University of South Bohemia

Faculty of Science

Overexpression of a tick salivary cysteine protease inhibitor in prokaryotic expression system

---

Bachelor Thesis in Biological Chemistry

Lovelyna Eromonsele

České Budějovice, 2015

Supervisor: Michail Kotsyfakis PhD.

Co. Supervisor: Mgr. Jan Kotál

Institute of Parasitology, Biology Centre CAS

Eromonsele L., 2015: Overexpression of a tick salivary cysteine protease inhibitor in prokaryotic expression system. BSc. thesis in English -29 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

This thesis focuses on the overexpression of a gene encoding a tick salivary protease inhibitor in a prokaryotic system (bacteria). It includes studies on the folding of the protein to its native state.

I hereby declare that I have worked on my Bachelor thesis independently and used only the sources listed in the bibliography. I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree to the publication of my Bachelor thesis, in full to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages. Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defence in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

…… ……..

Date, Place Signature

Acknowledgement

I would like to express my sincere thanks to my supervisor, Dr. Michalis Kotsyfakis for giving me the opportunity to work on a very interesting topic. It has been an honour and pleasure to work under your supervision.

Honest thanks to Mgr. Jan Kotál for his support, guidance and for giving me the motivation to work on my topic right from the onset and throughout the whole time of this project. I would also like to thank him for his patience, encouragement and willingness to help many times I needed help.

A very big thank you to all the members of '**Laboratory of Genomics and Proteomics of Disease Vectors**' for all their advice and support during my stay in the lab.

I am most grateful to my parents for their endless support and encouragement. Many thanks to my siblings, especially my twin sister for her endless support all through the time I spent in Budweis.

I am very grateful to my Husband, Osarobo Luis for his encouragement and support. No matter how far apart we were and no matter how tough things were, I knew he was just a phone call away to make me smile and give me the right advice.

Goals of the Work

1. To transform bacteria and overexpress the gene in them (pilot expression) to produce the tick protein cystatin G10.
2. To make a tick gene called cystatin G10, to be overexpressed in bacteria, thus producing high amount of a tick protein in bacteria.
3. Finally to refold the protein to its native state.

Glossary of Terms and Abbreviations

BCA	Bicinchoninic Acid
BSA	Bovine Serum Albumin
CatL	Cathepsin L
DNA	Deoxy Ribonucleic Acids
DTT	dithiotreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
IB	Inclusion Bodies
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
kDa	kilo Dalton
LB medium	lysogeny Broth Medium
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
RPM	Rotations Per Minute
RT	Room Temperature
SDS	Sodium dodecyl sulfate
SE	Standard Errors
TBS	Tris-Buffered Saline

Contents

1. Introduction

1.1. Ticks

Ticks are, after mosquitoes, the second most common vectors for disease transmission in most mammals, birds, and reptiles as well. Consequently, tick-borne diseases are very frequent. These animals feed by sucking the human blood, so they affect human health both directly through biting, stinging and the infestation of tissues, and indirectly, by transmitting diseases. Ticks are obligated blood-sucking insects with more than 800 species inhabiting the planet [1].

Ticks are classified in the phylum Arthropoda, class Arachnida, subclass Acarina, order Parasitiformes and suborder Ixodida. From the viewpoint of disease transmission to humans, the essential feature of ticks is their need to ingest a blood meal to transform to their next stage of development. Tick feeds by perching on any available host and insert its hypostome into the host's skin. Even though the host tries to defend against the ectoparasite, some tick species can ingest about 15 ml of blood while feeding [1].

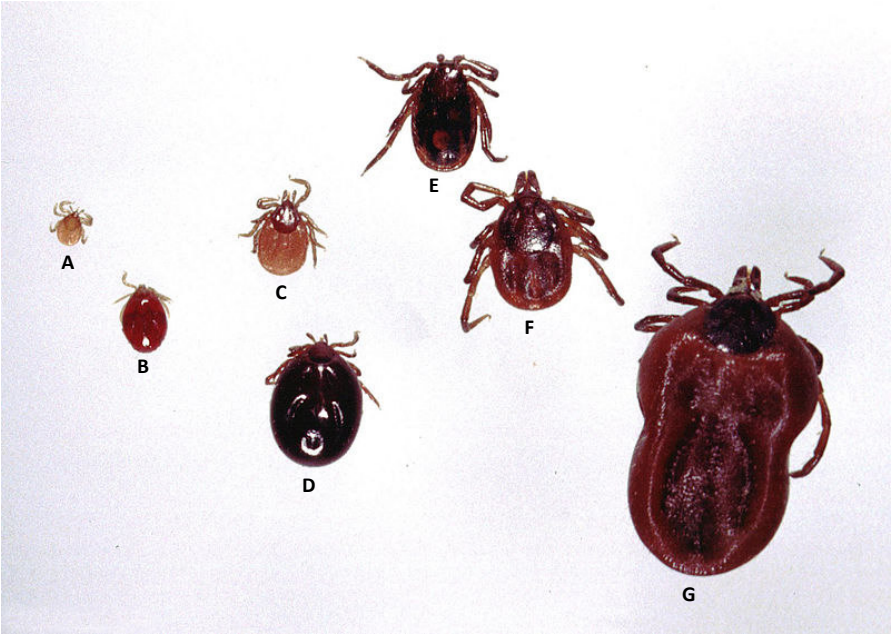

***Figure 1**:* Life cycle stages of *Ixodes ricinus* deer ticks. Starting from the left side, A - unfed larva; B - engorged larva; C - unfed nymph; D - engorged nymph; E - unfed male; F – unfed female; G - partially engorged female [2].

Tick borne diseases are everywhere and they may be similar or quite different between regions. Tick borne diseases commonly found in North America are Lyme disease, Human granulocytic

and monocytic ehrlichiosis, Babesiosis, Relapsing fever, Rocky Mountain spotted fever, Colorado tick fever, Tularemia, Q fever and Tick paralysis. The same diseases are seen in Europe as well with the addition of Boutonneuse fever and Tick-borne encephalitis. The most common tick borne disease in Africa, Spain, Saudi Arabia, Asia in and certain areas of Canada is the relapsing fever [3].

1.2. Tick saliva and pathogen transmission

The salivary glands of ticks contain active components which are able to render the haemostatic system of the host useless so that the tick is able to alter the host inflammatory and immune responses. Thus ticks are able to adapt firmly and are capable of transmitting a significant variety of pathogens, including bacteria, protozoa, viruses and nematodes. These pathogens invade the tick salivary glands and are transmitted to the host when the tick feeds on their blood. With a single bite, ticks can transmit multiple pathogens; a phenomenon that has resulted to some classic tick-borne diseases [3].

The tick salivary components can modulate the immune and haemostatic response of the host and as a result, the pathogens in the saliva are very infective. Other components in tick saliva have anti-haemostatic, anti-inflammatory and immunomodulatory properties that help ticks to feed on their host blood. It has also been observed that when ticks feed, new mRNAs are induced in the salivary glands which results in the synthesis of a variety of proteins which also play a key role for the completion of the tick feeding process and for the transmission of pathogens [4].

Due to the significance of the tick saliva in haemostasis, inflammation, immunity and pathogen transmission, the isolation of tick salivary components has become of great interest to researchers. The isolation of these molecules can help us to understand the role of saliva in blood feeding and pathogen transmission which paves the way for potential vaccine development to control tick-borne diseases. Characterization of a large number of genes from tick salivary glands could provide a better knowledge on how vaccines can be designed based on tick salivary antigens to control ticks and tick-borne diseases [4].

This thesis focuses on the overexpression of the gene encoding a tick salivary protease inhibitor and the subsequent refolding of the corresponding protein.

1.3. Cysteine proteases and one of their families - cathepsins

Cysteine proteases are proteolytic enzymes which are capable of catalysing the hydrolytic cleavage of the peptide bonds. They can be grouped as exopeptidases or endopeptidases. They are found in both, plants and animals, including vertebrates, arthropods, viruses and prokaryotes. Cathepsins are members of the lysosomal cysteine proteases and their role is in protein breakdown in lysosomes. They also participate in antigen presentation and intracellular protein outcome and as well take part in the proteolytic processing of proenzymes and prohomones, fertilization, cell proliferation and differentiation. Generally, the major role of cysteine proteases in the biology of parasitic organisms is catabolism. However imbalanced activity of these enzymes may result in diseases like rheumatoid arthritis, multiple sclerosis, neurological disorders, tumours and osteoporosis [5].

Cysteine proteases also play major roles in many physiological processes and their hyperactivity may result in several diseases. Nature has developed many strategies to protect organisms and the cells from unwanted proteolysis and one of these strategies is the control of proteolytic activities by inhibition [6]. Therefore, precise control of proteolytic processes is important for proper functioning of cells and the organisms as a whole. This can be achieved in various ways, from the regulation of protease expression, through specific degradation of mature enzymes to block their activity. Because of the possibilities of applications of selective proteinase inhibitors in therapy, the mechanisms responsible for inhibition are being thoroughly investigated by many researchers. Tick saliva contains cysteine protease inhibitors and thus disrupts the balance of the enzymes in the sites of tick feeding [7].

