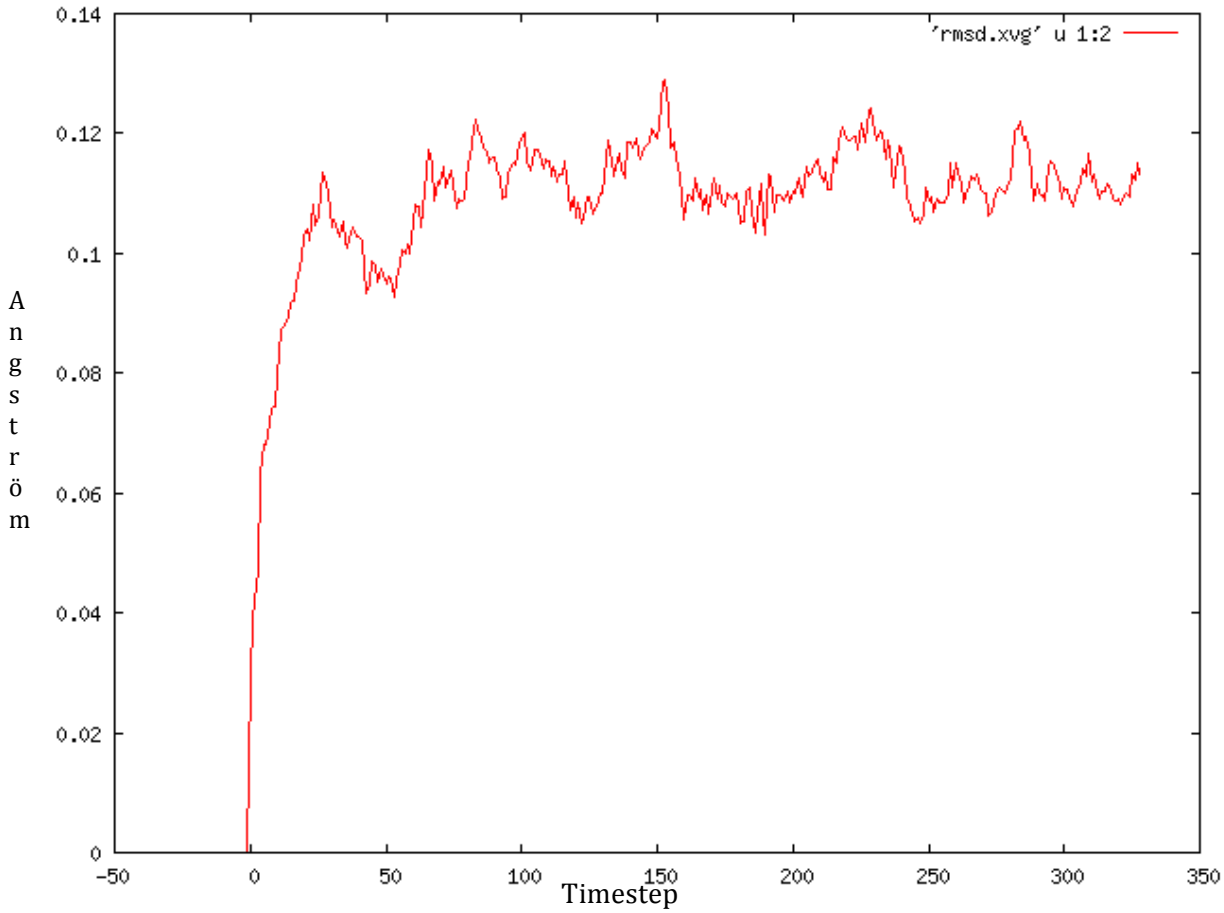


Figure 34 RMSD calculation after molecular dynamics on native CcMan5. The x-axis shows the step, while the y-axis shows the variation in angström during the simulation.

Based on the molecular dynamics movie, the RMSD (Figure 34) and RMSF (Figure 33) we can say that the native CcMan5 structure is very rigid. The RMSF show a variation of 0,04A to 0,16A for certain residues, these residues are not located near the active site. The RMSD shows first an increase but this is normal as the start value is zero. Our RMSD values are calculated relative to the structure presented in the minimized system. The first structure is the minimized one so that the difference is zero.

III.2.4.2 Molecular dynamics of the long deletion mutant

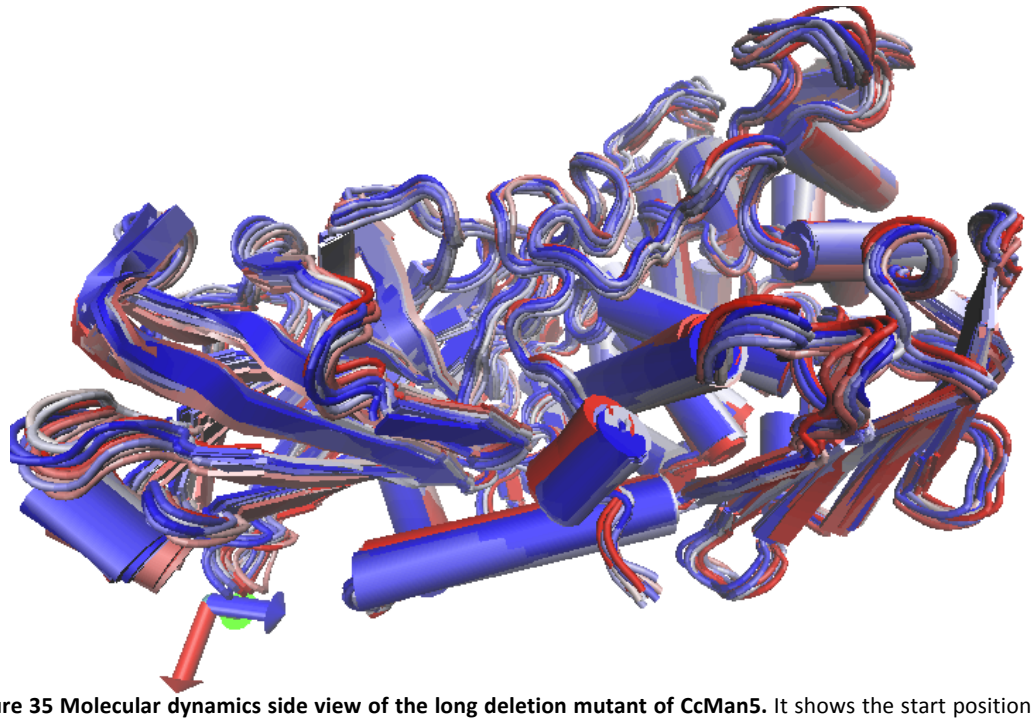


Figure 35 Molecular dynamics side view of the long deletion mutant of CcMan5. It shows the start position (blue) to the end position (red) in steps of 25 frames. The total simulation time was 1 ns in steps of 2fs with NVT conditions after 5000 steps of minimization.

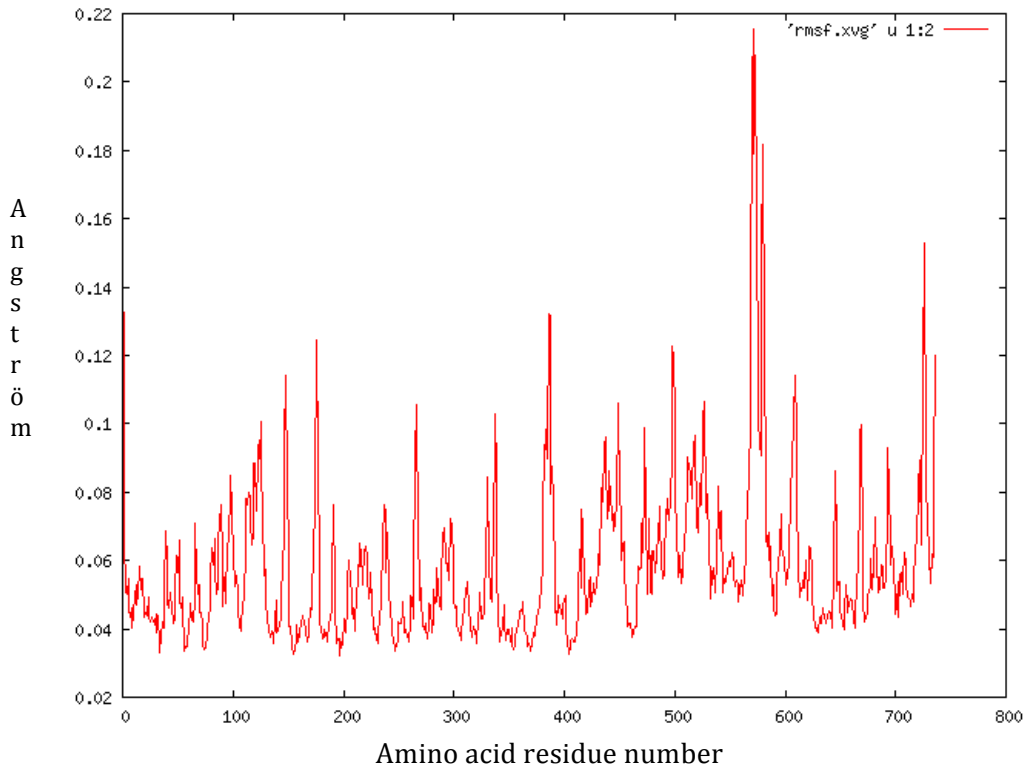


Figure 36 RMSF calculation after molecular dynamics on long deletion mutant of CcMan5. The x-axis shows the amino acid residue number, while the y-axis shows the variation in angström during the simulation.

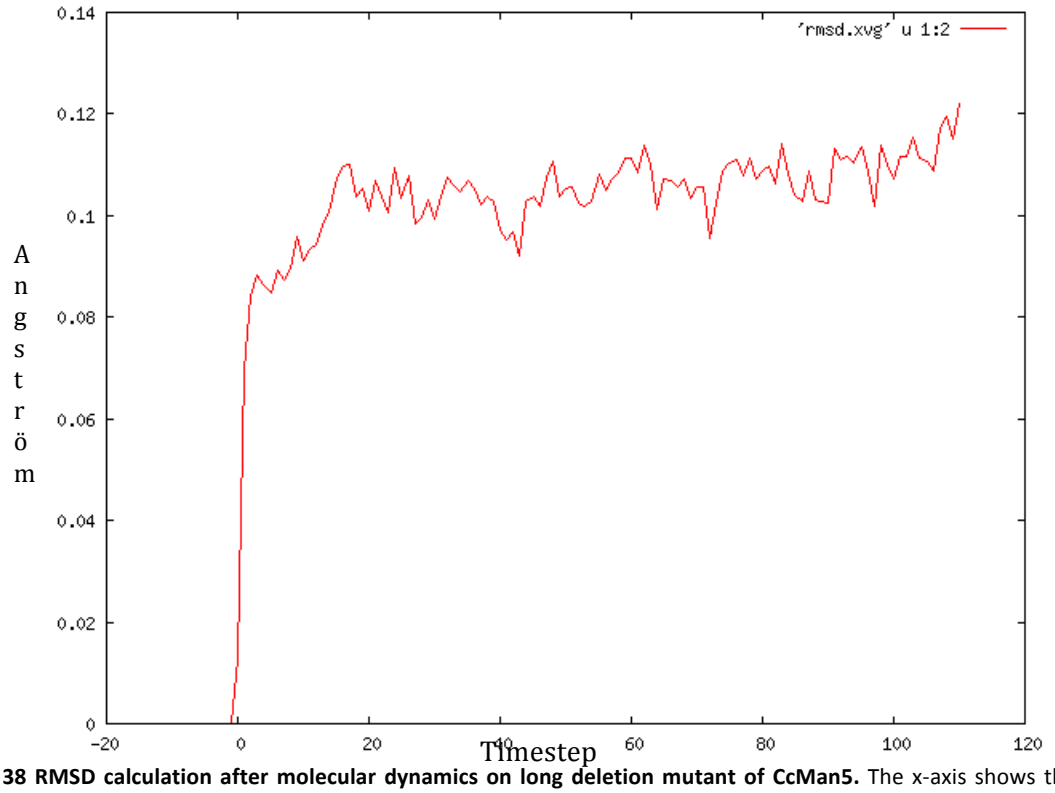


Figure 38 RMSD calculation after molecular dynamics on long deletion mutant of CcMan5. The x-axis shows the step, while the y-axis shows the variation in angström during the simulation.

The deletion mutant is as the native CcMan5 protein quite stable. The biggest difference is the peak around residues 580-590 in the RMSF curve (Figure 36 and Figure 37). This flexibility could be caused by the space that is made due to the deletion as these residues are nearby.

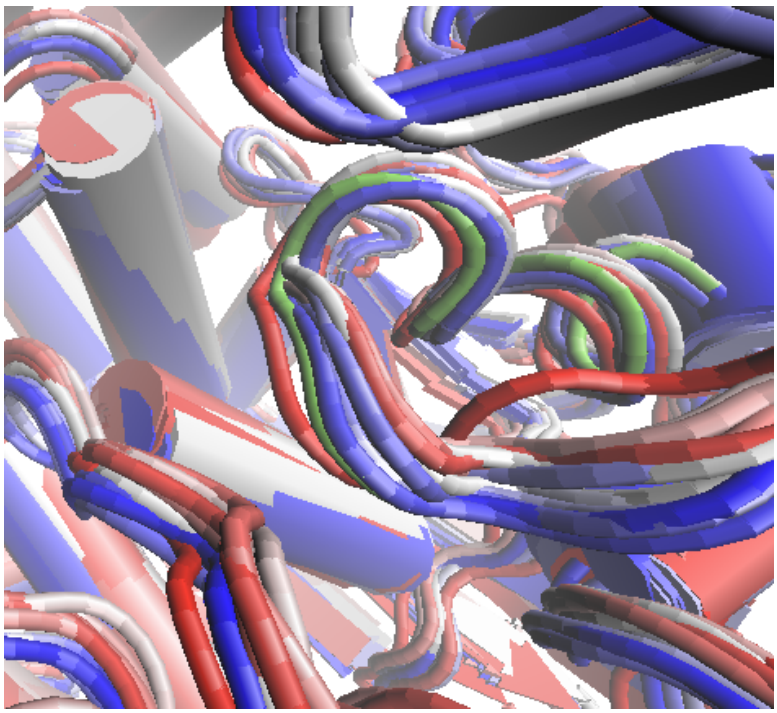


Figure 37 Residue 580-590 of the deletion mutant which have a RMSF value that differs more than the RMSF values for the native CcMan5.

III.2.4.3 Molecular dynamics of the Ala351 mutant

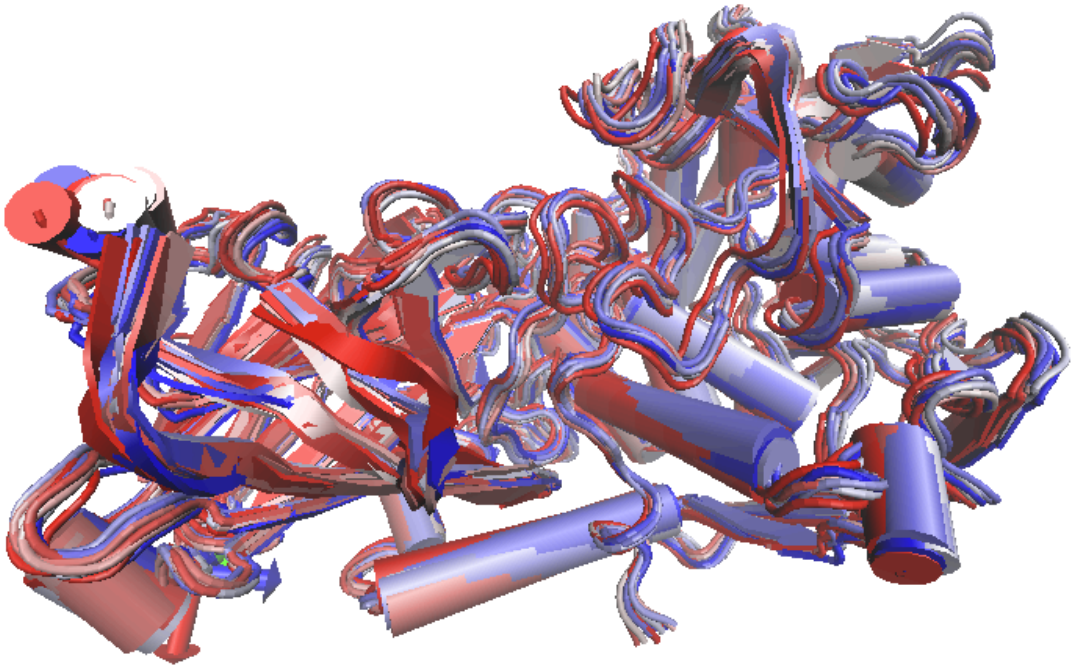


Figure 40 Molecular dynamics side view of the Ala351 mutant of CcMan5. It show the start position (blue) to the end position (red) in steps of 25 frames. The total simulation time was 1 ns in steps of 2fs with NVT conditions after 5000 steps of minimization.

