EX0421 independent project/degree project in biology E, 30
Masters thesis, 2009
Master programme in biology
ISSN 1651-5196 Nr 105

Crystallization attempt of TGG5 and mutant studies of TGG4 myrosinase from thale cress *(Arabidopsis thaliana)*

http://hfspmetabolomics.life.uiuc.edu/

Carlos Sibrian

Department of Plant Biology and Forest Genetics
Swedish University of Agricultural Sciences, Uppsala

EX0421 Självständigt arbete i biologi E, 30 hp
Masterarbete, avancerat E

ISSN 1651-5196 Nr 105
2009

Kristalliserings försök med TGG5 och mutant studier av TGG4 myrosinase från backtrav *(Arabidopsis thaliana)*

Crystallization attempt of TGG5 and mutant studies of TGG4 myrosinase from thale cress *(Arabidopsis thaliana)*

Carlos Sibrian
Masterprogrammet i biologi

Supervisors: Prof. Johan Meijer, Ass. profs. Jerry Ståhlberg and Bejai Sarosh
Examiner: Docent Folke Sitbon

Institutionen för växtbiologi och skogsgenetik
Uppsala Biocenter, SLU
Box 7080
75007 Uppsala

SLU, Sveriges lantbruksuniversitet, Uppsala

Institutionen for biologisk grundutbildning
EBC, Norbyv. 14/ BMC Husarg. 3, Uppsala Universitet
Norbyv 14/ Box 592, 752 36, 751 24 Uppsla

Abstract

Myrosinase is a glucosidase normally found in plants of the order Capparales and that is involved in the breakdown of glucosinolates releasing toxic breakdown products. Together myrosinase and glucosinolates form a plant defence system that deters pests and herbivores by a barrage of toxic breakdown products. Many of the crops containing the myrosinase-glucosinolate system are valuable vegetable- or oil-crops making studies on myrosinase very important.
TGG5 and TGG4 are myrosinases belonging to *Arabidopsis thaliana.* The advantages with using myrosinase from a model organism is the extensive knowledge of the organism and the many tools and methods developed. Crystallization attempt were performed with recombinant TGG5 (containing a his-tag) where the myrosinase was expressed in *Pichia pastoris* and purified using immobilized metal-ion affinity chromatography (IMAC) and either cation-exchange chromatography or gel filtration chromatography. No crystals were observed but it seemed like the $LiSO_4$ conferred stability to the myrosinase enzyme maybe due to interaction with the sulphate binding site of the enzymes usually used for binding the sulphate containing substrate. TGG4 is another myrosinase isozyme found in *A. thaliana* and is very similar to TGG5 sharing 97% sequence identity. TGG4 mutants were designed using the structure of a crystallised myrosinase from *Sinapsis alba* as reference. Mutants were modified with respect to active site and residues predicted to be involved in substrate binding and co-factor (ascorbate) binding. Some of the findings were the reduced catalytic activity encountered when mutating the glutamine residue (replaced by an glutamate residue in O-β-glucosidases) and that converting the glutamine back to glutamate did not confer any additional catalytic activity.

TABLE OF CONTENTS

1. Introduction

1.1. The Myrosinase-glucosinolate system in plants

The myrosinase-glucosinolate system is a plant defence system found among plants of the order Capparales and in particular in the Brassicaceae family. Glucosinolates and myrosinases are spatially separated but come in contact with each other during tissue damage usually caused by herbivores. The myrosinase then hydrolyses the glucosinolates giving rise to different toxic breakdown products (Windsor et al. 2005). These toxic breakdown products are deterring against herbivores such as birds, mammals, molluscs and insects (Mithen 2001a) and might also be effective against microorganism (Aires et al. 2008). Glucosinolates have been known for over hundred years but the molecular tools and the use of the model plant *Arabidopsis thaliana* has allowed advances which would have been hard to achieve using only biochemical methods (Mithen 2001b).

1.2. Glucosinolates

Glucosinolates are secondary metabolites related to cyanogenic glucosides and are primarily found in the order Capparales. Glucosinolates are characteristics of plants belonging to *Brassicaceae* such as *Brassica napus,* a major oil crop and has therefore attracted much attention (Rask et al. 2000). The core structure in glucosinolates consists of a β-thio-glucose moiety, a sulfonated oxime and a side chain of variable structure (Mithen 2001a). This side-chain is usually derived from amino acids (methionine, phenylalanine, valine, thryptophan, tyrosine, isoleucine, leucine or alanine) and chain extended homologous of phenylalanine and methionine (Kliebenstein et al. 2005). This side-chain is responsible for the classification into aliphatic, aromatic and indole glucosinolates depending on whether they are derived from aliphatic amino acids (isoleucine, leucine, valine, alanine and methionine), aromatic amino acids (tyrosine, phenylalanine) or indole (tryptophan) (Wittstock & Halkier 2002).

A. thaliana contains more than thirty different glucosinolates where the majority is derived from tryptophan and chain extended homologous of methionine although the type of glucosinolate and the amount might differ among ecotypes (Kliebenstein et al. 2005). Glucosinolates are not biologically active until hydrolysed by myrosinase enzymes. The glucosinolate then forms a thiohydroximate-O-sulfate and a D-glucose. The thiohydroximate-O-sulfate can be converted into different toxic products either spontaneously (isothiocyanates) or with the help of other proteins (nitriles). Available metal ions and local pH are other factors important in the formation of different toxic compounds (fig. 1)(Bones & Rossiter 2006).

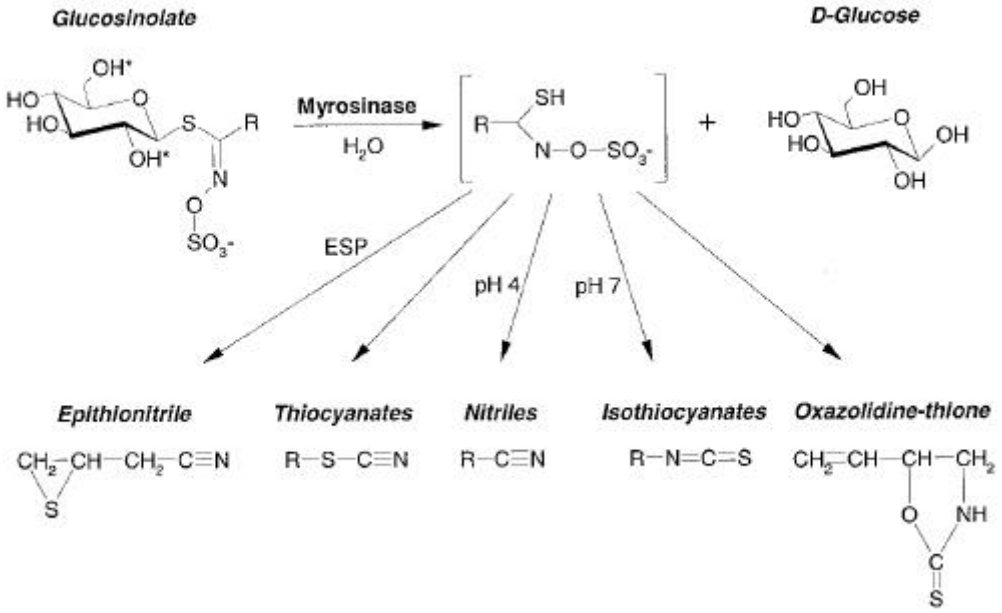


Fig. 1. Glucosinolates are hydrolysed by myrosinase producing a D-glucose and thiohydroximate-O-sulfate molecule. The thiohydroximate-O-sulfate can react further to produce different kind of toxic products.

1.3. Myrosinases

Myrosinase is a glucosidase and hydrolyses the thio-ether glycoside bond between carbohydrates and sulphur linked aglycone side chains present in glucosinolates. Glucosidases are grouped into families based on substrate specificity and occasionally on molecular mechanism. Myrosinases are grouped into the glycosyl hydrolase family 1, a family consisting mainly of O-β-glucosidases (Expasy).

The O-β-glucosidase members of the glycosyl hydrolase family 1 hydrolyse the O-ether glycosidic bond using two glutamate residues acting as catalytic proton donor/acid and catalytic nucleophile/base. Depending on the spatial position of the catalytic residues the hydrolysis occurs either with retention or inversion of the anomeric configuration (the attached group at the C-1 atom either changes position to the α- or β-form or retains it's position) (Henrissat 1997; Davies & Henrissat 1995). The β-glucosidases of glycosyl hydrolase family 1 uses retention (Expasy).

Myrosinases are the only known S-glucosidases and are capable of hydrolysing glycosidic bonds between glucosinolates and sulphur. Myrosinases differ from β-glucosidases in that the catalytic glutamate residue has been replaced with a glutamine residue; this means that myrosinases are missing a catalytic acid residue. Hydrolysis is still possible due to the use of water molecules positioned by the glutamine residue and the excellent leaving group abilities of the aglycon (Borderioux et al. 2005). This also means that myrosinases will not be able to cleave glucosinolates with poor aglycon leaving groups. The role of the catalytic acid/base (second glutamate) is believed to have been replaced by ascorbate that enters the empty active site left by the aglycon and activates the water molecule (Husebye et al. 2005).

High concentrations of ascorbate inhibit the myrosinase activity. This might be because of the fact that the aglycone of the glucosinolate and the ascorbate occupy almost the same site. Accordingly a high concentration of ascorbate makes it difficult for the glucosinolate to bind (Burmeister et al. 2000). Myrosinases are usually restricted to a special type of cells known as idioblasts or myrosin cells. These cells occur as scattered cells in seed, seedlings roots, stem, petal and leaf (Rask et al. 2000).

1.4. Three dimensional (3D) structure of myrosinases

The first myrosinase structure to ever be crystalized belonged to white mustard (*Sinapsis alba)* and it continuous to be the major reference in myrosinase structure analysis (fig. 2). Myrosinase is a dimer in solution and is stabilized by a Zn^{2+} ion. The myrosinase also has a $(\alpha\beta)_8$ barrel fold a common for members of family 1 glycosidases. The surface of the protein is heavily glycosylated and the protein increases its stability in several ways. These include three disulphide bridges, numerous salt bridges and hydrogen bonds between the main chain atoms and charged residues (Burmeister et al. 1997).

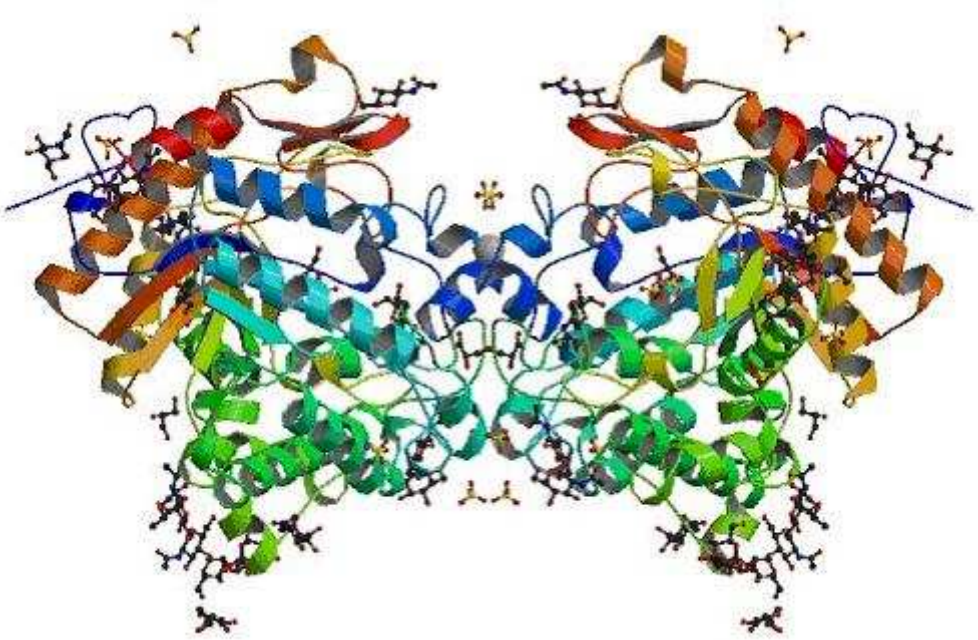


Fig. 2. The only available plant myrosinase structure is from *Sinapsis alba* (white mustard) and was elucidated by Burmeister et al (1997). Being the only available structure it serves as a major reference for work with myrosinases. The enzyme is a dimer with a stabilizing Zn^{2+} ion. The structure is annotated as 1MYR in the RCSB protein data bank.

The molecular masses of native myrosinases are estimated to be 135-150 kDa with the subunits often being around 59 kDa. Plant myrosinases are higly glycosylated proteins and it is estimated that the carbohydrate side chains make up between 10-20 % of the molecular mass of the myrosinase subunit. The carbohydrates are usually mannose, fucose and N-acetylglucosamine (Rask et al. 2000). Both the heavy glycosylation on the surface, the dimer formation and the many stabilizing bonds (disulphide bridges, hydrogen bonds etc) might be an adaptation to the dehydrated conditions occurring in mature seeds. But these conditions can differ between different isoforms expressed in other parts of the plant (Burmeister et al. 1997).

Another available myrosinase structure belongs to the cabbage aphid (*Brevicryne brassicae)* and has been found to have many similarities as well as differences. The most striking difference is that a glutamine has been replaced with glutamate, the acid-base residue most commonly found in O-glycosidases. This indicates that the aphid myrosinase has a different enzymatic reaction mechanism compared with plant myrosinases (Husebye et al. 2005).

1.5. Myrosinases in *Arabidopsis thaliana*

Arabidopsis has in contrast to other Brassicaceae species only one gene family coding for the six forms of myrosinases (TGG1-TGG6) found in this plant. TGG1 and TGG2 are both functional enzymes found in above ground tissues (Barth & Jander 2006). TGG3 is believed to be a pseudo gene and is found in stamens and petals (Zhang et al. 2002). TGG4 and TGG5 are only found in root tissue (according to transcript analysis) while TGG6 is found in the flowers particularly in the stamen (Barth & Jander 2006). In my project I expressed the TGG5 recombinant enzyme and different mutants of the TGG4 enzyme using the *Pichia pastoris* expression system from Invitrogen. The unprocessed TGG5 peptide has an estimated Mw of 57.469 kDa, contains 511 amino acids and a pI of 8.2. The unprocessed TGG4 peptide chain has a Mw of 57.542 kDa, contains 511 amino acids and has a pI of 8.2.

1.6. Aim of the work

The aim of this project was to increase the knowledge about myrosinase structure-function relationships. I made an attempt to crystallize the TGG5 myrosinase found in *A. thaliana* and to estimate its molecular mass using mass spectrometry. I also tried to perform mutant activity studies using TGG4 variants mutated at the active site and binding sites for ascorbate and the substrate (appendix 2). The general idea was that because TGG4 and TGG5 are almost identical any findings in either the crystallization or the mutant studies could be superimposed on the other (fig. 3). A sidetrack was made in order to identify a contaminant that due to its ubiquitous nature and nickel affinity created problems.

```
TGG4            MAIPKAHYSLAVLVLLFVVVSSSQKVCNPECKAKEPFHCDNTHAFNRTGFPRNFTFGAAT 60
TGG5            MAIPKAHYSLAVLVLLFVVVSSSQKVCNPECKAKEPFHCDNTHAFNRSGFPKNFTFGAAT 60
                ************************************************.***.********

TGG4            SAYQIEGAAHRALNGWDYFTHRYPEKVPDRSSGDLACDSYDLYKDDVKLLKRMNVQAYRL 120
TGG5            SAYQIEGAAHRALNGWDYFTHRYPEKVPDRSSADLACDSYDLYKDDVKLLKRMNVQAYRL 120
                ********************************.***************************

TGG4            SIAWSRVLPKGRLTGGVDENGITYYNNLINELKANGIEPYVTIFHWDVPQTLEDEYGGFL 180
TGG5            SIAWSRVLPKGRLTGGVDENGITYYNNLINELKANGIEPYVTIFHWDVPQTLEDEYGGFL 180
                ************************************************************

TGG4            STRIVEDYTNYAELLFQRFGDRVKFWITLNQPFSLATKGYGDGSYPPGRCTGCELGGDSG 240
TGG5            STRIVEDYTNYAELLFQRFGDRVKFWITLNQPLSLALKGYGNGSYPPGRCTGCELGGDSG 240
                *********************************.*** **** .******************

TGG4            VEPYTVAHNQLLAHAKTVSLYRKRYQKFQGGKIGTTLIGRWFAPLNEFSELDKAAAKRAF 300
TGG5            VEPYTVAHNQLLAHAKTVSLYRKRYQKFQGGKIGTTLIGRWFVPLNEFSELDKAAAKRAF 300
                ************************************************************

TGG4            DFFVGWFLDPLVYGKYPTIMREMVGDRLPEFTPEQSALVKGSLDFLGLNYYVTQYATDAP 360
TGG5            DFFVGWFLDPLVYGKYPTIMREMVGDRLPEFTPEESALVKGSLDFLGLNYYVSQYATDAP 360
                ************************************.*****************.*******

TGG4            PPTQLNAITDARVTLGFYRNGVPIGVVAPSFVYYPPGFRQILNYIKDNYKNPLTYITENG 420
TGG5            PPTQPNAITDARVTLGFYRNGSPIGVVASSFVYYPPGFRQILNYIKDNYKNPLTYITENG 420
                **** ***************** ******.******************************

TGG4            VADLDLGNVTLATALADNGRIQNHCSHLSCLKCAMKDGCNVAGYFAWSLMDNYEFGNGYT 480
TGG5            VADLDLGNVTLATALADNGRIQNHCSHLSCLKCAMKDGCNVAGYFAWSLMDNYEFGNGYT 480
                ************************************************************

TGG4            LRFGMNWVNFTNPADRKEKASGKWFSKFLAK 511
TGG5            LRFGMNWVNFTNPADRKEKASGKWFSKFLAK 511
```

Fig. 3. TGG4 and TGG5 are two very similar myrosinase isozymes both found in root tissue and containing 511 amino acids. When aligned using Clustal-W they show 97 % sequence identity. The yellow shaded letters represent the differences between TGG4 and TGG5.

2. Methods

2.1. Expression of recombinant protein using the *Pichia pastoris* expression kit

Pichia pastoris is a methylotropic yeast often used as host for expression of recombinant protein due to the fact that it can perform post-translational modifications and express a reasonable amount of proteins (Higgins 1995). One of the appeals with this expression system is that it has a strong inducible promoter activated in the presence of methanol. This promoter is coupled to the alcohol oxidase enzymes that are compartmentalized into the peroxisomes (Ozimek et al. 2005). When grown on methanol the peroxisomes may take up to 80 % of the total cell volume and the alcohol oxidase enzyme can make up to 30 % of the total protein in the yeast cell (Daly & Hears 2005). *P. pastoris* has two types of alcohol oxidases. AOX1 and AOX2, where AOX1 is responsible for the brunt of alcohol oxidase activity. This is the reason why the protein to be expressed is often coupled to the AOX1 promoter (Higgins & Cregg 1998). Proteins expressed by *P. pastoris* has the drawback that they are heavily glycosylated with N-linked oligosaccharide chains of mannose (8-14 mannose residues per side chain) but this is still much less then for the yeast *Saccharomyces cerevisiae* (50-150 mannose residues per side chain) (Nøhr et al. 2003).

The *P. pastoris* strain used was developed in a previous project using a GS115 strain and the pPIC3.5 expression vector containing the AOX1 promoter and has intracellular expression of the recombinant protein.

2.2. Inoculation and biomass accumulation of *P. pastoris*

The culture is often saturated with cells before the recombinant protein is expressed. Glycerol is a carbon source, which allows a faster growth and does not repress transcription as strongly as glucose, making it easier to induce expression when changing to methanol containing media (Inan & Meagher 2001).

One clone from the recombined *P. pastoris* GS115 strain (developed in a previous project) was inoculated in 5 ml BMGY at 28°C with shaking. When a satisfying turbidity was reached, 5 ml was transferred to a 500 ml Erlenmeyer flask containing 250 ml BMGY and grown at 28°C with shaking.

2.3. Inducing recombinant TGG5 expression in *P. pastoris*

To induce expression of the AOX1 promoter coupled protein the BMGY media must be removed and changed to a methanol-containing BMMY media. The methanol concentration of the media is set to 0.5 % because higher concentration might have toxic effects or decrease the growth of the yeast (Murray 1989; Guarna et al. 1997). New methanol is added on a daily basis to a concentration of 0.5 % to compensate for loss due to evaporation or consumption. Since alcohol oxidase has poor oxygen affinity it is important to assure good aeration of the culture (Higgins & Cregg 1998).

When the culture achieved an OD_{600} of 2-6, the culture was centrifuged at 3,000g for 5 min at room temperature using 50 ml Falcon tubes. The supernatant was removed and the pellet resuspended in BMMY until an OD_{600} of 1 was reached (~ 5×10^7 cells/ml). The BMMY medium contains 0.5 % methanol that induces expression of TGG5. New methanol is added each day to a final concentration of 0.5 %. The culture is grown in a Fernbach flask at 28°C with shaking. The flask is capped with cheesecloth and fresh methanol is added using a syringe.

2.4. Harvesting of cells

Since the recombinant protein has intracellular expression the yeast cells are collected and the supernatant removed.

Cells were harvested by centrifugation at 3,500g for 15 min using either 250 ml or 500 ml plastic bottles. The supernatant was poured out, the pellet resuspended in sterile water and then once again centrifuged at 3,500g for 15 min. The supernatant was removed and the pellet stored at -20°C until used.