1.4. Protein refolding and aggregation

The overexpression of a recombinant protein in a prokaryotic system (bacteria) can lead to the aggregation of insoluble proteins within inclusion bodies. Apart from the problem of purification of these inclusion bodies from the bacteria, the bigger task is to solubilize the inclusion bodies and refold the protein into its native structure to regain full biological activity. The precise conditions to obtain efficient refolding differ for each protein. Chaotropic agents like urea or guanidine are suitable agents for solubilizing and consequent refolding of the inclusion bodies. Also, reducing agents like DTT (dithiotreitol) are often used for the reduction of the disulphide bonds between the cysteines in the process of refolding [8].

Aggregation can lead to decreased yield during the protein refolding process. Aggregation of unwanted species can result from the folding of intermediates with hydrophobic patches which become exposed to an aqueous solvent. In addition, the ionic strength, pH, oxidation state, temperature and protein concentrations as well as hydrophobic, polar and chaotropic agents are all parameters which can result to undesirable aggregation [9].

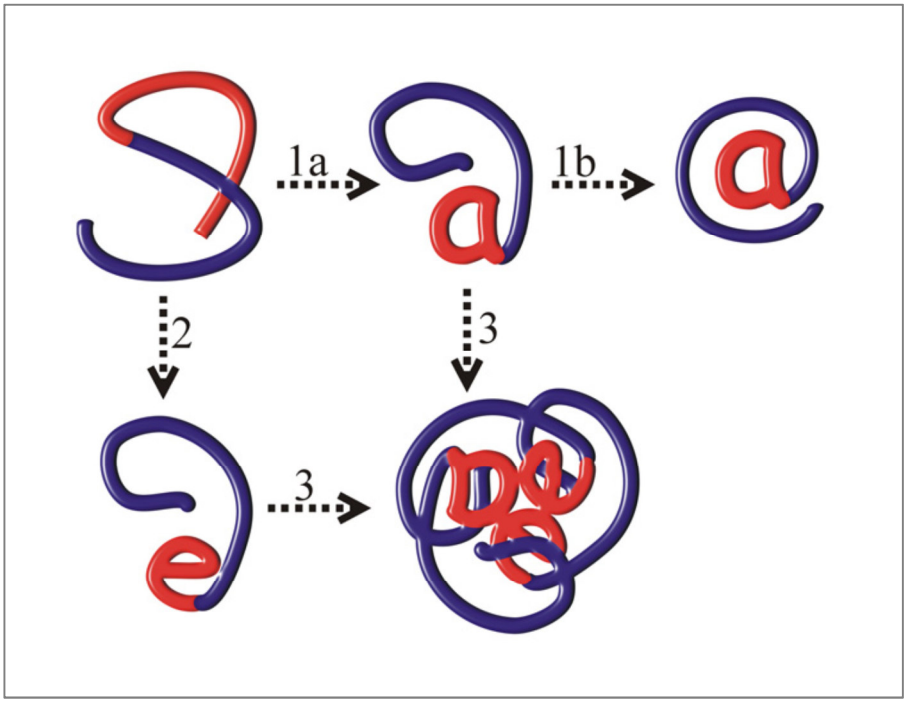

Figure 2:

Simplified version of correctly folded versus misfolded protein and the possible aggregation. The different parts in the figure are (1) correct protein folding pathway. (2) possible misfolding competition. (3) aggregation. The blue lines represent the hydrophilic solvent-exposed parts of the protein while the red lines represent the hydrophobic patches [9].

For a protein to be successfully folded, the accumulation of unwanted aggregates has to be prevented. Taking into account the possibility of aggregation, a continuous buffer system, refolding at low temperature and slow addition of denatured protein to refolding buffer etc., have to be considered in order to minimize misfolding and aggregation [8 – 10].

2. Materials

Table 1: Cells and cell culture reagents

Bacterial cultivation, protein expression, purification and refolding reagents and composition	
Competent cells	*Escherichia.coli* BL21(DE3)pLysS (Invitrogen)
LB agar	1.5% agar in LB medium
LB medium	LB Broth Miller (Amresco)
1000× Ampicillin	100 mg/mL Amp. In water (Sigma)
1000× chloramphenicol	35 mg/ml Chl. in water (Sigma)
1000× IPTG	1 M Isopropyl β-D-1-thiogalactopyranoside in water, (Invitrogen)
JET Quick-Plasmid Miniprep	Isolation of plasmid from bacteria (Genomed)

Table 2: Buffers, media and composition

Resuspending buffer with Triton	20 mM Tris-HCl, pH 8, 1% Triton-X 100 (Sigma)
Resuspending buffer	20 mM Tris-HCl, pH 8
Solubilizing buffer	6 M guanidine hydrochloride, 20 mM Tris, pH 8
Reducing agent	100× DTT (Dithiotreitol) 1 M, (Fermentas)
Refolding buffer	50 mM Tris-HCl, 300 mM NaCl, pH 7.4
TBS	20mM Tris, 150mM NaCl, pH 8.0 as the
Protein concentration estimation	Pierce® BCA Protein Assay kit (Thermo Scientific)
Aggregates estimation	ProteoStat® Protein aggregation assay kit (Enzo life Sciences)

Table 3 a: First set of 10× concentrated refolding buffers and composition

Refolding Buffers for Inclusion Bodies Refolding	
Buffer 1	3M NaCl, 0.2M NaAcetate, pH 5.5
Buffer 2	0.2M Tris, 3M NaCl, pH 8.5
Buffer 3	0.2M Tris, 3M NaCl, pH 6.8
Buffer 4	0.2M Tris, 3M NaCl, pH 8.0
Buffer 5	0.2M Tris, 0.1M NaCl, 5mM KCl, 10mM EDTA, pH 8.0
Buffer 6	0.2M Tris, 0.1M NaCl, 5mM KCl, 20mM $MgCl_2$, 20mM $CaCl_2$, pH 8.0
Buffer 7	0.2M Tris, 2.4M NaCl, 100mM KCl, 20mM $MgCl_2$, 20mM $CaCl_2$, pH 8.0
Buffer 8	0.2M Tris, 2.4M NaCl, 100mM KCl, 10mM EDTA, pH 8.0

Table 3 b: Second set of refolding buffers (1×)

Buffer 8 for protein refolding optimization with different NaCl concentration	
Buffer 8 /$_{20\ mM}$	20 mM Tris, 20 mM NaCl, 10 mM KCl, 1 mM EDTA, pH 8.0
Buffer 8 /$_{50mM}$	20 mM Tris, 50 mM NaCl, 10 mM KCl, 1 mM EDTA, pH 8.0
Buffer 8 /$_{100mM}$	20 mM Tris, 100 mM NaCl, 10 mM KCl, 1 mM EDTA, pH 8.0
Buffer 8 /$_{150mM}$	20 mM Tris, 150 mM NaCl, 10 mM KCl, 1 mM EDTA, pH 8.0
Buffer 8 /$_{240\ mM}$	20 mM Tris, 240 mM NaCl, 10 mM KCl, 1 mM EDTA, pH 8.0
Buffer 8 /$_{300\ mM}$	20 mM Tris, 300 mM NaCl, 10 mM KCl, 1 mM EDTA, pH 8.0

Table 4: SDS-PAGE

Sample buffer	4× NuPAGE LDS Sample buffer (Invitrogen)
Running buffer	1× NuPAGE Bis-Tris Running buffer (Invitrogen)
Reducing agent	10× NuPAGE Sample Reducing agent (Invitrogen)
Molecular Marker	Sea Blue Plus 2 Prestained Standard (Invitrogen)
Gel for SDS-PAGE	Nu-PAGE Bis-Tris gel, 4-12% (Invitrogen)
Staining solution	0.05% Coomassie Brilliant Blue R-250, 25% methanol, 10% acetic acid in distilled water
Destaining Solution	25% methanol and 10% acetic acid and dist. Water

Table 5: Cathepsin L activity measurement

Fluorogenic peptide	Z-L-R-AMC, Cat.N. ES008, (R&D systems)
Enzyme	Cathepsin L, Cat.N. 219382, activity 3000mU/ml, (Calibiochem)
Assay buffer	100 mM NaAc, pH 5.5, 100 mM NaCl, 1 mM EDTA, 1 mg/ml L-Cysteine (fresh), 0.01 % Triton ×-100

3. General and Experimental Methods

3.1. Transformation

Transformation is a process of a genetic transfer by which a plasmid DNA is incorporated into a recipient cell which causes a genetic change. A lot of prokaryotes are usually transformable, including some species of both gram-negative and gram-positive bacteria and also certain species of Archaea. In a prokaryotic cell, the DNA is usually present as a large single molecule, and can be obtained with a gentle lysis of the cell. The natural ability of plasmids to transfer genes between bacteria is the basis of transformation.