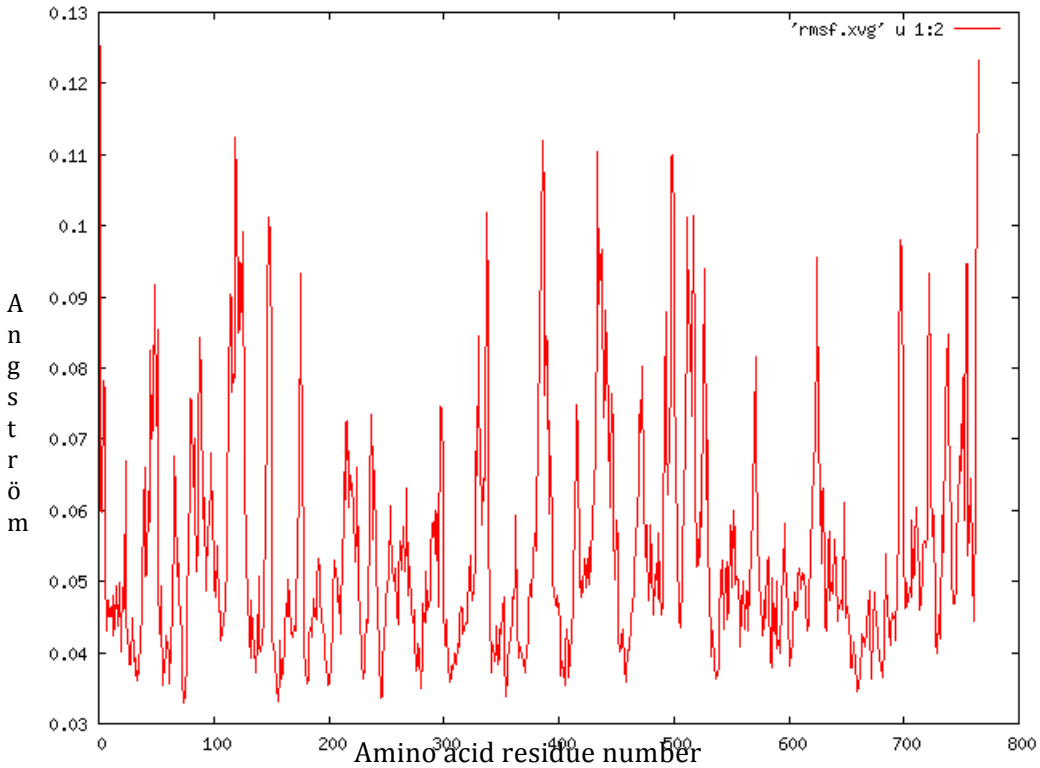


Figure 39 RMSF calculation after molecular dynamics on Ala351 mutant of CcMan5. The x-axis shows the amino acid residue number, while the y-axis shows the variation in angström during the simulation.

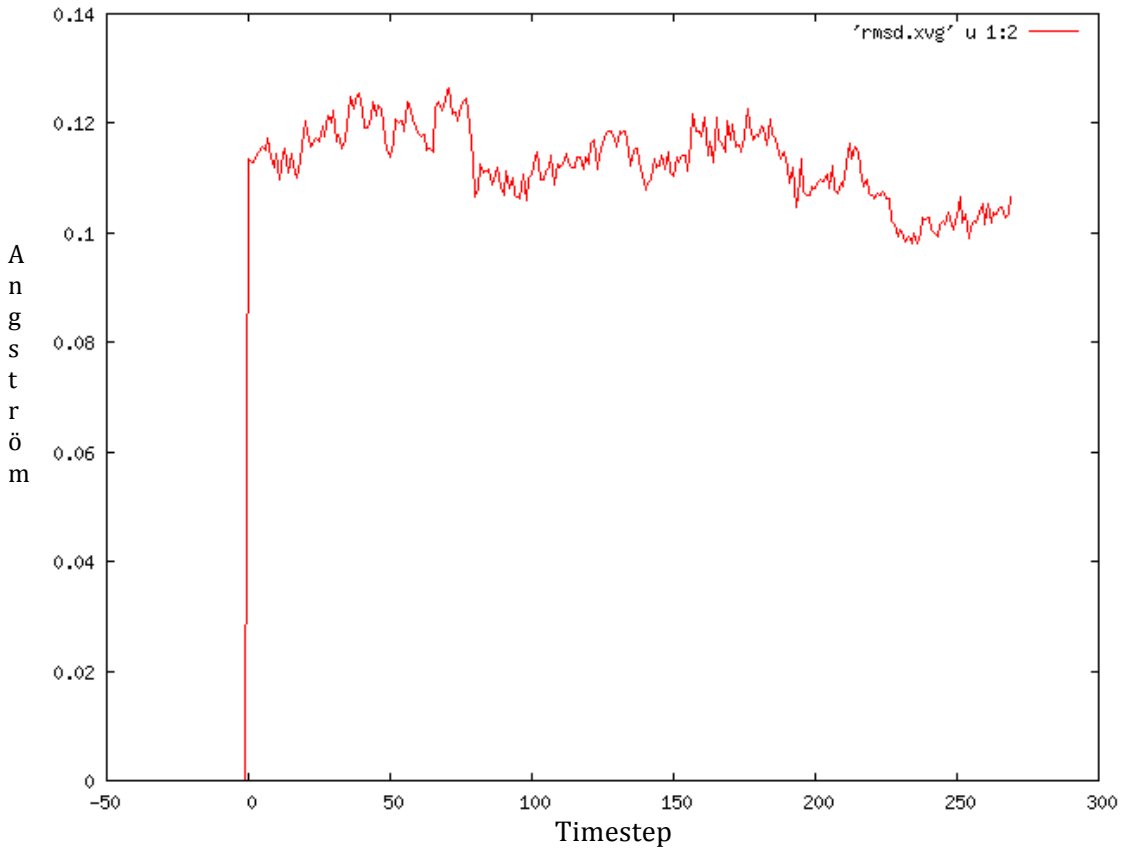


Figure 41 RMSD calculation after molecular dynamics on Ala351 mutant of CcMan5. The x-axis shows the step, while the y-axis shows the variation in angström during the simulation.

The Ala351 mutant RMSF (Figure 39) and RMSD (Figure 41) profiles are more similar to the native CcMan5 than the long deletion mutant profiles. This is straightforward as a point mutation is less invasive than a deletion of several amino acids. Also this mutant is quite stable.

III.2.4.4 Molecular dynamics of the Ala631 mutant

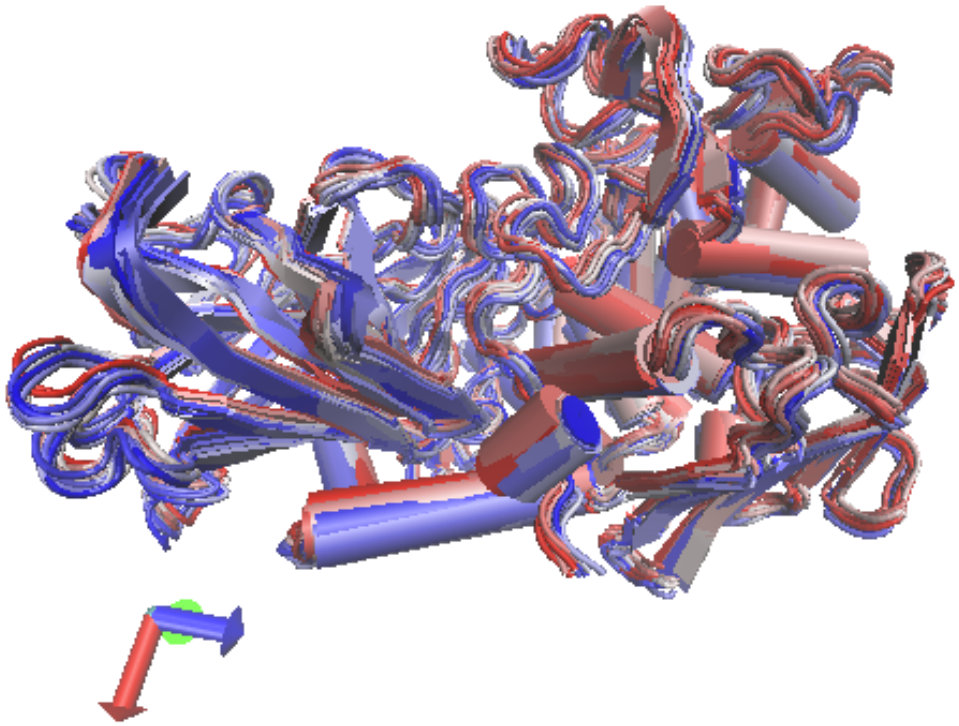


Figure 42 Molecular dynamics side view of the Ala631 mutant of CcMan5. It show the start position (blue) to the end position (red) in steps of 25 frames. The total simulation time was 1 ns in steps of 2fs with NVT conditions after 5000 steps of minimization.

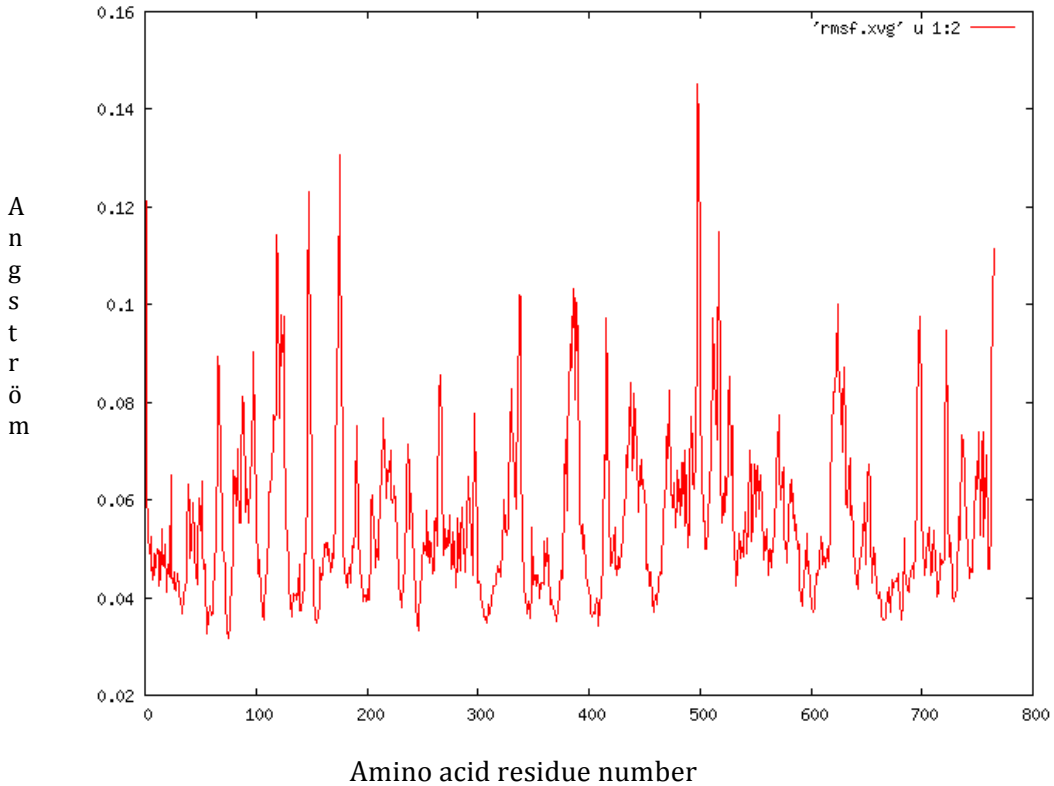


Figure 43 RMSF calculation after molecular dynamics on Ala631 mutant of CcMan5. The x-axis shows the amino acid residue number, while the y-axis shows the variation in angström during the simulation.

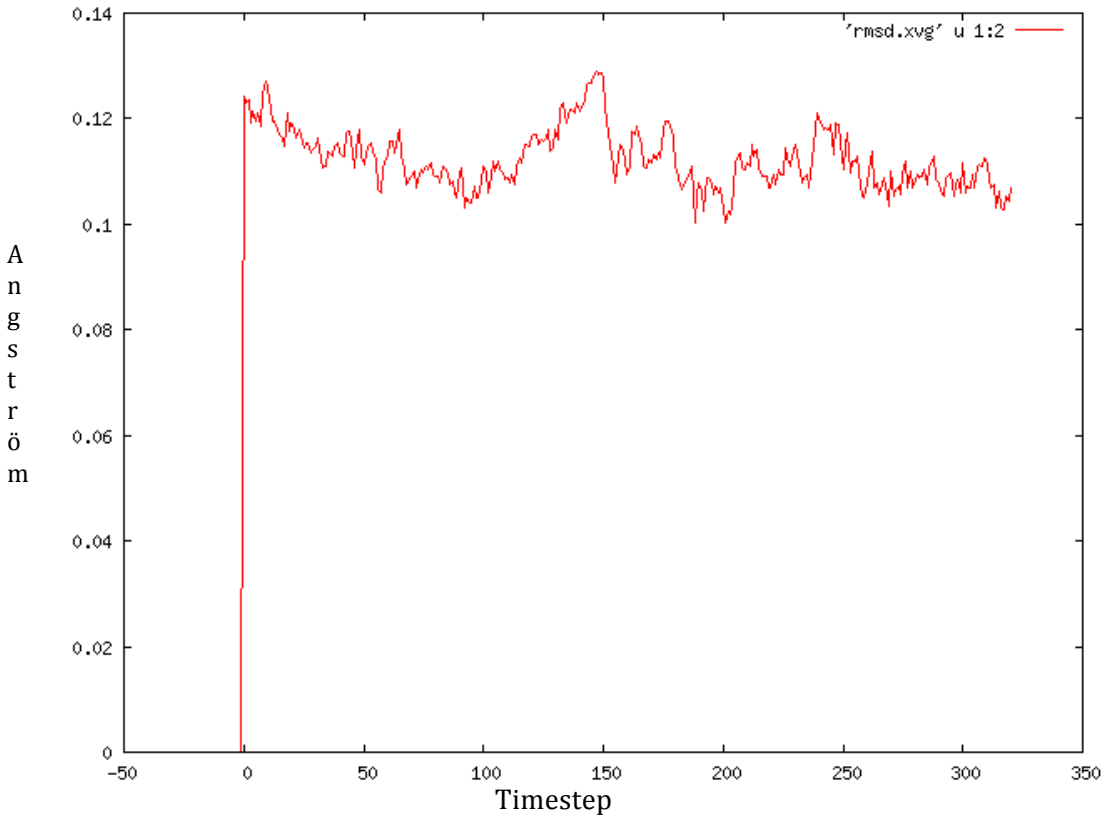


Figure 44 RMSD calculation after molecular dynamics on Ala631 mutant of CcMan5. The x-axis shows the step, while the y-axis shows the variation in angström during the simulation.

These curves (Figure 43 and Figure 44) tell us the same story as for the native CcMan5 and the other point mutant. They show also the stability of the mutated protein.

III.2.4.5 Molecular dynamics of the N-glycan

To check the possible movements of the sugar bonds we performed MD simulations on the two regio-isomers (Figure 45). The MD shows that the phosphates collapse into a kind of umbrella structure. In this situation the oxides are laying near each other with an angle of less than 10° and distance of less then 0.25A (Figure 46), which is a non-realistic distance. This is due to a defective description of the force fields of the phospates. The mannoses turn a lot around the intermediate bounds.

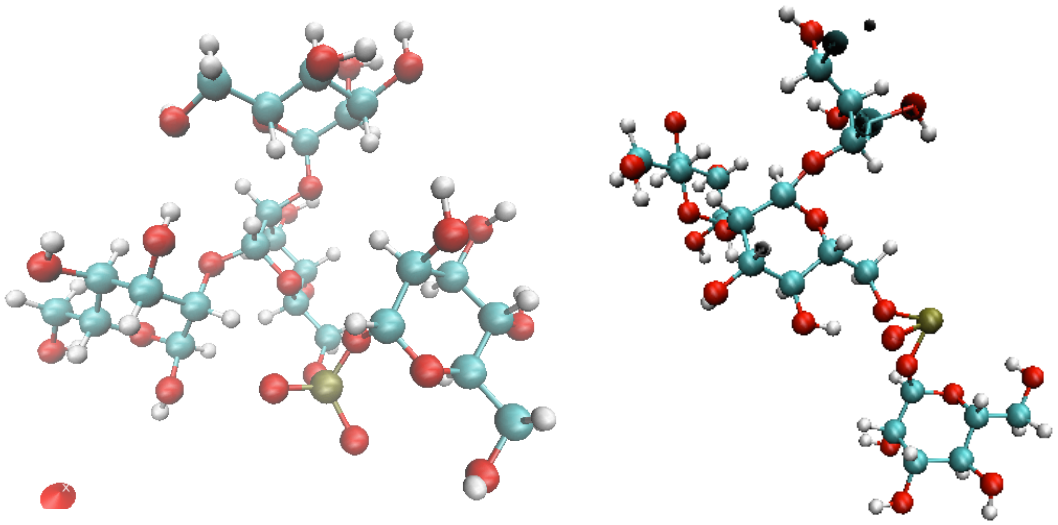


Figure 45 Molecular dynamics of regio-isomere (A) before (left) and after (right) simulation.

