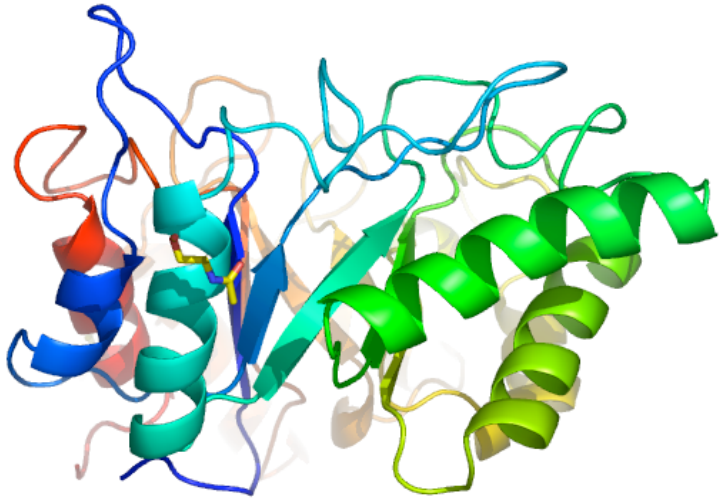
EX0565 Independent project in biology E, 30 HEC
Master thesis for the Master Programme in Biotechnology
at the Swedish University of Agricultural Science

Structure-Function Studies of Biomass Degrading Enzymes

— Crystallization of EndoT from *Hypocrea jecorina* and Cel5H from *Saccharophagus degradans*, and Enzyme Characterization of Cel5H



Jing Zhang

Department of Molecular Biology
Swedish University of Agricultural Sciences
Uppsala, 2009

Sveriges lantbruksuniversitet

MicroDrivE
Microbially Derived Energy

EX0565, Självständigt arbete i biologi, 30 hp, Avancerad nivå E

Structure-Function Studies of Biomass Degrading Enzymes
— Crystallization of EndoT from *Hypocrea jecorina* and Cel5H from *Saccharophagus degradans*, and Enzyme Characterization of Cel5H

Jing Zhang

Supervisors: Mats Sandgren, Jerry Ståhlberg, Saeid Karkehabadi, Dept of Molecular Biology, Swedish University of Agricultural Sciences (SLU) Sweden;

Examiner: Stefan Knight, Dept Molecular Biology, SLU

Institutionen för molekylärbiologi
SLU, Sveriges lantbruksuniversitet
Box 590
751 24 Uppsala

Examensarbete för Masterutbildning i Bioteknologi 2009
Vid:
KTH Bioteknologi
Kungliga Tekniska Högskolan
AlbaNova universitetscentrum 106 91 Stockholm

The MSc project was performed within the framework of the thematic research program "MicroDrivE - Microbially Derived Energy" at the Faculty of Natural Resources and Agricultural Sciences at SLU.

Key words: crystallography, EndoT, ligand, zinc ion, Cel5H, enzyme characterization, HPLC, endoglucanase activity, processivity

Uppsala, 2009

Table of Contents

Abstract

The purpose of this project was to investigate the structure-function relationship of two carbohydrate active enzymes, Endo-N-acetyl-β-(D)-glucosaminidase (EndoT) from *Hypocrea jecorina* and Cel5H from *Saccharophagus degradans*.

The enzyme Endo-T from *H. jecorina* is a highly specific endoglycosidase, which cleaves aspargine-linked mannose rich oligosaccharides, but not highly processed complex oligosaccharides from glycoproteins. Endo T cleaves the bond between two N-acetylglucosamine (GlcNAc) subunits found attached to glycosylated asparagine residue. The structure of native EndoT has in a previous study been determined. This crystal structure that then had been obtained was a structure with a zinc atom bound in the active site of the enzyme. To better understand the architecture of the active site of Endo-T, attempt was made in this study to determine the structure of a ligand-protein complex. For this purpose the protein was crystallized in the absence of zinc which otherwise could occupy the active site preventing the ligand from binding. Several ligands were tried in co-crystallization experiments, but no protein-ligand crystal complexes were obtained.

Cel5H from *Saccharophagus degradans 2-40*, a putative β-1,4 endoglucanase, belongs to glycoside hydrolase family 5. Very little is known about how this enzyme cleaves the substrate. In this study, we have tried to crystallize Cel5H with the aim to solve the three dimensional structure of the enzyme. Many attempts to crystallize the enzyme were performed, but no crystals were optained. To understand the substrate cleavage pattern of Cel5H, the function of enzyme was characterized by studying the hydrolysis pattern of the two substrates cellopentaose and cellohexose. The hydrolysis products were separated by reverse phase chromatography using a HPLC instrument. The results indicate that Cel5H has mainly endoglucanase activity and it has an preference for cleaving the substrate from the reducing end. The results were not conclusive enough here was not enough to if the enzyme was acting processively or not

Key words: crystallography, EndoT, ligand, zinc ion, Cel5H, enzyme characterization, HPLC, endoglucanase activity, processivity

1 Introduction

1.1 Lignocellulosic biomass

The increasing requirement of energy directly causes the depletion of the fossil fuel and excessive emission of green house gases, which further more bring up political, environmental and economic problems all over the world. In order to alleviate the dependence on crude oil, the search for a new renewable and clean energy is urgent. Nowadays, the most widely used renewable liquid fuel is bio-ethanol. From 2000 to 2007, the global yearly bio-ethanol production increased from 17.25 billion liters to more than 46 billion litters[1]. There are three main types of feedstock that can be utilized for bio-ethanol production: sucrose-containing material (e.g. sugar cane, sugar beet), starchy material (e.g. corn, wheat, rice, potato) and lignocellulosic biomass[1]. In the short term, sucrose and starchy material, which are mainly composed of crops, are the major sources; while in the long run, lignocellulosic biomass will most likely replace the former sources, due to its low cost and high abundance.

Examples of lignocellulosic biomasses are: wood, grass, agricultural residues, forestry wastes and daily garbage. There are three major components of lignocellulose: cellulose, hemicellulose and lignin[2]. Besides cellulose and hemicellulose, pectin is the third most abundant polysaccharide in plant cell walls. This polymer is made up of several kinds of monosaccharide such as galacturonic acid and L-rhamnose. Lignin is built up of several types of aromatic precursors. The composition of each component of the plant cell wall varies between different types of plant and even the age of the plant.

1.1.1 Cellulose

Cellulose is the basic building block of the cell wall of a plant. It comprises around 40-50% of the dry weight of a plant[1]. Cellulose is made up of repeating glucose units, which are connected exclusively via β-1,4 glucosidic bond. The number of repeating unit in the cellulose polymer varies between several hundreds to more than 10,000[2]. In the plant cell wall, the neighboring cellulose chains are linked firmly together with hydrogen bonds forming cellulose microfibrils, which in the plant cell wall are covered by hemicellulose, pectin and lignin to form the rigid nature of the plant cell wall[3]. This well-organized crystalline structure of cellulose and the association of it with other components in the cell wall render the plant certain mechanical strength and resistance to enzyme degradatio[4].

1.1.2 Hemicellulose and pectin

Hemicellulose makes up approximately 25-35% of wood dry weight[1]. It consists of several basic sugar units such as D-xylose, D-mannose, D-galactose, 4-O-methyl-glucuronic acid, L-arabinose and L-fucose, which build up xyloglycan, arabinoxylan, etc[5]. Compared to cellulose, hemicellulose does not only consist of s a homogenic glycan back bone chain with only one type of sugar such as xylan, glucan or mannan. Hemicellulose also has branches arranged in certain repetition, which varies depending on plant types or stages of growth[4]. Moreover, the degree of polymerization of hemicellulose is not as high as for cellulose. In addition, hemicellulose dose not aggregate to any higher extent therefore hemicellulose is easier degraded by enzyme than cellulose is.

Pectin is heterogeneous group of polysaccharides. It includes mainly homogalacturonan with homogeneous galacturonic acid liner chain, rhamnogalacturonan I with L-rhamnose and galacturonic acid in the main chain, and rhamnogalacturonan II forming dimmers via borate easters[5]. All pectin polymers have complex side chains, which are covalently linked to

cellulose or to themselves. The hemicellulose and the pectin are tightly filling in the matrix of the cellulose microfibrils, and therefore increases the tensile strength of the cell wall structure[6].

1.1.3 Lignin

Lignin is found in the secondary cell wall of plants, which is deposited inside of the primary cell wall. The structure of lignin polymer is complicated. The basic unit is phenylpropane, which is joined together by various linkages[3]. Inside the plant cell wall, the lignin cross-links to the branches of hemicellulose via ferulic acid ester linkages[1], which strengthens the mechanic force. Since lignin is non-water soluble and inert, it provides the plant resistance to water, oxidation and biological attack.[3].

1.2 Biomass degradation by enzymes from microorganisms

Fungi and bacteria are the major biomass degraders in our ecosystem. These microorganisms utilize cellullases, hemicellulases, pectin lyases or lignin degrading enzymes to deconstruct the well-organized cell wall structure[2]. Most of cellulolytic fungi and some bacteria secrete complexes of enzymes for cellulosic hydrolysis, which can be separated and recombined artificially for different purposes in industry. Another group of cellulolytic microorganisms bacteria do not secrete free extracellular enzymes into the environment surrounding the microorganism, instead, they have big multi component complexes called cellulosomes that are protruding outside the cell surfaces, on which cellulolytic and hemicellulolytic enzymes are arranged discretely[6]. The two enzymes studied in this project are from one of each group, respectively.
To enable for the microorganisms to degrade lignocellulosic biomass in an efficient way they need the synergistic collaboration of cellulases, hemicellulases and lignin degrading enzymes. Since cellulose is the most abundant component of the biomass, cellulases are vital in loosing the intact structure and producing fermentable sugars. There are three groups of celluloses. Exo-glucanases, also called cellobiohydrolases, which cleave off cellobiose from either reducing end or non-reducing end of a cellulose chain; Endoglucanases that make internal cut randomly in the middle of the cellulose polymer chain; and β-glycosidases that hydrolyze cellobiose, two linked glucose molecules, which is the smallest repeating unit in a cellulose chain into glucose[7, 8].

Hemicellulases have a heterogeneous structure, which can be broken down to its mono saccharide building blocks by a wide verity of different enzymes, for example xylan is hydrolyzed by endo-1,4-β-xylanase and β-xylosidase. Pectins digestion involves both hydrolases and lyases[9]. In nature there exist a wide range of different cellulolytic and carbohydrate active enzymes. The known ones of these have all been listed in a database called the CAZy (Carbohydrate Active Enzymes) databases. The carbohydrate active enzymes are classified into different families in this database depending on sequence similarity to other known enzymes and tertiary structure[10].

1.2.1 Endo-N-acetyl-β-(D)-glucosaminidase (EndoT) from *Hypocrea jecorina*

1.2.1.1 Hypocrea jecorina

Hypocrea jecorina, formerly named *Trichoderma reesei*, is a mesophilic soft-rot ascomycete fungus. It was firstly isolated by U.S. Army during World War II, since it caused deterioration of cotton cloth[11]. Nowadays, the strain has been modified and is widely used in industry for its superior ability of secreting extracellular cellulases and hemicellulases, which are widely

used in textile bleaching, paper industry and bioethanol production. Industrial strains are able to produce more than 40 g/L cellulosic hydrolases[13]. The genome sequence was published in 2008 by Martinez *et al*[12]. Compared to many other biomass degrading fungi, *H. jecorina* encodes fewer glycoside hydrolyze (GH) enzymes than most of these. However, due to its extraordinary ability to produce waste amounts of cellulolytic enzymes, it is still a hotspot for research studies worldwide.