2.5. Crude cell extract preparation

The cell pellet was weighed and then resuspended in a volume of start buffer roughly twice the weight of the cell pellet. This start buffer also contained several protease inhibitors such as 1 mM PMSF, 1 mM Benzamidine and 1 ml/100 ml of protease inhibitor cocktail (Roche). The resuspended cells were transferred to 2 ml tubes half filled with either glass beads (0.5 mm) or silica beads (0.5 mm). The cells in the tubes were disrupted using a bead beater. Shaking the tubes at maximum speed makes the beads rupture the tough yeast cells and homogenising the solution. The process is alternated between 45 sec with beat beating and 45 sec sample cooling in ice bath incubation and is repeated 5-7 times. The tubes were centrifuged at 12,000g for 10 min and the supernatant transferred into 34-SS tubes. The supernatant was centrifuged at 15,000g for 20 min at 4°C. This was done in order to remove cell debris that forms a pellet at the bottom of the tube. The supernatant was transferred to a new tube and the process repeated until no cell debris was visible. The supernatant was finally filtrated through a 0.20 µm filter and a 100 µl crude extract sample saved for activity assays while the rest was stored at 4°C.

2.6. Immobilized metal-ion affinity chromatography (IMAC) for purification of poly-histidine tagged recombinant proteins

IMAC is an affinity chromatography technique and separates proteins based on the reversible binding between amino acids and metal ions. Histidine, cysteine and tryptophan are residues with a natural affinity to metal ions, but one can also create affinity by attaching histidine residues to a protein (his-tag). The histidine residues are able to bind to bivalent cations such as Zn^{2+}, Co^{2+}, Ca^{2+}, Cu^{2+}, Fe^{2+} and Ni^{2+}. Matrixes with bivalent cation ligands are used to separate the recombinant protein from native proteins that usually do not have metal-affinity. Imidizole is the chemical of choice when reversing the binding to the ligand as it also has a metal-ion affinity (Charlton & Zachariou 2007). Purification of recombinant protein (6xHis) is performed using pre-packed columns, either Histrap Hp 5 ml column (GE healthcare) or Ni-NTA superflow cartridge 1 ml columns (Qiagen). Both columns are pre-packed with Ni^{2+} as ligand and are stripped and recharged after each use.

A pump (Gilson) with a flow rate of 1 ml/min was used. The column was equilibrated with 10 column volumes of start buffer (binding buffer) before loading the pre-treated yeast crude extract. The column was washed with 10 column volumes of washing buffer. Unspecific proteins were eluted with 5 column volumes elution buffer A. The flow-through was collected in one 50 ml tube. Proteins with affinity to the Ni^{2+} ligand were eluted with elution buffer B and collected in 1ml fractions. The whole procedure was performed either at room temperature or at 4°C. Fractions were stored at 4°C and later analysed for enzymatic activity and protein concentration. The columns were stripped and recharged by first washing with 5 column volumes water and then stripping the nickel ions with 10 column volumes stripping buffer. The column was washed once more with 10 column volumes water and then recharged with 2-2.5 ml 0.1M NiCl solution. The column was washed for a final time with 5 column volumes water and stored at 4°C.

2.7. Cationic exchange chromatography

Ion exchange chromatography relies on charge interactions between charged groups in the protein and the charges in the resin matrix. The charge of the protein is dependent on the pI where a pH under the pI will give it a positive charge and a pH over pI will give it a negative charge (Amersham Biosciences manual). In a cation exchange chromatography the charged amino acids consist of histidine (pKa 6.5), lysine (pKa 10) and arginine (pKa 12) while an anion exchange relies on the charged aspartic (pKa 4.4) and glutamic acid (pKa 4.4) (VHP high-perfomance ion-exchange columns manual). The start buffer is selected according to the isoelectric point with an optimal pH of 1-2 pH units from the isoelectric point. The start buffer should also have low ionic strength and to ensure good buffering capacity, the pKa should not be more then 0.5 pH units from the pH of the buffer used. The stationary matrix is made of anionic charged groups such as $–COO^-$, OPO_3^- and SO_3^- (Source 30Q and 30S instructions), elution is achieved by applying a stepwise increase in ionic strength of the elution buffer.

A 10 cm column packed with Source 30S (GE Healthcare) was used for cation exchange chromatography. The matrix consists of polystyrene/divinyl benzene and a S-resin (sulphate derivatives) as ion exchanger with the charged group $ROCH_2CHOHCH_2OCH_2CHOHCH_2SO_3^-$. The net charge of the protein has to be positive to bind to the resin. TGG5 has a pI of 8.2 therefore we use 20 mM sodium phosphate buffer with pH 6.75 (pKa 7) leaving the TGG5 protein positively charged. The ÄKTA-explorer system (GE Healthcare) was used for the cation exchange chromatography, this includes pre-equilibration of the column, running the protein sample and eluting the protein sample. The column was pre-equilibrated with buffer A and thereafter loaded with the sample and run with a flowrate of 1.5 ml/min. The bound protein was eluted with buffer B using a salt gradient going from 0-1 M (75 ml) and Na^+ as the displacer ion. A fraction-collector coupled to the ÄKTA-system was used to collect 2.2 ml fractions. UNICORN (GE Healthcare) software was used to detect the data in real-time.

2.8. Gel filtration chromatography

Gel filtration chromatography or size-exclusion chromatography is a method that separates proteins based on their molecular size. The gel media contain small pores that retains smaller molecules while larger molecules will be excluded from pores to varying extents (Bollag 1994). According to this scheme, the larger proteins are eluted first while smaller proteins are retarded and eluted later. The column matrix must be chosen such that it allows the best separation between the protein of interest and the contaminant. The protein of interest should be eluted first but not with the void volume, which contains molecules too large to enter any of the pores. The column itself should be long and narrow to allow good separation of the proteins (Stanton 2003). Due to the size-exclusion character of this chromatography method it is also a good way to change buffers as the proteins move faster through the column then the buffer molecules. It is also the most gentle separation method with high recovery.

Superdex 200 prep grade was the matrix used and has an optimal separation range between 10 kDa and 600 kDa. The column containing the pre-packed matrix was 60 cm long. The column was pre-equilibrated with 20 mM imidazole buffer (also containing 0.1M NaCl) with 0.5 ml/min flow rate. The same buffer and flow rate was used when eluting the protein. The ÄKTA-explorer system was used for the gel filtration chromatography. The protein sample was loaded into the injectionvalve of the system and was allowed to run for 20-40 min before starting the fraction-collector. Two ml fractions were collected for 5 hours or 150 ml. The column was run with 150 ml 20 % etanol before being stored at 4°C. The fractions were marked and stored at 4°C.

2.9. SDS-Polyacrylamide gel electrophoresis of proteins

SDS-PAGE is a widely used technique to separate proteins depending on their molecular mass. Proteins are mixed with the anionic detergent sodium dodecylsulfate (SDS) making the proteins form negatively charged complexes and thus allowing them to run through the gel. SDS binds approximately with a ratio of 1.4 gram SDS for each gram of protein (Smith 1994). The reducing agent (dithiothreitol or β-mercaptoethanol) was added to the samples to break inter- and intra-molecular disulphide bridges and thus making the protein lose its quaternary structure and dissociate into its subunits. This also allows the protein to form more uniform complexes with SDS (Blancher & Jones 2008).

Protein sample (2 µl) was mixed with 5.5 µl loading dye (200 µl loading buffer + 10 µl 1 M DDT). The mixture was heated for 5 min at 95°C, centrifuged briefly and then loaded on a 4-12 % "Ready-to-use" gel (C.B.S Scientific), which was mounted on a vertical running chamber. The running chamber was filled with running buffer (C.B.S Scientific). A voltage of 150 mV was run through the gel for 1 hour. The gel was then fixed in fixing solution for 5 min and incubated in Coomassie blue solution for at least 4 hours. The gel was destained with de-staining solution. A standard from Fermentas was used. When using the phastsystem (GE Healthcare), a somewhat different protocol was used. Protein sample (4 µl) was mixed with 1 µl loading buffer (5X buffer containing β-mercaptoethanol). The mixture was heated at 106°C for 5 min, centrifuged briefly and then loaded on a phastgel (8-25 %) and separated using the phastsystem workstation. The gel was stained with Coomassie blue solution, heated briefly in microwave for 10 seconds and incubated for 20-40 min. The gel was de-stained with de-staining solution (not the same as the one used for C.B.S Scientific gels). A LMW marker was prepared by mixing 2 µl LMW with 6 µl water, 2 µl loading buffer and then heated briefly at 106°C.

2.10. Western blot analysis

Western blotting involves the separation of protein on acrylamide gel and then transfer of the proteins from the gel to a membrane. The membrane is more stable and easier to work with then the acrylamide gel and also allows further analytical procedures (Walker 1986). In this case a primary anti-his antibody was used to bind recombinant proteins with his-tags, while a HRP secondary secondary antibody was used for detection.

One piece of PVDF membrane was cut slightly larger than the gel and soaked in 100% methanol for 10 min and then soaked in blotting buffer. The protein of interest was submitted to SDS-PAGE, the acquired gel was immersed in blotting buffer where after the PVDF membrane was placed onto the gel and made sure that no air bubbles were present. Whatman filter paper was cut into four rectangular pieces slightly larger then the gel. Two Whatman filter papers were dipped into blotting buffer until wet and the placed on top of the PVDF membrane. The gel, PVDF membrane and paper was turned to the other side where two more wet Whatman filter were placed on top. This paper-membrane-gel-paper sandwich was placed in a blot cell with the PVDF membrane side towards the positive electrode and submitted to electroblotting, which transfers the proteins from the gel to the membrane.
The blot (PVDF membrane) was then incubated in blocking buffer on a rocking table overnight at 4°C. The blot was briefly soaked in PBST to remove excess blocking buffer. The PBST was poured of and the blot incubated in 1.5 ml primary antibody diluted in PBST (1:3000). The blot was incubated for 1 hour at room temperature. The primary antibody was removed and the blot washed thrice in PBST (the blot was submerged in PBST for 10 min). The blot was then incubated in HRP conjugated secondary antibody (diluted in PBST to 1:1000) for 1 hour at room temperature. The blot was washed thrice in PBST (10 min). Excess PBST was poured off and the blot incubated in a small volume of ECL reagent

(50:50 mix of the two solutions) for 5 min. Excess ECL reagent was removed and the blot placed in a thin plastic film and the protein detected using chemiluminescence.

2.11. MALDI-TOF Mass spectrometry

The principle of mass spectrometry is that the path of charged particles (ions) will vary on their way through an electric and magnetic field depending on their charge-to-mass ratio. A mass spectrometer consist of three modules an ion source, which ionises the molecules, an analyser which sort the ions by their masses and a detector which measures the charge induced or the current produced when the charged particle hits a surface (Roepstorff 2005). During this project the MALDI-TOF variant of mass spectrometry was used. MALDI stands for matrix-assisted laser desorption ionisation and requires that the particle of interest was mixed with an excess of matrix molecules. The matrix usually consists of organic aromatic acids that after irradiation by a laser are sublimated allowing the non-volatile protein molecules to transfer into gas-phase as singly charged particles (Yates 2001). TOF or time-of-flight is the analyser used and uses an electric field to accelerate the charged particles and measure the time they take to reach the detector (Van Den Boom & Ehrich 2007). Åke Engström at Uppsala proteomics platform, BMC performed the MALDI-TOF procedure and I received the results in the form of mass spectra.

2.12. Peptide mass fingerprinting

In a peptide mass fingerprinting the protein is treated with proteolytic enzymes such as trypsin or pepsin and the resulting peptide fragments are analysed with mass spectrometry. Trypsin is an especially useful protease as it produces peptides in the 0.8– 4 kDa size range (Yates 2001).
The mass spectrum acquired can provide a fingerprint of the enzyme and the peptide mass values can be compared to other mass values from databases (Cottrell & Sutton 1996). Åke Engström performed peptide mass fingerprinting on the 38 kDa contaminant using MALDI-TOF and correlated the data to the MASCOT database.

Åke Engström also performed de novo sequencing on several peptide fragments by using the mass spectrometer and knowing the approximate masses of each amino acid (Gevaert & Vandekerckhove 2005).

2.13. De-salting

The protein was de-salted using a PD-10 desalting column (GE Healthcare) containing a Sephadex G-25 medium matrix (GE Healthcare) capable of separating high molecular weight substances (Mw > 5000 Da) from low molecular weight substances (Mw < 1000 Da). The column was equilibrated with 15 ml of the desired buffer and then 3 ml of the sample was applied and allowed to run through. Four ml of the desired buffer was added and the flow-through collected. Because low molecular mass ions run more slowly through the matrix, only the protein will be collected and suspended in the buffer of choice. The protein sample was stored at 4°C. The column was washed with 15 ml 0.02 % Na-Azide and then stored with 2-5 ml 0.02 % Na-Azide.

2.14. Concentration of protein

The protein was concentrated by membrane ultrafiltration. This procedure usually uses a single tube containing two compartments and a semi-permeable membrane with a molecular cut-off between the compartments. The solvent and low-molecular molecules are forced through the membrane by centrifugal forces while the protein remains in the upper compartment and thus gets more

concentrated. By adding a different solvent and continue the centrifugation until all ions from the previous solvent has passed the membrane, it is possible to change the buffer if needed. Vivaspin-20 (GE Healthcare) columns were used for concentration of the protein except when the volume was under 1 ml when a vivaspin-500 (GE Healthcare) column was used. The columns were washed with de-ionized water before use.

2.15. Enzymatic de-glycosylation

When expressed in *P. pastoris* the protein gets glycosylated with N-linked oligosacharide chains of mannose residues. The N-linked glycans are attached to asparagines through an amide bond (Harvey 2003). Treatment with α-mannosidase can cleave alpha-(1,2), (1,3) and (1,6) terminal residues and at a slower rate the 1,4 – linked terminal mannose residue (α-mannosidase product information, SIGMA). Pretreating with α-mannosidase might also make it easier for Endoglycosidase-H to cleave between the two innermost GlcAc residues of the N-linked glycan leaving one GlcAc residue attached to the asparagines (EndoH product information, SIGMA). Removing glycosidase residues might make it easier to crystallize the protein and to get Mass spectra using MALDI-TOF but there is also the possibility that the protein becomes unstable and unfolds.

The protein was diluted in 50 mM NaAcetate pH.5, and concentrated to a concentration of choice. ZnAcetate is added to a final concentration of 20 mM and to this 50 µg α-mannosidase was added. This mixture was sterile filtered through a 0.2 µm filter and then incubated at room temperature. Aliquots were taken before and after addition of α-mannosidase and stored at 4°C.
The TGG5 protein was left to crystallize together with 1:100 de-glycosylation protein either with α-mannosidase or with both α-mannosidase and endo-H. In the mixture containing α-mannosidase the final concentrations were 1 mM ZnAcetate, 2 mg/ml of TGG5 and 0.02 mg/ml of α-mannosidase. A similar mixture was prepared with the only difference that it contained 0.02 mg/ml endo-H. The mixture (2.4 µl) was set up together with 0.8 µl crystalisation buffer giving 3.2 µl drops. These drops were incubated at 20°C. The drops contained variable buffers but the pH was usually between pH 4.2 and 5.

2.16. Determination of protein concentration using A_{280}

Protein concentration can be estimated using the fact that aromatic amino acids, tryptophan and tyrosine and also disulfide bonded cysteine residues exhibit UV-light absorption at wavelengths of 280 nm (Simonian & Smith 2006). Out of these residues, tryptophan has the highest absorbance as it absorbs UV-light almost five times greater then tyrosine and 50 times more then disulfide bonded cysteine residues. This means that the protein measured must have aromatic amino acids especially tryptophan, because every protein has a different composition of these three UV-absorbing residues a coefficient must be used to be able to get comparable values. This coefficient is known as the molar absorption coefficient (ε) and is the sum of the three major UV-absorbing amino acid residues at 280nm (Simonian 2002).

$\varepsilon = ((n*\text{tryptophan})*5500)+((n*\text{tyrosine})+1490)+((n*\text{Cysteine})*125)$

n is the number of each residue while 5500, 1490 and 125 are the values for amino acid molar absorptivities at 280nm for the corresponding amino acids.

The absorbance of the protein was measured using a spectrophotometer (SHIMADZU UV 1800) using a spectrum between 340-240nm. The spectrophotometer was blanked at the spectrum of choice, a 1 cm width quartz cuvette was used and the volume was always 1 ml. The absorbance of the sample was

measured and peaks at 280nm were annotated. When large volumes can not be used for measuring protein concentration, a nanodrop spectrophotometer was used instead. The procedure is similar but instead of a spectrum only the 280nm wavelength was used and only 2 µl of the sample was needed to measure the protein concentration. The following equation can be used to calculate a protein concentration using the molar absorption coefficient, absorbance and molecular weight of the protein:

Concentration (mg/ml) = ($\mathbf{A}_{280}$*molecular weight) / (ε_{280} * 1 cm)

2.17. Determination of protein concentration using the Bradford assay

The Bradford assay is dye-binding technique and is a colorimetric assay as the colour of the test sample changes with the protein concentration. The dye Coomassie brilliant blue G-250 shifts its absorbance maximum from 465nm to 595nm upon binding to basic amino acids (arginine) and aromatic amino acids in acidic solution. Binding to protein changes the colour of the dye from reddish-brown to blue (the more protein the darker the blue colour) (Kruger 1996).

A microtiter plate protocol was used where the sample was prepared in a 96 well microplate and analysed at 595nm with a microplate reader (FLOUstar OMEGA). Bovine serum albumin (Fermentas) was used to set up a standard curve with known protein concentration to which the absorbance of the sample can be compared. The concentration range of the standards goes from 0.05 mg/ml through 01, 0.2, 0.3, 0.5, 0.8 and 1 mg/ml.

2.18. Enzyme assay of alcohol oxidase

P. pastoris has two genes coding for alcohol oxidase, AOX1 and AOX2. Out of these two, the AOX1 gene codes for the protein responsible for almost all alcohol oxidase activity (Gregg et al. 1989). AOX1 functions as a octamer of 8 identical subunits each of which binds a FAD cofactor (Ozimek et al. 2005). Each subunit has 663 amino acids (www.uniprot.org) and a molecular mass of ~ 72 kDa. When assembled in the functional octamer, the molecular mass is ~ 675 kDa (Ellis et al. 1985). When oxidizing methanol, Alcohol oxidase generates formaldehyde and hydrogen peroxide. Due to the toxicity of hydrogen peroxide, the process is carried out in peroxisomes to which the enzyme is confined (Gregg et al. 1989). The amount of H_2O_2 can be used as an indication of alcohol oxidase activity by the use of peroxidase enzyme and ABTS (2,2'-Azino-bis(3-ethylbenzthiazoline-6-sulfonic acid). The ABTS molecule functions as donor giving two water molecules and oxidized ABTS now a green soluble end-product.

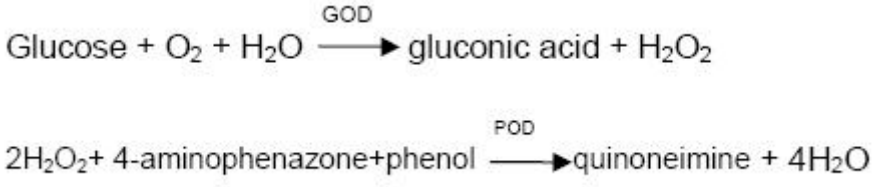
$$\text{Methanol} + O_2 \xrightarrow{\text{Alcohol Oxidase}} \text{Formaldehyde} + H_2O_2$$

$$H_2O_2 + \text{ABTS} \xrightarrow{\text{POD}} 2\,H_2O + \text{Oxidized ABTS}$$

ABTS (280 µl, 2 mM) was mixed with 1 µl peroxidase (250 unit/ml) and incubated at 25°C under shaking (Eppendorf thermomixer 300 rpm) for 5 minutes. Sample (2-10 µl) and 1 µl hydrogen peroxide (0.003%) was added to the mixture and again incubated at 25°C for 5 min. After adding 10 µl methanol (1%) incubation at 25°C shaking continued for 10 minutes. Absorbance was read at 405nm using a microplate reader (FLOUstar OMEGA). A negative control with 10 µl 100 mM potassium phosphate buffer pH 7.5 instead of protein sample was prepared. Crude extract of proteins from *P. pastoris* was used as positive control as it contains a large amount of AOX1.

2.19. Enzyme assay of myrosinase activity using sinigrin as substrate

Sinigrin is a glucosinolate and when hydrolysed by myrosinases the sulphur bond is cleaved giving secondary metabolites and glucose. The secondary metabolites vary depending on the glucosinolate type and environment and are also unstable, all bad characteristics when doing assays to determine enzyme activity (Kleinwächter & Selmar 2004). The enzyme activity assay was instead based on glucose and a GOD-POD method. The released glucose is oxidized by glucose oxidase (GOD) into gluconic acid and hydrogen peroxide. Hydrogen peroxide reacts under catalysis of peroxidase (POD) with 4-aminophenazone and phenol to form a red/violet quinineimine dye (Glucose GOD/PAP manual, RANDOX). GOD, POD, 4-aminophenazone and phenol are all included in the glucose reagent from RANDOX.



The protein sample was mixed with 50 mM Na-citrate pH 4.5 to a final volume of 25 µl and incubated at 37°C for 5 min. The sample was then mixed with 62 µl sinigrin (12.5 mM) and pre-heated to 37°C. The mixture was vortexed briefly and incubated at 37°C for 30 min. The reaction was stopped by heating at 95°C for 5 min and allowed to cool. After cooling, 250 µl Glucose reagent (Randox) was added to each sample and incubated at 37°C for 15 min. A 200 µl aliquot from each sample was transferred to a 96 well plate and analysed at 550nm with a microplate reader (FLOUstar OMEGA). A standard curve was made using glucose from a glucose test kit (Randox) to calculate the amount of released glucose from the samples. Myrosinase activity can be calculated using the following formula:

Total activity in X units (mmol glucose/min) = average of A550*F*dilution factor/30 min
Specific activity in Y mmol glucose/min mg = total activity / total protein
F = gradient from standard curve.

