#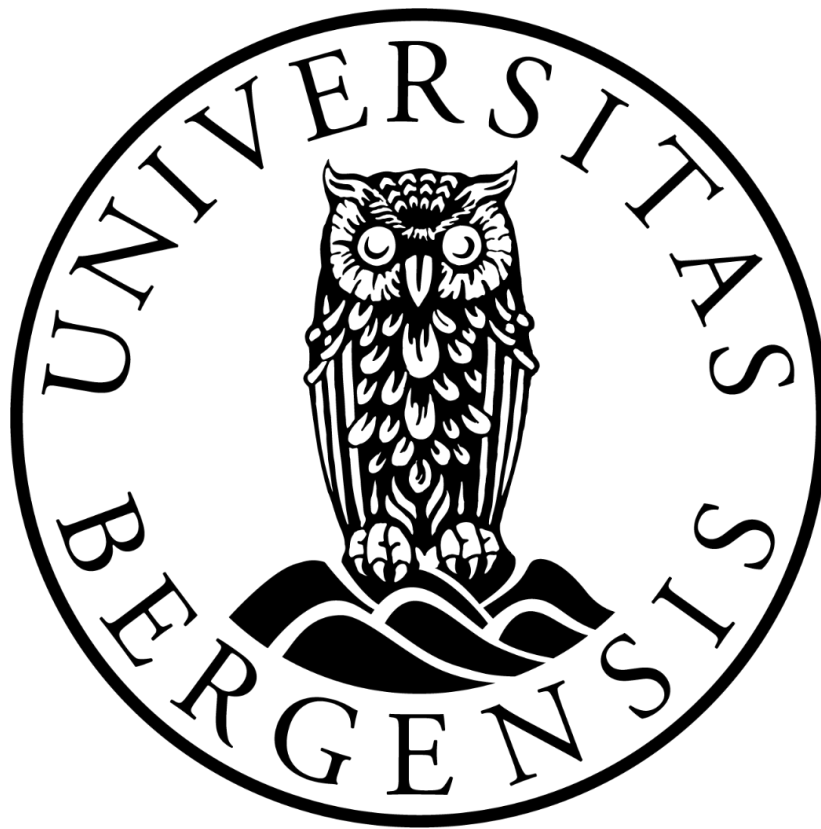 Metagenomics-based discovery of novel carbohydrate degrading enzymes from deep-sea hydrothermal *in situ* enrichments; expression, purification and characterization

Master of Science in Biology - Microbiology
Sondre Olai Spjeld

Supervisors:

Researcher Runar Stokke – Department of Biological sciences, University of Bergen

Prof. Ida Helene Steen – Department of Biological sciences, University of Bergen

Centre for Deep Sea research
Department of Biological Sciences
The University of Bergen
June 1st, 2021

Acknowledgements

Abstract

Deep sea hydrothermal vents located along ocean ridges harbor microorganisms with unique adaptations. These extremophiles have adapted to high temperature and extreme conditions. Carbohydrate acting enzymes with thermostable properties are an increasing interest for industrial usage, as their extreme nature enable processes to be performed more efficiently, lower contamination and overall cost. In this study we work with a metagenome that was retrieved from a hydrothermal vent located at Bruse vent site, and then annotated for carbohydrate active enzyme function. From the metagenome 34 sequences were annotated with Carbohydrate Binding Module 9 (CBM9) but lacking enzyme classes. These were further analyzed to determine if they contained signal peptide, transmembrane helices, and two putative xylanases were chosen for further analysis. Two putative xylanases and five annotated α-amylase were placed in pET21a vector, transformed and expressed in *E. coli* with heterologous protein expression. The target proteins were purified with histidine affinity tag, and enzyme activity was tested through plate assay with xylan and starch substrate. Neopullulanase from a published isolate Geobacillus sp. 12AMOR1 was also purified with histidine affinity tag, and the oligomeric state was estimated through size exclusion chromatography. The putative xylanases were expressed as inclusion bodies, purified under denatured conditions and dialyzed. No activity was observed when tested on plates containing xylan, incubated at 65°C for 4 hours. Two of the annotated α-amylases were expressed as soluble proteins, purified under denatured conditions and dialyzed. No activity was observed when tested on plates containing starch, incubated at 65°C for 4 hours. Based on SDS-PAGE and gel filtration the neopullanase's oligomeric state was estimated to be a dimer. To achieve active proteins for the putative xylanases we suggest the use of fusion partners to obtain soluble protein, for the α-amylases purification with new histidine tag or ion exchange chromatography to test and confirm the activity of these enzymes.

Table of contents

Abbreviations

AMOR – the Arctic Mid-Ocean Ridge
Amp – Ampicillin
BLAST – Basic Local Alignment Search Tool
CAZy – Carbohydrate-acting enzymes
CBM – Carbohydrate binding module
GH – Glycoside hydrolase
His-tag – Polyhistidine
IPTG – Isopropyl β- d-1-thiogalactopyranoside
mM – Millimolar
NCBI – The National Center for Biotechnology Information
RPM – Revolutions per minute
Rfc – Relative centrifugal force
SDS-PAGE – sodium dodecyl sulphate–polyacrylamide gel electrophoresis

1.Introduction

1.1 Bioprospecting of deeps-sea hydrothermal vents.

Deep-sea hydrothermal vents are volcanic vents known as black smoker located along the ridges where tectonic activity occur (Figure 1). These hydrothermal vents are known for their tall chimneys with a plum of black smoke exiting these vents (Dick, 2019). Fractures around these vents allow seawater to seep down in the sediment, where deep enough its super-heated and pushed out trough channels. Down in the sediment the water is loaded with minerals and dissolved gasses that is carried out with the water flow. This creates a unique geochemical profile when the mineral rich water is mixed with cold ocean water. Water exiting these chimneys can reach a temperature of 400°C due to the high pressure while temperature around the chimneys are cold (Webber et al., 2015). Around the chimneys the temperature and pH gradient are step. The sediment surrounding these vents with are also heated from below by the warm water exiting the vents and seeping outside.

Microorganisms that live in environments with extremes of high or low, temperature, pH, salinity or pressure are known as extremophiles. These microorganisms have adaptation and mechanisms that ensure survival in these conditions. High temperature denatures mesophilic enzymes as the heat breaks the bonds in the proteins. Thermophiles (45°C<) and hyperthermophiles (80°C<) have adaptions that stabilizes the proteins to avoid denaturing and remaining active. Acidophiles (pH < 5.5) and alkaliphile (pH > 8) maintain a neutral internal pH inside the cell compared to the environment their live inn (Madigan et al., 2019). Extremophiles can have multiple

adaptations, depending on the environment they are located in, with conditions in deep sea hydrothermal vents with high pressure, low pH and high temperatures are some extreme conditions.
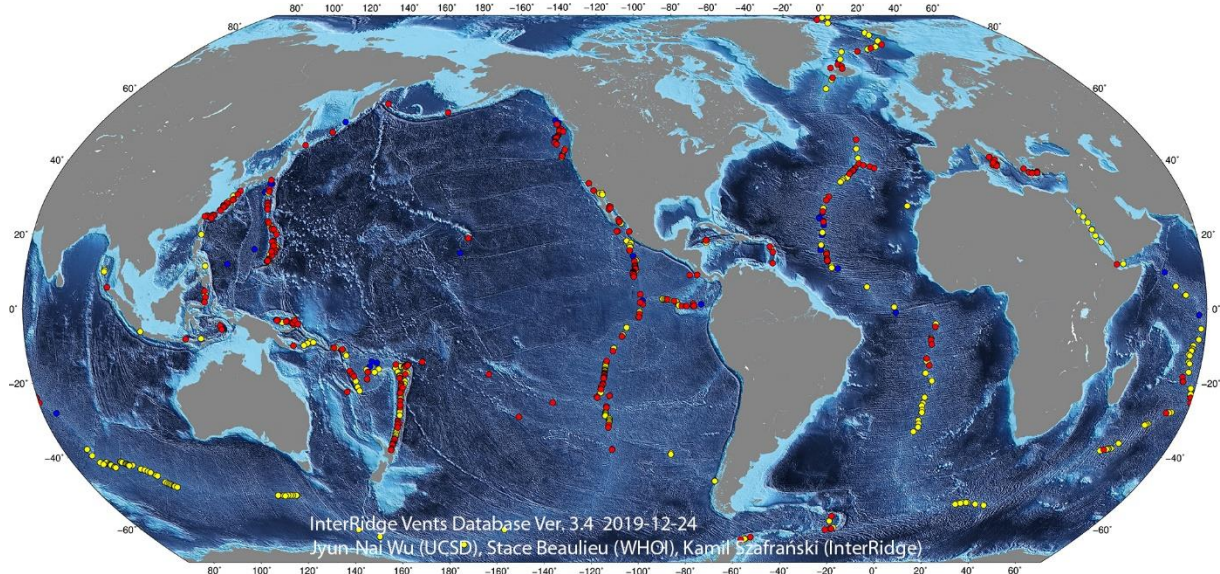
Figure 1 Map displaying the distribution of over 700 hydrothermal vents. Confirmed active vents (Red circles), active inferred (yellow circles) and extinct vents (Blue circles). Retrieved from: (Beaulieu et al., 2020)

Properties that these microorganisms harbor makes them attractive targets in biotechnology. Biotechnology is a technology based on biology and is applied various industrial processes or be utilized for scientific purposes (Madigan et al., 2019). With genetic engineering microorganisms can be manipulated to produce metabolites or other processes that can be harvested (Madigan et al., 2019).

Especially, enzymes of these extremophilic microorganisms have received much interest from Industry as their ability withstand extreme conditions and their cellular components adaptations. Applications and usage from pharmaceuticals, agriculture, pulp and paper industry to food and beverage (Demain & Adrio, 2007). Food beverage market has an increasing interest as processes that can be done at a higher temperature increases the efficiency, reduces amount of enzymes needed and therefore the overall cost. Enzymes with a higher optimum temperature lower the possibly of contaminant in the process (Sarmiento et al., 2015).

Metagenomics is the sequencing and analysis of the total genetic material from an environmental sample (Kodzius & Gojobori, 2015). This has been a well-recognized method to obtain genetic samples from extremophiles, as this eliminate the need to isolate and cultivate the microbes, as replicating their conditions and locating the microbes can be difficult (DeCastro et al., 2016). Metagenomics can be a starting point when bioprospecting for novel enzymes, as sample sites chosen reflects the properties wanted from the enzyme. Bioprospecting is the process of looking biological entities that can be used and improved for commercial use. With this approach for finding novel thermostable enzymes, the conditions from environment should reflect these properties.

With a metagenomic approach, the first step is taking the environmental sample from the source of interest. Sample is prepared and purified from contaminants,

DNA extraction ensures that all available DNA content in the sample is available and is achieved by cell lysis. Important to consider is factors that can hinder cell lysis and increase stability.

Cell lysis can be broken down in 4 categories, mechanical including bead beating and sonication, thermal with freezing and boiling, chemical with SDS and enzymatic with lysozyme. Each category has multiple way of rupturing cells, these mention are just a few examples (Felczykowska et al., 2015). The goal is to ensure that all cells burst and release their genetic material, a combination of different types of cell lysis can be combined be utilized.

Sequencing allows for the biological genetic information to be translated to digital information that can be interpreted and analyzed further, and advancements here is a major driving force that has allowed the advancement of metagenomics to happened. Illumina sequencing and next generation shotgun sequencing are methods that splits the metagenome into fragments that are sequenced simultaneously (Chiu & Miller, 2019). Overlapping fragments are assembled into continuous sequences (contigs). Predicted metagenome-assembled genomes (MAGs) are sorted in bins, and open reading frames (ORF) can be annotated (Kieser et al., 2020). Taxonomic annotation with homology search in databases used to map out the organisms in the sample and functional annotation to map out the activity present in the metagenome (Tamames et al., 2019).

Heterologous protein expression is a powerful tool in biochemistry, that gives the ability to utilize the machinery of the host cell to produce proteins of interest. Heterologous protein expression is the introduction of complementary DNA or RNA encoding for a protein from one species and expressed in another host species (Gagnon, 2010).

Commonly used in protein studies such as bioprospecting to obtain functional proteins of interest. Different organisms are used in heterologous protein expression and have different strengths and weaknesses, unicellular algae, yeast and bacteria are some common examples. *E. coli* is a widely used host organisms because of fast growth, low maintenance, cheap and huge variety of competent strains specialized for different expression (Rosano & Ceccarelli, 2014). Transferring gen of interest to *E. coli* is achieved with a vector contain the gene with other regulatory factors, selective factors and markers. pET-21a vector is a widely used expression vector used in *E. coli* that contains ampicillin resistance for selection[1], T7 promotor that regulates target gene expression and a His-tag that tags the expressed protein for protein purification. Vectors can be inserted in the host cell by transformation, the procaryotic mechanism for absorption of environmental DNA. Heat shock or electric shock are methods to increase the transformation rate.

The goal for heterologous protein expression is to obtain a large amount of functional correct folded protein from the gene of interest. Several factors can influence the success rate of protein expression. Extracellular proteins from prokaryotic origin contains a signal peptide, which is an N terminal peptide that helps export the protein. There are 3 known types of signal peptide. The first is signal peptide that is driven by

---

[1] https://www.merckmillipore.com/NO/en/product/pET-11a-DNA-Novagen,EMD_BIO-69436#documentation

ATP hydrolysis to export the unfolded protein out of the cell where the signal peptide is then removed. The second Sec/SPII lipoprotein signal peptide is transported by the Sec translocon and cleaved by Signal Peptidase II. The third Tat/SPI signal peptide is driven by the proton motive force and exported partly folded proteins that contains cofactors (Stahl et al., 2015). To avoid protein loss during heterologous protein expression these signal peptides are removed before synthesized in a vector. Membrane associated proteins contains a transmembrane helix, which is also important to check for and consider before doing protein expression, as this could affect the result (Wong et al., 2012).

Codon optimization is an important tool to obtain recombinant proteins by eliminating codon bias. Expression of foreign DNA in a host cell might encounter codon bias if codon optimization is not performed before protein expression. The frequency of codons used in a species is related to the corresponding tRNA and amino acids available during translation and protein expression (Fu et al., 2020). Codon usage vary to different degrees between organisms, and some may utilize different codons for the same amino acids, these rare codon are the limiting factor and can halter translation or protein expression. Codon optimization uses synonymous codons to avoid rare codons from the foreign DNA in the host organism and increasing protein expression (Mauro & Chappell, 2014).

1.2 Carbohydrate-acting enzymes in biotechnology

Carbohydrate-acting enzymes (CAZymes) have important metabolic roles in all forms of life and widely used in biotechnology such as biorefinery, food, feed, paper, pulp and detergent industry (Chettri et al., 2020; Kirk et al., 2002). CAZymes are divided in classes based on the activity, glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyses (PL), carbohydrate esterases (CE) and auxiliary activities (AA). Associated modules are carbohydrate binding modules (CBM) that binds to the substrate (Corrêa et al., 2020). To this day 88 families of CBM are described and in the CAZy database (Lombard et al., 2014). The two most abundant renewable biomass recourses on earth is cellulose and starch (Mischnick & Momcilovic, 2010). Finding novel enzymes with physicochemical properties to fully utilize the potential of these resources are high interest in biotechnology. Amylolytic are some popular enzymes used in the industry for the conversion of starch to monomer sugars with synorogenic mix of enzymes. Amylases are enzymes capable of this hydrolytic activity and are classified according to the specificity they operate (Fia et al., 2005). Xylan is main carbohydrate in hemicellulose found in plant cell walls (Subramaniyan & Prema, 2002). Xylanase is the enzyme that hydrolysis xylan, two known enzymes have this activity. Endo-1,4- β- xylanase with endohydrolysis of (1-4) β-D-xylosidic linkages in xylan and α-D-xylosidase with hydrolysis of terminal, non-reducing α-D-xylose residues with release of α-D-xylose (Dalmaso et al., 2015; Lombard et al., 2014)

1.3 (Meta)genomics-based enzyme discovery at AMOR

The genes focused on in this project are encoded by microorganisms from the Arctic Mid-Ocean Ridge (AMOR) vent fields. Three vent sites have been discovered 50 km north of the Jan Mayen fracture zone, the Troll Wall, Soria Moria vent site and Bruse vent site (Figure 2. 5) in The Jan Mayen hydrothermal Vent field (JMVF).The Bruse vent site is located 71°18'N, 05°42'W with a depth of 570 meters below the surface (Stokke et al., 2020).
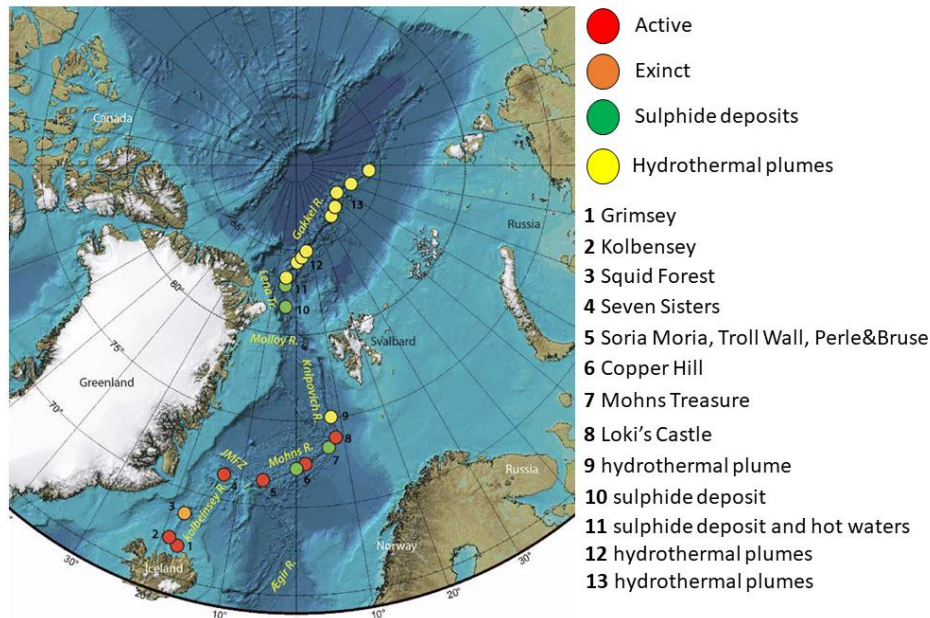
Figure 2 Map of the Artic mid-ocean ridge vent field. Source: Rolf B. Pedersen (Centres for Geobiology and Deep Sea Research, University of Bergen), Terje Bjerkgård (Geological Survey of Norway), 2016, Chapter 5: Sea-Floor Massive Sulphides in the Arctic Waters

In 2014 titanium incubators were placed in the hot sediment surrounding the hydrothermal vents at the Bruse vent site (Stokke et al., 2020) with an ROV (Remotely Operated Vehicle). The incubators had 3 chambers each with a volume of 16 ml arranged on top each other (Figure 3.B). They were filled with 16ml sediment from the vent site and supplemented with 1g substrate. The incubators of interest for this project were filled with 1g of unbleached Norwegian spruce (*Picea abies*) that had been subjected to sulfite pulping pre-treatment from Borregaard AS (Sarpsborg, Norway) termed the BALI process (Rødsrud et al., 2012). The substrate in this incubator had a content of 85% glucan, acid insoluble lignin 8% and hemicelluloses 3% (mannan and xylan) (Fredriksen et al., 2019), and named CGB6. The second incubator used in this project, CGB9, was filled with wheat grains (unpublished results). Both incubators were deployed in the hydrothermal sediment at Bruse with a temperature gradient ranging from 20°C at the sediment surface to 74°C deeper in the sediment, allowing the three chambers of each incubator to be subjected to different temperatures. The incubators were retrieved one year later in July 2015.
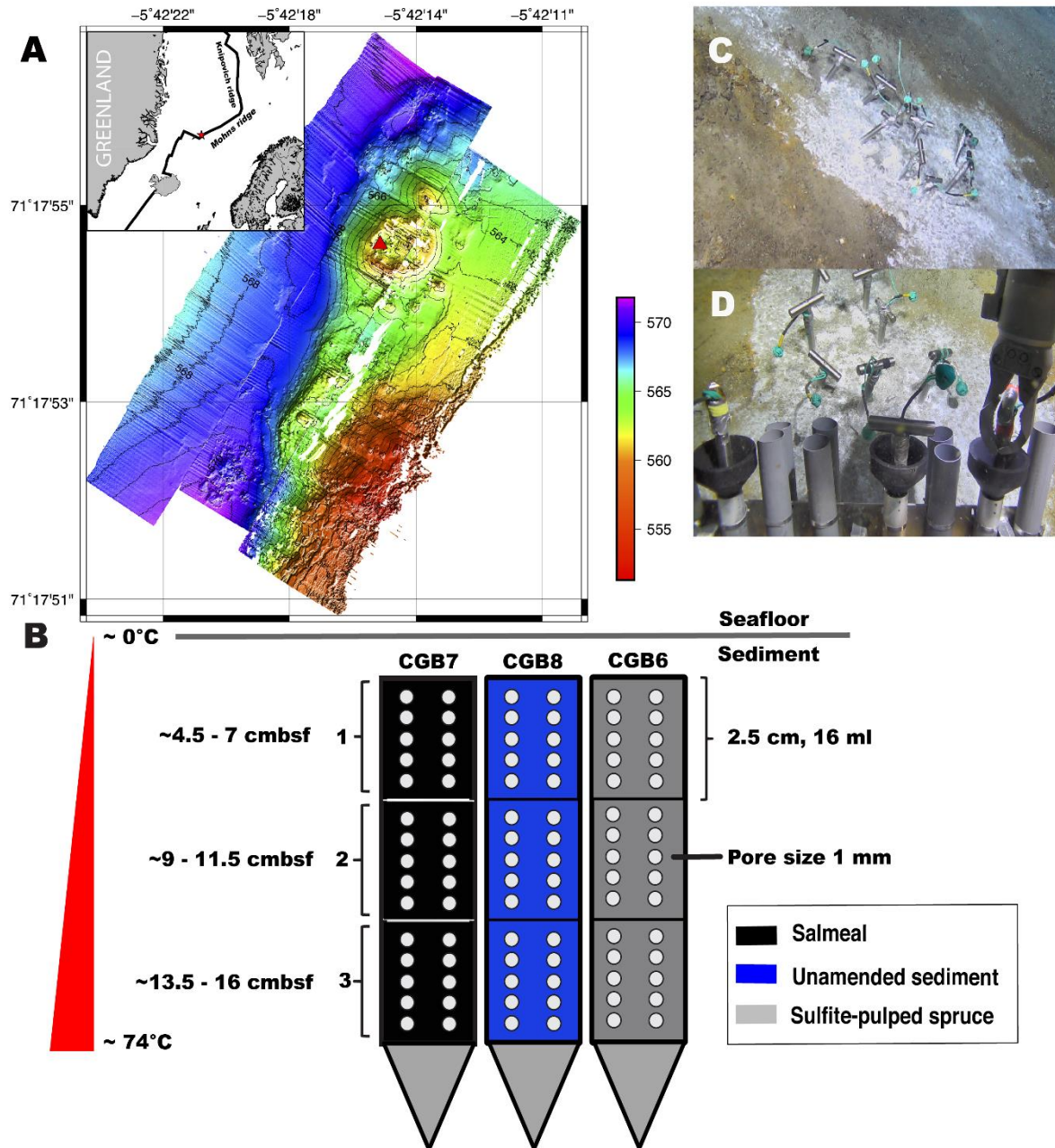
Figure 3 Overview of the in situ incubation performed at Bruse vents site. A) Bathymetry map showing depth of Bruse vent site. B) Schematic of CGB6 incubator with Sulfite-pulped spruce with depth location in the sediment centimeters below sea surface (cmbsf). C) Incubators placement in the hot hydrothermal sediment. D) Remotely operated vehicle used to deploy incubators in the sediment. Source (Stokke el al,. 2020)

Metagenomic sequencing of which in situ incubators and subsequent expression and characterization of enzyme have provided knowledge of 4 unique carbohydrate degrading enzymes (Stokke et al., 2020). Thermostable alginate lyase AMOR_PL7A (Vuoristo et al., 2019), thermostable alginate lyase AMOR_PL17A (Arntzen et al., 2021), hyperthermophilic novel GH9 cellulase (Stepnov et al., 2019), thermostable AMOR-GH10A xylanase (Fredriksen et al., 2019).

1.4 α-amylases GH13 and GH57

Amylase is an class of carbohydrate-degrading enzymes that hydrolyze starch polymers to glucose monomers, α-amylase is the enzyme that cleaves the internal α-1,4 polymer binding in starch (Dalmaso et al., 2015). α-amylases belonging to GH13 is the most characterized of the amylases (Janeček et al., 2014), however, amylases belonging to GH57 often have a thermophilic origin (Jeon et al., 2014). Identifying novel α-amylases is still a focus of interest for new and improved applications in biotechnology. In 2016, Wissuwa *et al.* published the isolation and genome sequence of a thermophilic starch-degrading *Geobacillus* strain, isolated from hot sediments at the Jan Mayen hydrothermal vent site (Wissuwa et al., 2016). In addition, an enzyme from the GH13 family and annotated as a neopullulanase, was expressed and purified (Internal Locus tag; Geob_1869, GenBank Locus tag GARCT_00679; AKM17981). The purified enzyme showed starch-degrading capacity and high thermal stability with a melting temperature of 76.4 °C. To further assess the molecular weight and oligomeric state of the GH13 from *Geobacillus* sp. 12AMOR1, the target gene cloned in the pOPINE_F expression vector was used for expression and purification in *E. coli* BL21. Furthermore, the purified GH13 from *Geobacillus* was used as a control in plate assays together with α-amylase from *Bacillus licheniformis* (Sigma-Aldrich).

From an *in situ* incubator (deployed and collected as for CGB6 described above) supplemented with starch as substrate (CGB9), sequences annotated as GH57 using the dbCAN annotation tool (http://bcb.unl.edu/dbCAN2/); run_dbCAN.py(Zhang et al., 2018), was selected. True amylases of the GH57 family from the dbCAN annotation was identified based on alignments with GH57 from Janecek and Blesak 2011 (Janeček & Blesák, 2011).

1.5 Aim:

This project aims to study biochemical properties of a selection of carbohydrate degrading enzymes originating from vent fields on the Arctic Mid-Ocean Ridge. The following sub-goals were set:

1) Bioinformatic analyses of 34 putative carbohydrate degrading enzymes with an identified carbohydrate binding module 9 (CBM9)
2) Based on this analysis, select 2 genes for cloning and expression trials
3) Purify and verify activity using a simple agar plate enzyme assay in comparison to a published thermostable GH10 xylanase from the Arctic Mid-Ocean Vent Fields - AMOR_GH10A (Fredriksen et al., 2019).
4) Perform expression trials on predicted proteins containing GH57 and do an agar plate enzyme assay to confirm the activity that reflect the protein domain.
5) Purify and estimate the oligomeric state of a neopullulanase encoded by *Geobacillus* sp. 12AMOR1 isolated from the Jan Mayen Vent Field.

2.Materials

2.1 Competent cells

Competent strain of *Escherichia coli (E. coli)* was used. Two types of BL21 (DE3) competent cells (Aligent Technologies, Santa Clara, USA) were used. *E. coli* BL21-Gold (DE3) competent cells, genotype *E. coli* B F- ompT hsdS(r$_B$- m$_b$-) dcm+ Tetr gal l(DE3) endA The. Encodes T7 RNA polymerase under the control of the lacUV5 promotor. The second, *E. coli* BL21-Gold (DE3) pLysS competent cells, genotype *E. coli* B F– ompT hsdS(rB– mB–) dcm+ Tetr gal l(DE3) endA Hte [pLysS Camr]

2.2 Metagenomics-based selection of target enzymes

Based on metagenomic analyses 34 CBM9-encoding protein sequences were provide by my supervisor Dr. Runar Stokke, and this selection was based on the following analyses performed by him. Metagenomic sequencing and assembly for the CGB6 incubator was conducted as previously described (Fredriksen et al. 2019). The metagenomic sequencing and assembly of the wheat grain incubator, CGB9, was conducted similarly (unpublished, Stokke et al.). Open reading frame (ORFs) prediction of all potential genes was performed using Prodigal v.2.6.3 with the *-p meta* option enabled. All potential protein coding genes were analyzed for carbohydrate-active enzymes (CAZymes) and carbohydrate binding modules (CBMs) using the standalone version of the dbCAN annotation tool (http://bcb.unl.edu/dbCAN2/); run_dbCAN.py (Zhang et al., 2018). As implemented in run_dbCAN.py, hmmscan and diamond blast was analyzed against the dbCAN HMMdb v8 (evalue cutoff 1E-5). Extracted hits were analyzed for signal peptides and transmembrane helices using standalone versions of SignalP5.0 (Almagro Armenteros et al., 2019) and TMHMM 2.0 (Krogh et al., 2001; Sonnhammer et al., 1998), respectively. In order to cluster functional groups of glycosyl hydrolases (GHs) from the CGB6 incubator, the protein hits were analyzed with the EFI-ENZYME SIMILARITY TOOL which uses similarity networks (SSN) for visualization of relationships amongst protein sequences https://efi.igb.illinois.edu/efi-est/ (Gerlt et al., 2015). Hence, related proteins are grouped together in iso-functional clusters dependent on the threshold set in the analysis. The EFI tool was run twice. The overall SSN analysis was performed using a similarity cutoff of 35% and visualized in Cytoscape v3.7.2. For this work, the largest node containing the unique signature of CBM9, a xylanase binding module, was selected (Figure 4.A). The second run of the EFI tool was set to a higher cutoff (SSN alignments of 80% identity) to differentiate between possible iso-families within the CBM9 enzymes (Figure 4.B).
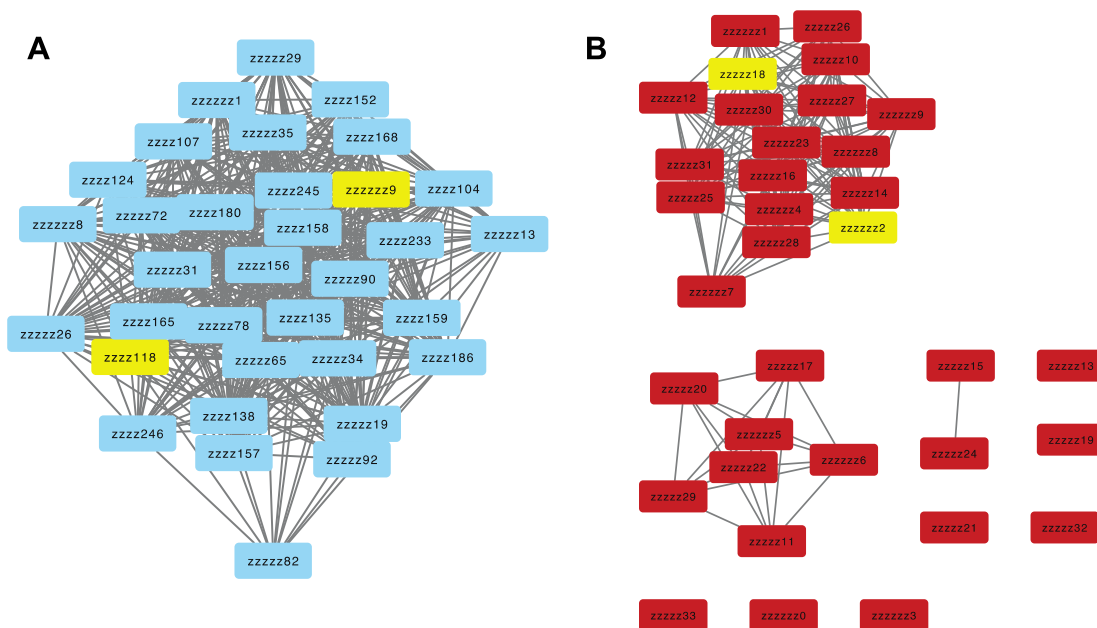
Figure 4. Isolated cluster node comprised of CBM9-containing sequences; A) 35 % identity cutoff and B) 80 % identity cutoff. The final selected sequences synthesized for expression are marked in yellow.