Competence in many naturally transformable bacteria is regulated and special proteins participate in the uptake and processing of DNA. These competence-specific proteins include a membrane-associated DNA-binding protein, a cell wall autolysin, and various nucleases. During natural transformation, competent bacteria reversibly bind DNA. Though afterwards, the binding becomes irreversible. Normally competent cells bind much more DNA than non-competent cells; i.e. as much as 1000 times or even more. Also the size of the transforming fragments is much smaller than that of the whole genome, and the fragments are further degraded during the uptake process [11].

BL21-(DE3)pLysS cells (*Escherichia coli* competent bacteria cells) were used as acceptors of plasmids. In 50 µl of competent bacteria cells, 2 µl of a commercially synthesized plasmid (200 ng) was added and incubated on ice for 30 minutes and then heat shocked by heating to 42 °C for exactly 1 minute. After other 2 minutes on ice, 125 µl of SOC medium was added and the tube was shaken for 1 hour at 37°C. The cells were then incubated overnight on an agar plate with antibiotics. One of the bacterial colonies that appeared on the plate was then picked and used for further experiments.

3.2. Protein Overexpression

One of the basic strategies used to produce a protein in large amount is the gene overexpression in *E. coli.* The technique is very simple to handle, cheap and less time consuming. To start the gene over-expression, the gene of interest must be cloned in an expression vector (plasmid). Another advantage of using *E. coli* expression system is that recombinant proteins can also be produced easily at high density and very rapidly in nutrient-rich media [12, 13].

For the pilot expression in a small scale, 50 ml of LB medium was used together with 50 µl of ampicillin and 50 µl of chloramphenicol and 1 ml of the cells cultured overnight in the same LB medium. The initial OD was about 0.1. The cells were then incubated at 37°C for about 2 to 3 hours until the OD of the bacterial culture reached approximately 0.7 and the expression was started by adding 50 µl IPTG. 1 ml of the cell culture was collected every hour up to the 8^{th} hour and also 1 ml was collected at the 24^{th} hour. All samples were centrifuged right after collection at 10,000xg for 10 minutes and the pellets were kept at -20°C.

The pellets were resuspended in 0.5 ml of resuspending buffer and then the cells were heat shock broken by placing in the heating block at 56°C and after 5 minutes, transferred and frozen in liquid nitrogen for 5 minutes. Heating/freezing was repeated in four cycles. Afterwards, the tubes were centrifuged at 10,000xg for 10 minutes and the pellets and supernatants were analysed on a gel using electrophoresis. With the gel analysis, the best time point was then noted and used for large scale overexpression experiments.

Protein overexpression was performed in 8 l of LB medium in eight flasks and each flask contained 1 l LB medium. 1 ml ampicillin and 1 ml chloramphenicol was added to each flask together with 25 ml of the cell culture grown overnight in the presence of the same antibiotics. The culture was incubated for 2 – 3 hours at 37 °C and thenIPTG was added. The bacterial cells were harvested 3 hours after the IPTG induction (based on the best time point found during the pilot experiments).

3.3. Inclusion Bodies Isolation

The isolation of inclusion bodies was performed by cell lysis with disruption using a sonication step followed by centrifugation. Isolated cells (from 8 l of LB medium) were dissolved in 2 l of 20 mM Tris, pH 8.0 and stirred for 1 hour or until no clumps of bacteria were apparent. The suspension was sonicated 3× 30s with ultrasonic waves at maximum power.

The resulting pellet was collected by centrifuging for 10 minutes at 10,000xg. The pellet was then re-suspended in 1 l of 20 mM Tris, 1% Triton, pH 8.0 and again sonicated 3× 30s with ultrasonic waves at maximum power. Incubation with this detergent-containing buffer with stirring followed for 1 hour in room temperature and centrifugation for 10 minutes at 10,000xg. Further purification of the pellet was achieved by four washing steps with 20 mM Tris, pH 8.0.

Solubilization of the inclusion bodies was performed before refolding of the protein, by dissolving 500 mg of the inclusion bodies in 3.1 ml 6M guanidine and 20 mM Tris solution. 31 µl of 1M DTT was added and the solution was shaken for 1 hour at RT.
Afterwards, the cytosolic fraction (disrupted cells in Tris), the membrane fraction (disrupted cells in Triton) and the inclusion bodies were all analysed with SDS-PAGE electrophoresis.

3.4. Electrophoresis: SDS – PAGE

Electrophoresis is a universal technique for the separation of proteins based on the migration of SDS-charged proteins in an electric field. It allows the visualization and separation of proteins and as well enables us to roughly estimate the number of different proteins in a mixture. It also enables us to determine the extent of the purity of a particular protein preparation. Generally electrophoresis of proteins is performed in a gel made up of cross-linked polymer polyacrylamide. The polyacrylamide gel always acts as a molecular sieve and it also slows the migration of proteins in proportion to their charge to mass ratio as well as their shapes [11].
Sodium dodecyl sulphate (SDS) electrophoresis is very common method which is often used to estimate the purity and molecular weight of protein. The amount of SDS that binds to most proteins is in proportion to the protein molecular weight (i.e. one molecule of SDS for every 2 amino acid residues). Also the intrinsic protein charge is very insignificant compared to the large net negative charge contributed by the bound SDS. Therefore, each protein is able to reach a similar charge-to-mass ratio. Additionally, almost all proteins assume the same shape since their native conformation is altered due to the bound SDS. Thus in the presence of SDS, proteins are separated exclusively on the basis of the mass/molecular weight (size), with the smaller polypeptides migrating more rapidly through the gel [11].

NuPAGE® SDS-PAGE Gels (4 - 12% polyacrylamide) were used for protein separation. Protein sample was mixed with NuPAGE sample buffer (4×) and NuPAGE reducing agent (10×). The protein mixture was denatured by heating at 70°C for 10 minutes. Electrophoresis was carried out at a constant voltage of 150 V for approximately 35 minutes in an electrophoretic tank filled with 1× SDS-PAGE running buffer until dye migrated to the bottom of the gel. Afterwards, the gels were stained with Coomassie Brilliant Blue for 30 minutes and the protein bands became visible after destaining in a Coomassie destaining solution (composition in table 4).

3.5. Protein Refolding and optimization

Pilot refolding was performed in 8 different refolding buffers (buffer compositions and pH in table 3a). Inclusion bodies were solubilized according to the above described protocol. Then, 312 µl of the solubilized protein was added to 50 ml of each buffer with constant stirring for 3 hours at room temperature and then stored over-night at 4°C. The following day, the amount of precipitates was compared in all buffers both before and after centrifugation at 10,000xg for 10 minutes. The comparison was done using the ProteoStat® Protein aggregation assay kit and also by eye. Based on the precipitate yield, the best buffer for further optimization was chosen merely by comparing the pellets and the amount of precipitates. The buffer containing smaller amount of precipitates and a smaller pellet suggests the best conditions for further optimization (pH, salt etc.).

Based on the result of the pilot refolding, another optimization step was performed with 6 different buffers which were all similar to the best buffer found in the previous step. Each buffer contained different NaCl concentration (details in table 3b) at the same pH 8.0.

The next step was optimization of the concentration of inclusion bodies which was also a vital step in the optimization. The variation of the inclusion bodies concentration used in this step was: 1, 0.5, 0.25, & 0.125 g/l.

The same procedure for solubilisation and refolding (as above-mentioned) was repeated. After high speed centrifugation, the sample supernatants were concentrated to about 500 - 800 µl using the Amicon Ultra-15 Centrifugal Filter units with 3kDa cutoff. Buffers in all samples were then exchanged for TBS and again concentrated to about 500 µl. The concentrated protein solution was transferred to Eppendorf tubes and centrifuged at 10, 000×g, for 10 minutes to remove precipitates and other high molecular weight impurities. The supernatant was transferred to new tubes and TBS was added to reach an equal final volume of 800 µl for all 6 samples. The success of the refolding was then determined using four different methods: Gel electrophoresis, protein concentration measurement using BCA method, aggregate measurement and protein inhibitory activity measurement.

3.6. Concentration Determination - BCA Method

BCA method involves a biuret reaction in which Cu^{+2} is reduced to Cu^{+1} by a protein in an alkaline medium. This method uses a unique reagent containing bicinchoninic acid with a highly sensitive and selective colorimetric detection of the cuprous cation (Cu^{+1}). In the first

part of the reaction called the biuret reaction, copper chelates with protein to form a light blue complex in an alkaline medium. The next reaction involves the colour development; BCA (bicinchoninic acid) reacts with the reduced copper (cuprous cation: Cu^{+1}) formed in the first step, and the resulting product is a purple complex solution. The resulting absorbance of 562 nm is nearly linear with increasing protein concentrations over a broad working range (20-2000 µg/ml) (Pierce™ BCA Protein Assay Kit).