2.20. Enzyme assay of myrosinase activity with addition of ascorbic acid

Ascorbic acid is a co-factor of myrosinase and functions as a nucleophile when involved in the hydrolysis reaction (Kleinwächter & Selmar 2004). Ascorbate has both a stimulating and inhibiting activity as high concentrations of ascorbic acid can inhibit the myrosinase (Burmeister et al. 2000). Ascorbate can have some negative effects on GOD-POD method assay as it reacts with hydrogen peroxide and inhibits the formation of quinineimine dye (Kleinwächter & Selmar 2004). Even if the values may not be absolute in the presence of ascorbate it can still give a good indication of differences between different myrosinases and mutants of myrosinases, as long as the ascorbic acid concentration is not too high.

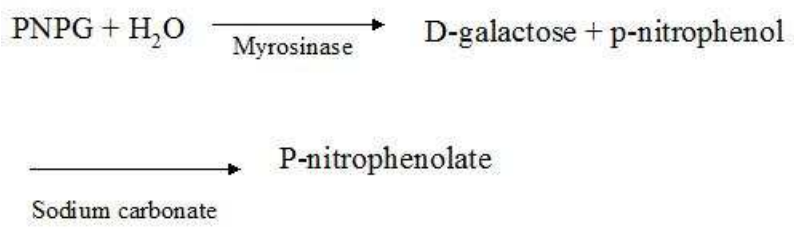
The protein sample was mixed with 50 mM Na-citrate pH 4.5 and incubated at 37°C for 5 min. The sample was mixed with ascorbic acid (pre-heated to 37°C) to a final volume of 25 µl. Sinigrin (62 µl, 12.5 mM pre-heated to 37°C) was added. The samples were mixed and incubated at 37°C for 30 min. The reaction was stopped by heating at 95°C for 5 min and allowed to cool. After cooling, 250 µl Glucose reagent (Randox) was added to each sample and incubated at 37°C for 15min. A 200 µl aliquot from each sample was transferred to a 96 well plate and analysed at 550 nm with a microplate

reader (FLOUstar OMEGA). Glucose standards from Randox were used to make a standard curve estimating the glucose released when myrosinase cleaves sinigrin. Ascorbic acid was freshly prepared before each assay and the concentration used ranged from 0.115 to 2.64 mM.

Total activity in X units (mmol glucose/min) = average of A550*F*dilution factor/30 min
Specific activity in Y mmol glucose/min mg = total activity / total protein
F = gradient from standard curve.

2.21. Enzyme assay of O-β-glucosidase activity of myrosinase

The protein sample was mixed with 50 mM Na-citrate, pH 4.5, to a final volume of 25 µl. To this mixture 25 µl, 5 mM p-nitrophenyl β-D-glucopyranoside (pNPG) was added and then incubated in a Eppendorf thermomixer (37°C, 350 rpm) for 30 min. The reaction was stopped by addition of 25 µl sodium carbonate and the samples transferred to a 96 well plate. The absorbance was analysed at 405 nm using a microplate reader (FLOUstar OMEGA). Free 4-nitrophenol was prepared and used to construct a standard curve to allow quantification.



2.22. Hanging drop crystalization

A hanging drop crystallization attempt was set up around crystallization values used earlier by Eva-Lena Andersson, which gave some protein crystal growth. The crystallization conditions used by Eva-Lena were 14-28% PEG1000, 0.2 M $LiSO_4$, 0.1 M $NaPO_4$/citrate buffer, pH 4.2. But I was adding de-glycosylation enzymes in some of the drops to remove mannosidase side-chains and could therefore not have phosphate in the buffer as it falls out in the presence of zinc. Therefore I substituted the $NaPO_4$ for sodium citrate buffer pH 4.2 and pH 5.0. Since there was no PEG1000 available I had to use PEG1500 instead. TGG5 concentrated to 2 mg/ml was used to set up two plates with 18 drops in each plate. The drops consisted of 2.4 µl protein+0.8 µl reservoir buffer. The reservoir contained 1 ml buffer. Six drops also contained 0.02 mg/ml α-mannosidase (1:100 of TGG5) and another six drops contained 0.02 mg/ml α-mannosidase and endoH (1:100). The plate was stored at 20°C.

2.23. Optimum solubility screen (OSS)

The OSS screen was used to test the protein compatibility at different variable conditions, in this case pH. Different 1 M buffers with pH ranging from 3 – 10 were used to test the proteins tendency to aggregate at different pH values. The reason 1 M buffers were used was to minimize any effect that the proteins original solvent may have on the screen. The difference between 20 mM imidazole buffer has less effect on 1 M buffer then on the recommended 0.1 M buffer. Hanging drop vapour diffusion was used to set up 23 drops. TGG5 protein concentrated to 2 mg/ml was used to set up drops consisting of 0.8 µl protein+0.8 µl reservoir buffer while the reservoir contained 50 µl 1 M buffer.

2.24. JCSGplus screen

A number of crystallization variables were tested using the JCSGplus screen and sitting drop vapour diffusion. The TGG5 protein sample had a concentration of 4 mg/ml and was mixed 1:100 α-mannosidase and 1:100 endo-H immediately before setting up the drops. Protein (0.8 µl) was added together with 0.8 µl reservoir buffer while 50 µl buffer was added to each reservoir.

3. Results

3.1. First TGG5 crystallization attempt

3.1.1. Expression of recombinant TGG5, crude extract preparation and first step in purification.

A previously prepared strain of *P. pastoris* was used for intracellular expression of recombinant his-tagged TGG5 under control of the AOX1 promoter. After cultivation of the cells to high density, expression was induced with methanol. Cells were harvested and disrupted for preparation of a crude protein extract that was fractionated on a Ni-NTA IMAC column.The fractions were tested for myrosinase activity and the protein concentration in the samples was estimated using the Bradford assay (Bio-Rad) (Table 1). Fractions B3-B11, the flow through and crude extract were analysed on a 4-12% polyacrylamide gel and stained with coomassie blue. The B5, B6 and B7 fractions gave two bands. One was between 60-70 kDa, which is presumed to be the size of the glycosylated TGG5, while the other band was a ~40 kDa contaminant. These three fractions were used for further purification.

Table1. Myrosinase fractions tested for enzyme activity and with estimated protein concentration

Elution B (1 ml fractions)	**Myrosinase activity (mmol /min)**	**Protein concentration (mg/ml)**
B3	1.7	0.18
B4	30.1	0.24
B5	70.3	0.48
B6	76	0.9
B7	39.3	0.62
B8	23.9	0.43
B9	10.4	0.25
B10	5.9	0.2
B11	2.6	0.13
Flowthrough	1.1	14.51
Crude extract	32.4	17.6

Two 50 ml tubes containing cell pellets prepared in the lab in previous experiments (not prepared by me) were used. The protein amount acquired from the culture was still a bit low and I wanted to increase the protein amount. The pellet was prepared and the crude extract ran through the Histrap 5 ml column in the same manners as before. The acquired fractions were then tested for myrosinase activity and the protein concentration estimated using the Bradford assay (table 2). Fractions B3-B10 were run on a gel together with the pooled flow-through. Two bands were observed, the TGG5 lying between 60-70 kDa while the contaminant lied between 30-40 kDa (fig. 4).

Table2. Myrosinase fractions tested for enzyme activity and with estimated protein concentration

Elution B (1 ml fractions)	**Myrosinase activity (mmol/min)**	**Protein concentration (mg/ml)**
B3	14.6	-0.24
B4	63.7	0.9
B5	65.8	2
B6	67.7	1.5
B7	69.2	1.1
B8	62.7	1.26
B9	44.5	0.66
B10	27.5	3.8E-2

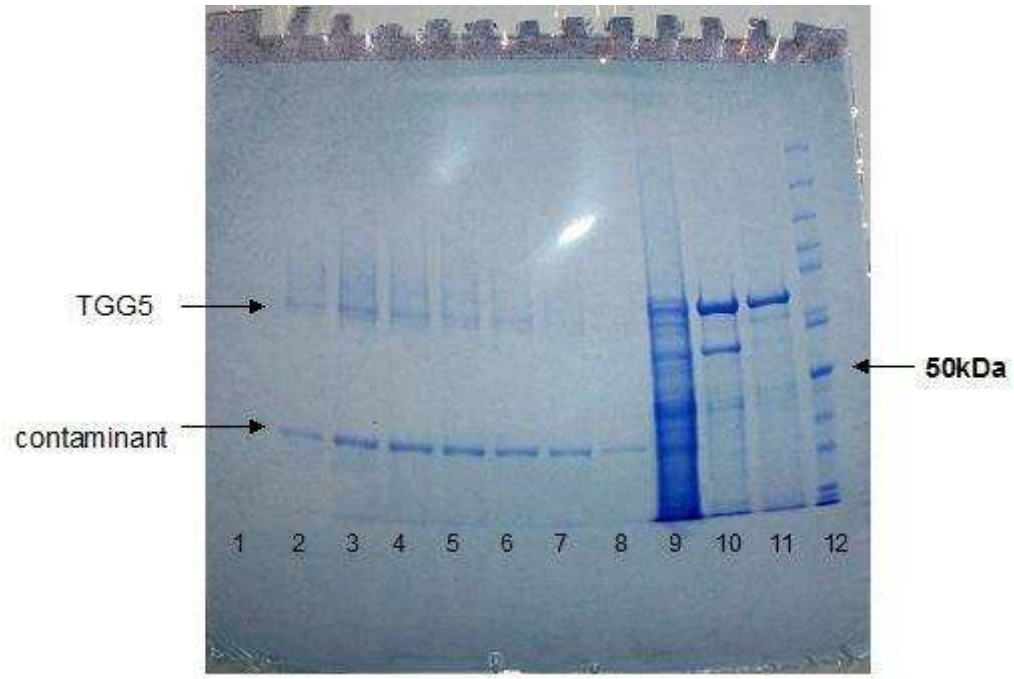


Fig. 4. Lanes 1 to 8 contain the proteins from fraction B3, B4, B5, B6, B7, B8, B9 and B10 respectively. Lane 9 is the flowthrough, lane 10 and 11 are not related to the expeiment while lane 12 contains the marker (Fermentas, pageruler^{TM} unstained protein ladder).

3.1.2. Second purification step

The fractions that gave bands were pooled together for further purification by ion-exchange chromatography. The fractions pooled were B5-B7 from the first purification and B4-B10 from the second purification. The protein amount was 9.34 mg in the pooled fractions but a great deal of this might actually be due to the contaminant. Cation-exchange chromatography was used for further purification and was performed using an ÄKTA-system at the SLU Molecular biology department at BMC. A salt gradient from 0 to 1 M was used for separation and the protein was eluted in 2.2 ml fractions at a flow rate of 1 ml/min. The software used to analyse the separation of proteins was the UNICORN program. The contaminant was eluted together with the flow-through while the TGG5 protein was eluted by increasing the salt gradient (0-1 M NaCl) (fig. 5).

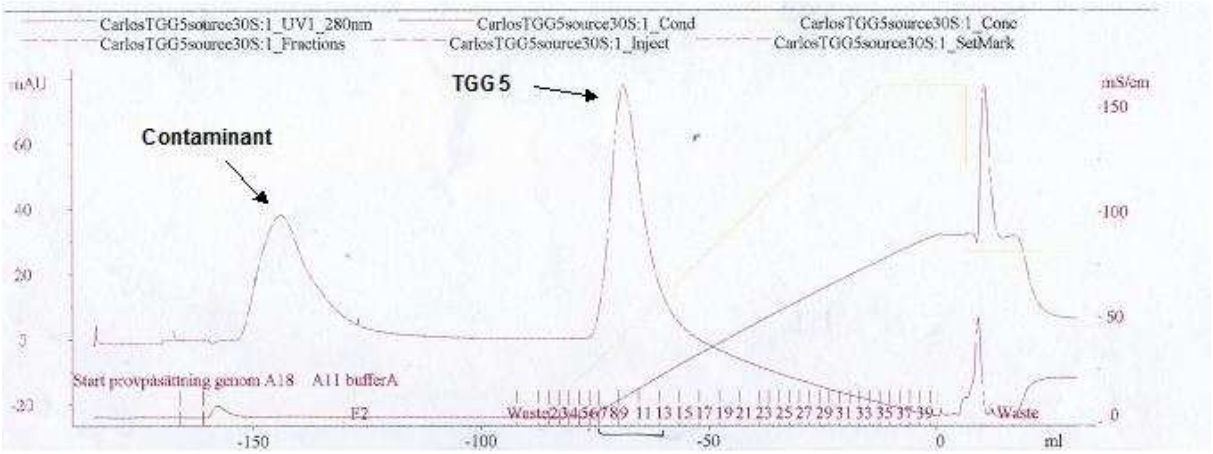


Fig. 5. The contaminant is eluted together with elution buffer A (20 mM sodium phosphate pH 6.75) while TGG5 is eluted with an increasing salt gradient. Fractions 6-12 were part of the TGG5 peak and saved for SDS-PAGE analysis.

Fractions 6 to 14 were run on two minigels using the Phastsystem workstation (GE Healthcare). Fraction 6, 12, 13 and 14 gave one weak band, fractions 7-9 gave a strong band and fraction 10 to 11 gave very strong bands. A low molecular weight marker was used and it indicated that the protein had a molecular mass between 45 kDa and 66 kDa. Some of the bands seemed to consist of two bands situated very close to each other. It is very possible that this is a result of heterogenous glycosylation. Such heterogenous glycosylation might be one of the reasons that it is difficult to crystallize the protein. The fractions were pooled and estimated to contain 2 mg/ml. The buffer was changed to 50 mM NaAcetate to allow deglycosylation with α-mannosidase. Treatment with α-mannosidase can cleave alpha-(1,2)-, (1,3)- and (1,6)- terminal residues giving more homogenous preparations. The pooled fractions were concentrated to 2 ml and to this 50 µg α-mannosidase and 48 µl zinc acetate was added. It was left at room temperature for two days, When inspecting the sample it looked like there was a little precipitation thus I added 2 ml sodium phosphate assuming that it was the NaAcetate buffer that was the cause. But I forgot that zinc falls out in the presence of phosphate and there was a lot of precipitation. I tried to remove the precipitate but ended up with only 0.4 mg/ml. There was no point in removing the cleaved mannosidase residues with ion-exchange and thus it ended the first crystallization attempt.

3.2. Second TGG5 crystalization attempt

3.2.1. Crude extract preparation and first step in purification

Crude extracts were prepared from tubes stored at -20°C that were made in the lab in previous experiments. I was told that there were no particular experiments planned for these cell pellets and therefore decided to use them for another attempt at protein crystallization. Six separate crude extract preparations were made with cell material from the tubes, the amount of protease inhibitor cocktail varied between the different preparations due to shortage. The IMAC step was performed at 4°C (the first attempt was performed at room temperature) and two types of pre-packed columns were used, either Histrap Hp 5 ml column or Ni-NTA superflow cartridge 1 ml columns. The columns were stripped and re-loaded with nickel between each run. From the SDS-PAGE gels made for each separation it seemed like a lot of the TGG5 protein did not bind. Therefore the flow-through from all separations was pooled and used in three separate IMAC runs to try to bind the remaining TGG5. Myrosinase assays were performed on all fractions acquired from each IMAC chromatography step. A problem with the standard curve linearity was experienced when trying to estimate the protein concentration using the Bradford assay. Instead I pooled all fractions with myrosinase activity (table 3)

and used SDS-PAGE protein separation to make a rough estimate on how much protein there was in the pooled samples (fig. 6).

Table 3. Eluted fractions with myrosinase activity were pooled together and tested for further myrosinase activity. These pooled fractions were to be used for further purification to remove the 30-40 kDa contaminant. The two remaining separations from the pooled TGG5 flowthrough were not tested for myrosinase activity.

TGG5 separate IMAC runs	**Pooled elution B fractions**	**Volume**	**Myrosinase activity (mmol /min)**
TGG5 (I)	B4, B5, B6, B7, B8, B9, B10, B11	7 ml	47.3
TGG5 (II)	B4, B5, B6, B7, B8, B9, B10	6.5 ml	27.3
TGG5 (III)	B4, B5, B6, B7, B8, B9, B10	6.5 ml	28
TGG5 (IV)	B4, B5, B6, B7, B8, B9, B10, B11	7.5 ml	37.3
TGG5 (V)	B2, B3, B4, B5	3.5 ml	40.1
TGG5 (VI)	B2, B3, B4, B5	3.5 ml	37.2
Pooled TGG5 flowthrough (I)	B2, B3, B4, B5	3.5 ml	18.7
Pooled TGG5 flowthrough (II)	B5, B6, B7	3 ml	
Pooled TGG5 flowthrough (III)	B5, B6, B7, B8	3.5 ml	

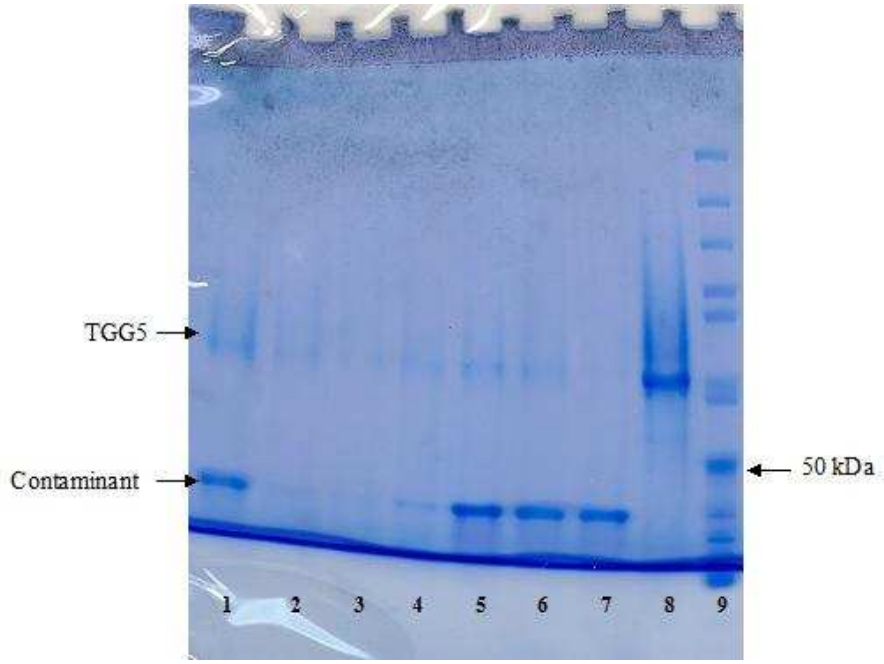


Fig. 6. Lanes 1-7 contained the following pooled fractions; TGG5 (I), TGG5 (II), TGG5 (III), TGG5 (IV), TGG5 (V), TGG5 (VI) and pooled TGG5 flow-through (I) respectively. Lane 8 separates a TGG4 protein unrelated to this experiment. Lane 9 contained the marker. The two remaining separations from the pooled TGG5 flow-through are not represented in the gel.

3.2.2. First gel filtration

The pooled fractions were divided into three samples to be purified separately, the reasoning behind this was that if mistakes were made I would always have backups. The first sample was 15 ml and underwent buffer change to 20 mM imidazole and 0.1 M NaCl with concentration column (Vivaspin 20). The sample was concentrated to 1.5 ml and was going to be used for gel filtration. I later learned that gel filtration changes the buffer automatically making the buffer change earlier unnecessary. An 0.5 ml aliquot of the sample was run through a Superdex 200 prepgrade column (1 ml was accidently spilled) using the ÄKTA system. UNICORN was used to record the curve, it seems like protein was lost through the initial buffer change and loss of the 1 ml (fig. 7). Both the contaminant and the TGG5 protein were collected in 2 ml fractions at a flowrate of 0.5 ml/min. A number of fractions were run on a SDS-PAGE mini gel with a 8-25% polyacrylamide gradient using the Phastsystem. These fractions corresponded to the first peak, second peak and fractions in between. No bands were observed on the gel probably because the fractions were too diluted (fig. 8). The fractions from this first gel filtration were not used for any further experiments.