The GH13 from *Geobacillus* sp. 12AMOR1, target gene cloned in the pOPINE_F expression vector was used for expression and purification in *E. coli* BL21 (DE3). Furthermore, the purified GH13 from *Geobacillus* was used as a control in plate assays together with α-amylase from *Bacillus licheniformis* (Sigma-Aldrich).

From the CGB9 incubator (deployed and collected as for CGB6 described above) supplemented with wheat as substrate, sequences annotated as GH57 using the dbCAN annotation tool (http://bcb.unl.edu/dbCAN2/); run_dbCAN.py(Zhang et al., 2018), was provided. True amylases of the GH57 family from the dbCAN annotation was identified based on alignments with GH57 from Janecek and Blesak 2011 (Janeček & Blesák, 2011).

In courtesy from our collaborators at NMBU Ås the characterized and published AMOR_GH10A (Fredriksen et al., 2019) was obtained as transformed colonies and used as control for the plate assays.

3. Methods

3.1 Sequence analysis and selection of putative xylanase CBM9 expression targets

The 34 sequences with a predicted xylanase carbohydrate binding domain were subjected to a set of bioinformatics analyses to provide a knowledge base for selection and evaluation of candidate sequences for cloning, expression, purification and biochemical characterization.

3.1.1 Screening for signal peptides and transmembrane helices (TMHs)

Signal peptides were predicted using both the SignalP5.0 (http://www.cbs.dtu.dk/services/SignalP/) and HMMER (v2.41.1;

https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan) webservers to find location and length on the sequence. Signal-P5.0 (Almagro Armenteros et al., 2019) were performed 3 times for each sequence, one for Gram positive, one for Gram negative and the final for archaea, as these sequences were of unknown prokaryote origin and Signal-P5.0 requires this option to be marked before it can run the prediction. It gives a prediction of what type of signal peptide that is present in the sequence and shows the cleavage site with a given probability.

Sequences that tested negative to containing signal peptide with signal-P5.0, was tested with OutCyte 1.0 server (Zhao et al., 2019). This is a tool that estimates if the proteins contain an unconventional protein secretion system. This will help to distinguish if the protein is internal or exterior (secreted).

In addition, The HMMER web server (V.2.41.1) were used for homology searches against the (UniPortKB) database for each sequence (Potter et al., 2018). The HMMER locates disorders, coiled-coil, transmembrane helix, signal peptide and protein family in the sequences. Each sequence without a transmembrane helix were run in the HMMER web tool; https://www.ebi.ac.uk/Tools/hmmer/search/phmmer. This was done to confirm that the sequences contained the carbohydrate binding domain, length and location of the domain was noted for each of the sequences.

Finally, the online server (TMHMM 2.0), which is an online tool to predict if a protein sequence contains a transmembrane helix (Department of Bio and Health Informatics, 2017, January 5), potential TMHs in the 34 target sequences were analyzed.

3.1.2 *Protein blast against NCBI and UniProt* databases

A functional search of the candidate protein sequences was performed with NCBI blast and UniProt web service to look for similar sequences based on the highest identity percentage value (Consortium, 2018; Coordinators, 2016). Blast against the NCBI database were performed with the blastp (protein-protein BLAST) algorithm (McGinnis & Madden, 2004). Blast results were sorted based on the highest percentage identity against the sequence that was tested, whereby function percentage identity and species was noted. The NCBI blast were run twice against two different databases. The first blast was done with Reference proteins (refseq_protein) database (O'Leary et al., 2016) performed 05.03.2020 and the second was run against the non-redundant protein sequences database performed 24.04.2020. Blast with the UniProt were performed against the UniProtKB database with 10E-threshold auto matrix with none filtering, this was performed 24.04.2020. Blast with UniProt were perform with the same protocol as the NCBI.

3.1.3 Theoretical predicted melting temperature, molecular weight and isoelectric point

The theoretical melting temperature of the sequences was predicted based on the composition of the sequences. Each candidate sequence was run in the melting temperature (T_m) Predictor website (http://tm.life.nthu.edu.tw/index.htm). The output indicates the theoretical T_m of the protein. The predictor estimates based on three

categories of T_m , bellow 55 °C, between 65 to 55 °C and higher than 65 °C, respectively.

Molecular weight was calculated with the online molecular weight calculation form (http://www.bioinformatics.org/sms2/protein_iep.html). This was performed for the purpose of having a potential base line to test produced protein later in the project. The protein Isoelectric point where determined with the Sequence Manipulation Suite online tool (Stothard, 2017, November 6). This tool calculates the theoretical isoelectric point for each protein, these values can be used to locate the proteins in a 2-D gel, and important when to know the pH for neutral net charge for each protein. The sequences were run in bulk with pK values from EMBOSS.

3.1.4 *Sequence alignment*

Global multiple sequence alignments were performed on candidate sequences without transmembrane helix. This was performed with the computer software MEGA-X (version 10.1.7) (Kumar et al., 2018). The alignment was performed using MUSCLE using the default options. The goal was to look for conserved regions in the sequences and selection of targets for expression.

3.1.5 Gene construct synthesis (GenScript)

Two CBM9 sequences were chosen to be ordered for gene synthesis at GenScript; CBM9_30710_4 and CBM9_302_17. To confirm that the sequences matched with their id, a global alignment was performed with EMBOSS NEEDLE, with an expected result of 100% match for both. Before running the needle alignment, the protein sequences need to be translated to nucleotide sequences, or translate the sequences sent to gene synthesis to amino acids, then aligning them in EMBOSS NEEDLE. This is achieved with EMBOSS Transeq (Madeira et al., 2019).The two target sequences were constructed with the pET-21a(+) vector with the cloning site NdeI/BamHI.

3.2 Selected gene targets

In total 9 different protein sequences were used, 8 were expressed and all were attempted protein purified (Table 1). Two CBM sequences with predicted xylanase activity were constructed for this research. 5 amylase sequences that had been constructed but has not been previously described before this research. One Neopullulanase isolated from a *Geobacillus* sp. performed and described in a previous research was further tested here (Wissuwa et al., 2016). One AMOR-GH10A with a known activity against xylanase were used as a positive control when testing activity of the CBM sequences.

Table 1: Overview of working sequences

Name	Working name	Vector	His-tag length and placement	Source
CGB6_2_contig_302_17	AMOR-CBM9-14	pET-21a(+)	C terminal (AHHHHHH)	This work
CGB6_2_contig_30710_4	AMOR-CBM9-15	pET-21a(+)	C terminal (AHHHHHH)	This work
CGB9_3_contig_2109_4	AMOR-GH57-1	pET-21a(+)	N terminal (HHHHHH)	This work
CGB9_3_contig_45_144	AMOR-GH57-2	pET-21a(+)	N terminal (HHHHHH)	This work
CGB9_3_contig_12143_4	AMOR-GH57-3	pET-21a(+)	N terminal (HHHHHH)	This work
CGB9_3_contig_66_86	AMOR-GH57-4	pET-21a(+)	N terminal (HHHHHH)	This work
CGB9_3_contig_16324_2	AMOR-GH57-5	pET-21a(+)	N terminal (HHHHHH)	This work
(MH727997)	AMOR-GH10A	pNIC-CH	C terminal (AHHHHHH)	(Fredriksen et al., 2019)
Neopullulanase (AKM17981)	12AMOR1-GH13	pOPINE_F	C terminal (HHHHHH)	(Wissuwa et al., 2016)

3.3 Buffer preparation

Buffer solutions for cell lysis and protein purification were prepared accordingly. 50mM Hepes (SIGMA-ALDRICH, St. Louis, USA), 300mM NaCl (SIGMA-ALDRICH, St. Louis, USA) were measured, scaled (AG64 METTLER TOLEDO) and mixed together with MiliQ water and 10% glycerol (SIGMA-ALDRICH, St. Louis, USA). pH was adjusted to 7.5. Three types of lysis buffers were prepared, for a full overview see (Appendix buffers)

Binding buffer and elution buffer used for protein purification were prepared accordingly. Binding buffer contained 20mM Hepes (SIGMA-ALDRICH, St. Louis, USA) and 500mM NaCl (SIGMA-ALDRICH, St. Louis, USA). Elution buffer contained the same, with the addition of 500mM Imidazole (Merk, Darmstadt, Germany). Chemicals measured and scaled, mixed with MiliQ water and pH adjusted to 7.5. Number of different binding and elution buffers were prepared for this research, for a full overview see (Appendix buffers)

Running buffer used for gel filtration were prepared accordingly. 20mM Hepes (SIGMA-ALDRICH, St. Louis, USA), 300 mM NaCl (SIGMA-ALDRICH, St. Louis, USA) were measured and scaled, mixed with MiliQ water and pH adjusted to 7.5.

To reach the correct the pH level for all the buffers used, 1M HCl and 1M NaOH solutions were used for adjustment. The pH was measured real time with a PHM210 standard pH meter (MeterLab, Lyon, France). All buffers were filtered with a vacuum chamber and 0.2µm nuclepore track-etch filter (Whatman).

3.4 Medium & plate preparation

3.4.1 Lysogeny broth (LB) and agar plates

LB and agar plates 10 g NaCl (SIGMA-ALDRICH, St. Louis, USA), 10 g tryptone (SIGMA-ALDRICH, St. Louis, USA), 5 g yeast extract (SIGMA-ALDRICH, St. Louis, USA). Components were measured and weighed then mixed with miliQ water to give the total volume a total of 1L. Autoclaved at 121°C for 20 min. Medium was cooled to 50 °C, where 2 tablets (25 mg per tablet) of ampicillin (Novagen, Temecula, USA) were added. To keep the potency the ampicillin the temperature cannot exceed 50°C.

Same procedure with agar plates with the addition of 20 g ager (SIGMA-ALDRICH, St. Louis, USA) per liter. After ampicillin was added, the was media poured into Petri dishes, 20ml in each plate. Plates were cooled for two hours without lid, to avoid condensation. Plates were stored with the lid, upside down at 4°C.

3.4.2 Terrific broth (TB) media

24 g yeast extract (SIGMA-ALDRICH, St. Louis, USA), 20 g tryptone (SIGMA-ALDRICH, St. Louis, USA) and 4ml glycerol (SIGMA-ALDRICH, St. Louis, USA) mixed with MiliQ water for a total volume 0.9L. Autoclaved at 121°C for 20 min. 0.1L phosphate buffer mixed containing 0.017M KH_2PO_4 and 0.072M K_2KPO_4. Buffer filtered as described in buffer preparation. Buffer mixed with media after autoclavation, and 50µg/ml kanamycin (Sigma) added.

3.4.3 Starch plates

Starch agar plate were made with phosphate buffer. 3g agar (SIGMA-ALDRICH, St. Louis, USA), 2g starch (SIGMA-ALDRICH, St. Louis, USA) mixed with MiliQ water to a volume of 100ml. Autoclaved and stored 75°C. Phosphate buffer pH 7,4 made with a mix of 2M solution 8,02ml K_2HPO_4 / 1,98 KH_2PO_4 and 1,17g NaCl mixed with MiliQ H_2O for a volume 100ml. Buffer heated to 75°C and mixed with starch agar solution and poured into petri dishes. Cooled until solidified and stored at 4°C.

3.4.4 Xylan plates

Xylan agar plate were made with phosphate buffer. 3g agar (SIGMA-ALDRICH, St. Louis, USA), 2g xylan from beechwood (SIGMA-ALDRICH, St. Louis, USA) mixed with MiliQ H_2O to a volume of 100ml. Autoclaved and stored 75°C. Phosphate buffer pH 5,8 made with a mix of 2M solution 0,85ml K_2HPO_4 / 9,15 KH_2PO_4 and 1,17g NaCl mixed with MiliQ H_2O for a volume 100ml. Buffer heated to 75°C and mixed with xylan agar solution and poured into petri dishes. Cooled until solidified and stored at 4°C.

3.5 Gene construct for CBM candidate

The two protein sequences that were chosen AMOR-CBM9-14 and AMOR-CBM9-15 for expression was sent to GenScript (genscript.com) to construct plasmids. The signal peptide sequence was removed from the sequences, with additional one alanine and 6 histidine HIS tag added on the C terminus. These were constructed in the pET-21a(+)

vector for each construct in the NdeI/BamHI cloning site. The pET-21a(+) vector contains gene for ampicillin resistance to allow for selective growth. Both proteins were codon optimized with GenSmart™ codon optimization Tool (genscript.com). The tool considers in more than 200 factors involved in gene expression including GC content and codon usage. The plasmids arrived as 4µg pellets in a vial.

3.6 Plasmid resuspension

Vials with vector stored at -20°C were thawed on ice, then centrifuged at 6000RPM fixed rotor Eppendorf (Centrifuge 5418 R, Hamburg, Germany) 1min. Resuspended with 80µl ultrapure water and incubated on ice for 15min, then vortexed 20 seconds, stored in freezer at -20°C.

3.7 Transformation

Competent cells were thawed on ice. 15µl competent cells mixed with 1µl 50 ng plasmid and incubated on ice for an additional 30 min. Heat shocked 20 seconds at 42°C in a water bath (Grant SUB Aqua 5, Cambridgeshire, England) and 2 minutes additional incubation in ice. Cell solution was mixed with 135µl preheated 37°C SOC (Outgrowth Medium New England BioLabs) and incubated 37°C 200RPM in a shaker incubator (innova 44, New Brunswick, Canada) for 60min. 60µl and 30µl of cell mixture were cultured on preheated to 37°C agar plates with amp. Incubated overnight at 37°C upside down, then stored at 4°C. Colonies were counted for each plate to see the success rate of the transformation. One plate was used as a control, cultured with competent cells without plasmid.

Transformation was performed with *E. coli* BL21-Gold (DE3) pLysS and *E. coli* BL21-Gold (DE3) competent cells on AMOR-CBM9-14 and AMOR-CBM9-15.

Transformation with *E. coli* BL21-Gold (DE3)competent cells performed on AMOR-GH57 constructs and a control with empty pet21a+ vector.

3.8 Expression and expression optimization

Protocol for protein expression

A single colony from transformed cells was picked from agar plate with a pipette and dropped in a 50ml falcon tube with 5ml LB + amp (100mg). Placed in shaker incubator (innova 44, New Brunswick, Canada), incubated overnight at 37°C 250 RPM. Expression culture 45ml LB + amp was inoculated with 5ml preculture. Then incubated at 37°C 250RPM to reach OD (optic density) between 0.6/0.8abs. OD was measured with UVmini-1340 spectrophotometer (SHIMADZU, Kyoto, Japan) at 600nm wavelength. 1ml medium was used as blank, then samples routinely measured up to the desired density. Each time 1ml sample placed in cuvette and ran in Uvmini. After OD was reached, samples were left to rest in the expression temperature 37°C 30 min. 200µl aliquots was taken from each culture, centrifuged 13000RPM, supernatant removed and pellet stored at -20°C, these are used as a reference when comparing before and after expression. When cultures had equilibrated temperature, protein expression was induced with 0.1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG),

and incubated at 37°C for 4 hours. After induction 100µl aliquots taken and treated as those prior to induction each hour after. Finally, expression cultures were harvested by centrifugation.

Several attempts were made to optimize the protein expression to yield high protein concentration and soluble proteins, with changes to the parameters (Table 2). Incubation temperature affects the growth rate of the culture, and the rate of expression(Rosano & Ceccarelli, 2014). Testing different temperatures for the incubation for the preculture, expression culture and protein expression. When lower temperatures were tested, longer incubation time was used to assess slower production. Different IPTG concentrations were tested to reduce over expression. The amount of inoculum used to inoculate the expression culture. Glycylglycine has proven to have an effect on increasing solubility on some recombinant proteins (Ghosh et al., 2004). This was replicated with AMOR-CBM9-14 and the AMOR-GH57 constructs. 5 cultures for each construct with 0M glycylglycine (Simga) 50mM glycylglycine, 200mM glycylglycine, 500mM glycylglycine and 1M glycylglycine in the expression culture. A total overview of the optimization steps tested (Table 2).

Expression of AMOR-GH10A was performed as described in Fredriksen er al.'s article 2019(Fredriksen et al., 2019). Transformed competent cells containing AMOR-GH10A were shipped from NMBU. Preculture made with 500ml Terrific broth (TB) with addition of 50 µg/ml kanamycin incubated 23°C overnight 200RPM. Expression was induced with 0.2mM IPTG the following day and additionally incubated for 24 hours.

Tabell 2 Overview of optimization and test in the lab for AMOR-CBM9 constructs and AMOR-GH57 constructs

		Pre culture			Expression culture			Expression	IPTG concentration	Harvesting	Cell lysis	Sonication
Sample	Competent cell	Source	Volume	Incubation temp, RPM	Inoculum	Temp and RPM	OD (Abs 600nm)	Temp, time and RPM	(mM)	Rcf and min	Lysis buffer	Amplitude, time, repititions
AMOR-CBM9-15 AMOR-CBM9-14	pLysS	Plate	5mL LB amp	37°C 250	45ml LB amp + 5ml preculture	37°C 250	0.7 0.7	26°C ON** 250	1 0.1*	4500, 10	50mM Hepes 300mM NaCl 10% glycerol (99%)pH 7.5	
AMOR-CBM9-15 AMOR-CBM9-14	pLysS	Plate	5mL LB amp	26°C 250	45ml LB amp + 5ml preculture	26°C 250	0.7 0.7	37°C 4H 250	1	4500, 10	50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5	
AMOR-CBM9-15 AMOR-CBM9-14	pLysS & Gold	plate	5mL LB amp	37°C 150	45ml LB amp + 5ml preculture	37°C 250	1.0 0.037 1.1 0.081	26°C ON 250	0.1	4500, 10	50mM Hepes 300mM NaCl 10% glycerol (99%)pH 7.5	
AMOR-CBM9-15 AMOR-CBM9-14	pLysS & Gold	plate	50ml LB amp	23°C 150	Used whole preculte	23°C 150	1.1 1.2 1.1 1.2	23°C ON 150	1	4500, 10	50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5	27% 10sec * 5
AMOR-CBM9-15 AMOR-CBM9-14	Gold	Plate	3ml LB amp	20°C 190	3.6ml LB amp + 400µl preculture	20°C 190	0.7 0.6	20°C ON 190	0.1	14500, 6	50mM Tris 500mM NaCl 5mM Imidazole pH 8	29% 10sec * 3
AMOR-CBM9-15 AMOR-CBM9-14	Gold	plate	5mL LB amp	37°C 150	45ml LB amp + 5ml preculture	37°C 150	0.6 0.6	20°C ON 150	1	4700, 20	50mM Tris 500mM NaCl 5mM Imidazole 10% Glycerol(99%) pH 8.5 50mM Hepes 300mM NaCl 10%	29% 10 sec * 3

											Glycerol(99%) pH7.5	
AMOR-CBM9-15 AMOR-CBM9-14	Gold	Plate	5mL LB amp	37°C 150	45ml LB amp + 5ml preculture	37°C 150	0.9 0.9	37°C 4H 150	0.1	4700, 20		
AMOR-CBM9-14	Gold	Plate	5mL LB amp	37°C 150	45ml LB amp (0-1M Glycylglycine) + 5ml preculture	37°C 150	0.8	20°C ON 150	0.5	7000, 20	50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5	30% 10 sec * 5
AMOR-GH57-1 AMOR-GH57-2 AMOR-GH57-3 AMOR-GH57-4 AMOR-GH57-5	Gold	Plate	4ml LB amp	37°C 150	9ml LB amp + 1ml preculture	37°C 150	0.7 0.6 0.7 0.6 0.6	20°C 3.5H 150/ 20°C ON 150	0.3	7000, 20	50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5	27% 10 sec * 3
AMOR-GH57-1 AMOR-GH57-2 AMOR-GH57-3 AMOR-GH57-4 AMOR-GH57-5	Gold	plate	3ml LB amp	37°C 200	4ml LB amp + 100µl preculture	37°C 150	0.5 0.5 0.5 0.4 0.6	16°C 3.5H 200	0.3	14500, 6	50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5	29% 10 sec * 3

AMOR-GH57-4	Gold	plate	3ml LB amp	37°C 200	10ml LB amp (0-1M glycylglycine)+ 250µl preculture	37°C 150	0.6 0.5 0.7 0.8 0.2	16°C ON 200	0.3	14500, 6	50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5	29% 10 sec * 3
AMOR CBM 9-4 AMOR CBM 9-17	Gold	Glycerol stock	3ml LB amp	37°C 200	9ml LB amp + 1ml preculture	37°C 200	1.1 1.0	16°C ON 200	0.1	7000, 20	50mM Tris 500mM NaCl 5mM Imidazole 10% Glycerol(99%) pH 8.5 50mM Hepes 300mM NaCl 10% Glycerol(99%) pH7.5	29% 10 sec * 3
AMOR CBM 9-4 AMOR CBM 9-17	Gold	Glycerol stock	4ml LB amp	37°C 200	9ml LB amp (0-200mM Glycylglycine) + 1ml preculture	37°C 200	0.6 0.7	16°C ON 200	0.2	7000, 20		
AMOR-GH57-4	Gold	Glycerol stock	4ml LB amp	37°C 200	9ml LB amp (0-200mM Glycylglycine) + 1ml preculture	37°C 200	0.6	37°C 4H 200	0.3	7000, 20		
AMOR-GH57-2 AMOR-GH57-4	Gold	Glycerol stock	4ml LB amp	37°C 200	9ml LB amp (0-200mM Glycylglycine) + 1ml preculture	37°C 200	0.7 0.8	16°C ON 200	0.3	7000, 20		
AMOR-GH57-2 AMOR-GH57-4	Gold	Glycerol stock	8ml LB amp	37°C 200	243 ml LB amp + 6.25ml preculture	37°C 200	0.6 0.6	16°C ON 200	0.3	7000, 10	50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5	29% 10 sec * 3
AMOR-GH57-2 AMOR-GH57-4	Gold	Glycerol stock	10ml LB amp	37°C 200	97.5ml LB amp + 2.5ml preculture	37°C 200	0.6 0.6	16°C ON 200	0.3	7000, 10	50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5	29% 10 sec * 3

| AMOR-GH57-2 AMOR-GH57-4 | Gold | plate | 3ml LB amp | 37°C 200 | 4ml LB amp + 100µl preculture/ 4ml LB amp (200mM glycylglycine) + 100µl | 37°C 200 | 0.7 0.5 | 16°C ON 200 | 0.3 | | | 50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5 | |
| AMOR-GH57-2 AMOR-GH57-4 | Gold | plate | 3ml LB amp | 37°C 200 | 4ml LB amp + 100µl precultue/ 4ml LB amp (200mM glycylglycine) + 100µl/ 100ml LB amp + 2.5ml precultue | 37°C 200 | 0.7 0.5 | 16°C ON 200 | 0.3 | | 7000, 20 | 50mM Hepes 300mM NaCl 10% glycerol (99%) pH 7.5 | 35% 10 sec * 7 |

* Duplicates tested with different concentration IPTG

** Overnight (ON)

*** Cultures contains Glycylglycine

3.9 Glycerol stock

Glycerol stock was made from both the CBM9 constructs and GH57. Before induction of the expression culture, 150µl culture was gently mixed with 150µl glycerol (SIGMA-ALDRICH, St. Louis, USA) to obtain a homologues mixture and stored at -80°C. Two duplicates for each construct were made.

3.10 Cell harvesting

Expression cultures ready to be harvested was cooled on ice. Then poured in falcon tubes and centrifuged (BECKMAN COULTER Allegra 2IR, Krefeld, Germany) 5500RPM with swingout orientation or 8000RPM fixed position 15-20min to achieve a solid pellet and clear supernatant. The supernatant was removed, and pellets weighed. Culture volume of 10ml or less was harvested at 13000RPM 5min (Eppendorf 5418R, Hamburg, Germany). Pellets were stored at -20°C.

3.11 Cell lysis

To retrieve the proteins, the cells had to be broken down without breaking the protein. Frozen pellets were thawed on ice. Then pellet was resuspended with lysis buffer (1ml per 0.1g pellet) and 25 mg/l lysosome and incubated on ice 30min. Samples sonicated (SONICS Vibra cell VCX 130 PB, Newtown, USA) 3-5 times with 27%-29% amplitude, in intervals of 8-10 sec. Different variations of sonication for each sample (Table 2). Aliquots were taken before samples were centrifuged. 2ml samples centrifuged 13000rpm 6 min and large samples < 2ml centrifuged 5500rpm 15-20 min.

3.12 SDS-PAGE

To test the success of the protein expression, aliquots from before induction with IPTG, after induction, lysed cell solution and lysed cell supernatant were tested with SDS-PAGE.

Cell pellet from before induction, after induction, lysed cell solution and supernatant after cell lysis samples were run in SDS-PAGE gel. Pellets were resuspended in 20µl lysis buffer and 10µl LDS sample buffer 4x (GenScript, Piscataway, USA). 16µl lysed cell solution was mixed with 4µl sample buffer. 20µl supernatant was mixed with 5µl sample buffer. All samples were heated in a thermomixer (Eppendorf. Hamburg, Germany) 96°C for 10min and spun down. Each well was loaded with 16µl/40µl sample and a variety of different ladders were used, precision plus protein dual color standard (BIO-RAD, Hercules, USA) broad multi pre-stained protein standard (GenScript, Piscataway, USA) and elite pre-stained protein ladder (ProteinArk, Sheffield, England).

Precast gels were used ExpressPlus 10%, Sure Page 12% and gradient 8-16% (GenScript, Piscataway, USA) placed in Mini PROTEAN Tetra cell (Bio Rad, Piscataway, USA). Inside between the gels and outside of the chamber surrounding the gels were filled with running buffer (Tris-MOPS-SDS). Gel ran 140V/170V 40/60 min. Gels washed in ionized water then stained with instantblue (RunBlue Bis-Tris protein gels) 60min 160RPM shaker. Stain removed and gels distained in water over

night on shaker 80RPM. Gels were observed and documented on blue pad light plate (Bio Helix, New Taipei city, Taiwan).

3.13 Protein purification

Protein purification was performed with ÄKTA start (Ge Healthcare, Illinois, USA) protein purification system. Cell pellets were thawed on ice and resuspended with 1ml lysis buffer pr 0.1g cell pellet sample with addition of 0.25mg/ml lysozyme (Sigma). Solution incubated on ice 30min. Cells lysed with ultrasonic processor (SONICS Vibra cell VCX 130 PB, Newtown, USA). Total cell lysate was then centrifuged 13000rpm/8000rpm 2-3min. Cell pellet suspension removed, and supernatant stored on ice.
ÄKTA start protein purification system was prepared accordingly. Flushed with MiliQ H_2O to remove storage ethanol 20% in the system loop. Elution buffer was then used to flush the loop followed by binding buffer. HisTrap column (5ml GE Healthcare) was then connected to the loop and flushed with binding buffer column five column volumes.

Sample supernatant injected in the sample loop. Method run selected then templates with Affinity chromatography. With following parameters chosen:

Column volume 5ml
Flow rate 5ml/min
Pressure limit 0.30mPA
Sample from pump
Sample volume 1-10mL
Equilibration volume 5cv (column volume)
Wash unbound volume 10-15 cv
Elution option Gradient
Target conc B 100%
Gradient volume 5cv
Fraction volume 1.5 ml

Flowthrough from unbound fraction taken in purification. Elution fractions sorted from 1-25 and stored in fridge 4°C with flowthrough. Elution fractions was tested on SDS-PAGE gel with flowthrough, sample from cell lysate and supernatant to test if the protein bound to the column.

Protein purification (Table 3) was performed as the workflow previously described with adjustments to each of the samples. Sample AMOR-CBM9-14, AMOR-GH57-2 and AMOR-GH57-4 was resuspended in lysis buffer B1 (Appendix Buffers). Sample AMOR-CBM9-15 was resuspended in lysis buffer B2 (Appendix Buffers). Sample AMOR-GH10A was resuspended in lysis buffer B3 (Appendix Buffers) without lysozyme. Sample 12AMOR1-GH13 was resuspended in binding buffer B9 (Appendix Buffers) without lysozyme. There were also used different binding and elution buffers with a full overview for each bellow (Table 3). Additional denatured purification was tested on AMOR-CBM9-14, AMOR-CBM9-15, AMOR-GH57-2 and AMOR-GH57-4 with 8M urea (Merck) added in the binding and elution buffer.

