UPPSALA
UNIVERSITET

UPTEC X 16 002
Examensarbete 30 hp
April 2016

A method for efficient synthesis of RNase A, using inteins

Ayda Zamany Company

UPPSALA UNIVERSITET

Degree Project in Molecular Biotechnology

Masters Programme in Molecular Biotechnology Engineering,
Uppsala University School of Engineering

UPTEC X 16 002	**Date of issue 2016-04**

Author

Ayda Zamany Company

Title (English)

A method for efficient synthesis of RNase A, using inteins

Title (Swedish)

Abstract

A protein with a complex structure can be challenging to produce recombinantly as it may become insoluble if the structure of the protein is not correctly formed. RNase A is a protein that is insoluble when produced recombinantly in a popular host like *Escherichia coli* due to its complex structure and for its ability to degrade its own transcript, which makes it cytotoxic. This study aims to use inteins to produce recombinant RNase A by dividing the gene encoding RNase A and fusing the two parts with the genes encoding inteins. Inteins are insertion sequences present in protein sequence, which can excise themselves and ligate the amino acids in the protein sequence present around them through the formation of a peptide bond. This process is called protein splicing and the goal was to use this process to produce an active RNase A from inactive parts using inteins. In this study, three out of eight RNase A-intein constructs were found insoluble from the tested conditions.

Keywords

RNase A, Recombinant protein, Split inteins, Inteins, Protein Splicing, Expressed Protein Ligation, Cis/trans-Splicing

Supervisors

Prof. Dr. Thomas Friedrich
Technical University of Berlin

Scientific reviewer

Ass. Prof. Lars-Göran Josefsson
Uppsala University

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **41**

Biology Education Centre Biomedical Center Husargatan 3, Uppsala
Box 592, S-751 24 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 471 4687

A method for efficient synthesis of RNase A, using inteins

Ayda Zamany Company

Populärvetenskaplig sammanfattning

Proteiner är en av de essentiella beståndsdelarna i allt levande och består av aminosyror som är bundna genom peptidbindningar. Proteiner har en tertiärstruktur som utgör veckningar vilket avgör dess funktionalitet. Att producera rekombinant protein innebär att transformera genen för ett önskat protein i en modellorganism som sedan manipuleras för att producera proteinet. Rekombinant produktion av protein är ett användbart sätt att producera proteiner då isolering av protein från vävnad eller annat biologiskt material kan vara omständligt. Dock finns det komplikationer vid användning av rekombinanta tekniker då det ofta är svårt att återskapa den naturliga tertiärstrukturen hos proteiner som i vissa fall är avgörande för dess funktion.

RNase A är ett protein som degraderar RNA och är användbart vid användning av olika laborativa tekniker. Att producera proteinet rekombinant genom att uttrycka det i *Escherichia coli* som modellorganism är utmanande då det är svårt att efterlikna dess tertiärstruktur som innehåller disulfidbindningar. Vid felaktiga bindningar i dess tertiärstruktur blir proteinet olösligt och förlorar sin funktionalitet.

I den här studien var målet att producera rekombinant RNase A genom att använda inteiner, vilket är proteiner som binder ihop delar av proteiner genom peptidbindningar (s.k. protein splicing) och återskapar fullt funktionella protein. Genen för RNase A är delad och ihopsatt med gener för inteiner. Dessa kloner uttrycks separat, mixas, och slutligen utförs protein splicing som binder ihop proteinhalvorna och genererar ett fullständigt RNase A.

Ett lyckat försök för att återskapa RNase A genom att använda intein skulle även kunna användas för andra proteiner som är olösliga vid rekombinant produktion.

Examensarbete 30 hp
Civilingenjörsprogrammet i Molekylär bioteknik
Uppsala universitet, april 2016

Abbreviations and glossary

Cis-splicing – Protein splicing carried out by one single intein segment.

CV – Column volume

Expressed protein ligation (EPL) – A process where a recombinant protein with a reactive thioester at its C-terminal is generated for ligation with a molecule or peptide having a -SH group at its N-terminal.

Intein – Protein segments that join protein parts through a peptide bond and excises itself.

IPTG – Isopropyl-beta-D-thiogalactopyranoside is a reagent that is used as a mimic of allolactose. It is a lactose metabolite used to induce transcription of the *lac* operon.

Protein splicing – The process where inteins join protein parts.

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

Trans-splicing – Protein splicing carried out by split inteins, which consists of two intein segments.

