

Table of contents

1. Abstract.....	3
2. Resumo alargado	4
3. Introduction	7
3.1 Background.....	7
3.2 Origin from <i>Archeoglobus fulgidus</i> strain 7324	7
3.2.1 Applications of thermostable enzymes.....	9
3.3 α - Amylase, the amylase family and starch.....	9
3.4 CGTase (Cyclodextrin glucoamylase).....	10
3.5 Cyclodextrins.....	12
3.6 <i>Escherichia coli</i> (<i>E.coli</i>)	12
3.6.1 NovaBlue.....	12
3.6.2 BL21 (DE3)	13
3.6.3 CodonPlus.....	13
3.6.4 BL21 Star (DE3).....	13
3.6.5 pRARE from Rosetta.....	13
3.7 pET-17b vector	14
3.8 T7 Lac promoter	14
4. Materials and Methods	15
4.1 Strain and plasmid construction	15
4.2 Preparation of Competent cells.....	15
4.2.1 Chemically competent cells.....	15
4.2.2 Electrocompetent cells.....	16
4.3 Transformation	16
4.3.1 Transformation by thermic shock.....	16
4.3.2 Transformation by electroporation	17
4.4 <i>E.coli</i> glycerol stocks.....	17
4.5 Plasmid DNA preparation	17
4.6 Site-Directed Mutagenesis.....	17
4.7 Agar plates for colony growth.....	18
4.8 Agarose Gel	18
4.9 PCR.....	19
4.10 Extration and quantification of DNA	19
4.11 Shake Flask Experiments.....	20
4.12 SDS-PAGE	20
4.13 Heat treatment.....	21
4.14 Batch cultivation.....	21
4.14.1 Preparation of materials and medium.....	21
4.14.2 Inoculum.....	22
4.14.3 The cultivation.....	22
4.14.4 Sampling and sample treatment.....	22
4.14.5 Induction.....	23
5. Results and Discussion	24
5.1 Plasmid construction.....	24
5.2 Codon usage	25
5.3 Protein expression.....	26
5.3.1 Expression in different media using the strain Codon Plus.....	27
5.3.2 Construction of a new expression strain.....	29

5.3.3	Different heat treatments	31
5.4	Site-Directed mutagenesis	33
6.	Conclusion	38
7.	Acknowledgement	39
8.	References	41

1. Abstract

The present work is referent to the optimization of the expression and production of a novel thermostable enzyme, CGTase (cyclodextrin glucanotransferase), from *Archaeal* origin.

The origin of this gene is an important point because it is a group that is unusual to have CGTase activity. Some of the *Archaeal* organisms are known as thermophiles. The enzymes of thermophiles are characterized by extreme stability and activity, large potential in industries like food and chemical industries and chemical synthesis. Also in biotechnological processes the increased reaction temperatures leads to reduction of viscosity, higher reaction rates and also reduced risk of contaminations. CGTase is a thermostable enzyme and is a member of the α -amylase superfamily. It is used to produce cyclodextrins that are used in pharmaceutical industries, it is an extracellular enzyme and degrades large starch molecules that can be used by the organism.

Four different strains were used to express this gene, where two of them were used to try to control the problem at the transcriptional or at the translational level.

After comparison between codons from *Escherichia coli*, codons from the gene (CGTase) and codons from pRARE (plasmid with rare codons), a mutation was introduced in those with less percentage of codons in *E.coli* and one of the rare codons in the sequence showed better expression. The sequence of the mutation revealed that the construct was not under the control of the T7 promoter, having relatively lower constitutive expression of the gene, not being significantly affected with the optimizations when they were effectuated.

Future work could be to put the gene of interest in the right direction (turning the gene), using the appropriate restriction enzymes and again try the right expression with BL21 Star (DE3).

Keywords: CGTase, expression, SDS-Page, BL21 Star (DE3), pRARE.

2. Resumo alargado

Este trabalho baseia-se na optimização da expressão de uma enzima termostável, a ciclodextrina glucanotransferase (CGTase) com origem no archaeon *Archaeoglobus fulgidus*.

A origem do gene é de extrema importância uma vez que há organismos com preferências diferentes no que diz respeito a codões, influenciando assim a expressão em hospedeiros heterológicos. Adicionalmente a comparação de sequências tem mostrado que a enzima pertence a um grupo que normalmente não tem actividade de CGTases.

Organismos do grupo *Archaea* têm características do domínio *Bacteria* e também do domínio *Eucariota*. Alguns destes organismos vivem a temperaturas perto dos 100°C e são conhecidos como termófilos. Geralmente as enzimas dos termófilos estão adaptadas às condições de crescimento dos seus organismos nativos, e assim, as enzimas de termófilos são caracterizadas por terem extrema estabilidade e actividade, tendo assim um grande potencial na indústria. A CGTase estudada neste trabalho é uma enzima termostável extracelular e é membro da superfamília α -amilase. As amilases cortam arbitrariamente no interior dos polímeros de amido, levando à formação de oligossacarídeos divididos e lineares. As aplicações das α -amilases são: liquefacção, sacarificação, utilização em detergentes para roupa, processamento têxtil, clarificação de cerveja ou de sumos de fruta, no pré-tratamento da alimentação animal na produção de etanol entre outras. As CGTases geralmente são utilizadas para produzir ciclodextrinas (CD) que são moléculas utilizadas em indústria farmacêutica onde a reacção catalítica específica para as CGTases é a de ciclização, onde a parte doadora que é retirada age também como aceitador formando assim as CD.

É muito importante escolher um biocatalisador que consiga produzir em quantidades satisfatórias num processo definido e determinado. Quando um gene é clonado num outro hospedeiro, e produzido recombinantemente para obter alta densidade celular em batch, fed-batch ou em modo contínuo, os rendimentos são normalmente elevados. É difícil escolher o tipo de promotor e hospedeiro a utilizar porque isso depende das características da proteína alvo. Também é possível produzir uma enzima no organismo nativo mas poderá ser difícil atingir melhores rendimentos e também pode tornar-se complicado regular a expressão da proteína. A expressão em sistemas bacterianos é mais interessante porque são espécies fáceis de cultivar, podem crescer mais rapidamente e ainda conseguem maior densidade celular com meios relativamente baratos. Hoje em dia vectores de clonagem e estirpes estão vastamente acessíveis. Contudo, a incapacidade das bactérias de glicosilar proteínas pode ser um ponto negativo para alguns estudos de expressão. *Echerichia coli* é o hospedeiro mais utilizado e melhor conhecido de todos os organismos para expressão de proteínas recombinantes. Podemos encontrar alguns problemas quando temos

expressão em *E.coli*, tais como baixos níveis de expressão, ou *folding* incorrecto (como corpos de inclusão). Contudo isso pode ser resolvido aplicando diferentes estratégias de produção ou utilizando diferentes tipos de estirpes. No presente trabalho, foram utilizadas algumas destas estratégias tendo sempre em vista aumentar a expressão e reduzir sempre que possível todos os problemas na produção. As diferentes estirpes utilizadas foram: 1. NovaBlue, hospedeiro ideal de clonagem pois tem grande eficiência de transformação resultando em elevados rendimentos de excelente qualidade de DNA plasmídeo. Esta estirpe não é contudo utilizada para expressão mas mesmo assim é uma boa estirpe para a construção inicial de clonagem. Também guarda a posição extracromossomal do plasmídeo se este for deficiente a nível da estratégia de recombinação; 2. BL21 é a estirpe mais utilizada para expressão em sistemas pET; 3. CodonPlus é uma estirpe que resolve problemas a nível de codões que estejam em falta, pode aumentar a expressão de proteínas fornecendo cópias adicionais de genes específicos de tRNA que possam ser raros em *E. coli*. Quando o mRNA dos genes alvo é sobre-expresso em *E.coli*, as diferenças em codões podem impedir a tradução por causa de um ou mais tRNAs que podem ser raros na população. Pode-se controlar genes ao induzir a expressão induzida por IPTG do promotor T7; 4. BL21 Star (DE3) é uma estirpe de *E.coli* que deriva da estirpe BL21 (DE3). É também apropriado para elevado nível de expressão de proteínas recombinantes, a diferença é que esta estirpe tem um aumento de estabilidade dos mRNAs; 5. pRARE é o plasmídeo retirado da estirpe Rosetta também derivada de BL21 e foi concebida para aumentar a expressão de proteínas *eucariotas* que contenham codões raros utilizados em *E.coli*, tal como CodonPlus. A diferença é que esta estirpe tem os genes para estes codões tRNA fornecidos no plasmídeo extracromossomal (pRARE) que é também integrado no cromossoma em CodonPlus. Auxiliando com estes tRNAs raros, a estirpe Rosetta é utilizada para tradução “universal” que pode ser de outra forma limitada pelos codões de *E.coli*.

Para estudar a expressão do gene CGTase neste trabalho foram testados então diferentes meios, diferentes estirpes, diferentes tratamentos e foi feita uma mutação baseada na comparação entre codões raros em *E.coli* e no gene CGTase. Quatro estirpes diferentes de *Escherichia coli* foram testadas onde duas delas foram utilizadas para aumentar a estabilidade na transcrição ou a produção a nível da tradução. Foi feita também uma nova estirpe para expressão (transformando a estirpe BL21 Star (DE3) com o plasmídeo pRARE), combinando assim o aumento da estabilidade nos RNAs com os codões raros de tRNA. Depois da comparação dos codões de *E.coli* com os codões do gene (CGTase), descobriu-se que muitos dos codões raros estavam a ser utilizados no gene de origem *Archaea*. Esta situação foi compensada com a adição de genes que predeterminam muitos dos codões raros que acontecem em *E. coli* (utilizando a estirpe CodonPlus ou o plasmídeo pRARE). Também foi feita uma mutação com o objectivo de remover os codões raros que estavam perto da extremidade 5' no gene CGTase. Muitos dos efeitos mais difíceis de contornar

são frequentemente observados quando múltiplos e consecutivos codões raros se encontram perto do N-terminal de uma sequência. As mutações silenciadas (remoção de codões raros no N-terminal) deram melhor expressão, e estes resultados foram conferidos por SDS-Page. No entanto a sequenciação revelou que a sequência na construção efectuada não estava sob o controlo do promotor T7, mas sim inserida a montante no plasmídeo. Consequentemente a expressão foi constitutiva e não foi significativamente afectada pela estratégia de optimização que foi efectuada ao longo do trabalho. Para trabalho futuro, deve incluir o ponto em que se irá colocar o gene a jusante no plasmídeo por PCR, ou seja, colocar o gene sob controlo do promotor T7, no início da sequência utilizando as enzimas de restrição apropriadas e mais uma vez tentar novamente a expressão com a estirpe BL21 Star (DE3) com o plasmídeo pRARE.

Palavras-chave: CGTase, expressão, SDS-Page, BL21 Star (DE3), pRARE.

3. Introduction

The aim of the project was the optimization of the expression of a novel glycosyl transferase (CGTase), originated from *Archeoglobus fulgidus* in *E. coli*.

3.1 Background

An initial expression in shake flasks using the pET-system (pET17b, expression under control of the T7/lac promoter) in *E.coli* BL21(DE3), has shown a relatively low expression level, and continued production in this system requires optimization. Optimization was planned to include expression in a set of strains, with different properties, and also with different media compositions. The more promising combinations of strain/medium were taken to cultivation in fed-batch, using a fed-batch system with probing of the response in DOT-level (Dissolved Oxygen Tension) to feed-pulses (to control glucose-concentration to relevant level), keeping the desired overall level of oxygen in the reactor by a combination of stirrer speed control and temperature control.

3.2 Origin from *Archeoglobus fulgidus* strain 7324

Archeoglobus fulgidus is a hyperthermophilic sulfate reducing archaeon and was isolated from hot (75°C) oil field waters from an oil production platform in the Norwegian sector of the North Sea. This strain was enriched on a complex medium and isolated on lactate with sulfate. The cells were non motile, irregular coccoid to disc shaped, and 0.3 to 1.0 µm wide [1]. The cells grow between 60 and 85 °C with an optimum at 76 °C. Lactate, pyruvate, and valerate plus H₂ can be utilized as carbon and energy sources with sulfate as electron acceptor [1]. Strain 7324 showed 100% DNA-DNA homology with *A. fulgidus* Z, and also on the basis of morphological, physiological, and serological features, the strain was classified as an *Archaeglobus sp.* *Archaeglobus sp.* has been selectively enriched and immunomagnetically captured from oil field waters from three different platforms in the North Sea [1].