1.2.1.2 Endo-N-acetyl-β-(D)-glucosaminidase (EndoT)

Endo-N-acetyl-β-D-glucosaminidase T (EndoT, E.C. 3.2.1.96) can be isolated from the culture medium which the soft-rot fungus *H. jecorina* is cultivated in. The enzyme EndoT belongs to the glycoside hydrolase (GH) family 18[10]. It is the firstly detected endo-glucosaminidase (ENGase) from a non-bacteria origin. The enzyme hydrolyzes the di-N-acetylchitobiosyl part of oligosaccharides linked to asparagine (Asn) residues of glycoproteins after they are secreted. The substrates for EndoT include oligomannosidic, phosphorylated and hypermannosylated type glycoproteins[13]. The molecular weight of the protein is about 33KD, and has a PI of 3. The theoretical values were calculated based on the Amino acid sequence on ExPASy proteomics server http://www.expasy.ch/.

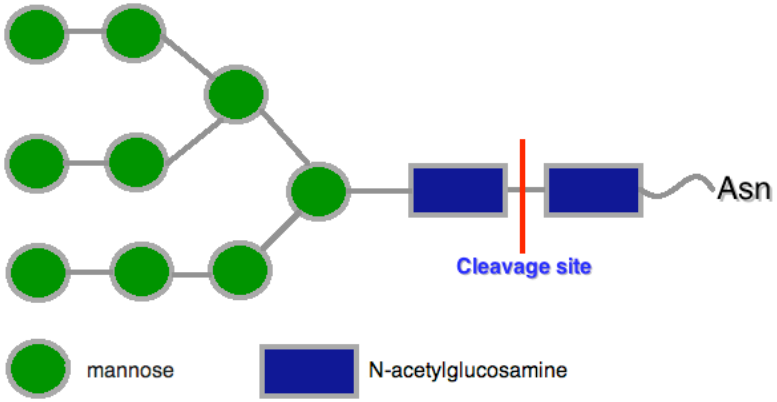


Figure 1. A schematic figure of an oligomannoside N-glycan. The cleavage site of the N-glycan by EndoT is between the two N-acetylglucosamines residues found bound closest to the glycosylated asparagine residue. The number of mannose residues attached to the varies between 3 to 10[14].

A homologous enzyme to EndoT, EndoH from *Streptomyces plicatus,* is commonly used for deglycosylation of proteins within glycoprotein research[14]. The use of EndoH for deglycosylation of glycoproteines is fairly expensive. EndoH and EndoT have similar substrate specificities. The biggest difference between the two enzymes is that EndoT is secreted as an extracellular protein, which EndoH isn't. The fact that EndoT is expressed as an extra-cellular protein makes it easier to be extracted and purified, compared to EndoH. One way to lower the cost for deglycosylating enzymes could be to start using EndoT instead of EndoH, since EndoT could be secreted to medium by expression host *Hypocrea jecorina* and extracted a lower price.

The first structure of EndoT was solved in a previous study by Mia C. Hertzberg, to a resolution of 1.3 Å[15]. This first Endo T structure did contain a zinc molecule bound in the active site of the enzyme. The overall structure of EndoT is a $(\alpha/\beta)_8$-barrel. The two catalytic residues of the enzyme are aspartic acid (Asp) 129 and glutamic acid (Glu) 131, respectively. In addition, tryptophan (Trp) 259 is supposed to locate the substrate chain into cleft[16]. Until now there was no structure of EndoT in complex with a substrate that had been published. A ligand complex structure would reveal information on how a ligand interacts with active side

residues in the enzyme, and which will in turn be a knowledge base for protein engineering of EndoT in the future.

1.2.2 Cel5H from Saccharophagus degradans

1.2.2.1 Saccharophagus degradans

Saccharophagus degradans 2-40 is a marine Gram-negative bacterium firstly isolated from a salt marsh cord grass *Spartina alterniflora* in Chesapeake Bay, USA[17]. *Saccharophagus degradans* 2-40 is an outstanding degrader in marine system that is reported to degrade at least 10 complex polysaccharides (CPs)[18], such as agar, chitin, cellulose, laminarin, which mostly come from algae, plants, bacteria and fungi. Instead of secreting isolated extracellular enzymes, the strain has surface protuberances resembling cellulosomes on which cellulolytic enzymes, and other carbohydrate active enzymes assemble and cooperate together when degrading biomasses in the surrounding of the organism[19]. *S. degradans* versatile degradation ability facilitates the carbon cycle in marine system[19]. The organism's different carbohydrate active enzymes could be utilized for bioethanol production, or within other industrial processes.

1.2.2.2 Cel5H

Cel5H from *S. degradans* has been classified as a putative β-1,4 endoglucanase (E.C. 3.2.1.4) (Uniprot Q21FN5). As many other carbohydrate active enzyme Cel5A has a catalytic domain, belonging to GH family 5, and a carbohydrate binding module (CBM), belonging to CBM family 6 [10]. The complete enzyme, including both the catalytic module and the CBM, has 630 amino acids[20]. Since *S. degradans* is a strong degrader of cellulosic material, utilizing enzymes from this organism has the potential of reduce the high cost for degrading plant biomass into smaller mono- or oligosaccharides in industrial processes. In order to understand the biological function of Cel5H and thereby also be able to understand the value of this enzyme within industrial processes, fundamental researches, including crystallization of the protein and biochemical characterization will be necessary to carry out.

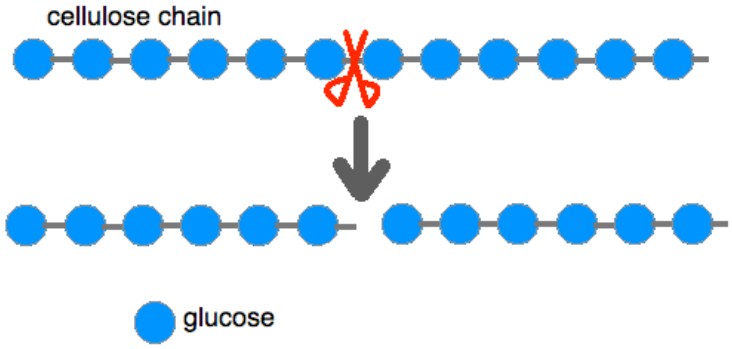


Figure 2. A schematic figure of how an endoglucanase cleaves a glucan chain. The red scissor shows one of the cleavage sites.

1.3 Methodology

1.3.1 Crystallography

X-ray crystallography has been used for determining the atomic structure of macromolecules since 1923. (REF) The methodology of this technology is based on that the electrons in an atom scatter X-ray. The diffraction pattern of one molecule indicates the spatial occupancy of

its electron, i.e., electron density. Thus, it is possible to determine the atomic structure of a protein structure through its x-ray diffraction pattern. The work to determine a structure of a protein starts from crystallizing a highly purified sample of the protein. If crystals of the protein are successfully acquired, one or more of these can be mounted in a loop and flash frozen in liquid nitrogen (100 °K). The crystal is then exposed to X-ray and a diffraction pattern of it can be collected. The acquired x-ray diffraction dataset that has been collected on the crystal will thereafter be processed, the space group and cell parameters of the crystal can thereby be determined. A structure solution will be chosen to determine the rough electron density map, which will be refined again and again until final structure is completed.

It is not possible to collect a diffraction pattern from one single protein molecule because the diffraction signal from one molecule is too weak to be detected. Since a crystal is a repeating unit of identical molecules with a certain pattern, it acts like an amplifier of the diffracted ray, thus, the diffraction signal gets amplified from each protein molecule in the protein crystal.

Crystallization, in another word is to form crystals. The most commonly used method for crystallizing a protein is the so-called vapor diffusion technique[21]. This is also the crystallization method that has been used all through this project. In this crystallization technique, a drop of a protein solution is placed on a glass lid or set on a plate and sealed into a closed system with a reservoir solution. The drop contains few to 10 µl of protein solution mixing with the same or similar volume of reservoir/crystallization solution that in most cases is a volume between 80-1000 µl. The difference in concentration between the drop and the reservoir drives the vapor diffusion to reach equilibrium, i.e., water molecules diffuse from the drop containing the protein to the reservoir solution. This diffusion of water molecules from the drop to the reservoir solution causes an increase in protein concentration in the drop and this increase in concentration may cause the protein in the drop to start nucleate and start forming protein crystals if the crystallization experiment is successful.

Many factors may influence the crystallization result. The purity of the protein should always be checked prior a crystallization experiment. Besides, protein concentration, buffer type, buffer concentration, pH and temperature are also important parameters within a crystallization experiment. The art of crystallization is to change these different parameters and try to find out the optimal combinations. There are several commercial screening kits available for seeking the initial conditions, such as the Core-96 JCSG+ crystallization screen used in this experiment, with which initial crystallization conditions can be found. After finding initial crystallization conditions for a protein of interest, these initial conditions need to be optimized by finely adjusting the parameters until crystals that are good enough for X-ray diffraction data collection have been obtained.

After good enough crystals have been obtained, they are mounted on a small supporter, soaked for a while in cryo-protectant (a nylon loop is used in the experiment), and flash frozen in liquid nitrogen. Freezing the crystal reduces radiation damage during the data collection when the protein crystals are exposed in a high energy X-ray beam at a home or synchrotron X-ray source.

The data collected during the exposure of the protein crystal in an X-ray beam is a so-called diffraction pattern. This diffraction pattern, which is an irradiation, can be recorded on a standard X-ray film, or nowadays most often using a ccd detector. The diffraction pattern can be used to calculate an electron density of the protein forming the crystal. The X-ray forming the diffraction from the crystal has amplitude, a wavelength and a phase angle. If all three parameters are known, the electron density map for the protein can be calculated from the diffraction pattern collected from a crystal exposed in the X-ray beam. Except for the phase angle, all needed parameters are usually easily extracted easily from the collected X-ray

diffraction data. Only the information of phase angle of the X-ray is lost during the data collection. This lost of phase angle is within X-ray crystallography called the phase problem. To be able to see the electron density of the molecule forming crystal this phase problem has to be solved.

There are three major routes used today within protein X-ray crystallography to solve the phase problem, these are called molecular replacement (MR), multiple isomorphous replacement (MIR) and anomalous dispersion (AD) methods. Which of these methods that is used to solve the three dimensional structure of a protein depend on if there exist known structures that are similar to the protein of interest or not.

In this thesis, the three dimensional structure of the protein under study was solved in a previous study[15]. This enabled us to use the MR method to solve the structure through all structure determining experiment within the study.

Within the MR method to determine the structure of a protein from a X-ray diffraction data-set, one calculates a electron density from the known protein structure, and use this calculated electron density to search for a solution of the unknown structure from the collected X-ray data-set.

Within the MIR and AD methods one make use of that heavy atoms gives a strong anomalous signal when exposed to X-rays when solving the phase problem for a diffraction dataset and thereby solve the structure of the protein. The MIR is performed by soaking or co-crystallizing crystals with a heavy atom, which hopefully does not affect the formation of crystal or the dimension of unit cell bound to the protein. A dataset is collected from a crystal soaked in a heavy metal and a dataset is collected from a crystal not soaked in a heavy metal. Then the differences between the two collected dataset are calculated and these difference between the two datasets are used to solve the phase problem for the structure. The Anomalous dispersion, methods consist of both single-wavelength (SAD) and multi-wavelength (MAD) methods. These methods utilize the fact that heavy-atom in protein gives stronger anomalous signal at specific wavelengths. During a MAD X-ray data collection one changes the wavelength of the X-ray beam generated by synchrotron, while during a SAD X-ray diffraction data collection one uses the same wavelength during the whole data collection.