Table of Contents

1 Introduction

1.1 Recombinant protein production

Proteins are important constituents in every living organism since they take part in many naturally occurring biological processes. They fulfill a wide range of functions within living organisms such as in transporting and storing biomolecules and catalyzing chemical processes. A traditional way of isolating protein is to purify it from animal or plant tissue. However, this requires kilograms of material and the yield of protein is very small. An alternative for producing protein is to use recombinant technologies where a DNA molecule encoding the protein of interest is inserted in a model organism such as bacteria using a plasmid in order to produce the recombinant protein of interest[1]. *Escherichia coli* is a commonly used host for expressing recombinant protein and will be used in this project.

One of the main challenges when producing a recombinant protein in *E. coli* is to correctly mimic its native tertiary structure. If the tertiary structure is not correctly formed, improper folding will occur, the protein will form inclusion bodies and thus be inactive. In other words, proteins generally tend to aggregate when non-native protein folding occurs[2].

1.2 RNase A – Insoluble and Cytotoxic

RNase A is a ribonuclease that catalyzes the degradation of RNA. It hydrolysis the RNA strands at internal phosphodiester bridges specifically for single-stranded RNA and cleaves the 3' bonds from pyrimidine. RNase A is useful in many biological applications due to its ability to degrade RNA. For instance, it is widely used when isolating DNA as it separates it from RNA.

RNase A is a small molecule with a chain of 124 amino acids and with a molecular mass of 13686 Da. RNase A has four disulfide bonds in its tertiary structure, which contributes to the high stability of the molecule[3]. The traditional way of isolating RNase A is to purify it from bovine pancreas. However, it is preferable to avoid animal material due to safety regulations and for therapeutic usage of recombinant protein. As previously stated, the production of recombinant proteins can be challenging. RNase A has shown to be insoluble when expressed recombinantly in *E. coli* due to the formation of inclusion bodies.
Recreating the activity of the protein requires correct folding including proper formation of disulfide bonds. If not, the protein can aggregate and create inclusion bodies that will make the protein insoluble. The recombinant way of obtaining RNase A is complicated due to low yield, improper folding and cytotoxicity. The expression

of RNase A in *E. coli* has shown to be cytotoxic due to the risk of degrading all sorts of RNA transcripts including its own. In order to produce RNase A recombinantly in an efficient way, it is necessary to overcome the issues of cytotoxicity and insolubility[1].

1.3 Inteins – Protein Splicing

Inteins are described in analogy to introns in RNA since they are protein insertion sequences that excise themselves from the host protein during posttranslational maturation. This process of posttranslational processing is called protein splicing[4]. They are mobile genetic elements and they exist in all three domains of life including viruses and bacteriophages. They occur in an even sporadic distribution among closely related species. Inteins tend to hitchhike as fragments of horizontally transferred genes, gene clusters, genome fragments, or even whole chromosomes[5]. The intein protein family is part of the Hint superfamily that contains three other families such as hog-hint and two types of Bacterial intein-like domains sharing the same structural fold and common sequence features[6].

The intein (*in*tervening pro*tein*) carries out the process of protein splicing where it excises itself from the host protein through cleavage of peptide bonds and ligates the flanking extein (*ex*ternal prot*ein*) segments with a new peptide bond. The process of protein splicing occurs spontaneously where the domain folds itself and requires no external factor such as energy source or cofactor[7]. Less than 5% of the identified genes coding for inteins are split inteins. The split inteins formed form two separate polypeptides: N-terminal and C-terminal intein (Int^N and Int^C), which upon assembly spontaneously carry out a trans-splicing reaction. Split inteins are spontaneously carrying out protein splicing due to the affinity between the Int^N and Int^C [7]. In cis protein splicing, in contrast, splicing is carried out by only one single intein polypeptide that is embedded in a host protein[8].

The overall protein splicing mechanism is the same for cis- and trans-splicing. The peptide bond between the N-extein and intein is activated by an N-to-S acyl shift that initiates a linear thioester intermediate. The thioester undergoes a trans-thioesterification where the end product is a branched thioester intermediate, which has reacted to the first residue in the C-extein, a cysteine. Lastly, the Asn residue in the C-terminal cyclizes and the branched thioester resolves where the N-extein and the C-extein join. A peptide bond will be formed as the thioester rearranges and the spliced protein is produced[8,9]. This process is illustrated in Figure 1.