The colour formation of the protein with BCA is as a result of the number of peptide bonds, the presence of cysteine, cystine, tryptophan and tyrosine and the molecular structure of protein. Protein concentration is thus generally determined with reference to standards, bovine serum albumin (BSA) according to the method described in the Pierce™ BCA Protein Assay Kit.

Following this method, series of 9 different dilutions of known concentration was prepared from the BSA protein stock solution as shown in the table 8. The working reagent was prepared by mixing in 50 parts of reagent A with one part of reagent B. 200 µl of working reagent was used for each of 9 BSA standards plus unknown samples in 2 replicates. 25 µl of each standard was pipetted into the microplate well, as well as 25 µl of the protein samples. Next, 200 µl of the working reagent was added to each well and mixed thoroughly by shaking the plate for 30 seconds. The plate was covered and incubated for 30 minutes at 37°C and then cooled to room temperature. Tecan infinite M200 microplate reader was used to measure the absorbance at 562 nm.

Table 8: Preparation of the Protein Standard

Vial	VoL. of Diluent µl	BSA/Sample µl	Final BSA Conc. µg/ml
A	0	60	2000
B	25	75	1500
C	100	100	1000
D	40	40 B	750
E	100	100 C	500
F	100	100 E	250
G	100	100 F	125
H	80	20 E	25
I	60	0	Blank

3.7. Aggregate Measurement

To determine the amount of aggregated protein, the ProteoStat® Protein aggregation assay kit was used. The ProteoStat® Protein aggregation assay kit contains a fluorescent dye that shows almost no fluorescence in the presence of monomeric protein. Whereas the intensity of the fluorescence increases 20 – 90 times upon binding an aggregate. The method involves measuring of the amount of aggregates in the sample; thereby indirectly determining the content of the native, monomeric protein. The more aggregates measured in a particular sample, the smaller the native protein. The aggregate amount was determined using the method described in the ProteoStat® Protein aggregation assay kit. Further details about the specific kit are not provided by the supplier for marketing reasons.

5 µl of each refolded protein solution was added into the microplate in 2 replicates. 95 µl of the ProteoStat detection reagent loading solution was added into each well and mixed by shaking the microplate. The plate was then incubated in the dark for 10 minutes at room temperature. Fluorescence intensity was read using Tecan infinite M200 microplate reader with 550nm as the excitation wavelength and 600nm as the emission wavelength.

3.8. Activity assay method

Enzyme assay can be used to investigate enzyme kinetics, which is the study of chemical reactions, catalysed by enzymes. To study the catabolic action of an enzyme, the reaction rate and the influence of varying assay conditions have to be investigated. Usually, the aim of measuring enzyme activity is to determine the reaction rate under certain conditions and then compare the activity between different samples. [14]

The activity of Cathepsin L was investigated using the Fluorogenic Peptide Substrate, Z-LR-AMC. The Cathepsin L stock solution was diluted twice by the factor of 50× in the Assay Buffer (composition in table 5) by diluting two times 1 µl of the stock solution in 49 µl of the Assay Buffer (2500 times dilution of the stock solution provided the diluted solution). 1 µl of the diluted solution of Cathepsin L was loaded in the 96-well microplate. 1 µl of each of the tested protein samples was then added (in varying concentrations) and 45. 5 µl Assay Buffer (46.5 µl in the 'no protein' positive control) were added to the wells which was then incubated at room temperature for 10 minutes under constant shaking with speed 600. Afterwards, 2.5 µl of the substrate was added in 250 µM final concentration to each well which (after shaking for 5 seconds) was then incubated in the Tecan Infinite M200 fluorimeter for 10 minutes at 30°C. The substrate hydrolysis rate was measured fluorometrically at excitation and emission wavelengths 365 nm and 450 nm with a cutoff at 435 nm. A statistical analysis of the observed inhibition of the enzyme in the presence of the protein was done using Excel.

In this work, the inhibitory activity of the Cathepsin L was done strictly by my supervisor while I followed him all the way through.

4. Results

4.1. Overexpression

One of the goals of this thesis was to make a tick gene to be overexpressed in a strain of bacteria. For the overexpression of the desired protein G10, a tick salivary cysteine protease inhibitor (cystatin) with a molecular weight approximately 14kDa, the BL21(DE3)pLysS *E. coli* strain was used. Prior to the overexpression, a plasmid with a gene of interest was transformed to the competent *E. coli* cell and the cells were incubated overnight on an agar plate with antibiotics. A single colony was picked and used for inoculation of 1ml LB medium with appropriate antibiotics as stated in the methods (section 3.2). This step was followed by the pilot expression which serves basically to find the best time point when the protein expression peaks.

Each of the 9 samples collected at different time point for both the pellet and the supernatant were analysed using NuPAGE gel electrophoresis.

As we can see in figure 3, the protein is already overexpressed after the first hour of cell induction and the overexpression level of the protein increases with time. The best expression time was chosen to be three hours after induction since the strength of the band does not increase anymore with increasing time of culture.

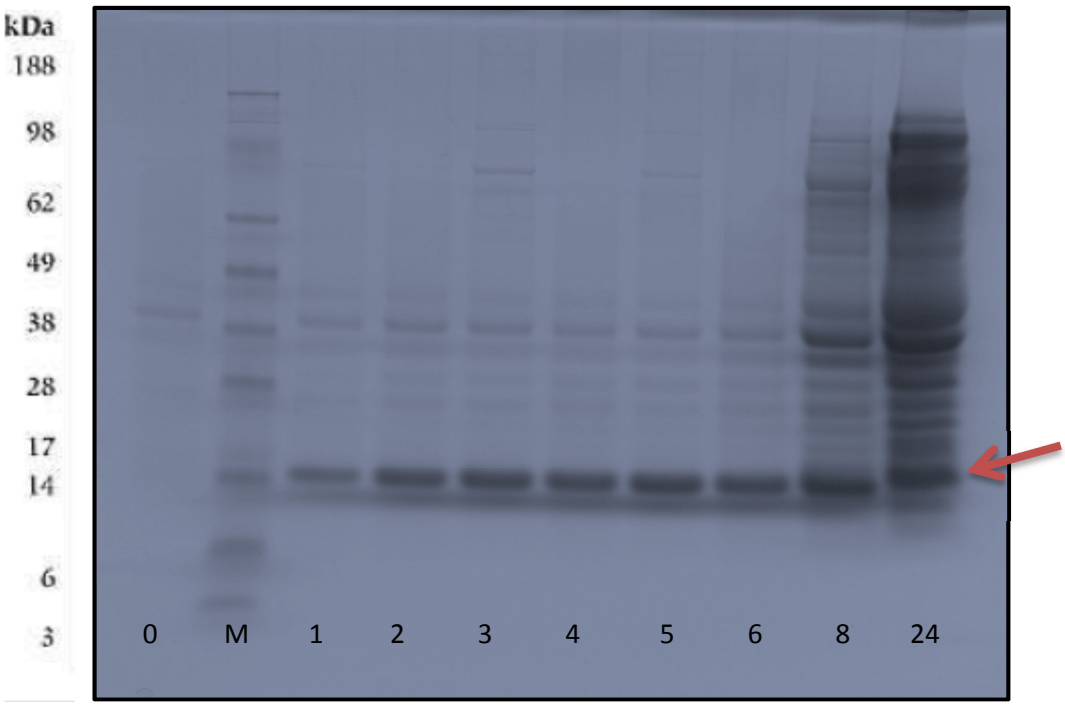

***Figure 3**:* The pilot expression experiment: the insoluble protein fraction of the bacteria was analysed by SDS-PAGE. Where 0: culture without induction with IPTG and M: molecular

weight marker. The subsequent numbers show cultures induced with 1mM IPTG. 1 - 6: cultures after one - six hours of induction, 8 and 24 are cultures after the 8th hour and 24th hour of induction. The protein of interest is shown in bands at 14KDa and marked with an arrow.

As seen from figure 4, the SDS-PAGE analysis of the soluble fraction of bacterial lysates shows no bands at all at 14 KDa. All further experiments from this point on were carried out with the pellet while the supernatant was discarded.