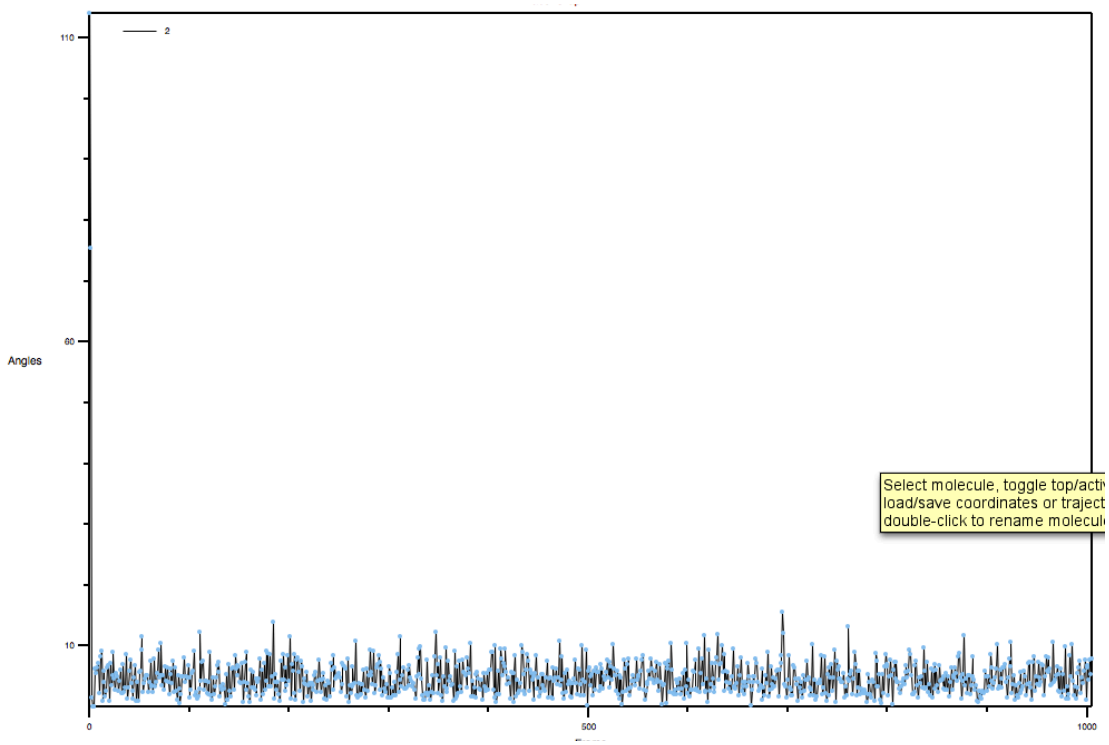


Figure 46 Angle between the two free oxides from the phosphate group of regio-isomere A. The first, original state had a value of 110° but from the first saved MD step on, the angle dropped to less then 10°.This included also a change from 2,5A to less then 0.25A.

III.2.5 Docking of the N-glycan

The next step in the process was combining the sugar tree with the protein. This was done by AutoDock Vina (Trott and Olson, 2010). AutoDock Vina is a docking program where the input files are pdbqt format. Pdbqt files store the atomic coordinates, partial charges and AutoDock atom types, for both the receptor and the ligand. We decided to use the command line scripts and not the graphic interface, as this was easier to handle and the results were the same. For the receptor we could choose between rigid or flexible. Both situations were performed. For the flexible receptor, we set Tyr535 and Gln536 flexible as these residues are lying near the active site. Then we needed also a configuration file where the receptor and ligand pdbqt files are declared. In this file we need to define the centre of the docking and the size of the box, where the ligand should fit in. The centre was defined as the (35.119, 21.558, 75,067) coordinate. The size of the box was 15x15x15 angstrom. As last we needed to define the number of results that must be shown and an exhaustiveness score. This exhaustiveness score is the number of trials that are given to find the optimal solution. The output file is also a pdbqt file with different best docking results, where the first is the best scoring conformation.

First we performed dockings for the simple mannose-6-phosphate-mannose structure. To know what the influence of the exhaustiveness score was, we started with score 8 and upgraded it stepwise to 300. This increased the calculation time from 1 min to 1 h but the best fitting score's stayed the same. We repeated the dockings several times. When we clustered the outcomes based on conformation, the best fitting structures clustered together and so on. This shows the reproducibility of AutoDock Vina. The simple structure is fitting in different conformations to the active site. The best two results, based on docking score, are visualised in Figure 47. The results showed no overall possible direction of where the N-glycan could go to. The outcomes must be visually evaluated based on what is structurally possible, based on the mannose units that are already in there but also on what has to come. Is there enough space left for the rest of the N-glycan? Are they close to Tyr620 and Gly72, the amino acids that made H-bonds with the N-glycan? As there were too many possibilities left, we decided to perform docking of both regio-isomers.

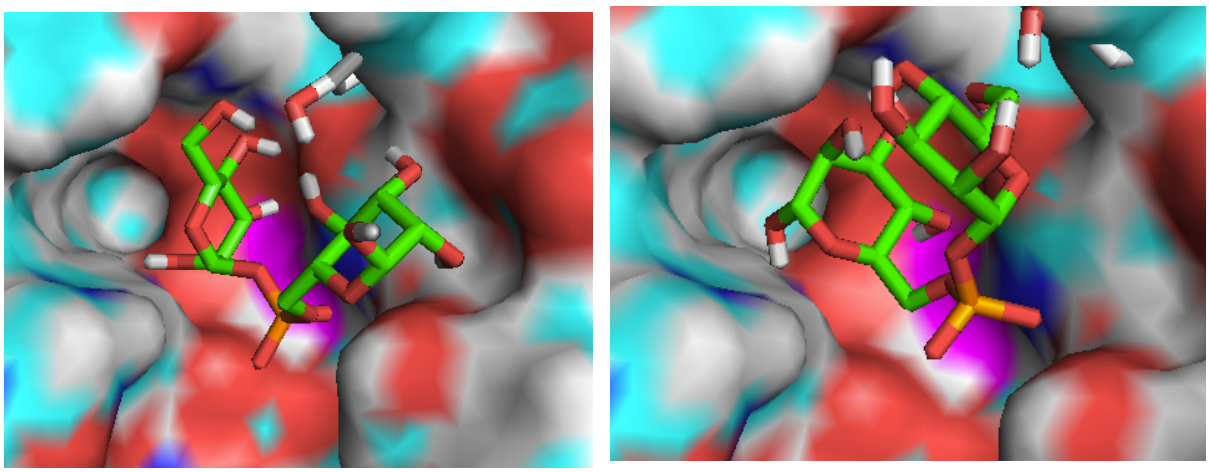


Figure 47 Docking of mannose-P-mannose on native CcMan5. These dockings were performed with AutoDock Vina with an exhaustiveness of 30. The receptor was set flexible for Tyr535 and Gln536. The centre was set as (35.119, 21.558, 75,067) and the box was 15x15x15 angstroms. The surface of the protein is visualised and the calcium is coloured magenta. The best result (left) had a score of -5.3 kJ/mol, while the second result (right) had a score of -5.2 kJ/mol.

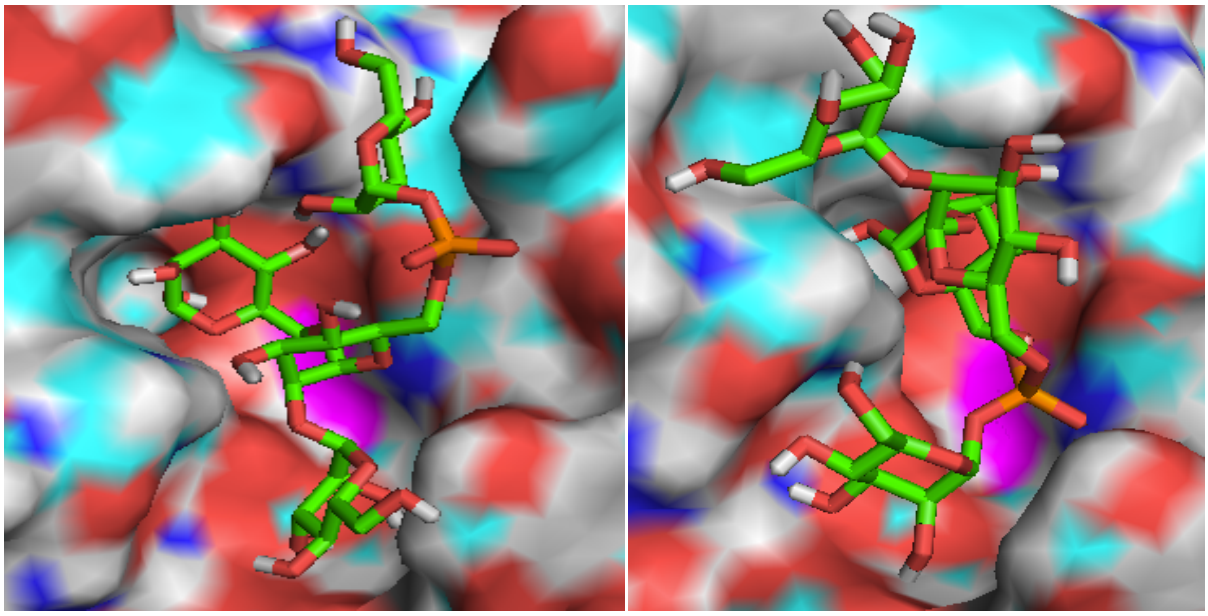


Figure 49 Docking of regio-isomere A of CcMan5. These dockings were performed with AutoDock Vina with an exhaustiveness of 200. The centre was set as (35.119, 21.558, 75,067) and the box was 20x20x20 angstroms. The surface of the protein is visualised and the calcium is coloured magenta. The best result (left) had a score of -7.3 kJ/mol, while the second result (right) had a score of -7.2 kJ/mol.

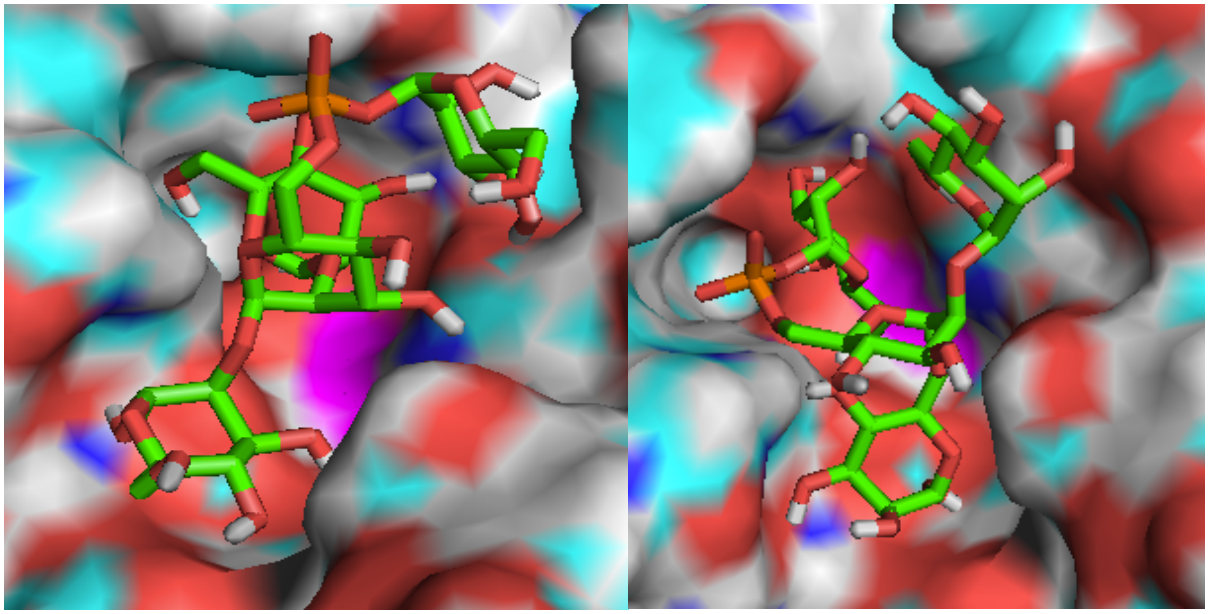


Figure 48 Docking of regio-isomere B of CcMan5. These dockings were performed with AutoDock Vina with an exhaustiveness of 200. The centre was set as (35.119, 21.558, 75,067) and the box was 20x20x20 angstroms. The surface of the protein is visualised and the calcium is coloured magenta. The best result (left) had a score of -8.6 kJ/mol, while the fifth result (right) had a score of -7.9kJ/mol.

As these regio-isomers were bigger we set the box larger (20x20x20 Angstrom). For both isomers the exhaustiveness was similar. In Figure 49 and Figure 48 some results of these dockings were visualised. As one could see, the left conformation is not possible as the phosphates are too far from the active centre. The right one's are more realistic as the mannose that needs to be uncapped is laying near the surface. But the structures give us no unambiguous answer of how the N-glycan binds. It seems that the regio-isomere B has overall better scores and that both regio-isomeres have preference for the opposite sites. As last, we could dock the whole N-glycan, we did this also. But the problem was that in AutoDock you couldn't choose that the mannose-6-phosphate-mannose have to be in a specific position near the active site. This resulted in a lot of poses in which enzymatic reaction on the N-glycan couldn't proceed or the (not-modelled) N-glycan tail would clash with the enzyme.

IV Discussion

LSDs are a group of more than 50 inherited monogenic disorders (Schultz et al., 2011). LSDs are mostly due to loss-of-function mutations in genes encoding lysosomal hydrolases resulting in impaired substrate degradation. A major lysosomal targeting pathway is based on modification of enzymes with Pi-6-Man moieties. The critical factor in therapeutic success of ERT is that the enzymes need to be modified with a high density of Pi-6-Man modifications on their N-glycans (Wraith, 2006; Zhu et al., 2004). This glycotope is recognized by the M6PR at the surface of human cells, leading to endocytosis of the receptor:enzyme complex. The enzyme is then delivered to the lysosomes, complementing the enzymatic defect. There is an ERT on the market to treat LSDs like Gaucher (Beutler et al., 1991), Fabry (Beck, 2002) and Pompe's disease (VanHove et al., 1996). Nowadays the treatment of LSDs is very expensive and not efficacious in many cases. This is due to the complex production systems in mammalian cells and low levels of Pi-6-Man modifications. At the moment Genzyme and Shire has a monopoly on the most enzymes used for ERT. Luckily the patents on ERT for Gaucher, Pompe and Fabry's disease have almost lapsed so that there is now room for new technologies.

CcMan5 is an important factor in the development of the enzymes for ERT in yeast. As it has an important role in the conversion of the yeast produced N-glycans to mammalian like. These N-glycans are needed for the uptake of the ERT by the M6PR towards the lysosomes. A higher conversion rate of these yeast N-glycans results in shorter incubation of the yeast supernatant containing the therapeutic enzyme, resulting in less dephosphorylation by the endogenous phosphatases. Enzymes with more Pi-6-Man will bind the M6PR better, which stimulates the uptake by lysosomes in the cells of the patients. More conversion results in a better uptake by the lysosomes of the patients. This reduces the amount of enzyme needed for the therapy. It results also in a higher efficiency of the enzyme, which is important for the life quality of the patient.

To improve the efficiency of the CcMan5 enzyme and to know more about his working mechanism, we produced mutations of the native enzyme. This was combined with computational analysis of the stability of these enzymes and docking experiments. The mutants that we analyzed were defined based on previous mutations, alignments en structural information. These included the Ala351Leu and Ala631Leu point mutants and a long (Y615-P642) and a short (S634-P642) deletion mutant.

IV.1 Discussion of experimental and computational results

IV.1.1 Production of the mutants

We started with classical site directed mutagenesis as described in Figure 16. This wouldn't succeed, except for the Ala631 mutant. The reason of this failure remained unclear as we changed different parameters and tried different times by multiple people. Before, they struggled also a lot with previous attempts for other mutations on the same protein. So we decided to use upcoming synthetic DNA synthesis. Nowadays synthetic DNA synthesis drops sharply in price and delivery time what makes it worthwhile to use it more and more. For this project we ordered a synthetic DNA fragment (1785bp) for the CcMan5 gene, where we

included more restriction sites. All the included sites are know as well cutting and are often sticky ends. The new restriction sites did not induces amino acid changes. This is very important as the amino acid sequence needed to stay the same to keep the activity. All the new codons were checked on appearance in *E. coli* as the translation efficiency needs to stay high. The detection of possible nucleotide changes, resulting in a restriction site, was manually performed. In combination with small fragments (300bp), where the mutation is already induced, it is possible to make mutated strains in one week. It is not recommended to use two blunt end cutters, as the fragments could be inserted in different ways and selection afterwards is needed. So with this system we succeeded to make our mutants in two weeks, while we tried a couple of months on the classical cloning system. The short deletion mutant wasn't made, as the long mutation would have more impact, and we first wanted to try if the synthetic system worked.