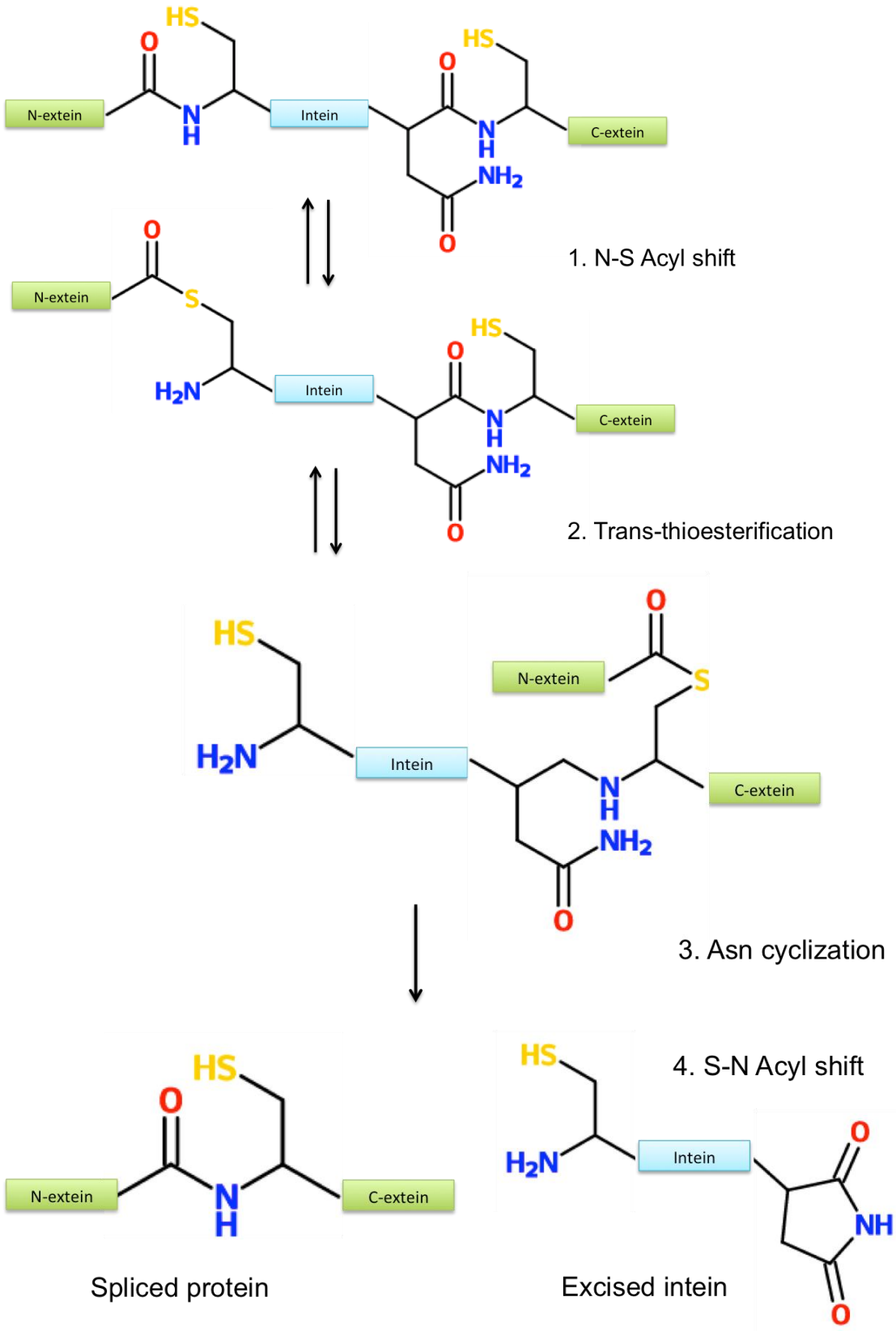


Figure 1 – Schematic illustration of the different steps in the protein splicing mechanism carried out by inteins. The protein splicing mechanism is summarized in four steps. In the first step **(1)**, the inteins are associated due to the N-S acyl shift **(2)** where a linear thioester intermediate is created by the N-S acyl shift. By cyclization of the Asn **(3)** at the C-terminal, the intein will be cleaved. In the last step **(4)** a peptide bond is formed between the exteins through a S-N acyl shift.

The inteins from the DnaE family have the potential for many applications in chemistry and biology. However, the general uses of inteins are limited due to some sequence preferences in order to splice properly. For instance, the C-extein should contain "CFN" tripeptide sequences in order to splice properly and an "AEY" tripeptide in the N-extein. The CFN and AEY tripeptides will remain in the final protein product after the splicing which - in the case that the critical amino acids were not present in the original sequence and had to be introduced deliberately - creates a mutated protein. The presence of CFN in C-extein has shown to contribute to a more efficient trans-splicing as the cysteine initiates the branched thioester intermediate in the splicing reaction[10].