Such surveys indicate that a significant amount of microbial diversity found in natural habitats has so far eluded cultivation in the laboratory [2]. Characterizing these microorganisms remains a critical goal for gaining a deeper and truer appreciation of microbial evolution, physiology, and ecology. *Archaea* were initially thought to preferentially occupy environments that are inhospitable to *Eucarya* and *Bacteria*, but their ecological and physiological diversity has turned out to be far greater than previously supposed [2].

Archaea include inhabitants of some of the most extreme environments growing at temperatures up to or even above 100 °C (thermophiles), some may live in extreme saline water such as halobacteria and some say that these organisms may be found in extraterrestrial environments (like on Mars) [3]. *Archaea* may not look different from any normal bacteria but are biochemically and genetically different [3]. Figure 1

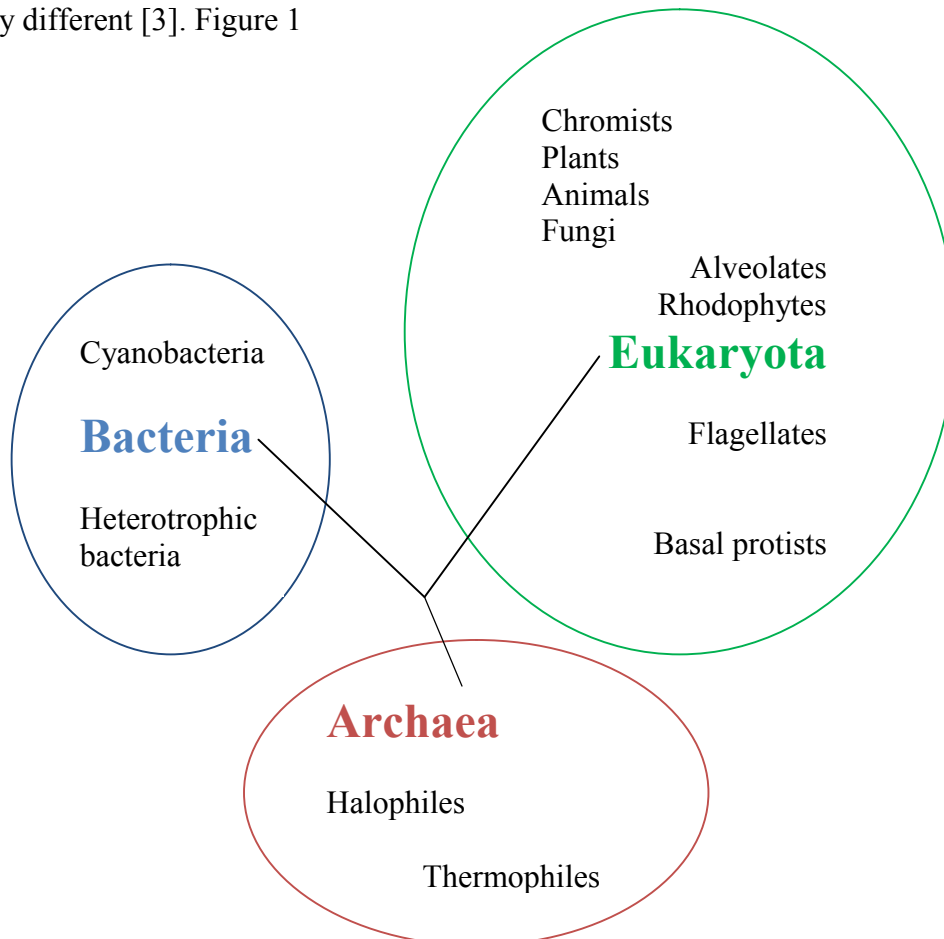


Figure 1 – The three domains of life: Archaea, Eukaryota and Bacteria. Archaea in general have characteristics from Bacteria domain and Eukaryota domain [3].

The hyperthermophilic archaeon *Archaeoglobus fulgidus* strain 7324 has been shown to grow on starch and sulfate and thus represents the first sulfate reducer able to degrade polymeric sugars [4]. *A. fulgidus* utilizes an unusual pathway of starch degradation involving cyclodextrins as intermediates. The pathway comprises the combined action of an extracellular cyclodextrin glucanotransferase (CGTase) converting starch to cyclodextrins and the intracellular conversion of cyclodextrins to glucose 6-phosphate via cyclodextrinase (CDase), maltodextrin phosphorylase (Mal-P), and phosphoglucomutase (PGM) [4].

3.2.1 Applications of thermostable enzymes

In general enzymes are adapted to the growth conditions of their native organisms. Consequently enzymes of hyperthermophiles are characterized by extreme stability and activity at high temperatures [5]. These enzymes, also referred as “extremozymes”, show a large potential in industry. Extremozymes are already used in food and chemical industries, i.e, for hydrolysis of starch, polymer degradation of detergents, food colouring and brewing. Especially thermophiles and hyperthermophiles offer a big source of novel enzymes for industrial application [5]. Running biotechnological processes with thermophile enzymes has many advantages: increased reaction temperature influences bioavailability and solubility of organic compounds; decreases viscosity and leads to higher reaction rates and also reduces the risk of contaminations. Thermostable DNA polymerases and DNA ligases are used in molecular biology. So far one of the main commercial application is thermostable *Taq* DNA polymerase of *Thermus aquaticus* used for polymerase-chain-reaction (PCR). Also in starch processing amylases, pullulanases, and glucosidases of thermophilic organisms play a decisive role. Further industrial applications for thermophilic enzymes include cellulose degradation, paper pulp bleaching, ethanol production and chemical synthesis [5].

3.3 α - Amylase, the amylase family and starch

α -Amylase has been named so, on the basis that it acts on amylase (linear α -1,4-linked polymer built from glucose entities, and one of the polymers in starch) [6]. Amylases cleaves arbitrarily in the interior of the starch polymer, which leads to the formation of linear and divided oligosaccharides. The applications of α -amylases are the starch liquefaction and saccharification, use as laundry detergent, textile processing, clarification of beer and fruit juices, pre-treatment of animal feed, anti-staling in bread and ethanol production [6].

Starch is an abundant renewable resource used widely in the food, pharmaceutical, textile, and paper industry. Amylolytic enzymes, e.g. α -amylases and cyclodextrin glucoamylases, modify starch in various ways and are some of the most used enzymes in industry accounting for 25% of the total enzyme market [6]. Starch from cultivated plants is used as a high-calorie food source and it is used in different food applications since it improves the functional properties of foods, such as gelling and pasting. It is also an abundant and accessible energy source that can be used for the manufacture of biofuels [6]. These applications require efficient degradation (hydrolysis), which can be made enzymatically. Starch hydrolysis is usually performed in a two-

step process of liquefaction and saccharification where liquefaction is the conversion of granular starch into soluble, shorter-chain-length dextrans [6].

The hydrolases and transferases that constitute the α -amylase family are multidomain proteins, but each has a catalytic domain in the form of a $(\beta/\alpha)_8$ -barrel, with the active site being at the C-terminal end of the barrel β -strands. The enzymes have mainly been assigned to three separate families – 13, 70 and 77, based on amino acids sequence similarities. But in the case of the CGTase used in this work, it belongs to a special group including few common sequences, the GH57 family. Each enzyme in family 13, 70 and 77 has one glutamic acid and two aspartic acid residues necessary for activity, while most enzymes of this family also contain two histidine residues critical for transition state stabilization [7]. One glutamic acid and one aspartic acid has also been shown to be necessary for catalysis in family 57 [8] but if additional sequence conservation occurs in this family remains to be investigated. The sequence-specificity and structure-specificity relationships described may provide useful pointers for rational protein engineering [7].

3.4 CGTase (Cyclodextrin glucanotransferase)

Cyclodextrin glucanotransferase (CGTase), figure 2, is an important industrial enzymes used to produce cyclodextrins (CDs). Many bacteria use starch as carbon and energy source for growth. To utilize this carbon and energy source, they produce a range of extracellular enzymes to convert the large molecules to utilizable small molecules. CGTase catalyzes the cleavage of the glycosidic bonds in starch, to produce these smaller circular molecules [9].

CGTases have received considerable attention, since cyclodextrins (CD) have many applications that previously were mentioned. The systematic name of CGTase is 1,4- α -D-glucan 4- α -D-(1,4- α -D-glucano)-transferase (cyclizing). It's a hexosyltransferase and belongs to the transferase group of enzymes. It catalyzes mainly transglycosylation reactions but can also hydrolyze starch [9]. The distinctive feature of this enzyme is the formation of cyclic products [9].

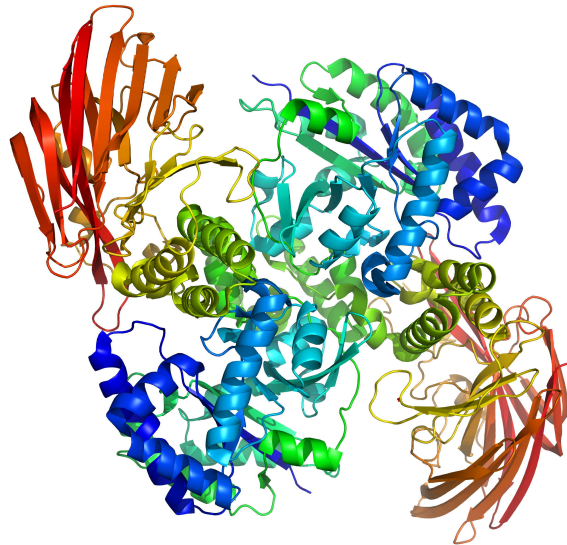


Figure 2 – Typical structure of an enzyme from the alpha-amylase family. The enzyme displayed here has two domains (one alpha/beta barrel (blue-green), and one domain composed of beta sheets (red)), and is shown as a dimer. An overall structure of this type would (based on sequence similarity and modelling) be likely for the CGTase studied in this work.

The catalytic reaction specific for CGTases is the cyclization reaction, in which the part of the donor that has been cleaved off also acts as the acceptor, resulting in the formation of a CD. The reverse reaction is also catalyzed by the same enzyme and is referred to as the coupling reaction [9].

CGTase is an extracellular enzyme which has been detected in a number of bacteria. Nearly 30 CGTase genes have been isolated since the first identification of this gene in *Bacillus macerans*. Some of these genes were found to be located in a gene cluster in the chromosome, such as the CGTase genes of *Klebsiella oxytoca*, which contain genes for the extracellular synthesis of CD, membrane transport and intracellular degradation of CD [9]. A complete genome sequencing analysis indicated that such a gene cluster may also exist in *Streptococcus pyogenes* [9].

CGTase from different sources show a clear similarity in their amino acid sequences, ranging from 47% to 99% [9]. However this level of sequence conservation is only valid for the enzymes classified under family 13. It is still not shown which of these conserved residues that also putatively exist in family 57.

Considering the distribution and physiological role, CGTase degrades large starch molecules into small molecules such as CD and oligosaccharides, which can be taken up and utilized by the organisms. Studies suggested a CD utilization system, where CDs are formed outside the bacterial

cells and transported through the cell membrane with the help of a specific transport protein. The CDs are then hydrolyzed by a cytoplasmic cyclomaltodextrinase (CDase) inside the cells [9].

3.5 Cyclodextrins

Cyclodextrins (CDs) are non-reducing cyclic glucose oligosaccharides resulting from the cyclomaltodextrin glucanotransferase (CGTase) catalyzed degradation of starch [10]. There are three common CDs with 6, 7 or 8 D-glucopyranosyl residues (α -, β -, and γ -cyclodextrin respectively) linked by α -1,4 glycosidic bonds. All three CDs have similar structures (that is, bond lengths and orientations) apart from the structural necessities of accommodating a different number of glucose residues [10]. They present a bottomless bowl-shaped (truncated cone) molecule stiffened by hydrogen bonding between the 3-OH and 2-OH groups around the outer rim. The hydrogen bond strengths are α -CD < β -CD < γ -CD [10].

3.6 *Escherichia coli* (*E.coli*)

It is important to choose a biocatalyst that can be produced in satisfactory quantity in a defined process. When the gene is cloned in another host, and produced in a recombinant way for high cellular density in batch, fed-batch or continuous mode, the yields are normally elevated. It is difficult to choose the type of promoter and host used because it depends on the target protein. Also it is possible to produce a target enzyme in the native organism but it could be difficult to have better yields and also could be tricky to regulate the expression [6]. The expression with bacterial systems is more interesting because there are species easier to cultivate and they can grow faster and with bigger density with cheaper media. Cloning vectors and host strains are vastly accessible today. However incapacity of bacteria to glycosylate proteins can be a drawback for some expression studies [6]. *E.coli* is the most used and known host of all prokaryotic organisms for recombinant proteins expression [6]. We can find some problems when we have expression in *E.coli*, like problems with low expression, or incorrect folding (as inclusion bodies), but this can be solved by applying different cultivation strategies, or by utilizing different types of host strains [6].