Table 3 Overview of sample runs and parameters during protein purification

Sample	Sample volume	Sonication	Binding and elution buffer*	Isocratic/gradient
AMOR-CMB9-14 (First run)	1.0 ml	29% amplitude 10 sec * 3	20mM Hepes 500mM NaCl pH 7.5	Gradient
AMOR-CBM9-14 (Second run)	1.5 ml	32% amplitude 10 sec * 4	20mM NaPO$_4$ 20mM imidazole 500mM NaCl 8M urea pH 7.5	Gradient
AMOR-GH57-2 (First run)	7ml	33% amplitude 10 sec * 3	20mM Hepes 25mM imidazole 500mM NaCl pH 7.5	Gradient
AMOR-GH57-4 (First run)	7ml	33% amplitude 10 sec * 3	20mM Hepes 25mM imidazole 500mM NaCl pH 7.5	Gradient
AMOR-GH57-2 (Second run)	9 ml	35% amplitude 10sec * 7	A: 20mM Hepes 25mM imidazole 500mM NaCl pH 7.5	Isocratic
AMOR-GH57-4 (Second run)	9 ml	35% amplitude 10sec * 7	20mM Hepes 25mM imidazole 500mM NaCl pH 7.5	Gradient
AMOR-GH57-2 (Third run)	10 ml	35% amplitude 10sec * 7	20mM Hepes 25mM imidazole 500mM NaCl pH 7.5	Gradient
AMOR-GH57-4 (Third run)	10 ml	35% amplitude 10sec * 7	A: 20mM Hepes 25mM imidazole 500mM NaCl pH 7.5	Gradient
AMOR-GH57-2 (Fourth run)	5ml flowthrough	Sonication not needed	10mM Hepes 10mM imidazole 500mM NaCl pH 7.5	Gradient
AMOR-GH57-2 (Fifth run)	1.8ml	29% amplitude 10sec * 5	20mM NaPO$_4$ 20mM imidazole 500mM NaCl 8M urea pH 7.5	Gradient

AMOR-GH57-4 (Fourth run)	1.8ml	29% amplitude 10sec * 5	20mM NaPO$_4$ 20mM imidazole 500mM NaCl 8M urea pH 7.5	Gradient
AMOR-CBM9-15 (First run)	1.5 ml	50% amplitude 30 sec * 5	20mM NaPO$_4$ 20mM imidazole 500mM NaCl 8M urea pH 7.5	Gradient
AMOR-CBM9-14 (Third run)	1.2 ml	50% amplitude 30 sec * 5	20mM NaPO$_4$ 20mM imidazole 500mM NaCl 8M urea pH 7.5	Gradient
12AMOR1-GH13 (First run)	7ml	50% amplitude 30 sec * 5	50mM Sodium Phosphate 300mM Sodium Chloride 10mM imidazole pH 8	Gradient
AMOR-GH10A (First run)	6.5 ml	50% amplitude 30 sec * 5	5mM imidazole 50mM Tris-HCl 500mM NaCl pH 8	Gradient

*Elution buffer content is same as binding buffer with exception of 500mM imidazole, only binding buffer is described in the table

A full overview of the different runs tested to purify the target protein (Table 3). The main difference was testing out different concentration of imidazole in the binding buffer.

3.14 Protein dialyses

Samples that were denatured during protein purification with urea, were attempted renatured with D-Tube Dialyzer Mega kit (Novagen, Darmstadt, Germany). This was tested on AMOR-CBM9-14 second run, AMOR-GH57-2 fifth run and AMOR-GH57-4 fourth run (Table 3). Tubes were prepared by incubating the inner membrane with 10ml milli-q water for 10 minutes. AMOR-CBM9-14 elution fraction T8-12 was used, in total 7ml sample placed D-tube, cap screwed in place and submerged in 0.4 L storage buffer B15 (Appendix Buffers) over night in a chilled room with a magnetic mixer.

AMOR-GH57-2 used elution fraction T6-T12 1,4ml from each with a total 9,8ml sample. To remove the elution buffer with imidazole and urea from the samples D-Tube Dialyzer Mega kits from (Novagen) was used on proteins denatured by Urea. 6-8kDa tubes used. Tubes was premoisten with miliQ water, incubated 10min in room temperature. Then 1.4ml from each fraction from the iMac used with a total volume 9.8ml sample in the tube. Tube was submerged in 900mL buffer and left over-night with a magnet stirrer.

AMOR-CBM9-14 that was purified with the urea buffer in the Äkta start, stored in the fridge 4°C in the urea buffer with imidazole. To change the buffer, the dialysis kit D-Tube Dialyzer 10ml MWCA 6-8 kDA from Novagen used. Fraction 8 to 12 was placed inside the premoisten tube, total 7,5ml sample volume. The tube submerged in the new storage buffer 0.4L (20mM NaPO$_4$,0.5M NaCl, 7,5pH). Placed over night in cooled room 4°C with magnetic stirrer.

3.15 Protein concentration

Protein samples were concentrated with Amicon® Ultra 30K device – 30,000 MWCO centrifugal filters (Merck, Darmstadt, Germany). New filters used for each sample and used according to the user guide. Filters prepared first with 4ml MiliQ water, centrifuged 4715RPM/4000G with swing out rotor 10min. Flow through water removed and 4ml sample loaded. Centrifuged 4000G 15min each time more sample is loaded. Flow through sample retrieved and removed. Then storage buffer loaded 4ml, centrifuged 4000G 15min, repeated 3-4 times to change the buffer in the sample. Protein sample is then retrieved with a pipette from the bottom of the filter. Samples kept on ice and stored in the fridge 4°C.

Protein sample from the dialysis tube removed and placed in a 15ml falcon tube kept on ice. To increase the protein concentration Amicon Ultra centrifugal filters from Merck Millipore used. 3ml protein sample loaded in the tube. Centrifuged 10 min 5000RPM, then additional 5ml sample loaded and centrifuged 10min 5000RPM. Final volume 200µL purified protein with 7.5mL flow trough. Tube stored in the fridge 4°C.

3.16 Measuring protein concentration

Protein concentrations were measured with Qubit^{tm} 4 Fluorometer Invitrogen by Thermo Fisher Scientific (Waltham, USA). Sample from before and after protein concentration was tested. 995µl Qbit buffer mixed with 5µl Qbit reagent was done to make the Qbit working solution. 3 different Qbit protein standards 10µl was mixed with 190µl working solution. Sample mixture 5µl protein sample before and after up concentration with 195µl working solution. Vortex 3 sec then incubated room temperature 15 min. The standards measured first to obtain the standard curve. Then the two samples were measured after telling the Qbit how much the sample was diluted.

3.17 Proteomics analysis of AMOR-CBM-14

Sample from the second run from protein purification of AMOR-CBM9-14 to confirm the protein present on the Gel. One band from the SDS-PAGE gel of AMOR-CBM9-14 was stapled out with x-tracta Gel Extraction (SIGMA-ALDRICH, St Louise, USA) and placed in sterile Eppendorf tube (2ml) and stored at -20°C. The band was delivered to PROBE (Proteomics Unit University of Bergen) for mass spectrometry analysis. This was done to confirm that the target protein is present in the band and was successfully purified.

E. coli BL21 (DE3) complete genome was retrieved from NCBI (ID:167 [22818]) delivered along with the sequence from AMOR-CBM9-14 in a fasta file.

3.18 Heat treatment

Protein solution from AMOR-GH57-2 and AMOR-GH57-4 were tested for heat tolerances. 800µl of supernatant from sonication of pellets AMOR-GH57-2 and AMOR-GH57-4, divided in two Eppendorf tubes 400µl each. Samples heated 60°C and 70°C30min in hot well thermomixer at 60°C and 70°C. Precipitated protein removed with centrifuging 13000RPM 3min. Samples tested on SDS-PAGE gel with supernatant prior to heat treatment.

Heat treatment was tested a second time at a later stage. Samples was taken from the flowthrough in the Äkta start affinity chromatography. 2ml flowthrough from AMOR-GH57-2 and AMOR-GH57-4, expressed with and without glycylglycine. 1ml from each placed in hot wells 70°C and 1ml in hot wells 80°C and heated 30min. Centrifuged 14000RPM 20min to remove precipitated proteins. Then tested on SDS-PAGE gels.

3.19 Plate assay of the samples

Concentrated purified protein was tested for activity against substrate. AMOR GH57_44 and AMOR-GH57-4 tested on 1% starch plate. Wells in the starch plate were made with pipet tip and protein loaded in each well. 2µg of protein loaded in each well from AMOR GH57_44, AMOR-GH57-4 and 12AMOR1-GH13. Alfa-amylase from *Bacillus licheniformis* (Sigma) used as positive control and AMOR GH10 used as negative control. Incubated 65°C 4 hours followed by staining with modified Gram's iodine reagent (1.0 g KI, 0.5 I, 300ml H_2o). The iodine stains the starch in the plate with a dark blue color, no staining indicates starch degradation and enzyme activity. After staining the results were documented and plates stored 4°C. Activity was also tested with 8µg protein in each well. 12AMOR1 activity was further tested for activity at 45°C, 55°C and 60°C with 4 hours incubation time followed by staining.

AMOR-CBM9-14 and AMOR-CBM9-15 was tested for activity on 1% xylan plate. 2 µg of protein loaded in wells in the plate, one from AMOR-CBM9-15 and two from AMOR-CBM9-14 with different treatments. AMOR-GH10A used as positive control and 12AMOR1 as negative control. Incubated 4 hours 65°C then inspected. Activity could be seen as clearing zones around the well, no change in color indicate no activity.

3.20 Gel filtration 12AMOR1-GH13

Gel filtration performed on 12AMOR1-GH13 with Äkta Start using HiLoad 16/600 Superdex200pg column. Sample filtered twice, second time with standard Bio-rad Gel filtration standard (Table 4).

The column was equilibrated using two column volumes of running buffer (20mM Hepes, 300mM NaCl, pH 7,5). Run parameter for gel filtration was 120ml column volume, 1ml/min flowrate, pressure limit 0.3 pascal, sample from loop, 1ml sample volume, 0.1 column volume equilibration volume, elution of 1 column volume and 4ml fractions.

1ml 30.2 µg protein sample was first ran without standard and elution fractions visualized with SDS-PAGE. Second run was performed with Biorad standards (Table 4) to estimate molecular weight. Standard was diluted with 2 ml running buffer and 400 µl diluted standard was mixed with 200 µl (0,56mg) of protein sample, in addition to 400 µl running buffer. 1 ml total protein sample (4.1mg) was loaded on the gel filtration column and run with standard parameters described above. Elution fractions with peaks in the chromatography was visualized using SDS-PAGE.

Table 4 Gel filtration standard components

Component	Molecular weight	Amount per vial (mg)
Thyroglobulin (bovine)	670,000	5.0
g-globulin (bovine)	158,000	5.0
Ovalbumin (chicken)	44,000	5.0
Myoglobin (horse)	17,000	2.5
Vitamin B12	1,350	0.5
Total		18

3.21 Calculation of molecular weight of 12AMOR1-GH13 from gel filtration chromatography

Chromatography from gel filtration with standards was used to calculate the molecular weight of the sample. With the chromatography acquired from gel filtration K_{ave} was calculated for each standard, elution volume (Ve) marked by each peak and subtracted from the void volume (Vo) divided by the column volume (Vt) subtracted by void volume (Ui, 1979). K_{ave} was plotted against log molecular weight of the standards in a scatter plot. Trendline was used to get the regression equation of the slope and obtain the log molecular weight of the sample.

$$Kave = \frac{(Ve - Vo)}{(Vt - Vo)}$$

4. Results

4.1 Sequence analysis and gene target selection of potential novel CBM9 containing xylanases

The CBM9 containing sequences were first analyzed for transmembrane helices as this were the first step in selecting candidates. Proteins that contain a transmembrane helix could be associated with the cell membrane and increase the chance for inclusion bodies, thus it was decided that these should not be chosen. In 8 of the CBM-containing sequences a transmembrane helix was predicted (Table 5). A domain search performed with the HMMER webserver (Johnson et al., 2010) showed that most of the candidates contained the CBM9.1 domain (family 9 carbohydrate binding module), in total 8 sequences had no hit on this CBM (Table 5). Theoretical melting

temperature (T_m), isoelectric point (pI) and molecular weight were tested on the candidates. The calculated T_m indicated that the targets were thermostable. 26 CBM9 containing sequences that did not have a predicted transmembrane helix was aligned in MEGA-X version (10.1.7) with the MUSCLE algorithm (Kumar et al., 2018) and visualized with the ESPript 3.0 web tool (Robert & Gouet, 2014). Based on the alignment, two candidates were chosen: AMOR-CBM9-14 and AMOR-CBM9-15 (Figure 5). These had conserved regions similar to the other candidates, while also being distinctive on their own. Both sequences contained a signal peptide (Sec/SPI type) at the beginning of the sequences with a likelihood of 0.98 for the type and length (Figure 6). The signal peptide was removed from the sequences prior to vector construction at GenScript (genscript.com) to avoid the protein being secreted by the *E. coli* during expression. The theoretical pI was re-calculated on the AMOR-CBM9-14 and AMOR-CBM9-15 modified sequences with his-tag and without signal peptide. The alteration in the sequence composition in both ends of the sequences could potentially alter the pI value. AMOR-CBM9-14 had theoretical pI of 8.98 and AMOR-CBM9-15 a pI of 6.65.

Blast was performed in the non-redundant and uniport database, and species with the highest similarity was recorded for each sequence. AMOR-CBM9-14 along with multiple of the other CBM9 candidate sequences shared highest similarity with *Caldithrix abyssi,* which is listed in the CAZy database with a protein containing the CBM9 domain and 4 other CBMs (Table 5). *Caldithrix abyssi* is also listed in CAZy database with 3 family's carbohydrate esterases, 24 families glycoside hydrolases, 12 families glycosyltransferases and two families polysaccharide lyases. The first cultivated representative of the group *Caldithrix abyssi* originated from a deep-sea hydrothermal vent (Sievert et al., 2000). *Bacteroidetes bacterium* were dominating hit in the analysis with 10 sequences with highest similarity against it. *Bacteroidetes* shows hits in CAZy database, and is described with CBM50, two families for carbohydrate esterase, 15 families of glycoside hydrolase and 11 glycosyltransferases. Hits for each sequence where for the most part identical between non-redundant and uniprot database.
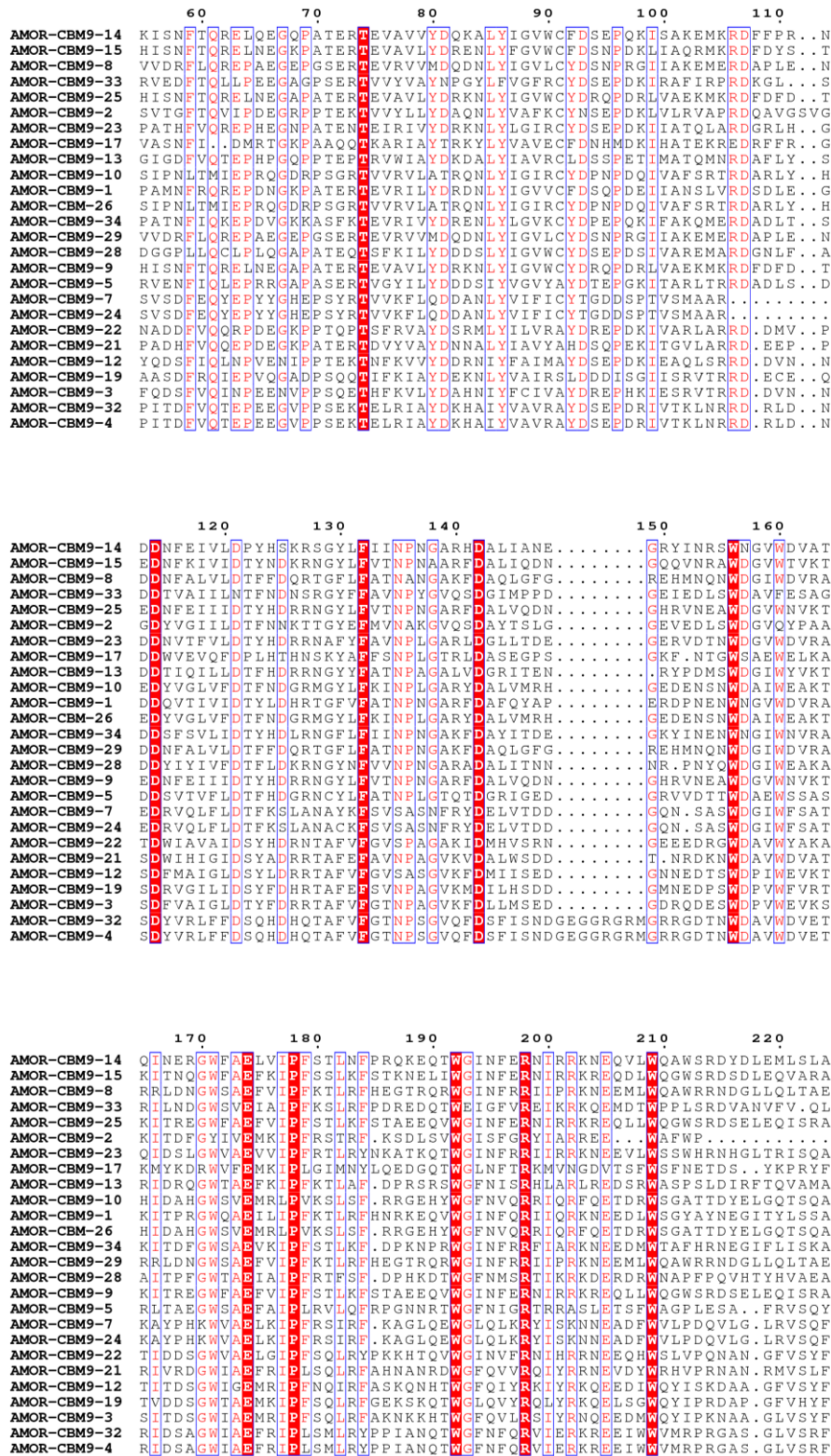
Figure 5: *Alignment of the CBM candidate sequences without transmembrane helices. Alignment performed in MEGA-X the output from ESPript 3.0. Alignment contains the CBM9 domain from no 60 at AMOR-CBM9-14.*

Figure 6: *Sequence analysis of AMOR-CBM9-14 and AMOR-CBM9-15 revealed both contains carbohydrate binding module family 9 domain (CBM9) able to bind to a rage of mono and disaccharides. Signal peptide located in the beginning of the sequence predicted to be Sec/SPII peptide, these were removed before vector construct.*

Table 5 Sequence analysis of the CBM9 containing sequences and AMOR-GH10A

Working name	Sequence Length	Transmembrane helix	Signal P HMMER	Signal P-5.0 Gram pos	Signal P-5.0 Gram neg	Signal P-5.0 Archaea	Domain/Pfam	NCBI Blast (nr)	Per indent(%)	Blast uniprot	Per indent (%)	Tm Predict (℃)	Molecular weight	Isoelectric Point (pI)
AMOR-CBM9-1	717	0	1-20	1-29 p:0,3073	1-20 p:0,9301	1-20 p:0,4751	CBM9_1	Bacteroidetes bacterium	82.08	Caldithrix abyssi DSM 13497	44.5	55 ~ 65	83504.23	6.72
AMOR-CBM9-2	294	0	1-27	1-20 p:0,8557	1-20 p:0,8951	1-20 p:0,6191	CBM9_1	Candidatus Latescibacteria bacterium	53.28	Candidatus Latescibacteria bacterium	53.3	Higher than 65	32612.38	7.61
AMOR-CBM9-3	880	0	1-19	1-19 p:0,4703	1-19 p:0,8307	1-19 p:0,7718	CBM9_1	Caldithrix abyssi	71.74	Caldithrix abyssi DSM 13497	71.7	55 ~ 65	101506.88	6.73
AMOR-CBM9-4	838	0	1-22	1-24 p:0,8489	1-24 p:0,7213	1-24 p:0,8018	CBM9_1	Bacteriodetes bacterium	81.84	Candidatus Poribacteria bacterium [2026781]	41.3	Higher than 65	96378.48	8.55
AMOR-CBM9-5	710	0	1-54	1-41 p:0,5389	1-41 p:0,4302	1-54 p:0,1703		Candidatus Solibacter sp.	58.72	Acidobacteria bacterium	49.6	55 ~ 65	82312.55	9.88
AMOR-CBM9-6	728	1	12-32	1-32 p:0,5603	other	1-32 p:0,3621		Caldithrix abyssi	72.53	Caldithrix abyssi	71.2	Higher than 65	84151.72	8.74
AMOR-CBM9-7	705	0	1-17	other	1-14 p: 0,3964	other	CBM9_1	Candidatus Hydrothermae bacterium	70.21	Candidatus Hydrothermae bacterium	70.2	55 ~ 65	80297.43	5.88
AMOR-CBM9-8	432	0	1-24	1-24 p:0,9488	1-24 p:0,9711	1-24 p:0,9617		Bacteriodetes bacterium	79.30	Deltaproteobacteria bacterium	60.6	55 ~ 65	49742.81	9.07
AMOR-CBM9-9	722	0	1-21	1-21 p:0,4642	1-21 p:0,8900	1-21 p:0,8610	CBM9_1	bacterium BMS3Abin03	71.65	bacterium BMS3Abin03	71.8	55 ~ 65	83106.12	8.32
AMOR-CBM9-10	720	0	1-23	1-23 p:0,9518	1-23 p:0,9785	1-23 0,9738		Bacteroidetes bacterium	71.57	candidate division KSB1 bacterium [2172550]	61.3	Higher than 65	82064.41	9.10
AMOR-CBM9-11	721	1	1-29	1-35 p:0,2349	other	1-30 p:0,2820		Caldithrix abyssi	64.16	bacterium BMS3Abin03	61.5	55 ~ 65	83691.47	5.30

AMOR-CBM9-12	881	0	1-20	1-25 p:0,3475	1-20 p:0,3837	1-25 p:0,4307	CBM9_1	Caldithrix abyssi	68.15	Caldithrix abyssi DSM 13497	68.1	Higher than 65	102189.40	5.55
AMOR-CBM9-13	732	0	1-24	1-24 p:0,1917	1-24 p:0,5717	1-24 p:0,4591	CBM9_1	Candidatus Solibacter sp.	60.56	Bryobacterales bacterium	55.4	Higher than 65	83809.98	9.86
AMOR-CBM9-14	718	0	1-20	1-20 p:0,8375	1-20 p:0,9632	1-20 p:0,9270	CBM9_1	Caldithrix abyssi	78.29	Caldithrix abyssi DSM 13497	76.9	55 ~ 65	83291.66	9.35
AMOR-CBM9-15	720	0	1-23	1-23 p:0,6292	1-23 p:0,9729	1-23 p:0,9169	CBM9_1	bacterium BMS3Abin03	74.75	bacterium BMS3Abin03	74.8	55 ~ 65	84449.97	7.83
AMOR-CBM9-16	758	1	1-38	1-36 p:0,4659	1-36 p:0,3971	1-36 p:0,5200	CBM9_1	Acidobacteria bacterium	52.72	Acidobacteria bacterium	48.6	55 ~ 65	85076.59	5.31
AMOR-CBM9-17	694	0	1-32	1-30 p:0,8939	1-30 p:0,9005	1-30 p:0,9252	CBM9_1	Candidatus Omnitrophica bacterium	58.25	Candidatus Omnitrophica bacterium [2035772]	58.2	55 ~ 65	82183.99	5.59
AMOR-CBM9-18	726	1	1-27	1-33 p:01448	other	1-27 p:0,3711	CBM9_1	Acidobacteria bacterium	47.18	bacterium BMS3Abin03	45.7	55 ~ 65	83352.04	5.91
AMOR-CBM9-19	879	0	1-19	1-19 p:0,7672	1-19 p:0,9859	1-19 p:0,9324	CBM9_1	Calditrichaeota bacterium	56.74	Calditrichaeota bacterium	56.8	Higher than 65	100937.61	7.20
AMOR-CBM9-20	716	1	1-29	1-30 p:0,7083	1-20 p:0,3986	1-30 p:0,6311	CBM9_1	bacterium BMS3Abin03	67.92	bacterium BMS3Abin03	67.9	55 ~ 65	82677.44	6.89
AMOR-CBM9-21	874	0	1-29	1-29 p:0,9109	1-29 p:0,8519	1-29 p:0,8944	CBM9_1	Bacteroidetes bacterium	76.69	candidate division KSB1 bacterium [2172550]	58.2	55 ~ 65	100213.81	9.87
AMOR-CBM9-22	854	0	1-22	1-24 p:0,3787	1-22 p:0,2425	1-24 p:0,3662	CBM9_1	Bacteriodetes bacterium	58.84	bacterium BM53Abin05	47.7	55 ~ 65	97497.08	8.86
AMOR-CBM9-23	729	0	1-31	1-36 p:0,9519	1-36 p:0,9240	1-36 p:0,9298	CBM9_1	Bacteroidetes bacterium	49.04	candidate division KSB1 bacterium [2172550]	48.7	55 ~ 65	83706.96	9.57
AMOR-CBM9-24	705	0	1-17	other	1-14 p: 0,3964	other	CBM9_1	Candidatus Hydrothermae bacterium	70.7	Candidatus Hydrothermae bacterium	70.1	55 ~ 65	80251.42	5.88

AMOR-CBM9-25	722	0	1-21	1-21 p:0,5146	1-21 p:0,8741	1-21 p:0,8654	CBM9_1	bacterium BMS3Abin03	71.65	bacterium BMS3Abin03	71.8	55 ~ 65	83093.08	7.43
AMOR-CBM9-26	720	0	1-23	1-23 p:0,9451	1-23 p:0,9779	1-23 p:0,9717		Bacteroidetes bacterium	71.72	candidatte division KSB1 bacterium [2172550]	61.4	Higher than 65	82098.34	8.94
AMOR-CBM9-27	712	1	1-20	1-20 p:0,2206	1-20 p:0,7030	1-20 p:0,4936	CBM9_1	Candidatus Aminicenantes bacterium	47.84	Candidatus Aminicenantes bacterium	47.8	Higher than 65	82192.57	8.15
AMOR-CBM9-28	745	0	1-31	1-31 p:0,4586	other	1--31 p:0,5412	CBM9_1	Planctomycetes bacterium	52.99	Planctomycetes bacterium	53.1	55 ~ 65	84025.60	5.76
AMOR-CBM9-29	721	0	1-24	1-24 p:0,9488	1-24 p:0,9711	1-24 p:0,9617		Bacteroidetes bacterium	73.66	candidate division KSB1 bacterium [2172550]	46.7	55 ~ 65	83536.82	9.39
AMOR-CBM9-30	693	1	1-23	1-23 p:0,7936	1-23 p:0,8062	1-23 p:0,8457	CBM9_1	Candidatus Poribacteria bacterium	81.24	Candidatus Poribacteria bacterium	59.9	55 ~ 65	80497.70	5.60
AMOR-CBM9-31	287	1	1-26	1-26 p:0,8205	1-26 p:o,8923	1-26 p:o,8936	CBM9_1	Candidatus Latescibacteria bacterium	89.25	Candidatus Latescibacteria bacterium	85.3	55 ~ 65	32515.59	9.82
AMOR-CBM9-32	838	0	1-22	1-24 p:0,8489	1-24 p:0,7213	1-24 p:0,8018	CBM9_1	Bacteroidetes bacterium	81.72	Candidatus Poribacteria bacterium [2026781]	41.4	Higher than 65	97216.31	8.55
AMOR-CBM9-33	714	0	1-26	1-23 p:0,7026	1-23 p:0,6416	1-26 p:0,4163		bacterium (candidate division B38)	38.00	bacterium (candidate division B38)	38.0	55 ~ 65	83185.33	4.73
AMOR-CBM9-34	705	0	1-19	1-25 p:0,5836	1-25 p:0,3034	1-25 p:0,6786	CBM9_1	Caldithrix sp.	47.78	candidate division KSB1 bacterium [2172550]	45.6	Higher than 65	82926.51	8.89
AMOR-GH10A	586	0	1-16	1-22 p:0,10	1-16 p:0,35	1-16 p:0,21	Glyco_hydro_10	Puniceioccaceae bacterium	46.40	Puniceioccaceae bacterium	46.6	Higher than 65	66133	9.0