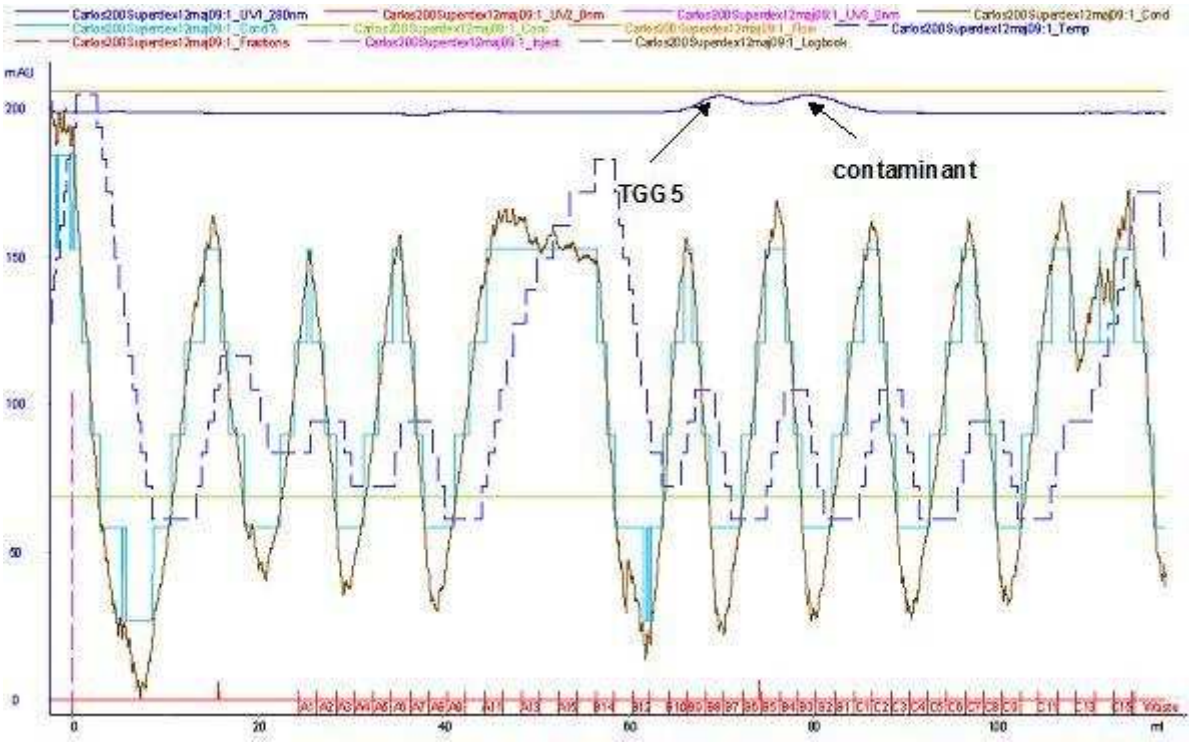


Fig. 7. The imidazole buffer of the sample gave a high background absorbance but it is still possible to observe the two peaks indicating the presence of the contaminant and TGG5. Gel filtration separates proteins based on differences in their size (roughly proportional to the molecular mass and this is the property that also allows buffer change) with the larger TGG5 protein (60-70 kDa) eluting before the smaller contaminant (30-40 kDa).

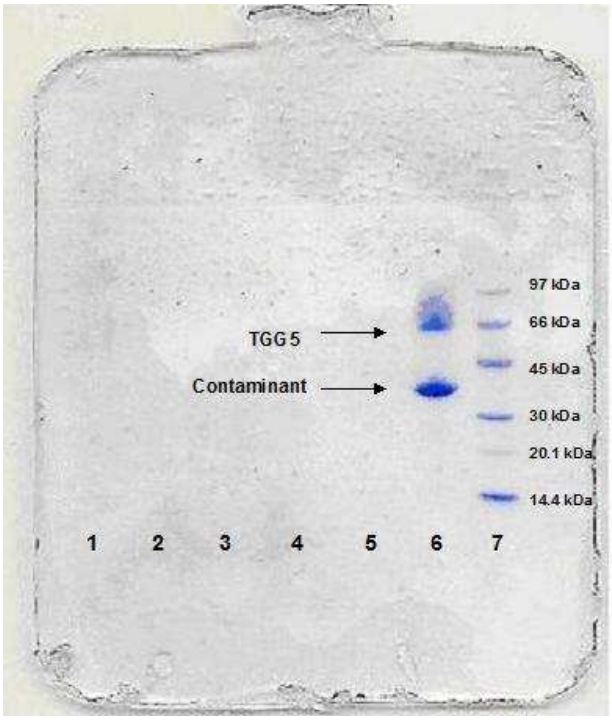


Fig. 8. Lane 1 contains the fraction before the first peak (TGG5), lane 2 contains the first peak, lane 3 has the second peak (contaminant), lane 4 the fraction between the two peaks, lane 5 the fraction after the second peak. Lane 6 contains concentrated TGG5 that have not been separated with gel filtration, and lane 7 a Low molecular weight marker used to approximate the molecular mass of the proteins.

3.2.3. Second gel filtration

I concentrated 10 ml of the pooled fractions to 1.5 ml for a second attempt at gel filtration but no buffer change was attempted as it is inherrent when running gel filtration. Two peaks were seen on the curve, the first one being TGG5 and the second the contaminant (fig. 9). A SDS-PAGE gel was run using a number of fractions but the protein was to diluted to give bands, except for a fraction containing the second peak (fig. 10). The first peak was fractionated into 0.5 ml fractions, thus there is small chance of getting a protein concentration big enough to give bands. I took the twelve 0.5 ml fractions making up the first peak (the small kink in the beginning not included) and concentrated it to 1 ml and a protein concentration of 0.24 mg/ml and analysed this sample with a SDS-PAGE minigel to make sure the sample only contained TGG5 (fig. 11).

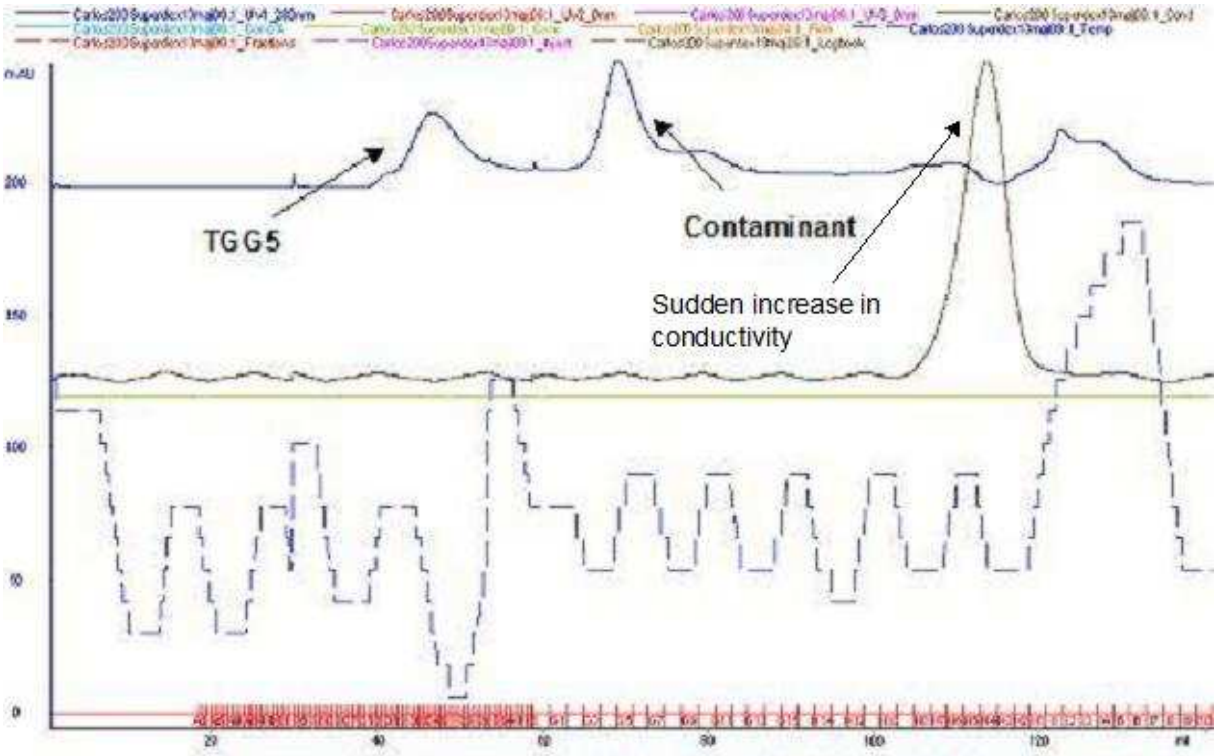


Fig. 9. The two peaks were more distinct this time with the larger TGG5 protein eluting before the smaller contaminant. The TGG5 protein was collected in 0.5 ml fractions while the second peak was collected in 2 ml fractions (there is no special reasoning behind this I just accidently set in the wrong parameters for fraction collecting). A sudden increase in conductivity resulted in two extra peaks occurring, most probably representing ghost peaks.

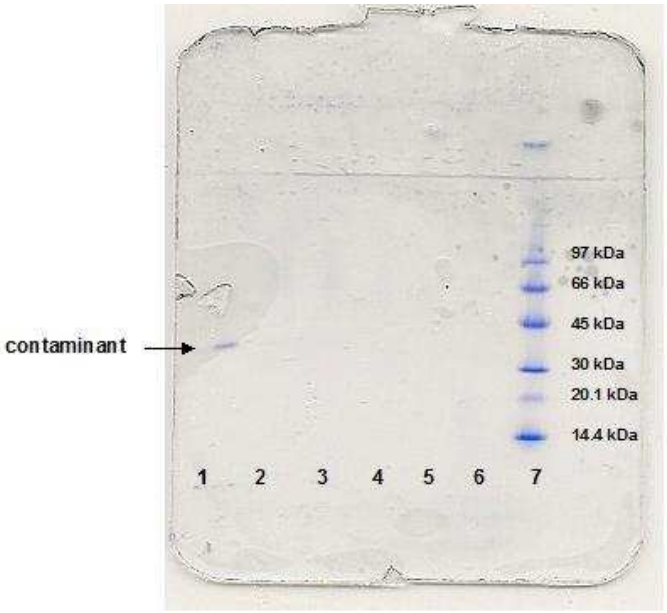


Fig. 10. The fractions from the second gel filtration were run on minigel using the phastsystem workstation. The first lane has a fraction containing the contaminant (second peak), lane 2 and 3 contain TGG5 (first peak), lane 4, 5 and 6 are fractions from the two extra peaks (occurring close to the sudden increase in conductivity), lane 7 has the low molecular weight marker. Only the contaminant gave a band, the TGG5 fractions were probably to dilute to give bands. The two extra peaks were probably a result of the sudden increase in conductivity and not due to presence of proteins.

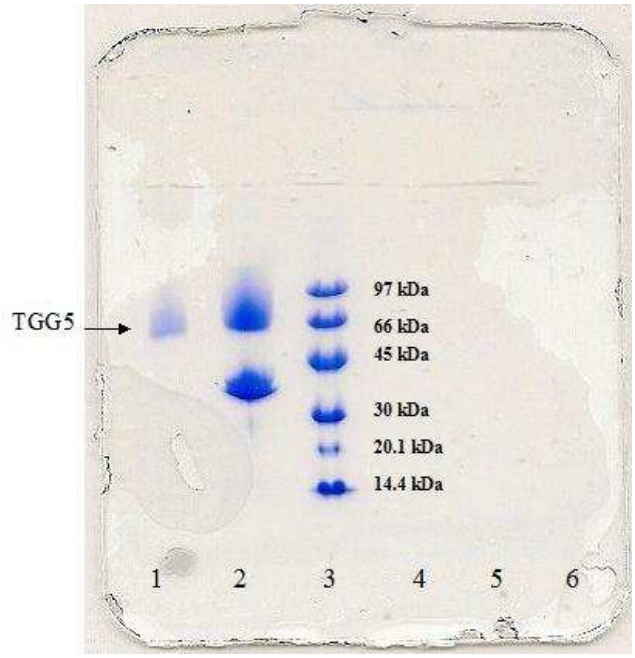


Fig. 11. TGG5 fractions from the second gel filtration were pooled and loaded on the first lane of a minigel, concentrated TGG5 that was not submitted to gel filtration was loaded on the second gel as positive control. The third lane contains a low molecular weight marker. Lane four, five and six are empty.

3.2.4. Hanging drop vapour diffusion crystallization

The pooled fractions of the first peak acquired from the second gel filtration were further concentrated to 100 µl and 2.44 mg/ml. This sample was to be used for hanging drop vapour diffusion. Previous crystallization attempts performed with 0.2 M Litiumsulphate, 0.1 M phosphate-citrate buffer, pH 4.2 and 20% PEG1000 had resulted in some protein crystal formation. The idea was therefore to set up drops around these conditions. The conditions had to be slightly modified because of availability of the chemicals (PEG1500 instead of 1000) and because some of the drops will contain α-mannosidase and α-mannosidase + endo-H. Previous experience have shown that phosphate makes the zinc needed for α-mannosidase to be active to fall out, therefore I changed phosphate-citrate for sodium-citrate, pH 4.2 and in some drops pH 5.0. In total 36 drops were set up using two plates (18 drops in each plate), 12 with only TGG5, 12 with TGG5 + α-mannosidase and 12 with TGG5 + α-mannosidase+ Endo-H. The drops consisted of 2.4 µl protein and 0.8 µl reservoir buffer while there was 1 ml buffer in each reservoir. The deglycosylation protein was 1:100 of the TGG5 protein amount. I could only observe the drops for 3-4 weeks but some tendencies could be observed. One is that PEG1500 concentrations over 15% gave precipitations, another is that $LiSO_4$ seems to make the proteins less likely to precipitate. The drops containing only TGG5 did not precipitate (unless the PEG1500 concentration is over 15%), 4 drops containing TGG5 + α-mannosidase remained clear while all drops containing both deglycosylation enzymes precipitated, except one drop containing $LiSO_4$.

3.2.5. Third gel filtration

The remaining 22.5 ml of the pooled TGG5 fractions were used for a third gel filtration. The same steps as before were taken with the separated proteins being collected in 2 ml fractions. The graph showed two peaks although the first peak containing the TGG5 proteins had an extra peak, it is probably the same as the kink seen in the second gel filtration graph although at a higher concentration (fig. 12). Fractions making up the first peak (TGG5), the extra peak in its vicinity and the second peak (30-40kDa contaminant) were separated on a SDS-PAGE minigel using the phastsystem workstation (fig. 13).

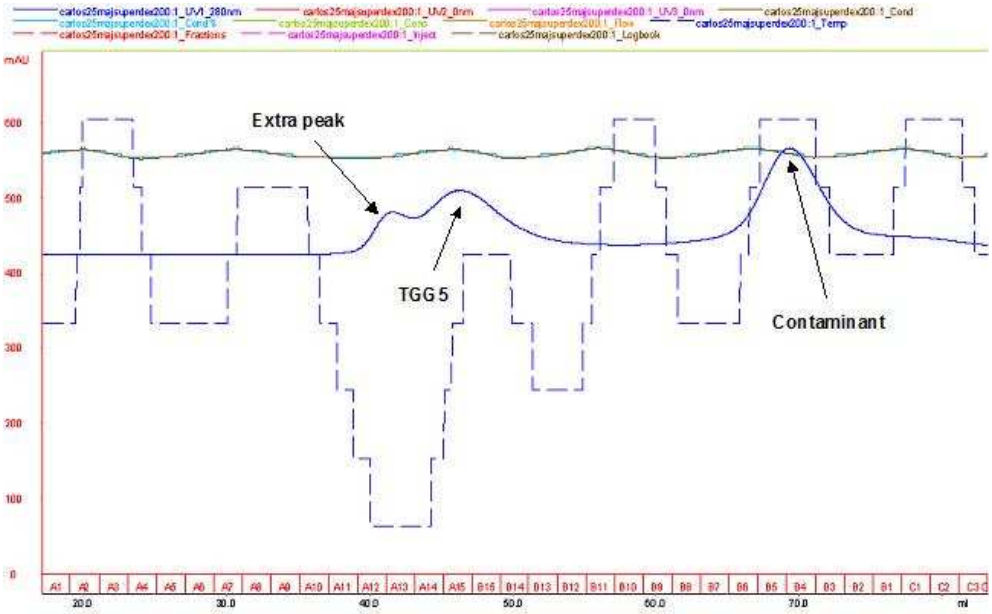


Fig. 12. The graph shows the two peaks characteristic for TGG5 and the contaminant but also a smaller third peak in proximity to the TGG5 peak.

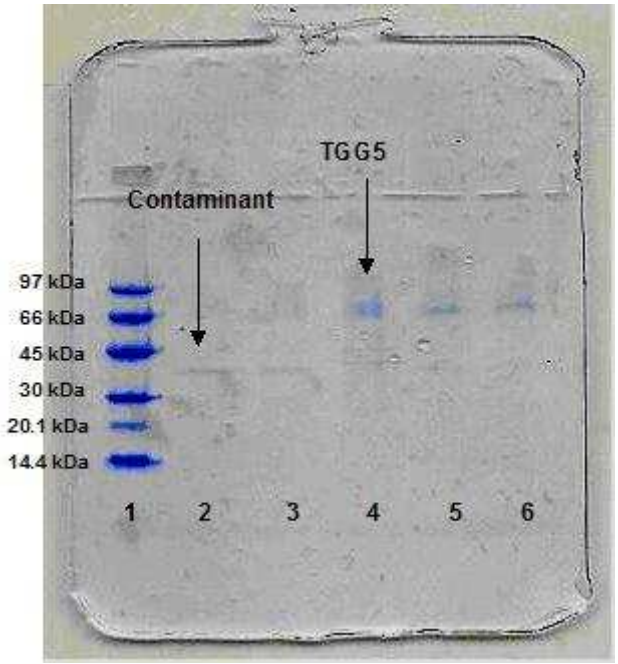


Fig. 13. Lane 1 contains the LMW marker, lane 2 contains the extra peak, lane 3 is a fraction between the extra peak and the first peak, lane 4, 5 and 6 are fractions of the first peak. It seems like the contaminant was not completely separated from the TGG5 protein (The small extra peak close to TGG5 in the third gel filtration graph is probably 30-40 kDa contaminant that did not separate). I would account this to a larger amount of contaminant in the sample rather then a difference in the gel filtration procedure.

3.2.6. Optimum solubility screen (OSS)

Four fractions making up the first peak were pooled and concentrated to 100 µl and a protein concentration of 4.81 mg/ml. Some of this sample were to be used to set up a optimum solubility screen in order to test TGG5 solubility in common buffers at different pH values (table 4). 1 M buffers were used to reduce any effect the 20 mM Imidazole buffer of the sample might have on the test. Hanging drop vapour diffusion was used and the drops consisted of 0.8 µl protein and 0.8 µl reservoir

buffer with 50 µl buffer in each reservoir. The buffer does not seem to have any major effect on the protein. neither does the pH, since no precipitation was observed in any of the drops.

Table4. The TGG5 protein solubility was tested in different buffers and a range of pH values.

1M Buffer pH range 3 – 10
Glycine pH 3.0
Citric acid pH 3.2
Formic acid pH 3.7
Citric acid pH 4.0
Na-acetate pH 4.5
Na/K phosphate pH 5.0
Na-citrate pH 5.5
Na/K phosphate pH 6.0
Bis-tris pH 6.0
MES pH 6.2
ADA pH 6.5
Bis-tris-propane pH 6.5
Cacodylate pH 6.5
AmAc PH 7.0
Na/K phosphate PH 7.0
HEPES pH 7.5
Tris pH 7.5
Bicine pH 8.0
Imidazole pH 8.0
Tris pH 8.5
CHES pH 9.0
CHES pH 9.5
CAPS pH 10

3.2.7. JCSGplus screen

The remaining 4.81 mg/ml TGG5 sample was used to set up a JCSGplus screen which tests a large variety of crystallization variables. The TGG5 protein sample was mixed with 1:100 α-mannosidase and 1:100 endo-H before setting up the drops. No Zinc was added as it might affect the screen. Sitting drop vapour diffusion was used and each drop contained 0.8 µl protein and 0.8 µl reservoir buffer, 50 µl buffer was added to each reservoir. The majority of the drops were either precipitated or clear except one of the drops in calcium chloride, 0.1 M bis-tris, pH 5.5 and 45% MPD. This drop contained a very small crystal of about 60 x 20 micrometer (fig. 14).

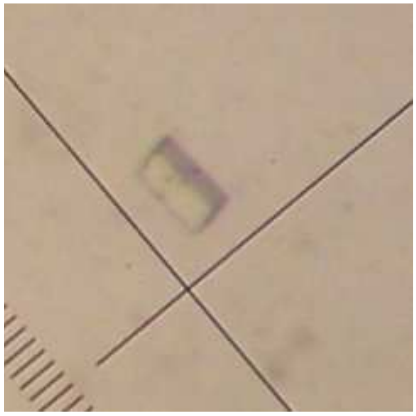


Fig. 14. A small crystal found in the JCSGplus screen.

3.3. TGG5 Molecular mass estimation using mass spectrometry

I wanted to do a rough molecular mass estimation of TGG5 before and after α-mannosidase and endoH deglycosylation. The MALDI-TOF procedure was going to be performed by Åke Engström and he preferred to have the proteins separated on a SDS-PAGE gel. TGG5 concentrated to 2.44 mg/ml was used to set up hanging drops: 1 µl with only TGG5, 1µl with 1:100 α-mannosidase and 1 µl with 1:100 α-mannosidase + 1:100 endoH. The protein was in its original 20 mM imidazole and 0.1 M NaCl buffer and 1 ml of this buffer was also added to the reservoir. The drops were incubated at 20°C for 6 days before being separated on a minigel using the Phastsystem (fig. 15). The gel was sent to Åke Engström for mass spectrometry. He was not able to estimate the mass from the samples on the gel and requested liquid samples for the three samples. I never threw anything away and was able to give him 5 µl TGG5, 4 µl TGG5 + α-mannosidase and 5 µl α-mannosidase + endoH. The deglycosylation enzymes still have a 1:100 ratio but the origin of the samples was hard to pinpoint. Åke was able to acquire mass spectrums from the TGG5 and the TGG5 + α-mannosidase samples but there was too much disturbances in the sample containing EndoH to give any mass spectrum. In the data from the TGG5 sample, two peaks could be seen very close together, the larger peak had the molecular mass 65.4 kDa and the smaller peak 63.6 kDa (fig. 16). The mass spectrum of the TGG5 + α-mannosidase gave only one peak with a molecular mass of 64.4 kDa (fig. 17), thus α-mannosidase must have trimmed of about 1 kDa of mannosidase glycans, the extra peak might have been a different glycosylation form of the TGG5 protein as its very common that proteins have heterogenous glycosylation (Daly & Hearn 2005). This might explain the double bands that sometimes occur on SDS-PAGE gels and some of the difficulties experienced when trying to crystallize the protein.