After obtaining the initial electron density map of a protein from an X-ray dataset, then one has to fit atoms of the protein into the calculated electron map manually by using some model building soft wares. In this thesis work the protein modeling program Coot[22] was used to build the protein structure model.

The structure model is refined, and the electron maps further improved, by the use of structure refinement programs such as Refmac[23]. The model is then further manually built and refined until a structure model of the protein of high enough quality is obtained.

1.3.2 Enzyme characterization by high performance liquid chromatography

High performance liquid chromatography (HPLC) is a chromatography separation method widely used for separation, purification, identification or quantification of chemical substances. In this method one make use of a matrix packed in a column (stationary phase), a pump to transport the analyte, a mobile phase, and a detector. The analyte dissolves in the mobile phase solvent and flows through the column. The interactions between analyte and solvent, and the analyte and the column determine the separation of the analyte. The advantages of the HPLC method compared to other liquid chromatographic methods in use is

that it in this method one is able to work under high pressure so as to speed up the flow rate, and the fine particles of the column material greatly increases the surface area, which allows the better resolution and separation of the particles in the analyte.

There are two main HPLC methods in use, and these are depending on the polarity of column and the solvent, these two main methods are called the normal phase and the reverse phase chromatography methods. The normal phase HPLC method uses a polar stationary phase and a non-polar mobile phase. Polar compounds in the analyte are adsorbed in the column for a longer time than non-polar compounds. Reverse phase column is applied in this project. Reverse phase, in contrary to normal phase, uses a non-polar stationary phase and polar mobile phase. Substance in the analyte which are more hydrophobic is attracted to the stationary phase for longer time, by van der Waals dispersion force, than those substances in the analyte that are less hydrophobic. Less polar molecules spend more time flowing with polar solvent by forming hydrogen bonds. In the experiment described in this thesis work, the analytes are mono or oligo-glucans. Based on the theory, the longer the polymer is, the later it gets eluted from the column. By varying the composition of mobile phase by time, the effect of separation can be optimized.

There are two different elution methods commonly used. These differ in the composition of the mobile phase and they are called isocratic flow and gradient elution. When the composition of the mobile phase is kept constant throughout the procedure, the elution is called isocratic. Although this method is simple, resolutions of later eluted compounds are problematic. Therefore, a gradient elution is applied in this experiment. In this case, the organic strength of the mobile phase is increasing in a linear or stepwise fashion, which results in the elution of all compounds with different polarity. In this experiment, since longer glucose chain has stronger polarity, these will be eluted later.

Instead of a ultra-violet detector normally used within HPLC separations, in this study a charged aerosol detection (CAD)[24] (ESA Biosciences, Inc., USA**)** technology was utilized. In this detector, the analyte eluted from the HPLC column is firstly nebulized and dried by nitrogen gas, and then charged with positive ions. When the positive particles fly through the sensitive electrometer, the generated signal is proportional to the quantity of analyte. The CAD detector has a relatively high sensitivity and it is possible to detect particle concentrations in the analyte in the nanogram range. The CAD detector has other advantages, such as high reproducibility, dynamic detection range, and broad applicability. Therefore, this detector was suitable for detect the tiny amount of oligosaccharides which were produced when longer sugar chain were hydrolyzed to shorter oligos in the enzymatic experiments described in this paper.

Objectives

The structure of EndoT in complex with a ligand

One aim of the aims for this thesis work was to solve the structure of the enzyme EndoT with a ligand bound in active site. EndoT is the first found fungal ENGase, and this new enzyme could potentially be used in different types of industrial and research applications, e.g. for the deglycosylation glycoproteins. The source organism of EndoT, *Hypocrea jecorina* is known for its strong ability of producing extracellular cellulolytic enzymes.

EndoH is the prevalent deglycosylation enzyme used in many research labs and since this ENGase is relatively expensive to purchase, its high price can add up to a substantial cost for a research lab if it is commonly used in the lab. If it could be shown that EndoT has higher the same or higher activity than the commonly used EndoH, the cost for using an ENGase in the lab could be substantially reduced, if the EndoT enzyme could be expressed in relatively large amounts by the expression host *H. jecorina*. Till now, the structure of apo-EndoT has been solved. If ligand can be visualized in the active site, the structure will not only help us

understand the mechanism of its biological function, but will also be a foundation for further rational designs.

Solve the native structure and study the biochemical function of Cel5H

The initial purpose of this part of the thesis work was to solve the structure of Cel5H, which is a family member of glycoside hydrolase family 5 (GH5). The enzyme Cel5A was chosen because the host *Saccharophagus degradans* has amazing ability to degrade many kinds of cellulosic substrate, and Cel5H is one of this organism's enzymes with high ability to degrade cellulose. The mechanism by which this enzyme degrades cellulose is not so well known. A three dimensional structure of the enzyme may reveal some hints on the hydrolytic mechanism of the enzyme, assisted by biochemical studies of the enzyme in which the products when incubating the cellooligos cellohexaose (G6) and cellopentaose (G5). Together with Cel5H is analyzing by HPLC.

2 Material and methods

2.1 EndoT and ligands

The EndoT protein was a sample that was available in the lab at the concentration of 16.5 mg/ml. The purity was determined to be high enough for crystallization experiment.

Eight different ligands were either co-crystallized with the EndoT protein, or soaked into the protein crystals prior X-ray data coolection. Six of those ligands were a kind gift from Hogeschool Gent, Belgium (No.1-6 in Table 1). These ligands were freeze-dried before delivered. The name and amount of each ligand is listed in table 1.

Table 1. Eight ligands used for crystallization

No.	Name of ligands	Amount
1	Mannopentaose-di- (N-acetyl)-D-glucosamine **$Man_{5-9}GlcNAc_2$**	1 nmol
2	Mannopentaose-N-acetyl-D-glucosamine **$Man_{5-9}GlcNAc$**	0.8 nmol
3	Mannopentaose-di-(N-acetyl)-D-glucosamine-asparagine **$Man_5GlcNAc_2Asn$**	0.6 nmol
4	N,N' diacetylchitobiose **Chitobiose**	2.35 µmol
5	Uncharged N-glycans from Cel7A released with Endo H **$GlcMan_{7-8}GlcNAc$**	Unknown
6	Charged N-glycans from Cel7A released with Endo H **(ManP)GlcMan7-8GlcNAc**	Unknown
7	Glucose	Sufficient
8	Cellobiose	Sufficient

Among the six ligands from Hogeschool Gent, $Man_{5-9}GlcNAc_2$, $Man_{5-9}GlcNAc$, and $Man_5GlcNAc_2Asn$ are natural substrates for the enzyme, while Chitobiose, $GlcMan_{7-8}GlcNAc$ and (ManP)$GlcMan_{7-8}GlcNAc$ are natural products. All ligand samples were dissolved and diluted to the concentration stated in Table 1 before being used in either the co-crystallization or crystal soaking experiments. Since the concentrations of $GlcMan_{7-8}GlcNAc$ and (ManP)$GlcMan_{7-8}GlcNAc$ samples were unknown, they were assumed to 100mM in 10µl Milli-Q water, and the amount was 1 µmol.

2.1.1 Crystallization

For EndoT crystallization experiments, the hanging drop vapor diffusion method was used throughout all crystallization experiments with this enzyme, and using the EasyXtal Tool crystallization plate (Qiagen, USA). Two different crystallization conditions were optimized to obtain crystals or crystal-ligand complex of EndoT. The initial crystallization conditions for EndoT were obtained within a previous bachelor's thesis by Mia C. Hertzberg[15]. The first condition attempted was 0.1M sodium acetate (NaAC) pH 5.0, 0.2 M zinc acetate (ZnAC), and 5%-18% Polyethylene glycol (PEG) 3350. It was important to get crystals that did not contain metals such as zinc, since zinc was found to block the active site from binding with ligands, therefore a second crystallization condition without a metal in the crystallization agent, 6%-12% PEG8000 and 6%-12% PEG1000, was used. The initial drop size in these crystallization varied from 3µl to 4µl, with half volume of 16.5 mg/ml EndoT protein sample. All crystallization experiments were incubated in 20°C.

All eight tested ligands were co-crystallized with the protein solution or soaked into crystals with prior flash freezing. Co-crystallizations were performed by adding one ligand solution into protein drops when these where prepared. Soaking was done by adding ligand solution into drops where crystals had already formed. Glucose and cellobiose, due to their lower price and sufficient amount, were tested more than the other ligands, by varying the concentrations of these ligands, crystallizing condition and the two ligand-binding methods. The other six ligands were co-crystallized or soaked into crystals obtained in the condition without zinc (only PEG8000 and PEG1000). Since $Man_{5-9}GlcNAc_2$, $Man_{5-9}GlcNAc$ and $Man_5GlcNAc_2Asn$ are nature substrates for the enzyme, and they will therefore be hydrolyzed by the enzyme. In order to inactivate the enzyme to enable binding a ligand in the active site of this that was not hydrolyzed, soaking experiments were carried out in a cryo solution at pH3.5, buffered by citric acid, a pH at which the enzyme is most likely no longer active. Table 2 below shows the conditions tested for each ligand, including the crystallizing condition, ligand binding methods and the concentration of ligand solution.

Table 2. Conditions for ligands

Ligands	Conditions	Soaking/co-crystallization	Ligand concentration
$Man_{5-9}GlcNAc_2$	No Zn	Soaking	40 μM*
$Man_{5-9}GlcNAc$	No Zn	Soaking	40 μM*
$Man_5GlcNAc_2Asn$	No Zn	Soaking	50 μM*
Chitobiose	No Zn	Co-crystallization	18 mM
		Soaking	54 mM
$GlcMan_{7-8}GlcNAc$	No Zn	Soaking	NA
$(ManP)GlcMan_{7-8}GlcNAc$	No Zn	Soaking	NA
Glucose	With Zn	Co-crystallization	20 mM
	No Zn	Co-crystallization	20 mM
		Soaking	45 mM
Cellobiose	With Zn	Co-crystallization	10 mM
	No Zn	Co-crystallization	10 mM
		Soaking	10 mM, 25 mM

Note: "*" represents soaking proceeded in citric acid buffer, pH 3.5.

2.1.2 Crystal mounting and freezing

Cryo-protectants were prepared before freezing the crystals. Two cryo solutions were used in this project. For the condition with zinc, 0.1 M NaAC, 0.2 M ZnAC, 35% PEG3350 and the same concentration of the corresponding ligand was used. For the condition consisting only PEG, 35% PEG3350, 10% glycerol and the same concentration of corresponding ligand was used.

When the EndoT crystals in a drop had reached a acceptable size, these crystal were transferred from the original drop to a drop of cryo-protectant solution and kept there for apr. 1min. For those crystals soaked in a cryo-solution at pH 3.5, including a ligand, the incubation time was elongated to 90 minutes. During ligand soaking, the crystals were visually inspected with a microscopy for signs of damage. After that, the crystals were mounted on a loop and flash frozen by plunging the loop with the crystal into liquid nitrogen (100 K).

2.1.3 Data collection and process

The X-ray diffraction data sets were collected using two different x-ray sources: 1) at a home source X-ray rotating anode generator (Rigaku Rotaflex RTP300) equipped with a Mar345 IP detector locating in the lab. 2) at a synchrotron X-ray source, beamline I911-3 and I911-5 at the MAX-lab in Lund, Sweden.

The X-ray datasets were processed with the integration program iMosflm 1.0.0[25]. The integrated data was scaled with the program Scala in the CCP4i program suite[26]. After that, molecular replacement was used to solve the structure by running Phaser[27] in the CCP4i suite. The previously solved EndoT structure with zinc molecules was used as the search model for the MR solution. The structure model refinement was done by alternately running refinement program Refmac5[23] and the structure model building program Coot[22]. Cartoons of the final structure models where prepared using the structure model visualizing program PyMOL[28].