4.2 Sequence analysis of alpha amylase sequences

Sequence analysis was performed on GH57 sequences with 12AMOR1-GH13. To verify that all AMOR-GH57 sequences contain the Glyco_hydro_57 domain (GH57) (Figure 7), domain search was performed with HMMER webserver (Johnson et al., 2010). No predicted signal peptide found on the N terminal, so one additional analysis was performed with OutCyte (Zhao et al., 2019) to predict if the sequences contained an unconventional protein secretion system (UPS) (Table 6). AMOR-GH57-2 and AMOR-GH57-4 were categorized as UPS, indicating that they might have a system for protein secretion without the use of conventional signal peptide. The sequences were thermostable, and AMOR-GH57-4 had a predicted thermostability higher than 65 °C based on (T_m) calculated. The pI varied between 5.2 to 8.2 between the GH57 sequences (Table 6). Blast with NCBI non-redundant database and uniport revealed that the five alpha-amylase sequences all had similarity around 80% with proteins from archaea. There was no dominating similarity in the blast results, AMOR-GH57-1, AMOR-GH57-4 and AMOR-GH57-5 had similarity with proteins from archaea phyla Euryarchaeota. AMOR-GH57-1 with hit against *Thermoplasmata archaeon* and AMOR-GH57-5 with hit against *Methanosarcinales archaeon* are both described in the CAZy database with a GH57 module. *Thermoplasmata archaeon* also has a Glycosyltransferase family. *Methanosarcinales archaeon* is also described with two CBM families, 8 different glycosyl hydrolase, 8 glycosyltransferase and one polysaccharide lyases.
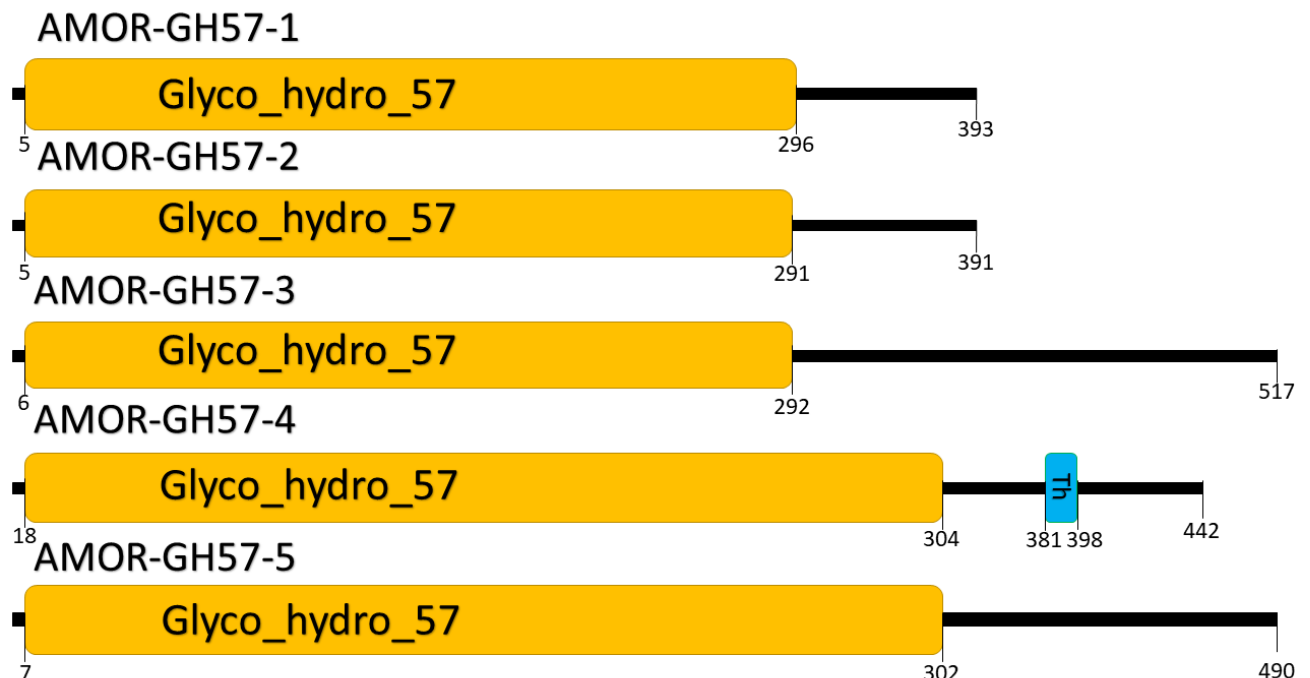
Figure 7: *Sequence analysis of the five α-amylase sequences. All the sequences contain the Glyco_hydro_57 (GH57) domain from the family glycoside hydrolases. AMOR-GH57-4 has an additional transmembrane helix located 381-398*

Table 6 *Sequence analysis of the GH57 containing sequences and 12AMOR1-GH13*

Working name	Sequence Length	Transmembrane helix	Signal P HMMER	Signal P-5.0 Gram pos	Signal P-5.0 Gram neg	Signal P-5.0 Archaea	OutCyte 1.0 / score	Domain/Pfam	NCBI Blast (nr)	Per indent (%)	Blast uniprot	Per indent (%)	Tm Predictor (℃)	Molecular weight	Isoelectric Point (pI)
AMOR-GH57-1	393	0	no	no	no	no	Intracellular 0.9345	Glyco_hydro_57	Thermoplasmata archaeon	97.46	Thermoplasmata archaeon	97.5	55-65	46421	5.4
AMOR-GH57-2	391	0	no	no	no	no	UPS 0.5236	Glyco_hydro_57	Candidatus Diapherotrites archaeon	90.54	Candidatus Diapherotrites archaeon	90.5	55-65	46119	6.9
AMOR-GH57-3	517	0	no	no	no	no	Intracellular 0.5471	Glyco_hydro_57	Candidatus Bathyarchaeota archaeon	81.01	Candidatus Bathyarchaeota archaeon	81.0	55-65	60979	5.2
AMOR-GH57-4	442	0	no	no	no	no	UPS 0.502	Glyco_hydro_57	Euryarchaeota archaeon	76.02	Euryarchaeota archaeon	76.0	Higher than 65	52638	8.2
AMOR-GH57-5	490	0	no	no	no	no	Intracellular 0.5478	Glyco_hydro_57	Methanosarcinales archaeon	73.83	Methanosarcinales archaeon	73.83	55-65	57692	5.4
12AMOR1-GH13	588	0	no	no	no	no	Intracellular 0.5484	Alpha amylase	Geobacillus sp DSP4a	99.83	Bacillus caldolyticus	99.7	Higher than 65	68254	5.5

4.3 Transformation of samples

Transformation was performed with the same procedure on all the constructs using competent cells without plasmids as controls. Growth of visible colonies on the agar plates containing ampicillin indicates that the *E. coli* has successfully taken the pET-21a plasmid containing both the construct sequence and ampicillin resistance (Figure 8). The control plates with competent cells that had not been transformed showed no growth, indicating that the selective media only support growth for those cells that have taken the plasmid. Both BL21 Gold and BL21 pLysS competent cells were successful in the transformation, but comparing the plates and number of colonies, the amount was much greater with pLysS then with Gold (Table 7). While the colonies were much smaller and more difficult to count those with pLysS.
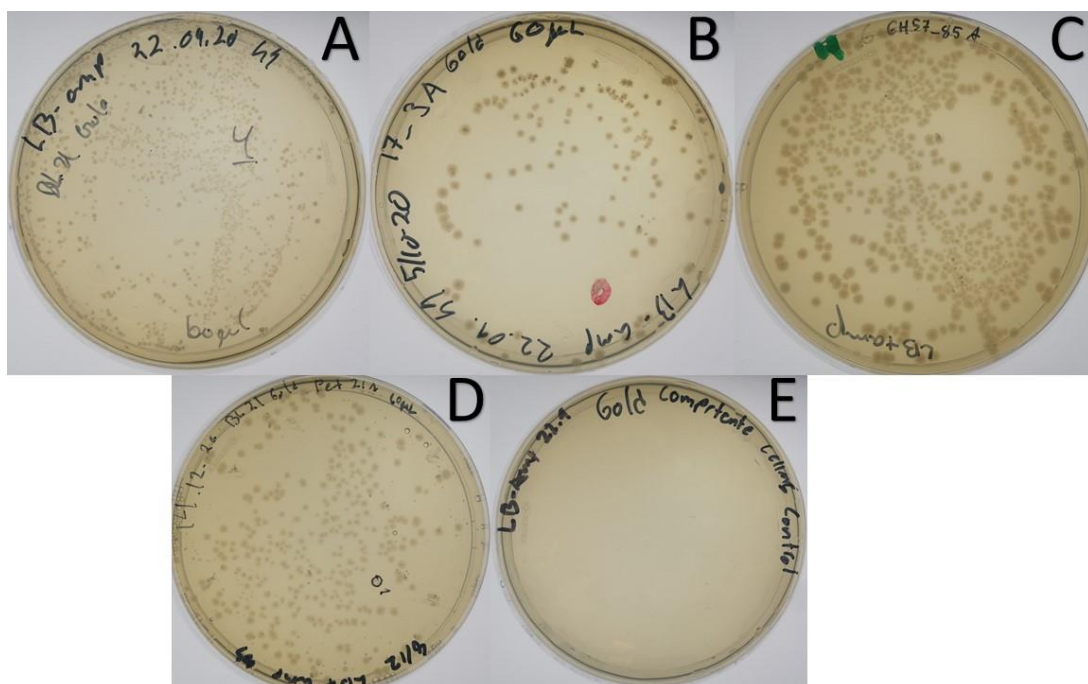
Figure 8: Overview of the different transformations performed cultured on agar plates containing ampicillin. A) transformation of E. coli BL21-Gold (DE3) pLysS competent cells with AMOR-CBM9-15, B) transformation of E. coli BL21-Gold (DE3) competent cells with AMOR-CBM9-14, C) transformation of E. coli BL21-Gold (DE3) cells with AMOR-GH57-4, D) positive control with pet21a plasmids in E. coli BL21-Gold (DE3) cells, and E) negative control used for each transformation using competent cells that has not been transformed with a plasmid.

Table 7 Overview of the *E. coli* BL21-Gold (DE3) pLysS and *E. coli* BL21-Gold (DE3) competent cell transformations.

Working name and competent cell	Number of colonies 60μL	Number of colonies 30μL
AMOR-CBM9-15 (pLysS)	1314	660
AMOR-CBM9-14 (pLysS)	1322	678
AMOR-CBM9-15 (Gold)	114	33
AMOR-CBM9-14 (Gold)	149	79
AMOR-CBM9-15 (Gold)	206	431

AMOR-CBM9-14 (Gold)	81	26
AMOR-GH57-3 (Gold)	166	104
AMOR-GH57-5 (Gold)	268	242
AMOR-GH57-1 (Gold)	120	76
AMOR-GH57-2 (Gold)	221	127
AMOR-GH57-4 (Gold)	714	278
Pet21a+ (Gold) (both 60µl)	339	229

4.4 Expression of CBM9 containing constructs

Expression of AMOR-CBM9-15 and AMOR CBM9-14 was achieved with both *E. coli* BL21-Gold (DE3) pLysS and BL21-Gold (DE3) competent cells (Figure 9-10). The major difference between pLysS and Gold was the time it took to achieve optical density. Gold had a stable predictable growth and achieving an OD$_{600nm}$ of 0,6/0,8 2 hours after inoculation with 10% preculture or 3 hours with 2.5% inoculum. In comparison, pLysS with 10% inoculum used 4 to 6 hours reaching desired density. Hence, the *E. coli* BL21-Gold competent cells were selected as expression host.

A series of expression parameters was tested during protein expression to increase the success for soluble proteins for the two selected AMOR-CBM9 targets. In addition to standard conditions at 37°C 4 hours, lowering the incubation temperature to 20°C and 16°C overnight was performed. These growth conditions were performed to see the effect on expression levels and the effect on potential soluble proteins. The optimization tests showed that expression at 20°C overnight was the lowest temperature with detected expression of target proteins, even if increasing the IPTG concentration from 0,1mM to 0,2mM at 16°C overnight. Visible expression on SDS-PAGE was seen on cultures tested between 20°C and 37°C, glycylglycine had no effect on solubility (Figure 11), while 16°C was tested twice with no visible expression of the target protein. AMOR-CBM9-15 had a theoretical pI of 6.65, and was tested with lysis buffers of different pH; 7.5, pH 8 and pH 8.5, to try and stabilize the protein in the cell lysate.
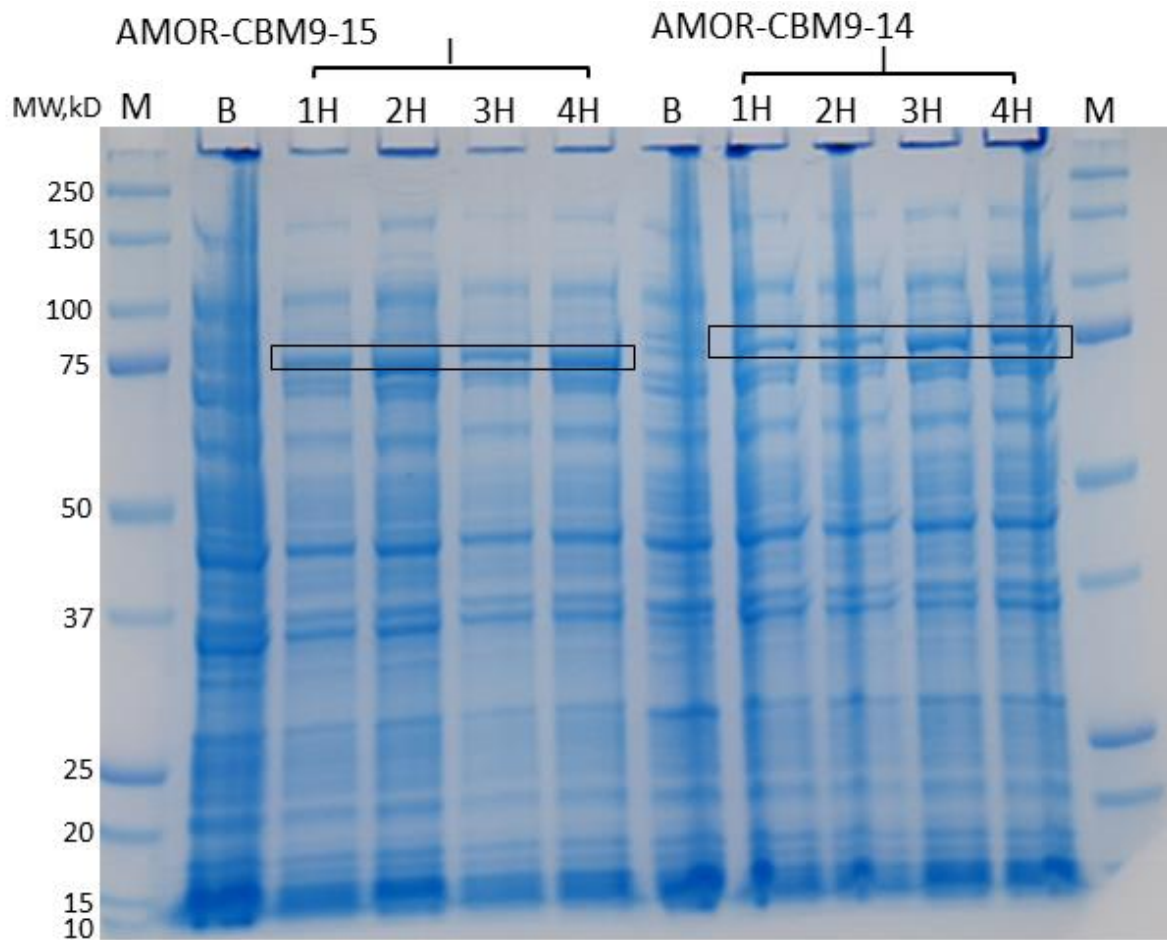
Figure 9: *SDS-PAGE of AMOR-CBM9-15 and AMOR-CBM9-14 expressed with E. coli BL21-Gold (DE3) pLysS competent cells. Samples separated on GenScript ExpressPluss PAGE Gels 10%. Target protein expressed is highlighted with black boxes around the bands. M (marker); BIO-RAD precision plus protein dual color standards, B (before induction); fraction before expression, I (induced); fraction taken each hour (H) after expression was induced. Expression was performed at 37°C for 4 hours.*
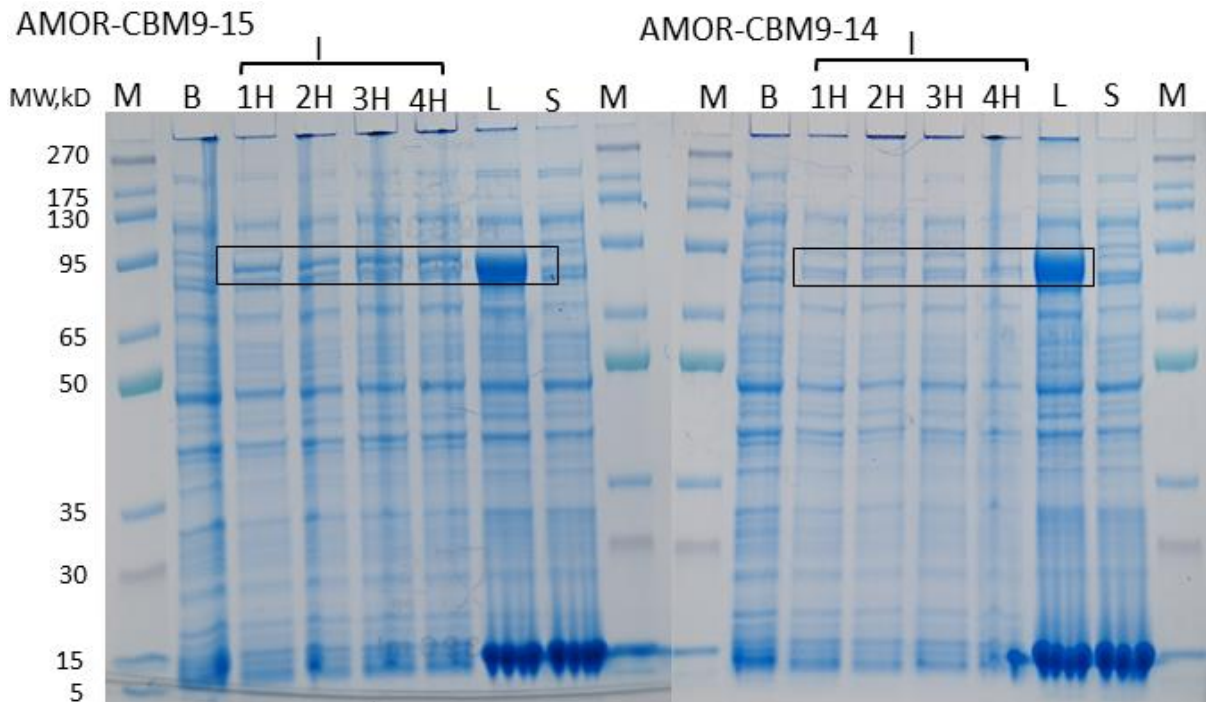
Figure 10: *SDS-PAGE of AMOR-CBM9-15 and AMOR-CBM9-14 expressed with E. coli BL21-Gold (DE3) competent cells. Samples separated on GenScript ExpressPluss PAGE Gels 10%. The target protein is highlighted with a black box around the bands. M (marker); GenScript broad multi pre-stained protein standards, B (before induction); fraction before expression, I (induced) fraction taken each hour (H) after expression was induced, L (lysate); fraction harvested and sonicated, S (soluble); fraction separated from lysate. Expression was performed 37°C 4 hours.*
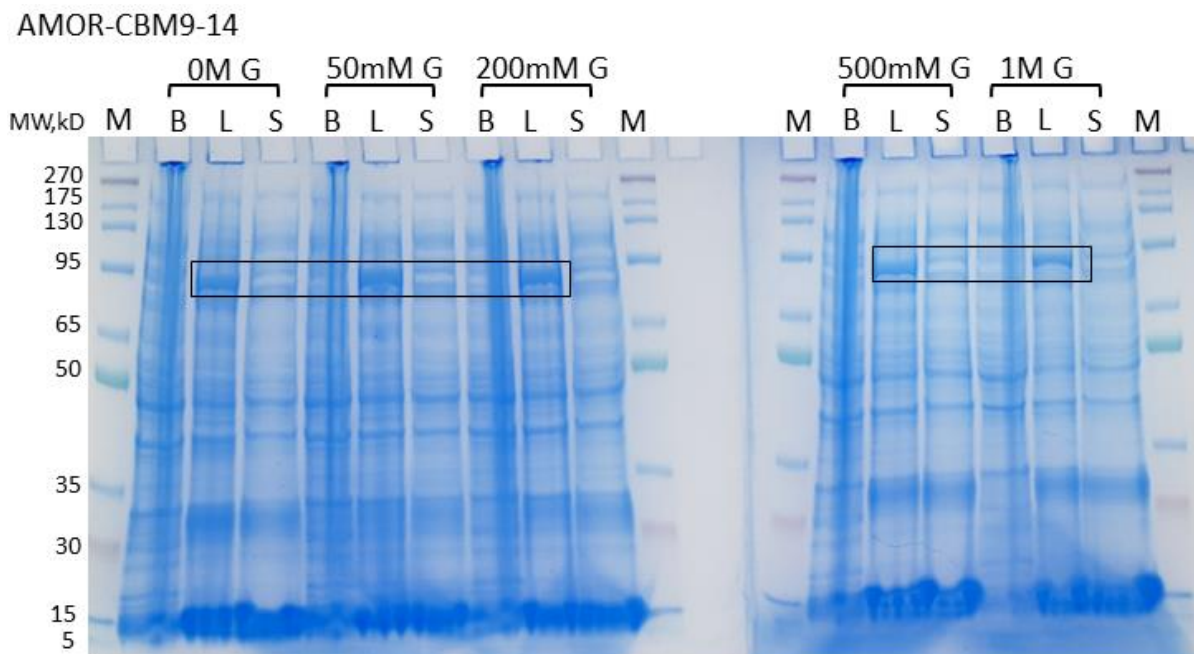
Figure 11: *SDS-PAGE of AMOR-CBM9-14 expressed with Glycylglycine in E. coli BL21-Gold (DE3) competent cells. Samples separated on GenScript ExpressPluss PAGE Gels 10%. The target protein is highlighted with a black box around the bands. M (marker); GenScript broad multi pre-stained protein standards, G (glycylglycine); added in 4 different cultures, ranging from 50mM to 1M, B (before induction); fraction of expression, L (lysate); harvested and sonicated fraction. S (soluble) fraction separated from lysate. Expression was performed 20°C overnight.*

4.5 Purification of CBM9-containing constructs

The first attempt on purifying the AMOR-CBM9-14 showed promising result on the chromatography from the Äkta start with a good peak between elution fraction T5 to T17 (Figure 12.A). When tested on SDS-PAGE there was several different bands with varying size in the observed chromatography peak, with the band at 50 kDa the most abundant (Figure 12.B). Since the target protein has a theoretical molecular weight of 82 kDa, this indicated that observed bands are *E. coli* host proteins. Hence, the purification of AMOR-CBM9-14 was not successfully.

A second run on purifying the AMOR-CBM9-14 was initiated by modifying the elution buffer and binding buffer with 20mM imidazole and 8M urea (Figure 13). The protein sample was applied to the His-trap column directly from the cell lysate to avoid losing the target protein due to aggregation in the solubilization step. The addition of imidazole and urea would denature and avoid *E. coli* protein binding to the His-trap column. Although also denaturing the target protein, this was an effort to increase the possibility that AMOR-CBM9-14 bind to the His-trap column, and subsequently renatured by removing the imidazole and urea by changing the buffer. The chromatography showed a very low peak during the elution process (Figure 13.A). This was compared on an SDS-PAGE gel with proteins from an uninduced sample, the cell lysate, and the soluble fraction from induced cells, which resulted in a single band from T6 to T14 (Figure 13.B). To confirm that this was the AMOR-CBM9-14 target protein, the band was cut from the gel and sent to the Proteomics unit at UIB (PROBE; https://www.uib.no/rg/probe). The mass spectrometry results from PROBE confirmed the presence of AMOR-CBM9-14 and thus, successfully purified (Appendix Proteomic analysis). As a result, denaturation with the addition of imidazole and urea directly to the cell lysate showed promise in the process of purifying the protein compared to the result with the standard protocol.