3.6.1 NovaBlue

NovaBlue is a K-12 strain ideally suited as an initial cloning host due to its high transformation efficiency, blue/white screening capability (with appropriate plasmids) and recA endA mutations, which result in high yields of excellent quality plasmid DNA [11]. This strain is hence not used

for expression, but is instead a good strain for the initial cloning of a construct. It also keeps the extrachromosomal location of the plasmid as it is recombination deficient.

3.6.2 BL21 (DE3)

BL21 is the most widely used expression host and has the advantage of being deficient in both *lon* (3) and *ompT* proteases [11]. An expression strain must have the appropriate regulatory proteins for expression of cloned genes. To improve yields of expressed proteins from the cloned genes, expression strains are often deficient in some proteases. For the pET system, expression hosts are λ DE3 lysogens, which have the inducible gene for T7 RNA polymerase, allowing selective induction of the target protein [12].

3.6.3 CodonPlus

To resolve the codon bias problem that can hinder the expression of heterologous proteins in *E. coli* hosts, we can select BL21-CodonPlus™ competent cells as an expression host. This derivative can improve protein expression by supplying additional copies of specific tRNA genes that are rare in *E. coli*. When the mRNA of heterologous target genes is overexpressed in *E. coli*, differences in codons usage can impede translation due to the demand for one or more tRNAs that may be rare or lacking in the population [13]. BL21-CodonPlus (DE3) cells can induce the IPTG-inducible expression of T7 promoter controlled genes. It has the T7-polymerase present, as described for BL21(DE3) above [14].

3.6.4 BL21 Star (DE3)

The BL21 Star (DE3) *E. coli* strains are also derived from the BL21 (DE3) strain. The BL21 Star (DE3) is also suitable for high-level recombinant protein expression. The difference is that this strain has an increase in stability of mRNAs. This result in a sometimes observed higher basal expression of heterologous genes in BL21 Star strains compared to other BL21 strains [15].

3.6.5 pRARE from Rosetta

This host strain is another BL21 derivative designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*, just like the strain CodonPlus described above. The difference is that this strain has the genes for these tRNA-codons supplied in an extrachromosomal plasmid (pRARE), while it is integrated in the chromosome in CodonPlus. By supplying these rare tRNAs, the Rosetta strains provide for “universal” translation, which would otherwise be limited by the codon usage of *E. coli* [16]. In particular Arg codons, Ile and Leu

codon all represent less than 8% of their corresponding population of codons. The Rosetta strain carry the pRARE plasmid and supply tRNAs for the codons AUA, AGG, AGA, CUA, CCC, and GGA on a chloramphenicol-resistant plasmid [13].

3.7 pET-17b vector

The pET17b vector (from Novagen) carries an N-terminal T7 Tag sequence followed by a region of useful cloning sites.

Based on the T7 promoter-driven system, the pET System has been greatly expanded and now includes over 23 vector types, 11 different *E. coli* host strains for efficient detection and purification of target proteins [17].

3.8 T7 Lac promoter

The pET plasmids contain an expression cassette in which the gene of interest is inserted behind an extremely strong promoter from the *E. coli* bacteriophage T7. In the absence of the T7 polymerase, this promoter is completely shut off. For expression, the pET plasmids are transformed into bacteria strains that typically contain a single copy of the T7 polymerase on the chromosome in a λ lysogen (the most commonly used lysogen is known as DE3). The T7 polymerase is under the control of the Lac-UV5 lac promoter. When cells are grown in media without lactose, the lac repressor (lacI) binds to the lac operator and prevents transcription from the lac promoter. When lactose is the sole carbon source, or when the lactose analog IPTG is added to the media, lactose (or IPTG) binds to the repressor and induces its dissociation from the operator, permitting transcription from the promoter. Finally, addition of glucose to the culture media contributes to repression of the T7 RNA polymerase via the mechanism of catabolite repression [18].

4. Materials and Methods

Since the aim of the work is the optimization of the production of the *A. fulgidus* CGTase in an *E. coli* strain, we analyzed the work done by Antje Labes and Peter Schönheit [4], because they tried the same strain and the same gene, the CGTase, and we want to improve the expression of the same protein.

4.1 Strain and plasmid construction

E. coli expression vector pET-17b is from the Novagen Company. The insert, the gene encoding the CGTase, was kindly supplied by Prof. Schönheit (Universität Kiel, Germany). The CGTase encoding gene was amplified and inserted in pET-17b and transformed into electrocompetent *E. coli* cells [BL21 (DE3), NovaBlue, CodonPlus, BL21 Star (DE3) and in the end of this study, cotransformed with the helping plasmid, pRARE in BL21 Star] which were used in expression studies. Two specific silent mutations were also attempted in the CGTase gene, with the QuickChange II Site-Directed Mutagenesis Kit and the resulting plasmid expressed in BL21 Star (DE3) with the coplasmid pRARE.

4.2 Preparation of Competent cells

Competent cells to allow transformation were made and the protocols adapted for transformation either by thermic shock (chemically competent cells), or electric shock (electrocompetent cells).

4.2.1 Chemically competent cells

A loopful of *E. coli* cells was added to 20 mL of LB medium [19] and incubated overnight at 37 °C. The next day 100 µL of the overnight culture was incubated in 40 mL of LB medium and grown at 37 °C until $OD_{550} \sim 0,5$ (~3h). At this point the cells were cooled. The cells were swirled on ice for 3-5 min followed by: centrifugation at 4 °C and 8500 rpm for 8 min. The cell pellet was resuspended carefully in 10 mL of chilled 0,1 M $MgCl_2$ during 5-10 minutes, and again centrifuged at 4 °C and 8500 rpm. This was repeated twice, and the pellet was finally suspended in 1800 µL of 0,1 M $CaCl_2$ and 300 µL of 80% glycerol. Aliquots of 100 µL each were stored at -80 °C.

4.2.2 Electrocompetent cells

An overnight culture of 20 mL was prepared by shaking at 37 °C and the cells were added to fresh medium so that OD₆₂₀ would be 0.1. The cultivation curve was plotted and the cultivation was stopped when the OD₆₂₀ reached 0.5. 500 mL of mixture (with LB medium) was put on ice water bath for 15-30 minutes and the bottles centrifuged at 4 °C. The cells were centrifuged at 8000 rpm for 10 minutes at 4 °C and the supernatants were poured off and the pellet resuspended in 1 volume of 10% of cold sterile glycerol. The cells were centrifuged again at 8000 rpm for 10 minutes at 4 °C and again the supernatant were poured off and the pellet resuspended in 0,5 volume of 10% of cold sterile glycerol. The cells were centrifuged a third time at 8000 rpm for 10 minutes at 4°C and the supernatant were poured off and the pellet resuspended in 0,1 volume of 10% of cold sterile glycerol and transferred to smaller tubes. In the end of the process, the cells were centrifuged at 3750 rpm in the cold room for 10 minutes and the supernatants were poured off and the pellet resuspended in 0,002 volume of 10% cold sterile glycerol and divided into cryo tubes of 50 µL in each and the tubes were stored immediately at -80 °C.

4.3 Transformation

Two methods were used to introduce plasmid DNA into competent cells, which are described below.

4.3.1 Transformation by thermic shock

The competent cells were removed from -80 °C and thawed on ice for 20 minutes. 1-2µL of plasmid DNA was added to 100 µL of competent cells and swirled gently. The competent cells stored on ice for 30 minutes and an heat shock was done to the competent cells at 42 °C for 30 s. The competent cells stored on ice for 10 minutes and 200 µL of SOC or LB medium was added to the cells. The competent cells were incubated at 37 °C for 1 h and 200 µL of competent cells were spread and plated on selection plates with the appropriate antibiotic (LB with 100 µg/mL ampicillin). The plate was incubated overnight at 37 °C. A negative control without plasmid should be included and also a positive control (pUC18 plasmid).

4.3.2 Transformation by electroporation

Electrocompetent cells (50-100 μ L per reaction) and cuvettes (Bio-Rad Gene Pulser Cuvette, 0.2 cm, cat.no.1652086) were chilled on ice for 15 min. Plasmid DNA 1-5 μ L is added to the electrocompetent cells and the mixture stirred carefully with a pipette tip. The cells are then transferred to a cuvette and incubated on ice for 30 minutes. The cuvette is carefully dried, and the cells electroporated with Bio-Rad Gene pulser at 25 μ F, 400 Ω , 10-12 kV/cm. The time constant should be in the range 5-10 ms. 1 mL of SOC or LB medium pre-heated to 37 $^{\circ}$ C is added to each cuvette, followed by gentle mixing, and incubation at 37 $^{\circ}$ C for 1 h. The cells are streaked into two plates (150 μ L on each) with appropriate antibiotic and incubated overnight at 37 $^{\circ}$ C.

A negative control, without plasmid and a positive control (pUC18 plasmid) should be included.

4.4 *E.coli* glycerol stocks

This method was used to preserve cells. A single colony was picked of the clone and grown overnight in the appropriate selective liquid medium (e.g., LB ampicillin). A label was made for the construct and placed into a sterile screw cap microcentrifuge tube. 800 μ L of the overnight culture was added to 200 μ L of 80% of sterile glycerol in the sterile screw cap microcentrifuge tube and vortexed and the glycerol stocks were frozen at -80 $^{\circ}$ C.

4.5 Plasmid DNA preparation

Plasmids were prepared, by an alkaline lysis method using the NucleoSpin Plasmid kit (from Clontech).

A volume of 1-5 mL of saturated *E.coli* LB overnight culture was pelleted in a standard benchtop microcentrifuge for 30 s at 11,000x g. The purification proceeded according to the instructions of the manufacturer, and the plasmid was finally eluted using either 50 μ L of AE buffer or distilled water.

4.6 Site-Directed Mutagenesis

Site-directed mutagenesis was made using the QuickChange II site-directed mutagenesis kit (from Stratagene).

The sample reactions for mutant strand synthesis were prepared in a PCR tube (table 1) and left on ice for 2 minutes. The template has to be prepared from a methylating *E.coli* strain (like NovaBlue), as the template is selectively digested using the enzyme DpnI after the amplification.

Table 1 – Sample reaction for mutant strand synthesis

Reagent	Final Concentration	Volume (μL)
DI water	-	28,7
10x reaction buffer	1x	5
Primer #1	125ng	1,15
Primer #2	125ng	1,15
dNTP mix	kit conc	1
Template	(5-50ng)	12
<i>PfuUltra</i> DNA polymerase	2.5U/ μL	1
Total		50

The PCR reaction for introducing the mutations and amplifying the gene fragment was set up using the parameters defined in table 2.

Table 2 – Cycling parameters

Segment	Cycles	Temperatura ($^{\circ}\text{C}$)	Time
1	1	95	30 sec
2	16	95	30 sec
		72	1 min
		68	8 min

After amplification, the template is digested with Dpn I by adding 1 μL Dpn I (10 U/ μL) directly to the amplification reaction which is incubated at 37 $^{\circ}\text{C}$ for 1 h. The digest is then transformed to *E. coli* strain BL21 Star (DE3) with the same procedure as the point 3.3.1. Plasmids from the resulting clones are purified using step 3.5.

4.7 Agar plates for colony growth

LB agar (Merck) was used as medium for the plates, and a resistance necessary for each experiment was added: ampicillin, chloramphenicol, zeocin, and also ampicillin and chloramphenicol.

4.8 Agarose Gel

DNA fragments (which are negatively charged) were separated according to size by agarose electrophoresis. The gels were made and separated in 50 mL buffer 1x TAE and 0.4 g of agarose. Agarose was melted (3 minutes) in the microwaves, and upon cooling but before gelling 2 μL

ethidium bromide was added and the gel solution poured into a tray, in which combs were inserted to obtain wells. After approximately 15 minutes the gel was solid and could be used.

4.9 PCR

The polymerase chain reaction (PCR) results in the selective amplification of a chosen region of DNA molecule. The pET-17b plasmid containing the CGTase gene was amplified by PCR. Two short oligonucleotides must hybridize to the DNA molecule, one to each strand of the double helix. These oligonucleotides, which act as primers for DNA synthesis reaction delimit the region that will be amplified, so the primers are the key to the success or failure of the PCR experiment [20]. Table 3 shows the conditions of the PCR amplification and in table 4 the mixture and components for the reaction are shown.