2.2 Cel5H

Two different protein batches of Cel5H catalytic domain were used for initial crystallization screening experiments. One of these two different batches of Cel5H was an old batch of this enzyme that was the remains from a previous study of the enzyme the lab. The new batch of Cel5A with a protein concentration of 4.97 mg/ml, buffering with 50 mM sodium phosphate, 0.25 M ammonium sulfate, pH 6.5. The total volume was 825 µl. Both of the Cel5H enzyme batches used were kind gifts from The University of Maryland (USA)

2.2.1 Purification

A purification of the enzyme was only carried out for the new batch of enzyme. In this purification, a SuperDex75 (GE Healthcare, Uppsala, Sweden) size exclusion column was used for removing impurities of low molecular weight contaminating enzymes found in the protein sample delivered from The University of Meryland. During the size exclusion purification the buffer was changed to 30 mM Bis-Tris pH 6.5 and 25 mM sodium chloride (NaCl), which also was the running buffer used during the purification. After the size exclusion purification the elution fractions were analyzed by running a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using an 8-25% gradient Phast gel (GE Healthcare) and a Phast gel electrophoresis system (GE Healthcare). Elusion fractions containing protein of right size were pooled and concentrated to a final protein concentration of 12.73 mg/ml using a vivaspin 20 centrifuge concentrator (GE Healthcare) with a molecular cutoff of 10 kDa.

2.2.2 Cel5H Crystallization Screening and optimizations

For the old batch of Cel5H a Core 96-JCSG+ screen (Qiagen, USA), and a PEG/Ion HT (Hampton research, USA) screen was initially set up using a Greiner 96 Well CrystalQuick crystallization plates for sitting drop (Greiner Bio-one, Belgium). The crystallization drops were prepared using Douglas instrument XYZV crystallization robot (Douglas instruments, Berkshire, U.K.). The size of the crystallization drops was 0.6 µl. These were prepared by mixing equal amounts of protein solution and crystallization agent (0.3+0.3µl). Different Cel5H protein concentration 8.01 mg/ml, 10.68 mg/ml and 15.12 mg/ml, were tested. For the new batch of enzyme, in addition to Core 96-JCSG and PEG/Ion screen, an ammonium sulfate grid screen (Hampton Research) was also used.

Using the Core 96-JCSG+ screen an initial crystallization condition for Cel5H was identified, this was 0.2 M potassium nitrate (KNO_3) pH 6.9 and 20% PEG3350. To further improve the quality of the crystals, this initial crystallization condition was optimized by making a grid screen, in which the concentration of KNO_3 was varied from 0.15 M to 0.25 M

and the PEG3350 concentration was varied from 18% to 24%. After 10 days, seeding of the new drops that then were still clear using the crystals that was obtained in the initial crystallization experiments as nucleation seeds.

2.2.3 Enzyme characterization

In order to carry out an initial characterization of the hydrolysis mechanism of Cel5H, the ligands cellohexaose (G6) and cellopentaose (G5) were chosen to be tested as substrates for Cel5H. An high performance liquid chromatography (HPLC) system, Agilent 1200 series LC (Agilent Technologies, Sweden), equipped with a C-18 column, was used to separate oligosaccharides, and a Corona Charged Aerosol detector (ESA Biosciences, USA) was used for detecting the substances released from the C-18 column. Before running the reaction product after incubating cellooligos with Cel5H, a standard mixture with defined concentration of glucose (G1), cellobiose (G2), cellotriose (G3), cellotetraose (G4), cellopentaose and cellohexaose was injected on to the HPLC system to find out which elution gradient that should be used to get best possible separation of the glucan products. The gradient used was a combination of Milli-Q water and acetonitril (AcN). Table 3 show the gradient was identified as best and also used in the following hydrolysis experiments for measuring the different amounts of different cellooligo products.

Table 3. Method used in HPLC analysis

Time (min)	Milli-Q (%)	Acetonitril (%)
Initial	100	0
1	100	0
15	85	15
17	5	95
18	5	95
19	100	0
21	100	0

The hydrolysis reaction was carried out at room temperature. Since the peak of the original buffer overlapped with the peak of glucose and also broadened the peak of G4, the hydrolysis experiments were carried out in Milli-Q water. Different molar concentrations of G5 and G6 were tested, in combined with different concentrations of the enzyme, to figure out the most appropriate reaction rate which easily could be monitored, by the CAD detector mounted after the HPLC system.

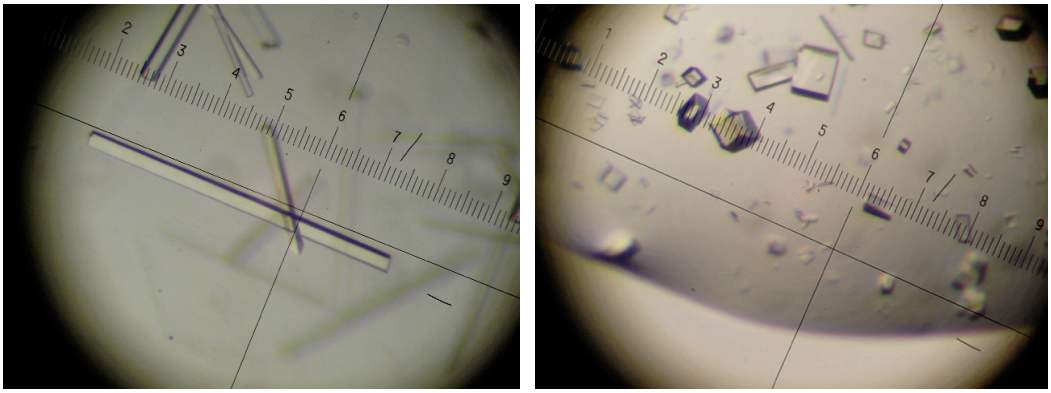
Each HPLC separation of the hydrolysate took approximate 25 min. The best combinations of ligand and enzyme was found to be 67 μM G5 with 10.15 nM Cel5H, and 50 μM G6 with 1.52 nM Cel5H. Untreated and defined concentrations of G5 and G6 were injected and separated on the HPLC system and the detected mass concentration of these ligands were used as the concentration of these ligands at time point zero of the Cel5H hydrolysate experiments. The Cel5H enzyme was mixed with appropriate amounts of the substrates, incubated in room temperature for around 10 min, and then injected into the HPLC system and the amounts of the hydrolytic products were determined. The additional incubation times before injection samples of the reaction mixtures on to the HPLC system were: 30 minute, 4 hours, and overnight.

3 Results

3.1 EndoT

3.1.1 EndoT crystallization

EndoT crystals were successfully obtained both with and without zinc atoms added in the crystallization agent. The best crystals of EndoT, which were used for X-ray data collection, were found in the following two crystallization conditions: a) 0.1 M NaAc pH5.0, 0.2 M ZnAC, 8% PEG 3350; b) 6% PEG8000 and 6% PEG1000. The crystals found growing in these two crystallization conditions were differing in morphology. The crystals found growing in the crystallization condition including zinc solution had a rod-shape (**Figure 3a**), whereas crystals found in the crystallization solution without added zinc were cubic, shorter and wider (**Figure 3b**). All ligands were co-crystallized with EndoT solution or soaked into formed EndoT crystals. There was no crystal damage during soaking step.



(b)

Figure 3. Photos of EndoT crystals grown in two different crystallization conditions. (a) EndoT crystals grown in 0.1 M NaAc pH5.0, 0.2 M ZnAc and 8% PEG3350 (rod shaped). (b) EndoT crystals grown in 6% PEG8000 and 6% PEG 1000 (cubic).

3.1.2 EndoT data collection and processing

The EndoT X-ray diffraction datasets were either collected with an in house X-ray source at the lab or at the synchrotron in Lund. Except for EndoT in complex with $GlcMan_{7\text{-}8}GlcNAc$, which failed to diffract in good resolution, all other datasets on ligand soaked crystal were successfully collected. The failed experiment was not repeated since the concentration of $GlcMan_{7\text{-}8}GlcNAc$ was unknown, and the rest amount of the ligand was not enough for more attempts.

All datasets were integrated and scaled successfully. All crystals had the same space group and cell dimension. All the structures with one exception were solved by molecular replacement using the program Phaser[27] (a program in the CCP4i package). The map of EndoT in complex with chitobiose failed to be calculated, because the distance between the crystal and the detector at the lab was wrong and could afterwards not be determined. However, calculated electron density maps from all other collected datasets show no ligand binding in the active site. So besides EndoT soaking with G2, the other structures were not refined.

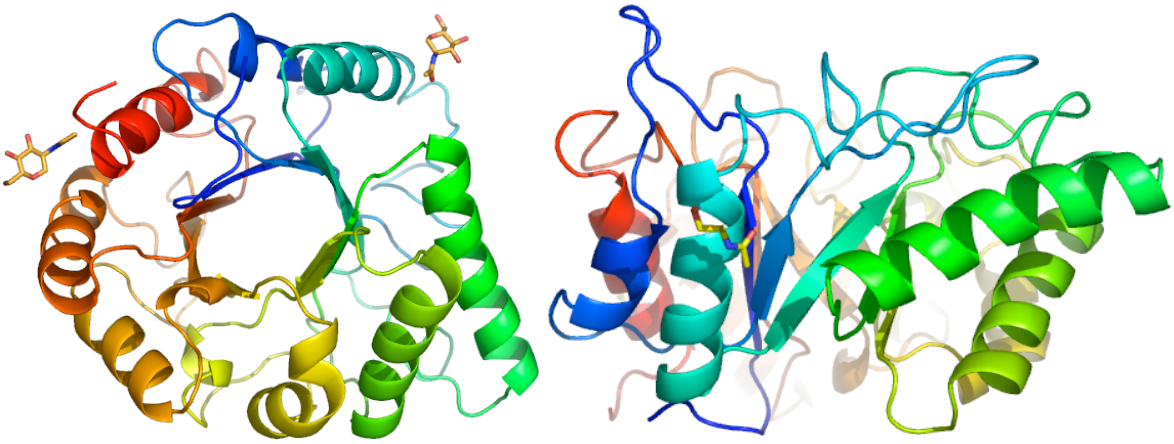
The structure without zinc atoms (EndoT-free) was firstly solved and refined in this study. The diffraction dataset was collected from an EndoT crystal soaking with G2, but no ligand was found bound in the active site. The data were collected at MAX-lab, Lund at the beamline I911-2 to the resolution of 1.8 Å. The space group of the new structure as well as the cell dimension is the same as the previously solved EndoT-zinc structure; $P2_1$ and a = 35.38 Å, b = 63.75 Å, c = 59.31, β = 100.96°. Molecular replacement using the EndoT-zinc structure as search model was applied to solve the zinc-free EndoT structure. After refinement of the structure model, the R_{free} value and R-value was 21.93 and 17.16, respectively. The overall structure is shown in Figure 4.

The structure of the zinc-free EndoT structure is an $(\alpha/\beta)_8$-barrel with the active site of the enzyme sitting on top of one side of the β-barrel (figure 4). The active site residues consist of the two residues Asp 129 and Glu 131. The structure has two N-acetyl glucosamine (NAG) molecules attached to it, one is found bound to to Asn 70 and the other one to Asn 240. The β/loop/α units 5 and 6 are irregular, where the loops are shorter and an α helix is missing.