The third purification run of AMOR-CBM9-14 and the first purification of AMOR-CBM9-15 were adjustments of the second run of AMOR-CBM9-14. Sonicated lysate samples were resuspended in binding buffer B7 (Appendix Buffers), vortexed to mix the solution, and then applied to the His-trap column. The first purification of AMOR-CBM9-15 showed a small elution peak between fraction T7-T13 and bands of similar size as aliquot taken from the cell lysate (Figure 14.A). Bands in the flowthrough showed that the sample only obtained a partly binding to the column and some of the proteins were washed out (Figure 14.B). The third purification of AMOR-CBM9-14 had similar results as the first purification of AMOR-CBM9-15, with a small elution peak between fraction T7-T13 and weak bands on the SDS-PAGE that was also present in the flowthrough fraction (Figure 15). The eluted fractions, T7-T13, for AMOR-CBM9-15 and AMOR-CBM9-14 (Figure 14-15) was collected and concentrated with Amicon ultra centrifugal filter. Elution buffer were changed to storage buffer B16 (Appendix Buffers).
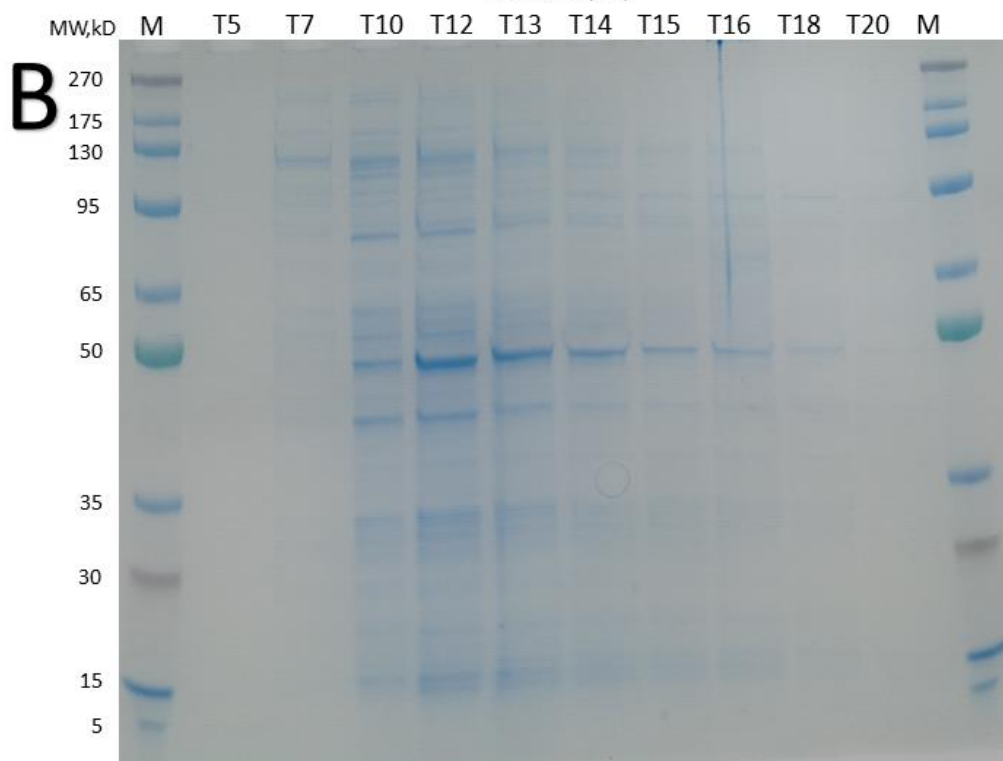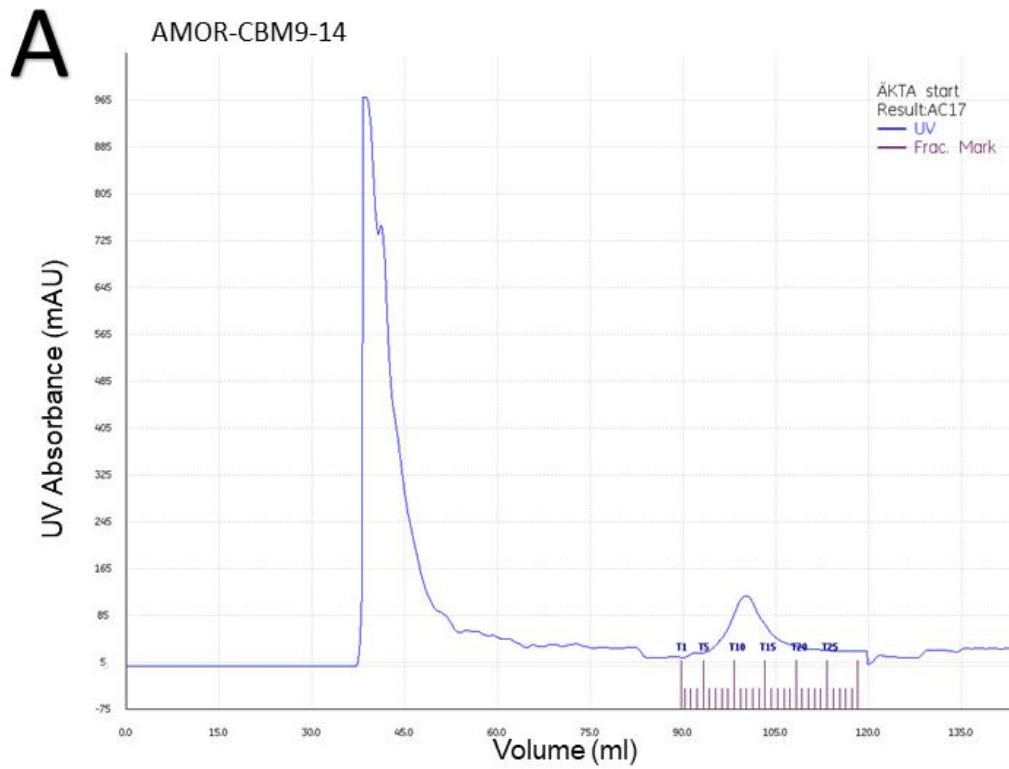
Figure 12: *First purification of AMOR-CBM9-14 using a His-trap column. A) chromatography from Äkta start with the unbound flowthrough peak (36-54 ml) and elution peak (T1-T18), B) SDS-PAGE with elution fractions (T), separated on a 10% GenScript ExpressPluss PAGE gel. M (marker); GenScript broad multi pre-stained protein standard.*
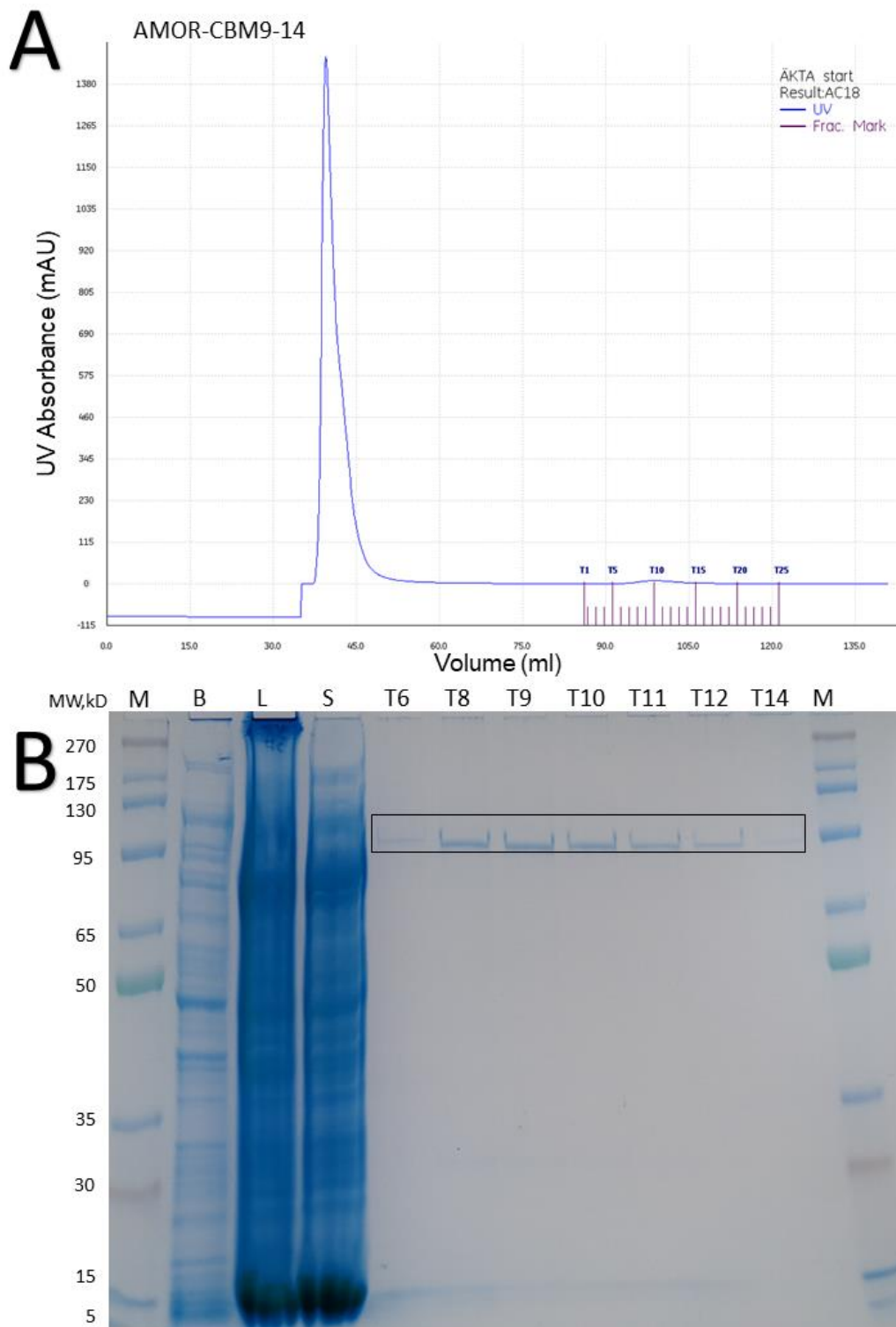
Figure 13*: Second purification of AMOR-CBM9-14 using a His-trap column. A) chromatography from Äkta start with the unbound flowthrough peak (36-52 ml) and elution peak (T5-T15), B) SDS-PAGE with elution fractions (T6-T14), M (marker); GenScript broad multi pre-stained protein standard, B (before induction); fraction of expression, L (lysate); harvested and sonicated fraction, S (soluble) fraction separated from lysate. Samples was separated on a 10% GenScript ExpressPluss PAGE gel. The target protein is highlighted with a black box around the bands.*
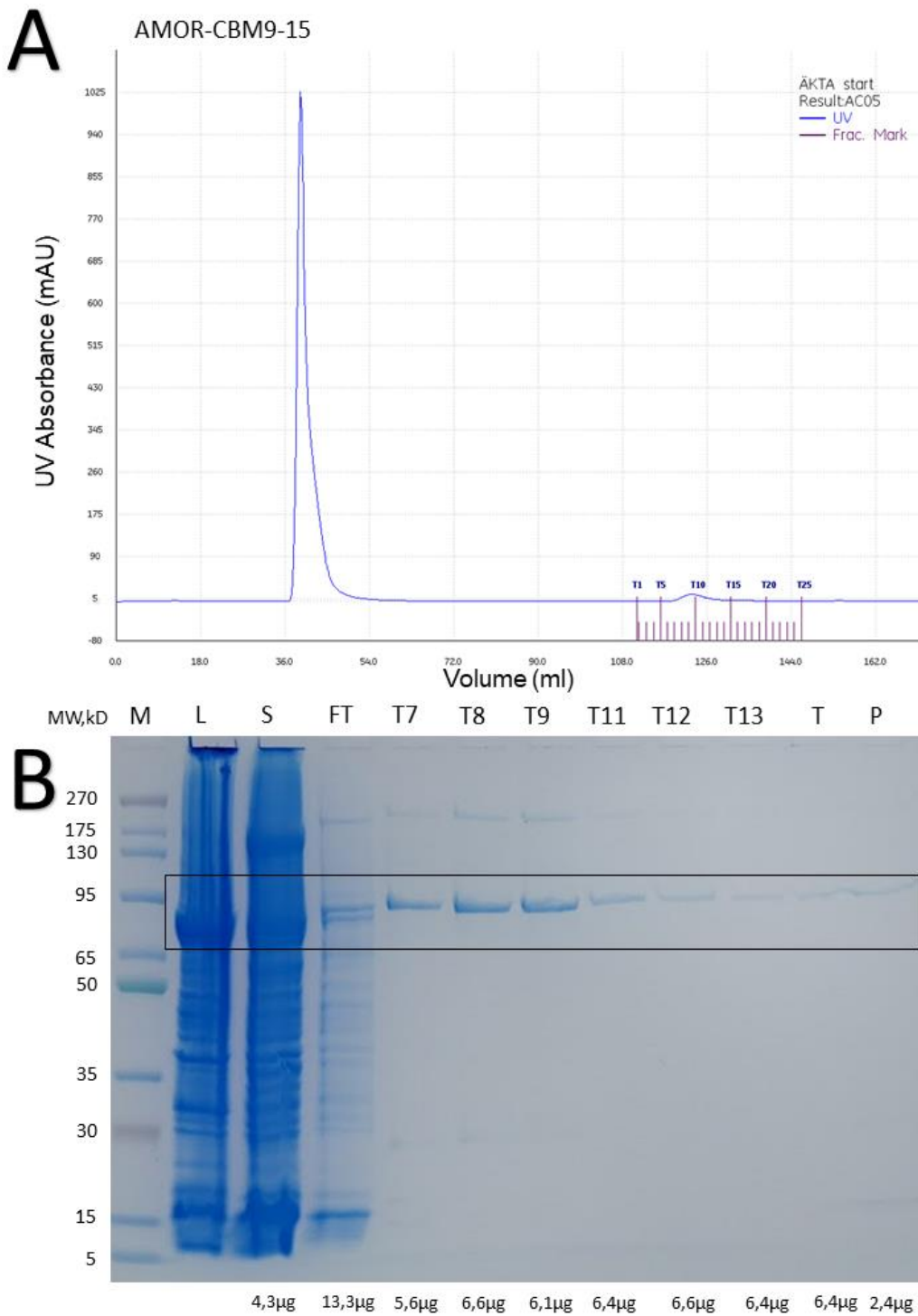
Figure 14: *First purification of AMOR-CBM-15 using a His-trap column. A) chromatography from Äkta start with the unbound flowthrough peak (36-48 ml) and elution peak (T7-T25), B) SDS-PAGE with elution fractions (T7-T13), L (lysate); harvested and sonicated sample, S (soluble); fraction separated from lysate, FT (flow through); unbound fraction, T (elution peak); containing all fraction within the peak, P (protein); elution peak concentrated with centrifugal filtration, M (marker); GenScript broad multi pre-stained protein standard. Samples separated on (GenScript SurePage, Bis-Tris 8-16%), protein mass calculated from protein concentration and amount loaded. The target protein is highlighted with a black box around the bands.*
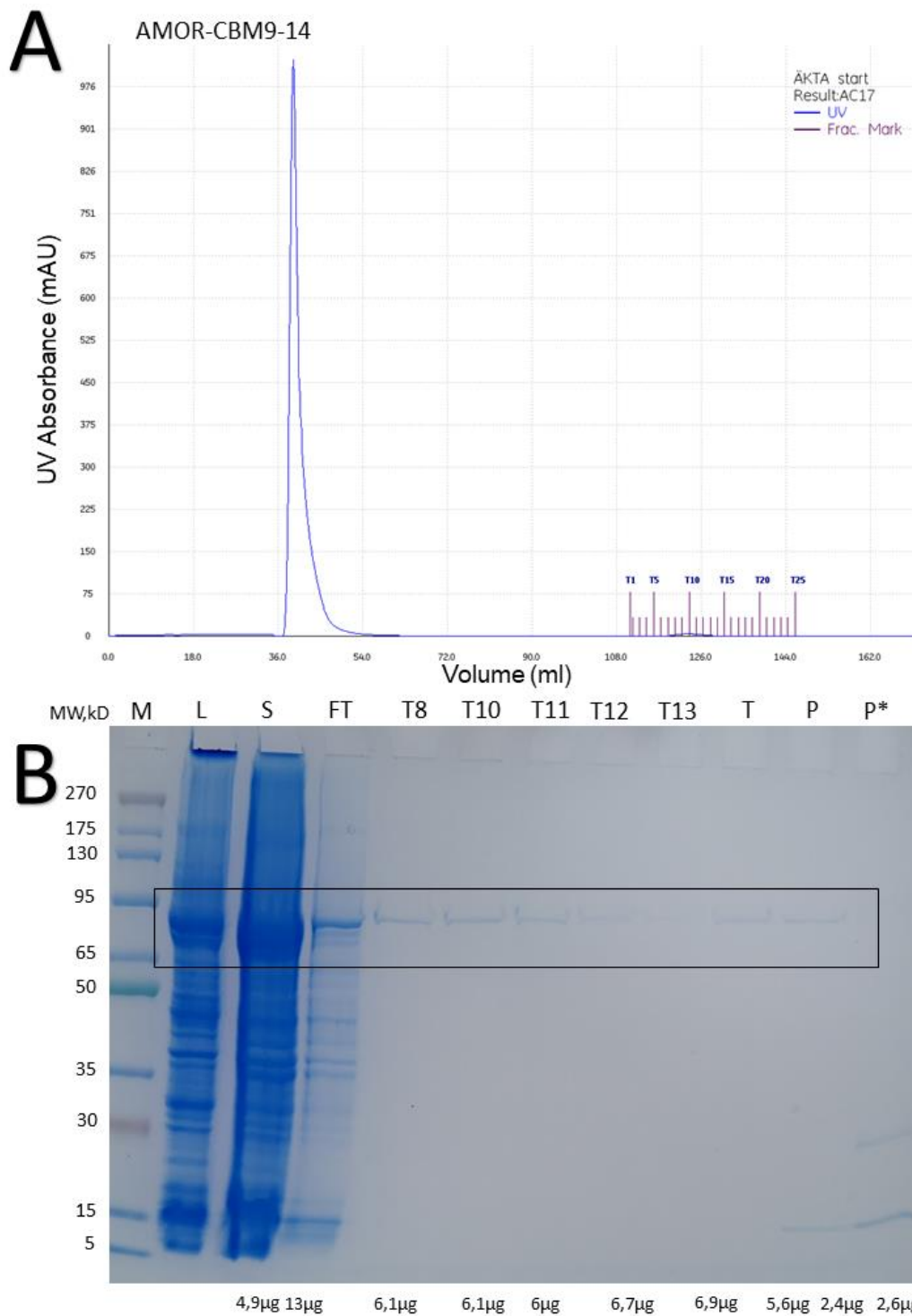
Figure 15: *Third purification of AMOR-CBM9-14 using a His-trap column. A) chromatography from Äkta start with the unbound flowthrough peak (36-54 ml) and elution peak (T7-T13), B) SDS-PAGE with elution fractions (T7-T13), L (lysate); harvested and sonicated sample, S (soluble); fraction separated from lysate, FT (flow through); unbound fraction, T (elution peak); containing all fraction within the peak, P (protein); elution peak concentrated with centrifugal filtration, P* (protein) from second purification included for comparison, M (marker); GenScript broad multi pre-stained protein standard. Samples separated on (GenScript SurePage, Bis-Tris 8-16%), protein mass calculated from protein concentration and amount loaded. The target protein is highlighted with a black box around the bands.*

The control xylanase characterized in Fredriksen *et al. (Fredriksen et al., 2019)*. AMOR-GH10A, was purified during the first run (Figure 16). With a large band in the soluble fraction, no modification was done to the purification process (Figure 16.B). From the chromatography the protein was eluted between fraction T7 to T25 (Figure 16.A). Fraction T9 to T16 was collected for concentration with the Amicon ultra-4 centrifugal filters. These fractions were chosen since they had the highest absorbance indicating that these contained most of the protein. Elution fractions were tested on SDS-PAGE with samples from the cell lysate, soluble fraction and flowthrough (Figure 16.B). The flowthrough fraction showed that most of the protein bind to the column while present in all the elution fraction T9 to T16. During the protein concentration, the buffer was changed from the elution buffers to a storage buffer B16 (Appendix Buffers). The concentration of the final purified AMOR-GH10A (1.2 ml) was 2.66mg/ml.
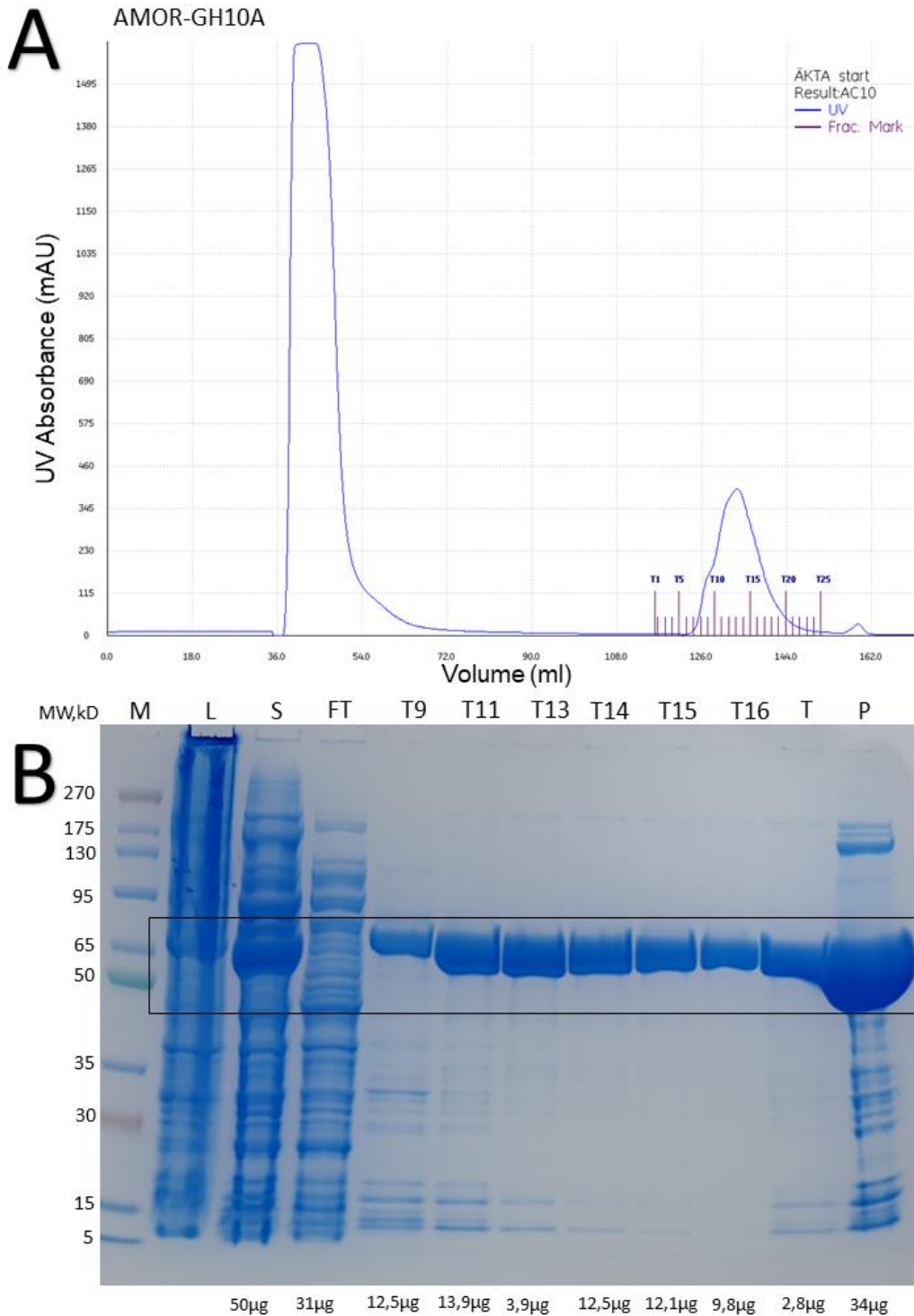
Figure 16: *Purified AMOR-GH10A using a His-trap column. A) chromatography from Äkta start with the unbound flowthrough peak (36-72 ml) and elution peak (T7-T25), B) SDS-PAGE with elution fractions (T9-T16), L (lysate); harvested and sonicated sample, S (soluble); fraction separated from lysate, FT (flow through); unbound fraction, T (elution peak); containing all fraction within the peak, P (protein); elution peak concentrated with centrifugal filtration, M (marker); GenScript broad multi pre-stained protein standard. Samples separated on (GenScript SurePage, Bis-Tris 8-16%), protein mass calculated from protein concentration and amount loaded. The target protein is highlighted with a black box around the bands.*

4.6 Plate assay for enzyme xylanase activity

Enzyme activity was tested on agar plates containing 1% xylan (Figure 17). 2µg of purified protein of AMOR-CBM9-14 (second and third purification) and AMOR-CBM9-15 (first purification) was used for the agar plate assay. The purified AMOR-GH10A xylanase was used as a positive control and the purified 12AMOR1-GH13 amylase was used as a negative control. After incubation at 65°C for 4 hours, the plate was inspected for clearing zones around the wells. The positive control, AMOR-GH10A, showed a visible clearance zone around the well, indicating activity for xylan degradation (Figure 17, nr .1). In comparison, the purified target enzymes, AMOR-CBM9-14 and AMOR-CBM9-15 (Figure 17, nr.3-5), and the negative GH13 control (Figure 17, nr.2), showed no indication of activity with the absence of discoloration around the wells on the plate.
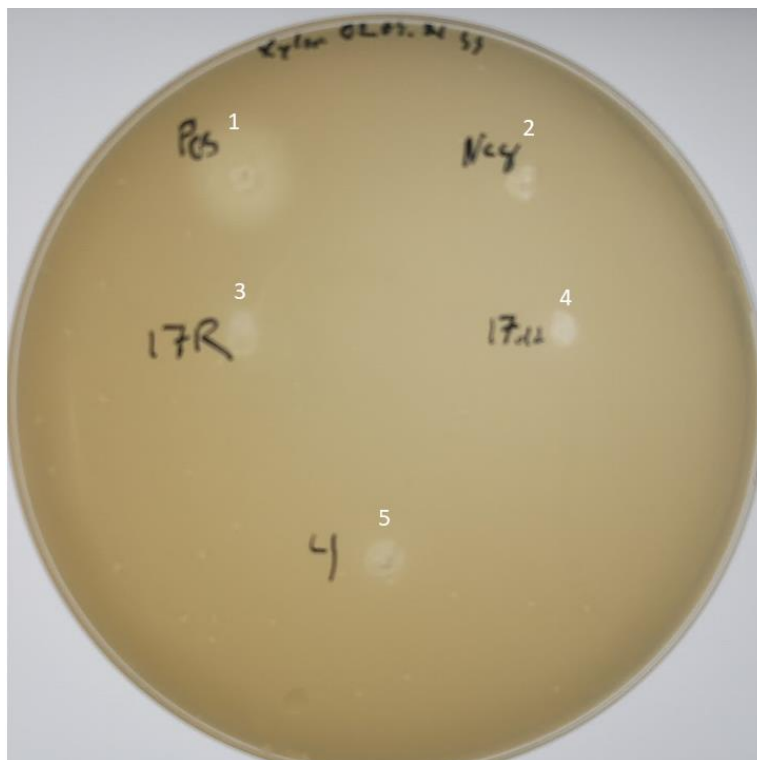
Figure 17: Plate assay on agar plates containing 1% xylan and phosphate buffer pH 5.8 loaded with 2µg protein in each well and incubated 65°C 4H. 1 positive control AMOR-GH10A. 2 negative control 12AMOR1-GH13. AMOR-CBM9-14 second run. AMOR-CBM9-14 third run. AMOR-CBM9-15.

4.7 Expression of α-amylase constructs

The five α-amylase constructs were first tested with expression at 20°C overnight. Expression was visible on lysate fraction for all the constructs. A second try was done with expression at 16°C overnight with the successful expression of all five constructs (Figure 18-19). In addition, AMOR-GH57-2 showed a band in the supernatant (S), indicating that the protein was soluble.
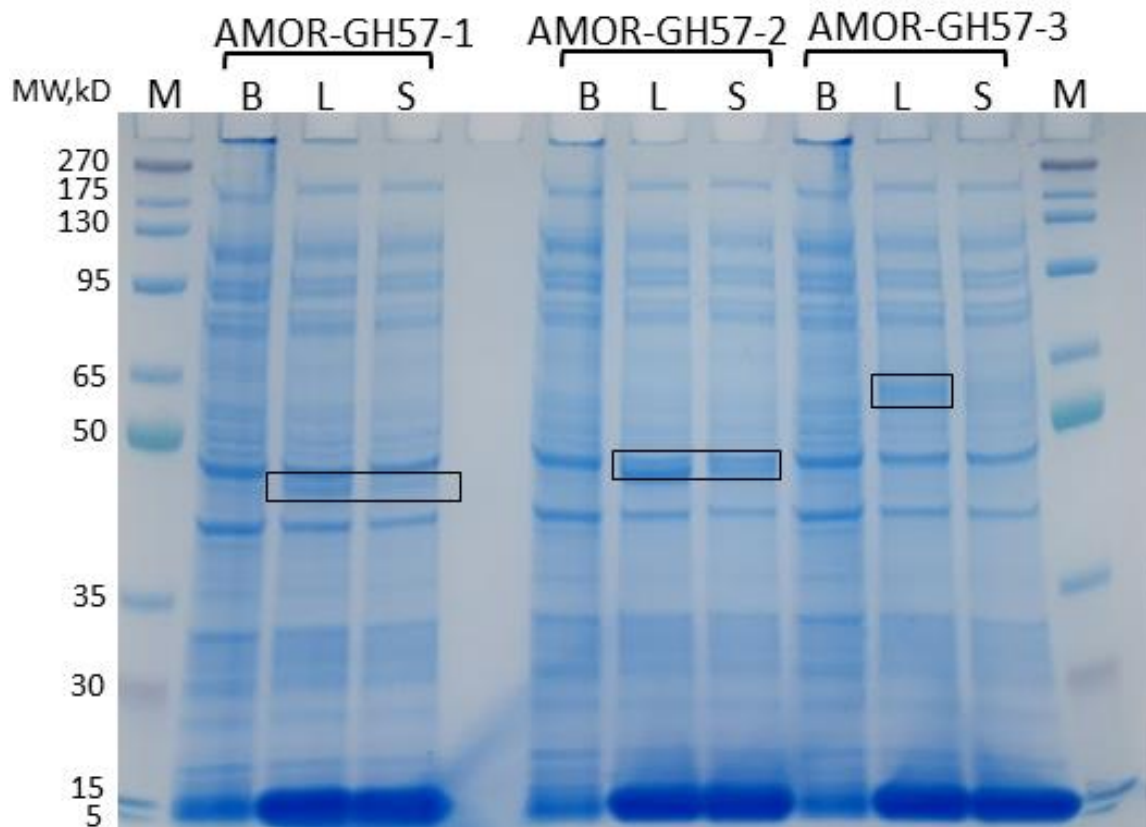
Figure 18 *SDS-PAGE of AMOR-GH57-1, AMOR-GH57-2 and AMOR-GH57-3 expressed with E. coli BL21-Gold (DE3) competent cells. Samples separated on (GenScript ExpressPlus 10%), M (marker); GenScript broad multi pre-stained protein standards, B (before induction); fraction before expression, L (lysate); harvested and sonicated samples, S (soluble); fraction separated from lysate. Target proteins expressed is highlighted with black boxes around the bands. Expression was performed 16°C overnight.*
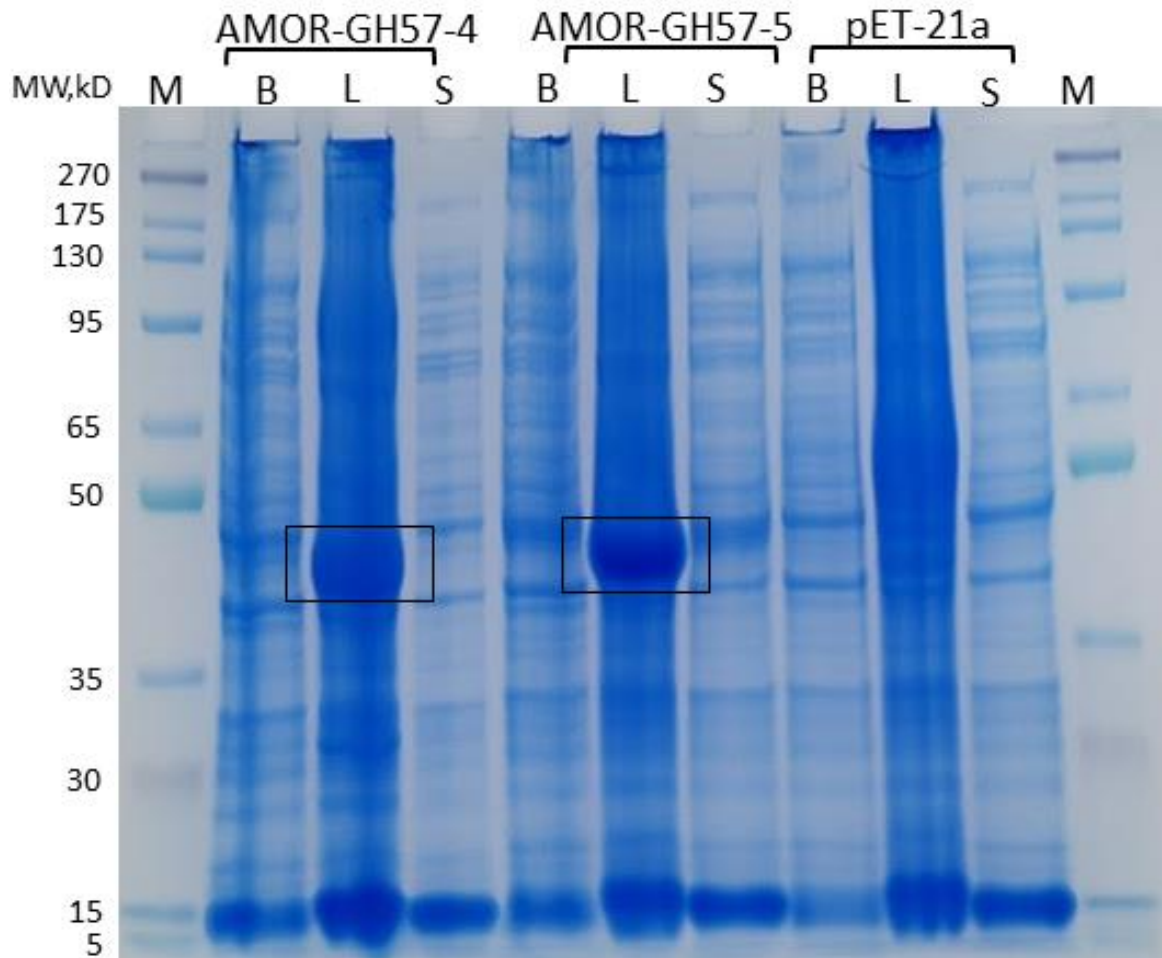
Figure 19 *SDS-PAGE of AMOR-GH57-4, AMOR-GH57-5 and pET-21a control expressed with E. coli BL21-Gold (DE3) competent cells. Samples separated on (GenScript ExpressPlus 10%), M (marker); GenScript broad multi pre-stained protein standards, B (before induction); fraction before expression, L (lysate); harvested and sonicated samples, S (soluble); fraction separated from lysate. Target proteins expressed is highlighted with black boxes around the bands. Expression was performed 16°C overnight.*

In order to enhance the solubility of AMOR-GH57-4 during expression, glycylglycine was tested to enhance solubility, concentration of (0-1M) was added in the expression media. When comparing proteins in lanes before induction (Figure 20.(B)), with proteins in the cell lysate (L) and soluble (S) of AMOR-GH57-4, the target protein is visible in the both the cell lysate and supernatant (Figure 20). Thus, indicating that the target protein is expressed and soluble, respectively. Furthermore, the addition of glycylglycine also had an effect on the growth of the cultures. Cultures with 0M and 50mM glycylglycine reach target optical density (OD) of 0.6-0.8 $_{600nm}$ after 4 hours, while the addition of 200mM glycylglycine used 5 hours and 0.5M glycylglycine used 6.5 hours. 1M glycylglycine grew very slowly and was induced after 6.5 hours at an OD of 0.24 $_{600nm}$.
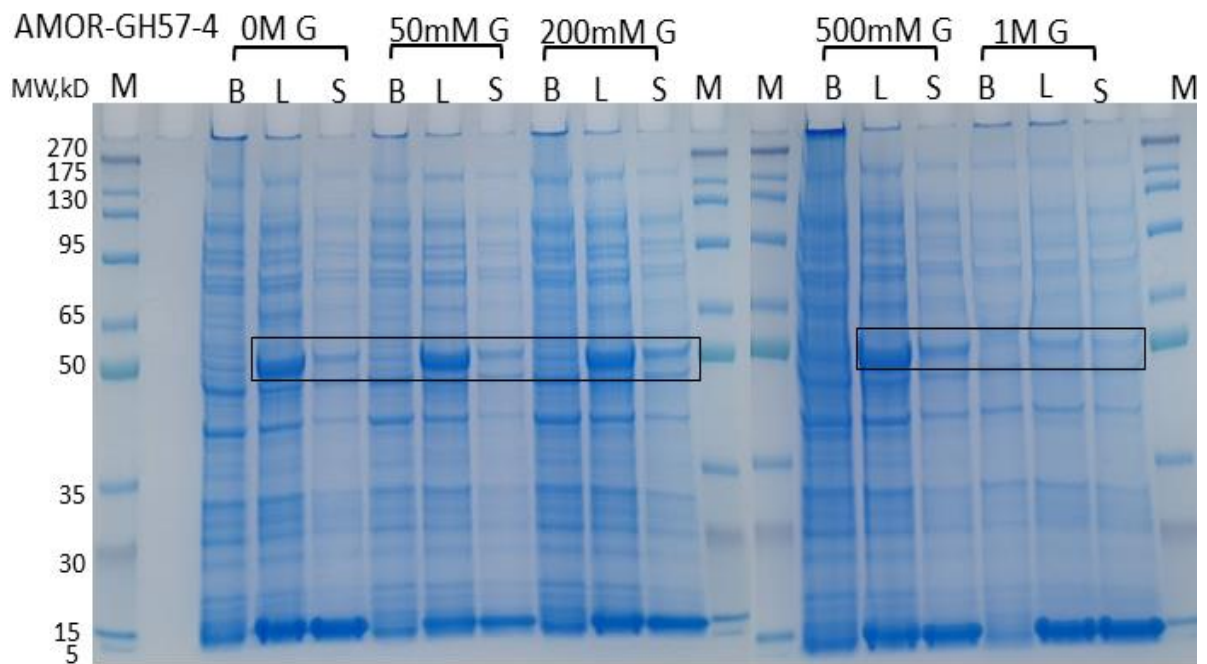
Figure 20 SDS-PAGE of AMOR-GH57-4 expressed with E. coli BL21-Gold (DE3) competent cells with the addition of glycylglycine. Samples separated on (GenScript ExpressPlus 10%), M (marker); GenScript broad multi pre-stained protein standards, B (before induction); fraction before expression, L (lysate); harvested and sonicated samples, S (soluble); fraction separated from lysate, G (glycylglycine) added in expression culture ranging from 50mM to 1M concentration. Target proteins expressed is highlighted with black boxes around the bands. Expression was performed 16°C overnight.