Table 3 – Cycling parameters

Segment	Cycles	Temperatura (°C)	Time
1	1	98	30 sec
2	30	98	10 sec
		72	1 min
		68	10 min

Table 4 –Master Mix

	Co	Cf	Volume (µL)
H2O			28.5
Buffer HF	5x	1x	10
dNTPs	10mM	200µM	1
Primer F	10mM	0.5µM	2.5
Primer R	2/100	0.5µM	2.5
Plasmid			5
Fusion enzyme		0.02U/µL	0.5

4.10 Extration and quantification of DNA

The PCR products were separated in an agarose gel and thereafter, extracted using the QIAEX II Agarose Gel extraction protocol (Qiagen, Germany) according to the manufacturers instructions. The PCR products were purified and quantified spectrophotometrically at 260 nm (Nanodrop, Saveen & Werner).

4.11 Shake Flask Experiments

E.coli inoculum cultures (5 mL) were prepared in a 15 mL tube. The tube contained 5 mL of medium (LB), 100 μ L of frozen inoculum, 5 μ L of ampicillin and 3.4 μ L of chloramphenicol (if pRARE was used) (100 mg/mL) was added and was used after overnight at 37°C at 200rpm.

Different shake flask experiments were done. All the experiments were carried in 1 L sterile baffled shake flasks using the shaker at 37 °C at 200 rpm and a working volume of 200 mL. The different media compositions are presented in the next table:

Table 5 – Shake flask media for growth

Component	LB	YT	NYAT
Bacto-Tryptone	10g	16g	
Bacto-Yeast Extract	5g	10g	
NaCl	10g	5g	
DI Water	~970mL	~960mL	~776mL
5xNYAT salts			200mL
50% Glucose			20mL
1M MgSO ₄			2mL
Trace Elements			2mL
Ampicillin/Chloramphenicol		100mg/mL	
Total		1L	

The induction of gene expression with IPTG (Isopropyl β -D-1-thiogalactopyranoside) was performed when the OD₆₂₀ was at ~0.5.

Optical density was measured, and samples for protein analysis were taken 0, 1 h, 2 h, 3 h and 4 h after induction. The samples were stored on ice prior to centrifugation (13000 x g during 5min. The supernatants were separated from the pellets and both of them were stored at -20 °C.

4.12 SDS-PAGE

The protein production was analyzed by sodium dodecyl sulphate polyacrylamide gel (10%) electrophoresis (SDS-Page) [21]. The gels were stained with Coomassie Brilliant Blue G250 and destained with a solution consisting in 40% methanol and 10% acetic acid.

Polyacrylamide gel electrophoresis is a method to separate protein mixtures and to determine molecular weights of proteins. The anion detergent (SDS) binds to proteins and denatures them. Due to their negative charge, protein-SDS complexes move towards the anode during electrophoresis [5].

The sample, treated previously with 20 mM Tris-buffer, sonicated during 1 minute with 60 cycles of amplification and centrifuged during 30 minutes at 4 °C and 13000 x g. The supernatant was separated from the pellet and a mix of 80 μ L of sample and 40 μ L of sample buffer were boiled

during 30 minutes and loaded into the BioRad equipment. The electrophoresis was run at a voltage of 100 V during the first 15 minutes and increased up to 200 V. The standard used was Prestained SDS-PAGE standards, broad range (figure 3).

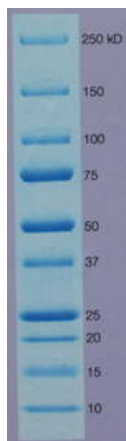


Figure 3 – Marker used for measurements of protein expression

Enzyme activity was measured by overlay agarose 1.5% gel (zymogram), containing 0.05% starch in buffer. The SDS gel was washed with 100 mM Tris-HCl plus 5 mM CaCl₂, pH 8.5, containing 2.5% Triton X-100 for 20 minutes. After this step, the gel was washed only in buffer for 20 minutes and covered with the overlay gel, and incubated for 30-45 minutes at 60°C. The agarose gel was stained in 1% Congo Red solution (and also in Lugol) and destained with 1 M NaCl.

4.13 Heat treatment

Some enzyme samples (shown subsequently in the results) were heated in a heating block at 65 °C and 70 °C at different times: 15 min, 30 min, 45 min and 60 min and then centrifuged 5 min at 13000 x g to precipitate denatured proteins. The result of the heat treatment was checked by loading the supernatants and the pellets on the SDS-Page analysis.

4.14 Batch cultivation

4.14.1 Preparation of materials and medium

Before the cultivation, the vessels for transfer and base adjustment, the parts of the NYAT medium (MgSO₄, trace elements (like Li, Rb, Cs, Ba, Sr) and glucose) that should be separately sterilized and 0.9% NaCl for dilution were autoclaved (120 °C, 20 min). The pO₂ electrode was calibrated and aseptically rinsed with 70% ethanol.

The reactor (3 L stirred tank reactor) was assembled, the pH-electrode was mounted in the reactor and medium (NYAT) was added and sterilized in the reactor (in a volume corresponding to 2L, of

the total medium volume of 2.5 L). The salts included 2.0g/L $(\text{NH}_4)_2\text{SO}_4$, 14.6g/L K_2HPO_4 , 3.2g/L $\text{NaH}_2\text{PO}_4\cdot\text{H}_2\text{O}$ and 0.5g/L $(\text{NH}_4)_2\text{H}\cdot\text{citrate}$. The reactor, the NYAT-salts and pH electrode were then autoclaved at 121°C for 45minutes.

4.14.2 Inoculum

All the sterile solutions were added into a 100 mL sterile measuring cylinder: 20 mL 5x NYAT salts stock solution (10 g/L $(\text{NH}_4)_2\text{SO}_4$, 73 g/L K_2HPO_4 , 16 g/L $\text{NaH}_2\text{PO}_4\cdot\text{H}_2\text{O}$, 2.5 g/L $(\text{NH}_4)_2\text{H}\cdot\text{citrate}$), 2 mL 50% of glucose, 0,2mL of 1M MgSO_4 , 0.2mL of trace elements, 100 μL ampicillin stock solution (100mg/mL, stored at -20°C) and sterile DI-water was added until a final volume of 100mL. In the end poured into a sterile baffled shake flask. Also 1.0mL of freeze inoculum of *E.coli* containing the chosen gene construct was added. The flask cultivation of the inoculum was started at 30°C, on a rotary-shaker, approximately 18h before the 2.5L cultivation start.

4.14.3 The cultivation

The reactor was put on the motor stand and all sensors and tubing, including temperature sensor were connected (temperature was set at 37°C, and controlled by a water flow system). Appropriate amounts of Glucose, MgSO_4 and trace elements were added to give a final composition of the medium as indicated in table 5 (but considering the final volume of 2.5 L). 50 mL of diluted NH_3 (25% NH_3 diluted in sterile DI-water 1:1) was added to the base transfer vessel in a fume hood, and connected with a tubing to the reactor and pump. In the end, the fermentor was supplied with 3 L/min of air. The pO₂ was kept over 40% by automatic adjustment of the stirring by the Phantom program. Ampicillin was added to a final concentration of 100 $\mu\text{g}/\text{mL}$ via the septum using sterile syringe and 0.1 mL of antifoam (polypropylene glycol) was also added to the reactor via the septum using a sterile syringe. During sampling, 10 mL samples were withdrawn from the sample port, first one full tube which was discarded, and then another tube with the fresh sample which was saved. Off-line measurements of OD (620 nm, using 0.9 % NaCl as a blank) were made at 30 min intervals. At the cultivation start, the pH electrode was recalibrated using only the pH 7 setpoint value. The inoculum was then added via the transfer vessel and the sampling procedure was started.

4.14.4 Sampling and sample treatment

As described above, one full tube was always discarded before the samples were taken. The samples were rested on ice, during the preparation procedure. Three analyses were made OD₆₂₀

measurement, the glucose and finally recombinant protein estimation (SDS-PAGE). The samples for glucose and protein estimation were centrifuged and supernatants and pellets stored in different eppendorf tubes. Part of the sample was used to control the external pH and doublecheck the pH electrode connected to the fermentor.

Optical density OD_{620} was measured every 30 minutes, using 1 mL cuvette and Pharmacia spectrophotometer at 620 nm and if necessary, the samples were diluted with 0.9% NaCl.

Glucose was measured every 60 min using an Accutrend-sensor blood-glucose meter and if necessary diluted with 0.9% NaCl to be within the optimum range of 10-20 mM.

4.14.5 Induction

IPTG with a molecular weight of 238.3 g/mol was prepared to result in a final concentration of 1 mM considering a total reactor volume (2500 mL sampling volume at induction time 0). At $OD_{620}=3$, filter sterilized IPTG-solution was added to the reactor, via the septum using a sterile syringe, and a sample was taken immediately thereafter.

5. Results and Discussion

5.1 Plasmid construction

First, presence of the gene of interest in construction of the plasmid pET-17b given by the Labes & Schoenheit [4], was verified by restriction enzyme digestion. The cells *E.coli* BL21 (DE3), carrying this construction (plasmid+gene) were grown on LB agar plates and after that 5 different colonies were chosen, and grown overnight in liquid LB cultures. Thereafter the plasmids were extracted; a digestion was done with the restriction enzymes (BamHI and XhoI) and run in agarose gel to prove that the gene supplied was really there. (Figure 4)

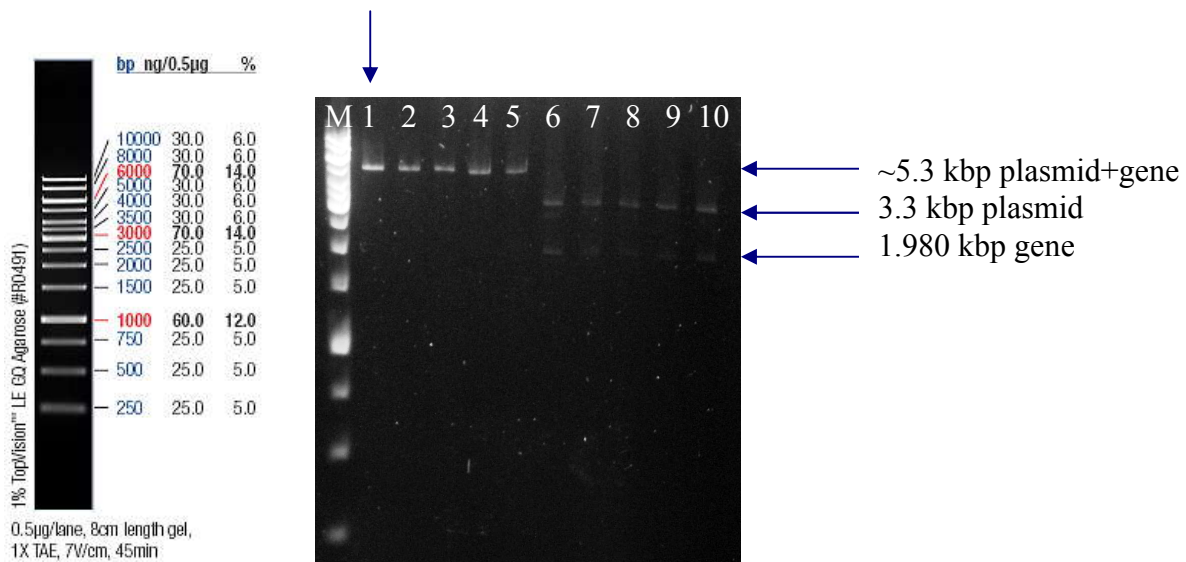


Figure 4 – Plasmid with the CGTase gene with the size 1.980 kbp from five different colonies [4]. 1 until 5 undigested plasmids and 6 to 10 double-digested plasmid with 2 separate bands of the plasmid at 3.3 kbp and the gene of interest at 1.980 kbp.

The gel showed a gene fragment of the expected size (1.980 kbp), and also the plasmid pET17b fragment (3.3 kbp), proving that the construction was right, and giving permission to continue the experiments. The purified plasmids were transformed to *E.coli* strains, BL21 (DE3) (for expression) and Nova Blue (for storage), and the best colony was recovered, shown in the gel (figure 4, marked with an arrow).

Following transformation, the clones obtained were checked for presence of the target gene using PCR amplification with gene specific primers matching the 5'- and 3' ends of the gene, to check if the gene insertion had the expected size of 1.980 kbp [4] (Figure 5).