Inteins can be utilized in protein ligation applications such as Expressed Protein Ligation (EPL). In EPL, the inteins are used to ligate a peptide or molecule having –SH at its N-terminal[11]. Inteins (except DnaE) have not been commonly used in applications due to their slow reaction kinetics and low ligation yield. However, the discovery of the fast DnaE inteins proposes to overcome these difficulties and find use in novel applications[12].

EPL is the most widely used intein-based technology for site-specific modification of a protein. Artificially fused inteins can be used for generating recombinant protein thioester derivatives (at the C-terminal) for EPL[12]. Intercepting the thioester with the thiol group of cysteine at the N-terminal initiates protein ligation that generates a new peptide bond[13, 14].

1.4 Aim of study

This project aimed to develop a new and efficient method to produce active recombinant RNase A using a commonly used expression host, *E. coli*. As previously mentioned, recombinant production of proteins can be a challenge due to commonly observed aggregation of the foreign protein in *E. coli* and due to cytotoxicity of some recombinant protein.

RNase A is a protein with a complex structure due to four disulfide bonds and faces the insolubility and cytotoxicity issues when expressed in *E. coli*. This is why it is of value to find an efficient method for producing RNase A recombinantly without using costly techniques or multiple steps that offer low protein yield and require heavy workload. With an efficient method of producing recombinant RNase A, a higher yield of the protein can be achieved.

The Ph.D. student supervising this project generated the needed DNA designs that are used in the study. My goal is to find optimal conditions of expression and lysis to obtain RNase A-intein fusion proteins in soluble form. After successful expression

and cell lysis, selected fusion proteins are purified. Afterward, the proteins will be tested by *in vitro* trans- and cis-splicing to investigate the ability to create active RNase A using inteins fused with inactive truncated RNase fragments.

1.5 Project Description

In this project, a proposed novel and an efficient way of producing recombinant RNase A was tested. To overcome the insolubility and cytotoxicity issue when producing recombinant RNase A in *E. coli*, the gene encoding full-length RNase A was split and fused with inteins at the cDNA level. The gene fragments of a truncated RNase A fused with intein was expressed in *E. coli* and tested for solubility of created fusion proteins. Soluble fusion protein was tested for their ability to generate active RNase A through two different strategies that are further described in this section.

In both strategies, the gene encoding for RNase A was divided at an amino acid position to create two inactive halves. The first strategy proposes to synthesize an active mutant RNase A variant by trans-splicing from its inactive fragments comprising both half of the protein sequence fused via sequence-optimized splicing sites with split intein sequences as shown in Figure 2.

In the second strategy, the RNase A was split right before an internal cysteine residue close to its C-terminal end at DNA level and fused with an artificially joined split intein required for EPL with a small part of RNase which was not fused with intein (to be produced e.g. by standard peptide synthesis). Both strategies divide the gene encoding RNase A at an amino acid position, which are separate from the active residues needed for the active enzyme. Thus, the fusion proteins will be inactive and have no cytotoxicity associated with RNase A.

DnaE split inteins are a family of split inteins that come from cyanobacteria. The inteins used in this project belong to the DnaE family of inteins and were shown to have a high efficiency and reaction rate[10]. Three pairs of the DnaE split intein from different species X^N-X^C, Y^N-Y^C and Z^N-Z^C (each pair containing Int^N and Int^C) was tested in the first strategy and two DnaE cis-splicing generating "fused" inteins A and B were used in the second strategy. They are named X, Y, Z, A and B to maintain confidentiality.

The strategies involved trying different *E. coli* strains, expression media and lysis buffers to test the solubility of truncated RNase A-intein fusion proteins.

1.5.1 Strategy 1 – Synthesis of mutant RNase A using trans-splicing/split inteins

Strategy 1 aims to obtain an active mutant of RNase A by ligating its two inactive parts. The two inactive parts of RNase A were fused at gene level with DNA encoding for different DnaE split intein pairs. The two fused proteins, each containing one of the split intein pairs, were expressed separately in *E. coli.*

The codon optimized DNA encoding for RNase A was split into two fragments containing amino acid residues 1-65 and 66-124 (see Figure 2) of RNase A. The fragment comprising residues 1-65 was fused with one of the split intein pairs Int^N and residues 66-124 was fused with Int^C.

When the fused genes have been expressed as protein and purified from *E. coli*, the fusion proteins will be mixed. Due to the affinity of the split inteins, they will undergo trans-splicing and excise themselves out and join the amino acid on both sides of the intein pairs, thus ligating the two halves of RNase A. After expression and trans-splicing, an active mutated RNase A protein will be generated as shown in Figure 2. At position 65, the DNA sequence coding for residues EY was added after the native A residue coding as the AEY sequence is needed on the N-extein for efficient trans-splicing. At position 67, the DNA encoding for the K residue within the native CKN sequence was mutated to F as CFN is required at the C-extein for efficient trans-splicing.