IV.1.2 Expression of the mutants compared with molecular dynamics results

As the mutations were made, expression was induced by IPTG induction in *E. coli* BL21. First we tested with a rough cell extract if there was some of our protein (Figure 19). We did this also for the CcMan5.2 as codon usage could influence the expression. CcMan5.1 expressed as well as CcMan5.2 so that we could use CcMan5.2 from this point on. Then different cell fractions were extracted to visualise where the CcMan5 protein was located (Figure 20 and Figure 21). From these figures we can conclude that the majority of the protein stays in the soluble and insoluble fraction, only a small part is excluded to the periplasm. This periplasmatic fraction is used for the activity test, as the quantity of other proteins is low.

For the long deletion mutant we hardly see protein in the periplasm and the medium fraction. This could be due to incorrect folding. As the MD results show, the long deletion mutant is less stable than the native CcMan5 (Figure 36 and Figure 38). But this says only something about an already folded structure. To study the *ab inito* folding, the computational time would be a limiting factor as the protein is quite large. But we could use some database-driven approaches such as Rosetta (Leaver-Fay et al., 2011) or iTasser (Roy et al., 2010). The other mutants are as well expressed as the native CcMan5. This was also seen in the MD simulations, as the RMSF curves were almost identical. It's also logic that, most of the time, a point mutation would have less impact on the overall structure then a long deletion. As the periplasmatic fraction is used for purification and treatment of the enzymes for ERT, it is required that the protein is expressed in this fraction. So this long deletion mutant is not usefull.

IV.1.3 Activity analysis results

For the activity test we used DSA-FACE, a fluorophore based carbohydrate identification technique. For the calibration we used dextran and RNase B, while our negative control were sugars obtained from a MNN4 *Yarrowia lypolitica* overexpressed strain. We loaded the periplasm samples from CcMan5.1, CcMan5.2, the long deletion mutant and the two point mutants (Figure 22 and Figure 23). The CcMan5.1 and CcMan5.2 profiles were identical, as expected due to the fact that their amino acid content is identically. As the expression is the same and the activity stays, we can use CcMan5.2 instead of CcMan5.1.

The Ala631 profile is also identically to the native CcMan5 profiles. This implicates that it is also active. If the mutant has the same activity, we could say that this amino acid is not

needed for N-glycan binding. This could be an indication that the N-glycan binds the opposite direction. The Ala351 mutant is missing the $GlcNAc_2Man_8$ double uncapped peak at 1500 in the profile after one-hour incubation but the peak appears in the overnight profile. The protein needs more time so that in all probability the Ala351 mutant has a slower conversion. The leucine is filling a small hole near the active site, which could result in less robust binding of the N-glycan. As the mutation is still active, the Ala351 residue is not essential in the reaction pathway. The long mutant is not active, which is logical as almost no protein was available in the periplasmatic fraction. To compare the results in more detail we should calculate the concentration based on western blots, where after we could compare dilution series of each mutant with the native form. With these dilution series we can conclude if the activity is really higher or lower. At this point we can only say if there is activity or not.

IV.1.4 Autodock Vina results

The docking results show us that regio-isomer B binds with a higher free energy than regio-isomer A with a difference of 1kJ/mol. This could explain the reason for more single uncapping. Docking of the two regio-isomers and the N-glycan showed us also that a lot of positions are possible, although not every position is perfect for the reaction as Tyr620 and Gln72 cannot make H-bonds. The pointed directions are in both ways. The A regio-isomer is more directed away from the Ala631 residue and the B regio-isomer towards the Ala631 residue. A very good conformation isn't found.

IV.1.5 Purification of CcMan5

The purification of CcMan5 was based on the available N-terminal His-tag. If we look at the Coomassie gels of different steps in the purification of CcMan5 (Figure 24), we can conclude that with this one step process almost only pure CcMan5 stays in solution. But the quantities of CcMan5 protein were too low. The reason for failure can be that we used only small quantities of start culture (1l), where normal 10l is used. In the long deletion mutant purified fraction, the concentration was much higher but there were also more other proteins in the purified fraction. As the long deletion mutant was purified as last on the same column, we suppose that the rest fractions of the point mutants stayed on the column. We can conclude that one-step purification

IV.2 Conclusion

The deletion mutant did not express in the periplasmatic fraction, probably due to folding problems. The Ala351Leu point mutant is presumably less active then the wild type and Ala631Leu mutant. The Ala351Leu mutation is filling a small hole near the active site, which could result in less robust binding of the N-glycan. As the mutation is still active, the Ala351 residue is not essential in the reaction pathway. If the Ala631Leu mutant has the same activity, we could say that this amino acid is not needed for N-glycan binding. This could be an indication that the N-glycan binds the opposite direction. The purification with IMAC is enough to purify the proteins but the start culture needs to be higher.

The docking results showed that it is possible that the two regio-isomers bind the active site in an opposite manner. The free bindings energy of the best results of one of the regio-isomers is 1kJ/mol higher than the other. This could be an explanation for the favourable uncapping of just one mannose from the Man-Pi-6-Man's on enzymes for ERT.

IV.3 Use of combinatorial approach

The combination of computational and experimental work is often used these days. The force of the computers and the improved databases results in better predictions. As seen in this project, computational work is a guideline for experimental work. It does not replace it. The real proof must still be given by wet lab set-ups. Molecular dynamics can give an idea of the stability of a protein but they can't predict protein folding. Therefore more computationaly intensive programs are developed. The dockings are a good indication for what's happening but we must be careful with the results. Still, I think that a combination of experimental and computation work could be useful.

IV.4 Future perspectives

IV.4.1 Future experimental perspectives for CcMan5

It is still too early to jump into real conclusions. Further investigation on CcMan5 and its N-glycan binding is needed, to be sure of the hypothesis. First we need to make dilution series of the mutants and compare these with dilution series of the native form to see whether the activity is really higher or lower. For the activity test we used the periplasm extracts. Another thing that could be tested is the use of the purified protein, therefore it is recommended to do the purification again with a larger start volume of medium so that the concentration is higher.

Some more mutations could be made, as the system based on synthetic synthesis is faster. A possible mutation could be a deletion mutant in the on the other side of the active site. Another one could be the short deletion mutant that could have less folding problems. If these could be expressed in the periplasmatic fraction, we could learn more about the binding of the N-glycan. The next step should be cloning of the CcMan5 gene in the yeast strains where the enzyme for ERT is already available. But success is not guaranteed.

IV.4.2 Future computational perspectives

To dock the full N-glycan we need an option to determine a position for the Man-Pi-6-Man so that less false results are obtained. Another option might be to start from a good docking result of the regio-isomers and build op the fully N-glycan. With molecular dynamics we could simulate the movement of the N-glycan and detect what the minimal energy positions are. When a good conformation is found, QM/MM could be performed to study the reaction mechanism. In this approach, the reactive part of the molecular system is described by a quantum mechanical (QM) method, whereas the rest of the system is described using molecular mechanics force fields (MM). This enables us to simulate a reaction in a protein at an affordable computational cost. Thermodynamic and kinetic information could be obtained by determining the free energy profile for the entire reaction. The latter could be calculated with the "nudged elastic band" (NEB) method, which tries to find the minimum energy path, between reactant, intermediate and product states. Based on the free energy profile, the reaction rate can be estimated. And so, the effect of structural alterations in the receptor on this reaction rate can be determined, opening up the possibility to perform *in silico* screening of CcMan5 mutations, near the active centre.

IV.4.3 Future perspectives for ERT production in yeast

Production of enzymes for ERT in yeast could be a huge impact on the live quality of patients, so it is needed to investigate more in this domain. CcMan5 is not specific for one enzyme for ERT but for a whole bunch of enzymes for different LSDs as they also need the M6PR delivery to the lysosome. The development of new therapies for LSDs that are not yet possible to treat with ERT will become easier in yeast production systems. The advantage of CcMan5 over other uncapping solutions (Urbanelli et al., 2011; Zhu et al., 2009) is the fact that no chemical reaction pathways are needed. The CcMan5 gene could be cloned in yeast, where it works directly on the ERT produced in the same strain.

New therapeutics for orphan diseases, including LSDs, has exclusivity of 7 years on the market in the VS and 10 years in Europe. This makes it more attractive to the pharmaceutical industry. Oxyrane Belgium NV is already interested in the production of enzymes for ERT in yeast cells. They are specialised in production of human proteins in microorganisms. One of their microorganisms of interest is *Yarrowia lipolytica*.

Thereby the discussion rises in the Netherlands if the health insurance should pay the cost of these expensive therapies. They put forward that the results of the therapies, living a couple years longer, are not balanced with the input of money. Others proclaimed that person with LSD's did not choose for the disease, so we need to treat them as we treat every patient with another disease. This is an ethical question without a good, unambiguous answer. This indicates again the need for cheaper alternatives. As conclusion we can say that a lot of work is done but some more needs to be done.

V Samenvatting

Lysosomale opslag ziekten (LSDs) zijn een groep van meer dan 50 overerfbare, monogenische aandoeningen (Schultz et al., 2011). LSDs zijn meestal het gevolg van gebrekkige substraat degradatie door verliesmutaties in genen coderend voor lysosomale hydrolases. Een belangrijke pathway om enzymen naar het lysosomoom te transporteren is gebaseerd op modificaties van deze enzymen met Pi-6-Man groepen. Een kritische factor in het succes van ERT is dat de enzymen een hoge densiteit aan Pi-6-Man modificaties op hun N-glycanen nodig hebben (Wraith, 2006; Zhu et al., 2004). Dit glycotoop wordt herkend door de M6PR aanwezig op het oppervlak van humane cellen, waardoor het receptor:enzym complex wordt opgenomen door endocytose. Het enzym wordt dan afgeleverd in het lysosoom waar het enzymatisch defect wordt gecomplementeerd. Er is een ERT op de markt voor Gaucher (Beutler et al., 1991), Fabry (Beck, 2002) en de ziekte van Pompe (VanHove et al., 1996). Op dit moment zijn de behandelingen voor LSD zeer duur en vaak niet efficient. Dit komt door dat het productie in zoogdiercellen complex is en het aantal Pi-6-Man modificaties laag is. Genzyme en Shire hebben momenteel een monopolie over de productie van enzymen gebruikt voor ERT. Gelukkig lopen de patenten voor ERT voor Gaucher, de ziekte van Pompe en Fabry bijna af, waardoor er ruimte is voor nieuwe technologiën.

CcMan5 is een belangrijke factor wanneer de enzymen voor ERT in gist worden geproduceerd. Het zorgt namelijk voor de omzetting van de in gist geproduceerde N-glycanen naar zoogdierachtige. Deze N-glycanen zijn nodig voor de opname van het ERT door de M6PR in de lysosomen. Een hogere conversiesnelheid van deze gist N-glycanen zal resulteren in een minder lange incubatie met het gist supernatans, wat de therapeutische enzymen bevat, resulterend in minder defosforylatie door de endogene fosfatases. Enzymen met meer Pi-6-Man zullen beter binden met de M6PR waardoor de opname in de lysosomen van de patiënten beter is. Dit vermindert de hoeveelheid enzym dat nodig is voor therapie. Het resulteert ook in een hogere efficiëntie van het enzym, wat belangrijk is voor de levenskwaliteit van de patiënt.

Om de efficiëntie van het CcMan5 enzym te verbeteren en meer te weten te komen over de werking, produceerden we mutanten van het enzym. Dit werd gecombineerd met computationele analyse van de stabiliteit van deze enzymen en docking experimenten. De mutanten die we analyseerden werden gedefinieerd met de kennis van vorige mutanten, alligneringen en structurele informatie. De mutanten waren de Ala351Leu en de Ala631Leu punt mutant maar ook de lange (Y615-P642) en korte (S634-P642) deletie mutant.

V.1 Discussie van experimentel en computationele resultaten

V.1.1 Productie van mutanten

We startte met klassieke plaatsgerichte mutagenese zoals beschreven in Figuur 16. Dit was niet echt een succes, behalve voor de Ala631 mutant. De rede van dit falen is onverklaarbaar aangezien we verschillende parameters uitprobeerden alsook verschillende herhalingen door meerdere personen. Voordien had men ook al problemen om mutaties aan te maken op hetzelfde gen. We beslisten om het opkomende synthetisch synthese van DNA te gebruiken, wat reeds sterk in prijs en levertijd is gedaald waardoor het steeds meer en meer

gebruikt kan worden. Voor dit project bestelden we een synthetisch DNA fragment (1785bp) voor het CcMan5 gen, waar we meer restrictiesites aan toevoegden. Alle toegevoegde sites staan gekend als goed verknippend en zorgen meestal voor overhangende eindjes. De nieuwe restrictiesites werden zodanig gekozen dat de aminozuren niet veranderden. Dit is zeer belangrijk aangezien deze sequentie dezelfde moet blijven om de activiteit te behouden. Voor al de nieuwe codons werd het voorkomen in *E. coli* nagekeken, aangezien de translatie efficiëntie hoog moet blijven. De zoeken naar mogelijk nucleotide veranderingen, resulterend in een restrictiesite, werd handmatig gedaan. In combinatie met kleine fragmenten (300bp), waar de mutaties reeds aanwezig zijn, is het mogelijk om gemuteerde stammen te maken in slechts een week. Het is niet aangewezen om gelijktijdig twee enzymen te gebruiken die resulteren in stompe eindjes want hierdoor kunnen de fragmenten in verschillende richtingen worden geïntegreerd waardoor extra selectie genoodzaakt is. Met dit systeem slaagden we erin om de mutanten te maken in twee weken, terwijl we enkelen maanden probeerden met het klassieke klonering systeem. De korte deletie mutant maakten we niet omdat we eerst de lange mutatie wouden uittesten. Deze mutatie zou meer impact hebben en we wouden ook eerst het systeem uittesten.

V.1.2 Expressie van de mutanten vergeleken met de moleculaire dynamica resultaten

Wanneer we de gemuteerde plasmiden hadden aangemaakt, induceerden we expressie door IPTG in de *E. coli* BL21 stam. Eerst werd via een ruw extract getest of er wel degelijk eiwit aanwezig was (Figuur 19). Dit werd ook gedaan voor de CcMan5.2 stam om te kijken of het codon gebruik geen invloed had op de expressie. CcMan5.1 wordt even goed aangemaakt als CcMan5.2 zodat CcMan5.2 kan gebruikt worden. Daarna werd voor elke mutant en het natieve eiwit verschillende fracties bereid om na te gaan waar het CcMan5 eiwit zich bevond (Figuur 20 en Figuur 21). Uit deze figuren kunnen we besluiten dat de meerderheid van het eiwit in de oplosbare en onoplosbare fractie blijft. Enkel een klein deel wordt naar het periplasma gestuurd. De periplasmatische fractie wordt gebruikt voor de activiteitstesten omdat de hoeveelheid aan andere eiwitten laag is.

Voor de lange deletie mutant is er geen eiwit in het periplasma en het medium aanwezig. De reden hiervoor is wellicht niet correcte opvouwing. Als we kijken naar de MD resulaten, is daar ook de lange deletie mutant minder stabiel dan het natieve CcMan5 (Figuur 36 en Figuur 37), maar dit zegt alleen iets over reeds opgevouwen structuren. Om *ab initio* opvouwing te bestuderen, is de computationele tijd de limiterende factor. We kunnen wel gebruik maken van databasegedreven benaderingen zoals Rosetta (Leaver-Fay et al., 2011) of iTasser (Roy et al., 2010). De puntmutanten worden wel even goed aangemaakt. Dit konden we ook zien aan de MD simulaties, waar de RMSF grafieken ongeveer identiek waren. Het is ook logisch dat een puntmutant minder impact heeft op de structuur dan een lange deletie mutant. Doordat de periplasmatische fractie gebruikt word voor zuivering en behandelig van den enzymen voor ERT, is het vereist dat het eiwit wordt aangemaakt in deze fractie. Dit maakt dat de lange deletie mutant niet bruikbaar is.

V.1.3 Activiteitsanalyse resultaten

Voor de activiteitstesten maakten we gebruik van DSA-FACE, een fluorofoor gebaseerde carbohydraat identificatie techniek. Voor de ijking gebruiken we dextraan en RNase B, terwijl we als negatieve controle suikers van de MNN4 *Yarrowia lypolytica* overexpressie mutant gebruikten. We laden de periplasma stalen van CcMan5.1, CcMan5.2, de lange deletie

mutant en de puntmutanten (Figuur 22 en Figuur 23).De CcMan5.1 en CcMan5.2 profielen waren identiek, zoals verwacht want hun aminozuur sequentie is dezelfde. Dit geeft ons nog een extra bevestiging dat we CcMan5.2 kunnen gebruiken in plaats van CcMan5.1.