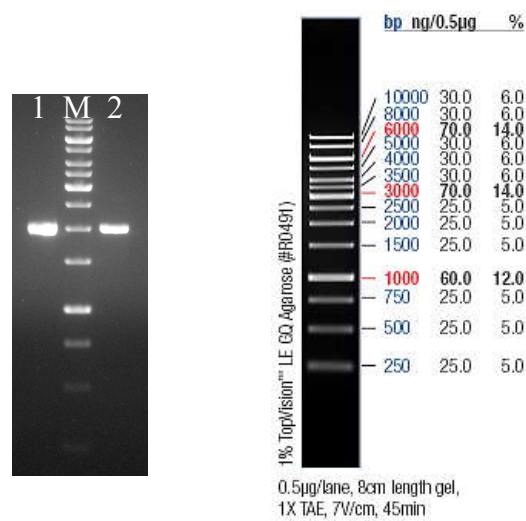


Figure 5 – CGTase gene after PCR amplification with 1.980 kbp (1 and 2).

The agarose gel (figure 5) shows that the CGTase gene was in the right size, 1,980 kbp [4] and the PCR resulted in a good concentration, because the bands were strong and well defined.

In that way, it was possible to continue the process, and analyze the expression of the protein.

5.2 Codon usage

As the gene originates from an organism (*A.fulgidus*) that is phylogenetically different from *E.coli* a comparison was done between the codons used in *E.coli* [24] and the codons in the gene encoding the CGTase. The data was exposed in graphics to have a good view of *E.coli* and CGTase codons (Figure 7). From the analysis we can conclude that the codons ATA (Ile), CTA (Leu), CCC (Pro), AGA (Arg), AGG (Arg) and GGA (Gly) are less common in *E.coli* and could lead to low expression. Nevertheless these codons are supplied in *E.coli* strain Codon Plus, or alternatively in the plasmid pRARE, and use of this strain or plasmid could be a strategy to try to improve the expression.

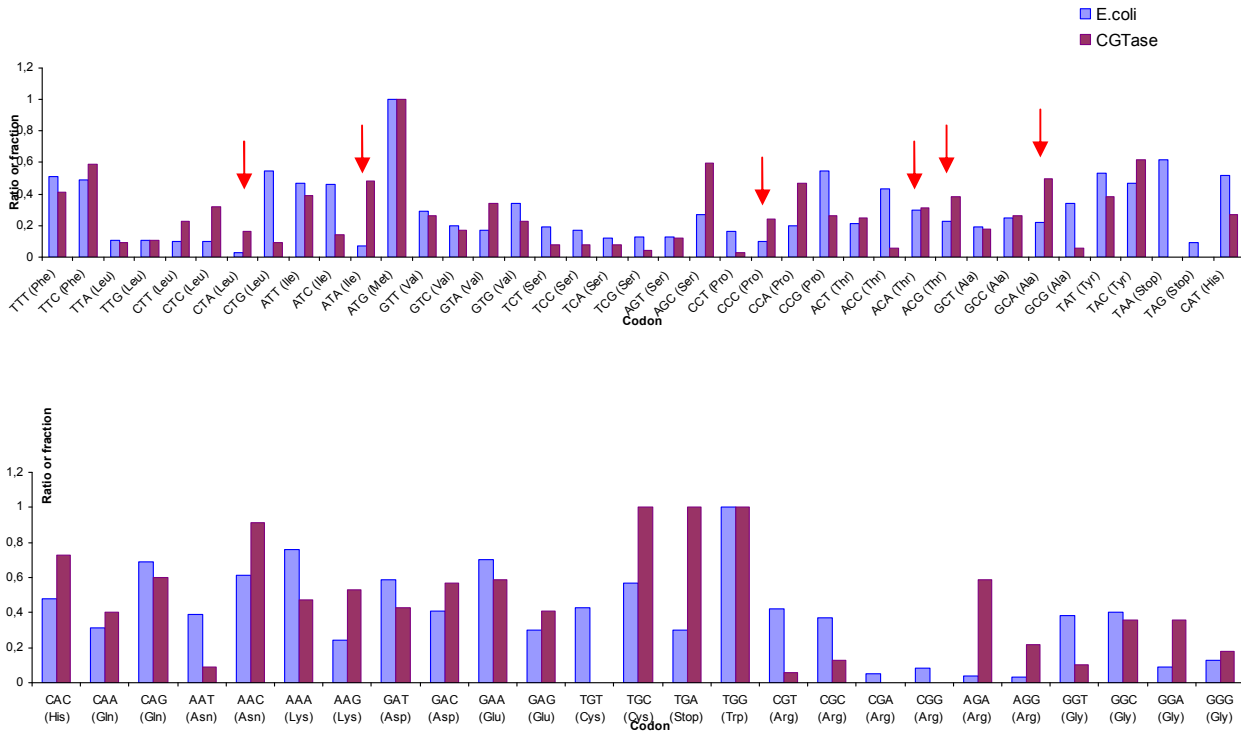


Figure 6 - Comparison between CGTase codons and *E.coli* codon (the second graph is continuation of the first one)

5.3 Protein expression

As stated in the Background (2.1), in the beginning of the experiments, the first *E.coli* strain utilized was the BL21 (DE3), but examination of the electrophoresis gels, after separation of the expressed proteins after induction in this strain, it showed that the expression was not so good. It was weak, so, therefore the gene construct was transformed to modified *E.coli* strains, to compensate for rare codons, and also to evaluate if the mRNA needed stabilization. New cultivations were performed in the two new strains, Codon Plus (supplying rare codons), already used by Sørensen *et al.* [22] and later the strain BL21 Star (DE3) used by Lopez *et al.*, [23] with the purpose of increasing the expression of the gene in question. Thereafter, to the last strain a co-plasmid (the pRARE from *E.coli* strain Rosetta- also supplying rare codons) was transformed with the intention to improve expression. All expression trials were analyzed by SDS-Page as previously mentioned.

The size of the CGTase encoded by the 1980bp-gene is ~80 kDa, consisting of 659 amino acids [4].

5.3.1 Expression in different media using the strain Codon Plus

Different media were compared, LB medium, YT medium and NYAT medium (table 5) to verify which one could be the better to express the protein. Cultivations were made in shake flasks, in which growth was followed by measuring the optical density at 620 nm, and protein production was followed using SDS-PAGE.

NYAT and YT medium were first compared for growth. On this analysis we see that the YT medium results in higher cell density than the NYAT medium (Figure 7). This was expected because the NYAT is a minimal glucose based medium, while the YT medium is a richer medium that contains both amino acids and glucose. Even so, based on the evaluation of the expression (see below) the NYAT medium was later chosen for all the remaining evaluation.

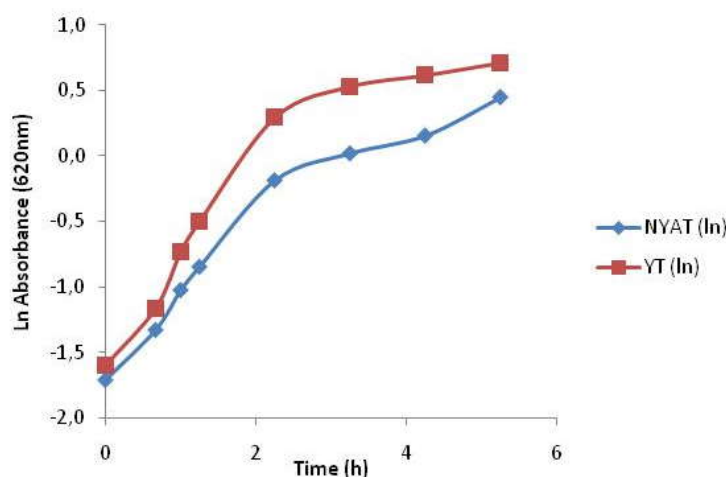


Figure 7 – Comparison between YT medium and NYAT medium

The target gene is described to be cloned without leader-sequence under control of the T7/lac promoter [4], for extracellular production in *E.coli*. First, it is good to conclude if our protein of interest is preferentially in the soluble fraction obtained after cell lysis, and if the protein expression is properly controlled by the promoter. Expression is supposed to start after induction with IPTG, and no target protein should be visible at $t=0$. The cells were induced when the OD₆₂₀ nm was approximately 0.7. Expression of the active protein is expected in the soluble fraction, but if the protein is not properly folded or too hydrophobic (causing aggregation) it could instead be in inactive form in the insoluble fraction. Nevertheless, in all experiments we could see that the protein was leaking before the induction. In figure 8 we see that the band is present at $t=0$, not so strong, but it is there, and also, the protein might aggregate and appear also in the insoluble fraction.

Comparison of the three different media, all with the same concentration of the sample, shows that there is not any big difference in expression (judged by the contrast of the target bands) which in all cases is rather low, but we see that the defined minimal NYAT medium (C) may give slightly higher expression than the complex LB and YT media (A and B).

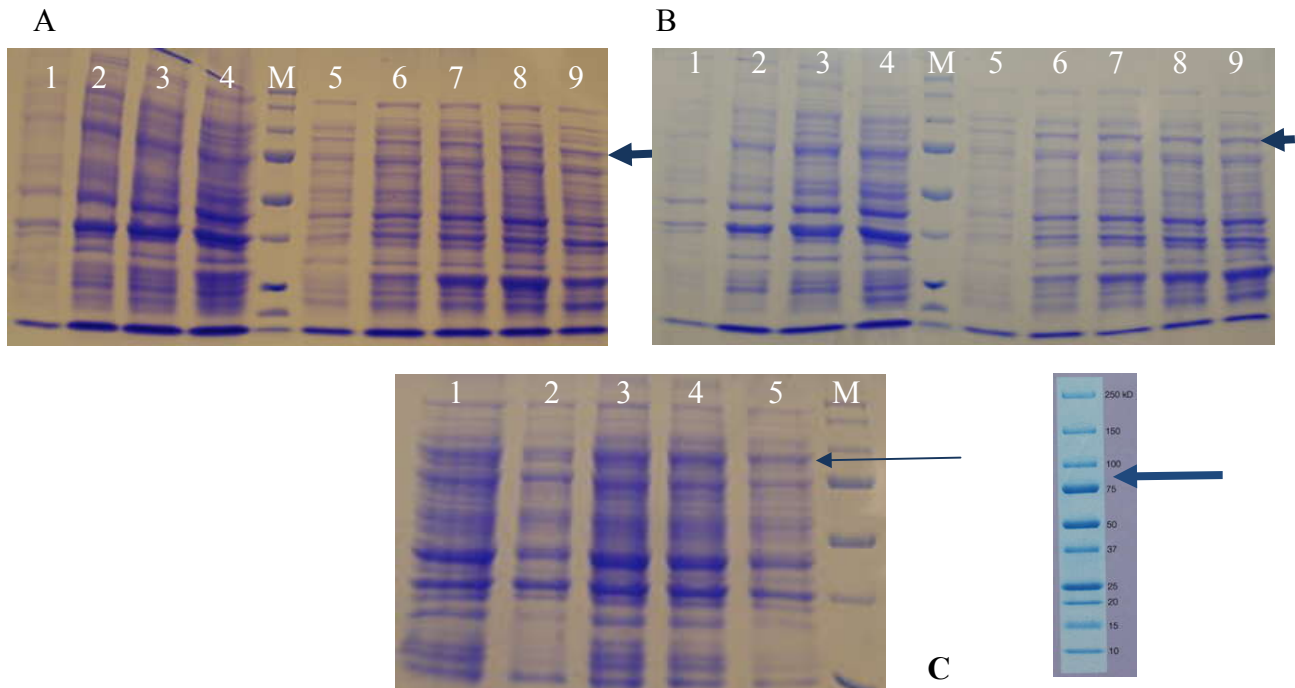


Figure 8 – The CGTase in CodonPlus is 80 kDa [4] marked by the arrow in the picture. The insoluble fraction in A1, A2, A3, A4 and B1, B2, B3 and B4, the soluble fraction A5, A6, A7, A8, A9 and B5, B6, B7, B8, B9 in LB medium (A) and YT medium (B). In the NYAT medium (C) the soluble fraction in C1, C2, C3, C4 and C5. All the soluble samples are $t=0$, $t=1$, $t=2$, $t=3$ and $t=4$ (after induction with IPTG), while only $t=1$ to $t=4$ is shown for the insoluble fraction. The composition of each medium is in table 5.

We can also conclude that the majority of the target protein appears in the soluble fraction, as expected but some also appears in the insoluble fraction, and that means that the protein has problems to fold in the cell. Another observation was that the protein expression was leaky appearing also at time zero, meaning that the protein is expressed before the induction by IPTG. This is unusual and not expected, because this protein was expected to be under the control of T7 promoter. Personal communication with A. Labes, however revealed that the German group had the same problem, and our results hence reproduced theirs.