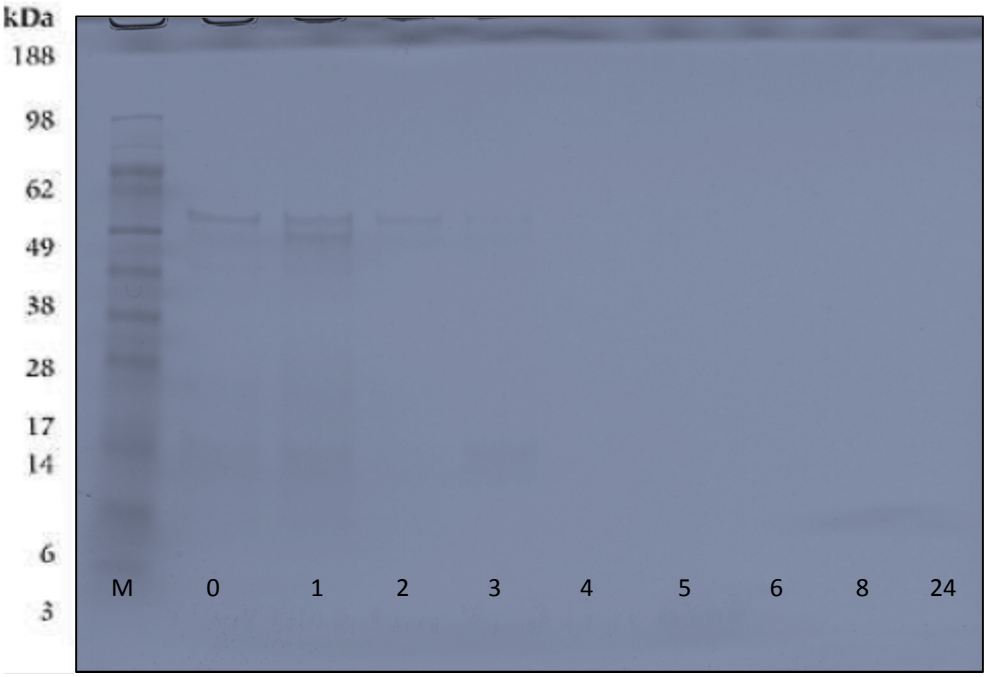

***Figure 4**:* The pilot expression of the supernatant analysed with SDS-PAGE. M represents the molecular weight marker. The subsequent numbers, though with no protein bands of the appropriate Mw represent protein induction from 1st – 6th hour, the 8th hour and 24th hour.

4.2. Inclusion Bodies Isolation

With the result of the pilot expression, the protein overexpression was scaled up to 8 litres a day. The cytosolic fraction (disrupted cells in Tris), the membrane fraction (disrupted cells in Triton) and the inclusion bodies were separated according to the method described in section 3.3. The abundance of G10 in these fractions was analysed using SDS-PAGE to confirm again the fact that G10 is overexpressed in the insoluble fraction, thus in the inclusion bodies, as we can see in figure 5.

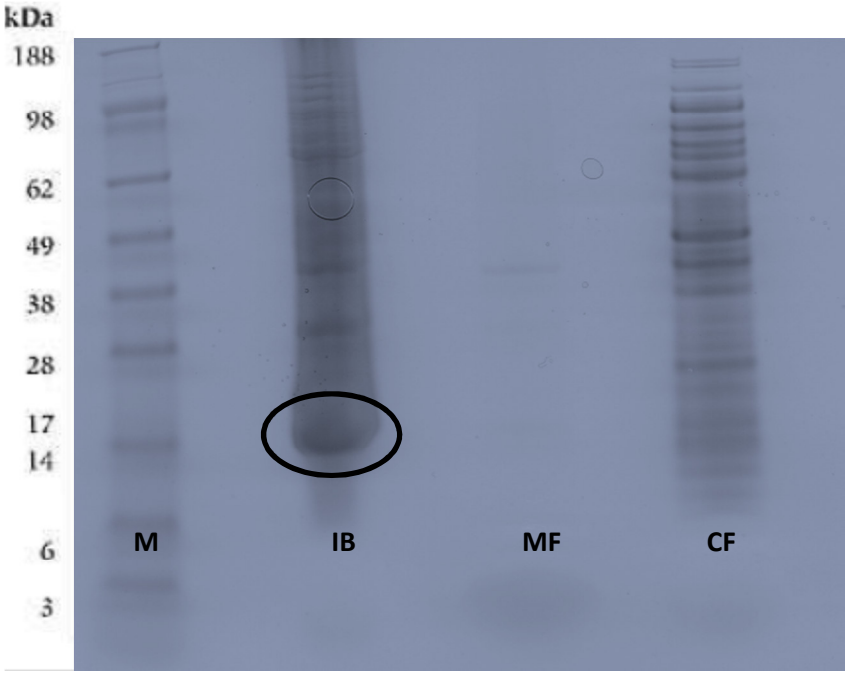

***Figure 5*:** The presence of the target protein in the different fractions.
M: molecular weight marker, **IB**: inclusion bodies, **MF**: membrane fraction, **CF**: cytosolic fraction. The circle shows the band corresponding to the target protein in the inclusion bodies.

Figure 5 shows that the overexpressed protein is found in the inclusion bodies and not in the cytosolic fraction or in the membrane fraction of the bacterial lysates. The cytosolic and membrane fractions were discarded. The inclusion bodies were stored at -20°C for protein refolding.

4.3. Protein refolding and buffer optimization

The overexpressed protein was refolded in 8 different buffers with different compositions and pH (table 3a), using the method for protein refolding described in section 3.5 at the IB concentration 1 mg/ml of refolding buffer.

The refolding conditions in each of the 8 buffers were determined mainly by mere pellet comparison. Buffer 5 (0.2 M Tris, 0.1 M NaCl, 5 mM KCl, 10 mM EDTA at pH 8.0) and buffer 8 (0.2 M Tris, 2.4 M NaCl, 100 mM KCl, 10 mM EDTA, at pH 8.0) were chosen as the best conditions.

Further optimization was carried out with these buffer 5 and buffer 8 in order to determine the one buffer with the optimal refolding conditions. To achieve this, protein concentration in both

buffers as well as the amount of aggregated protein was measured. The fluorescence of the aggregates in the refolding buffers as well as the protein concentration is shown in table 9.

Table 9:

Protein concentration measurement & Aggregate measurement for protein sample.

<>	Buffer 5	Buffer 8		<>	Buffer 5	Buffer 8
mg/ml	0.0365	0.0617		Fluorescence units	324	293

Based on this result, buffer 8 with the composition (0.2 M Tris, 2.4 M NaCl, 100 mM KCl, 10 mM EDTA, at pH 8.0) was chosen as the best one since it gave better refolding results than buffer 5 with higher not-precipitated protein concentration and less aggregates as shown in the table 9.

4.4. Optimization of NaCl concentration in the refolding buffer

For further refolding, all other parameters of the buffer 8 were maintained while the concentration of the NaCl was varied for the next optimization. For this, 50 ml of buffer with (20, 50, 100, 150, 240 and 300) mM salt concentration were used as described in the method section (table 3b). The BCA method was used to estimate the total protein concentration and the amount of aggregated protein was estimated using the ProteoStat® Protein Aggregation Assay Kit. The samples were also analysed on a gel using electrophoresis and their inhibitory activity against cathepsin L was measured. The following figures 6 – 9 show the results.

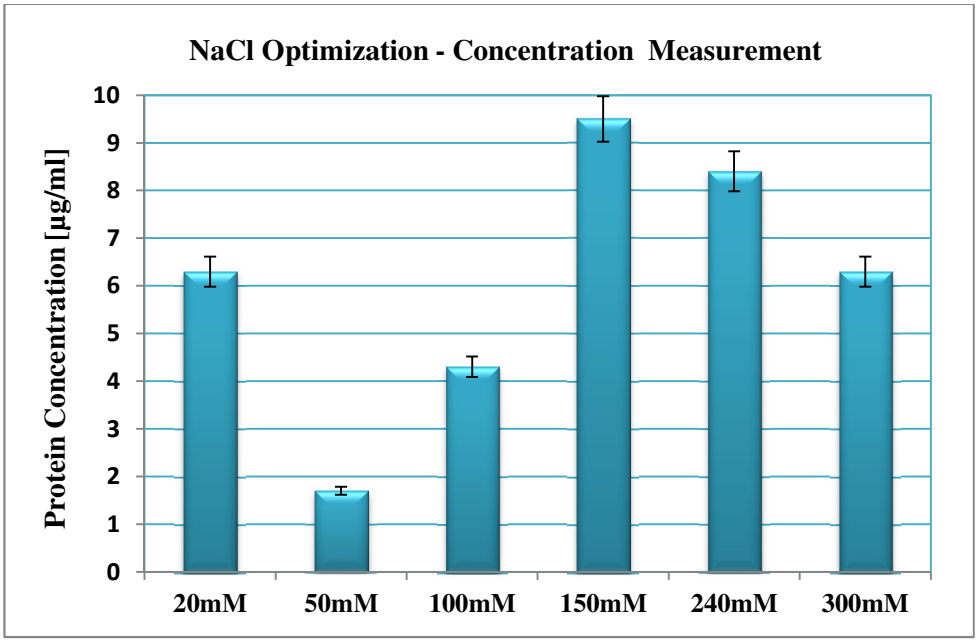

Figure 6: Total concentration of protein in refolding buffer 8 with different concentration of NaCl +/- SE.