Het Ala631 profiel is ook identiek aan de natieve CcMan5 profielen. Het eiwit is dus ook actief. De Ala351 mutant mist, in het profiel na 1 uur incubatie, de $GlcNAc_2Man_8$ piek waarbij de twee eindstandige mannoses zijn verwijdert. Deze piek is echter wel aanwezig in het profiel van de overnacht geïncubeerde stalen. Het eiwit heeft dus meer tijd nodig, wat laat uitschijnen dat deze mutant minder actief is. De Ala351Leu mutatie vult naast de actieve site een kleine gaatje op, wat zou kunnen zorgen voor minder binding van het N-glycaan. Omdat de mutatie nog steeds actief is, zal het Ala351 geen essentiële rol hebben in de reactie pathway. De lange deletie mutant is niet actief, maar dat is logisch aangezien er bijna geen eiwit in de periplasmatische fractie aanwezig was. Om meer in detail te weten te komen, moeten we de concentraties berekenen met behulp van western blots. Waarna we verdunningsreeksen van de mutanten met het natieve CcMan5 kunnen vergelijken zodat we kunnen zeggen of de activiteit hoger of lager is. Op dit punt kunnen we alleen zeggen of er activiteit is of niet.

V.1.4 Autodock Vina resultaten

De docking resultaten tonen dat regio-isomeer B met een hogere vrije energie bindt dan regio-isomeer A. Het verschilt bedraagt 1kJ/mol. Dit kan de reden zijn voor meer enkele ontkapping. Het docken van de twee regio-isomeren en het N-glycaan toont aan dat veel posities mogelijk zijn maar niet elke posities is correct doordat Tyr620 en Gln72 geen waterstofverbindingen kunnen maken. Beide richtingen ten opzichte van de actieve site zijn mogelijk. Het A-regio isomeer is vaker gericht weg van het Ala631 residue en het B-regio isomeer naar het Ala631 residue. Een zeer goede conformatie werd niet gevonden.

V.1.5 Zuivering van CcMan5

Voor de zuivering van CcMan5 werd gebruik gemaakt van de N-terminale His-tag. Wanneer we naar de Coomassie gels van de verschillende stappen kijken (Figuur 24), kunnen we stellen dat na het éénstaps-proces bijna alleen CcMan5 in de oplossing blijft. Maar de hoeveelheid eiwit was laag. Dit kunnen we verklaren door het feit dat we gestart waren van 1L cultuur terwijl normaal 10L wordt gebruikt. In de gezuiverde fractie van de lange deletie mutant was de concentratie hoger maar er waren ook meer onzuiverheden. De deletie mutant werd als laatste gezuiverd over dezelfde kolom. We veronderstellen dat de restfracties van de puntmutanten en eventueel ook afbraak producten op de kolom bleven. We kunnen besluiten dat het éénstaps-proces met IMAC voldoend is om het eiwit te zuiveren maar de hoeveelheid startcultuur moet hoger zijn.

V.2 Conclusie

De deletie mutant werd niet aangemaakt in het periplasma, wellicht door problemen met de opbouwing. De Ala351Leu puntmutant is wellicht minder actief dan het wildtype en de Ala631Leu mutant. De Ala351Leu mutatie voelt een kleine uithollig naast de actieve site op, wat kan resulteren in minder sterke binding van het N-glcyaan. Omdat de mutatie nog actief is, is het Ala351 residue niet essentieel in de reactie pathway. Als de Ala631Leu mutant de zelfde activiteit heeft, dan kunnen we ervan uitgaan dat dit aminozuur niet nodig is voor N-glycaan binding. Dit kan een indicatie zijn voor het feit dat het N-glycaan in de andere

richting bindt. De zuivering met IMAC is voldoende, maar het volume van de startcultuur moet hoger zijn.

De docking resultaten tonen aan dat het mogelijk is dat de twee regio-isomeren de actieve site in een tegenovergestelde richting binden. De vrije bindingsenergie van de beste resultaten van het een regio-isomeer verschilt 1kJ/mol met het ander. Dit kan verklaren waarom vaker slechts een mannose van de Man-Pi-6-Man's op de enzymen voor ERT wordt verwijdert.

VI Material and Methods

VI.1 Experimental part

VI.1.1 Preparation of plasmid with mutated gene

VI.1.1.1 Site Directed mutagenesis for Ala 631 mutant

The original pLSAHCcMan5domain plasmid was available, containing the CcMan5 domain under control of a pL-promoter and selectable by ampiciline antibiotics resistance. First a 50 µl PCR was performed (QuikChange site-directed mutagenesis kit, (Agilent)), composed of 1µl Phusion polymerase, 50ng original pLSAHCcMan5domain plasmid, 10mM dNTPs, 5X Phusion Buffer and 150ng of each primer (ccggtagcagctctctggttccgtcaggtggtggtg and ccacctgacggaaccagagagctgctaccggttgc, Integrated DNA Technologies, for Ala631 mutant). The annealing temperature was set on 65°C for 35s. Analysis was done with agarose gel (1,3%) electrophoresis in TAE (Tris-Acetate-EDTA) buffer. As reference the 1kb BenchTop ladder (Promega) was used. Fragments were separated on 105V during 50 min and visualised with ethidium bromide and Gel Doc 2000 (BioRad). The fragments were purified out of the agarose gel with Nucleospin® Gel and PCR Clean-up (Macherey-Nagel). A DpnI digest was performed to remove the original methylated, not mutated plasmids. Therefore 1µl DpnI was added to the 50µl PCR reaction for 1h at 37°C. The resulting plasmids were transformed by a heat shock method to compatible *E. coli* MC1061 cells. To test the correctness, the clones were grown up in 5ml Luria Bertani (LB) medium with ampicillin (100µg/ml). Then the plasmids were purified with QIAprep® Miniprep kit (Qiagen) and the concentrations were measured with Nanodrop (ND-1000, Isogen Life Science). To assure that the Ala631 mutation occurred and no secondary mutations, the gene was sequenced by VIB facilities using T7 forward and T3 reverse primers.

VI.1.1.2 Restriction digest for Ala 351 and long deletion mutant

The new synthetic gene (CcMan5.2) and shorter fragments, with Ala351 mutation and deletion, were ordered at GenScript. First a digest on the CcMan5.2 gene in the pUC57 and on the pLSAHCcMan5domain was performed in 20µl with NsiI (Promega) and BglII (Promega) in compatible buffer D (Promega). The fragments, respectively 2,708 bp and 1,798 bp, were purified out of agarose gel (1,3%) with the Nucleospin® Gel and PCR Clean-up kit (Macherey-Nagel). The fragments were ligated to one other with T4 DNA ligase (Fermentas). Afterwards the ligation reaction was transformed to *E. coli* MC1061 λ. To obtain a larger quantity of plasmid the QIAprep® spin Miniprep kit (Qiagen) on 6ml of culture was used. For the Ala351 mutations a restriction digest with SpeIGQ (Promega) and HindIII (Promega) in buffer E (Promega) was performed on the pLSAHCcMan5.2 plasmid and the pUC57 with synthetic fragment. The 5,045 bp and 310 bp long fragments were excised. After gel clean up, the fragments were ligated with T4 DNA ligase (Fermentas) and transformed again to *E. coli* MC1061 λ. For the deletion mutant first KpnI (Promega) in buffer J (Promega) and afterwards NcoI (Promega) in buffer D (Promega) was used on both plasmids. In between the two restriction digests, a Nucleospin® clean up (Macherey-Nagel) was done to remove the first buffer. After both digests we loaded the fragments on agarose gel (1,3%) in TAE buffer. After gel clean up, the T4 DNA ligase (Fermentas) reaction was transformed to *E. coli*

MC1061 λ. To check the deletion mutant we did a colony PCR with T7 forward and T3 reverse primers and both mutated genes were sequenced with the same primers by VIB facilities.

VI.1.2 Induction and lysis

After cloning, the different mutated *CcMan5* genes were transformed into *E. coli* BL21 + pICa2 expression strain. The transformed strains were grown in Luria Bertani (LB) media with ampicillin (100μg/ml) and kanamycin (50μg/ml) overnight at 28°C. The day after, the cultures were brought to an optimal density (OD) of 0.2. After two hours at 20°C they reach an OD of 0.5 and were induced overnight with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). To test if there was induction, the cells were broken with glass beads and analyzed by SDS-PAGE and Western blotting with a mouse monoclonal anti-pentaHis antibody (Qiagen) and anti-mouse DyLight 800 (Thermo Scientific) on the Odyssey system.

VI.1.3 Isolation of different cell fractions

The genes were expressed in 2 ml LB culture of *E. coli* BL21 with ampicillin (100μg/ml) and kanamycin (50μg/ml) overnight at 20°C. IPTG was added the day after at an OD 0.5. The medium was removed by centrifugation at 5,000 rpm for 5 min at 4°C and the supernatant was discarded. The cells were lysed by adding 1.5ml buffer (30mM Tris/HCl pH8, 20% sucrose, 1mM EDTA) to the pellet and centrifugation at 12,500 rpm for 10 min at 4°C. The supernatant was discarded and 1.5ml of cold 5mM $MgCl_2$ was added. After 10 min shaking in an ice bath and centrifugation at 12,500 rpm for 10 min at 4°C, the periplasm was collected. Then 0,4ml cold 20mMTrisHCl pH 7.5 and 4μl 10mg/ml lysozyme (fc=100 μg/ml) were added to the pellet. The solution was incubated at 30°C for 15 min, where after the samples were sonicated on ice for 4 times 15 s. Then we centrifuged at 14,000 rpm for 10 min. The obtained supernatant is the soluble fraction. Then 100μl of 20mM TrisHCl pH 7.5 was added to the remaining pellet and it was centrigugated for 10 min at 10,000 rpm. We repeated this step, discarded the supernatans and added 450 ml 1% SDS to resolve the pellet. This is the insoluble fraction. The samples were loaded onto SDS-PAGE to perform Coomasie colouring and Western blotting.

VI.1.4 Activity analysis

First activity was measured on the periplasm extracts. The samples (3.6μl) (in duplicate) were added to a mix with GAA sugars obtained from a MNN4 overexpressing strain, 20mM $CaCl_2$ and 100mM Hepes pH7.5. This was incubated for 1h at 37°C and the reaction was stopped by incubating it for 5 min at 100°C. The duplicate was incubated overnight at 37°C. To analyze the activity 2μl was used for DSA-FACE.

To compare the activity of each mutation variant, the concentration of CcMan5 in the periplasm was measured by making dilution series starting from 5μl fraction. These series were loaded onto a SDS-PAGE gel and analysed by Western Blot. The fractions and PNGase standards (150ng and 300ng) were loaded on Tricine-SDS-PAGE gels (12%) and visualised by Western Blot.

VI.1.5 Purification

The mutated *CcMan5* genes were expressed in *E. coli* strain BL21 + pICa2 with the pLSAHCcMan5domain expression vector. The bacteria of the Ala351, Ala631 and long

deletion were grown in 40 ml of LB medium with ampicillin (100μg/ml) and kanamycin (50μg/ml) overnight at 28°C. From these precultures three 350 ml cultures for each mutant were started with OD 0.2 in LB medium. When they reached an OD of 0.5, 1mM IPTG was added overnight at 20°C. Cells were harvested and frozen at -20°C. After thawing, the cells were gently resuspend at a concentration of 3ml/g in 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 1mM PMSF and 10 μg/ml DNaseI. The periplasmic fraction was prepared by stirring the cell suspensions for 1h at 4°C and was isolated by centrifugation at 18,000 x g for 30 min. The supernatant was applied to a 1 ml Ni-Sepharose 6 FF column (GE Healtcare), equilibrated with 20 mM NaH_2PO_4 pH 7.4, 300 mM NaCl, 20 mM imidazole, 0.1% CHAPS. The column was eluted with 20mM NaH_2PO_4 pH7.4, 20mM NaCl, 400mM imidazole, 0.1% CHAPS.

The concentrations of the purified proteins was measured by BCA Protein Assay Kit (Thermo Scientific, Pierce). After generating a standard curve, the concentration of the protein was calculated.

VI.2 Computational part

VI.2.1 Topology file

We started with the 2XSG.pbd from the PDB site with chain A en B. With the following script, executable in vmd with tcl, we created a PDB and the PSF or topology file of chain A and the associated calcium ions. We solvated the protein in a water box of size 10x10x10 angstroms, similarly for the mutated protein variants. Deletion protein variants were obtained by removing the corresponding amino acids from the original PDB file. To introduce mutations, a mutate patch statement was employed.

```
mol new 2XSGaddedH.pdb
set ProtA [atomselect top "protein and chain A"]
$ProtA writepdb ProtA.pdb
set Calc [atomselect top "not protein and name CA"]
set Cxyz [$Calc get {x y z}]
set Cx [lindex $Cxyz 0]
set Cy [lindex $Cxyz 1]
set Cz [lindex $Cxyz 2]
mol delete all
package require psfgen
topology top_all27_prot_na.rtf
pdbalias residue HIS HSE
pdbalias atom ILE CD1 CD
segment A {
pdb ProtA.pdb
}
coordpdb ProtA.pdb A
segment C {
residue 1 CAL
}
coord C 1 CAL "$Cx $Cy $Cz"
guesscoord
writepdb CcMan5.pdb
writepsf CcMan5.psf
package require solvate
solvate CcMan5.psf CcMan5.pdb -o CcMan5s -t 10
quit
```

VI.2.2 Energy Minimization

All parameters and input files were defined in a configuration script. The run time, nodes and cpu were determined in a shell script. Energy minimization of the molecules was performed with NAMD and for 5000 steps. The input files were the PDB and PSF files created with tcl in vmd. The parameter file, par_all27_prot_na.prm, contains the CHARMM forcefield parameters for all amino acids. To set the periodic boundary conditions properly, the centre of the protein was measured and defined in the configuration file. The computations were executed at the Stevin supercomputer infrastructure of the Ghent University.

Fragments of the configuration script:

```
#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################
structure         CcMan5s.psf
coordinates       CcMan5s.pdb
set temperature    300
set outputname     CcMan5
firsttimestep      0

#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm              on
parameters         par_all27_prot_na.prm
temperature         $temperature

# Periodic Boundary Conditions
cellBasisVector1    89.7    0.   0.0
cellBasisVector2     0.   78.2   0.0
cellBasisVector3     0.    0   104.5
cellOrigin        45.3 22.7 76.6

#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################

# Minimization
minimize           5000
```

VI.2.3 Molecular dynamics

The overall script for MD is quite the same as the script for energy minimization, some additional parameters needed to be set. First a timestep must be defined. As we are dealing with quite a big protein, we declared H-bonds non flexible and this implicated that the recommended timestep was 2fs. As writing every for every step the results would be to computational demanded, we decided to write only every 1000 steps. In totally we wanted to perform 10^6 steps. The temperature for the MD simulations was set at 300K or 37°C.

```
# Integrator Parameters
timestep          2.0
rigidBonds        all
nonbondedFreq     1
fullElectFrequency  1
stepspercycle     10

# Output

outputName        $outputname

restartfreq       1000    ;# 500steps = every 1ps
dcdfreq           1000
xstFreq           1000
outputEnergies    500
outputPressure    1000


#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################


# Minimization

reinitvels    $temperature
minimize      5000
run           500000
```

VI.2.4 RMSD and RMSF plots

To calculate the RMSD and RMSF plots we used the GROMACS module on the hpc. But first we saved the dcd file, the output of the MD simulation, as a pdb file with vmd. For RMSF we decided to calculate the flexibility of each residue so that calculations were done for each Cα. For the RMSD value we calculated the movement over time of the Cα's. This gave us a xvg file.

module load GROMACS

RMSF calculation: g_rmsf –f x.pdb –s x.pdb

RMSD calcution: g_rms –f x.pdb –s x.pdb

Now we needed to convert these xvg files to a plot in gif format. For this step we used Gnuplot on the hpc.

gnuplot

set term gif

set output "x.gif"

plot 'x.xvg' u 1:2: w l

VI.2.5 AutoDock Vina

First we made the pdbqt file for the rigid receptor:

source ~/initAutodock.sh

pythonsh $ADTlib/prepare_receptor4.py –r receptor.pdb –v –U nphs_lps_waters

Then we made the pdbqt file for the flexible receptor:

pythonsh $ADTlib/prepare_receptor4.py –r receptor.pdb –v –s :A:TYR535_GLN536

Now we prepared the pdbqt file for the ligand. First the water molecules in the pdb-files must be removed.

pythonsh $ADTlib/prepare_ligand4.py –l ligandpdb –R

The configuration file is composed as follow:

receptor = receptor.pdbqt
flexible= flexrec.pdbqt
ligand =ligand.pdbqt

center_x = 35.119
center_y = 21.558
center_z = 75.067

size_x = 15
size_y = 15
size_z = 15

num_modes = 15
exhaustiveness = 60

References

Akeboshi, H., Kasahara, Y., Tsuji, D., Itoh, K., Sakuraba, H., Chiba, Y., and Jigami, Y. (2009). Production of human beta-hexosaminidase A with highly phosphorylated N-glycans by the overexpression of the Ogataea minuta MNN4 gene. Glycobiology *19*, 1002-1009.