5.3.2 Construction of a new expression strain

To try to improve the expression, the strain BL21 Star (DE3) was also used because this strain increases the stability of mRNAs [15], and this could be an answer to our problem. In addition, the same strain was supplied with the plasmid pRARE when the target gene contains codons that are rarely occurring in *E.coli* expression will increase when tRNAs for the rare codons that exists in pRARE plasmid are present. These constructions were compared with CodonPlus (which contains the same rare codon tRNAs integrated in the chromosome, but which does not contain the mRNA stabilization). As a noneexpressing control, the plasmid pET-17b without insert was used.

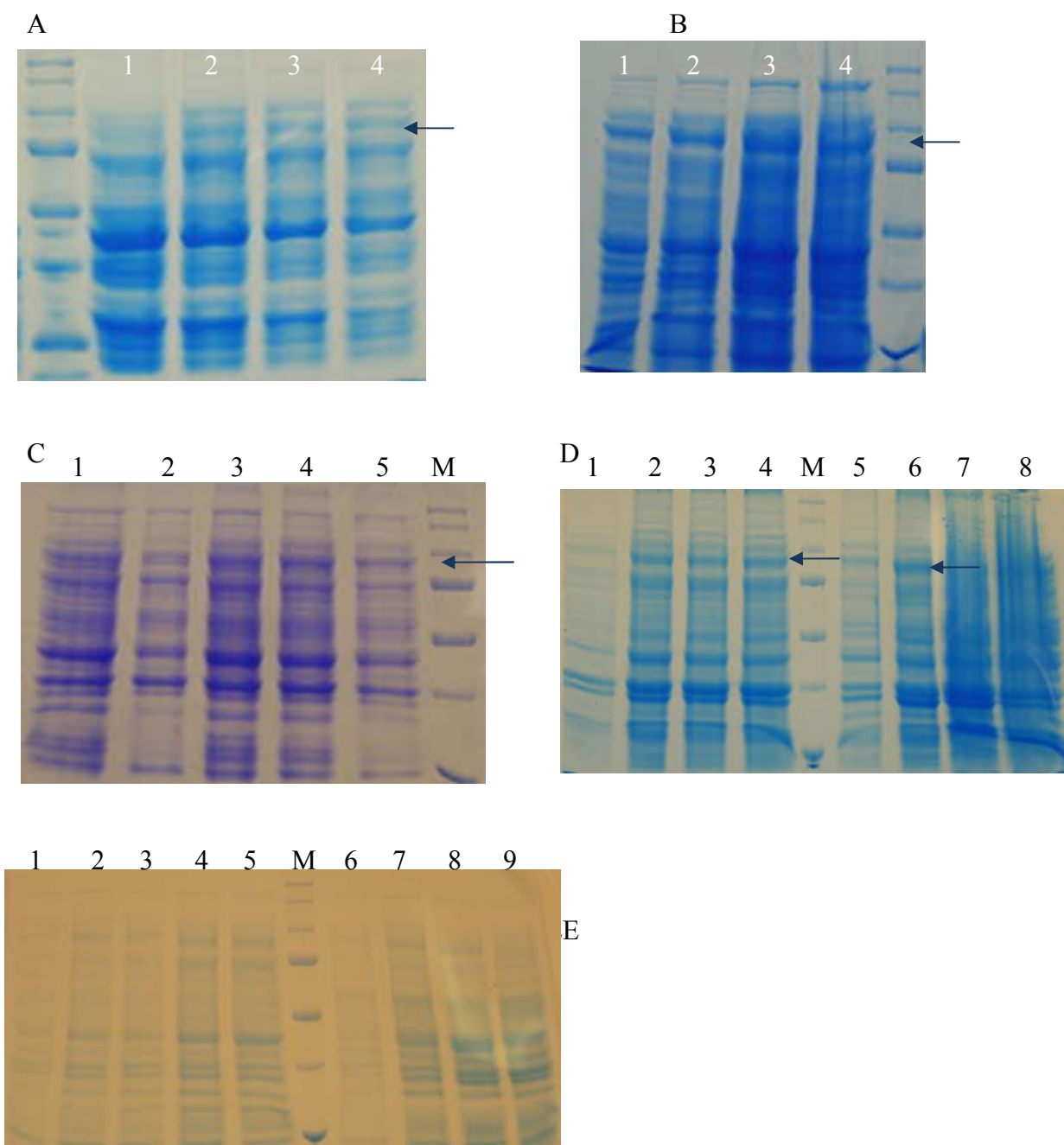


Figure 9 – Soluble fraction of CGTase in BL21 Star (DE3) (A1, A2, A3 and A4), the same fraction but in BL21 Star (DE3) supplemented with the pRARE in B1, B2, B3 and B4 and in

Codon Plus (C1, C2, C3, C4 and C5). In D1, D2, D3 and D4 the insoluble fraction of CGTase in BL21 Star (DE3) with the pRARE and after the marker the same fraction but the plasmid pET-17b and the pRARE (D5, D6, D7 and D8). The (E) display the soluble fraction of the CGTase in BL21 Star (DE3) with the pRARE before the marker (E1, E2, E3, E4 and E5) and the plasmid pET17b with pRARE after the marker (E6, E7, E8 and E9).

When comparing the expression in Codon Plus figure 9-C and the figure 9-A we can not see a significant increase of the expression but we observe that the CGTase in BL21 Star (DE3) with pRARE (Figure 9A) has better expression than without the pRARE (Figure 9B). This means that stabilization of the mRNA was not likely the problem as the BL21 Star (DE3) was supposed to fix it. However, the lack of difference between Codon Plus and BL21 Star indicated that presence of rare codons also did not efficiently improve expression. A small improvement was however seen when combining both mRNA-stabilization and presence of rare codons.

The insoluble fraction of CGTase produced in BL21 Star (DE3) with the pRARE showed that a relatively large proportion of the protein may be insoluble (Figure 9E). However, the nonexpressing control (Figure 9E, after the marker) shows that a protein with the same molecular mass as the CGTase (80 kDa) is expressed that may interfere with the evaluation. It can thus be proteins encoded by the pRARE that are appearing at positions near to or overlapping the expected position of the CGTase.

The plasmids from the two strains were isolated, to confirm the stability of the genetic constructs. Samples from the strain with 2 plasmids (construct encoding the CGTase and the pRARE, Figure 10, left) were compared with samples from the strain only harbouring the CGTase encoding construct (Figure10, right). We can see clear difference, proving that the expected plasmid composition in the respective strain is maintained.

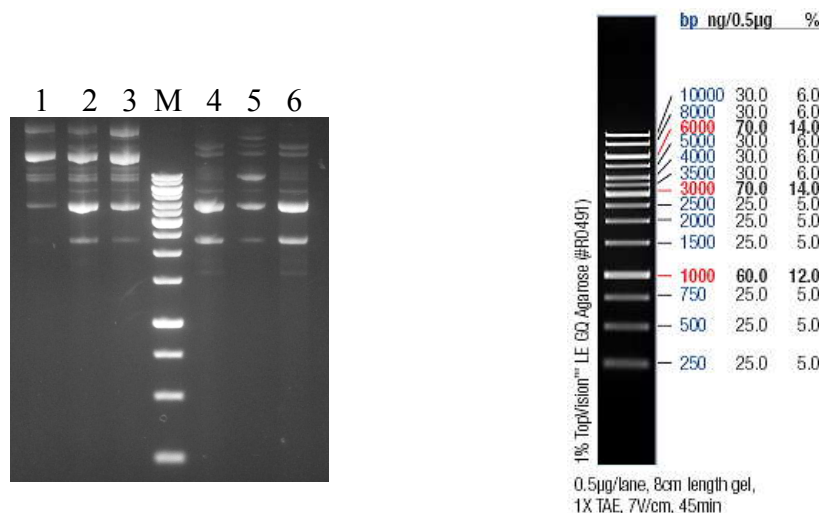


Figure 10 – CGTase gene with and without the plasmid pRARE. 1, 2, 3 with pRARE and after the marker 4, 5, 6 without pRARE.

5.3.3 Different heat treatments

The cell pellets were collected 4h after induction, and treated following the same heat-treatment that Labes and Schönheit [4] did (70 °C, 60 min). The results were not the expected because it seems that the CGTase is destroyed by the heat-treatment, showing weak bands. Subsequently different temperatures (65 °C and 70 °C) and also different times of treatment (15', 30', 45', 60') were tried. Heat-treatment was done on different cell-pellets from the host BL21 Star (DE3), grown and induced under identical conditions, but carrying different vectors: [1. pET17b-CGTase and pRARE 2. pET17b-CGTase 3. pET17b pRARE 4. pRARE] and the results are shown in figures 11 and 12.

Heat-treatment at 70°C, for 60 min (Figure 11A), resulted in bands that were really weak and almost disappearing. Nevertheless, in the t=4 of the CGTase reaction in BL21 Star (DE3) with pRARE (A4), a weak band is observed after treatment at the same temperature, with higher intensity in the culture coexpressing pRARE than without this coplasmid. That means that presence of pRARE was a relatively good strategy because it improved the expression of the gene. However, we can also see that a band appears very close to the band of interest, and when comparing with figure 9D this band also appears when the gene is not there, so interfering bands may lead to the conclusion of higher expression, but these interfering bands are removed by heat, leaving only CGTase that appears after heat-treatment.

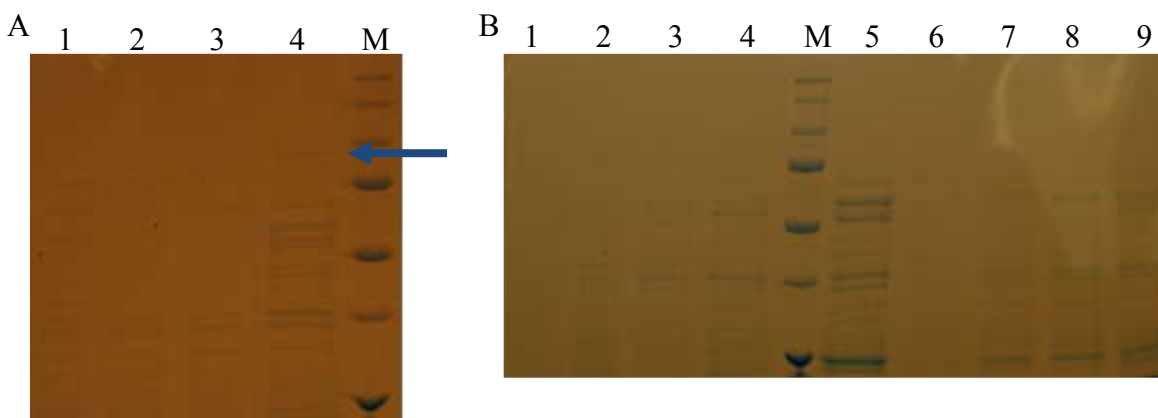


Figure 11 – In (A) CGTase produced in BL21 Star (DE3) without pRARE treated at 70 °C in A1, A2, A3. In A4 we have only the t=4 of BL21 Star (DE3) with pRARE, without production of the CGTase treated in the same temperature. In B1, B2, B3 and B4 the plasmid pET (without insert) and pRARE treated at 70 °C. After the marker only the pRARE treated at 65 °C (B5, B6, B7, B8 and B9).

As was expected, (figure 11B, lane 1-4), using the proteins from a cultivation with the empty pET-vector combined-with pRARE treated at 70 °C, there is no band in the position corresponding to the CGTase. A comparison with the BL21 Star (DE3) also with pRARE, was unfortunately inconclusive because the bands were truly weak.

Trying different times, and in the three different combination in BL21 Star (DE3) [1. pRARE, 2. pET with pRARE and 3. pET with CGTase-gene and pRARE], the results were again unexpected. In the non-treated samples, (figure 12-A) with 1. pRARE, 2. pET and pRARE and 3. pET with CGTase-gene and pRARE the band pattern is relatively similar indicating low expression of target protein, but in the first group of three (A1, A2, A3) the A3, with the CGTase, the band of interest is more visible. Even though, one interfering band (with same size as the expected CGTase) appears in the control, harbouring a vector without the CGTase (A1 and A2).

After 15 min of heat treatment, the intensity of the bands decrease and it is quite difficult to observe and compare. Nonetheless we detect the same band that we saw in the strains with the pRARE (without the CGTase) appearing in lane A4, (15 min treatment) but it is not possible to see the band corresponding at CGTase in the third column (A6) of this treatment. That means that the treatment degrades the protein, but this is contradictory to the work of Labes and Schönheit, because they could treat the same protein at 70 °C during 60 min [4]. Using the other temperature we have the same result (data not shown).