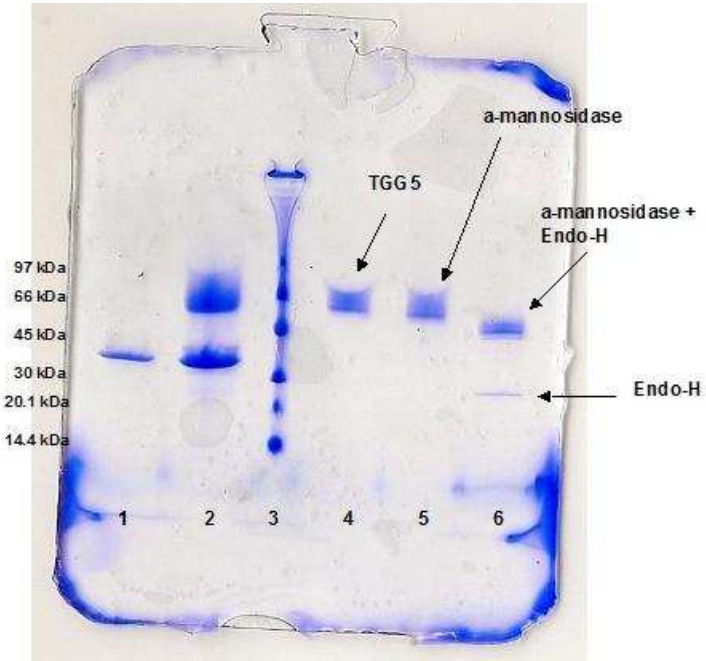


Fig. 15. Lane 1 contains the contaminant (to be used in another experiment), lane 2 contains concentrated TGG5 (before gel filtration), lane 3 contains a LMW marker, lane 4 contains TGG5, lane 5 contains α-mannosidase treated TGG5 while lane 6 contains TGG5 that had been treated both with α-mannosidase and EndoH. The contaminating band in lane 6 is probably EndoH as it corresponds to EndoH's Molecular mass of 28.5 kDa.

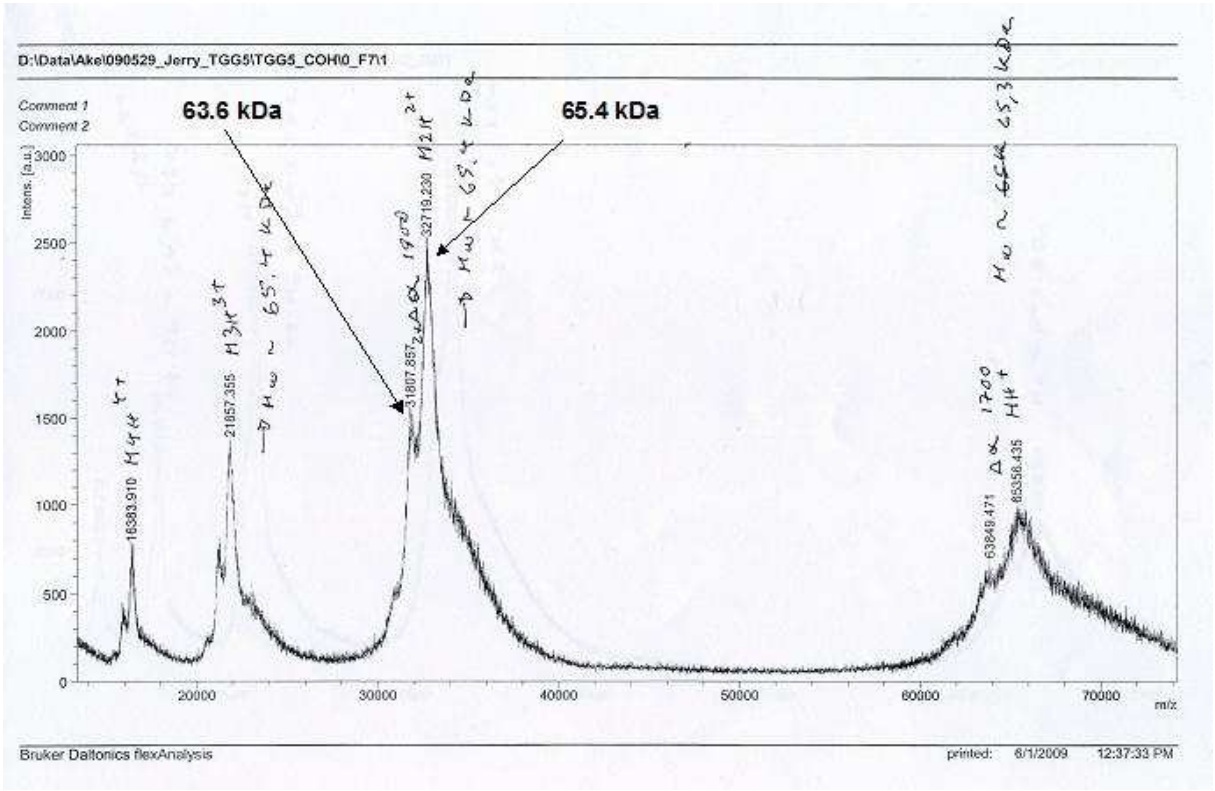


Fig. 16. The larger 65.4 kDa peak is presumed to be the TGG5 protein and the smaller 63.6 kDa peak might represent a different glycosylation variant of the same protein.

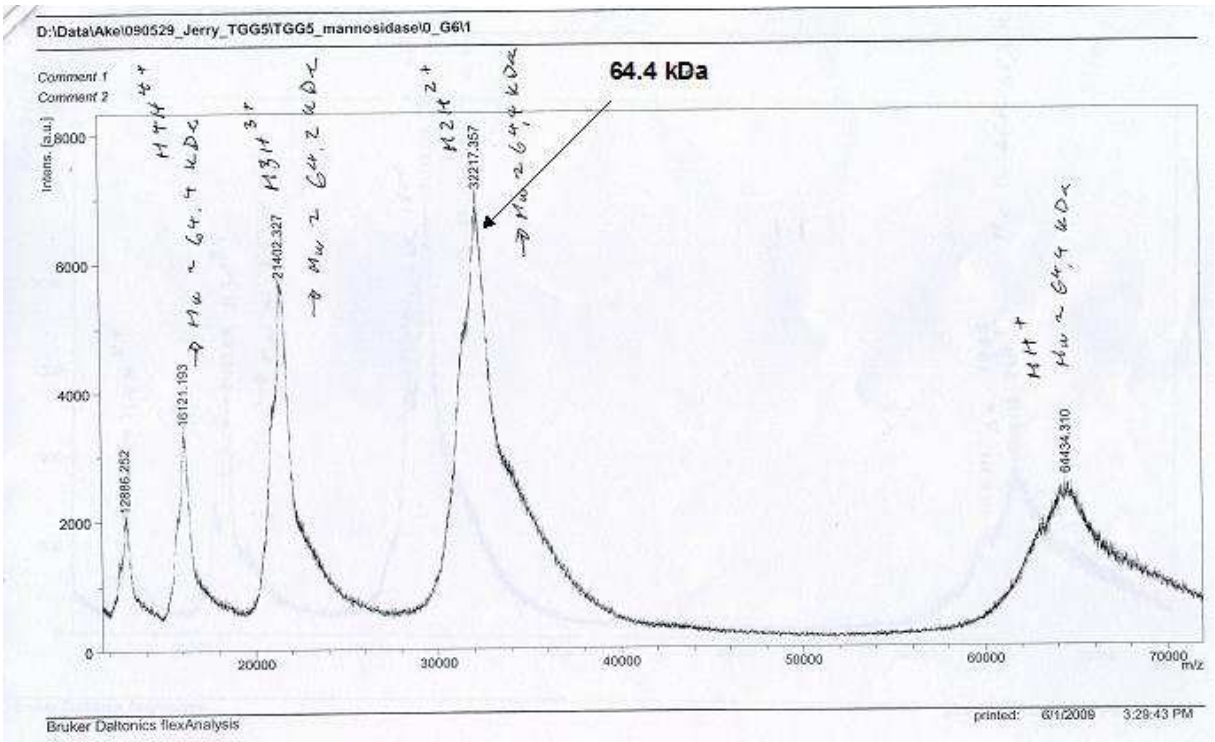


Fig. 17. The mass spectrum of the α-mannosidase treated TGG5 gave a 1 kDa reduction in the molecular mass (65.4 kDa → 64.4 kDa). The 63.4 kDa peak is absent in the mass spectrum and my guess is that it was trimmed to 62.4 kDa and became undistinguishable from the noise in the mass spectrum.

.

3.4. Determination of the identity of the contaminant

A series of experiments were done to determine the identity of the 30-40 kDa contaminant, the experiments included western blot analysis, alcohol oxidase assays, gel filtrations, SDS-PAGE and MALDI-TOF to estimate the mass and peptide mass fingerprinting.

3.4.1. Western blot analysis

Western blotting was performed using a TGG5 sample with myrosinase activity and a primary polyclonal antibody with histidine affinity was used to bind the 6xhis tag of the recombinant TGG5 protein. An HRP conjugated secondary antibody was used to bind the primary antibody and allow visualization. When visualized it did not seem like there was any TGG5 present to bind to, but unexpectedly the his-binding antibody bound to the 30-40 kDa contaminant. When running the gel a magic marker was used.

3.4.2. Enzyme assay of alchohol oxidase

The contaminant has been hypothesized to be AOX1, a protein highly abundant when *P. pastoris* is grown with methanol as the only carbon source. I decided to test the activity of fractions acquired from the IMAC purification containing both the TGG5 protein and the 30-40 kDa contaminant. The same test was also performed on fractions acquired from gel filtration where the TGG5 protein and the contaminant were purified into individual fractions. A crude extract of proteins from *P. pastoris* was used as positive control as it was assumed to contain AOX1. A sample with 100 mM potassium phosphate buffer instead of protein was used as a negative control. The colorimetric assay used to measure the alcohol oxidase activity quickly overshoot in the crude extract while no activity was observed in any of the other samples (table 5). Therefore it is hihly unlikely that the contaminant is an alcohol oxidase enzyme.

Table 5. Triplicates were made for every sample and the contaminant was tested at a lower (2 µl) and a higher concentration (10 µl). TGG5 (I) is a sample consisting of all pooled fractions from an IMAC purification, while a concentrated first peak contained only TGG5. When measured at 405 nm for absorbance it is quite clear that there is no alcohol oxidase activity whatsoever.

Sample	Absorbance (405nm)
Concentrated second peak (2 µl)	-0.026
Concentrated second peak (2 µl)	0.012
Concentrated second peak (2 µl)	-0.001
Concentrated second peak (10 µl)	-0.031
Concentrated second peak (10 µl)	-0.039
Concentrated second peak (10 µl)	-0.039
TGG5 (I) (2 µl)	-0.026
TGG5 (I) (2 µl)	0.012
TGG5 (I) (2 µl)	-0.001
TGG5 Crude extract (2 µl)	3.339
TGG5 Crude extract (2 µl)	3.339
TGG5 Crude extract (2 µl)	3.339
Concentrated first peak (2 µl)	-0.026
Concentrated first peak (2 µl)	0.012
Concentrated first peak (2 µl)	-0.001

3.4.3. MALDI-TOF

A contaminant containing fraction was run on a SDS-PAGE minigel using a Phastsystem and sent for mass estimation and peptide mapping (fig. 15). The mass was estimated to approximately 37 kDa (fig. 18) correlating quite well with the mass estimated from the SDS-PAGE minigels (e.g. fig. 15). The contaminant was also peptide mass fingerprinted and the data was run up against the Mascot database but no protein was found in the database. Åke Engström performed de novo sequencing on the contaminant translating three peptides into their amino acid sequence (table 6). These three peptide sequences were Blasted and they all gave the same result; Mitochondrial alcohol dehydrogenase isozyme III. The alignment was more or less 100%. Therefore the contaminant most likely represents an alcohol dehydrogenase isozyme.

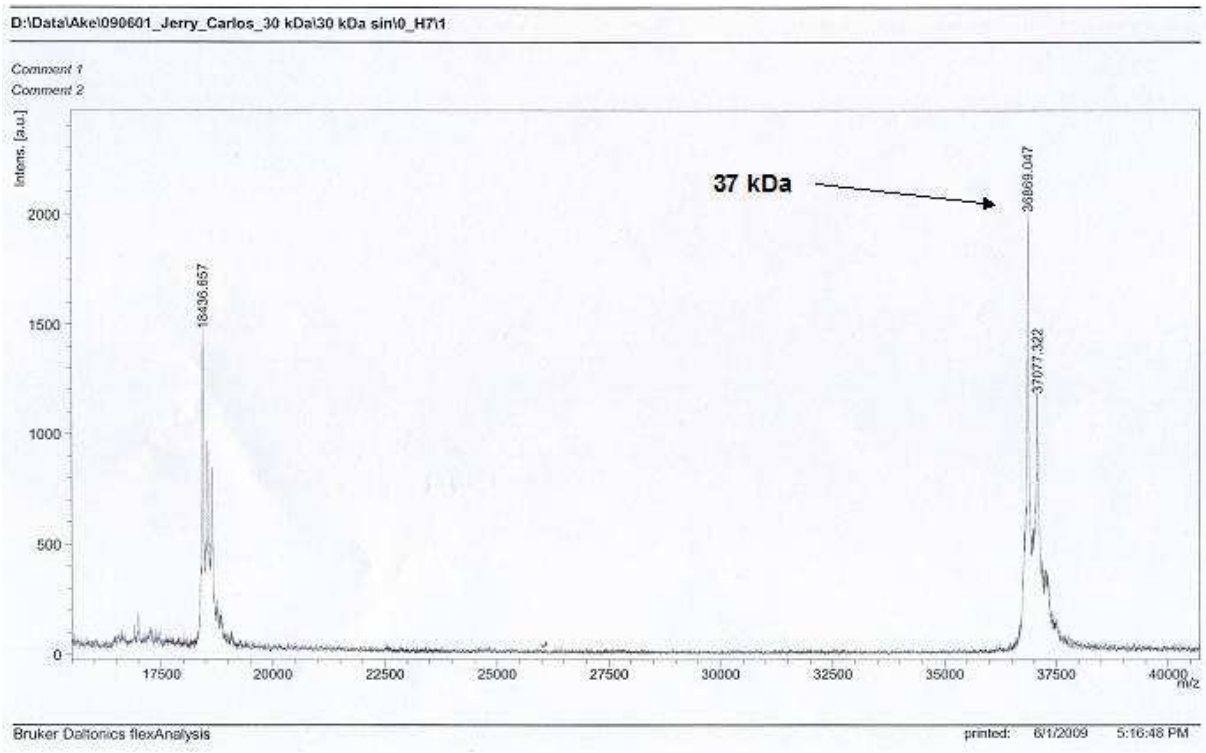


Fig. 18. The mass of the contaminant was estimated to 37 kDa using MALDI-TOF.

Table 6. Translated amino acid sequences.

Sequence 1	VYEQMEAGAIIGR
Sequence 2	ADISGFTHDGSFQQYATADATQAAR
Sequence 3	AINQSVQYVR

3.5. Myrosinase mutant activity studies - first attempt

Five mutants of TGG4; E418A, E418H, S214A, I278A, E418Q and a TGG4 wild type were sample prepared for separation by immobilized metal-ion affinity chromatography (IMAC). The separations were carried out at room temperature although the samples and all elution buffers were kept on ice. Out of these the S214A mutant was too unstable to give good readings while the E418H and I278A mutants were only stable enough to be tested once for myrosinase activity using sinigrin (S-β-glucoside) as a substrate. The TGG4 wild type and E418Q mutant were stable enough to be tested for myrosinase activity using sinigrin and pNPG a substrate usually hydrolysed by the O-β-glucosidases.

An attempt was made on testing how different concentrations of ascorbate affect the myrosinase enzymatic activity but the results were very dubious and will not be described furher. The mutants were made in previous experiments by other members of the laboratory. I merely used the cell pellet of the mutants for myrosinase activity studies.

3.5.1. Myrosinase activity using sinigrin as substrate

The myrosinase activity of the mutants was tested using sinigrin, a S-β-glucoside, and the myrosinase activity calculated as mmol glucose released per minute (Units/ml) as glucose, sulfate and an aglycone is released when sinigrin is hydrolysed. The glucose will only be read as total activity because the samples contained the 30-40 kDa contaminant and it was uncertain how large part of the total protein concentration was due to the contaminant. The TGG4 wt, E418Q, E418H and I278A had a total activity of 32.8, 17,11.5 and 0.87 units/ml, respectively. It seems like the change of the glutamate active site to glutamine and histidine did not take away all activity while a change in isoleucine, a proposed aglycone binding site, gave a big reduction in activity (fig. 19).

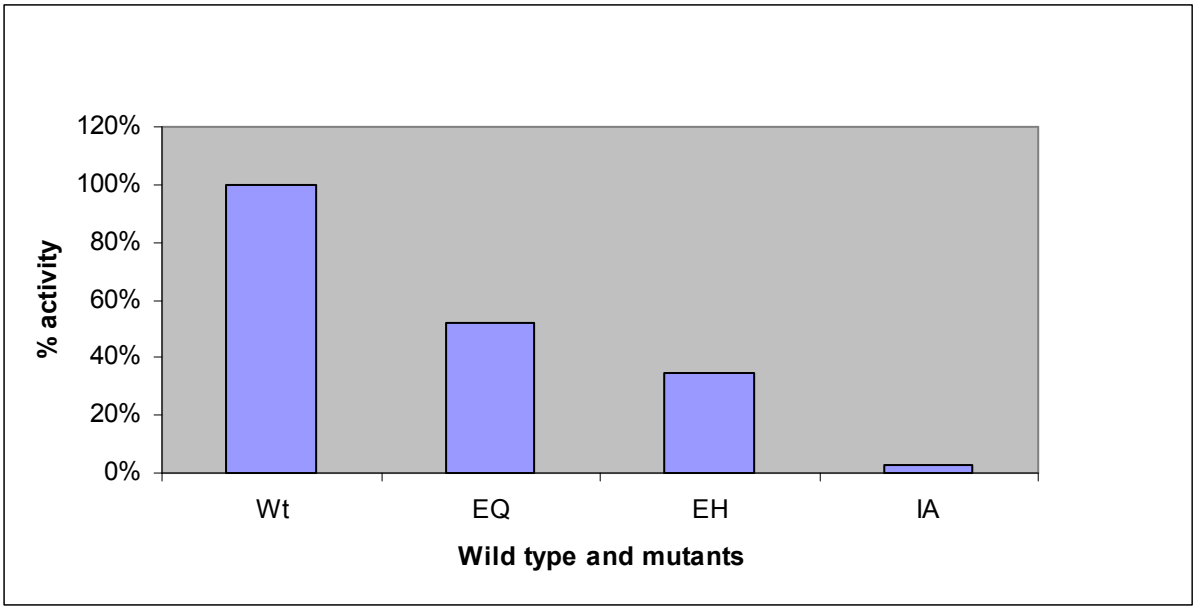


Fig. 19. The activity as mmol glucose/min was converted to percent activity setting the wt as 100% and comparing the mutants with the wild type activity. The E418Q mutant showed only 52% activity while the E418H mutation had 35% of the wt myrosinase activity. The I278A mutation had the biggest loss of activity giving only 3% of the normal activity.

3.5.2. O-β-glucosidase activity using pNPG as substrate

pNPG is an O-β-glucoside and commonly hydrolysed by O-β-glucosidases. Out of the mutants only the TGG4 wt and the E418Q mutant were stable enough to be further tested. The assay is similar to the myrosinase assay with the difference that the amount of free aglycone is measured at 405nm as released p-nitrophenol and converted to mmol p-nitrophenol/min. TGG4 gave 1.56 units/ml while the E418Q mutant gave 5.91 units/ml, thus the enzymatic activity in the mutant was increased almost four-fold by substituting the glutamate residue with glutamine (fig. 20).

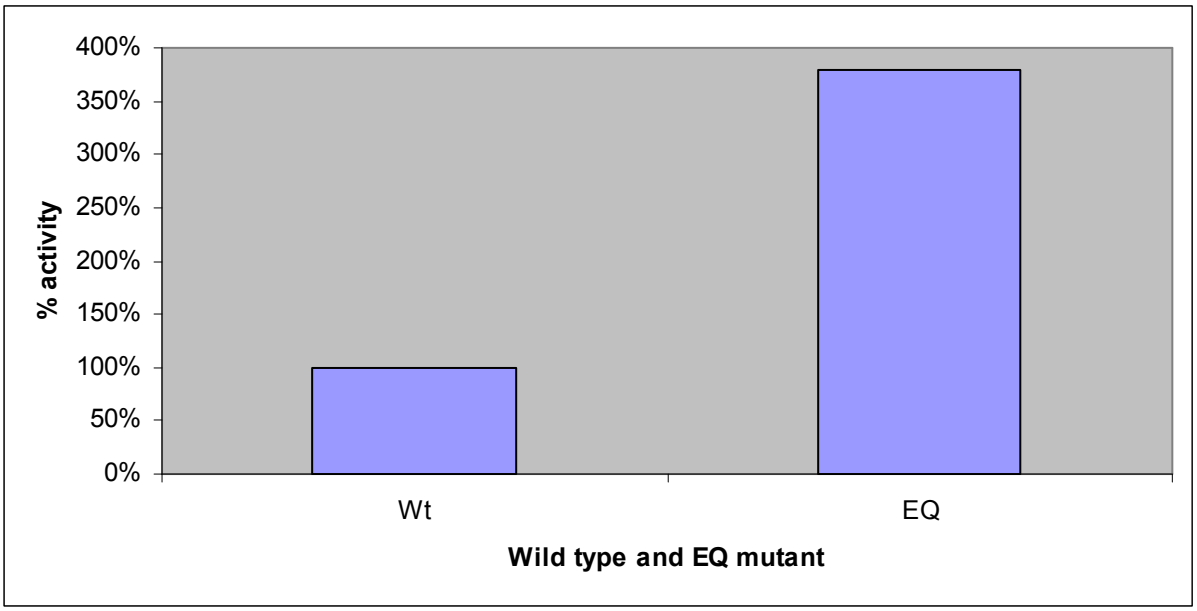


Fig. 20. The activity in mmol p-nitrophenol/min was converted to relative activity setting the wild type activity as 100 percent. The mutant had 379% activity and was thus 279% more active then the wild type.