Ausems, M.G., Verbiest, J., Hermans, M.P., Kroos, M.A., Beemer, F.A., Wokke, J.H., Sandkuijl, L.A., Reuser, A.J., and van der Ploeg, A.T. (1999). Frequency of glycogen storage disease type II in The Netherlands: implications for diagnosis and genetic counselling. Eur J Hum Genet *7*, 713-716.

Bach, G., Eisenberg, F., Jr., Cantz, M., and Neufeld, E.F. (1973). The defect in the Hunter syndrome: deficiency of sulfoiduronate sulfatase. Proc Natl Acad Sci U S A *70*, 2134-2138.

Beck, M. (2002). Agalsidase alfa--a preparation for enzyme replacement therapy in Anderson-Fabry disease. Expert Opin Investig Drugs *11*, 851-858.

Beutler, E., Kay, A., Saven, A., Garver, P., Thurston, D., Dawson, A., and Rosenbloom, B. (1991). Enzyme replacement therapy for Gaucher disease. Blood *78*, 1183-1189.

Beutler, E., Nguyen, N.J., Henneberger, M.W., Smolec, J.M., McPherson, R.A., West, C., and Gelbart, T. (1993). Gaucher disease: gene frequencies in the Ashkenazi Jewish population. Am J Hum Genet *52*, 85-88.

Bijvoet, A.G., Kroos, M.A., Pieper, F.R., de Boer, H.A., Reuser, A.J., van der Ploeg, A.T., and Verbeet, M.P. (1996). Expression of cDNA-encoded human acid alpha-glucosidase in milk of transgenic mice. Biochim Biophys Acta *1308*, 93-96.

Bijvoet, A.G., Kroos, M.A., Pieper, F.R., Van der Vliet, M., De Boer, H.A., Van der Ploeg, A.T., Verbeet, M.P., and Reuser, A.J. (1998). Recombinant human acid alpha-glucosidase: high level production in mouse milk, biochemical characteristics, correction of enzyme deficiency in GSDII KO mice. Hum Mol Genet *7*, 1815-1824.

Bijvoet, A.G., Van Hirtum, H., Kroos, M.A., Van de Kamp, E.H., Schoneveld, O., Visser, P., Brakenhoff, J.P., Weggeman, M., van Corven, E.J., Van der Ploeg, A.T.*, et al.* (1999). Human acid alpha-glucosidase from rabbit milk has therapeutic effect in mice with glycogen storage disease type II. Hum Mol Genet *8*, 2145-2153.

Blanz, J., Groth, J., Zachos, C., Wehling, C., Saftig, P., and Schwake, M. (2010). Disease-causing mutations within the lysosomal integral membrane protein type 2 (LIMP-2) reveal the nature of binding to its ligand beta-glucocerebrosidase. Hum Mol Genet *19*, 563-572.

Brady, R.O., Gal, A.E., Kanfer, J.N., and Bradley, R.M. (1965). The metabolism of glucocerebrosides. 3. Purification and properties of a glucosyl- and galactosylceramide-cleaving enzyme from rat intestinal tissue. J Biol Chem *240*, 3766-3770.

Braulke, T., and Bonifacino, J.S. (2009). Sorting of lysosomal proteins. Biochim Biophys Acta *1793*, 605-614.

Brooks, A.I., Chattopadhyay, S., Mitchison, H.M., Nussbaum, R.L., and Pearce, D.A. (2003). Functional categorization of gene expression changes in the cerebellum of a Cln3-knockout mouse model for Batten disease. Mol Genet Metab *78*, 17-30.

Bultron, G., Kacena, K., Pearson, D., Boxer, M., Yang, R., Sathe, S., Pastores, G., and Mistry, P.K. (2010). The risk of Parkinson's disease in type 1 Gaucher disease. J Inherit Metab Dis *33*, 167-173.

Callewaert, N. (2011). Course Notes: Biomolecular Production Methods (UGent).

Callewaert, N., Geysens, S., Molemans, F., and Contreras, R. (2001). Ultrasensitive profiling and sequencing of N-linked oligosaccharides using standard DNA-sequencing equipment. Glycobiology *11*, 275-281.

References

Callewaert, N., Vervecken, W., Van Hecke, A., and Contreras, R. (2002). Use of a meltable polyacrylamide matrix for sodium dodecyl sulfate-polyacrylamide gel electrophoresis in a procedure for N-glycan analysis on picomole amounts of glycoproteins. Anal Biochem *303*, 93-95.

Campbell, C.H., and Rome, L.H. (1983). Coated vesicles from rat liver and calf brain contain lysosomal enzymes bound to mannose 6-phosphate receptors. J Biol Chem *258*, 13347-13352.

Cardone, M., Porto, C., Tarallo, A., Vicinanza, M., Rossi, B., Polishchuk, E., Donaudy, F., Andria, G., De Matteis, M.A., and Parenti, G. (2008). Abnormal mannose-6-phosphate receptor trafficking impairs recombinant alpha-glucosidase uptake in Pompe disease fibroblasts. Pathogenetics *1*, 6.

Cho, A., Kim, S.J., Lim, B.C., Hwang, H., Park, J.D., Kim, G.B., Jin, D.K., Lee, J., Ki, C.S., Kim, K.J*., et al.* (2012). Infantile Pompe disease: clinical and genetic characteristics with an experience of enzyme replacement therapy. J Child Neurol *27*, 319-324.

Clarke, L.A., Wraith, J.E., Beck, M., Kolodny, E.H., Pastores, G.M., Muenzer, J., Rapoport, D.M., Berger, K.I., Sidman, M., Kakkis, E.D*., et al.* (2009). Long-term efficacy and safety of laronidase in the treatment of mucopolysaccharidosis I. Pediatrics *123*, 229-240.

Corchero, J.L., Mendoza, R., Lorenzo, J., Rodriguez-Sureda, V., Dominguez, C., Vazquez, E., Ferrer-Miralles, N., and Villaverde, A. (2011). Integrated approach to produce a recombinant, His-tagged human alpha-galactosidase A in mammalian cells. Biotechnol Prog *27*, 1206-1217.

Cox, T.M. (2010). Gaucher disease: clinical profile and therapeutic developments. Biologics *4*, 299-313.

Cregg, J.M., Madden, K.R., Barringer, K.J., Thill, G.P., and Stillman, C.A. (1989). Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. Mol Cell Biol *9*, 1316-1323.

Dahms, N.M., and Hancock, M.K. (2002). P-type lectins. Biochim Biophys Acta *1572*, 317-340.

De Duve, C., Pressman, B.C., Gianetto, R., Wattiaux, R., and Appelmans, F. (1955). Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. Biochem J *60*, 604-617.

De Pourcq, K., Tiels, P., Van Hecke, A., Geysens, S., Vervecken, W., and Callewaert, N. (2012). Engineering Yarrowia lipolytica to Produce Glycoproteins Homogeneously Modified with the Universal Man(3)GlcNAc(2) N-Glycan Core. PLoS One *7*, e39976.

Demain, A.L., and Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv *27*, 297-306.

Desnick, R.J., and Schuchman, E.H. (2002). Enzyme replacement and enhancement therapies: lessons from lysosomal disorders. Nat Rev Genet *3*, 954-966.

Dutsch, M., Marthol, H., Stemper, B., Brys, M., Haendl, T., and Hilz, M.J. (2002). Small fiber dysfunction predominates in Fabry neuropathy. J Clin Neurophysiol *19*, 575-586.

Ellgaard, L., and Helenius, A. (2003). Quality control in the endoplasmic reticulum. Nat Rev Mol Cell Biol *4*, 181-191.

Engel, A.G., Gomez, M.R., Seybold, M.E., and Lambert, E.H. (1973). The spectrum and diagnosis of acid maltase deficiency. Neurology *23*, 95-106.

Eskelinen, E.L., Tanaka, Y., and Saftig, P. (2003). At the acidic edge: emerging functions for lysosomal membrane proteins. Trends Cell Biol *13*, 137-145.

References

Federico, A., Capece, G., Cecio, A., D'Auria, N., Di Iorio, G., Ronsisvalle, L., and Di Natale, P. (1981). Sanfilippo B syndrome (MPS III B): case report with analysis of CSF mucopolysaccharides and conjunctival biopsy. J Neurol *225*, 77-83.

Ferri, K.F., and Kroemer, G. (2001). Organelle-specific initiation of cell death pathways. Nat Cell Biol *3*, E255-263.

Fogher, C. (2003). Expression of lysosomal enzymes in plant seeds.

Frenkel, D., and Smit, B. (2002). Understanding molecular simulation : from algorithms to applications, 2nd edn (San Diego ; London, Academic).

Furbish, F.S., Steer, C.J., Krett, N.L., and Barranger, J.A. (1981). Uptake and distribution of placental glucocerebrosidase in rat hepatic cells and effects of sequential deglycosylation. Biochim Biophys Acta *673*, 425-434.

Futerman, A.H., and van Meer, G. (2004). The cell biology of lysosomal storage disorders. Nat Rev Mol Cell Biol *5*, 554-565.

Germain, D.P. (2010). Fabry disease. Orphanet J Rare Dis *5*, 30.

Germain, D.P., Benistan, K., and Angelova, L. (2010). X-linked inheritance and its implication in the diagnosis and management of female patients in Fabry disease. Rev Med Interne *31 Suppl 2*, S209-213.

Ghosh, P., Dahms, N.M., and Kornfeld, S. (2003). Mannose 6-phosphate receptors: new twists in the tale. Nat Rev Mol Cell Biol *4*, 202-212.

Giugliani, R., Carvalho, C.G., Herber, S., and de Camargo Pinto, L.L. (2011). Recent Advances in Treatment Approaches of Mucopolysaccharidosis VI. Curr Pharm Biotechnol *12*, 956-962.

Gritti, A. (2011). Gene therapy for lysosomal storage disorders. Expert Opin Biol Ther *11*, 1153-1167.

Hancock, M.K., Haskins, D.J., Sun, G., and Dahms, N.M. (2002). Identification of residues essential for carbohydrate recognition by the insulin-like growth factor II/mannose 6-phosphate receptor. J Biol Chem *277*, 11255-11264.

Hannun, Y.A., and Obeid, L.M. (2002). The Ceramide-centric universe of lipid-mediated cell regulation: stress encounters of the lipid kind. J Biol Chem *277*, 25847-25850.

Hers, H.G. (1963). Alpha-Glucosidase Deficiency in Generalized Glycogen-Storage Disease (Pompes Disease). Biochemical Journal *86*, 11-&.

Hesselink, R.P., Gorselink, M., Schaart, G., Wagenmakers, A.J., Kamphoven, J., Reuser, A.J., Van Der Vusse, G.J., and Drost, M.R. (2002). Impaired performance of skeletal muscle in alpha-glucosidase knockout mice. Muscle Nerve *25*, 873-883.

Hobbs, J.R., Hugh-Jones, K., Barrett, A.J., Byrom, N., Chambers, D., Henry, K., James, D.C., Lucas, C.F., Rogers, T.R., Benson, P.F.*, et al.* (1981). Reversal of clinical features of Hurler's disease and biochemical improvement after treatment by bone-marrow transplantation. Lancet *2*, 709-712.

Hollak, C.E., van Weely, S., van Oers, M.H., and Aerts, J.M. (1994). Marked elevation of plasma chitotriosidase activity. A novel hallmark of Gaucher disease. J Clin Invest *93*, 1288-1292.

Holtzman, E. (1989). Lysosomes (New York ; London, Plenum).

Honing, S., Sosa, M., Hille-Rehfeld, A., and von Figura, K. (1997). The 46-kDa mannose 6-phosphate receptor contains multiple binding sites for clathrin adaptors. J Biol Chem *272*, 19884-19890.

References

Hopkin, R.J., Bissler, J., Banikazemi, M., Clarke, L., Eng, C.M., Germain, D.P., Lemay, R., Tylki-Szymanska, A., and Wilcox, W.R. (2008). Characterization of Fabry disease in 352 pediatric patients in the Fabry Registry. Pediatr Res *64*, 550-555.

Hug, G., and Schubert, W.K. (1967). Lysosomes in type II glycogenosis. Changes during administration of extract from Aspergillus niger. J Cell Biol *35*, C1-6.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph *14*, 33-38, 27-38.

Hwu, W.L., Chien, Y.H., Lee, N.C., Chiang, S.C., Dobrovolny, R., Huang, A.C., Yeh, H.Y., Chao, M.C., Lin, S.J., Kitagawa, T.*, et al.* (2009). Newborn screening for Fabry disease in Taiwan reveals a high incidence of the later-onset GLA mutation c.936+919G>A (IVS4+919G>A). Hum Mutat *30*, 1397-1405.

Jack, R.M., Gordon, C., Scott, C.R., Kishnani, P.S., and Bali, D. (2006). The use of acarbose inhibition in the measurement of acid alpha-glucosidase activity in blood lymphocytes for the diagnosis of Pompe disease. Genet Med *8*, 307-312.

Johnson, K.F., and Kornfeld, S. (1992). A His-Leu-Leu sequence near the carboxyl terminus of the cytoplasmic domain of the cation-dependent mannose 6-phosphate receptor is necessary for the lysosomal enzyme sorting function. J Biol Chem *267*, 17110-17115.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys *79*, 926-935.

Journet, A., Chapel, A., Kieffer, S., Roux, F., and Garin, J. (2002). Proteomic analysis of human lysosomes: application to monocytic and breast cancer cells. Proteomics *2*, 1026-1040.

Karplus, M., and Petsko, G.A. (1990). Molecular dynamics simulations in biology. Nature *347*, 631-639.

Kint, J.A. (1970). The enzyme defect in Fabry's disease. Nature *227*, 1173.

Koeberl, D.D., and Kishnani, P.S. (2009). Immunomodulatory gene therapy in lysosomal storage disorders. Curr Gene Ther *9*, 503-510.

Kroos, M., Pomponio, R.J., van Vliet, L., Palmer, R.E., Phipps, M., Van der Helm, R., Halley, D., and Reuser, A. (2008). Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating. Hum Mutat *29*, E13-26.

Kroos, M.A., Pomponio, R.J., Hagemans, M.L., Keulemans, J.L., Phipps, M., DeRiso, M., Palmer, R.E., Ausems, M.G., Van der Beek, N.A., Van Diggelen, O.P.*, et al.* (2007). Broad spectrum of Pompe disease in patients with the same c.-32-13T->G haplotype. Neurology *68*, 110-115.

Kukol, A. (2008). Molecular modeling of proteins. In Methods in molecular biology v 443 (Totowa, N.J., Humana Press), pp. 1 online resource (xi, 390 p.).

Lachmann, R.H. (2011). Enzyme replacement therapy for lysosomal storage diseases. Curr Opin Pediatr *23*, 588-593.

Lazzarino, D.A., and Gabel, C.A. (1989). Mannose processing is an important determinant in the assembly of phosphorylated high mannose-type oligosaccharides. J Biol Chem *264*, 5015-5023.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W.*, et al.* (2011). ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol *487*, 545-574.

References

Lee, J.H., Yu, W.H., Kumar, A., Lee, S., Mohan, P.S., Peterhoff, C.M., Wolfe, D.M., Martinez-Vicente, M., Massey, A.C., Sovak, G.*, et al.* (2010). Lysosomal proteolysis and autophagy require presenilin 1 and are disrupted by Alzheimer-related PS1 mutations. Cell *141*, 1146-1158.

Lee, K., Jin, X., Zhang, K., Copertino, L., Andrews, L., Baker-Malcolm, J., Geagan, L., Qiu, H., Seiger, K., Barngrover, D.*, et al.* (2003). A biochemical and pharmacological comparison of enzyme replacement therapies for the glycolipid storage disorder Fabry disease. Glycobiology *13*, 305-313.

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.

Li, C.M., Park, J.H., Simonaro, C.M., He, X., Gordon, R.E., Friedman, A.H., Ehleiter, D., Paris, F., Manova, K., Hepbildikler, S.*, et al.* (2002). Insertional mutagenesis of the mouse acid ceramidase gene leads to early embryonic lethality in homozygotes and progressive lipid storage disease in heterozygotes. Genomics *79*, 218-224.

Lindorff-Larsen, K., Piana, S., Dror, R.O., and Shaw, D.E. (2011). How fast-folding proteins fold. Science *334*, 517-520.

Lodish, H.F. (2008). Molecular cell biology, 6th edn (New York, W.H. Freeman).

MacDermot, K.D., Holmes, A., and Miners, A.H. (2001a). Anderson-Fabry disease: clinical manifestations and impact of disease in a cohort of 60 obligate carrier females. J Med Genet *38*, 769-775.

MacDermot, K.D., Holmes, A., and Miners, A.H. (2001b). Anderson-Fabry disease: clinical manifestations and impact of disease in a cohort of 98 hemizygous males. J Med Genet *38*, 750-760.