From figure 6, the buffer with 150 mM NaCl contains the highest protein concentration, reaching 9.5 µg/ml while 50 mM NaCl gave the lowest total protein concentration.

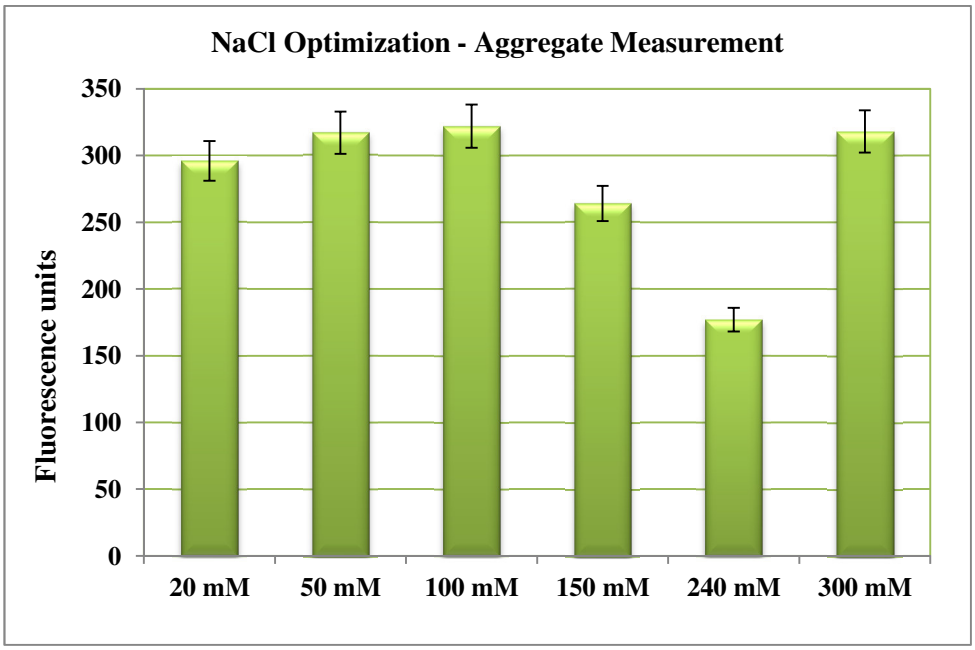

***Figure 7*:** Amount of aggregated protein in the refolding buffer 8 with different NaCl concentration +/- SE.

Figure 7, shows the amount of aggregates upon refolding. Buffer with 240 mM NaCl concentration has significantly less amount of aggregates followed by that of 150 mM NaCl.

Buffers with 300 mM, 100 mM and 50 mM have fluorescence of approximately 320 arbitrary fluorescence units (much higher) which means, these buffer compositions are not so favourable for the protein refolding.

All the samples were then run on gel to determine the best condition of buffer for refolding. The result is presented in the figure 8.

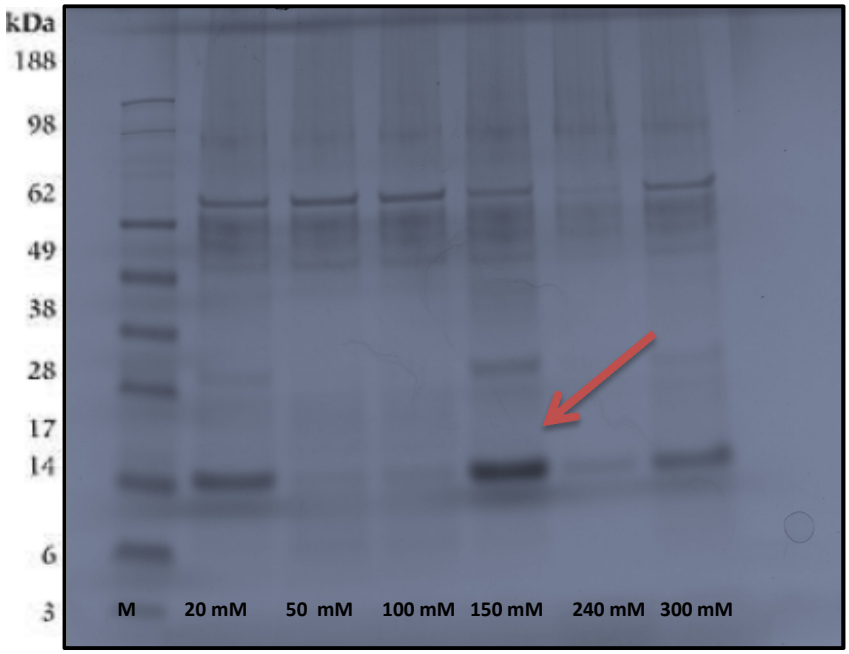

***Figure 8*:** Scan of a gel with protein refolded in buffer 8 with different NaCl concentration. M: molecular weight marker, 20 – 300 mM: molarities of NaCl concentration. The arrow shows the strongest band corresponding to G10.

From figure 8**,** it is obvious that the buffer with 150 mM NaCl has the best refolding conditions since the protein band is the strongest here. Therefore, based on the SDS-PAGE analysis, the best buffer composition for protein refolding is the buffer with 150 mM NaCl concentration, followed by the 20 mM NaCl which has the second strongest band while the 300 mM NaCl is also fairly considerable.

Figure 9 shows how G10 affected the enzymatic activity of cathepsin L. The inhibition activity decreases with increasing NaCl concentration in the refolding buffer. While samples refolded at 20-100mM NaCl suppress the CatL activity to the level that corresponds to less than 2 fluorescence units, other samples reach values from 4.3 to 8.8 arbitrary fluorescence units.

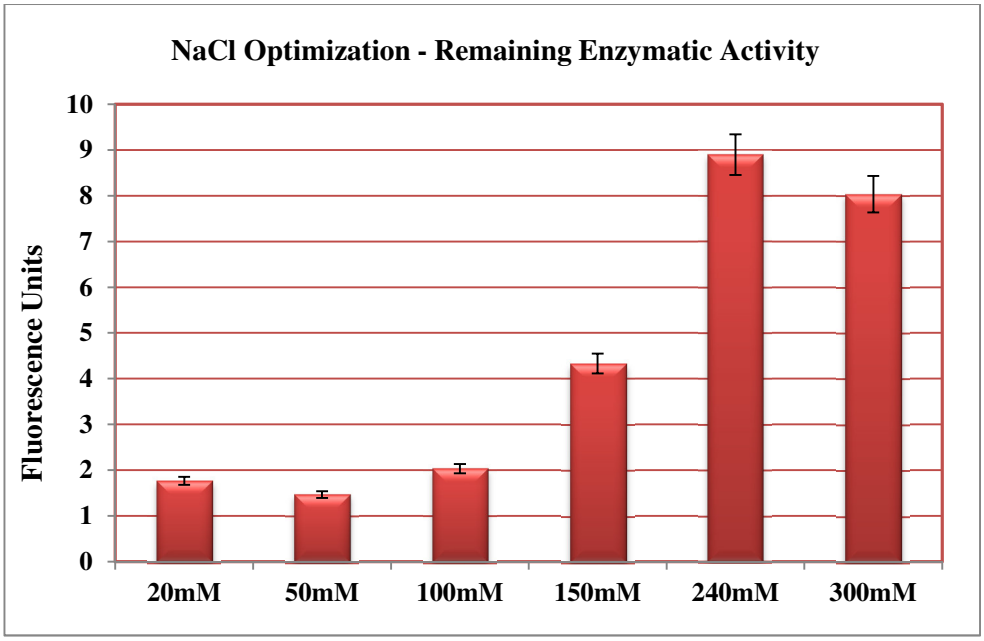

Figure 9: The remaining enzymatic activities of cathepsin L affected by G10 refolded in different refolding buffers +/- SE. G10 samples refolded at low NaCl concentration (20-100mM) show the lowest remaining CatL activity – and thus the strongest CatL inhibition. CatL inhibition then decreases (activity increases) when using G10 refolded at higher NaCl concentration.

Based on the results of all the employed methods, 150mM NaCl was chosen as the best condition. It shows the strongest band on the gel, highest protein concentration and also less aggregates than most of the other buffers. The only method that did not point directly to 150mM was the activity test, where this condition was worse than 3 other. We have however chosen 150mM NaCl, because we preferred higher yield of protein. Further optimization was then carried out with buffer 8 with 150 mM NaCl.

4.5. Inclusion bodies concentration optimization

As the next step, concentration of the inclusion bodies was optimized in the buffer that showed the best results in the previous steps. The final concentration of IB ranged between the values of 0.125 mg/ml, 0.25 mg/ml, 0.5 mg/ml and 1 mg/ml. The composition of the used refolding buffer is 20 mM Tris, 150 mM NaCl, 10 mM KCl, and 1 mM EDTA, at a pH of 8.0, as mentioned above. Refolding conditions were evaluated in the same way as in the previous step. The following figures 10 – 13 show the results.