3.6. Mutant myrosinase activity studies – second attempt

Mutants used in the previous experiment were grown in 10 ml BMGY in 50 ml falcon tubes. The BMGY was removed and the mutants were expressed in 10-15 ml BMMY in 50 ml falcon tubes. The reason I grew the mutants in falcon tubes was that earlier I had experienced problems with contamination and smaller volumes are easier to keep sterile. The seven mutants grown were K218A, E418D, S214A, F475A, Q211E, Q211A and E418A. All these mutants were sample prepared and the protein purified using IMAC. A TGG4 wild type sample prepared in previous experiment was taken from a -20°C freezer, sample prepared and purified in the same manners as for the mutants. The IMAC procedure was performed at 4°C as it seems the myrosinase protein becomes unstable when purified at room temperature. When separating the proteins on a SDS-PAGE gel it seems like very little protein bound to the column (fig. 21). All mutants were stable enough to be tested for myrosinase activity with both sinigrin and pNPG as substrate. The potential enzymatic activation by ascorbate was tested using sinigrin as substrate. The total activity of the myrosinase was used because the sample was not pure enough to give a protein concentration specific for myrosinase.

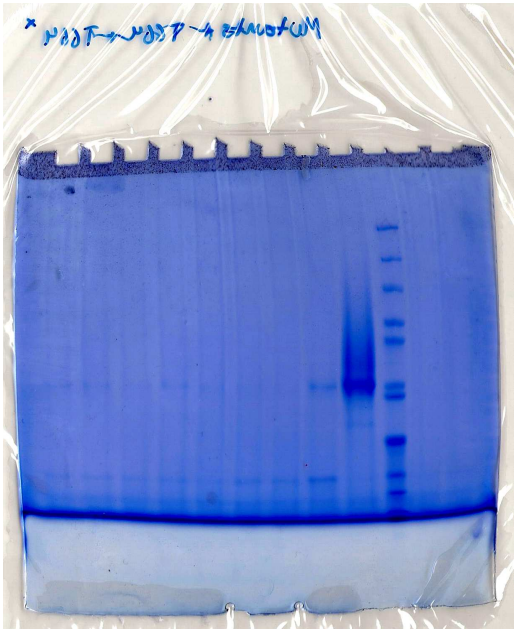


Fig. 21. From left to right the following samples are present; K218A, E418D, S214A, F475A, Q211E, Q211A, E418A, TGG4 (WT), unrelated sample and the marker. The bands are weak indicating that there is very little protein in the sample. The good news is that the mutants seem to have a similar amount making them more comparable.

3.6.1. Myrosinase activity using sinigrin as substrate

Like the first attempt, sinigrin was used as the substrate and the amount of released glucose was measured at 550nm and the absorbance value converted to mmol glucose/min. TGG4 wt, F475A, S214A, Q211A, E418D, Q418E, K218A, and E418A mutant had an activity of 18.3, 10.7, 3.15, 1.51, 7.26, 5, 6.3 and 1 mmol glucose/min. The F475A and E418D mutant retained more or less half their activity while the E418A and Q211A mutants lost almost all their activity. The three remaining mutants (S214A, Q211E and K218A) had activities ranging between 17-35% of the original activity (fig. 22).

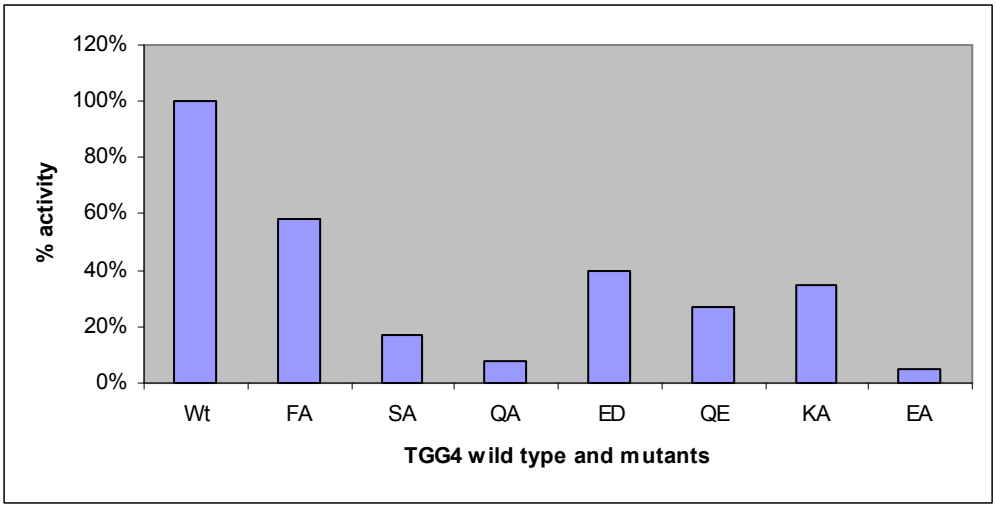


Fig. 22. From the graph it seems like F475A, E418D and K218A retained most of their activity at 58%, 40% and 35% respectively. S214A and Q211E mutants formed an intermediate group at 27% and 17%, while Q211A and E418A retained the least activity at 8% and 5%, respectively.

3.6.2. O-β-glucosidase activity using pNPG as substrate

An activity assay performed using pNPG as substrate showed that all mutants except Q211E and E418D had activity.TGG4 wt, E418A, S214A, F475A, Q211A and K218A, showed an total activity of 0.47, 0.14, 0.2, 0.28, 0.3, and 0.31 mmol p-nitrophenol/min, respectively. These are values considerably lower than when using a S-β-glycoside, the natural substrate of myrosinases. When converting the values into percent activity it also seems like the difference between the mutants and the wild type is less than with sinigrin with F475A, Q211A and K218A giving a little more then half the activity of the wild type, and E418A and S214A giving a little less then half (fig. 23).

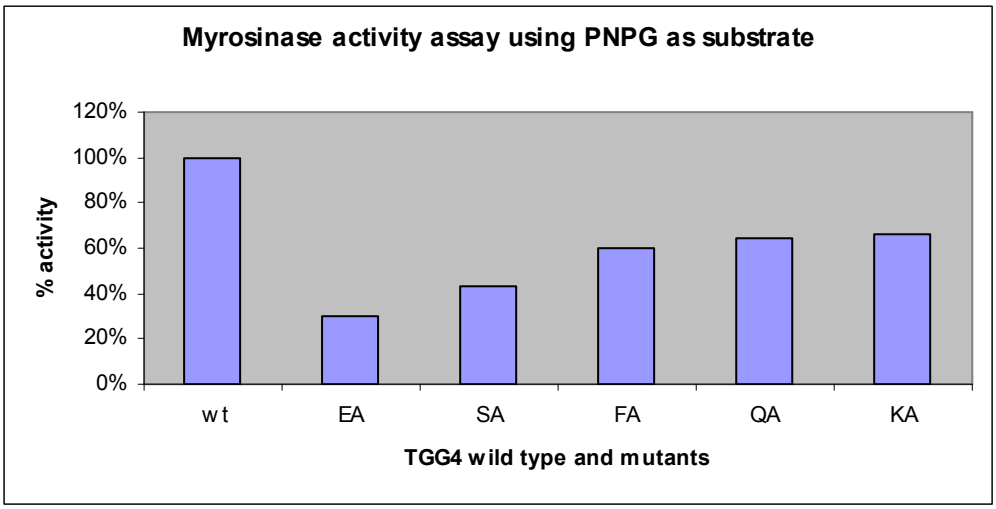


Fig. 23. The three mutants F475A, Q211A and K218A have a similar activity at 60%, 64% and 66%, respectively. E418A and S214A have a little less activity at 30% and 43%, respectively.

3.6.3. Myrosinase activity using sinigrin as substrate with addition of ascorbate

Ascorbate is a co-factor in the myrosinase enzyme and can increase the myrosinase activity at certain concentrations. This increase in activity varies with different myrosinases. The TGG4 wild type and all the mutants were tested for myrosinase activity using sinigrin as substrate and different concentrations of ascorbate. The ascorbate solution was prepared right before every assay and samples, sinigrin and ascorbate were pre-heated before the assay. The ascorbate concentrations used were 0.115, 0.46, 1.1, 1.8 and 2.64 mM. It is the glucose released upon sinigrin hydrolysis that is measured at 550nm and converted to mmol glucose/min. This value was converted to percent activity using TGG4 wt as 100% (fig. 24). The values were also converted to percent activity using the mmol glucose/min values received when performing the assay without addition of ascorbate having this basal value as 100% (fig. 25).

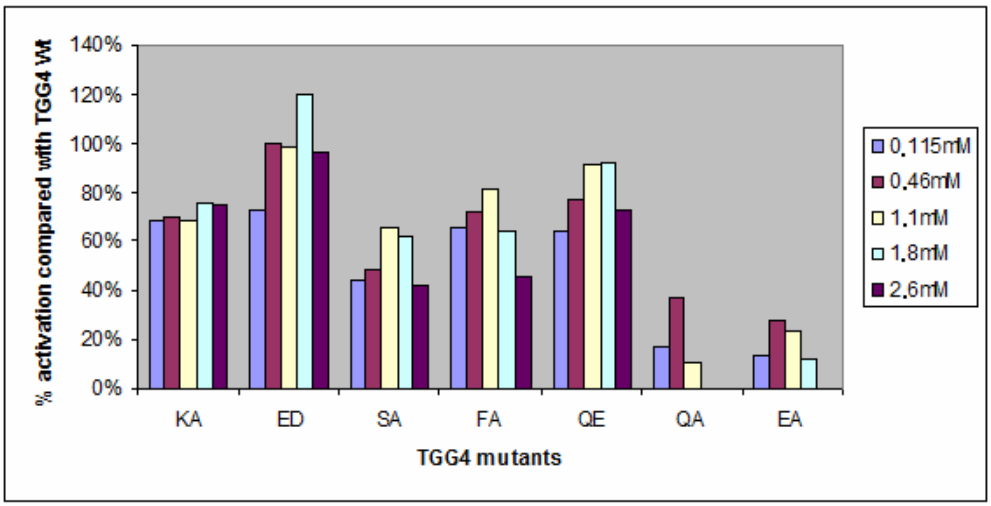


Fig. 24. When looking at the graph it is noticeable that E418D has a higher activity then the wild type even reaching 120% at 1.8mM thus being 20% more active then the wild type. Q418E is another active mutant reaching 91-92% while K218A, S214A and F475A are stuck between 50 and 80%. Q211A and E418A are still the least active mutants only reaching between 10-40% activation.

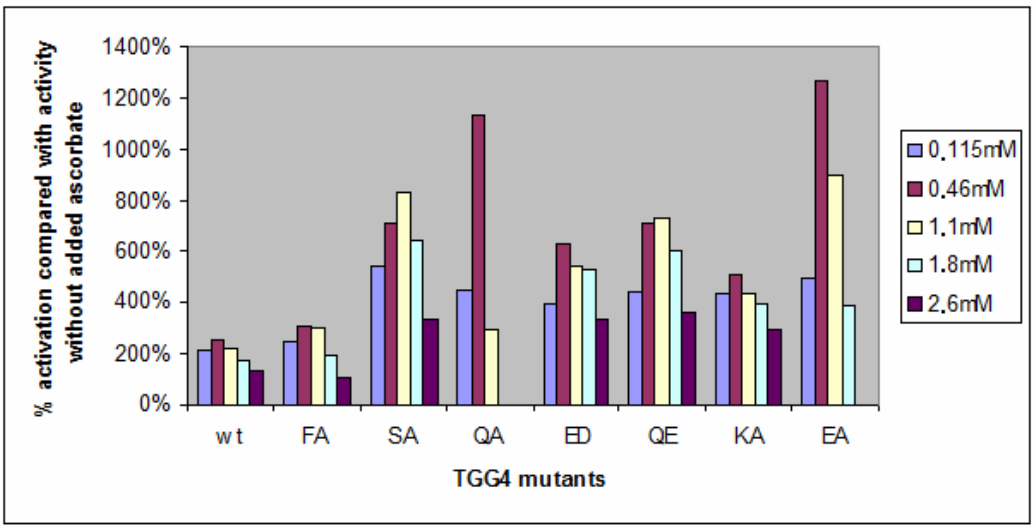


Fig. 25. Another way to measure the activation by ascorbate is to set the activity reached in the absence of ascorbate to 100% and then add the activity gained under the influence of ascorbate. When measured like this it seems like the wild type is the least activated form only reaching 251% activation at its highest point of 0.46 mM ascorbate. The most activated are instead E418A and Q211A both reaching over 1000% .

4. Discussion

4.1. TGG5 crystalization attempt

A *P. pastoris* GS115 strain had been transformed in previous experiments to contain a TGG5 gene expressed behind the AOX promoter known to be one of the strongest promoters in nature (Ozimek et al. 2005). The resulting recombinant strain was grown in BMGY media to gain mass and switched to BMMY media to express the recombinant protein. Being a methylotropic yeast, *P. pastoris* is capable to grow on media containing methanol as the only carbon source. It is important to keep the methanol level at 0.5% and new methanol is thus added on a daily basis to make up for methanol consumed and methanol evaporated. However it is important to not get carried away as even a concentration exceeding 3% might be detrimental (Murray et al. 1989). The recombinant TGG5 protein had a 6xhis tag engineered at its N-terminal making immobilized metal-ion affinity chromatography (IMAC) an ideal purification step. Purification with IMAC is not sufficient to obtain a pure protein since a 30-40 kDa contaminant (later identified as a alcohol dehydrogenase isozyme) co-eluted with TGG5.

Ion-exchange chromatography was used to separate TGG5 and the contaminant but resulted in quite low concentration of the protein, probably due to the buffer exchange step. This was enough to be concentrated into a concentration more suited for crystallization but I also wanted to deglycosylate the TGG5 protein to make it more homogenous and hopefully easier to crystallize. Some protein was lost due to phosphate and zinc precipitating during enzymatic deglycosylation with α-mannosidase and when trying to separate the glycans from TGG5 with ion-exchange chromatography, the rest of the protein was lost.

I made a second attempt to purify TGG5 but was running short on time. Therefore I used TGG5 cell pellets stored at -20°C for purification with IMAC. In contrast to the last time I performed the IMAC step at 4°C instead of room temperature because I was hoping to get a more stable preparation. Although the proteins might have become more stable it seemed like a lot of the TGG5 protein did not bind to the column as seen on SDS-PAGE gels, containing the elution samples, crude extract and flow-through. It might be that the proteins adopt a different structure at 4°C making the his-tag less exposed and thus less likely to bind. To remove the contaminant I used gel filtration chromatography mostly because this chromatography method automatically changes the buffer of the samples. I divided my protein into three samples to be purified individually because I did not want to loose all protein due to any accident. The sample used for the first gel filtration was lost because of an accident when loading the sample into the ÄKTA-system, the second and third attempt were more successful. The sample acquired at the second attempt was used to set up hanging drop for vapour diffusion crystallization. The sample was concentrated to 2.44 mg/ml, a concentration not very ideal for crystallization but I gave it a try anyway. The crystals were set up around conditions that Eva-Lena Andersson used as these resulted in some crystals although not big enough to give a molecular structure of TGG5. The conditions she used were 0.2 M litiumsulphate, 0.1 M phosphate-citrate buffer, pH 4.2 and 20% PEG 1000. Some adjustment had to be made because I was going to include 1:100 α-mannosidase in some drops and 1:100 α-mannosidase + 1:100 EndoH in other drops and from previous experiments it was concluded that zinc falls out in the presence of phosphate, zinc being necesseray for α-mannosidase to reach full activity. Therefore the phosphate-citrate buffer was changed to sodium-citrate buffer. The PEG1000 was also changed to PEG1500 because PEG1000 was not available. No crystals were observed but it seemed like the presence of α-mannosidase did not make the protein precipitate. All drops containing EndoH (except one) precipitated but if this is because TGG5 become unstable due to the loss of the glycan side-chains or the presence of EndoH was not clear. The only drop containing EndoH that did not precipitate was the one containing LiSO4. It is possible that the sulphate group confers some kind of stability to the myrosinase enzyme considering that it binds the sulphate group of its natural substrate (glucosinolates). The TGG5 sample from the third gel filtration was concentrated to 4.81 mg/ml and used to set up a buffer screen (optimum solubility screen) and a screen to test various crystallization variables. For the buffer screen, hanging drops of TGG5 were set up using 1 M buffers with a pH ranging from 3-10. No precipitation was observed suggesting that a wide variety of buffers and pH values can be used when purifying the protein without a major loss due to precipitation. In the JCSGplus screen, TGG5 was mixed with 1:100 α-mannosidase and 1:100 EndoH but no zinc was added. A small crystal was observed in the drop containing calcium chloride, 0.1 M bis-tris, pH 5.5 and 45% MPD. It is uncertain if the crystal is a protein, considering the involvement of calcium chloride it might be a salt.

4.2. TGG5 molecular mass estimation

When expressing TGG5 in *P. pastoris* the protein gets glycosylated with N-linked mannose residues giving the protein a molecular mass different from the 57.469 kDa molecular mass that the 511 amino acids would give. An attempt was made to get an estimate of the molecular mass of the glycosylated TGG5 after treatment with α-mannosidase and after treatment with both α-mannosidase and EndoH. These samples were used for MALDI-TOF mass spectrometry, first as bands separated on a SDS-PAGE minigel and when that did not work, as liquid samples. A molecular mass of 65.4 kDa was estimated for the glycosylated TGG5, although a smaller peak of 63.6 kDa was found in very close proximity to the 65.4 kDa peak. This extra peak might be a differently glycosylated form of TGG5 as two bands have been spotted before when separating TGG5 on SDS-PAGE gels and it is not uncommen that proteins undergo heterogeneous glycosylation. The α-mannosidase treated TGG5 had a molecular mass of 64.4 kDa and thus trimmed away 1 kDa of N-linked glycans from the protein. Only one peak was present and this could mean that either the protein was more homogenously glycosylated or that trimming away 1 kDa of N-linked glycans made the mass spectrometer unable to register it (as it already was a very small peak). No mass spectrum was recorded for the TGG5 protein treated with both α-mannosidase and Endo-H.

4.3. Determination of the identity of the contaminant

When trying to purify the TGG5 protein expressed using the Pichia expression kit a very abundant 38 kDa contaminant was encountered. This contaminant was first presumed to be a degraded form of the AOX gene as it is the protein most abundant when *P. pastoris* is grown on methanol containing media. Although this seemed a bit farfetched when the AOX protein is known to give 72 kDa bands (Ellis et al. 1985) and occurs as a octameric protein (Higgins & Cregg 1998) it is unlikely that all eight subunits would get degraded. Thus after the IMAC purification step I performed a colorimetric AOX assay to confirm whether any alcohol oxidase was present. All fractions from the IMAC purification gave negative results. Earlier I had also performed a western blot analysis using primary anti-his antibodies and found that they bound to the 38 kDa band, thus I formed a new theory that this contaminant was actually a degraded TGG5 subunit containing the his-tag. But an abundant ~ 40 kDa protein has been encountered before when purifying proteins from *P. pastorsis* (Ellis et al. 1985). One report even found out that it had nickel affinity making it stick to IMAC columns (Law et al. 2001). This nickel affinity could explain the difficulties in removing the 38 kDa contaminant using IMAC. When removing contaminants with IMAC it would probably be a good idea to use washing buffer with higher imidazole concentration since the his-tagged protein should bind stronger than the contaminant and be less eluted.
The contaminant was further characterized by running the samples through gel filtration separating TGG5 and the contaminant running the contaminant as shown by SDS-PAGE. The contaminant was then used for mass estimation and peptide mass fingerprinting. The contaminants mass was estimated to 37 kDa correlating well with the molecular mass observed upon SDS-PAGE. I also received three peptide sequences from peptide mass fingerprinting. When running a BLAST on these sequences a 100% correlation to mitochondrial alcohol dehydrogenase isozyme III, a 350 amino acids long protein of *P. pastoris,* was found. An alcohol dehydrogenase was also found by another lab group and they correlated the ubiquitous nature to its role in detoxifying methanol (Law et al. 2001). However, a 40 kDa protein has also been observed when *P. pastoris* was grown on ethanol but it is unknown if it is the same alcohol dehydrogenase and not another isozyme (Ellis et al. 1985).

4.4.1. Mutant myrosinase activity studies

Mutants created in previous experiments were expressed in *P. pastoris* and purified by IMAC, first at room temperature, but when it seemed to make the enzyme unstable I switched to 4°C.
The mutants were tested for enzymatic activity using both sinigrin (S-β-glucoside) and pNPG (O-β-glucoside) as substrate. The mutants were also tested by addition of different concentrations of ascorbate using sinigrin as substrate. The mutation sites where chosen based on conserved residues found in the 1E71 crystal structure of myrosinase from *Sinapsis alba* (Bourderioux et al. 2005; see also alignment in appendix 3)

The Glutamine residue replacing the glutamate normally found in β-glucohydrolases (see appendix 3) is believed to be involved in binding of the sulphate group in glucosinolates as well as ascorbate, it is also believed to be crucial for the positioning of the water molecule allowing hydrolysis of the substrate.