Mahalingam, K., Janani, S., Priya, S., Elango, E.M., and Sundari, R.M. (2004). Diagnosis of mucopolysaccharidoses: how to avoid false positives and false negatives. Indian J Pediatr *71*, 29-32.

Marks, D.L., and Pagano, R.E. (2002). Endocytosis and sorting of glycosphingolipids in sphingolipid storage disease. Trends Cell Biol *12*, 605-613.

Martiniuk, F., Mehler, M., Pellicer, A., Tzall, S., La Badie, G., Hobart, C., Ellenbogen, A., and Hirschhorn, R. (1986). Isolation of a cDNA for human acid alpha-glucosidase and detection of genetic heterogeneity for mRNA in three alpha-glucosidase-deficient patients. Proc Natl Acad Sci U S A *83*, 9641-9644.

Matsuoka, Y., Senda, Y., Hirayama, M., Matsui, T., and Takahashi, A. (1988). Late-onset acid maltase deficiency associated with intracranial aneurysm. J Neurol *235*, 371-373.

Mayes, J.S., Scheerer, J.B., Sifers, R.N., and Donaldson, M.L. (1981). Differential assay for lysosomal alpha-galactosidases in human tissues and its application to Fabry's disease. Clin Chim Acta *112*, 247-251.

Mehler, M., and DiMauro, S. (1977). Residual acid maltase activity in late-onset acid maltase deficiency. Neurology *27*, 178-184.

Meikle, P.J., Hopwood, J.J., Clague, A.E., and Carey, W.F. (1999). Prevalence of lysosomal storage disorders. JAMA *281*, 249-254.

Mikosch, P., and Hughes, D. (2010). An overview on bone manifestations in Gaucher disease. Wien Med Wochenschr *160*, 609-624.

Muenzer, J. (1986). Mucopolysaccharidoses. Adv Pediatr *33*, 269-302.

Nakamura, K., Hattori, K., and Endo, F. (2011). Newborn screening for lysosomal storage disorders. Am J Med Genet C Semin Med Genet *157*, 63-71.

References

Odani, T., Shimma, Y., Tanaka, A., and Jigami, Y. (1996). Cloning and analysis of the MNN4 gene required for phosphorylation of N-linked oligosaccharides in Saccharomyces cerevisiae. Glycobiology *6*, 805-810.

Okumiya, T., Kroos, M.A., Vliet, L.V., Takeuchi, H., Van der Ploeg, A.T., and Reuser, A.J. (2007). Chemical chaperones improve transport and enhance stability of mutant alpha-glucosidases in glycogen storage disease type II. Mol Genet Metab *90*, 49-57.

Olson, L.J., Zhang, J., Dahms, N.M., and Kim, J.J. (2002). Twists and turns of the cation-dependent mannose 6-phosphate receptor. Ligand-bound versus ligand-free receptor. J Biol Chem *277*, 10156-10161.

Parkinson-Lawrence, E.J., Shandala, T., Prodoehl, M., Plew, R., Borlace, G.N., and Brooks, D.A. (2010). Lysosomal storage disease: revealing lysosomal function and physiology. Physiology (Bethesda) *25*, 102-115.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. Journal of Computational Chemistry *26*, 1781-1802.

Platt, F.M., Neises, G.R., Dwek, R.A., and Butters, T.D. (1994). N-butyldeoxynojirimycin is a novel inhibitor of glycolipid biosynthesis. J Biol Chem *269*, 8362-8365.

Ramaswami, U. (2011). Update on role of agalsidase alfa in management of Fabry disease. Drug Des Devel Ther *5*, 155-173.

Reczek, D., Schwake, M., Schroder, J., Hughes, H., Blanz, J., Jin, X., Brondyk, W., Van Patten, S., Edmunds, T., and Saftig, P. (2007). LIMP-2 is a receptor for lysosomal mannose-6-phosphate-independent targeting of beta-glucocerebrosidase. Cell *131*, 770-783.

Robbins, S.L., Kumar, V., and Cotran, R.S. (2010). Robbins and Cotran pathologic basis of disease, 8th edn (Philadelphia, Pa., Saunders/Elsevier).

Roberts, D.L., Weix, D.J., Dahms, N.M., and Kim, J.J. (1998). Molecular basis of lysosomal enzyme recognition: three-dimensional structure of the cation-dependent mannose 6-phosphate receptor. Cell *93*, 639-648.

Rohrer, J., and Kornfeld, R. (2001). Lysosomal hydrolase mannose 6-phosphate uncovering enzyme resides in the trans-Golgi network. Mol Biol Cell *12*, 1623-1631.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc *5*, 725-738.

Samaj, J., Read, N.D., Volkmann, D., Menzel, D., and Baluska, F. (2005). The endocytic network in plants. Trends Cell Biol *15*, 425-433.

Schiffmann, R., Warnock, D.G., Banikazemi, M., Bultas, J., Linthorst, G.E., Packman, S., Sorensen, S.A., Wilcox, W.R., and Desnick, R.J. (2009). Fabry disease: progression of nephropathy, and prevalence of cardiac and cerebrovascular events before enzyme replacement therapy. Nephrol Dial Transplant *24*, 2102-2111.

Schultz, M.L., Tecedor, L., Chang, M., and Davidson, B.L. (2011). Clarifying lysosomal storage diseases. Trends Neurosci *34*, 401-410.

Shaaltiel, Y., Bartfeld, D., Hashmueli, S., Baum, G., Brill-Almon, E., Galili, G., Dym, O., Boldin-Adamsky, S.A., Silman, I., Sussman, J.L.*, et al.* (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.

Shaw, D.E., Deneroff, M.M., Dror, R.O., Kuskin, J.S., Larson, R.H., Salmon, J.K., Young, C., Batson, B., Bowers, K.J., Chao, J.C.*, et al.* (2008). Anton, a special-purpose machine for molecular dynamics simulation. Commun Acm *51*, 91-97.

Spada, M., Pagliardini, S., Yasuda, M., Tukel, T., Thiagarajan, G., Sakuraba, H., Ponzone, A., and Desnick, R.J. (2006). High incidence of later-onset fabry disease revealed by newborn screening. Am J Hum Genet *79*, 31-40.

Staretz-Chacham, O., Lang, T.C., LaMarca, M.E., Krasnewich, D., and Sidransky, E. (2009). Lysosomal storage disorders in the newborn. Pediatrics *123*, 1191-1207.

Talsma, M.D., Kroos, M.A., Visser, G., Kimpen, J.L., and Niezen, K.E. (2002). A rare presentation of childhood pompe disease: cardiac involvement provoked by Epstein-Barr virus infection. Pediatrics *109*, e65.

Tikkanen, R., Obermuller, S., Denzer, K., Pungitore, R., Geuze, H.J., von Figura, K., and Honing, S. (2000). The dileucine motif within the tail of MPR46 is required for sorting of the receptor in endosomes. Traffic *1*, 631-640.

Tomatsu, S., Gutierrez, M.A., Ishimaru, T., Pena, O.M., Montano, A.M., Maeda, H., Velez-Castrillon, S., Nishioka, T., Fachel, A.A., Cooper, A*., et al.* (2005a). Heparan sulfate levels in mucopolysaccharidoses and mucolipidoses. J Inherit Metab Dis *28*, 743-757.

Tomatsu, S., Okamura, K., Maeda, H., Taketani, T., Castrillon, S.V., Gutierrez, M.A., Nishioka, T., Fachel, A.A., Orii, K.O., Grubb, J.H*., et al.* (2005b). Keratan sulphate levels in mucopolysaccharidoses and mucolipidoses. J Inherit Metab Dis *28*, 187-202.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem *31*, 455-461.

Umapathysivam, K., Hopwood, J.J., and Meikle, P.J. (2001). Determination of acid alpha-glucosidase activity in blood spots as a diagnostic test for Pompe disease. Clin Chem *47*, 1378-1383.

Urbanelli, L., Magini, A., Polchi, A., Polidoro, M., and Emiliani, C. (2011). Recent developments in therapeutic approaches for lysosomal storage diseases. Recent Pat CNS Drug Discov *6*, 1-19.

Valenzano, K.J., Khanna, R., Powe, A.C., Boyd, R., Lee, G., Flanagan, J.J., and Benjamin, E.R. (2011). Identification and characterization of pharmacological chaperones to correct enzyme deficiencies in lysosomal storage disorders. Assay Drug Dev Technol *9*, 213-235.

van der Ploeg, A.T., and Reuser, A.J. (2008). Pompe's disease. Lancet *372*, 1342-1353.

VanHove, J.L.K., Yang, H.W., Wu, J.Y., Brady, R.O., and Chen, Y.T. (1996). High-level production of recombinant human lysosomal acid alpha-glucosidase in Chinese hamster ovary cells which targets to heart muscle and corrects glycogen accumulation in fibroblasts from patients with Pompe disease. P Natl Acad Sci USA *93*, 65-70.

Vedder, A.C., Linthorst, G.E., van Breemen, M.J., Groener, J.E., Bemelman, F.J., Strijland, A., Mannens, M.M., Aerts, J.M., and Hollak, C.E. (2007). The Dutch Fabry cohort: diversity of clinical manifestations and Gb3 levels. J Inherit Metab Dis *30*, 68-78.

Vedder, A.C., Strijland, A., vd Bergh Weerman, M.A., Florquin, S., Aerts, J.M., and Hollak, C.E. (2006). Manifestations of Fabry disease in placental tissue. J Inherit Metab Dis *29*, 106-111.

Vervecken, W., Kaigorodov, V., Callewaert, N., Geysens, S., De Vusser, K., and Contreras, R. (2004). In vivo synthesis of mammalian-like, hybrid-type N-glycans in Pichia pastoris. Appl Environ Microbiol *70*, 2639-2646.

Vijay, S., and Wraith, J.E. (2005). Clinical presentation and follow-up of patients with the attenuated phenotype of mucopolysaccharidosis type I. Acta Paediatr *94*, 872-877.

Weinreb, N.J., Charrow, J., Andersson, H.C., Kaplan, P., Kolodny, E.H., Mistry, P., Pastores, G., Rosenbloom, B.E., Scott, C.R., Wappner, R.S*., et al.* (2002). Effectiveness of enzyme replacement therapy in 1028 patients with type 1 Gaucher disease after 2 to 5 years of treatment: a report from the Gaucher Registry. Am J Med *113*, 112-119.

References

Winkel, L.P., Van den Hout, J.M., Kamphoven, J.H., Disseldorp, J.A., Remmerswaal, M., Arts, W.F., Loonen, M.C., Vulto, A.G., Van Doorn, P.A., De Jong, G.*, et al.* (2004). Enzyme replacement therapy in late-onset Pompe's disease: a three-year follow-up. Ann Neurol *55*, 495-502.

Wokke, J.H., Ausems, M.G., van den Boogaard, M.J., Ippel, E.F., van Diggelene, O., Kroos, M.A., Boer, M., Jennekens, F.G., Reuser, A.J., and Ploos van Amstel, H.K. (1995). Genotype-phenotype correlation in adult-onset acid maltase deficiency. Ann Neurol *38*, 450-454.

Wraith, J.E. (2006). Limitations of enzyme replacement therapy: current and future. J Inherit Metab Dis *29*, 442-447.

Yu, Z., Sawkar, A.R., Whalen, L.J., Wong, C.H., and Kelly, J.W. (2007). Isofagomine- and 2,5-anhydro-2,5-imino-D-glucitol-based glucocerebrosidase pharmacological chaperones for Gaucher disease intervention. J Med Chem *50*, 94-100.

Yurimoto, H., Oku, M., and Sakai, Y. (2011). Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. Int J Microbiol *2011*, 101298.

Zhu, Y., Jiang, J.L., Gumlaw, N.K., Zhang, J., Bercury, S.D., Ziegler, R.J., Lee, K., Kudo, M., Canfield, W.M., Edmunds, T.*, et al.* (2009). Glycoengineered acid alpha-glucosidase with improved efficacy at correcting the metabolic aberrations and motor function deficits in a mouse model of Pompe disease. Mol Ther *17*, 954-963.

Zhu, Y., Li, X., Kyazike, J., Zhou, Q., Thurberg, B.L., Raben, N., Mattaliano, R.J., and Cheng, S.H. (2004). Conjugation of mannose 6-phosphate-containing oligosaccharides to acid alpha-glucosidase improves the clearance of glycogen in pompe mice. J Biol Chem *279*, 50336-50341.

Zhu, Y., Suits, M.D., Thompson, A.J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K.W., Xiang, Y., Siriwardena, A.*, et al.* (2010). Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont. Nat Chem Biol *6*, 125-132.

Zimran, A., Altarescu, G., Philips, M., Attias, D., Jmoudiak, M., Deeb, M., Wang, N., Bhirangi, K., Cohn, G.M., and Elstein, D. (2010). Phase 1/2 and extension study of velaglucerase alfa replacement therapy in adults with type 1 Gaucher disease: 48-month experience. Blood *115*, 4651-4656.

Protocols

1 Media

Luria Bertani (LB)

0,5% NaCl
1% Bacto™ Tryptone
0,5% Bacto™ Yeast Extract

Luria Bertani (LB)-agar

0,5% NaCl
1% Bacto™ Tryptone
0,5% Bacto™ Yeast Extract
1,5% Bacto™ Agar

2 PCR

Phusion polymerase PCR

5µl 10X Buffer
1µl DNA (50ng)
1µl Forward Primer
1µl Reverse Primer
1µl dNTP mix (10mM) (Promega)
1µl Phusion polymerase (New England Biolabs)
40µl MQ

Extension on 72°C and 1min/kb

Colony PCR

2µl 10X Buffer
1µl Forward Primer
1µl Reverse Primer
0.4µl dNTP mix (10mM) (Promega)
0.6µl MgCl2
0.2µl Taq Polymerase (Invitrogen)
14.8µl MQ

Add tip with culture

Extension on 72°C and 1min/kb

3 Ligation reactions

T4 ligase

17µl DNA in MQ
2µl 10X T4 Buffer (Promega)
1µl T4 ligase

Room temperature overnight

4 DpnI reaction

1µl DpnI
50µl PCR reaction

37°C for 1 hour

5 Restriction Digest

500ng-1µg DNA
1,5µl Enzyme1
1,5µl Enzyme2
2µl Compatible Buffer
0,2µl BSA
Add MQ until 20µl

1 hour at 37°C

6 Transformation to *E. coli* MC1061

DNA on cells
10 min on ice
1,5 min on 42°C
2 min on ice
add 900µl LB medium
1 hour at 37°C

Maximum 200µl on LB-agar plates + Antibiotic

37°C Overnight

7 Nucleospin® Gel and PCR Clean-up (Machery & Nagel)

Add Buffer NTI (~400µl). If green (pH6-7) add more until it turns yellow (pH<6).
Dissolve agarose gel at 50°C (5-10 min)
Load sample on Nucleospin® column
Centrifuge 30s at 11000 rpm
Discard flow trough
Add 700µl Buffer NT3 to wash silica membrane
Centrifuge 30s at 11000 rpm
Discard flow trough
Centrifuge 30s at 11000 rpm to dry silica membrane
Place the Nucleospin® column into a new eppendorf
Elute DNA by adding 30 µl MQ

8 Periplasmatic fraction isolation

1-2ml of IPTG induced culture
Centrifuge 5min at 5000 rpm and 4°C
Collect supernatant (=medium)
Add 1,5ml buffer (30mM TRIS/HCl pH8, 20% sucrose, 1mM EDTA pH8) to the pellet
Centrifuge 10min at 12500 rpm and 4°C
Discard supernatant

Add 1,5ml cold 5mM$MgCl_2$
10min shaking in ice bath
Centrifuge 10min at 12500 rpm and 4°C
Collect supernatant (=periplasm)
Add 0,4ml cold 20 mM TrisHCl pH 7.5 and 4 µl 10 mg/ml lysozyme to the pellet
Incubate 15 min at 30°C
Sonicate 4 times 15 s
Centrifuge 10 min at 14000 rpm
Collect supernatant (=soluble fraction)
Add 100µl 20M TrisHCl pH 7.5 to the pellet
Centrifuge 10 min at 10000 rpm
Add 100µl 20M TrisHCl pH 7.5
Centrifuge 10 min at 10000 rpm
Add 450µl 1% SDS (=insoluble fraction)

9 QIAprep® spin Miniprep kit

Centrifuge 1-5ml at 10000 rpm for 3 min.
Discard supernatant and resuspend pellet in 250µl Buffer P1.
Add 250µl Buffer P2 and mix by inverting the tube 4-6 times until the solution turns blue.
Add 350µl Buffer N3 and mix by inverting the tube 4-6 times until the solution turns colorless.
Centrifuge 10 min at 13000 rpm.
Apply supernatant to the QIAprep® spin column by decanting.
Wash the QIAprep® spin column by adding 0,75ml Buffer PE.
Centrifuge 1 min and discard flow-trough.
Centrifuge for 1 min to remove residual wash buffer.
Transfer the QIAprep® spin column to the collection tube.
Elute the DNA by adding 50µl MQ to the centre of the QIAprep spin column, let stand for 1 min and centrifuge for 1 min.