On figure 12-B, in the lane corresponding to the treated protein (B1), the band of interest appears but incredibly weak, and also other bands are visible, with molecular masses of 75 kDa, 83 kDa and in the 49 kDa size.

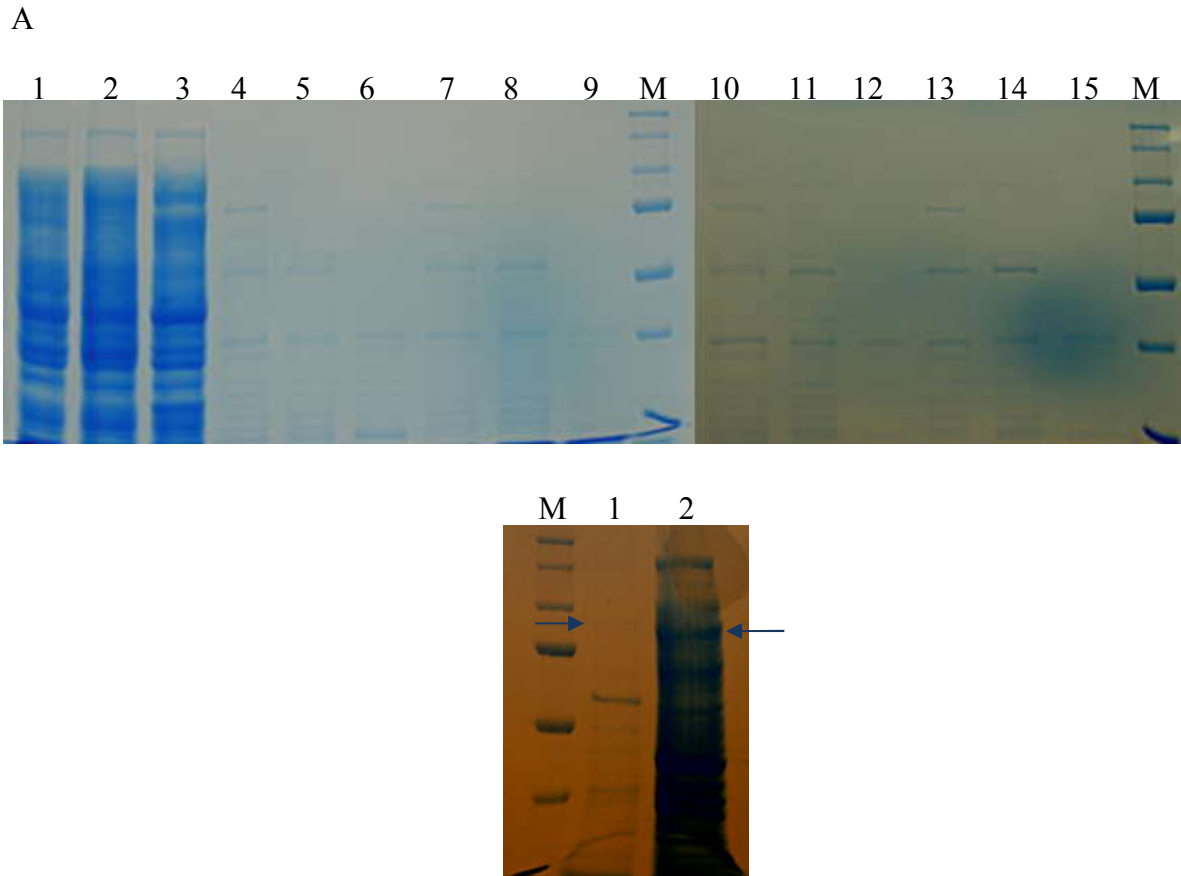


Figure 12 – In A1, A4, A7, A10, A13 the BL21 Star (DE3) with pRARE, A2, A5, A8, A11, A14 the pET with pRARE and A3, A6, A9, A12, A15 the BL21 Star (DE3) with both pRARE and CGTase. In A1, A2, A3 the samples are without treatment and in A4, A5, A6 with 15 min, A7, A8, A9 with 30 min, A10, A11, A12 with 45 min and A13, A14, A15 with 60 min of treatment. In (B) the BL21 Star (DE3) is with both pRARE and CGTase (B1) and was treated at 70 °C and B2 the same construction but without treatment.

Taken together this must mean that we still have a very low expression of our target protein, and in order to understand why and promote better expression, mutagenesis introducing silent mutations replacing rare codons was introduced.

5.4 Site-Directed mutagenesis

The next step was to mutate some of the rare codons in the gene to try to improve expression of the CGTase. In that way, mutations were introduced in the 5'-part of the gene, which has been suggested as most important for the heterologous expression level, and rare codons were replaced by commonly used in *E.coli*. It is interesting to note that rare codons (in groups of 2) are present in the beginning of the sequence (Figure 13). These locations were targeted for silent mutagenesis using the primer pairs SDM-cgtaseF and R and SDM-cgtaseF2 and R2 (Figure 13).

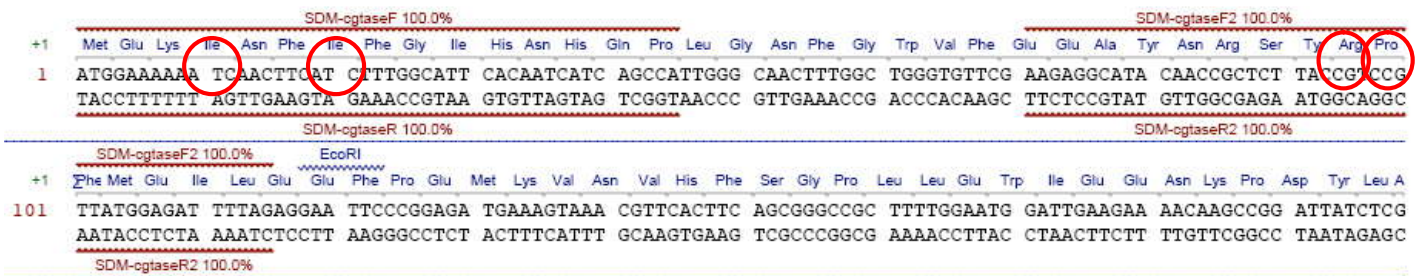


Figure 13 – CGTase 5'-sequence with primers for mutation indicated by lines above and below the sequence.

Before the site-directed mutagenesis, the CGTase-gene in BL21 Star (DE3) was again digested with the restriction enzymes BamHI and XhoI to prove presence of the vector, the insert and both, and the results are shown in Figure 14.

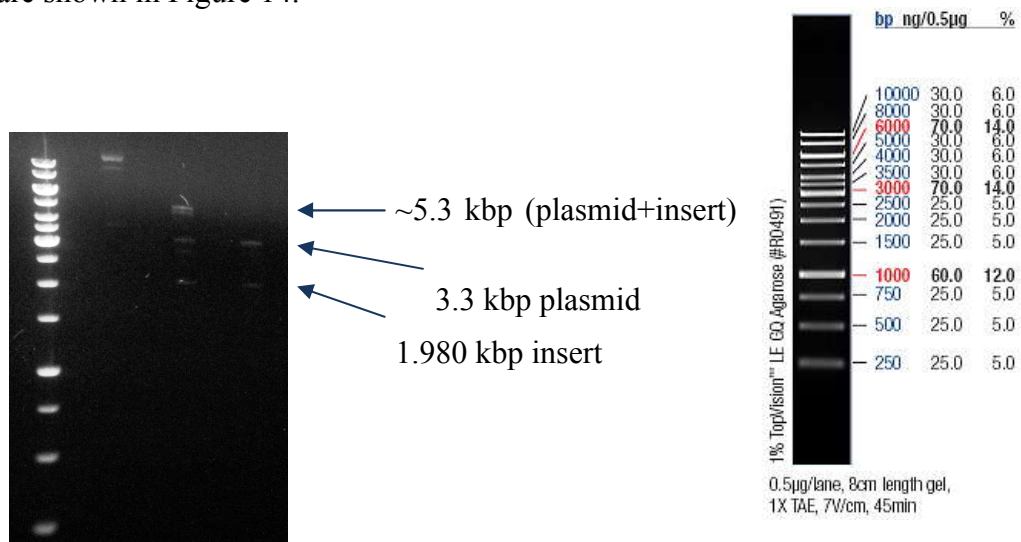


Figure 14 – Digestion of the plasmid. From left to right is the marker, the undigested vector, vector digested with BamHI and finally with BamHI and XhoI.

The insert, the plasmid and the construction positional had the appropriate size, because we have the total size of the plasmid and the insert = 5.3 kbp, the plasmid pET-17b has = 3.3 kbp the insert, with 1980 bp.

The mutagenesis reaction was made using the PCR-based method Quick change (Stratagene), using the primers shown (Figure 13) and the plasmid template was isolated from the methylating strain NovaBlue. After PCR amplification (table 3) and before the digestion with DnpI, the samples were loaded in agarose gel and run to check if amplification occurred (figure 15).

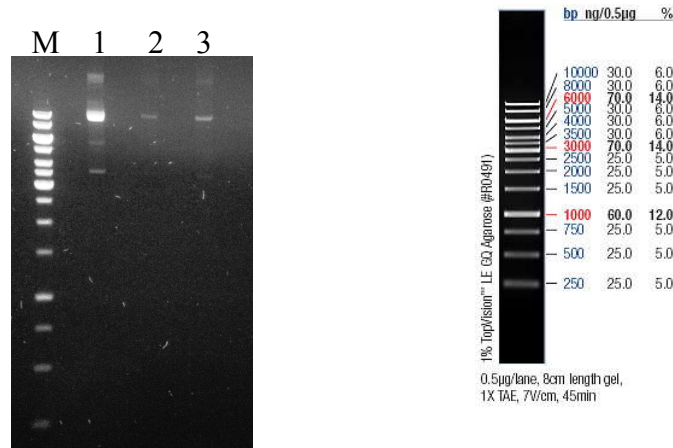


Figure 15 – In the gel picture, 1 the template (used in the amplification), 2 the first mutation, and 3 the second mutation

The 2 colonies were picked and grown overnight in LB medium with ampicillin and the plasmids were extracted and loaded in agarose gel (figure 16).

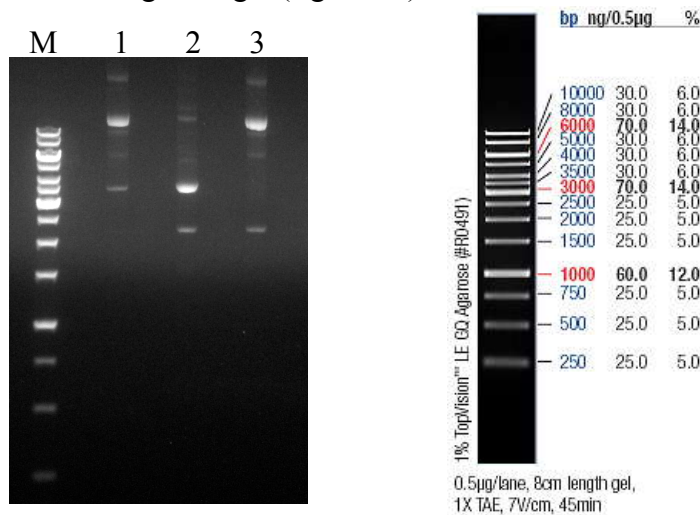


Figure 16 – Second mutation after plasmid extraction. First is the marker column, 1 the template, 2 the plasmid of the colony 1 and 3 the plasmid of the colony 2.

On figure 16, the CGTase gene is in the right size in both colonies. The clones were grown in shake flasks, induced 4h after which the cells were harvested and lysed. After all sample treatment, the samples were loaded in the SDS-Page (Figure 17):

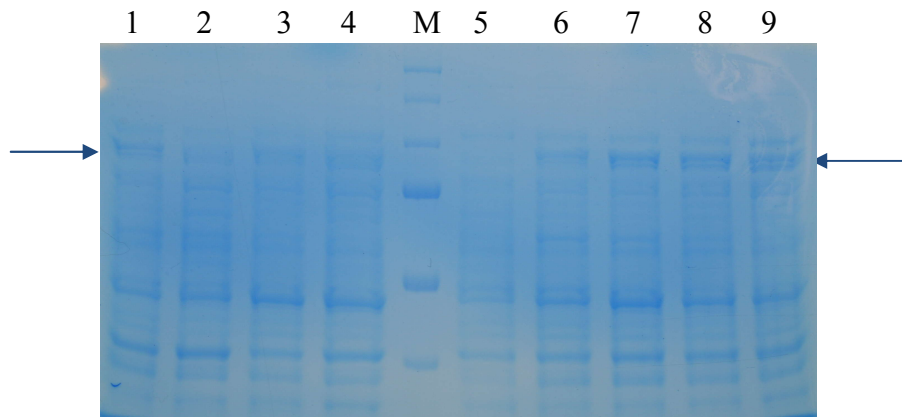


Figure 17 – The soluble fraction of the CGTase in BL21 Star (DE3). Before the marker (1, 2, 3, 4) the colony 1, after the marker (5, 6, 7, 8, 9) the colony 2.