Q211E is a mutation that replaces the glutamine residue for glutamate thus resembling the acid/base complex of O-β-gucosinolates. But it does not seem like having a full acid/base pair gives any advantage as the enzyme activity was only a third when using sinigrin as a substrate and not measurable using pNPG. The loss of glutamine might make it harder for the enzyme to position the water molecule accurately. Another possibility is that the myrosinase has problem binding the substrate with glutamate due to unfavourable charges (both glutamate and the sulphate group are negatively charged). Some of the activity was restored when ascorbate was added (80% of wt activity) maybe because the ascorbate molecule helped with the positioning of the water molecule or binding of the substrate.

An Q211A mutant was also tested, the alanine mutation being a way to replace an important residue with a small and uncharged residue. The activity assay using sinigrin gave as expected very low activity, only a tenth of the wild type activity. A higher activity was reached when pNPG was used as substrate, giving 64% activity. When having ascorbate added, the Q211A mutant reached only 37% of the full activity (100% being the activity the wild type had at the same ascorbate concentration). This suggests that the glutamine residue is very important for the myrosinase to function properly but I do not know if it is the reduced ability to bind the SO4/ascorbate or the positioning of the water molecule that is responsible for the low activity of the mutants. I can not explain why it had higher activity when pNPG was used as a substrate either.

E418Q is a mutation that replaces the acid residue of the catalytic site to glutamine. When testing the activity using sinigrin as substrate, the activity was halved, an expected result as glutamate is involved in the catalytic activity. An unexpected result when using pNPG (O-β-glucoside) was that the activity was almost threefold the activity of the wild type. If you compare this with the Q211E mutant, where a glutamine was changed to glutamate, and the absence of activity that followed, it almost seems like glutamate makes it harder for the myrosinase to hydrolyse O-β-glucosides. It is possible that glutamate makes it harder for the protein to position the water molecule for hydrolysis. Another explanation is that glutamate hinders binding of the substrate, either sterically or by its negative charge. I was not able to perform an ascorbate assay on this mutant, therefore no data for its interaction with ascorbate is available.

Aspartic acid and histidine are much like glutamic acid, common amino acids in active sites much due to their ability to change their charge depending on the environment and act as either proton donors or proton acceptors. Glycosyl hydrolases using aspartic acid as proton donors have been encountered although the same can not be said about histidine. Nevertheless, two mutants were tested, E418D and

E418H, replacing the glutamic acid with either aspartate or histidine. When testing these two mutants for myrosinase activity using sinigrin as substrate they both gave an activity around 40%. This might mean that even if histidine is not a natural residue in glycosyl hydrolases it does confer some activity.

The E418H mutant was purified by IMAC at room temperature. This made the mutant very unstable and it lost all its activity within one day making any further tests impossible. The E418D mutant was tested for activity using pNPG as a substrate, but much like the Q211E mutant no activity was measured. Thus it might be that the negatively charged side chain of the glutamate and aspartate residue is interfering with binding of the substrate. The E418D mutant was tested for enzymatic activity under addition of ascorbate using sinigrin as substrate. Surprisingly enough the enzymatic activity was higher than for the wild type and at some point even gave 20% higher activity than the wild type.

E418A is a mutation that replaces the proton donor with the non-reactive alanine. Both the assay using sinigrin and sinigrin with ascorbate gave low activity. If you compare it with the results from the E418D mutant it seems like ascorbate can restore the activity as long as there is an available proton donor. The E418A mutant had some activity using pNPG as a substrate and I guess the aspartate of the E418D mutant made it unable to use pNPG as substrate.

The Serine 214 residue is believed to bind the sulphate group of the glucosinolate and establishing hydrogen bonds together with other residues (e.g. Gln211 and Lys218). The S214A mutant had very low activity (17% of normal wt activity) using sinigrin as substrate and 43% of wt activity using pNPG as substrate. This is probably due to an inability to bind the SO_4 group and it might also affect the ability of other residues to bind the sulphate group.

The lysine residue corresponding to the Arg 194 residue in 1E71 is believed to have a weak interaction with the SO_4 group (through a positive charge). The K218A mutant had 35% of wt activity using sinigrin and 66% using pNPG. I guess that because the interaction is believed to be weak it does not affect the binding of SO_4 as strongly as the S214A mutation. The lysine residue is believed to only play a weak electrostatic role when binding to ascorbate but still the K218A mutant loses ~ 30% activity when using sinigrin as substrate with addition of ascorbate.

Phe 475 is located in the wall of the glucosyl binding site nearest to the 3-hydroxyl of the glucose residue. The F475A mutant had 58% of wt activity using sinigrin as substrate and 60% using pNPG. In presence of ascorbate it had lost ~ 30% activity. One edge of the phenyl ring of Phe 475 is in contact with the 3-hydroxyl of the glucose moiety, while the rest of the sidechain is buried in the wall of the glucosyl binding site where it fits snugly. When mutated to alanine, the sidechain will not fill up the hole and may not be as firmly fixed in position. This may destabilise the polypeptide chain in this region and affect adjacent amino acids like the preceding residue, Glu 466, that is helping to position the glucosyl unit by hydrogen bonding to its 6- and 4-hydroxyls. As a consequence the mutation is likely to primarly affect the binding of the glucosyl unit of the substrate. Ascorbate binds > 9 Å away from the Phe 475, but could be affected by more distant effects of structural perturbations caused by the mutation.

The isoleucine 278 residue forms a hydrophobic surface together with tyrosine and phenylalanine that interacts with the SO_4 group. The I278 mutant was purified by IMAC at room temperature and was therefore very unstable, I managed to get a reading of 3% activity compared to wt but I do not know if this low activity is due to the mutation or the instability of the myrosinase. Ile 278 is also believed to interact with the ascorbate co-factor but the unstable mutant could not be tested for myrosinase activity under addition of ascorbate.

4.4.2 Ascorbate vs sinigrin

The sinigrin was always used at 12.5 mM while the ascorbate concentrations used were 0.115, 0.46, 1.1, 1.8 and 2.6 mM, respectively. From the experiments it seemed like the 0.46 mM ascorbate concentration gave the highest catalytic activity for the K218A, E418D, Q211A, E418A and wild type. The F475A mutant showed a similar activation as for the Q418E mutant at both 0.46 and 1.1 mM ascorbate. The S214A differed from the other mutants in that its highest activation was reached at 1.1 mM ascorbate. The Q211A and E418A mutants both lost activity quickly when higher concentrations of ascorbate was added, with Q211A losing all activity at 1.8 mM and E418A at 2.6 mM. The wild type differed from most mutants in that it reached a high activation at 0.115 mM ascorbate and only a slight increase in catalytic activity at 0.46 mM. The highly activated E418D was similar to Q211A and E418A in that its catalytic activity increased greatly between 0.115 mM and 0.46 mM. The K218A on the other hand differed in that it lost less catalytic activity when higher ascorbate concentration was used (appendix 4).

5. Conclusions

The crystallization was not very successful, but for further crystallization attempts it can be noted that gel filtration similar to ion exchange chromatography, gives a good separation of TGG5 and the contaminant and makes buffer exchange unnecessary. It also turned out that TGG5 is stable in a broad spectrum of buffers. Perhaps two-dimensional polyacrylamide gel electrophoresis could be performed to assure that only TGG5 remains after purification. The conditions yielding the small crystal should be used for further protein crystallisation attempts, assuming the crystal is a genuine protein crystal and not a salt crystal. The crystal is too small for x-ray diffraction experiments on the in-house x-ray diffractometer at BMC. It may hopefully be possible to test the crystal at a powerful synchrotron radiation source (e.g. ESRF, Grenoble, France) in a near future.When crystallizing the protein, LiSO4 could be added to avoid TGG5 precipitation as it seemed to have a stabilizing effect, if another attempt at mass estimation is made ion-spray mass spectrometry could be an alternative to MALDI-TOF as it may result in less noise. From the mutant studies it was observed that most of the mutants were capable of hydrolysing pNPG although not very efficient. The only exceptions were Q211E and E418D that were incapable to hydrolyse any pNPG. Among the mutants, Q211A, E418A and S214A had the greatest reduction in catalytic activity,.Q211A and E418A are already known to be important residues based on the 1MYR structure but it seems like S214A and its proposed sulphate-group binding ability is also quite important. The ascorbate co-factor proved to be very effective in increasing the catalytic activity of the mutants with the E418D mutant even exceeding the activity of the wild type. Some of the mutants seemed to have acquired a changed stimulating-inhibiting reaction to ascorbate compared to the wild type. This changed inhibiting-stimulating relation could be further tested by using other ascorbate concentrations and ascorbate analogs. The mutant activity studies should be repeated to make sure the result is reproducible and also to obtain the V_{max} and K_m values. The IMAC procedure should be performed at 4°C as room temperature seemed to be detrimental to the mutant protein. By increasing the imidazole concentration in the washing buffer (e. g. 100 mM) an attempt to remove the contaminant could be made making further purification steps unnecessary. Gel filtration and ion-exchange could further purify the protein, allowing observation of specific activity instead of total activity, although there is a risk that the mutant protein will not survive the purification.

Acknowledgements

I would like to thank Johan Meijer for giving me the opportunity to perform my master thesis at his lab. I would also like to thank my co-supervisors Jerry Ståhlberg and Sarosh Bejai for all their assistance and Åke Engström for the mass spectrometry.

6. References

Aires A, Mota VR, Saavedra MJ, Monteiro AA, Simões M, Rosa EAS and Bennet RN. 2009. *Initial in vitro* evaluations of the antibacterial activities of glucosinolate enzymatic hydrolysis products against plant pathogenic bacteria. Journal of Applied Microbiology 106: 2096-2105

Barth C and Jander G. 2006. Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. The Plant Journal 46: 549-562

Blancher C and Jones A. 2008. SDS-PAGE and western blotting techniques. Methods in Molecular Medicine 57: 145-162

Bollag DM. 1994.Gel-filtration chromatography. Methods in Molecular Biology 36: 1-9

Bones AM and Rossiter JT. 2006. The enzymic and chemically induced decomposition of glucosinolates. Phytochemistry 67: 10053-1067

Bourderioux A, Lefoix M, Gueyrard D, Tatibouët A, Cottaz S, Artz S, Burmeister WP and Rollin P. 2005.The glucosinolate-myrosinase system. New insights into enzyme-substrate interactions by use of simplified inhibitors. Organic & Biomolecular Chemistry 3: 1872-1879

Burmeister WP, Cottaz S, Driguez H, lori R, Palmieri S and Henrissat B. 1997. The crystal structures of *Sinapsis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. Structure 5:663-675

Burmeister WP, Cottaz S, Rollin P, Vasella A and Henrissat B. 2000. High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base. The Journal of Biological Chemistry 275:39385-39393

Charlton A and Zachariou M. 2007. Immobilized meal ion affinity chromatography of histidine-tagged fusion proteins. Methods in Molecular Biology 421:137-150

Cottrell JS and Sutton CW. 1996. The identification of electrophoretically separated proteins by peptide mass fingerprinting. Methods in Molecular Biology 61:67-82

Daly R and Milton Hearn MTW. 2005. Expression of heterologous proteins in Pichia pastoris: a useful experimental tool in protein engineering and production. Journal of Molecular Recognition 18:119-138

Davies G and Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. Structure 3:853-859

Ellis SB, Brust EPF, Koutz PJ, Waters AF, Harpold MM and Gingeras TR. 1985. Isolation of alcohol oxidase and two other methanol regulatable genes from the Yeast *Pichia pastoris*. Molecular and Cellular Biology 5:1111-1121

Gevaert K and Vandekerckhove J. 2005. Protein characterization: Analytical approaches and applications to proteomics. In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd: Chichester http://www.els.net/, doi: 10.1038/npg.els.0005725

Gregg JM, Madden KR, Barringer KJ, Thill GP and Stillman CA. 1989. Functional Characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Molecular and Cellular Biology 3: 1316-1323

Guarna MM, Lesnicki GJ, Tam BM, Robinson J, Radziminski CZ, Hasenwinkle D, Boraston A, Jervis E, MacGillivray RTA, Turner RFB and Kilburn DG. 1997. On-line monitoring and control of methanol concentration in shake-flask cultures of pichia pastoris. Biotechnology and Bioengineering 3:279-286

Harvey DJ. 2003. Identification of sites of glycosylation. Methods in Molecular Biology 211: 371-383

Henrissat B and Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases.Current Opinion in Structural Biology 7:637-644

Higgins DR. 1995. Overview of protein expression in *Pichia pastoris.* Current Protocols in Protein Science 5.7.1 – 5.7.18

Higgins DR and Cregg JM. 1998. Introduction to *Pichia pastoris.* Methods in Molecular Biology 103, pichia protocols, ISBN 0-89603-421-6

Husebye H, Artzt S, Burmeister WP, Härtel FV, Brandt A., Rossiter JT and Bones AM. 2005. Crystal structure at 1.1 Å resolution of an insect myrosinase from *Brevicoryne brassicae* shows its close relathionship to β-glucosidases. Insect Biochemistry and Molecular Biology 35: 1311-1320

Inan M and Meagher MM. 2001. Non-repressing carbon sources for Alcohol oxidase (AOX1) promoter of *Pichia pastoris.* Journal of Bioscience and Bioengineering 6:585-589

Kleinwächter M and Selmar D. 2004. A novel approach for reliable activity determination of ascorbic acid depending myrosinases. Journal of Biochemical and Biophysical Methods 59: 253 – 265

Kliebenstein DJ, Kroymann J and Mitchell-Olds T. 2005. The glucosinolate-myrosinase system in an ecological and evolutionary context. Current Opinion in Plant Biology 8:264-271

Kruger NJ. 1996. The Bradford method for protein quantitation. The Protein Protocols Handbook. Doi:10.1007/978-1-60327-259-9_4

Law RHP, Robinson HC, Rowly MJ and Mackay R. 2001. The ubiquitous 38 kDa contaminant in glutamic acid decarboxylase preparation from the cytosol of *Pichia pastoris* after immobilised metal ion affinity chromatography is an alcohol dehydrogenase. Biothechnology Letters 23: 697-703

Mithen R. 2001a. Glucosinolates and their degradation products. Advances in Botanical Research 35: 213 – 232

Mithen R. 2001b. Glucosinolates – biochemistry, genetics and biological activity. Plant Growth Regulation 34: 91-103

Murray WD, Duff SJB, and Lanthier PH. 1989. Induction and stability of alcohol oxidase in the methylotrophic yeast *Pichia pastoris.* Applied Microbiology and Biotechnology 32: 95-100

Nøhr J., Kristiansen K and Krogsdam AM. 2003. Protein expression in yeasts. Methods in Molecular Biology 232: 111-125

Ozimek P, Veenhuis M and Van der Klie IJ. 2005. Alcohol oxidase A complex peroxisomal , oligomeric flavoprotein. FEMS Yeast Research 5: 975-983

Rask L, Andreasson E, Ekbom B, Eriksson S, Pontoppidan B and Meijer J. 2000. Myrosinase: gene family evolution and herbivore defense in *Brassicacea.* Plant Molecular Biology 42:93-113

Roepstorff P. 2005. Mass spectrometry instrumentation in proteomics. In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd: Chichester http://www.els.net/, doi: 10.1038/npg.els.0006194

Simonian MH. 2002. Spectrophotometric determination of protein concentration. Current Protocols in Cell Biology. Doi:10.1002/0472913.fab0103s04

Simonian MH and Smith JA. 2006. Spectrophotometric and colorimetric determination of protein concentration. Current Protocols in Molecular Biology. doi:10.1002/0471142727.mb1001as76

Smith BJ. 1994. SDS-polyacrylamide gel electrophoresis of proteins. Methods in Molecular Biology 32: 23 – 34

Stanton P. 2003.Gel filtration chromatography. Methods in Molecular Biology 251:55-73

Van den Boom D and Ehrich M. 2007. Discovery and identification of sequence polymorphisms and mutations with MALDI-TOF MS. Methods in Molecular Biology 366:287-306

Walker JM. 1986. Protein (western) blotting with immunodetection. In experiments in Molecular Biology. Biological Methods. Ed: Robert J. Slater. Humana Press, Clifton (New Jersey) doi10.1007/978-1-60327-405-0_17

Windsor AJ, Reichelt M, Figuth A, Svatoš A, Kroymann J, Kliebenstein DJ, Gershenzon J and Mitchell-Olds T.2005.Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). Phytochemistry 66: 1321-1333

Wittstock U and Halkier BA. 2002. Glucosinolate research in the Arabidopsis era. Trends in Plant Science 6:263-270

Yates JR. 2001. Mass spectrometry in biology. In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd: Chichester http://www.els.net/, doi: 10.1038/npg.els.0000999

Zhang J, Pontoppidan B, Xue J, Rask L and Meijer J. 2001. The third myrosinase gene TGG3 in *Arabidopsis thaliana* is a psuedogene specifically expressed in stamen and petal. Physiologia Plantarium 115: 25-34

Web adresses and product manuals

Source 30Q and 30S instructions, 18-1106-03 AD, GE Healthcare

Endoglycosidase H from *Streptococcus plicatus*, recombinant expressed in *E. coli.* Product information, SIGMA Chemical Company.

α-Mannosidase from *Canavalia ensiformis* (jack bean) proteomics grade, SIGMA-Aldrich

Principles and applications of high-performance ion-exchange chromatography for bioseparations
VHP high performance ion-exchange columns, VYDAC

Ion exchange chromatography principles and methods, Amersham biosciences

Glucose (Gluc-PAP) GOD/PAP (liquid) manual, RANDOX

www.uniprot.org/uniprot/Q9URI7

www.uniprot.org/uniprot/Q9URI8

APPENDIX 1

BLAST results of contaminant protein partial sequences

```
>emb|CAY69102.1|  Mitochondrial alcohol dehydrogenase isozyme III [Pichia pastoris
GS115]
Length=350

 Score = 45.6 bits (100),  Expect = 6e-04
 Identities = 13/13 (100%), Positives = 13/13 (100%), Gaps = 0/13 (0%)

Query  1    VYEQMEAGAIIGR  13
            VYEQMEAGAIIGR
Sbjct  331  VYEQMEAGAIIGR  343


>emb|CAY69102.1|  Mitochondrial alcohol dehydrogenase isozyme III [Pichia pastoris
GS115]
Length=350

 Score = 76.6 bits (173),  Expect = 3e-13
 Identities = 24/25 (96%), Positives = 25/25 (100%), Gaps = 0/25 (0%)

Query  1    ADISGFTHDGSFQQYATADATQAAR  25
            AD+SGFTHDGSFQQYATADATQAAR
Sbjct  117  ADLSGFTHDGSFQQYATADATQAAR  141


>emb|CAY69102.1|  Mitochondrial alcohol dehydrogenase isozyme III [Pichia pastoris
GS115]
Length=350

 Score = 35.4 bits (76),  Expect = 0.68
 Identities = 10/10 (100%), Positives = 10/10 (100%), Gaps = 0/10 (0%)

Query  1    AINQSVQYVR  10
            AINQSVQYVR
Sbjct  254  AINQSVQYVR  263
```

APPENDIX 2

TGG4 mutations based on the 1E71 crystal structure from *Sinapsis alba*

```
TGG4            MAIPKAHYSLAVLVLLFVVVSSSQKVCNPECKAKEPFHCDNTHAFNRTGFPRNFTFGAAT 60
TGG5            MAIPKAHYSLAVLVLLFVVVSSSQKVCNPECKAKEPFHCDNTHAFNRSGFPKNFTFGAAT 60
1E71            ----------------------DEEIT---CQENLPFTCGNTDALNSSSFSSDFIFGVAS 35
                                     .:::    *: : ** *.**.*:* :.*. :* **.*:

TGG4            SAYQIEGAAHRALNGWDYFTHRYPEK-VPDRSSGDLACDSYDLYKDDVKLLKRMNVQAYR 119
TGG5            SAYQIEGAAHRALNGWDYFTHRYPEK-VPDRSSADLACDSYDLYKDDVKLLKRMNVQAYR 119
1E71            SAYQIEGTIGRGLNIWDGFTHRYPNKSGPDHGNGDTTCDSFSYWQKDIDVLDELNATGYR 95
                *******:  *.** ** ******:*  **:...* :***:. ::.*:.:*..:*. .**

TGG4            LSIAWSRVLPKGRLTGGVDENGITYYNNLINELKANGIEPYVTIFHWDVPQTLEDEYGGF 179
TGG5            LSIAWSRVLPKGRLTGGVDENGITYYNNLINELKANGIEPYVTIFHWDVPQTLEDEYGGF 179
1E71            FSIAWSRIIPRGKRSRGVNEKGIDYYHGLISGLIKKGITPFVTLFHWDLPQTLQDEYEGF 155
                :******::*:*: : **:*:** **:.**. *  :** *:**:****:****:*** **

TGG4            LSTRIVEDYTNYAELLFQRFGDRVKFWITLNQPFSLATKGYGDGSYPPGRCTG----CEL 235
TGG5            LSTRIVEDYTNYAELLFQRFGDRVKFWITLNQPLSLALKGYGNGSYPPGRCTG----CEL 235
1E71            LDPQIIDDFKDYADLCFEEFGDSVKYWLTINQLYSVPTRGYGSALDAPGRCSPTVDPSCY 215
                *..:*::*:.:**:* *:.*** **:*:*:**  *:. :***..  .****:     .

TGG4            GGDSGVEPYTVAHNQLLAHAKTVSLYRKRYQKFQGGKIGTTLIGRWFAPLNEFSELDKAA 295
TGG5            GGDSGVEPYTVAHNQLLAHAKTVSLYRKRYQKFQGGKIGTTLIGRWFVPLNEFSELDKAA 295
1E71            AGNSSTEPYIVAHHQLLAHAKVVDLYRKNYT-HQGGKIGPTMITRWFLPYNDTDRHSIAA 274
                .*:*..*** ***:*******.*.****.*  .******.*:* *** * *: .. . **

TGG4            AKRAFDFFVGWFLDPLVYGKYPTIMREMVGDRLPEFTPEQSALVKGSLDFLGLNYYVTQY 355
TGG5            AKRAFDFFVGWFLDPLVYGKYPTIMREMVGDRLPEFTPEESALVKGSLDFLGLNYYVSQY 355
1E71            TERMKEFFLGWFMGPLTNGTYPQIMIDTVGERLPSFSPEESNLVKGSYDFLGLNYYFTQY 334
                ::*  :**:***:.**. *.** ** : **:***.*:**:* ***** ********.:**

TGG4            ATDAPPPTQLN---AITDARVTLGFYRN-GVPIGVVAP--------SFVYYPPGFRQILN 403
TGG5            ATDAPPPTQPN---AITDARVTLGFYRN-GSPIGVVAS--------SFVYYPPGFRQILN 403
1E71            AQPSPNPVNSTNHTAMMDAGAKLTYINASGHYIGPLFEKDKADSTDNIYYYPKGIYSVMD 394
                *  :* *.: .   *: ** ..* : .  *  ** :          .: *** *: .:::

TGG4            YIKDNYKNPLTYITENGVADLDLGNVTLATALADNGRIQNHCSHLSCLKCAMKDG-CNVA 462
TGG5            YIKDNYKNPLTYITENGVADLDLGNVTLATALADNGRIQNHCSHLSCLKCAMKDG-CNVA 462
1E71            YFKNKYYNPLIYVTENGISTP--GDENRNQSMLDYTRIDYLCSHLCFLNKVIKEKDVNVK 452
                *:*::* *** *:****::    *: .   :: *  **:  ****. *: .:*:   **

TGG4            GYFAWSLMDNYEFGNGYTLRFGMNWVNFTNPADRKEKASGKWFSKFLAK 511
TGG5            GYFAWSLMDNYEFGNGYTLRFGMNWVNFTNPADRKEKASGKWFSKFLAK 511
1E71            GYLAWALGDNYEFNKGFTVRFGLSYIDWNNVTDRDLKKSGQWYQSFISP 501
                **:**:* *****.:*:*:***:.::::.* :**. * **:*:..*::
```