Table 4. Data collection, processing and structure refinement statistics

EndoT	NATIVE
A. Data collection and processing statistics	
Beamline[a]	I911-2
Wavelength (Å)	1.037
No. of images	180
Oscillation range (°)	1.0°
Space group	$P2_1$
Cell parameters (a, b, c: Å;)	35.38, 63.75, 59.31 beta 100.96
Resolution range (Å)	24.7 - 1.77
Resolution range outer shell	1.87-1.78
No. of observed reflections	168163
No. of unique reflections	25096
Average multiplicity	3.7 (3.3)
Completeness (%)[b]	99.5 (96.6)
R_{merge} (%)[c]	6.9 (25.5)
I/σ(I)	14.3 (4.8)
***B**. Refinement and final structure model statistics*	
Resolution used in refinement (Å)	24.7 - 1.77
Reflections in: working & test set	23684 & 1250
R[d] & R_{free} factor (%)	16.96 & 21.72
Residues in protein	287
Protein atoms	2218
Waters	282
Residues with double conformations	3
N-glycosylation (NAG) atoms	2
RMSD bond lengths from ideal (Å)	0.0087
RMSD bond angles from ideal (°)	1.185
Ramachandran outliers (%)	1.06

[a]Beamline at MAX-lab, Lund, Sweden
[b]Numbers in parentheses are for the highest resolution bins.
[c]$R_{merge} = \Sigma_{hkl} \Sigma_i | I - < I > | / \Sigma_{hkl} \Sigma_i | I |$.
[d]$R = \Sigma | |F_o| - |Fc| | / \Sigma |F_o|$; the final R-factor is given.



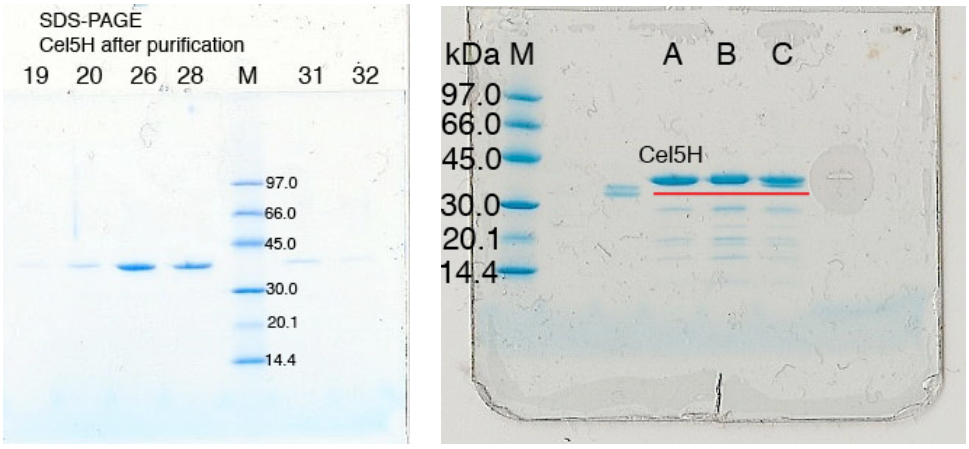
(a) (b)

Figure 4. Cartoon representation of the structure of EndoT without zinc atoms (EndoT-free). The two NAG molecules are also shown. (a) Bottom view of the EndoT $(\alpha/\beta)_8$-barrel shape. (b) Side view, with the missing and twisted α helices (light green).

3.2 Cel5H

3.2.1 Purification

Cel5H was purified by carrying out a size exclusion chromatography (SEC) purification, using a SuperDex75 column. The buffer of the protein sample was changed to 30 mM Bis-Tris pH 6.5, 25 mM NaCl during the SEC purification. The fractions at the edge and the middle of the elution peak were pure, as determined by SDS-PAGE gel electrophoresis (Figure 5a). But after that all the fractions containing EndoT had been pooled and concentrated, the sample still contained impurities (Figure 5b). In spite of this, the impurities were less after purification and the protein should be pure enough for crystallization.



(a) (b)

Figure 5. Photos of SDS-PAGE gel runs of EndoT. (a) The fractions at the edge and middle of the main elution peak. (b) The result of purification of Cel5H. Lane A, B and C are Cel5H new batch after purification, before purification and the old batch, respectively.

3.2.2 Cel5H Crystallization

For the old batch of Cel5H only one crystal had been found growing in the Core 96-JCSG+ Suite screen, solution A12 of the screen, which is 0.2M potassium nitrate pH 6.9, 20% PEG 3350. However, the following seeding attempts using this crystal failed to induce any crystal growth in any further crystallization optimized screens carried out. For the purified new batch of Cel5H, no crystal was found in any of the crystallization screens tested.

3.2.3 Cel5H Enzyme characterization

It was shown that a mixture of different oligo-cellulose could be well separated when these separated by HPLC using the method listed in Table 3. Thus, this cellooligo separation method was used for all further HPLC characterizations of cellooligos. Two oligosaccharides, G5 and G6 were used as substrates to investigate the hydrolytic pattern of Cel5H, which is classified as a putative endoglucanase that hydrolyzes the β-1,4 glucosidic bond between glucose unit in cellulose and cello-oligos such as G5 and G6. The final goal for the Cel5H hydrolysis experiment was to identify which end preference Cel5H had, i.e. acting from the reducing or the non-reducing end of the cellooligo, if the enzyme had some endo-processive activity, or if the enzyme is a pure endoglucanase.

The cleavage pattern of the cellooligo G5 was investigated to determine the right experiment conditions for Cel5H and the cellooligos since this ligand is relatively cheap, and sufficient amounts of G5 did exist in the lab. The G5 cellooligo ligand was also used to determine if the Cel5H had a glucan chain end preference, i.e. reducing or non-reducing end preference, for its hydrolytic activity. It was already known, based on homology to other glycoside hydrolase family 5 enzymes, that this enzyme has a retaining mechanism. Based on this knowledge there exist two possible scenarios after treating G5 with Cel5H, these possible scenarios are depending on from which end Cel5H starts to digest the ligand. If the enzyme starts, or has a preference for, to cleave off cellobiose units from the reducing end of G5, then the product cellobiose will contains a normal ratio of α-, β-configuration of its anomeric carbon at the reducing end of it (36% and 64% respectively), and the product cellotriose will have an increased amount of molecules with β-configuration of the anomeric carbon, compared to the amount of molecules with α- configuration. If the enzyme instead cuts off cellobiose units from the non-reducing end of G5, then the scenario will be opposite, then will the product cellotriose have a normal ratio of α-, β-configuration (36% and 64% respectively), and the product cellobiose will have an increased amount of molecules with β-configuration.

If one instead uses G6 for hydrolysis studies with Cel5H then there exist four possible scenarios:

*If the enzyme cleaves off cellobiose from the non-reducing end of G6 (Figure 6, solution 1), then will the product cellobioses have an increased amount of molecules with β-configuration of the anomeric carbon.

* If the enzyme cleaves off cellobiose from reducing end (Figure 6, solution 3), then will the product cellotetraoses have an increased amount of molecules with β-configuration of the anomeric carbon.

*If Cel5H is an endoglucanase with no chain end preference (Figure 6, solution 2 and 2', which differ in from which end the hydrolysis proceeds), the most common product will be cellotrioses.

By analyzing the ratio of the β- and the α- anomers of the products G4 or G2，it should be possible to determine from which end of a glucan chain Cel5H preferably acts. The

processivity could be determined by comparing the relative amount of G2 and G4 during hydrolysis.
Figure 6 shows the possible cleavage patterns, when using G6 as substrate for Cel5H.

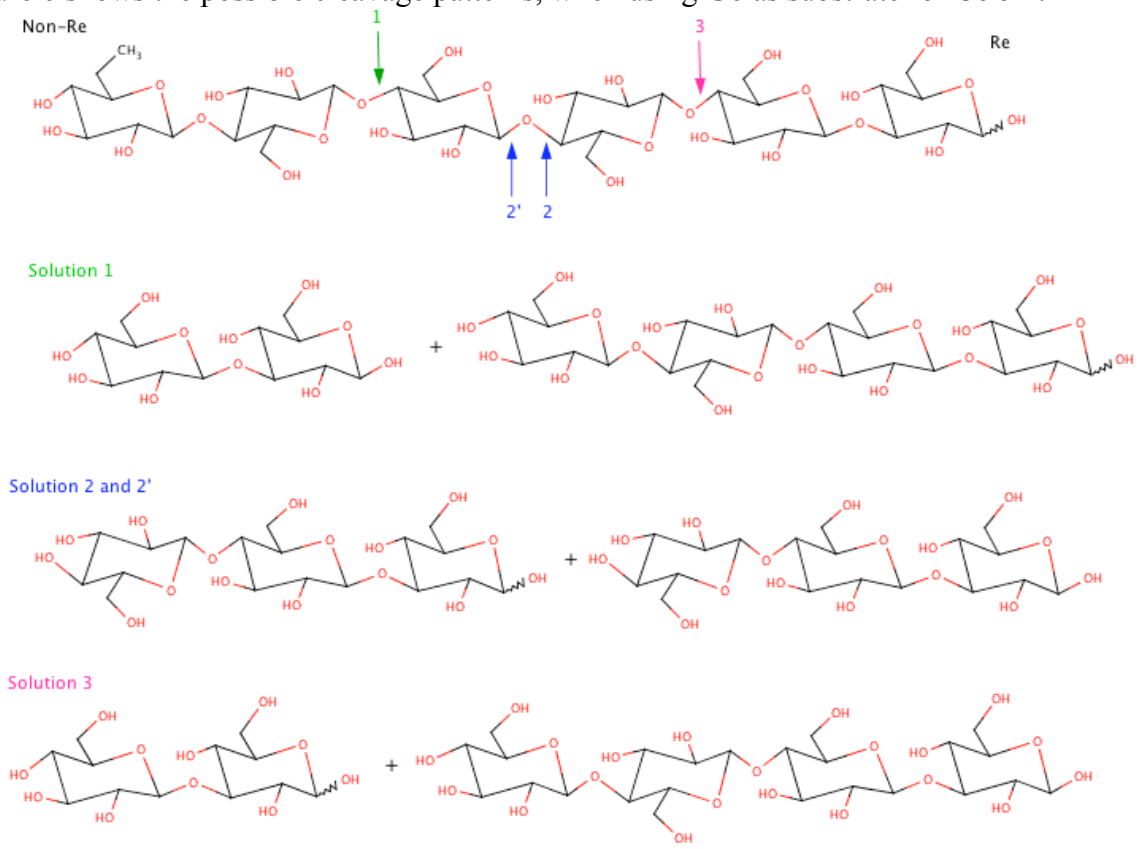


Figure 6. The possible cleavage pattern scenarios when incubating the ligand G6 with Cel5H. There are four possible scenarios: solution 2 and 2' will be true if there are increased amounts of G3 in the product. Solution 1 and 3 can be proven by studying the anomeric conformation of the newly formed oligo-saccharide.

The reaction rate of hydrolysis of G5 was much slower than hydrolysis of G6 if the same molar of Cel5H and substrate cellooligos were used. Even increase the amount of enzyme Cel5H in the hydrolysis of G5, G5 could not be totally consumed after 72 hours, the result could only show G2 and G3 were the hydrolates. The resolution of the peaks of G2 and G3 was not high enough to tell the conformation of the newly formed cellooligos. In spite of this, the hydrolysis of G5 could still prove that Cel5H has endoglucanase activity.