The addition of glycylglycine to the growth media was also tested for the 4 other GH57 target proteins, however limited to 0M, 50mM and 200mM glycylglycine due to the reduced growth rate of the expression host at higher concentrations. AMOR-GH57-2 got similar results as AMOR-GH57-4 with the band visible in the cell lysate and soluble fraction (Figure 21).
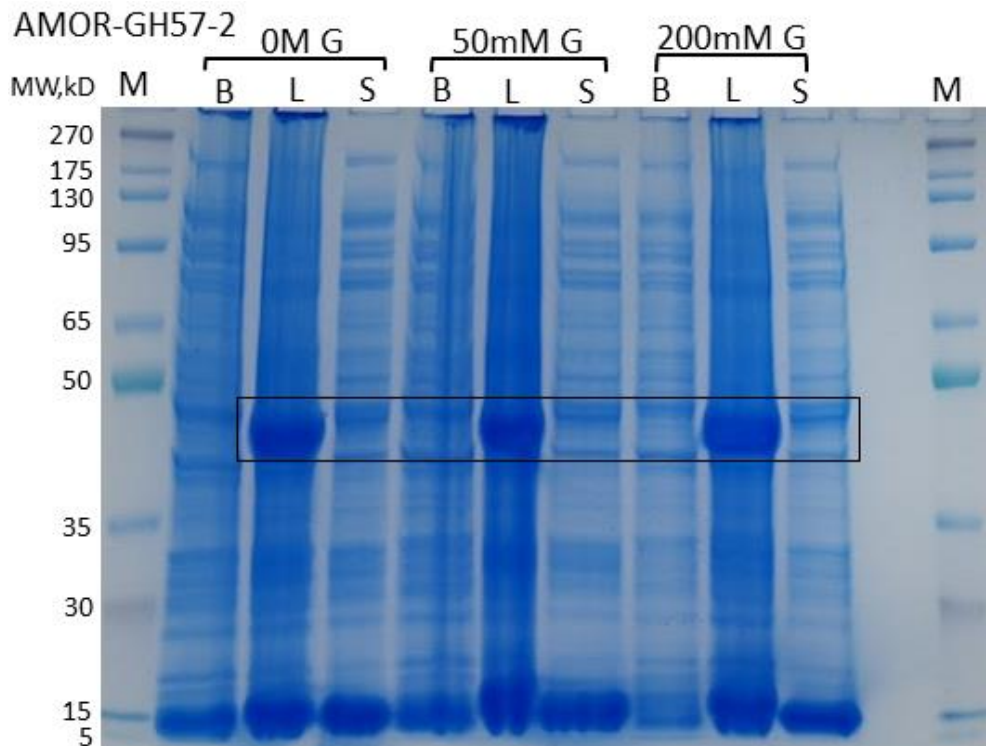
Figure 21: *SDS-PAGE of AMOR-GH57-2 expressed with E. coli BL21-Gold (DE3) competent cells with the addition of glycylglycine. Samples separated on (GenScript ExpressPlus 10%), M (marker); GenScript broad multi pre-stained protein standards, B (before induction); fraction before expression, L (lysate); harvested and sonicated samples, S (soluble); fraction separated from lysate, G (glycylglycine) added in expression culture ranging from 50mM to 200mM concentration. Target proteins expressed is highlighted with black box around the bands. Expression was performed 16°C overnight.*

4.8 Purification of GH57 and GH13 α-amylases

The AMOR-GH57-2 and AMOR-GH57-4 was subjected to heat treatment to test if the target proteins would not precipitate and could be used as an additional purification step (Figure 22). The first heat treatment at 60°C and 70°C was not conclusive since the SDS-PAGE gel showed bands drifting between the wells. A second attempt for each construct was performed with 1ml of flowthrough from protein purification for 30min at 70°C and 80°C and compared to a supernatant that was not heat treated. Results showed heat treatment precipitated both *E. coli* proteins and both AMOR-GH57-2 and AMOR-GH57-4 (Figure 22).
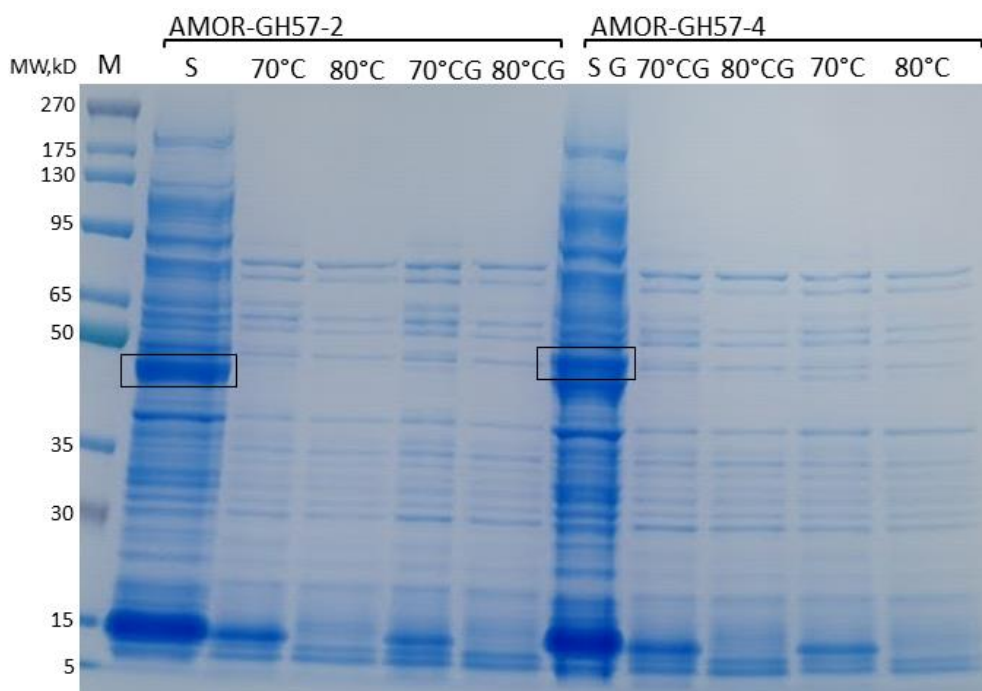
Figure 22 SDS-PAGE of heat-treated AMOR-GH57-2 and AMOR-GH57-4. Protein samples heated 30 min in thermomixer, tested at (70°C) and (80°C), G (glycylglycine); samples expressed with 200mM glycylglycine, S (sample) unbound flowthrough fraction from protein purification, M (marker) GenScript broad multi pre-stained protein standards. Samples separated on (GenScript SurePage, Bis-Tris 8-16%), precipitated protein were removed by centrifugation before loading on gel.

Both AMOR-GH57-2 and AMOR-GH57-4 were chosen to be purified since these had the largest bands in the soluble supernatant fraction after expression. First purification was done with samples that was expressed with 200mM glycylglycine at 16°C overnight. The binding buffer was modified with 25mM imidazole to avoid *E. coli* proteins to bind to the column. The first purification for both samples showed no visible elution on the chromatography and no visible bands on the SDS-PAGE. Inspecting the flowthrough fraction, showed that the protein did not bind to the his-trap column. The purification was re-run under identical conditions to test if there was a user error, however, the result was identical with no protein binding the his-trap column. A third purification was performed from cells expressed without glycylglycine to test if this prevented binding to the column. Both AMOR-GH57-2 and AMOR-GH57-4 was washed out with the flowthrough with no difference in results against samples with glycylglycine. The flowthrough from AMOR-GH57-4 third run was diluted in binding buffer without imidazole, with a total concentration of 10mM imidazole in the sample. Then purified again as fourth run with binding buffer contain 10mM imidazole and a flowrate at 2ml/min. However, similar results were observed, with the target protein washed out in the flowthrough with the unbound *E. coli* proteins.

Purification of denatured target proteins was then tested using binding and elution buffers containing 8M urea. AMOR-GH57-2 had a small elution peak in the chromatography and some visible gel bands in the elution fractions (Figure 23.A).
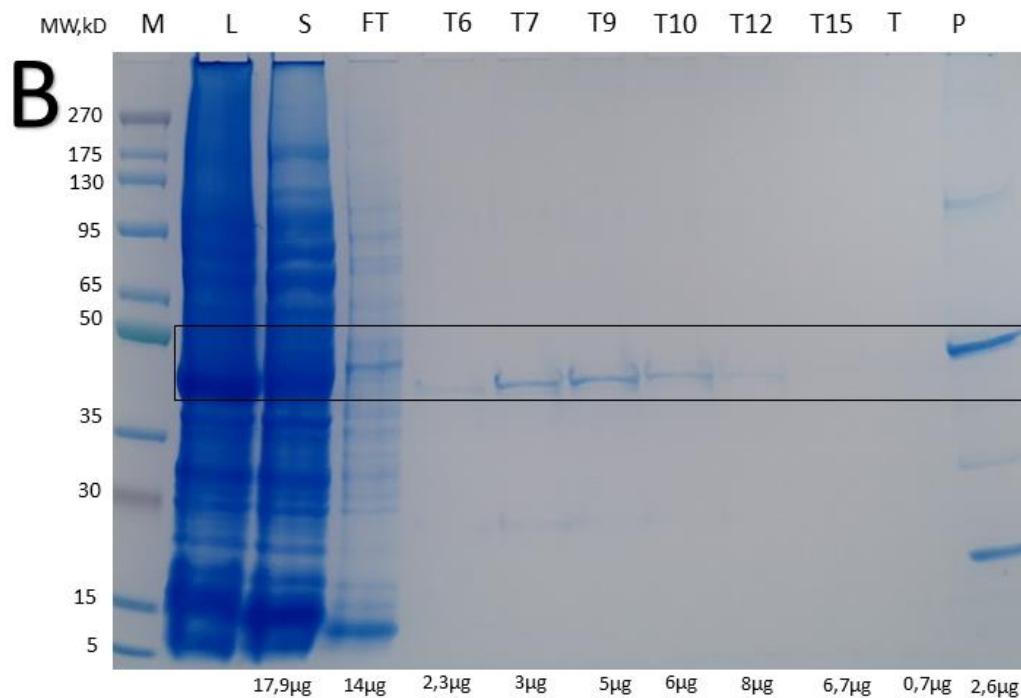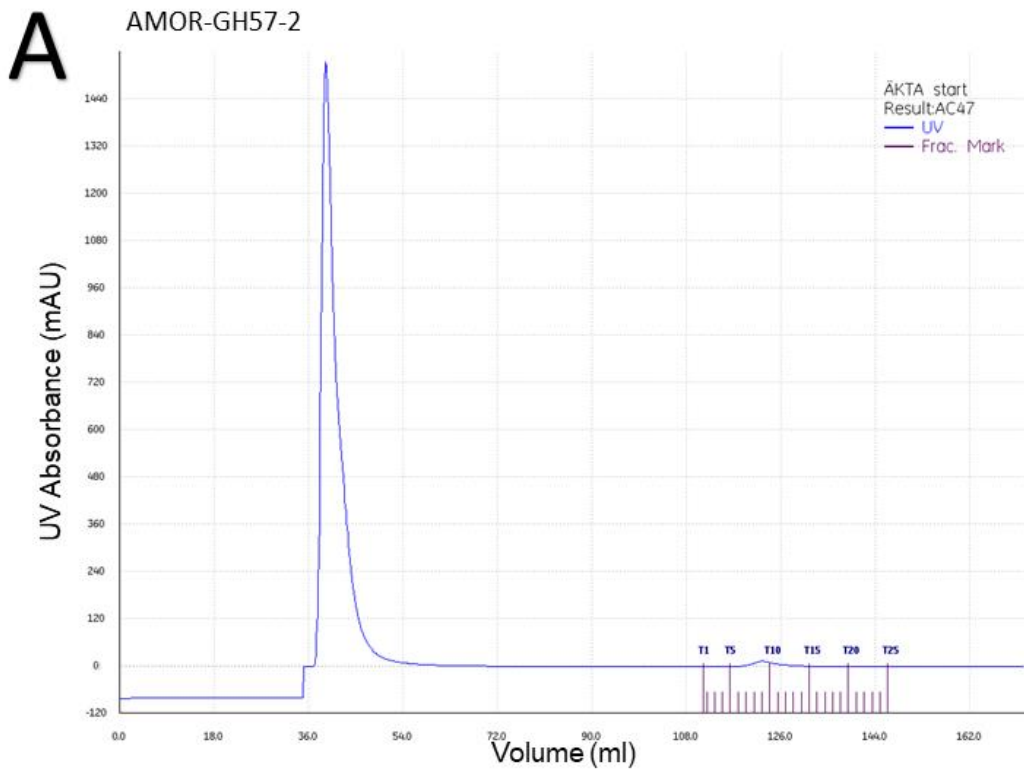
Figure 23: *Fifth run purifying AMOR-GH57-2 using a His-trap column. A) chromatography from Äkta start with the unbound flowthrough peak (36-54 ml) and elution peak (T6-T15), B) SDS-PAGE with elution fractions (T9-T16), L (lysate); harvested and sonicated sample, S (soluble); fraction separated from lysate, FT (flow through); unbound fraction, T (elution peak); containing all fraction within the peak, P (protein); elution peak concentrated with centrifugal filtration, M (marker); GenScript broad multi pre-stained protein standard. Samples separated on (GenScript SurePage, Bis-Tris 8-16%), protein mass calculated from protein concentration and amount loaded. Target protein are highlighted with black box around the bands.*

However, the target protein, AMOR-GH57-2, as determined by size, was also present in the flowthrough (Figure 23.B). Hence, partly binding to the column. AMOR-GH57-4 had similar results with a small elution peak with two weak bands T11 and T13, and presence in flowthrough fraction (Figure 24).
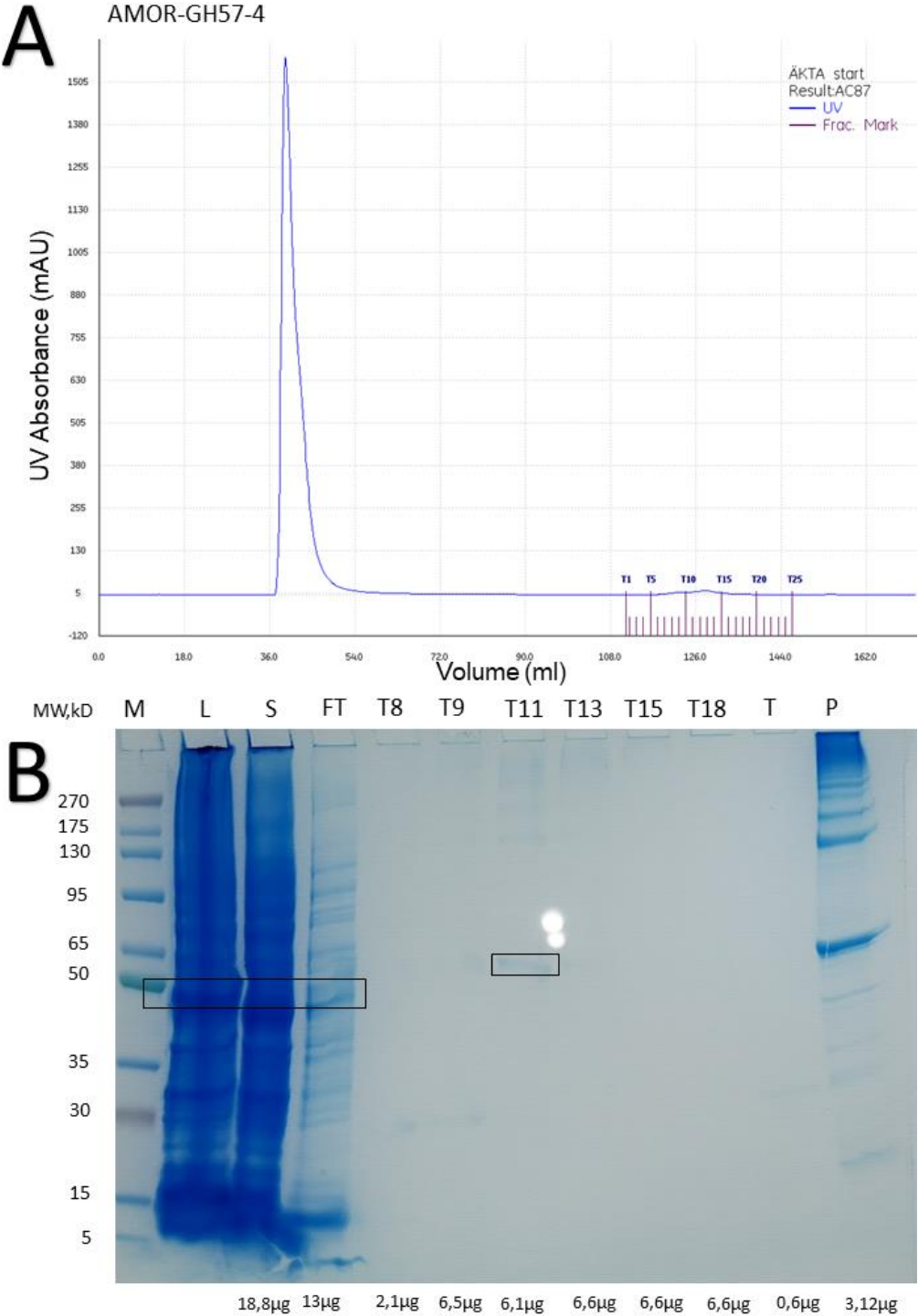
Figure 24: *Fourth run purifying AMOR-GH57-4 using a His-trap column. A) chromatography from Äkta start with the unbound flowthrough peak (36-72 ml) and elution peak (T8-T18), B) SDS-PAGE with elution fractions (T9-T16), L (lysate); harvested and sonicated sample, S (soluble); fraction separated from lysate, FT (flow through); unbound fraction, T (elution peak); containing all fraction within the peak, P (protein); elution peak concentrated with centrifugal filtration, M (marker); GenScript broad multi pre-stained protein standard. Samples separated on (GenScript SurePage, Bis-Tris 8-16%), protein mass calculated from protein concentration and amount loaded. Target protein are highlighted with black box around the bands.*

AMOR-GH57-2 and AMOR-GH57-4 were renatured with protein dialysis overnight and concentrated with centrifugal filtration, resulting in 200µl of AMOR-GH57-2 with a concentration of 102µg/ml and 200µl of AMOR-GH57-4 with a concentration of 122µg/ml.

The control amylase 12AMOR1-GH13 was purified as previously described in Wissuwa et al. (Wissuwa et al., 2016), using an additional heat treatment step on the cell lysate which removed some of the *E. coli* proteins in the sample and not effecting the target protein (Figure 25). Although a weak band was observed in the flowthrough at 65kDa, most of the GH13 protein bound to the column. After centrifugal filtration, 500µl of 12AMOR1-GH13 with a concentration of 15,1mg/ml was obtained.
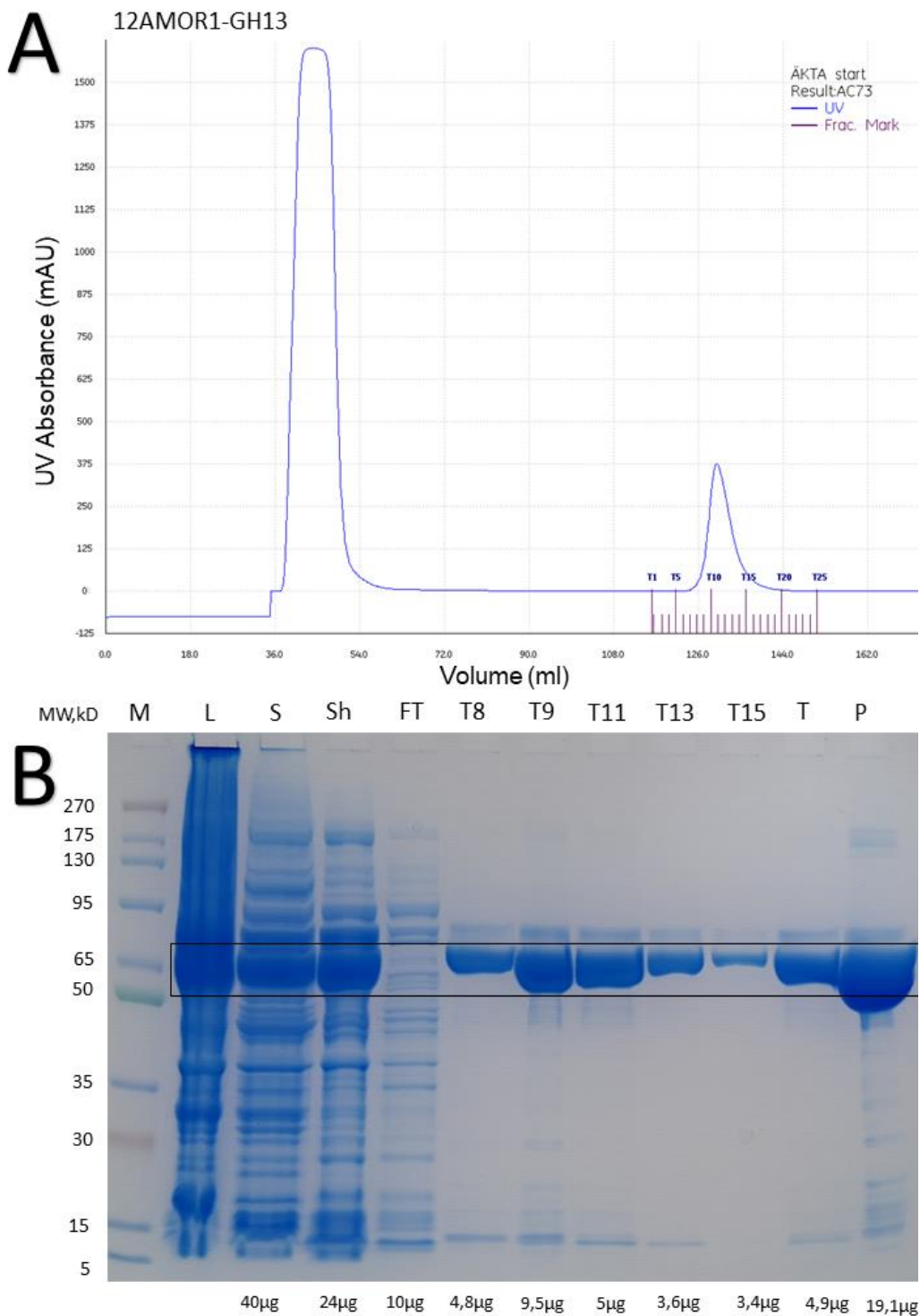
Figure 25*: Purified 12AMOR1-GH13 using a His-trap column. A) chromatography from Äkta start with the unbound flowthrough peak (36-72 ml) and elution peak (T7-T25), B) SDS-PAGE with elution fractions (T9-T16), L (lysate); harvested and sonicated sample, S (soluble); fraction separated from lysate, FT (flow through); unbound fraction, T (elution peak); containing all fraction within the peak, P (protein); elution peak concentrated with centrifugal filtration, M (marker); GenScript broad multi pre-stained protein standard. Samples separated on (GenScript SurePage, Bis-Tris 8-16%), protein mass calculated from protein concentration and amount loaded. Target protein are highlighted with black box around the bands.*

4.9 Gel filtration and molecular weight calculation of 12AMOR1-GH13

Gel filtration of 12AMOR1-GH13 without standards showed that the protein sample from the purification contained low contamination. During gel filtration, a small amount of a larger protein was eluted before 12AMOR1-GH13 as seen in the chromatogram (Figure 26.A).
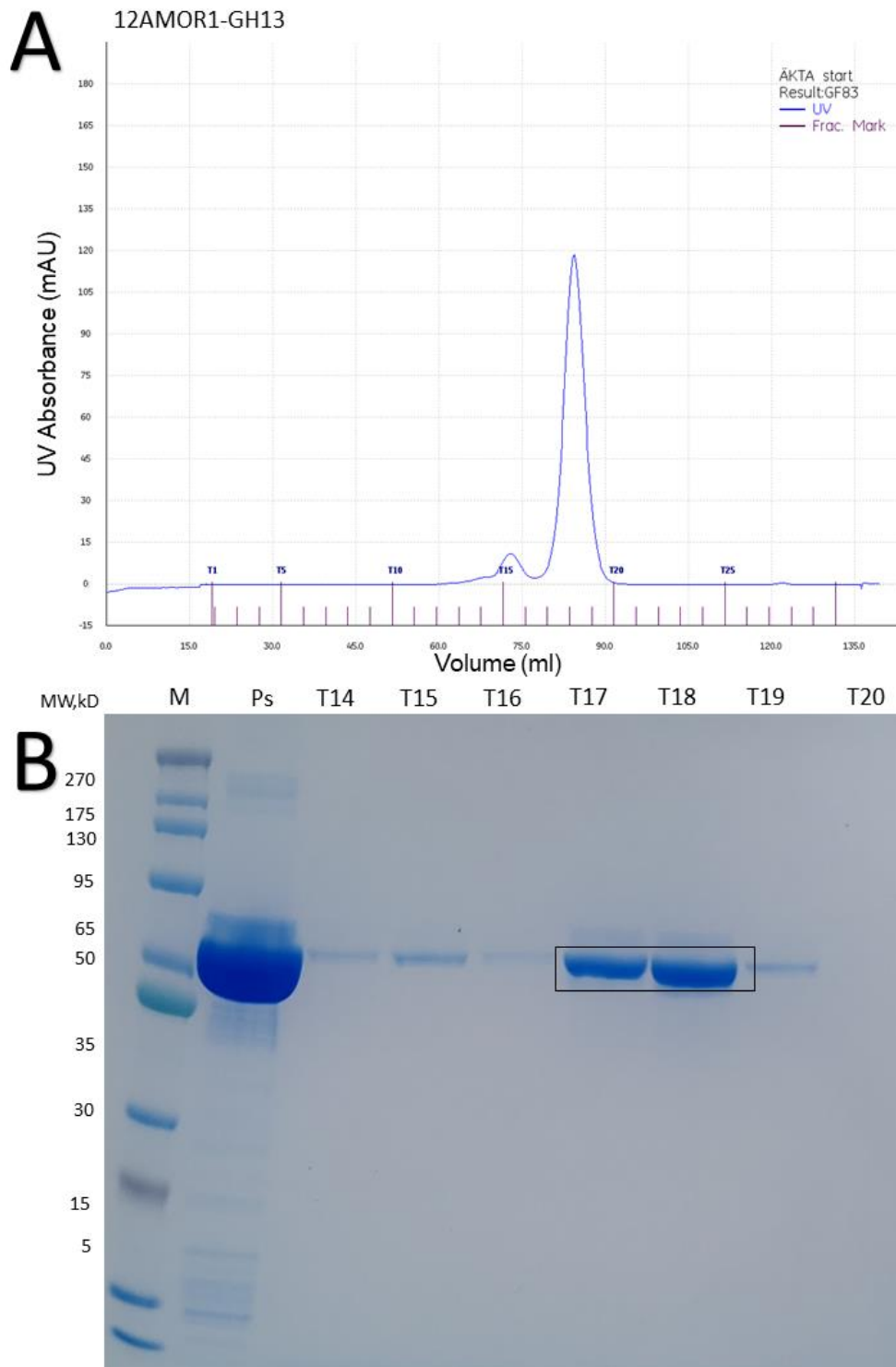
Figure 26 Gel filtration of 12AMOR1-GH13 without standards. A) chromatogram from Äkta start with HiLoad 16/600 Superdex200pg column, with exclusion peak (T17-T19). SDS-PAGE with exclusion fractions (T14-T20), Ps (protein sample); purified with His-trap column and concentrated, T14-T20 (exclusion fractions);from gel filtration, M (marker); GenScript broad multi pre-stained protein standard. Samples separated on (GenScript SurePage, Bis-Tris 8-16%). Target protein is highlighted with black box around the bands.