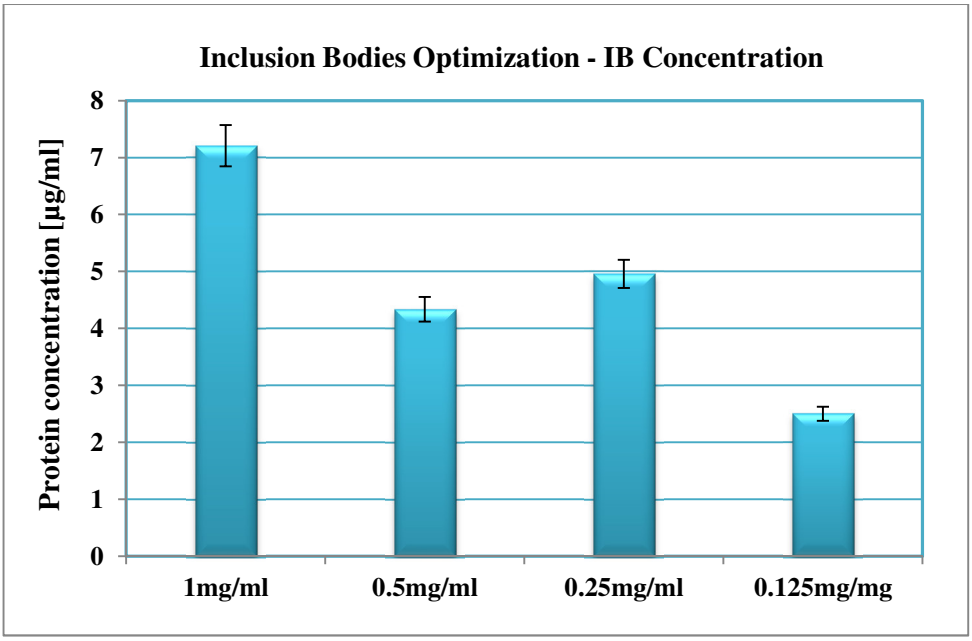

Figure 10: Total concentration of refolded protein in refolding buffer with different concentrations of IB +/- SE. The concentrations on the x-axis represent the initial concentration of the inclusion bodies.

As shown in figure 10, the amount of refolded protein decreases with decreasing IB concentration. The 1 mg/ml IB concentration shows the most refolded protein, while that of 0.125 mg/ml has the least. With these results, it is obvious that the total refolding yield in terms of final protein concentration depends on the amount of the initial IB concentration used.

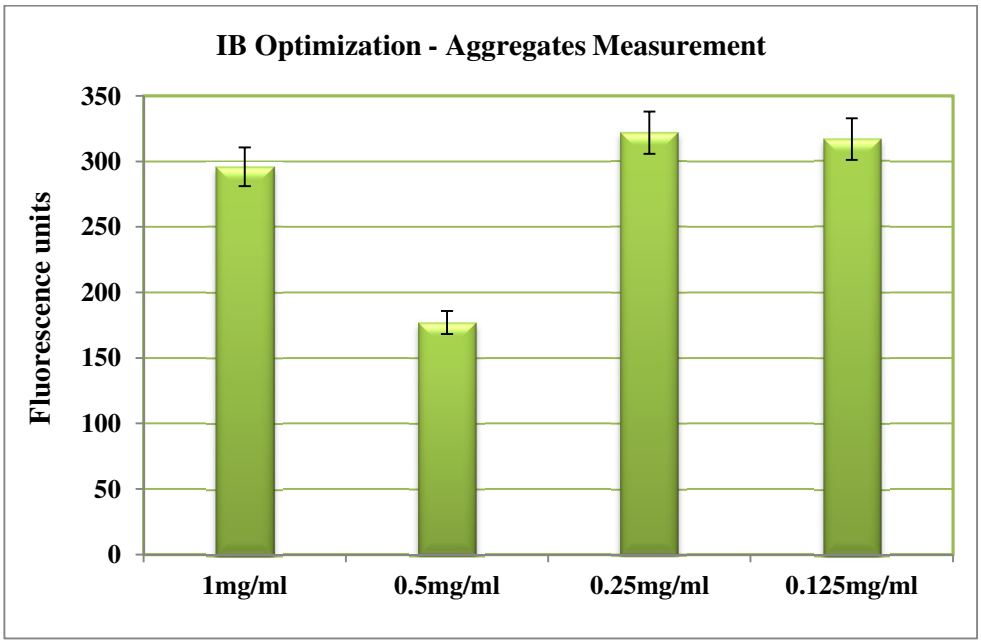

***Figure 11*:** The amount of aggregates present in the refolding buffers with different IB concentration +/- SE.

As seen in figure 11, IB concentration 0.5 mg/ml contains smaller amount of aggregates, almost half the amount of aggregates when compared to other conditions.

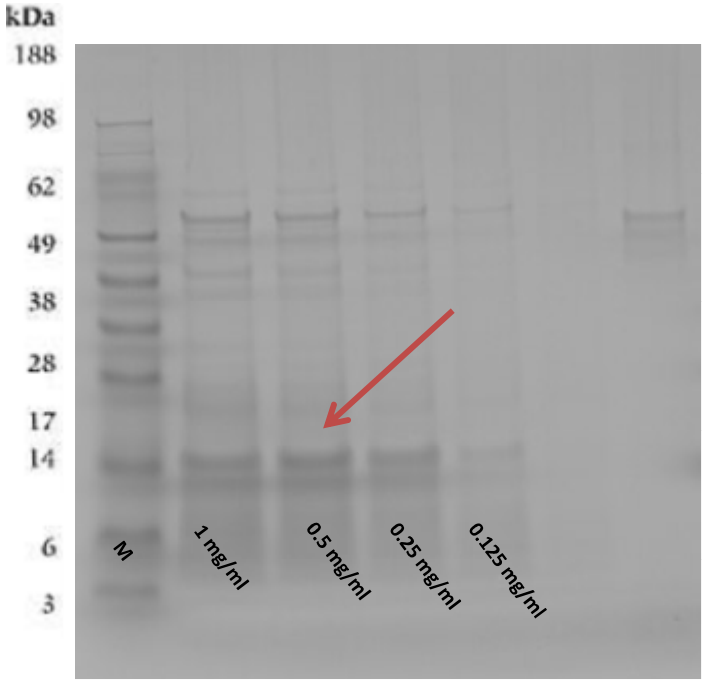

***Figure 12*:** Bands of G10 refolded at various IB concentrations in the best refolding buffer. M represents the marker, while the numbers represent the IB concentration.

Figure 12 shows that the protein bands with 0.25 mg/ml and 1 mg/ml are fairly strong bands while that of the 0.5 mg/ml is stronger, and thus the IB are best refolded at this concentration. There is almost no band with the 0.125 mg/ml IB concentration which means that this IB concentration will not be ideal for refolding.

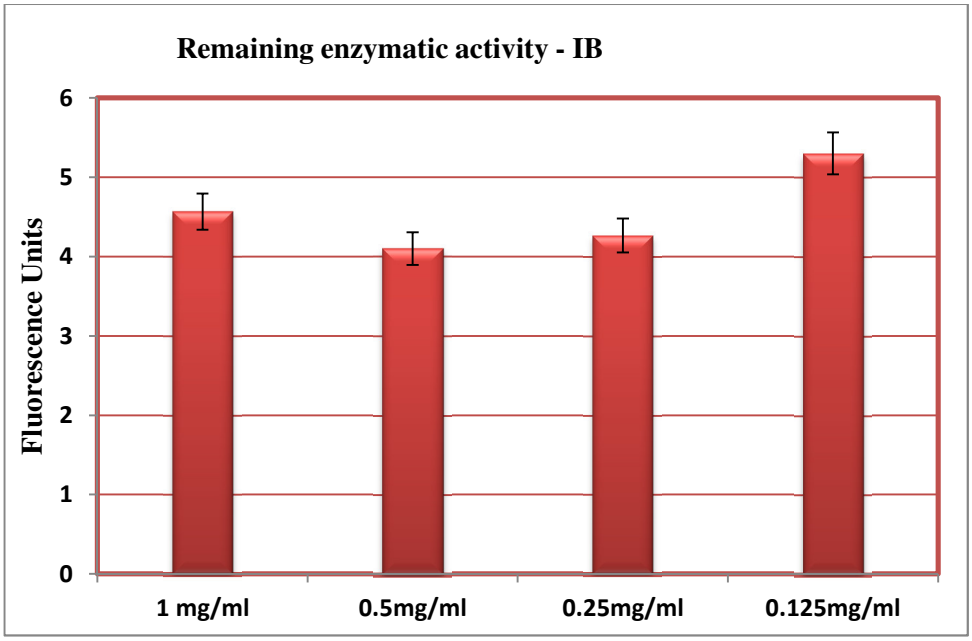

Figure 13: The remaining enzymatic activities of cathepsin L with different IB concentration +/- SE. The numbers below the graph represent the IB concentration during refolding.