The hydrolysis of G6 was successfully carried out. Figure 7 and 8 shows the variation of amount of the hydrolates and substrate G6 in different time intervals: 10-30 min in figure 7 and 1-24hour in Figure 8. Both figure show the increase of G3, which meet with the scenario 3 that Cel5H is an endoglucanase (Figure 6, solution 2 and 2'). Cellooligos G2 and G4 increased during the process of hydrolysis, this proves that Cel5H cleaves both from the end and in the middle of the cellulose chain. If the cutting is in the middle of a cellooligo (figure 6, solution 2 or 2'), one G6 will be converted to two G3 molecules; if cutting is from the end of a cellooligo (Figure 6, solution 1 or 3), one G6 will give out one G2 molecule and one G4 molecule. By comparing the half amount G3 and the amount of G2, the preferred position of cleavage can be studied. In Figure 7 and 8, the half amount of G3 was close to the amount of G2 in every check point, so the enzyme Cel5H does not have preference on choosing the cleavage site.

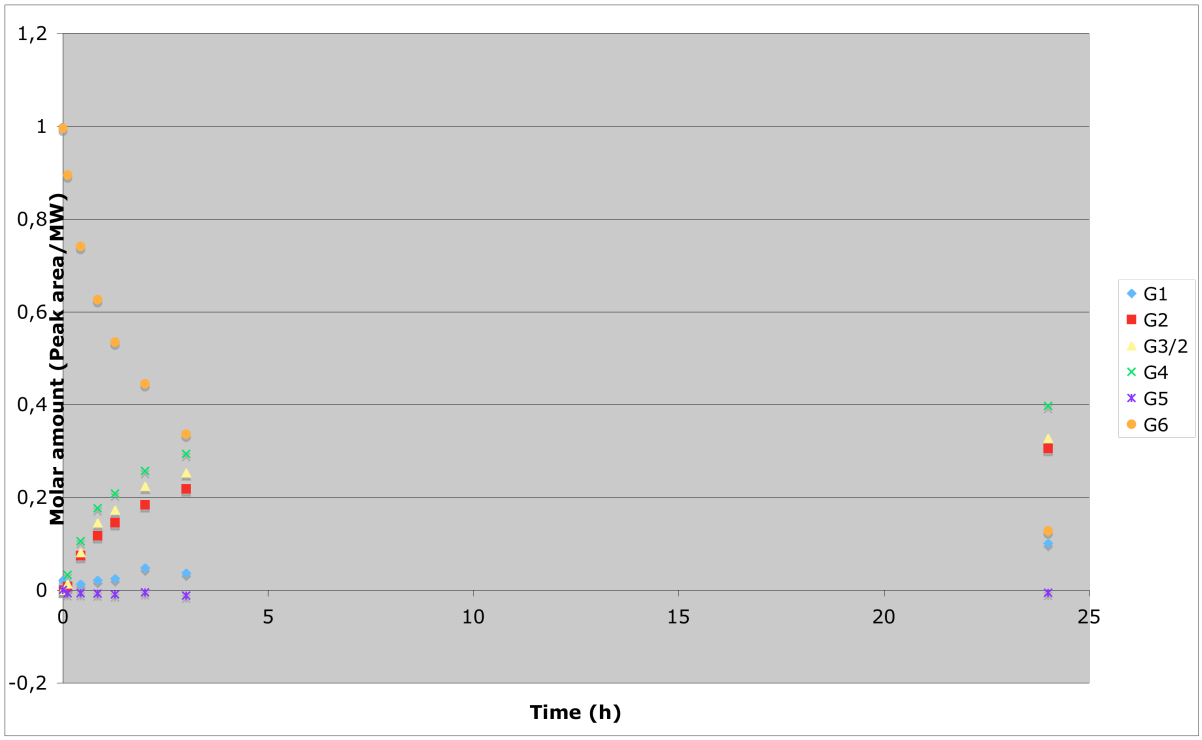


Figure 7. Products and substrate after hydrolysis of 50µM G6 by 1,9nM Cel5H catalytic domain. The injection interval in the first 3 hours was 30 min. The last injection was after one day.

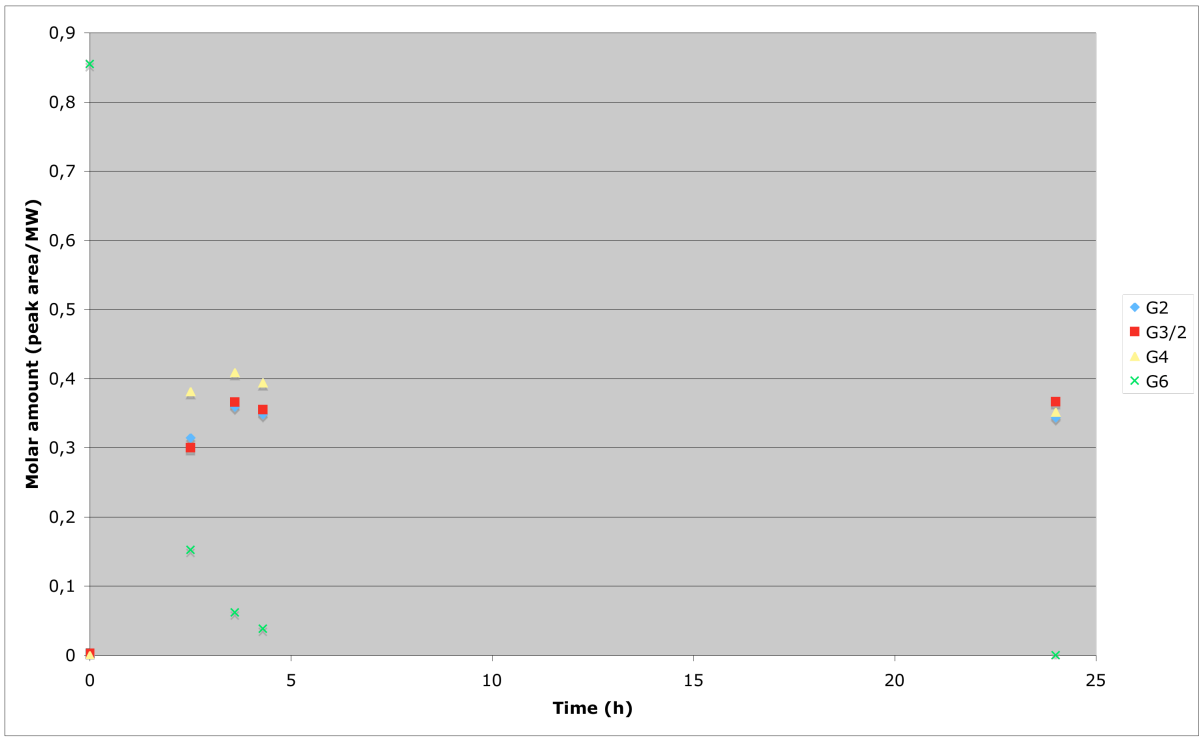
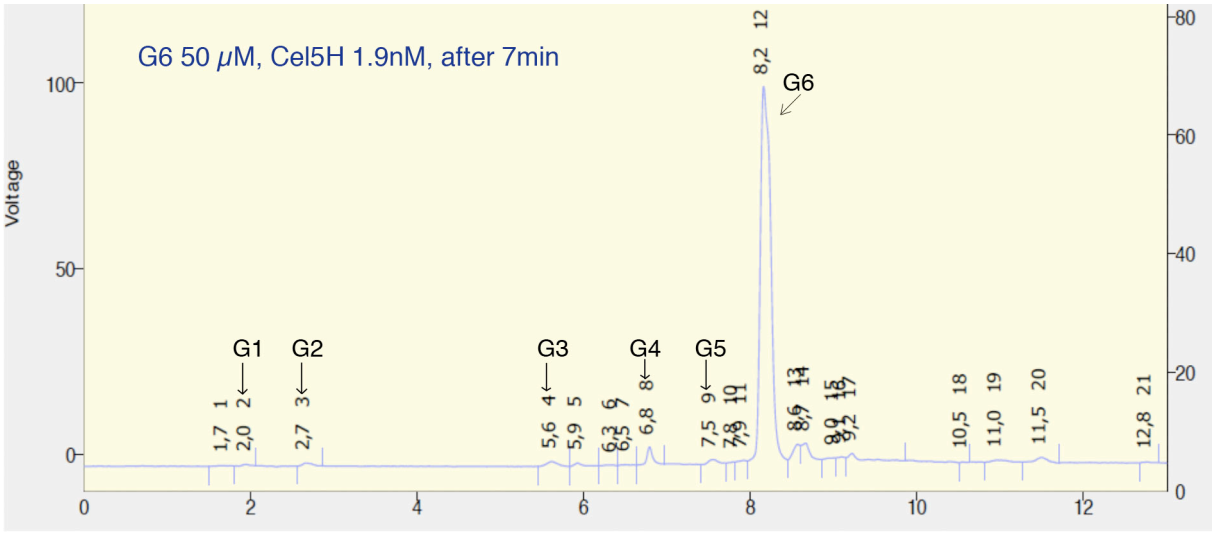
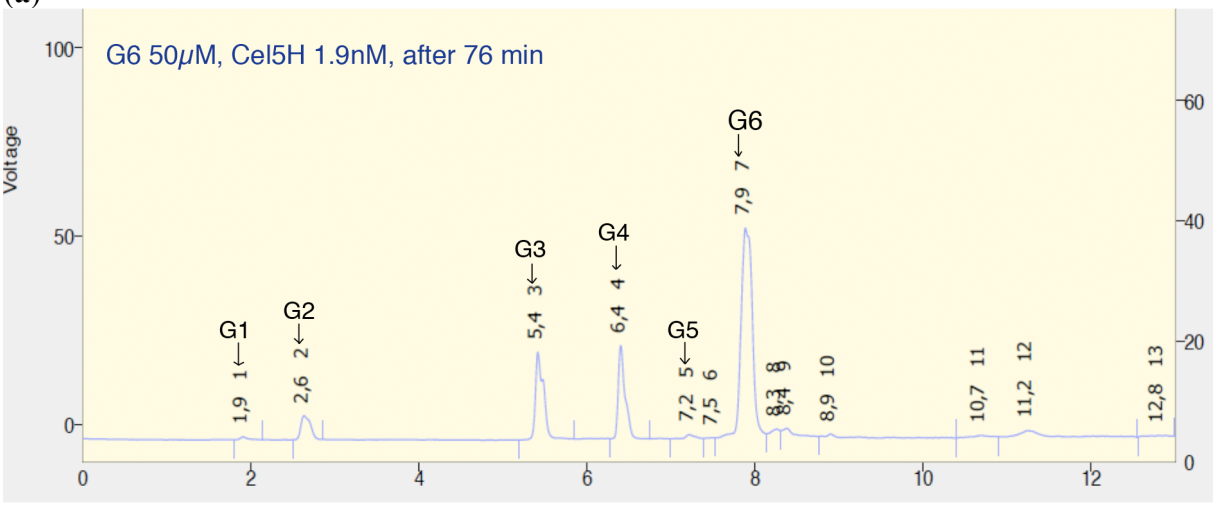


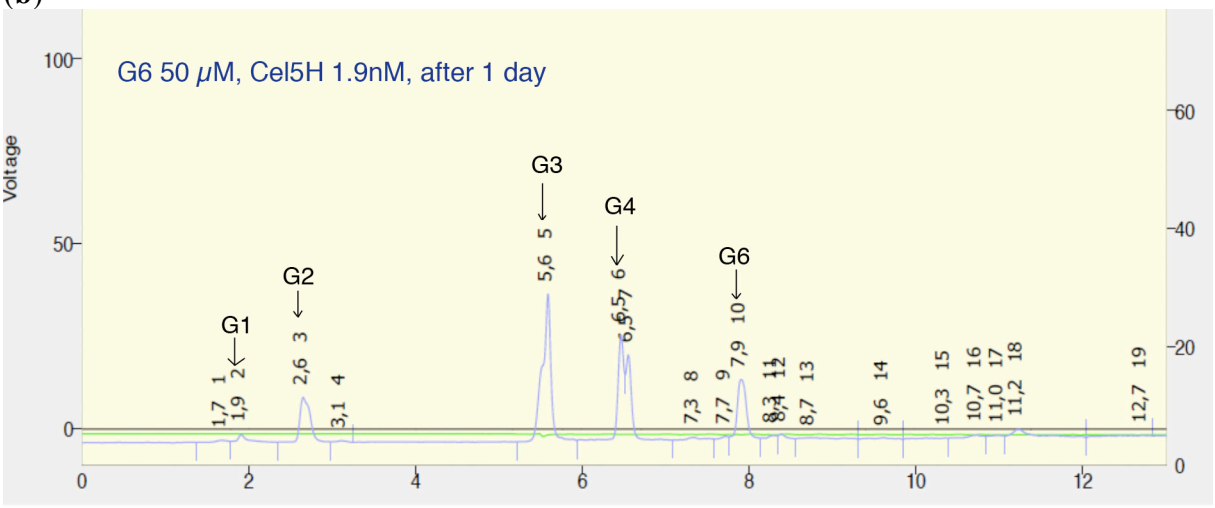
Figure 8. Products and substrate after hydrolysis of 50µM G6 by 1.52nM Cel5H catalytic domain.