Glutamine mutations: Q211E, Q211A (Gln 187 in Sinapsis alba 1E71)
Active site mutations: E418Q, E418D, E418H, E418A (Glu 409)
binding of hydrophobic part of Aglycone (Allyl, Benzyl or indole group): F475A (Phe 473)
Ascorbate/SO_4 binding: S214A (Ser190) K218A (Arg194), I278A (Ile 257)

APPENDIX 3

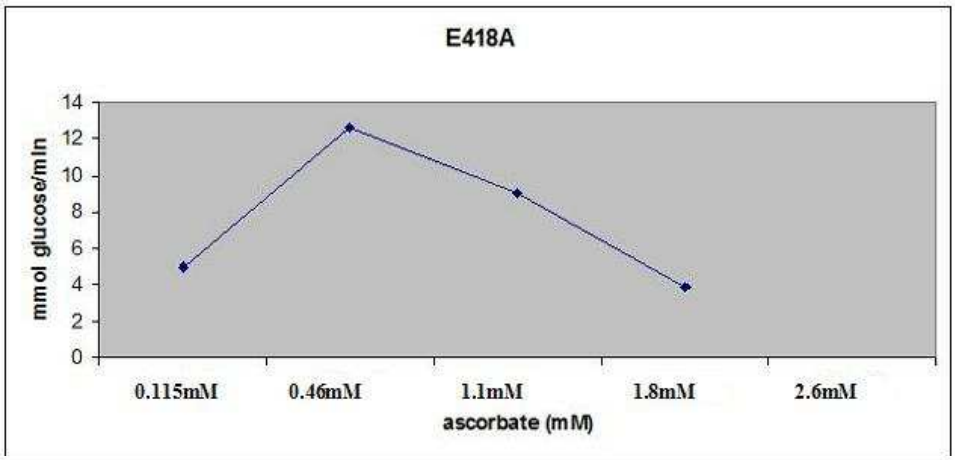
Amino acid sequence comparison of plant glycosidases

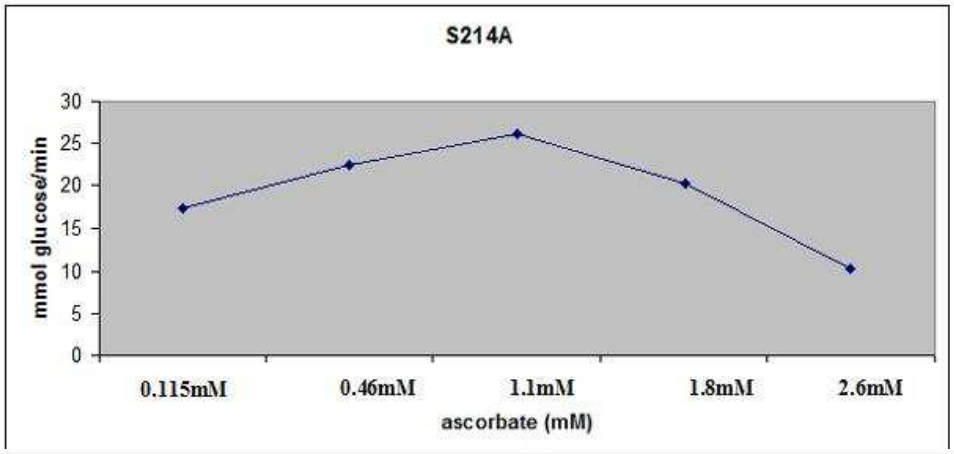
```
TGG4            MAIPKAHYSLAVLVLLFVVVSSSQKVCNPECKAKEPFHCDNTHAFNRTGFPRNFTFGAAT 60
TGG5            MAIPKAHYSLAVLVLLFVVVSSSQKVCNPECKAKEPFHCDNTHAFNRSGFPKNFTFGAAT 60
1CBG            ------------------------------FKPLPISFDDFSDLNRSCFAPGFVFGTAS 29
1E71            -------------------------DEEITCQENLPFTCGNTDALNSSSFSSDFIFGVAS 35
                                          *:  .:   :* : *. .* **.*:

TGG4            SAYQIEGAAHRALNG---WDYFTHRYPEKVP-DRSSGDLACDSYDLYKDDVKLLKRMNVQ 116
TGG5            SAYQIEGAAHRALNG---WDYFTHRYPEKVP-DRSSADLACDSYDLYKDDVKLLKRMNVQ 116
1CBG            SAFQYEGAAFEDGKGPSIWDTFTHKYPEKIK-DRTNGDVAIDEYHRYKEDIGIMKDMNLD 88
1E71            SAYQIEGTIGRGLNI---WDGFTHRYPNKSGPDHGNGDTTCDSFSYWQKDIDVLDELNAT 92
                **:* **:  .  :    ** ***:**:*   *: ..* : *.:  ::.*: ::. :*

TGG4            AYRLSIAWSRVLPKGRLTGGVDENGITYYNNLINELKANGIEPYVTIFHWDVPQTLEDEY 176
TGG5            AYRLSIAWSRVLPKGRLTGGVDENGITYYNNLINELKANGIEPYVTIFHWDVPQTLEDEY 176
1CBG            AYRFSISWPRVLPKGKLSGGVNREGINYYNNLINEVLANGMQPYVTLFHWDVPQALEDEY 148
1E71            GYRFSIAWSRIIPRGKRSRGVNEKGIDYYHGLISGLIKKGITPFVTLFHWDLPQTLQDEY 152
                .**:**:*.*::*:*: : **:.:** **:.**. :  :*: *:**:****:**:*:***

TGG4            GGFLSTRIVEDYTNYAELLFQRFGDRVKFWITLNQPFSLATKGYGDGSYPPGRCTG---- 232
TGG5            GGFLSTRIVEDYTNYAELLFQRFGDRVKFWITLNQPLSLALKGYGNGSYPPGRCTG---- 232
1CBG            RGFLGRNIVDDFRDYAELCFKEFGDRVKHWITLNEPWGVSMNAYAYGTFAPGRCSDWLKL 208
1E71            EGFLDPQIIDDFKDYADLCFEEFGDSVKYWLTINQLYSVPTRGYGSALDAPGRCSPTVDP 212
                 ***. .*::*: :**:* *:.*** **.*:*:*:  .:. ..*. .  .****:

TGG4            CELGGDSGVEPYTVAHNQLLAHAKTVSLYRKRYQKFQGGKIGTTLIGRWFAPLNEFSELD 292
TGG5            CELGGDSGVEPYTVAHNQLLAHAKTVSLYRKRYQKFQGGKIGTTLIGRWFVPLNEFSELD 292
1CBG            NCTGGDSGREPYLAAHYQLLAHAAAARLYKTKYQASQNGIIGITLVSHWFEPASK-EKAD 267
1E71            SCYAGNSSTEPYIVAHHQLLAHAKVVDLYRKNYTH-QGGKIGPTMITRWFLPYNDTDRHS 271
                   .*:*. *** .** ****** .. **:..*   *.* ** *:: :** * .. .. .

TGG4            KAAAKRAFDFFVGWFLDPLVYGKYPTIMREMVGDRLPEFTPEQSALVKGSLDFLGLNYYV 352
TGG5            KAAAKRAFDFFVGWFLDPLVYGKYPTIMREMVGDRLPEFTPEESALVKGSLDFLGLNYYV 352
1CBG            VDAAKRGLDFMLGWFMHPLTKGRYPESMRYLVRKRLPKFSTEESKELTGSFDFLGLNYYS 327
1E71            IAATERMKEFFLGWFMGPLTNGTYPQIMIDTVGERLPSFSPEESNLVKGSYDFLGLNYYF 331
                  *::*  :*::***: **. * **  *   * .***.*:.*:*  :.** ********

TGG4            TQYATDAP---PPTQLNAITDARVTLGFYRN-GVPIGVVAPS--------FVYYPPGFRQ 400
TGG5            SQYATDAP---PPTQPNAITDARVTLGFYRN-GSPIGVVASS--------FVYYPPGFRQ 400
1CBG            SYYAAKAPR-IPNARPAIQTDSLINATFEHN-GKPLGPMAASS------WLCIYPQGIRK 379
1E71            TQYAQPSPNPVNSTNHTAMMDAGAKLTYINASGHYIGPLFEKDKADSTDNIYYYPKGIYS 391
                : **  :*     :.     *:  .  : .  *  :* :  .        :  ** *: .

TGG4            ILNYIKDNYKNPLTYITENGVADLDLGNVTLATALADNGRIQNHCSHLSCLKCAMKDG-C 459
TGG5            ILNYIKDNYKNPLTYITENGVADLDLGNVTLATALADNGRIQNHCSHLSCLKCAMKDG-C 459
1CBG            LLLYVKNHYNNPVIYITENGRNEFNDPTLSLQESLLDTPRIDYYYRHLYYVLTAIGDG-V 438
1E71            VMDYFKNKYYNPLIYVTENGIS--TPGDENRNQSMLDYTRIDYLCSHLCFLNKVIKEKDV 449
                :: *.*::* **: *:****          .   :: *  **:    **  :  .: :

TGG4            NVAGYFAWSLMDNYEFGNGYTLRFGMNWVNFTNPADRKEKASGKWFSKFLAK 511
TGG5            NVAGYFAWSLMDNYEFGNGYTLRFGMNWVNFTNPADRKEKASGKWFSKFLAK 511
1CBG            NVKGYFAWSLFDNMEWDSGYTVRFGLVFVDFKNNLKRHPKLSAHWFKSFLKK 490
1E71            NVKGYLAWALGDNYEFNKGFTVRFGLSYIDWNNVTDRDLKKSGQWYQSFISP 501
                ** **:**:* ** *:..*:*:***: ::::.*  .*. * *.:*:..*:
```

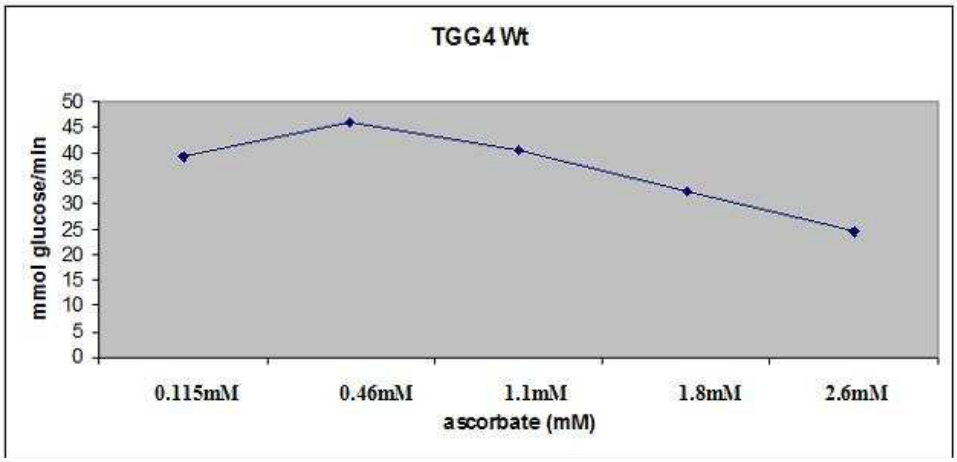
Active site
Glutamine
Substrate binding

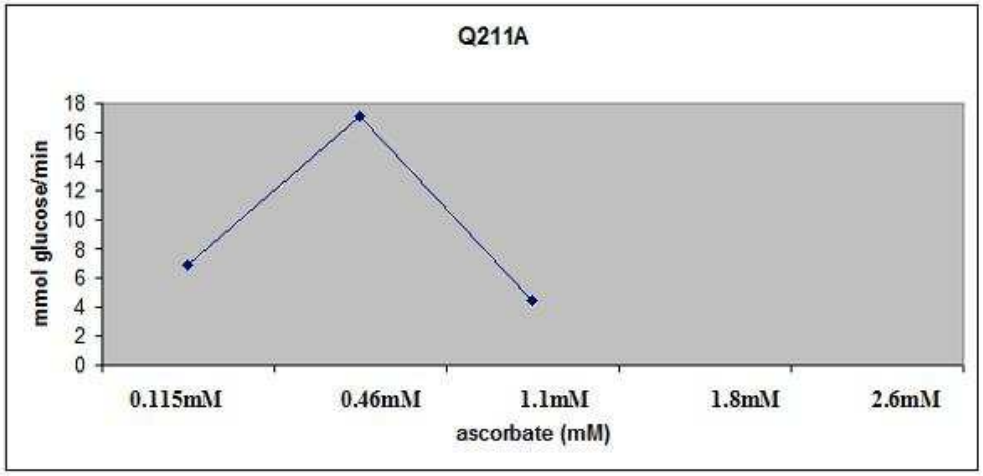
APPENDIX 4

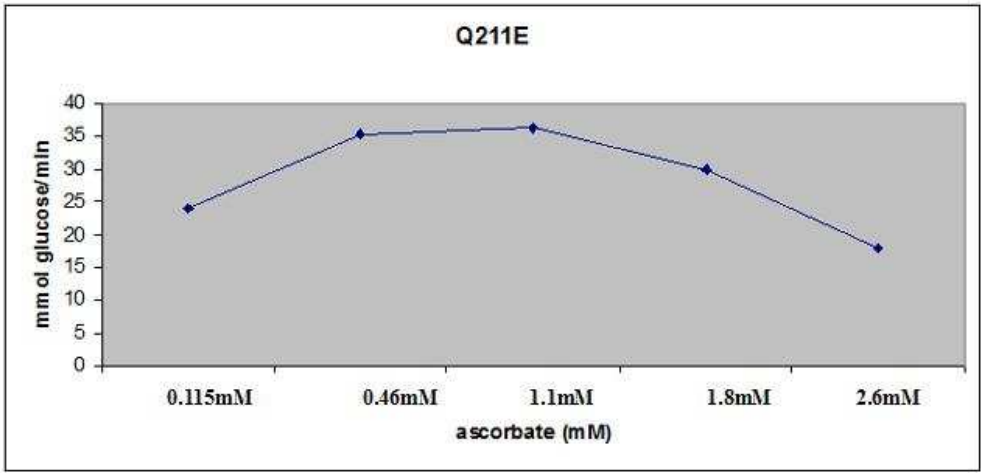
Graphs displaying enzyme activity towards sinigrin as substrate with different concentrations of ascorbate

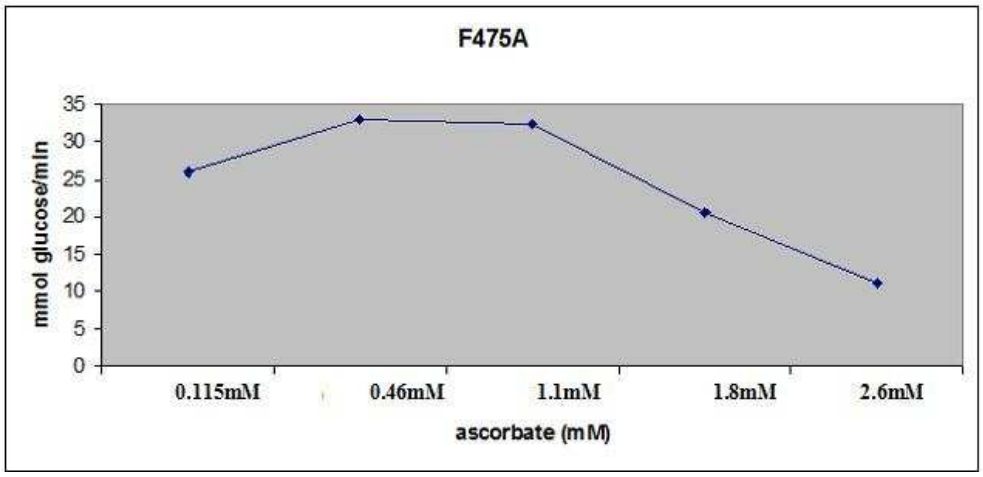


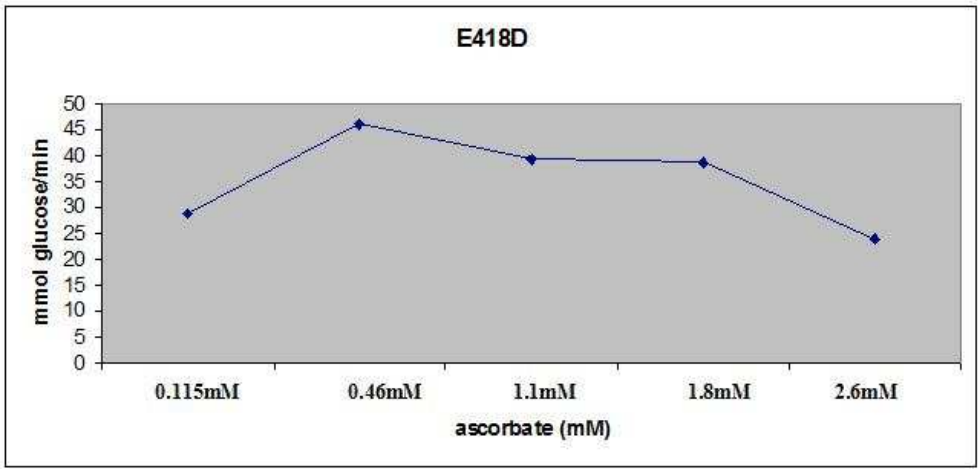


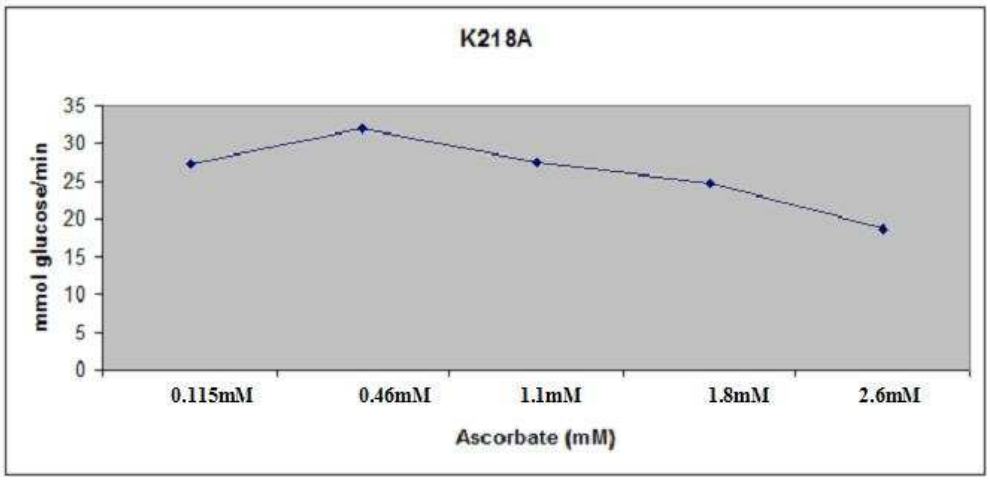





Q211E
mmol glucose/min
40
35
30
25
20
15
10
5
0
0.115mM
0.46mM
1.1mM
1.8mM
2.6mM
ascorbate (mM)


F475A
mmol glucose/min
35
30
25
20
15
10
5
0
0.115mM
0.46mM
1.1mM
1.8mM
2.6mM
ascorbate (mM)


E418D
mmol glucose/min
50
45
40
35
30
25
20
15
10
5
0
0.115mM
0.46mM
1.1mM
1.8mM
2.6mM
ascorbate (mM)


K218A
mmol glucose/min
35
30
25
20
15
10
5
0
0.115mM
0.46mM
1.1mM
1.8mM
2.6mM
Ascorbate (mM)