10 Agarose gel

1,5% Agarose
Add TAE buffer

TAE Buffer

40mM Tris
20mM Acetic Acid
1mM EDTA
Add MQ

11 Acrylamide gels (12%)

Running Gel (for 4 of 1,5 mm)

13,2ml MQ
12ml 3X Tris Buffer
10,8ml 40X Acrylamide/Bisacrylamide

360µl 10% APS
36µl TEMED

Stacking Gel

9,4 ml MQ
3,75 ml 3X Tris Buffer
1,9 ml Acrylamide/Bisacrylamide
150µl 10% APS
15µl TEMED

BioRad Mini-Protean Tetra Electrophoresis System

Clean glasses with ethanol and put in casting system. Test for leakage.
Prepare running gel and stacking gel solutions. Add TEMED as last.
Pour running gel between glasses (+- 4/5). Add isopropanol to have a straight front.
If clotted, remove isopropanol and add stacking gel until full.
Put the comb between the glasses. Prevent air bubbles.

Remove glasses from casting system and put into Electron Assembly.
Add Tricine Cathode Buffer in the inner tank.
Remove combs and load samples.
Add Tricine Anode Buffer in the outer tank.
Run for 2h at 100-115V.

Sample preparation

Add 5X Laemli to samples.
Boil for 5-10 min on 96°C.

Tricine Gel Buffer (3X)

3M Tris
10mM SDS
Adjust to pH 8,45 with HCl
Add MQ

Tricine Cathode Buffer (inner tank)

100mM Tris
100mM Tricine
3,5mM SDS
Add MQ
pH should be around 8,25

Tricine Anode Buffer (outer tank)

100mM Tris
Adjust to pH 8,9 with HCl

Laemli (5x)

333mM SDS
25mM Tris pH6,8
1,5mM Bromphenol Blue

40,5mM DTT
44% glycerol
Add MQ

12 Coomassie

Coomassie Brilliant Blue

45% Methanol
9,15% Acetic Acid
1,5mM Coomassie Brilliant Blue
Add MQ

Destain

30% Methanol
7% Acetic Acid
Add MQ

Protocol

Add Coomassie Brilliant Blue on acrylamide gel
Remove Coomassie Brilliant Blue after 20 min
Add destain and wash frequently until desired colour
Fixate with 20% ethanol / 4% glycerol
Conserve acrylamide gel if needed with BioRad Gel Air dry system

13 Westernblot

Transfer Buffer

16,7mM Tris
200 mM Glycine
20% Methanol
Add MQ

PBS buffer (10X)

Solve Dulbecco's Phosphate Buffered Saline (DPBS) (Lonza) in 1l MQ

Protocol

Load maximum 2µl Precision Plus Protein Standards (Biorad)
Cut away stacking gel section
Put 3 Watmann sheets (7x10cm), 1 Nitrocellulose membrane (6x10cm), acrylamide gel and 3 Watmann papers on Semi-Dry transfer Unit (Hoeffer) after putting them into transfer buffer
Add some more transfer buffer and roll over to remove air bubbles
Blot 1h10 on 50A per gel
Add 2,5% milk in PBS for 30 min
Discard and add anti-His in 10ml PBS + Tween20 for 60min
Discard and wash 3 times 10 min with PBS + Tween20
Add anti-mouse (Dylight800, ThermoScientific) in 10ml PBS + Tween20 for 45min
Discard and wash 3 times 10 min with PBS + Tween20

Add PBS
Measure light intensity with Odyssey

Stripbuffer

62,5mM Tris/HCl pH 6,8
2% SDS
100mM β-Mercaptoethanol
Add MQ

Protocol Stripping

Put 30 s in microwave and 30 min at 50-56°C.
Wash 2 times 10 min with MQ, 2 times with PBS + Tween20 and 2 times with MQ.
Restart from adding antibody1.

14 Purification

Lysisbuffer

50mM NaH_2PO_4
300mM NaCl
1mM PMSF and complete inhibitor tabs
10μg/ml DNaseI
Set pH 8

A-buffer Ni

20mM NaH_2PO_4
300mM NaCl
0,1% CHAPS
Set pH 7,4

B-buffer Ni

20mM NaH_2PO_4
20mM NaCl
400mM imidazole
0,1% CHAPS
Set pH 7,4

Stripcolumn buffer

20mM NaH_2PO_4
500mM NaCl
50mM EDTA
Set pH 7,5

Grow *E. coli* strain BL21 + pICa2 in 40 ml of LB medium with antibiotics overnight at 28°C.
Start tree 350 ml cultures with OD 0.2 in LB medium.
Weight until OD of 0.5 is reached.
Add 1mM IPTG overnight at 20°C.
Prepare periplasm by adding 3ml/g lysisbuffer and stirring for 1u at 4°C.
Supply periplams to a 20 ml Ni-Sepharose 6 FF column (GE Healtcare)

Equilibrated with buffer A and 8% buffer B.
Collect protein with 50% and 100% buffer B.

15 Pierce BCA Protein Assay Kit

Making dilution serie from 2,000 μg/ml until 0 μg/ml of BSA.
Add 200 μl of WR reagents to 50 μl of standards (triplicate) and samples (duplicate).
Put for 30min at 37°C.
Measure absorbance at or near 562 nm on a plate reader.
Deduct the average of the blank standard replicates from all other individual standards and samples.

Prepare a standard curve for each BSA standard.
Use this curve to determine sample concentration.

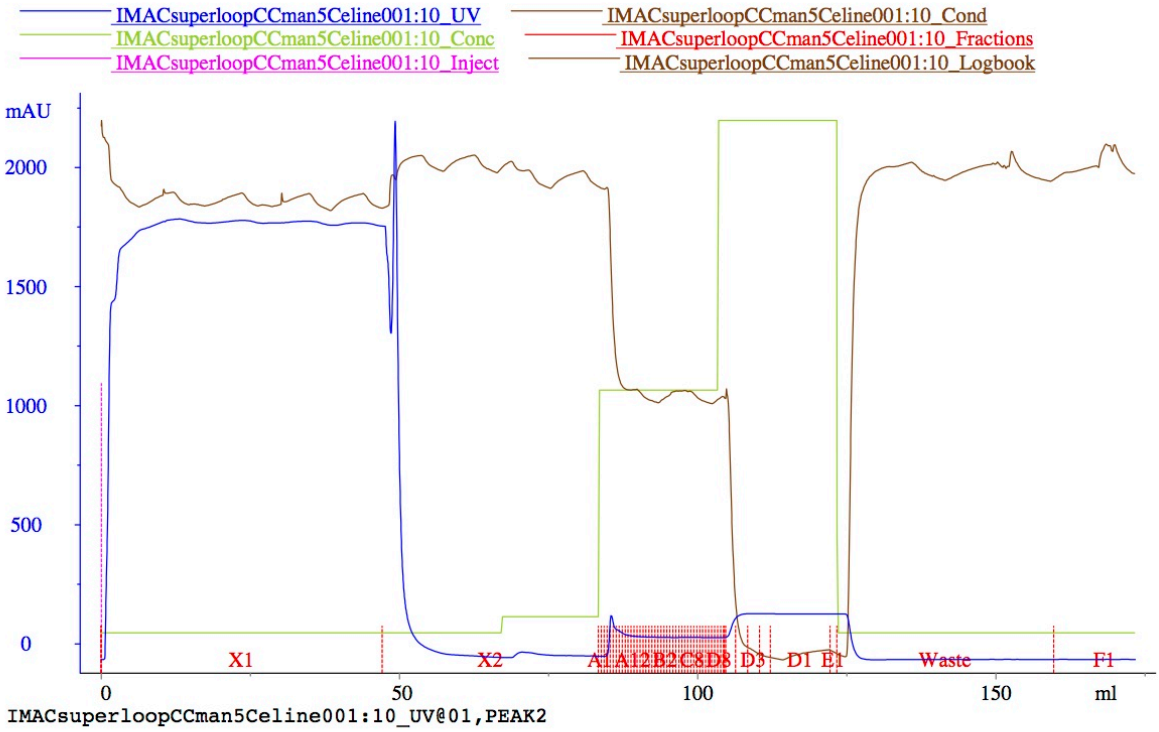
16 Activity test for GAA sugars

3.6μl periplasm
0.4μl GAA sugars
0.5μl 20mM CaCl2
0.5μl 100mM Hepes pH7.5.

This was incubated for 1h at 37°C.
The reaction was stopped by incubating it for 5 min at 100°C.

To analyze the activity 2μl was used for DSA-FACE.

Addendum

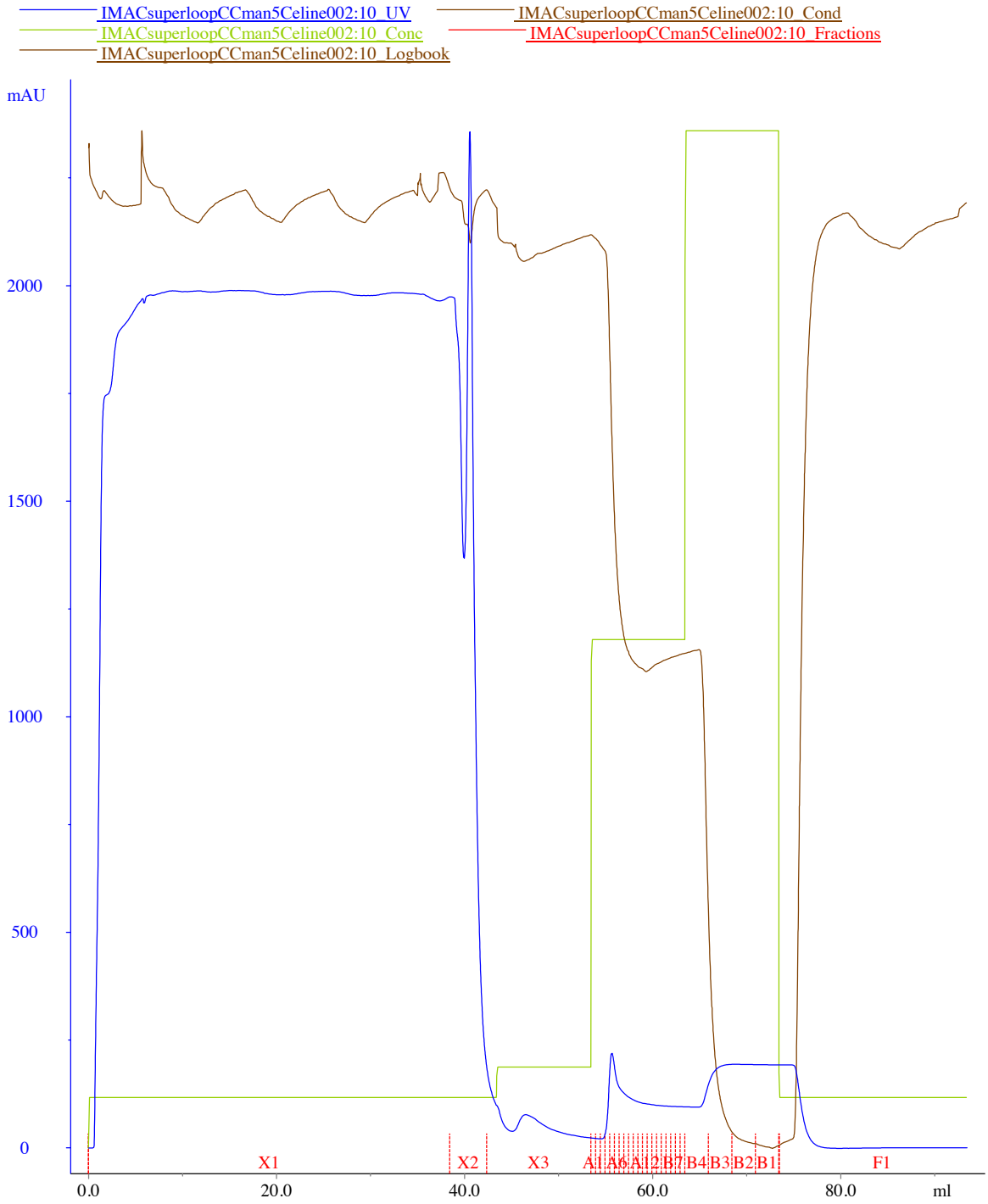


IMACsuperloopCCman5Celine001:10_UV@01,PEAK2

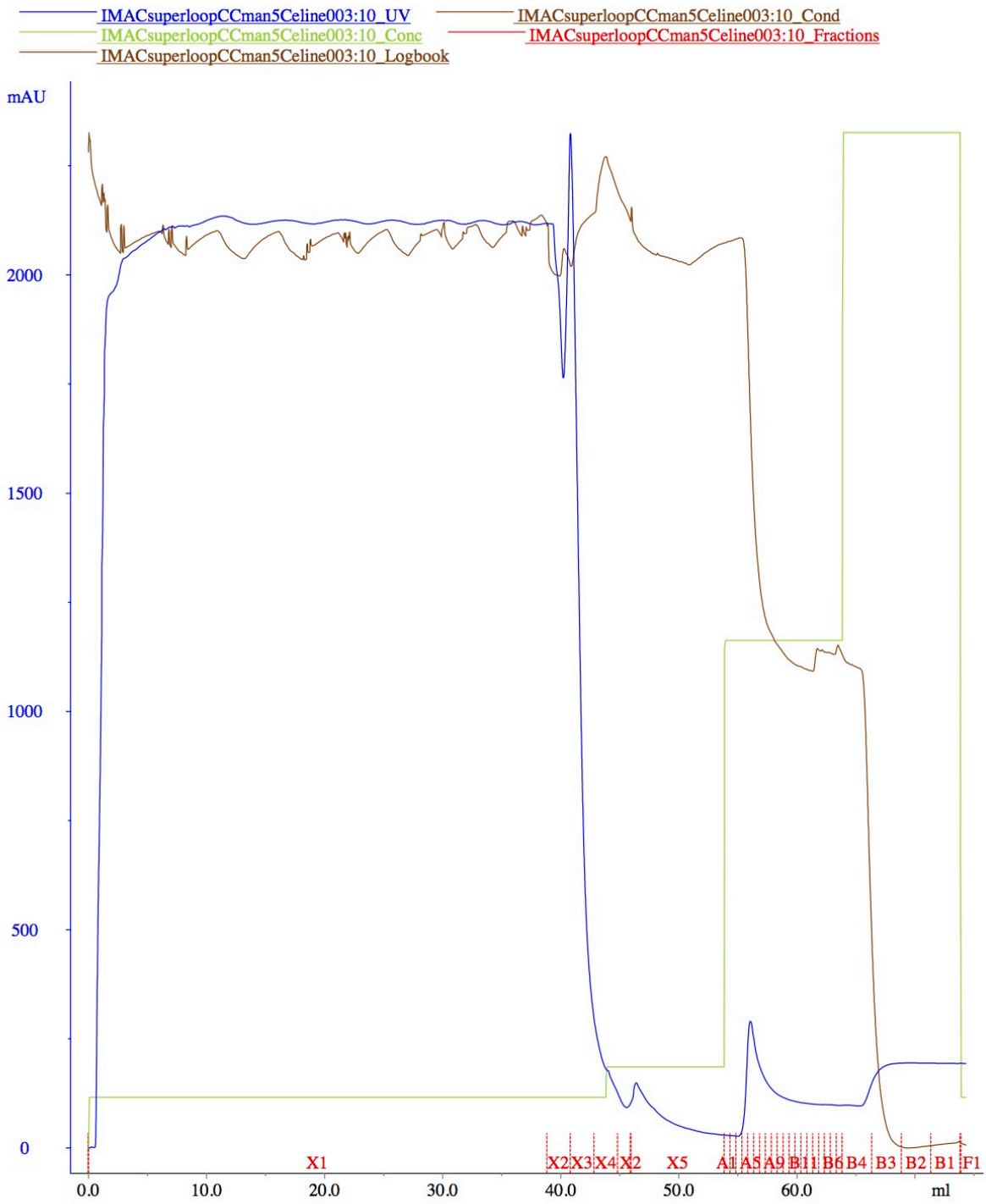
No	Ret	Area	Height

Total number of detected peaks = 47
Total area = 19839.0353 mAU*ml
Area in evaluated peaks = 0.0000 mAU*ml
Ratio peak area / total area = 0.000000
Total peak width = 0.00 ml
Column height = 2.50 cm
Column Vt = 0.86 ml
Calculated from : IMACsuperloopCCman5Celine001:10_UV
Baseline : IMACsuperloopCCman5Celine001:10_UV@01,BASEM
Peak rejection on:
Maximum number of peaks: 20

Addendum 1 Overview of the purification process of the Ala351 mutant



Addendum 2 Overview of the purification process of the Ala631 mutant



Addendum 3 Overview of the purification process of the long deletion mutant