(**a**)



(**b**)



(**c**)

Figure 9. Chromatograms from HPLC. The reaction condition was 50 μM G6 digested by 1.9nM Cel5H. (**a**) is after 7 min, (**b**) is after 76min and (**c**) is after 24 hours

The HPLC chromatogram shows not only the amount of each product, and also rough ratio of α and β anomers. Throughout the whole reaction, G2 had the same ratio between the two anomeric configurations, while fore the products G3 and G4 the β anomer configuration was

mote than the α anomer (the ligand with β configuration is eluted earlier then than with α) in the beginning when the reaction rate was fast (figure 9, a and b). After some time, the two anomeric configurations converted to reach the equilibrium state, which comprises 40% α configuration and 60% β configuration (figure 9, c). This phenomenon indicates that Cel5H starts cutting from the reducing end of the cellulose chain, thus, solution 2 and 3 in Figure 6 are proven to be right.

The processivity of Cel5H could not be visualized directly from either the reaction curve or chromatogram. If Cel5H would be a processive enzyme, then should the enzyme keep attackin on the same chain after the first initial cut on this cellulose chain. So a formula for how Cel5H would hydrolyze G6 if this enzyme was processive would be:

G6 → G4 + G2 → 3 ×G2

This means the molar amount of G4 should be equal to or lower than G2, and at the end of the reaction, there should be no G4 at all. However, the amount of G4 was always higher than G2. Therefore, given the data achieved within this study there is no enough evidence to state if Cel5H is an endo-processive enzyme or not.

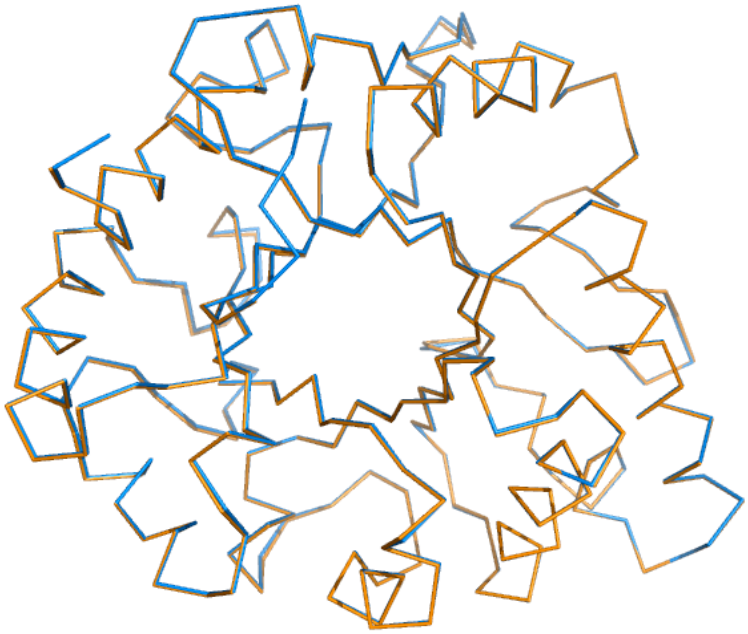
4 Discussion

4.1 EndoT

All eight ligands listed in Table 2 have within this study been tried in co-crystallize with the protein or soaked into crystals, and all the datasets for complexes with a good diffraction were collected in the lab on a home X-ray source or at a synchrotron source at MAX-lab. Except for EndoT-chitobiose, all the datasets were successfully processed and the structures were solved by molecular replacement using the EndoT structure determined in a previous studdy[16] as model. All collected X-ray data-set hade the same space group and the same cell dimension as the data for the initial structure model. Inspection of the elctron density maps of the collected potential ligand complexes revealed that there was no ligand bound at the active site, i.e., Asp 129 and Glu 131[16], in any of the structures. The new EndoT structure without zinc atoms was solved by molecular replacement and refined to a resolution of 1.8 Å. The current R_{free} and R-value for this EndoT model is 21.93 and 17.16 respectively.

When comparing the crystal of the EndoT with bound zinc and the EndoT zinc-free, the crystal of EndoT zinc-free was shorter and wider, although they still have same space group and cell dimension. The structure comparison between the EndoT with bound zinc and the EndoT zinc-free shows very subtle differences (Figure 10). The structure fold for both structures is a $(\alpha/\beta)_8$ β-barrel. There are two NAG residues found attached to the enzyme, one is bound to Asn 70 and the second one to Asn 240. The active site residues consist of Asp 129, Glu 131 and Trp 259. Asp 129 and Glu 131 are the catalytic nucleophile and the proton doner residues respectively. Trp 259 at the active site is the residue that aids the substrate to get a correct orientation in the active site to enable the enzyme to hydrolyse the bond between two glucose residues. The new zinc-free EndoT structure has four additional residues at the N-terminal end (Ala, Glu, Pro and Thr) and one additional (Leu) at C-terminal end (Figure 10) compared to the previously solved EndoT structure. In most cases, especially in hydrophobic regions within protein, the positions of water molecules are conserved between the two structures. The existence of a bound zinc atoms in one of the two structures gives in hand that this structure model has more bound water molecules because each zinc atom coordinates several new water molecules and these are also several more well defined water molecules found in the interface between two symmetric related protein molecules in the EndoT structure with zinc bound to it. There are a total six zinc atoms bound to each one of the two EndoT molecules in the structure model. One of those occupied the active site and the other are at the interface between neighboring protein molecules in the protein crystal. Given that the zinc in the active site may hinder the substrate molecule, initiated the trials to get EndoT crystals without zinc added to the crystallization solution. In the EndoT structure with a zinc bound in the active site, the side chain of Glu 131 points directly to the zinc (2.08 Å) while the same residue in EndoT without zinc the side chain of Glu 131 has moved away slightly (2.67 Å away from the position of the zinc, Figure 11). There is no obvious difference for Asp 129 and Trp 259 between the structure of EndoT with and without bound zinc. Water molecules in this region are not in the same position with each other. Strong density was found around zinc displaying a “Y” shape and this appears also around the other zinc atoms. Considering that ZnAc and NaAc were used as crystallizing buffer, it may be an acetate ion bound at this position. This was fitted into the electron density and it fitted well with the extra electron density found positioning the active site of the enzyme. The rest of the zinc atoms found bound to the structure were almost all positions close to Asp, Asn or His residues in enzyme, these amino acids has all stronger ability to attract electrons. In all cases the bound zinc atoms coordinates more water molecules and thus make the structure tight and more

stable. This fact could potentially explain why the EndoT structure with zinc has better resolution (1.3 Å) than the structure without zinc (1.8 Å).



Figures 10. Cα cartoon overlay of the EndoT structures wit and without bound zinc. The orange model is the structure with zinc bound to the enzyme, and the blue one is the structure without bound zinc atoms. The zinc-free EndoT structure has more residues at both terminal ends.

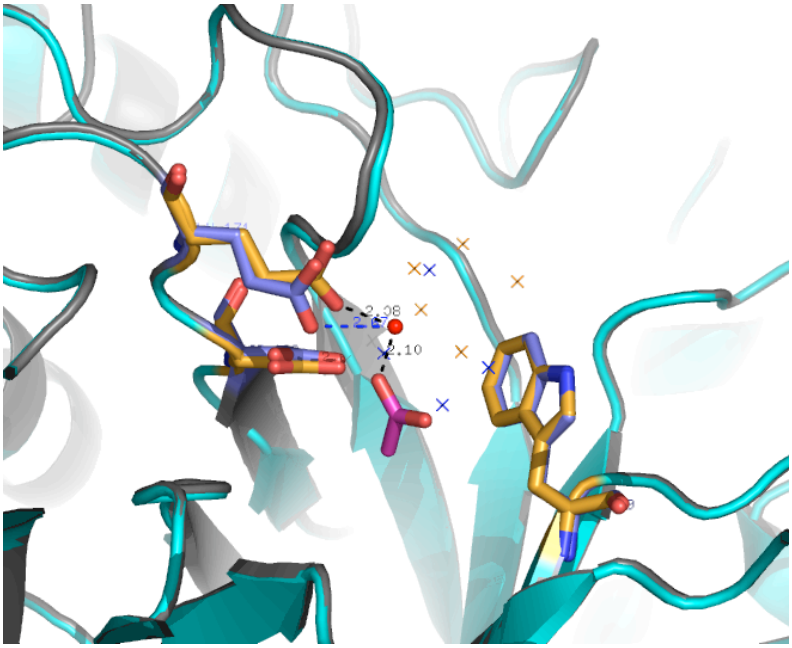


Figure 11. Comparison of two EndoT structures at the active site. Yellow sticks show the structure with bound zinc atoms, and blue sticks show the structure without zinc. The red spot in the middle is zinc atom. The pink sticks is acetate molecule found bound in the active site.

The reason for why a ligands could not bind the protein is unclear. It is possible that the enzyme did not have enough affinity to those ligands tested within this studdy, especially the ligands glucose and cellobiose, which neither one is a substrates for nor a product of EndoT.

Three of the ligands are natural substrates for EndoT. These ligands were soaked into crystals of EndoT at low pH, in which the enzyme is supposed to be inactivated. However, the protein may still have been active at this low pH and thus have digested these substrates. The residual hydrolytic activity of EndoT at the low pH used was never tested.

Potential future new experiment with EndoT could include:
1) Measure binding affinity to all ligands.
2) Enzyme characterization, in order to find out and confirm at which condition the enzyme has lost it hydrolytic activity completely, then use this condition for new enzyme ligand complex crystallization experiments.
3) Prolong soaking time, since the crystal can exist at low pH condition.
4) Use active-site mutants, which bind ligand without hydrolyzing it.
5) Use special ligands, which facilitate binding but blocks further hydrolysis.

4.2 Cel5H

Two crystallization screens, Core 96-JCSG+ Suite and PEG-ION, were used for initial crystallization experiments with the two batches of Cel5H protein. Only one crystal was found growing in these screens, and the condition for this crystal was 0.2M potassium nitrate pH 6.9, 20% PEG 3350. Trials was carried out to improve this found initial crystallization condition, but no new crystals were found in any of these, even after seeding. SDS-PAGE analysis showed that nether of the two Cel5 samples were extremely pure. Other purification method should be tested, such as ion-exchange chromatography, to get a purer enzyme to crystallize. In addition to this, the protein concentration in the sample was lower than the expected one. A good starting point for how concentrated a protein should before setting up crystallization experiments with this is that 35-50% of the drops in the initial tested Core 96-JCSG+ should at least have some tendency of nucleation in them.