The inhibition activity is a little stronger using sample refolded at 0.5 µg/ml concentration. However there are no big differences among these four samples.

Based on the given data, we have chosen the IB concentration 0.5mg/ml as the best condition. It shows the strongest band on the gel, it contains less aggregates and as well gave the lowest remaining CatL activity thus the best inhibition. Nevertheless, total protein yield is not so favourable with this refolding condition since it is worse than using 1 mg/ml and comparable to 0.25 mg/ml. On the other hand, the lower yield in comparison to 1 mg/ml is compensated by only half consumption of the inclusion bodies.

5. Conclusion and Discussion

One of the aims of this thesis was to make a tick gene to be overexpressed in bacteria and then produce high amount of the tick protein. We can conclude that the protein Cystatin G10 was strongly expressed in the *E. coli* bacterial strain BL21(DE3)pLysS and the best expression time is the third hour after protein induction with IPTG, as seen in figure 3.

Another area, this thesis focused on, was refolding optimization to find out the best conditions to successfully refold protein G10 produced in the bacterial inclusion bodies. Determination of the best refolding condition was carried out with different approaches. As seen from the results presented in figures and tables in section 4, protein refolding is highly dependent (among other parameters) on buffer compositions and the initial concentration of IB as described also in the literature [15, 16].

First of all, the salt component of buffer was analysed by running samples of the refolded proteins at different NaCl concentrations on a gel, the total protein concentration in the refolded protein solution was measured, the amount of aggregated protein in each refolded conditions was measured and finally the inhibitory activity of the refolded protein was determined. At the end of the analysis, the best NaCl concentration for G10 refolding was determined successfully.

For the different NaCl concentrations optimized, G10 refolds best at 150 mM NaCl concentration as shown in figure 6, which means that the ionic strength of refolding buffer plays a major role in G10 refolding [15]. Further optimization using the best refolding buffer also shows that initial IB concentration has major impact in successful refolding of G10. If we focus our interest only on total protein experimental yield, we can conclude from figure 10 that G10 was best refolded at 1 mg/ml initial IB, however, it could be argued that 0.5 mg/ml initial IB concentration is preferred if we consider aggregation and inhibitory activity of G10.

On the other hand, there are other approaches that point towards successful protein optimization which could still be investigated in future experiments. The influence of other parameters like pH, temperature and the ionic strength of refolding buffer with a different salt could be employed for further analysis [15]. The preferred refolding conditions can further be optimized at different pH to find out if G10 refolding efficiency increases by slightly increasing/decreasing pH of refolding buffer or if G10 simply refolds best at pH near the protein pI. Nevertheless, these steps could not be carried out due to the demanding schedule of the study program and the short time available for the project.

Cathepsin L has been proved to be actively inhibited by G10. Some cysteine protease inhibition was expected based on the protein sequence, but could have been proven only after producing a small amount of the protein in its native form. These results also confirm that G10 could be an immunomodulator, playing a role in altering the immune response by inhibiting CatL.

The results of the thesis together with literature, confirm that inhibitors of cathepsin L can act as immunomodulators thus making G10 a good protein to be further investigated. One of the approaches would be to test G10 immunomodulatory properties, for instance the ability of G10 to suppress lymphocyte proliferation and recruitment and to suppress inflammation [3]. Another approach would be to test G10 as a potential vaccine against ticks and thus preventing tick borne diseases transmission.

In conclusion, the inhibition of harmful cysteine proteases in health and disease prevention is very important and has become an important area of focus for many biologists. This is because maintaining the right equilibrium between cysteine proteases and complexes with their inhibitors is very crucial in all living systems. [17-18]

In this thesis, I have learned many new methods which would help me in future. I was successful in the task of protein expression. Although there were difficulties in refolding, because G10 was a rather very difficult protein to refold in solution, several steps of refolding condition optimization were successful.

6. Appendix

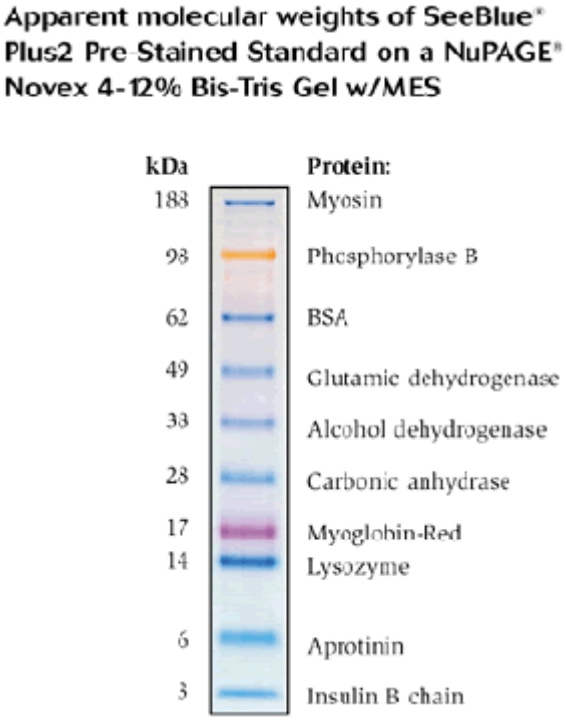

Figure 13: The molecular weight marker See Blue® Plus2. The molecular weight marker (indicated as **M**) in the gel pictures is compared with this standard. Diagram is adapted from www.lifetetchnologies.com

7. References

[1] **Sanson T., Taylor J. P. III**; *Tick-Borne Diseases*. Medscape. MD LLC. 2013

[2] **Walker A.**; *Ixodes-ricinus-life-cycle*. Wikimedia Commons. 2012

[3] **Schwarz A, Valdés J. J., Kotsyfakis M.**; *The role of cystatins in tick physiology and blood feeding*. Ticks and Tick-borne Diseases. 2012. 117–127

[4] **VALENZUELA J. G.**; *Exploring tick saliva: from biochemistry to 'sialomes'*. Parasitology. 2004. 129

[5] **Rzchon M., Chmiel D. and Stec-Niemczyk J.**; *Modes of Inhibition of Cysteine Proteases*. Acta Biochimica Polonica. 2004

[6] **Kotsyfakis M., Horka H., Salat J., and Andersen F. J.**; *The crystal structures of two salivary cystatins from the tick Ixodes scapularis and the effect of these inhibitors on the establishment of Borrelia burgdorferi infection in a murine model*. Molecular Microbiology. 2010. 77(2), 456–470

[7] **Kakegawa H., Nikawa T., Tagami K., Kamioka H., Swith H., Sumitani K., Kawata T., Drobnic-Kosorok M., Lenarčič B., Turk V., Katunuma B.**; *Participation of cathepsin L on bone resorption*. FEBS Lett. 1993. 321, 247-250

[8] **Clark E. D. B.**; *Refolding of Recombinant Proteins*. Current Opinion in Biotechnology. 1998. 9(2), 157-63

[9] **Vallejo F. L. and Rinas U.**; *Strategies for the recovery of active proteins through refolding of bacterial inclusion body protein*. Microbial Cell Factories. 2004. 3, 11

[10] **Sambrook J. and Green M. R.**; *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. 2001

[11] **Bruce A., Johnson A., Lewis J., Raff M., Roberts K. and Walter P.**; *Molecular Biology of the Cell, 5th Edition*. New York: Garland Science. 2008

[12] Pierce Protein Biology; *Overview of Cell Lysis and Protein Extraction*. Thermo Fisher Scientific Inc. 2015

[13] **Bird P. I., Pak S. C., Worrall D. M., and Bottomley S. P.**; *Production of recombinant serpins in Escherichia coli*. Methods. 2004. 32(2), 169–176

[14] **Scopes R. K.;** *Enzyme Activity and Assays*. Encyclopedia of Life Sciences. 2002

[15] **Rudolph R., Lilie H.**; *In vitro folding of inclusion body proteins*. FASEB J. 1996. 10(1), 49-56

[16] **Yamaguchi S., Yamamoto E., Mannen T. and Nagamune T.**; *Protein refolding using chemical refolding additives*. Biotechnol J. 2013. 8(1), 17-31

[17] **Turk V. and Bode W.**; *The cystatins: protein inhibitors of cysteine proteinases*, FEBS lett. 1991. 285(2), 213-219

[18] **Grzonka Z., Jankowska E., Kasprzykowski F., Kasprzykowska R., Lankiewicz L., Wiczk W., Wieczerzak E., Ciarkowski J., Drabik P., Janowski R., Kozak M., Jaskólski M., and Grubb A.**; *Structural Studies of Cysteine Proteases and Their Inhibitors*. Acta Biochemica Polonica. 2001. 48, 1