Both colonies appear to better express a protein of the expected size, but best in clone 2. It indicated that the silent mutation was a good strategy to improve the expression of the CGTase, but should ideally also be tried the mutation in CGTase in a BL21 Star (DE3) with the pRARE, because we will have the RARE codons helping those in less percentage in *E.coli* and in CGTase.

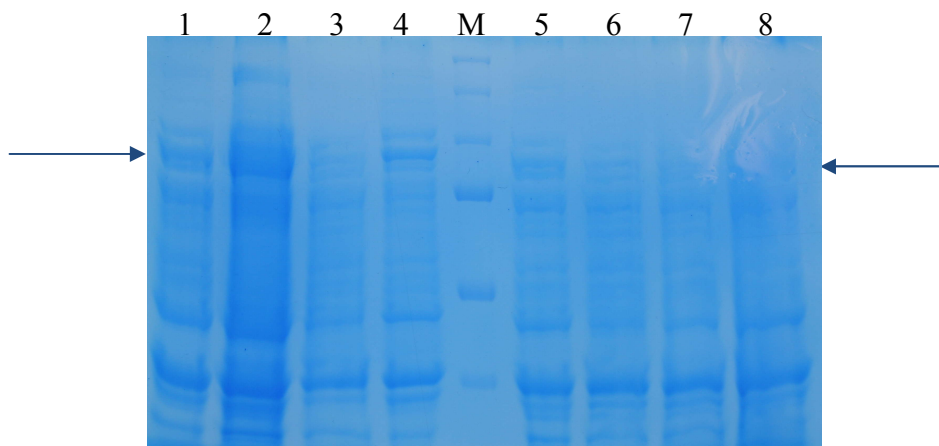


Figure 18 – The insoluble fraction of colony 1 before the marker (1, 2, 3, 4) and colony 2 after the marker (5, 6, 7, 8), in BL21 Star (DE3) of the mutant 2.

However, after mutagenesis, a higher proportion of protein was found in the insoluble fraction a protein band of the expected size was seen on gels displaying the insoluble fraction (Figure 18), and that means the protein may have problems to fold. In addition, heat treatment also lead to protein loss, (figure 19, only a very weak band was visible) and the results of the expression still remained unsolved.

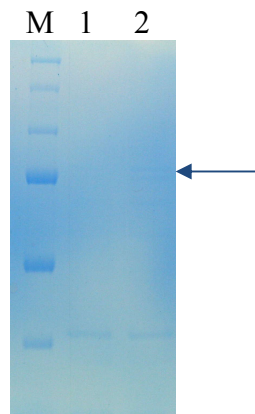


Figure 19 – Treatment at 70 °C during 30 minutes of the CGTase in BL21 Star (DE3) of the colony 1 (1) and colony 2 (2) of the mutant 2

The pET17b-CGTase encoding was then subjected to sequencing using vector specific primers (matching the T7-promoter and terminator). The results were quite strange because no sequence matched right with the sequence that we prepared previously. Use of the search engine (BLAST) [28] however confirmed presence of the CGTase encoding gene from *Archaeoglobus fulgidus* but also revealed that the sequence was not in the right direction. It was cloned in reverse. The gene was hence not under the control of the T7 promoter. This explains the bands appearing before the induction with IPTG, as well as the low expression that we had in all the experiments, showing the enzyme to be produced in very low amounts constitutively. The next step should hence be to again PCR-amplify the gene using primers containing the correct restriction sites for optimized cloning under control of the T7-promoter.

6. Conclusion

The CGTase from *Archaeoglobus fulgidus* is an unusual type of CGTase with a novel type of structure, and is hence an interesting selection for characterization. The gene however has a number of codons that are rarely used by *E.coli*, and this could result in low expression of the target protein.

In this work a strategy for production in *E.coli* was used, that involves supplying rare codons as well as stabilizing mRNA. The work also includes the construction of a new expression strain, with these two features combined.

The supplementation with the rare codons and/or the mRNA stabilization gave surprisingly low effects on the expression, but this was later explained when sequencing revealed that the construct was cloned in reverse. The currently used constructs were hence only having constitutive expression of the gene and were not affected by the T7-promoter.

Decreased expression has previously been observed when multiple consecutive rare codons are near the N-terminal of a coding sequence and mutation of at least one of these rare codons in the CGTase sequence showed a better expression even using the reversed construct, indicating that this should be promising for future studies, using an optimized construct cloned under the control of the T7 promoter.

7. Acknowledgement

First of all I want to thank God for taking care of me all the time, for His love and help in all my life.

I would like to thank Associate Professor Dr. Eva Nordberg Karlsson for giving me the opportunity to carry out this thesis project under her guidance and supervision at the Department of Biotechnology at Lund University.

I also would like to express my gratitude to the PhD student Tania Pozzo of the Department of Biotechnology at Lund University, for being my practical supervisor and for your inspiration, friendship and support during this work, not forgetting the enjoyable “fikas”, as well as to the other members of the Fermentation Group, Christina Wennerberg and Sami Khan for supporting me with everything that I needed.

Furthermore, I would like to thank Laura Saavedra from the department of biochemistry to the good times and for the availability to helping me in my work, also to Javier Pastén, Héctor Suarez, Thuoc Van, Victor Ibarhim, Laura Fernandes, Marlene Gaitán and Thuy Thi from the department of biotechnology for the everyday laboratory life.

Moreover I would like to thank Professor Doutor Raúl Barros from the Algarve University for being my assisting supervisor and for all inspiration during my entire course.

In addition, special thanks goes to Heber and Therese Teixeira, Jansen Pinto, Daniel Olsson and all the family that I won here in Sweden, to all the good moments, friendship, lunches and good arrangement, great trips, and helping me with everything that I needed were the everyday life would not have been so lovely.

Especially, I would like to give my special thanks to Evandro whose patient love enabled me in all my course and also to complete this work.

Finally, I want to thank my mother, brother, grandmother and friends for always believing in me and encouraging me to experience new things, for support and motivation. And I want to use the opportunity to dedicate my work to my mother and in memory of my father, since they always believed in me, and they struggled throughout my life for my well being.

Without any of you, the experience of writing my master thesis here in Lund would not have been so good. Thank you all!!

Catarina

8. References

- [1] Beeder, J., Nilsen R., Rosnes, J., Torsvik T. and Lien, T., (1994) “*Archaeoglobus fulgidus* Isolated from Hot North Sea Oil Field Waters”, *Appl. Environ. Microbiol.* 60, 1227-1231.
- [2] DeLong, Edward F., (2003), *ASM News / Volume 69, Number 10, Pages 503-511.*
- [3] <http://www.dac.neu.edu/biology/k.bergman/WebsiteBarney/Sema/sema.htm> (accessed on 27 July 2008).
- [4] Antje Labes and Peter Schönheit, (2007) “*Unusual Starch Degradation Pathway via Cyclodextrins in the Hyperthermophilic Sulfate-Reducing Archaeon Archaeoglobus fulgidus Strain 7324*”, *J. Bacteriol.* 189, 8901-8913.
- [5] Kroll, K., (2007), “*Characterization of thermostable β -glucosidase TnBglB and evaluation of possibilities in transglycosylation reactions*”, Degree project, Department of Biotechnology, Lund University, Sweden.
- [6] Turner, P., “*Exploring thermostable glycoside hydrolases – Amylases and β -glucosidases*”, (2007), Department of Biotechnology, Lund University, Sweden.
- [7] MacGregor E. A., Janeček Š., Svensson B., (2001), “*Relationship of sequence and structure to specificity in the α -amylase family of enzymes*”, *Biochim, Biophys. Acta – Prot Struct. Mol. Enzymol.* 1546, 1-20.
- [8] Imamura, H., Fushinobu, S., Yamamoto, M., Kumasaka, T., Jeon, B.S., Wakagi, T., Matsuzawa, H. (2003) “*Crystal structures of 4- α -glucanotransferase from *Thermococcus litoralis* and its complex with an inhibitor*”. *J.BIOL.CHEM.* 278, 19378-86.
- [9] Qi Q.S., Zimmermann W., (2005) “*Cyclodextrin glucanotransferase: from gene to application*”, *Appl. Microbiol. Biotechnol.* 66, 475-485.
- [10] <http://www.lsbu.ac.uk/water/cyclodextrin.html> (accessed on 7 August 2008).
- [11] <http://www.emdbiosciences.com/SharedImages/Novagen/HostStrains.html> (accessed on 3 August 2008).
- [12] Donna Daniels, Novagen Technical Service Scientist, (2006), “*Answers to questions about Selecting E. coli cloning hosts*”, Novagen, in *Innovations Volume 24*, Pages 14 and 15.

- [13] Novy N., Drott D., Yaeger K., Mierendorf R., (2001), “*Overcoming the codon bias of E.coli for enhanced protein expression*”, Newsletter of Novagen Inc., Volume 1, Number 12, Pages 1-3.
- [14] Carstens C.P, Bonnardel J., Allen R., Waesche A., (2001) “*High-level expression of heterologous proteins in E. coli - BL21-CodonPlus™ Cells Correct Expression Problems Caused by Codon Bias*”, Strategies, Vol. 14, Pages 50 to 52, Stratagene, (accessed on 3 August 2008).
- [15] Anonymous, (2002), “*BL21 Star (DE3) one Shot, BL21 Star (DE3)pLysS One Shot, Chemically Competent Cells*”, Invitrogen life technologies (Catalogue nos. C6010-03, C6020-03, Version D)
- [16] Anonymous, (2004), “*Competent Cells, What a difference a strain makes*”, Novagen, Competent cells, Merck Biosciences, Pages 1-12, (<http://chem-faculty.ucsd.edu/joseph/Protocols/Competent%20cells.pdf>) (accessed on 3 August 2008).
- [17] Mierendorf R., Yeager K., Novy R., (1994) “*The pET System: Your Choice for expression*”, Newsletter of Novagen Inc, Volume1, (Number 1), pages 1-3.
- [18] John Altman, (5/13/2002), “*Autoinduction of Expression in the T7 Expression System*”, Altman Laboratory at the Emory Vaccine Center, Protocols for cellular immunology, http://www.microbiology.emory.edu/altman/f_protocols/f_tetramers/autoind_annot.html, (accessed on 3 August 2008).
- [19] Sambrook, J., Fritsch, E. F. and Maniatis, T.: *In Ford, N. (ed.), Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
- [20] Tómasdóttir V., (2005), “*Genetic and biochemical characterization of the novel short-chain alcohol dehydrogenase ADH101 from a thermophilic Bacillus strain from Iceland*”, Degree project, Department of Biotechnology, Lund University, Sweden.
- [21] Laemmli, U.K., (1970), “*Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4*”, Nature 227, 680-685.
- [22] Sørensen, H.P., Mortensen, K.K., (2005) “*Advanced genetic strategies for recombinant protein expression in Escherichia coli*”, J. Biotechnol, 115, 113-128.
- [23] Lopez, P.J., Marchand, I. Joyce, S.A., Dreyfus, M., 1999. “*The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo*”. Mol. Microbiol. 33, 188-199.

- [24] Modified from Maloy, S., V. Stewart, and R. Taylor. 1996. “*Genetic analysis of pathogenic bacteria.*” Cold Spring Harbor Laboratory Press, NY.
- [25] Mateus B., (2004) “*Expression and production of a thermostable xylanase from Rhodothermus marinus in the methylotrophic yeast Pichia pastoris*”, Degree project, Department of Biotechnology, Lund University, Sweden,.
- [26] G. Brent Irvine, (2002) “*Molecular weight estimation for native proteins using high-performance size-exclusion chromatography*”, The Protein Protocols Handbook, 2nd Edition, J.M.Walker, Humana Press Inc., Totowa, NJ, chapter 84.
- [27] GE Healthcare, “*Gel filtration, Principles and methods*”, Handbooks from GE Healthcare, GE Healthcare Bio-Sciences AB, 18-1022-18 AJ 01/2007, Sweden.
(www.gelifesciences.com/protein-purification)
- [28] <http://blast.ncbi.nlm.nih.gov/Blast.cgi> (accessed on 18 August 2008).