The idea with the enzymatic characterization study was to give an overall idea of the cleavage pattern of Cel5H on a cellooligo ligand. This was tested by incubating Cel5H with two cellooligosaccharides G5 and G6, and then analyzing the ratio of the formed and also the anomeric configuration of the formed products (Figure 6). The products were separated with reverse phase HPLC and detected with a CAD detector. The ligand G5 was initially used to find suitable reaction conditions that could be monitored by HPLC, but the results from these experiments were not clear. The hydrolysis experiments with the ligand G6 successfully proved that Cel5H has endoglucanase activity, which was consistent with the result from G5 digestion. The hydrolysis experiments with G6 did also show that Cel5H preferably starts to cleave cellulose chain from the reducing end. This was shown by studying the ratio of the anomeric configuration of the formed products right after the cleavage. It was difficult to conclude if Cel5H is an endoprocessive enzyme or if this enzyme is a pure endoglucanase by looking at reaction curves (figure 7 and 8).

Strictly speaking, a processive enzyme is supposed make successive cleaving events on one chain, which means it should not stop cutting until G6 is totally converted to G2 or G1. So the reaction can be expressed as the following equations if the possible of a transglycosylation reactions is ignored:

$$\mathbf{G6 + E \rightarrow G6 \bullet E \rightarrow} \begin{cases} \mathbf{G4 \bullet E + G2 \rightarrow G2 \bullet E + 2 \times G2 \rightarrow E + 3 \times G2} & \mathbf{(1)} \\ \mathbf{G3 \bullet E + G3 \rightarrow E + 2 \times G3} & \mathbf{(2)} \end{cases}$$

The equation is under the assumption that the enzyme could slowly digest G4 but not G3. The equation (1) shows processive activity, in which the first reaction is much faster than the second (underlined in the equation), because G4 is not the preferred substrate for the enzyme if compared with G6. In the case of that Cel5H would be a processive enzyme, at the beginning, the position of the cleavage is random. After the first cleavage event, if the enzyme is processive, then should the enzyme stay bound to the glucan chain and cleave the newly formed substrate G4 in a second cleavage event instead of seeking for free G6, meanwhile some free enzyme molecules will associate with G3. This will mean that there will be less free enzyme in the mixture than before, which will results in a decrease of the reaction rate. And because G4 is degraded to G2 if the enzyme is processive, then we should see a decrease of G4 and increase of G2 after G6 has been consumed. A good sign for that G6 is being depleted in the reaction mixture is that G3 is no longer being accumulated in the sample.

The aim of the experiment in which 50μM of G6 was digested with 1.52nM enzyme was to find out what would happen after that all G6 had been consumed. In figure 8 we can estimate that G6 was depleted after approximately 5 hours. In the next 20 hours, the amount of G4 went slightly down to the same level as G2, G3 had a slightly increase, and the levels of G2 kept the same. These results fail to meet with our theory that degradation of G4 to G2 causes the increase of G2 and decrease of G4, if the enzyme Cel5H is processive. Further more, the change of each product was too tiny to draw any conclusion. The reason for this is probably due to that the CAD detector was not sensitive enough to detect the amounts of G2 formed., G2 has a smaller total mass than the mass of equivmolar amount of other longer polymeric sugars has. The hydrolytic experiment carried out with Cel5H has to be repeated to get data that is statistically significant enough to give any insights about the hydrolytic action of Cel5H. The experiments with G5 fail to give good results, this was mainly due to that the G5 standard had trace of G4 and G3 in it. These trace amounts of G3 and G4 may greatly confuse the ratio between the two anomeric configurations of the newly formed G3 molecules. The second reason for the bad results with G5 was that the reaction rate with this ligand was too slow, so the change of the anomeric configuration of the newly formed product is much harder to detect.

To improve the characterization part of the stud, one could initially carry out the hydrolytic experiments in a proper buffer and thereby enhance the hydrolytic reaction rate of the enzyme. Using a buffer in the reaction mixture would also have the effect that the enzyme is stabilized, so that it is possible to repeat the experiment. Secondly, longer oligo-celluloses should be used to check the processivity of Cel5H, since Cel5H has much lower specificity for G4 and G5. Thirdly, increase the injection volume in the HPLC characterizations. In the c HPLC experiment carried out in this study, a manual injection was always used, and which all had a total sample volume of 25 μl. If an auto-injection system would have been used, the volume could have been increased to 100 μl. If the same concentration of enzyme and substrate would be used in these experiments as for the experiments carried out in this study, then the amount of each product would be larger than the amounts of products analyzed by The HPLC runs in this study. Having larger amounts of products in the experiment samples would give more reliable and reproducible values.

Although the crystallization of Cel5H failed to give any positive result, the enzyme has been proven to have endoglucanase activity. In the future study, it would be interesting to study the kinetic parameters of Cel5H in greater detail and compare these with the ones of the activity with other endoglucanases from the same Glycoside hydrolase family.

5 Acknowledgement

It is a great honor to express my gratitude to all those who gave me the possibility to complete this thesis. Firstly I want to thank my supervisor Jerry Ståhlberg and Mats Sandgren, I would not have chance to start this project without your admission. I owe my deeply gratitude to my mentor Saeid Karkehabadi, who supported my project in many ways. He helped me finish the plan, taught me basic principles of crystallography, and cheered me up when I was depressed by failure. Without him, I still have no idea of structural biology. I would also like to thank Nils Mikkelsen who taught me how to process the data, and Henrik Hansson who taught me to operate the HPLC, and Äkta purifier system. Finally, I want to thank all the other lab colleagues for your encouragement, suggestions and also jokes that accompanied me all through this project.

6 Reference

1. Gray, K. A.; Zhao, L.; Emptage, M., Bioethanol. *Curr Opin Chem Biol* 2006, *10* (2), 141-6.

2. Klemm, D.; Heublein, B.; Fink, H. P.; Bohn, A., Cellulose: fascinating biopolymer and sustainable raw material. *Angew Chem Int Ed Engl* 2005, *44* (22), 3358-93.

3. Perez, J.; Munoz-Dorado, J.; de la Rubia, T.; Martinez, J., Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *Int Microbiol* 2002, *5* (2), 53-63.

4. Northcot.Dh, Chemistry of Plant-Cell Wall. *Annual Review of Plant Physiology* 1972, *23*, 113-&.

5. Cosgrove, D. J., Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology* 2005, *6* (11), 850-861.

6. Bayer, E. A.; Belaich, J. P.; Shoham, Y.; Lamed, R., The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology* 2004, *58*, 521-554.

7. Lee, J., Biological conversion of lignocellulosic biomass to ethanol. *J Biotechnol* 1997, *56* (1), 1-24.

8. Rabinovich, M. I.; Pinto, R. D.; Abarbanel, H. D.; Tumer, E.; Stiesberg, G.; Huerta, R.; Selverston, A. I., Recovery of hidden information through synaptic dynamics. *Network* 2002, *13* (4), 487-501.

9. Jayani, R. S.; Saxena, S.; Gupta, R., Microbial pectinolytic enzymes: A review. *Process Biochemistry* 2005, *40* (9), 2931-2944.

10. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 2009, *37* (Database issue), D233-8.

11. Reese, E. T., History of the cellulase program at the U.S. army Natick Development Center. *Biotechnol Bioeng Symp* 1976, (6), 9-20.

12. Martinez, D.; Berka, R. M.; Henrissat, B.; Saloheimo, M.; Arvas, M.; Baker, S. E.; Chapman, J.; Chertkov, O.; Coutinho, P. M.; Cullen, D.; Danchin, E. G.; Grigoriev, I. V.; Harris, P.; Jackson, M.; Kubicek, C. P.; Han, C. S.; Ho, I.; Larrondo, L. F.; de Leon, A. L.; Magnuson, J. K.; Merino, S.; Misra, M.; Nelson, B.; Putnam, N.; Robbertse, B.; Salamov, A. A.; Schmoll, M.; Terry, A.; Thayer, N.; Westerholm-Parvinen, A.; Schoch, C. L.; Yao, J.; Barabote, R.; Nelson, M. A.; Detter, C.; Bruce, D.; Kuske, C. R.; Xie, G.; Richardson, P.; Rokhsar, D. S.; Lucas, S. M.; Rubin, E. M.; Dunn-Coleman, N.; Ward, M.; Brettin, T. S., Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). *Nat Biotechnol* 2008, *26* (5), 553-60.

13. Fujita, K.; Nakatake, R.; Yamabe, K.; Watanabe, A.; Asada, Y.; Takegawa, K., Identification of amino acid residues essential for the substrate specificity of Flavobacterium sp. endo-beta-N-acetylglucosaminidase. *Biosci Biotechnol Biochem* 2001, *65* (7), 1542-8.

14. Rao, V.; Guan, C.; Van Roey, P., Crystal structure of endo-beta-N-acetylglucosaminidase H at 1.9 A resolution: active-site geometry and substrate recognition. *Structure* 1995, *3* (5), 449-57.

15. Hertzberg, M. C., Purification , crystallization and structure determination studies of the enzymes Cel5A, Cel5B and EndoT from *Hypocrea jecorina Bachelor's project, Swedish University of Agricultural Sciences, Uppsala* 2008.

16. Jacobsén, E., Purification, crystallization and structure determinatino of enzymes used for biomass conversion in to fermantable sugars. *Bechelor, Swedish University of Agricultural Sciences, Uppsala* 2008.

17. Andrykovitch, G.; Marx, I., Isolation of a New Polysaccharide-Digesting Bacterium from a Salt Marsh. *Appl Environ Microbiol* 1988, *54* (4), 1061-1062.

18. Ekborg, N. A.; Gonzalez, J. M.; Howard, M. B.; Taylor, L. E.; Hutcheson, S. W.; Weiner, R. M., Saccharophagus degradans gen. nov., sp. nov., a versatile marine degrader of complex polysaccharides. *Int J Syst Evol Microbiol* 2005, *55* (Pt 4), 1545-9.

19. JGI, DOE Joint GEnome Institute. *http://genome.jgi-psf.org/micde/micde.home.html*.

20. Weiner, R. M.; Taylor, L. E., 2nd; Henrissat, B.; Hauser, L.; Land, M.; Coutinho, P. M.; Rancurel, C.; Saunders, E. H.; Longmire, A. G.; Zhang, H.; Bayer, E. A.; Gilbert, H. J.; Larimer, F.; Zhulin, I. B.; Ekborg, N. A.; Lamed, R.; Richardson, P. M.; Borovok, I.; Hutcheson, S., Complete genome sequence of the complex carbohydrate-degrading marine bacterium, Saccharophagus degradans strain 2-40 T. *PLoS Genet* 2008, *4* (5), e1000087.

21. Bergfors, T., Protein Crystallization: Techniques, Strategies, and Tips. 1999.
22. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004, *60* (Pt 12 Pt 1), 2126-32.
23. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997, *53* (Pt 3), 240-55.
24. ESA Biosciences, I., Corona CAD. *http://coronacad.com/corona.htm*.
25. Leslie, A. G. W., Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* 1992, *No. 26*.
26. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 1994, *50* (Pt 5), 760-3.
27. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J Appl Crystallogr* 2007, *40* (Pt 4), 658-674.
28. DeLano, W. L., The PyMOL Molecular Graphics system. *DaLano Scientific, Palo Alto, CA, USA* 2002.