Fraction T17 and T18 was then concentrated with ultra-centrifugation. To calculate the molecular weight of 12AMOR1-GH13, a second gel filtration was performed spiked with standards (Bio-Rad). Chromatography was used to calculate the elution volume of the standards to get K_{ave} values (Figure 27.A).
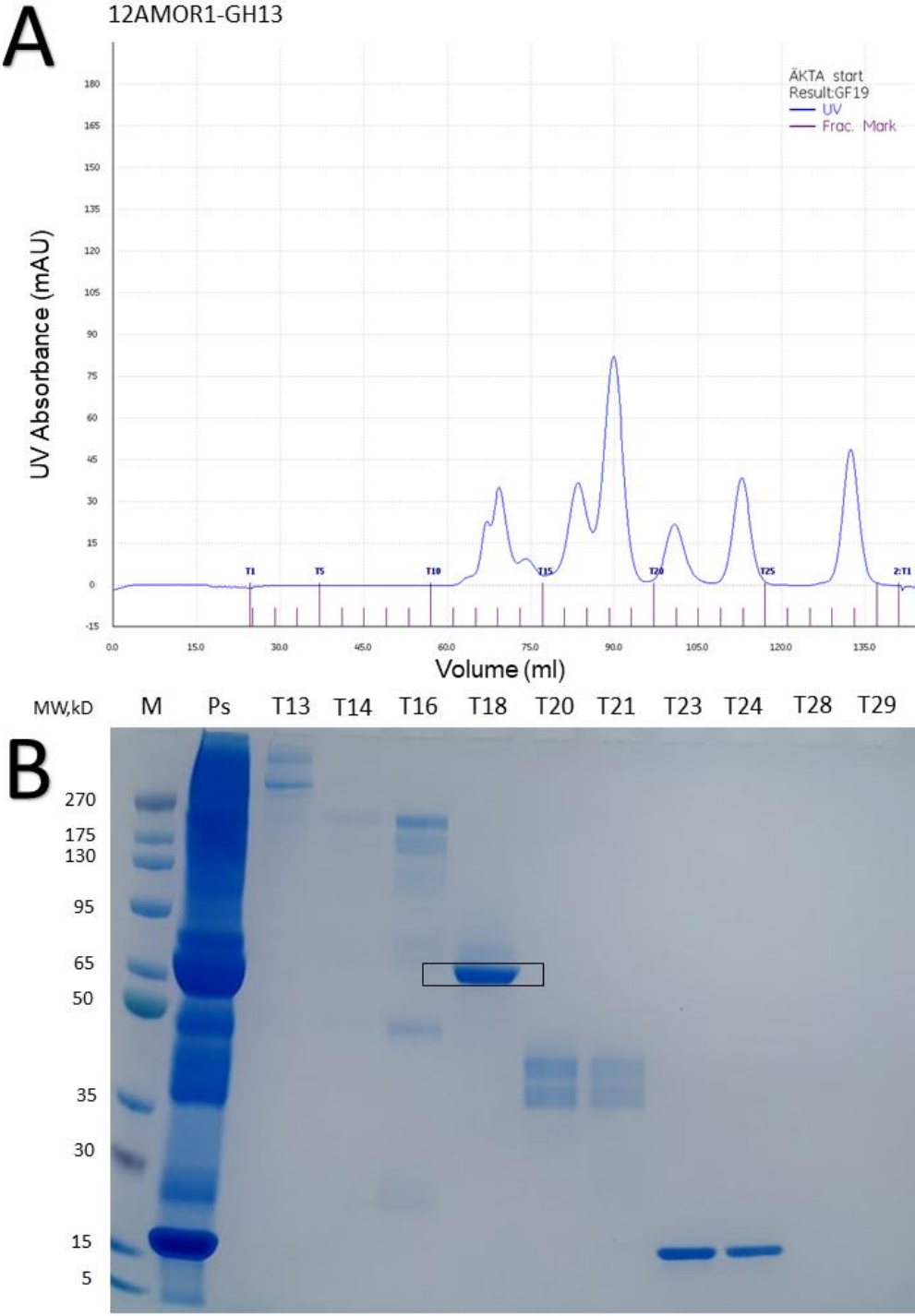
Figure 27: *Gel filtration of 12AMOR1-GH13 standards (BioRad). A) chromatogram from Äkta start with HiLoad 16/600 Superdex200pg column, with exclusion peak (T11-T29). SDS-PAGE with exclusion fractions (T14-T20), Ps (protein sample); purified with His-trap column and protein standards, T13-T29 (exclusion fractions); from gel filtration, M (marker); GenScript broad multi pre-stained protein standard. Samples separated on (GenScript SurePage, Bis-Tris 8-16%). Target protein is highlighted with black box around the band.*

These values were plotted in a scatter plot with log of the molecular weights of the protein standards. The trendline was further used to calculate the molecular weight of 12AMOR1-GH13 (Appendix Gel filtration scatter plot). The molecular weight was estimated to be 105.5 kDa. The oligomeric state of 12AMOR1-GH13 is estimated to a dimer, as native state is almost double of the denatured state.

4.10 Plate assay for α-amylase activity

The purified protein from denaturation purification of AMOR-GH57-2 and AMOR-GH57-4 was tested with 12AMOR1-GH13 for activity on agar plates containing 1% starch, using α-amylase from *Bacillus lichenifromis* (SIGMA-ALDRICH, St Louis, USA) as positive control and AMOR-GH10A as a negative control (Figure 28). 2µg of each purified protein was loaded in each well on the agar plate and incubated for 4 hours 65°C and subsequently stained with an iodin mixture (1g KI, 0.5g I, 300ml H_2O). A clearance zone observed around the positive control and 12AMOR1-GH13 showed that the starch had been degraded (Figure 28).
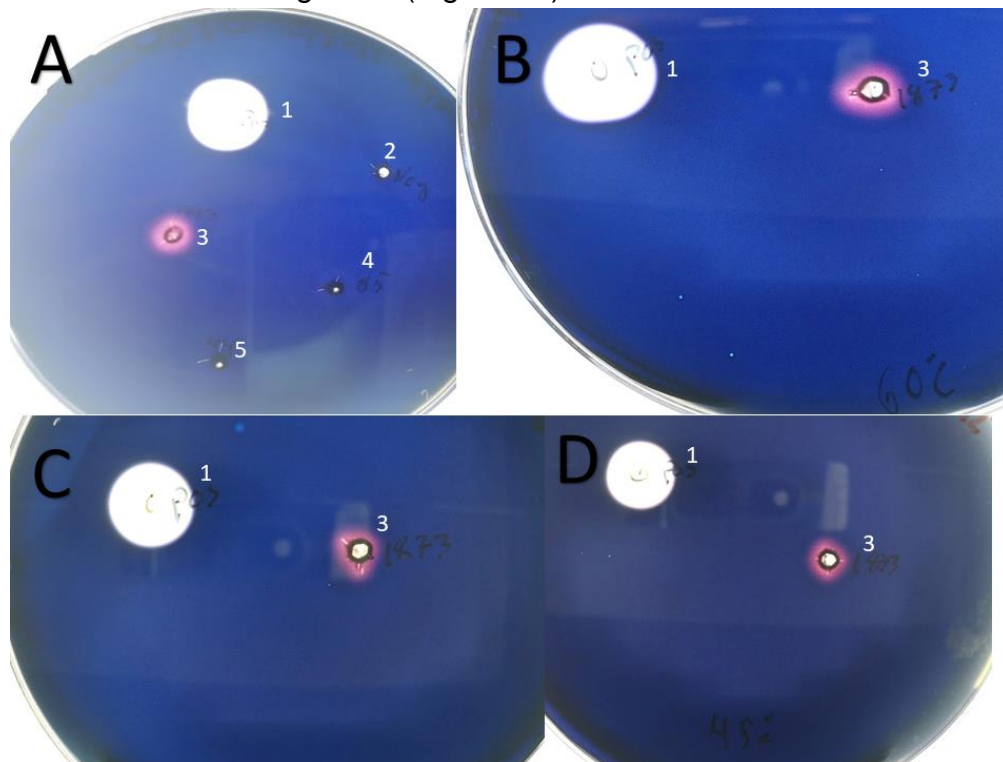
Figure 28: Agar plate with 1% starch and phosphate buffer pH 7.4. 2µg protein loaded in each well and incubated for 4 hours before staining with iodine solution. 1) Bacillus licheniformis α-amylase from Sigma as positive control. 2) AMOR-GH10A xylanase as negative control. 3) 12AMOR1-GH13. 4) AMOR-GH57-4. 5) AMOR-GH57-2. A: incubated at 65°C. B: incubated at 60°C. C: incubated 55°C. D: incubated at 45°C.

Even after storing 12AMOR1-GH13 in the fridge at 4°C the clearance zone increased in size, indicating activity at low temperature. 12AMOR1-GH13 was also tested at 60°C, 55°C and 45°C for 4 hours (Figure 28.B,C and D). The results show that 12AMOR1-GH13 has activity on starch in the temperature range 65°C to 45°C. AMOR-GH57-2 and AMOR-GH57-4 showed no activity on starch at 65°C.

5. Discussion

5.1 Sequenced-based analysis of selected CAZYmes

Sequence's analysis was performed to examine for thermostability, signal peptide and transmembrane helices. This was done to narrow down the candidate selection for expression on the CBM9 containing sequences. Blast results for CBM9 containing sequences (Table 5) showed that *Bacteroidetes bacterium* and *Caldithrix abyssi* were the dominating hits. Both have been described as part of the biodiversity in hydrothermal vents (Fedosov et al., 2006; Gavande et al., 2021). Both are also listed in the CAZy database with multiple GH families (Lombard et al., 2014), expected with the annotation of these CBM9 containing sequences.

Sequence analysis revealed that the GH57 sequences contained the Glycoside hydrolase class without a carbohydrate binding module present. Cazy enzymes does not need a CBM to function, while shows that these helps to increase the activity with binding to substrate and can affect the thermostability. CBMs can be 30 to 200 amino acids in size and up to 3 CBMs on one protein (Shoseyov et al., 2006).

Homology blast of GH57 sequences showed that three of the sequences had similarity to proteins from the Euryarchaeota phyla. AMOR-GH57-2 is related to the superphylum DPANN and AMOR-GH57-3 to the superphylum TACK (Dombrowski et al., 2019; Guy & Ettema, 2011). Euryarchaeota has been reported from hydrothermal vents with CAZymes activity (Li et al., 2015). These results fit with the assumption of thermophilic origin, as many of the hits have been found at hydrothermal vents.

5.2 Novel CBM9 xylanases at AMOR?

CBM9 is associated with xylanases showing endo-1,4-β-xylanase activity (EC 3.2.1.8), and to date this activity is found in the following GH families (3, 5, 6, 8, 9, 10, 11, 16, 18, 26, 30, 43, 51, 62, 98 and 141, http://www.cazy.org/). The primary sequence of CBM9 comprises 170 residues and found primarily in xylanases, with one cellulose-binding function detected so far (Lombard et al., 2014). Here we expect the CBM9 containing sequences to have a putative xylanase activity based on the CBM9 domain. To test this, two sequences were expressed, purified and tested for activity on plates containing xylan. AMOR-CBM9-14 and AMOR-CBM9-15 were both successfully expressed in both BL21 DE3 Gold and pLysS competent cells in the temperature range of 20°C to 37°C. BL21 DE3 Gold were chosen for further expression due to a faster growth rate. Expression was tested twice at 16°C, including an increase in IPTG concentration to 0.2 mM, with no observed expression. For the CBM9 target genes induction at 20°C was the lowest temperature for observed expression. Performing expression at lower temperatures has previously been shown to promote solubility by decreasing the amount of produced proteins (Schein & Noteborn, 1988).

To overcome insoluble protein aggregates of the CBM9 proteins, expression was performed with the dipeptide glycylglycine added in the media. This approach was inspired by a paper published on the effect of glycylglycine on soluble proteins (Ghosh et al., 2004). However, no yield of soluble protein was observed after expression of AMOR-CBM9-14 (Figure 11). A major difference with AMOR-CBM9-14 and AMOR-CBM9-15, compared to the work by Ghosh et al., was insoluble proteins after expression. Hence, enhancing the solubility with the addition of glycylglycine may only be successful if the target protein is partly soluble. Furthermore, AMOR-CBM9-15 was also tested with 3 different lysis buffers with the pH range 7.5 – 8.5 with no effect.

After different attempts on optimizing the expression of both AMOR-CBM9-14 and AMOR-CBM9-15 to obtain soluble protein with no result, attempts to purify the samples were performed. Although having insoluble proteins as visualized on gels, an attempt on purifying with a his-trap column was performed on AMOR-CBM9-14. The result showed a mix of proteins with the most abundant around 50kDa in fraction T12-T16 from the elution peak (Figure 12). Since AMOR-CBM9-14 is expected to be 82 kDa, this indicated that the column showed unspecific binding to *E. coli* proteins and not the target. Since AMOR-CBM9-14 was not expressed as a soluble protein, we assume that the expressed target proteins aggregated as inclusion bodies inside the cell. This is not an uncommon occurrence in heterologous protein expression (Rosano & Ceccarelli, 2014) and one of the largest obstacles in biotechnology. One approach, dating back to 1961, in a study by Anfinsen and Haber (Anfinsen & Haber, 1961), is the addition of the denaturing agent urea in the binding and elution buffer, introducing reversible denaturation of proteins. Using urea as denaturing agent in the cell lysate after expression resulted in the successful purification of a denatured AMOR-CBM9-14, (Figure 13), confirmed by mass spectrometry at the proteomics unit, UiB. Similar results were obtained for AMOR-CBM9-15 (Figure 14). Renaturing of both targets followed the procedure of the control AMOR-GH10A. Since the control originates from the same metagenome as the target CBM9 proteins, these conditions might be better suited.

When proteins are expressed but not soluble in heterologous protein expression, it becomes a bottleneck for further analysis. Similar studies done on a larger scale with 506 CAZymes from procaryotic origin, resulted in 323 soluble proteins (Helbert et al., 2019). In the current study, lowering the expression rate by adjusting the IPTG inducer concentration, in two different strains of competent *E. coli* was performed to overcome this obstacle. Furthermore, codon optimization was performed with GenScript GenSmart tool to limit the possibility for codon bias when expressed in *E. coli*. Another approach on synthesized genes would be codon harmonization. This is a method which aims to replicate the codon usage frequencies with the expression host (Punde et al., 2019). This allows the expression to be similar to the origin organisms. To achieve codon optimization on AMOR-CBM9-14 and AMOR-CBM9-15 a profile could be made based on the metagenome-assembled genomes (MAG) from which the target enzymes belong. Thus, obtaining a profile to match the codon frequency to the expression host and optimizing protein expression and folding.

Other strategies that have proven successful in heterologous protein expression is the use of fusion proteins (Rosano & Ceccarelli, 2014). A study of a thermophilic cyclomaltodextrinase, showed that optimized expression increased the solubility in connection with the NusA fusion protein, and activity increased with the fusion protein cleaved of (Turner et al., 2005). Helbert et al (Helbert et al., 2019) tested 24 of the insoluble proteins with 4 fusion proteins, DsbC, thioredoxin, maltode-binding protein and CpB, without achieving solubility. Fusion proteins are not a guaranty for obtaining solubility during protein expression, however, could be helpful as part of the expression toolbox. Selecting the right fusion partner when testing is also something to consider as they vary in size and can affect the protein differently (Costa et al., 2014).

5.3 Difficulties working with GH57 amylase sequences

The GH57 candidates AMOR-GH57-2 and AMOR-GH57-4 showed the largest soluble fraction after protein expression at 16°C overnight and were chosen to be purified with the His-trap column. Similar to the CBM9 targets, all of the GH57 target proteins were expressed using glycylglycine to enhance the amount of soluble protein. Overall, glycylglycine enhanced the solubility of GH57 target proteins (Figure 20 and 21), however, using 1M glycylglycine showed a negative effect on the growth rate of the host, increasing the time to reach desired optical density for induction.

Assessing if there was an increase in the amount of soluble protein with increasing glycylglycine concentration were difficult based on the SDS-PAGE gels alone. The study which described the increase in soluble protein with glycylglycine, concluded that 1M glycylglycine yielded the highest amount of soluble fraction (Ghosh et al., 2004). However, the major difference from the current study was that the glycylglycine experiments were not performed with terrific broth media, which could have solved or reduced the growth limitation caused by the increase of glycylglycine.

The first purifications of AMOR-GH57-2 and AMOR-GH57-4 were performed on cultures expressed with 200mM glycylglycine, with no visible protein bound to the His-trap column and that were washed out with the flowthrough fraction. To check if the glycylglycine influenced binding to the His-trap column, the purification was repeated with samples expressed without glycylglycine. However, with no difference in result. To test if lower flowrate and imidazole concentration could help, the flowthrough fraction from AMOR-GH57-2 was diluted with binding buffer to a final concentration of 10mM imidazole and re-applied to the his-trap column. This had no effect.

Although both AMOR-GH57-2 and AMOR-GH57-4 were soluble, the His-tag were not able to bind during purification. Two possible explanations to why these proteins did not bind to the column; the first could be a loss of the his-tag after expression, the second that the His-tag was folded inside the protein, and thus unavailable for attachment to the column. Similar to the CBM9 targets, purification with denatured AMOR-GH57-2 and AMOR-GH57-4 using urea should make the His-tag available to bind to the column. Results showed that although some of the protein for both AMOR-GH57-2 (Figure 23) and AMOR-GH57-4 (Figure 24) bound to the column, the protein was still present in the flowthrough with the unbound fraction. These results support

the theory that the His-tag in the native form of the proteins is not available to bind to the nickel within the column.

Heat treatment experiments as a purification step of AMOR-GH57-2 and AMOR-GH57-4 at 60°C, 70°C and 80°C. However, after multiple trials, the results were inconclusive. The reasoning for the heat treatments as a purification step was based on the origin of these sequences. These originated from the deepest chamber (chamber three) of the CGB9 *in situ* incubator (Stokke et al. 2020), where the temperature increases with the depth in the sediment, thus inferring thermostable properties of the GH57 target enzymes. Hence, an increased temperature would remove the *E. coli* proteins since they are mesophilic. When looking at the heat treatment performed of the 12AMOR1-GH13 protein (Figure 25), there is a difference of 14 µg of protein in the soluble fraction and heat-treated soluble fraction, with the target protein present in both fractions. This allows for better purification as less protein from *E. coli* could interfere and bind to the His-trap column.

Overall, the difficulties during purification experiments of AMOR-GH57-2 and AMOR-GH57-4 were most likely due to the His-tag being unavailable for binding to the column. If time would have allowed, synthesizing these genes with a C-terminal his-tag could have been performed, and a better solution then trying to optimize renaturing conditions. Protein denaturation and refolding, if successful, will also decrease the protein concentration. C-terminal expression and purification has been reported successful on other GH57 sequences (Zhang et al., 2019). Other purification methods which do not rely on an affinity tag could also be considered such as the purification and characterization of a thermostable α-amylase from *Bacillus cereus* where the protein was successfully purified with ion exchange chromatography (Annamalai et al., 2011). Ion chromatography uses the protein net charge to separate the proteins (Vella, 1987). If ion exchange chromatography would be considered, heat treatment at 60°C on the samples first to remove part of the *E. coli* protein should be aimed for. This would increase the possibility to purify the protein when less *E. coli* protein is present. Gel filtration could also supplement the purification by filtering out proteins larger and smaller than the targets.

5.4 Isolating active proteins from CBM9 and GH57 candidates

Plate assays for protein activity against xylan showed no activity on AMOR-CBM9-14 and AMOR-CBM9-15 (Figure 17), with same results for plate assays with starch for AMOR-GH57-2 and AMOR-GH57-4 (Figure 28). The plate tests were not conclusive enough to conclude that the proteins had no activity against the substrate tested. The major problem is the treatment done with the proteins after expression. When denaturing the proteins, this require that the proteins need to be renatured to their native state. Here renaturation was tested with dialysis and buffer change. Activity was tested with these constructs on 65 °C for 4 hours. Additional tests at 55 °C would have been a good supplement to rule out if they were not thermostable at this temperature. However, due to low protein yields from protein purification, there was not enough sample for the experiment within the time of the project.

Two buffers for renaturing and buffer exchange were tested for the CBM9 targets, the second with the storage buffer used on AMOR-GH10A. Renaturation of AMOR-GH57-2 and AMOR-GH57-4 was also tested with dialysis. Difficulties with refolding proteins are identifying the best ionic concentration and pH that suit the target protein. Testing the samples for correct folding is also a bottleneck. Here the test was done with plate activity assay. If the result were positive, this approach would have been good enough. However, with negative results from the plate assays it is difficult to assess if correct folding has been achieved.

Other approaches with protein dialysis have shown that stepwise dialysis may increase the fraction of active protein. With stepwise dialysis the protein sample is equilibrated in multiple stages in buffer that gradually decreases in concentration, allowing the protein to first partly refold in dialysis with a lower urea concentration. Then dialysis with a buffer without urea to induce refolding back to native structure (Yamaguchi & Miyazaki, 2014). Protein loss is problem with protein denaturation and renaturation. AMOR-CBM9-14 and AMOR-CBM9-15 went from 9.8ml with a protein concentration of 0.220 mg/ml and 0.250 mg/ml respectively to 200µl 0.095 mg/ml and 500µl 0.095 mg/ml. AMOR-GH57-2 and AMOR-GH57-4 went from 9.8ml with a concentration 0.0292 mg/ml and 0.0260 mg/ml respectively to 200µl 0.102 mg/ml and 0.122 mg/ml. This shows that protein loss was a problem.

5.5 Oligomeric state of the thermostable *Geobacillus* sp. 12AMOR1 neopullanase

The difficulties with expressing AMOR-CBM9 and purifying the AMOR-GH57 target proteins shows that protein expression and purification from environmental metagenome samples is not straightforward that could be implemented the same way for different proteins. However, AMOR-GH10A that originates from the same metagenome as the CBM9 targets, and the 12AMOR1-GH13 which originates from a bacterial isolate, were successfully purified and active after first attempt.

From the theoretical molecular weight calculation of 12AMOR1-GH13 it is estimated to be 68kDa, which fits with the SDS-PAGE where the band is close to the 65kDa marker (Figure 25.B). Using gel filtration spiked with standards, the molecular weight was calculated to be 105.5kDa. During gel filtration the protein is purified in its native state, hence, when the size is near double of the denatured state this suggest that the protein is a dimer. Knowing the oligomeric state of the protein can often be used to determine the activity and binding affinity of the protein (Hashimoto & Panchenko, 2010). Gel filtration with size exclusion is not the most accurate method to calculate molecular size, however, it gives a good estimate. For more accurate measurements analytical ultracentrifugation may be performed on the samples. Analytical ultracentrifugation can be performed as sedimentation velocity or sedimentation equilibrium, both can be used to calculate molecular weight of proteins (Cole et al., 2008)

In the paper from Wissuwa et al. the growth of *Geobacillus* sp. was in the temperature range of 40-70°C and the melting temperature for 12AMOR1-GH13 was estimated to be 76,4°C. The plate assay for starch degradation showed that 12AROR1-GH13 was

active in the range that was tested between 40-65°C. Higher temperatures was not tested as agar plates start to melt above 65°C, although other solutions such as gelrite in the plates could substitute agar at high temperatures. If time would have allowed in the current project, the pH range of 12AROR1-GH13 could have been analyzed since knowing the optimal activity parameters are important when considering proteins for their enzyme activity to be utilized in industrial processes.

6. Conclusion

Two selected CBM9 sequences was selected for expression, purification and activity analysis. Expression was achieved in the temperature range 20°C – 37°C while protein aggregated as inclusion bodies. However, purification was achieved under denaturing conditions. Refolding of the target proteins against xylan resulted in no activity observed at 65°C.

Two annotated GH57 α-amylase sequences were expressed as soluble protein both with and without glycylglycine in the media. Purification was achieved under denaturing conditions. Although refolding resulted in soluble proteins, no activity against starch was observed at 65°C.

The 12AMOR1-GH13 neopullulanase was purified in the native state with a monomeric molecular weight estimated with SDS-PAGE at 65kDa, and oligomeric state estimated to be dimeric.

7. Future perspectives

The novelty in finding new enzymes for biotechnology is important amongst others for efficient processes. Identifying their ecological role could also strengthen our ability to fully utilize the enormous potential of deep-sea hydrothermal enzymes. Mapping enzyme targets to metagenome-assembled genomes from the same samples enables a further study of the ecological roles of secreted CAZy enzymes. To further confirm the activity of the putative xylanases, expression with a variety of fusion partners should be considered to obtain the proteins in a native state. A larger study on substrate and temperature screening of both the CBM9 and the GH57 targets should also be conducted since we know little about substrate specificity. AMOR-GH57-2 and AMOR-GH57-4 should be tested with new affinity tag placement. Testing Hit-tag placement on the C-terminal or considering ion exchange chromatography to purify the proteins and avoiding denaturation. The 12AMOR1-GH13 neopullulanase activity should be further tested to find optimal pH range.

8. References

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019, Apr). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol, 37*(4), 420-423. https://doi.org/10.1038/s41587-019-0036-z

Anfinsen, C. B., & Haber, E. (1961, 1961/05/01/). Studies on the Reduction and Re-formation of Protein Disulfide Bonds. *Journal of Biological Chemistry, 236*(5), 1361-1363. https://doi.org/https://doi.org/10.1016/S0021-9258(18)64177-8

Annamalai, N., Thavasi, R., Vijayalakshmi, S., & Balasubramanian, T. (2011). Extraction, Purification and Characterization of Thermostable, Alkaline Tolerant α-Amylase from Bacillus cereus. *Indian journal of microbiology, 51*(4), 424-429. https://doi.org/10.1007/s12088-011-0160-z

Arntzen, M., Pedersen, B., Klau, L. J., Stokke, R., Oftebro, M., Antonsen, S. G., Fredriksen, L., Sletta, H., Aarstad, O. A., Aachmann, F. L., Horn, S. J., & Eijsink, V. G. H. (2021, Feb 26). Alginate Degradation: Insights Obtained through Characterization of a Thermophilic Exolytic Alginate Lyase. *Appl Environ Microbiol, 87*(6). https://doi.org/10.1128/aem.02399-20

Beaulieu, S. E., Szafra, & Ski, K. M. (2020). *InterRidge Global Database of Active Submarine Hydrothermal Vent Fields Version 3.4* PANGAEA. https://doi.org/10.1594/PANGAEA.917894

Chettri, D., Verma, A. K., & Verma, A. K. (2020). Innovations in CAZyme gene diversity and its modification for biorefinery applications. *Biotechnology reports (Amsterdam, Netherlands), 28*, e00525-e00525. https://doi.org/10.1016/j.btre.2020.e00525

Chiu, C. Y., & Miller, S. A. (2019, 2019/06/01). Clinical metagenomics. *Nature Reviews Genetics, 20*(6), 341-355. https://doi.org/10.1038/s41576-019-0113-7

Cole, J. L., Lary, J. W., P Moody, T., & Laue, T. M. (2008). Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods in cell biology, 84*, 143-179. https://doi.org/10.1016/S0091-679X(07)84006-4

Consortium, T. U. (2018). UniProt: a worldwide hub of protein knowledge. *Nucleic acids research, 47*(D1), D506-D515. https://doi.org/10.1093/nar/gky1049

Coordinators, N. R. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic acids research, 44*(D1), D7-D19. https://doi.org/10.1093/nar/gkv1290

Corrêa, C. L., Midorikawa, G. E. O., Filho, E. X. F., Noronha, E. F., Alves, G. S. C., Togawa, R. C., Silva-Junior, O. B., Costa, M. M. d. C., Grynberg, P., & Miller, R. N. G. (2020, 2020-October-06). Transcriptome Profiling-Based Analysis of Carbohydrate-Active Enzymes in Aspergillus terreus Involved in Plant Biomass Degradation [Original Research]. *Frontiers in Bioengineering and Biotechnology, 8*(1179). https://doi.org/10.3389/fbioe.2020.564527

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014, 2014-February-19). Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system [Review]. *Frontiers in Microbiology, 5*(63). https://doi.org/10.3389/fmicb.2014.00063

Dalmaso, G. Z. L., Ferreira, D., & Vermelho, A. B. (2015). Marine extremophiles: a source of hydrolases for biotechnological applications. *Marine drugs, 13*(4), 1925-1965. https://doi.org/10.3390/md13041925

DeCastro, M.-E., Rodríguez-Belmonte, E., & González-Siso, M.-I. (2016). Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes. *Frontiers in Microbiology, 7*, 1521-1521. https://doi.org/10.3389/fmicb.2016.01521

Demain, A. L., & Adrio, J. L. (2007, 2007/12/04). Contributions of Microorganisms to Industrial Biology. *Molecular Biotechnology, 38*(1), 41. https://doi.org/10.1007/s12033-007-0035-z

Department of Bio and Health Informatics. (2017, January 5). *Predictor of transmembrane helices in proteins* http://www.cbs.dtu.dk/services/TMHMM/

Dick, G. J. (2019, 2019/05/01). The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nature Reviews Microbiology, 17*(5), 271-283. https://doi.org/10.1038/s41579-019-0160-2

Dombrowski, N., Lee, J.-H., Williams, T. A., Offre, P., & Spang, A. (2019). Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiology Letters, 366*(2). https://doi.org/10.1093/femsle/fnz008

Fedosov, D. V., Podkopaeva, D. A., Miroshnichenko, M. L., Bonch-Osmolovskaya, E. A., Lebedinsky, A. V., & Grabovich, M. Y. (2006, 2006/03/01). Investigation of the catabolism of acetate and peptides in the new anaerobic thermophilic bacterium Caldithrix abyssi. *Microbiology, 75*(2), 119-124. https://doi.org/10.1134/S0026261706020020

Felczykowska, A., Krajewska, A., Zielinska, S., & Łoś, J. (2015, 02/13). Sampling, metadata and DNA extraction - Important steps in metagenomic studies. *Acta biochimica Polonica, 62*. https://doi.org/10.18388/abp.2014_916

Fia, G., Giovani, G., & Rosi, I. (2005). Study of β-glucosidase production by wine-related yeasts during alcoholic fermentation. A new rapid fluorimetric method to determine enzymatic activity. *Journal of Applied Microbiology, 99*(3), 509-517. https://doi.org/https://doi.org/10.1111/j.1365-2672.2005.02657.x

Fredriksen, L., Stokke, R., Jensen, M. S., Westereng, B., Jameson, J. K., Steen, I. H., & Eijsink, V. G. H. (2019, Mar 15). Discovery of a Thermostable GH10 Xylanase with Broad Substrate Specificity from the Arctic Mid-Ocean Ridge Vent System. *Appl Environ Microbiol, 85*(6). https://doi.org/10.1128/aem.02970-18

Fu, H., Liang, Y., Zhong, X., Pan, Z., Huang, L., Zhang, H., Xu, Y., Zhou, W., & Liu, Z. (2020, 2020/10/19). Codon optimization with deep learning to enhance protein expression. *Scientific Reports, 10*(1), 17617. https://doi.org/10.1038/s41598-020-74091-z

Gagnon, K. (2010). Chapter 9 - Measuring Electroneutral Chloride-Dependent Ion Fluxes in Mammalian Cells and in Heterologous Expression Systems. In F. J. Alvarez-Leefmans & E. Delpire (Eds.), *Physiology and Pathology of Chloride Transporters and Channels in the Nervous System* (pp. 149-157). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-374373-2.00009-1

Gavande, P. V., Basak, A., Sen, S., Lepcha, K., Murmu, N., Rai, V., Mazumdar, D., Saha, S. P., Das, V., & Ghosh, S. (2021, 2021/02/04). Functional characterization of thermotolerant microbial consortium for lignocellulolytic enzymes with central role of Firmicutes in rice straw depolymerization. *Scientific Reports, 11*(1), 3032. https://doi.org/10.1038/s41598-021-82163-x

Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., & Whalen, K. L. (2015, Aug). Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta, 1854*(8), 1019-1037. https://doi.org/10.1016/j.bbapap.2015.04.015

Ghosh, S., Rasheedi, S., Rahim, S. S., Banerjee, S., Choudhary, R. K., Chakhaiyar, P., Ehtesham, N. Z., Mukhopadhyay, S., & Hasnain, S. E. (2004, Sep). Method for enhancing solubility of the expressed recombinant proteins in Escherichia coli. *Biotechniques, 37*(3), 418, 420, 422-413. https://doi.org/10.2144/04373st07

Guy, L., & Ettema, T. J. (2011, Dec). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol, 19*(12), 580-587. https://doi.org/10.1016/j.tim.2011.09.002

Hashimoto, K., & Panchenko, A. R. (2010). Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proceedings of the National Academy of Sciences, 107*(47), 20352-20357. https://doi.org/10.1073/pnas.1012999107

Helbert, W., Poulet, L., Drouillard, S., Mathieu, S., Loiodice, M., Couturier, M., Lombard, V., Terrapon, N., Turchetto, J., Vincentelli, R., & Henrissat, B. (2019). Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. *Proceedings of the National Academy of Sciences, 116*(13), 6063-6068. https://doi.org/10.1073/pnas.1815791116

Janeček, Š., & Blesák, K. (2011, 2011/07/24). Sequence-Structural Features and Evolutionary Relationships of Family GH57 α-Amylases and Their Putative α-Amylase-Like Homologues. *The Protein Journal, 30*(6), 429. https://doi.org/10.1007/s10930-011-9348-7

Janeček, Š., Svensson, B., & MacGregor, E. A. (2014, 2014/04/01). α-Amylase: an enzyme specificity found in various families of glycoside hydrolases. *Cellular and Molecular Life Sciences, 71*(7), 1149-1170. https://doi.org/10.1007/s00018-013-1388-z

Jeon, E.-J., Jung, J.-H., Seo, D.-H., Jung, D.-H., Holden, J. F., & Park, C.-S. (2014, 2014/06/10/). Bioinformatic and biochemical analysis of a novel maltose-forming α-amylase of the GH57

family in the hyperthermophilic archaeon Thermococcus sp. CL1. *Enzyme and Microbial Technology, 60*, 9-15. https://doi.org/https://doi.org/10.1016/j.enzmictec.2014.03.009

Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010, 2010/08/18). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics, 11*(1), 431. https://doi.org/10.1186/1471-2105-11-431

Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M., & McCue, L. A. (2020, 2020/06/22). ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinformatics, 21*(1), 257. https://doi.org/10.1186/s12859-020-03585-4

Kirk, O., Borchert, T. V., & Fuglsang, C. C. (2002, 2002/08/01/). Industrial enzyme applications. *Current Opinion in Biotechnology, 13*(4), 345-351. https://doi.org/https://doi.org/10.1016/S0958-1669(02)00328-2

Kodzius, R., & Gojobori, T. (2015, Dec). Marine metagenomics as a source for bioprospecting. *Mar Genomics, 24 Pt 1*, 21-30. https://doi.org/10.1016/j.margen.2015.07.001

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001, Jan 19). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol, 305*(3), 567-580. https://doi.org/10.1006/jmbi.2000.4315

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018, Jun 1). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol, 35*(6), 1547-1549. https://doi.org/10.1093/molbev/msy096

Li, M., Baker, B. J., Anantharaman, K., Jain, S., Breier, J. A., & Dick, G. J. (2015, 2015/11/17). Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nature Communications, 6*(1), 8933. https://doi.org/10.1038/ncomms9933

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014, Jan). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research, 42*(Database issue), D490-495. https://doi.org/10.1093/nar/gkt1178

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019, 2019/07//). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research, 47*(W1), W636-W641. https://doi.org/10.1093/nar/gkz268

Madigan, M. T., Madigan, M. T., Bender, K. S., Buckley, D. H., Sattley, W. M., Stahl, D. A., & Brock, T. D. (2019). *Brock biology of microorganisms* (Fifteenth edition.; Global edition. ed.). Pearson.

Mauro, V. P., & Chappell, S. A. (2014). A critical analysis of codon optimization in human therapeutics. *Trends in molecular medicine, 20*(11), 604-613. https://doi.org/10.1016/j.molmed.2014.09.003

McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic acids research, 32*(Web Server issue), W20-W25. https://doi.org/10.1093/nar/gkh435

Mischnick, P., & Momcilovic, D. (2010). Chemical Structure Analysis of Starch and Cellulose Derivatives. In D. Horton (Ed.), *Advances in Carbohydrate Chemistry and Biochemistry* (Vol. 64, pp. 117-210). Academic Press. https://doi.org/https://doi.org/10.1016/S0065-2318(10)64004-8

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., & Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research, 44*(D1), D733-D745. https://doi.org/10.1093/nar/gkv1189

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018, 2018/07//). HMMER web server: 2018 update. *Nucleic acids research, 46*(W1), W200-W204. https://doi.org/10.1093/nar/gky448

Punde, N., Kooken, J., Leary, D., Legler, P. M., & Angov, E. (2019, 2019/10/19). Codon harmonization reduces amino acid misincorporation in bacterially expressed P. falciparum proteins and improves their immunogenicity. *AMB Express, 9*(1), 167. https://doi.org/10.1186/s13568-019-0890-6

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research, 42*(W1), W320-W324. https://doi.org/10.1093/nar/gku316

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology, 5*, 172-172. https://doi.org/10.3389/fmicb.2014.00172

Rødsrud, G., Lersch, M., & Sjöde, A. (2012, 2012/11/01/). History and future of world's most advanced biorefinery in operation. *Biomass and Bioenergy, 46*, 46-59. https://doi.org/https://doi.org/10.1016/j.biombioe.2012.03.028

Sarmiento, F., Peralta, R., & Blamey, J. M. (2015, 2015-October-20). Cold and Hot Extremozymes: Industrial Relevance and Current Trends [Review]. *Frontiers in Bioengineering and Biotechnology, 3*(148). https://doi.org/10.3389/fbioe.2015.00148

Schein, C. H., & Noteborn, M. H. M. (1988, 1988/03/01). Formation of Soluble Recombinant Proteins in Escherichia Coli is Favored by Lower Growth Temperature. *Bio/Technology, 6*(3), 291-294. https://doi.org/10.1038/nbt0388-291

Shoseyov, O., Shani, Z., & Levy, I. (2006). Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. *Microbiology and Molecular Biology Reviews, 70*(2), 283-295. https://doi.org/10.1128/mmbr.00028-05

Sievert, S. M., Kuever, J., & Muyzer, G. (2000). Identification of 16S ribosomal DNA-defined bacterial populations at a shallow submarine hydrothermal vent near Milos Island (Greece). *Applied and environmental microbiology, 66*(7), 3102-3109. https://doi.org/10.1128/aem.66.7.3102-3109.2000

Sonnhammer, E. L., von Heijne, G., & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol, 6*, 175-182.

Stahl, D. A., Buckley, D. H., Bender, K. S., Brock, T. D., Martinko, J. M., & Madigan, M. T. (2015). *Brock biology of microorganisms* (14th ed., Global ed. ed.). Pearson.

Stepnov, A. A., Fredriksen, L., Steen, I. H., Stokke, R., & Eijsink, V. G. H. (2019). Identification and characterization of a hyperthermophilic GH9 cellulase from the Arctic Mid-Ocean Ridge vent field. *PloS one, 14*(9), e0222216-e0222216. https://doi.org/10.1371/journal.pone.0222216

Stokke, R., Reeves, E. P., Dahle, H., Fedøy, A.-E., Viflot, T., Lie Onstad, S., Vulcano, F., Pedersen, R. B., Eijsink, V. G. H., & Steen, I. H. (2020, 2020-February-21). Tailoring Hydrothermal Vent Biodiversity Toward Improved Biodiscovery Using a Novel in situ Enrichment Strategy [Original Research]. *Frontiers in Microbiology, 11*(249). https://doi.org/10.3389/fmicb.2020.00249

Stothard, P. (2017, November 6). *Sequence Manipulation Suite: Protein Isoelectric Point*. http://www.bioinformatics.org/sms2/protein_iep.html

Subramaniyan, S., & Prema, P. (2002, 2002/01/01). Biotechnology of Microbial Xylanases: Enzymology, Molecular Biology, and Application. *Critical Reviews in Biotechnology, 22*(1), 33-64. https://doi.org/10.1080/07388550290789450

Tamames, J., Cobo-Simón, M., & Puente-Sánchez, F. (2019, 2019/12/10). Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes. *BMC Genomics, 20*(1), 960. https://doi.org/10.1186/s12864-019-6289-6

Turner, P., Holst, O., & Karlsson, E. N. (2005, 2005/01/01/). Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in Escherichia coli by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expression and Purification, 39*(1), 54-60. https://doi.org/https://doi.org/10.1016/j.pep.2004.09.012

Ui, N. (1979, 1979/08/01/). Rapid estimation of the molecular weights of protein polypeptide chains using high-pressure liquid chromatography in 6 m guanidine hydrochloride. *Analytical Biochemistry, 97*(1), 65-71. https://doi.org/https://doi.org/10.1016/0003-2697(79)90328-2

Vella, F. (1987). Modern concepts in biochemistry (fifth edition) R C Bohinski. pp 739. Allyn & Bacon, Boston. 1987. £19.95 ISBN 0-205-08852-X and 0-205-10555-6 (Int student edn). *Biochemical Education, 15*(3), 160-160. https://doi.org/https://doi.org/10.1016/0307-4412(87)90061-6

Vuoristo, K. S., Fredriksen, L., Oftebro, M., Arntzen, M. Ø., Aarstad, O. A., Stokke, R., Steen, I. H., Hansen, L. D., Schüller, R. B., Aachmann, F. L., Horn, S. J., & Eijsink, V. G. H. (2019, 2019/03/13). Production, Characterization, and Application of an Alginate Lyase, AMOR_PL7A, from Hot Vents in the Arctic Mid-Ocean Ridge. *Journal of Agricultural and Food Chemistry, 67*(10), 2936-2945. https://doi.org/10.1021/acs.jafc.8b07190

Webber, A. P., Roberts, S., Murton, B. J., & Hodgkinson, M. R. S. (2015). Geology, sulfide geochemistry and supercritical venting at the Beebe Hydrothermal Vent Field, Cayman Trough. *Geochemistry, Geophysics, Geosystems, 16*(8), 2661-2678. https://doi.org/https://doi.org/10.1002/2015GC005879

Wissuwa, J., Stokke, R., Fedøy, A. E., Lian, K., Smalås, A. O., & Steen, I. H. (2016). Isolation and complete genome sequence of the thermophilic Geobacillus sp. 12AMOR1 from an Arctic deep-sea hydrothermal vent site. *Stand Genomic Sci, 11*, 16. https://doi.org/10.1186/s40793-016-0137-y

Wong, W.-C., Maurer-Stroh, S., Schneider, G., & Eisenhaber, F. (2012). Transmembrane helix: simple or complex. *Nucleic acids research, 40*(Web Server issue), W370-W375. https://doi.org/10.1093/nar/gks379

Yamaguchi, H., & Miyazaki, M. (2014). Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules, 4*(1), 235-251. https://doi.org/10.3390/biom4010235

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., & Yin, Y. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic acids research, 46*(W1), W95-W101. https://doi.org/10.1093/nar/gky418

Zhang, X., Leemhuis, H., & van der Maarel, M. J. E. C. (2019). Characterization of the GH13 and GH57 glycogen branching enzymes from Petrotoga mobilis SJ95 and potential role in glycogen biosynthesis. *PloS one, 14*(7), e0219844-e0219844. https://doi.org/10.1371/journal.pone.0219844

Zhao, L., Poschmann, G., Waldera-Lupa, D., Rafiee, N., Kollmann, M., & Stühler, K. (2019, 2019/12/19). OutCyte: a novel tool for predicting unconventional protein secretion. *Scientific Reports, 9*(1), 19448. https://doi.org/10.1038/s41598-019-55351-z

9. Appendices
9.1 Appendix Buffers

B1 - Lysis buffer:
50mM Hepes
300mM NaCl
10% glycerol (99%)
MiliQ water
pH 7.5

B2 - Lysis buffer:
50mM Tris
500mM NaCl
5mM Imidazole
10% Glycerol (99%)
MiliQ water
pH 8.5

B3 - Lysis buffer
50mM Tris
500mM NaCl
5mM Imidazole
MiliQ water
pH 8

B4 - Binding buffer
20mM Hepes
500mM NaCl
MiliQ water 1L
pH 7.5

B5 - Binding
buffer (10mM imidazole)
20mM Hepes
500mM NaCl
10mM Imidazole
MiliQ water 1L
pH 7.5

B6 - B Binding
buffer (20mM imidazole)
20mM Hepes
500mM NaCl
25mM Imidazole
MiliQ water 1L
pH 7.5

B7 - Binding buffer
(Urea)
20mM NaPO$_4$
20mM Imidazole
500mM NaCl
8M Urea
pH 7.5

B8 - Binding
buffer (AMOR-GH10A)
5mM Imidazole
50mM Tris HCl
500mM NaCl
pH 8

B9 - Binding
buffer (12AMOR-GH13)
50mM Sodium Phosphate
300mM Sodium Chloride
10mM imidazole
pH 8

B10 - Elution buffer
20mM Hepes
500mM NaCl
500mM Imidazole
pH 7.5

B11 - Elution buffer
(Urea)
20mM NaPO$_4$
20mM Imidazole
500mM NaCl
8M Urea
pH 7.5

B12 - Elution buffer
(AMOR-GH10A)
500mM Imidazole
50mM Tris HCl
500mM NaCl
pH 8

B13 - Elution buffer
(12AMOR1-GH13)
50mM Sodium Phosphate
300mM Sodium Chloride
500mM Imidazole
MiliQ water

B14 - Running buffer:
0.02M Hepes
300mM NaCl
MiliQ water
pH 7.5

B15 - Storage buffer
0.02M NaPO$_4$
0.5M NaCl
pH 7.5

B16 - Storage buffer
50mM Sodium Acetate
500mM NaCl
pH 5.6

B17 - Storage buffer
20mM Potassium
Phosphate
100mM Sodium Chloride
pH 7.5

B18 - Phosphate buffer
(DSC)
1M KH$_2$PO$_4$
1M K$_2$HPO4
pH 7.4

9.2 Appendix Media

<u>M1 - Lysogeny Broth (LB)</u>
10g NaCl
10g Tryptone
5g Yeast extract
MiliQ H_2O to final volume 1L

<u>M2 - Terrific Broth (TB)</u>
24g Yeast extract
20g Tryptone
4ml Glycerol (99%)
Phosphate buffer TB (100ml)
0,017M KH_2PO_4/
0,072M K_2HPO_4
MiliQ H_2O to final volume 1L

<u>M3 – Agar plates</u>
10g NaCl
10g Tryptone
5g Yeast extract
20g Agar

<u>M4 – Starch plate</u>
3g Agar
2g Starch
2M phosphate buffer
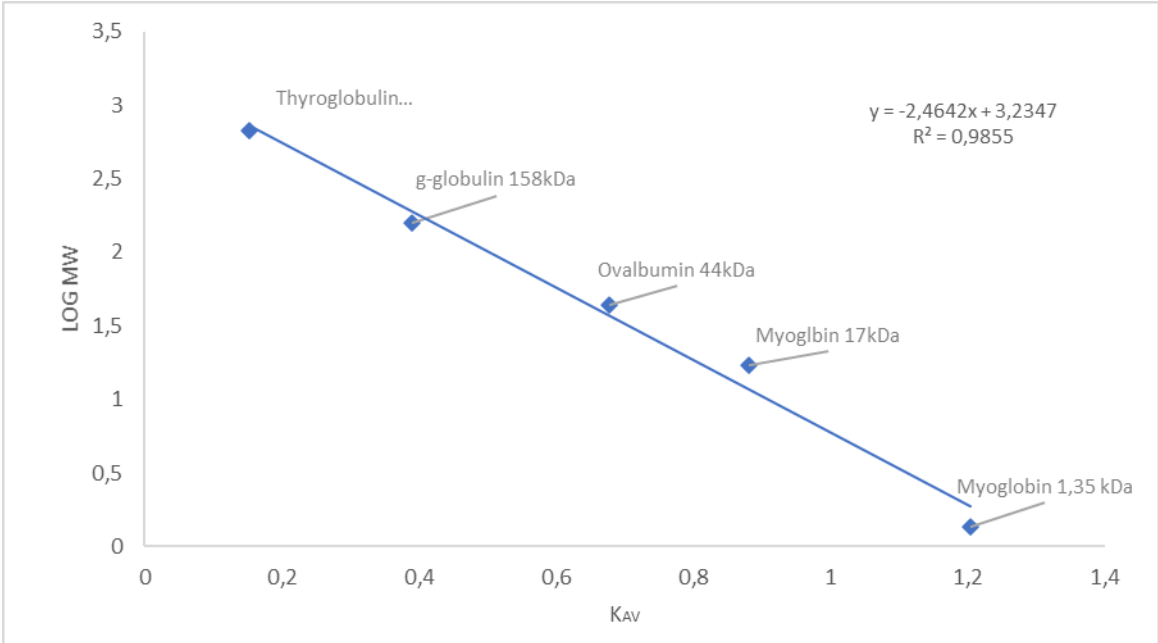200ml H_2O

<u>M5 – Xylan plates</u>
3g Agar
2g Starch
2M phosphate buffer
200ml H_2O

Results *from proteomic analysis*

Checked	Protein FDR Conf. Comb.	Protein FDR Conf.	Master	Master Protein	Accession	Description	Exp. q-value Comb.	Exp. q-value	Coverage	# Peptides	# PSMs	# Unique	# AAs	MW [kDa]	calc. pI	Score Seq.	Score	# Peptides	# Pept.	Found in Sa.	# Protein Gr.
FALSE	**High**	**High**	**Master**	**Master Protein**	**17 AMOR-CBM9-14**	Trypsin - Sus scrofa (Pig).	**0**	**0**	**77**	**64**	**259**	**64**	**707**	**82,2**	**8,95**	**13643,09**	**547,52**	**42**	**64**	**High**	**1**
FALSE	High	High	Master	Master Protein	P04264	SWISS-PROT:P04264 Tax_Id=9606 Gene_Symbol=KRT1 Keratin, type II cytoskeletal 1	0	0	52	10	72	10	231	24,4	7,18	3971,64	176,01	5	10	High	1
FALSE	High	High	Master	Master Protein	P35527	SWISS-PROT:P35527 Tax_Id=9606 Gene_Symbol=KRT9 Keratin, type I cytoskeletal 9	0	0	47	27	51	25	644	66	8,12	2483,78	106,26	16	25	High	1
FALSE	High	High	Master	Master Protein	P35908	SWISS-PROT:P35908 Tax_Id=9606 Gene_Symbol=KRT2 Keratin, type II cytoskeletal 2 epiderma	0	0	45	15	21	15	623	62,1	5,3	466,11	83,12	4	14	High	1
FALSE	High	High	Master	Master Protein	P13645	SWISS-PROT:P13645 Tax_Id=9606 Gene_Symbol=KRT10 Keratin, type I cytoskeletal 10	0	0	43	23	41	17	645	65,8	8	2097,57	81,3	14	21	High	1
FALSE	High	High	Master	Master Protein	P13647	SWISS-PROT:P13647 Tax_Id=9606 Gene_Symbol=KRT5 Keratin, type II cytoskeletal 5	0	0	41	21	35	18	593	59,5	5,21	1416,23	71,05	10	20	High	1
FALSE	High	High	Master	Master Protein	P02538	SWISS-PROT:P02538 Tax_Id=9606 Gene_Symbol=KRT6A Keratin, type II cytoskeletal 6A	0	0	21	13	20	5	590	62,3	7,74	964,59	32,53	7	11	High	1
FALSE	High	High	Master	Master Protein		[gene=ompF] [locus_tag=ECD_00933] protein=outer membrane porin 1a(Ia;b;F)] [protein_id=...	0	0	16	10	18	1	564	60	8	874,49	30,35	6	10	High	1
FALSE	High	High	Master	Master Protein	P02533	SWISS-PROT:P02533 Tax_Id=9606 Gene_Symbol=KRT14 Keratin, type I cytoskeletal 14	0	0	26	9	13	9	362	39,3	4,96	533,92	25,67	4	9	High	1
FALSE	High	High	Master	Master Protein		[gene=fusA] [locus_tag=ECD_03191] protein=protein chain elongation factor EF-G, GTP-bindi...	0	0	10	5	8	5	704	77,5	5,38	325,13	13,07	3	5	High	1
FALSE	High	High	Master	Master Protein		[gene=lpp] [locus_tag=ECD_01646] protein=murein lipoprotein] [protein_id=ACT43504.1] [lo...	0	0	49	3	7	3	78	8,3	9,25	384,8	12,58	3	3	High	1
FALSE	High	High	Master	Master Protein	P02768-1	SWISS-PROT:P02768-1 Tax_Id=9606 Gene_Symbol=ALB Isoform 1 of Serum albumin precursor	0	0	13	6	7	5	609	69,3	6,28	236,84	12,14	2	5	High	1
FALSE	Medium	High	Master	Master Protein	P02769	SWISS-PROT:P02769 Tax_Id=9606 Gene_Symbol=... Bovine serum albumin precursor	0,033	0	10	5	6	2	472	51,6	5,16	95,64	11,48	1	5	High	1
FALSE	High	High	Master	Master Protein	P60712	SWISS-PROT:P60712 (Bos taurus) Actin, cytoplasmic 1	0	0	26	1	2	1	81	8,7	4,56		8,99		1	High	1
FALSE	Medium	High	Master	Master Protein	Q7RTS7	SWISS-PROT:Q7RTS7 Tax_Id=9606 Gene_Symbol=KRT74 Keratin, type II cytoskeletal 74		0,02	6	2	2	1	529	57,8	7,71	128,87	6,04	1	2	High	1
FALSE	High	High	Master	Master Protein	Q6NXH9	TREMBL:Q6NXH9 Tax_Id=10090 Gene_Symbol=KRT73 Keratin 73	0,018	0	9	2	2	1	233	26	10,27	133,75	3,85	1	2	High	1
FALSE	High	High	Master	Master Protein		[gene=rpsC] [locus_tag=ECD_03165] protein=30S ribosomal subunit protein S3] [protein_id...	0	0,017	4	1	2	1	233	26	6,52	149,36	3,48	1	1	High	1
FALSE	Medium	High	Master	Master Protein	P34955	SWISS-PROT:P34955 (Bos taurus) Alpha-1-antiproteinase precursor	0,041	0	4	4	4	4	416	46,1	6,37	101,47	7,97	1	1	High	1
FALSE	High	High	Master	Master Protein		[gene=glgB] [locus_tag=ECD_03284] protein=1,4-alpha-glucan branching enzyme] [protein_id...	0	0	7	4	4	4	728	84,3	6,37	191,94	7,59	1	1	High	1
FALSE	High	High	Master	Master Protein	P07477	SWISS-PROT:P07477 Tax_Id=9606 Gene_Symbol=PRSS1 Trypsin-1 precursor	0	0	8	1	3	1	247	26,5	6,51	279,92	6,91	2	5	High	1
FALSE	High	High	Master	Master Protein	P02769	SWISS-PROT:P02769 (Bos taurus) Bovine serum albumin precursor	0	0	5	2	4	1	607	69,2	6,18		3,1		1	High	1
FALSE	n/a	Medium	Master	Master Protein		[gene=rcdA] [locus_tag=ECD_00260] protein=reactive chlorine stress species (RCS) resistance	0	0,015	7	1	1	1	332	36,8	5,99		3,03		1	High	1
FALSE	Medium	Medium	Master	Master Protein		[gene=lysC] [locus_tag=ECD_03896] protein=lysine-sensitive aspartokinase 3] [protein_id=A...	0	0,015	3	1	2	1	449	48,5	5,11	129,22	3,01	1	1	High	1
FALSE	Medium	Medium	Master	Master Protein		[gene=argG] [locus_tag=ECD_03308] protein=argininosuccinate synthetase] [protein_id=ACT...		0,015	7	1	1	1	447	49,9	5,45		3		1	High	1
FALSE	Medium	Medium	Master	Master Protein		[gene=tufB] [locus_tag=ECD_03856] protein=translation elongation factor EF-Tu 2] [protein_...	0,014	0,014	4	1	1	1	394	43,3	5,45		2,94		1	High	1
FALSE	n/a	Medium	Master	Master Protein		[gene=rpsB] [locus_tag=ECD_00167] protein=30S ribosomal subunit protein S2] [protein_id=...	0,014	0,014	6	1	1	1	241	26,7	7,14		2,87		1	High	1
FALSE	Medium	Medium	Master	Master Protein		[gene=yjgE] [locus_tag=ECD_08591] protein=DUF2233 family protein] [protein_id=ACT4548A...	0,04	0,04	8	1	1	1	254	28,1	9,2		2,84		1	High	1
FALSE	Medium	Medium	Master	Master Protein		[gene=katE] [locus_tag=ECD_01701] protein=catalase HPII, heme d-containing] [protein_id=A...	0,038	0,013	1	2	2	1	753	84,1	5,82	97,53	2,78	1	1	High	1
FALSE	Medium	Medium	Master	Master Protein	Q5D862	SWISS-PROT:Q5D862 Tax_Id=9606 Gene_Symbol=FLG2 Filaggrin-2	0,022	0,013	0	2	2	1	2391	247,9	8,31	111,78	2,72	1	1	High	1
FALSE	Medium	Medium	Master	Master Protein		[gene=slyD] [locus_tag=ECD_03200] protein=FKBP-type peptidyl prolyl cis-trans isomerase (...	0,042	0,026	5	1	2	1	196	20,8	5,05	105,1	2,54	1	1	High	1
FALSE	Medium	Medium	Master	Master Protein		[gene=metL] [locus_tag=ECD_03826] protein=Bifunctional aspartokinase/homoserine dehyd...	0,038	0,025	1	1	2	2	810	88,9	5,59	97,92	2,38	1	1	High	1
FALSE	Medium	n/a	Master	Master Protein	P02662	SWISS-PROT:P02662 Alpha-S1-casein - Bos taurus (Bovine).		0,021	6	1	1	1	199	23	4,94		2,17		1	High	1
FALSE	Medium	n/a	Master	Master Protein		[gene=fbaB] [locus_tag=ECD_02025] protein=fructose-bisphosphate aldolase class I] protein	0,021	0,021	3	1	1	1	350	38,1	6,73		2,16		1	High	1
FALSE	Medium	n/a	Master	Master Protein		[gene=tnaA] [locus_tag=ECD_03592] protein=tryptophanase/L-cysteine desulfhydrase, PLP-d	0,02	0,02	3	1	1	1	471	52,7	6,23		2,11		1	High	1
FALSE	Medium	Medium	Master	Master Protein	ENSEMBL	[gene=b] [locus_tag=ECD_00881] protein=putative QLD family ATP-dependent endonucle		0,02	3	1	1	1	451	50,1	5,06		2,11		1	High	1
FALSE	n/a	Medium	Master	Master Protein		[gene=nfiE] [locus_tag=ECD_02923] protein=fused deadenyl)transferase/adenylyltransfera...	0,023	0,019	4	1	1	1	552	63,5	9,26		0		1	High	1
FALSE	Medium	n/a	Master	Master Protein		[gene=plsB] [locus_tag=ECD_03913] protein=glycerol-3-phosphate O-acyltransferase] [protei...	0,023		2	1	1	1	946	108,3	5,2	118,22			1	High	1
FALSE	Medium	n/a	Master	Master Protein		[gene=acs] [locus_tag=ECD_03941] protein=acetyl-CoA synthetase] [protein_id=ACT45731.1]	0,022	0,013	2	1	1	1	807	91,3	8,38	117,33			1	High	1
FALSE	Medium	n/a	Master	Master Protein	Q3SX28	TREMBL:Q3SX28;Q5R48 (Bos taurus) similar to alpha-tubulin isoform 1	0,021	0,021	2	1	1	1	652	72	5,8	108,29			1	High	1
FALSE	Medium	n/a	Master	Master Protein	Q3SX28	TREMBL:Q3SX28;Q5R48 (Bos taurus) Tropomyosin 2	0,039	0,02	6	1	1	1	284	33	4,69	98,68			1	High	1
FALSE	Medium	n/a	Master	Master Protein		[gene=ydeI] [locus_tag=ECD_01496] protein=inactive PmcC family protein] [protein_id=ACT4...	0,037	0,037	11	1	1	1	172	18,3	4,88	96,98			1	High	1

9.4 Appendix Gel filtration scatter plot

scatter plot used to calculate the molecular weight for 12AMOR1-GH13. Log of molecular weight of the known standards plotted against the calculated K_{ave} for the standards. Trendline used to calculate the molecular weight of 12AMOR1-GH13