The RNase A created will have a modified AEYCFN instead of the native ACKN (see Figure 2). These mutations are proposed to cause minimal disorder of the RNase A structure as the added or changed amino acids are within a flexible loop region and not involved in the active site[15].

The new design that contains the mutated residues AEY and CFN will cause minimal disorder of the RNase A structure and its activity will remain since the mutations are located away from residues that are involved in the catalysis (His12, His119, Lys41, Asp121, Gln11, and Phe120)[15].

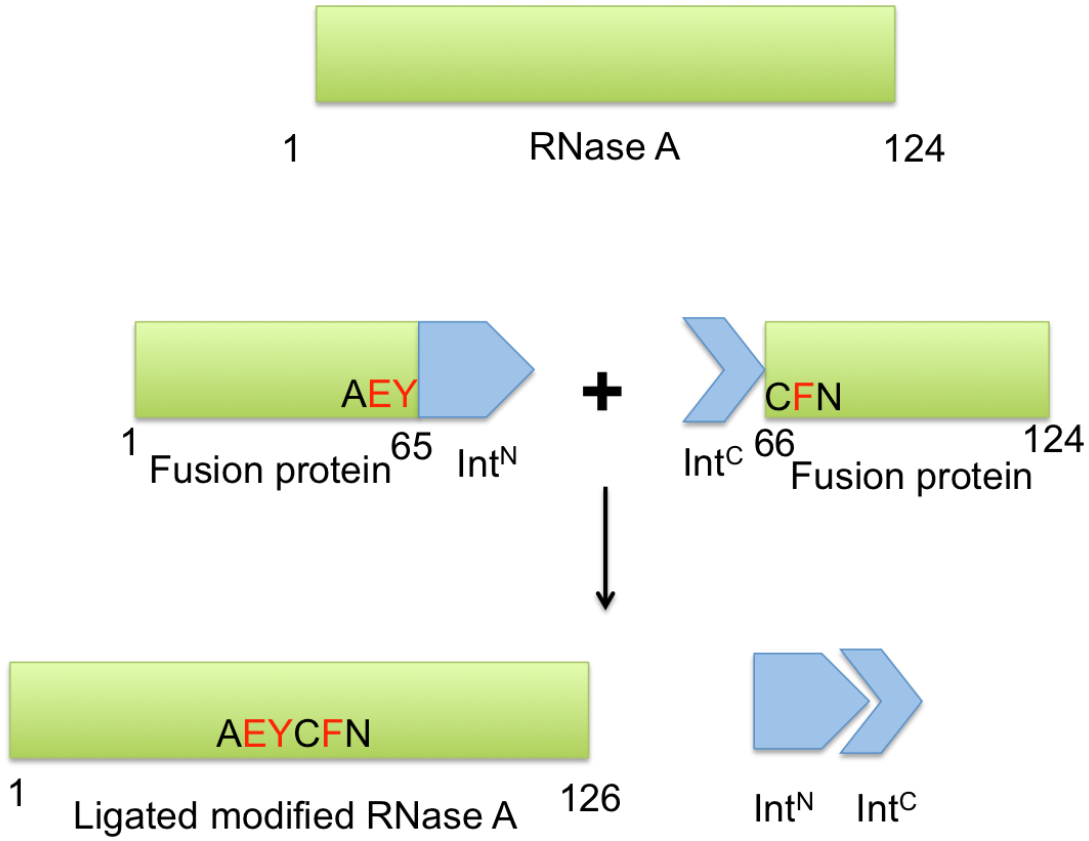


Figure 2 – Illustration of the fused gene encoding for RNase A and split intein used in the first strategy. The gene for RNase A was divided and fused with split inteins Int^N and Int^C. Each fusion protein was expressed separately and mixed. Due to the affinity of the split intein, the RNase A parts were ligated through a peptide bond as the split inteins will excise themselves. The protein contains mutations (colored in red).

1.5.2 Strategy 2 – Synthesis of recombinant RNase A by expressed protein ligation using cis-splicing

In strategy 2 the gene encoding for RNase A was split in a different position compared to strategy 1. As shown in Figure 3, the gene was split between amino acid positions 110 and 111. The fragment with residues 1-110 will most likely not be cytotoxic when expressed in *E. coli* as it does not contain all the amino acids needed for the catalytic activity. At the DNA level, the fragment with residues 1-110 has its C-terminal fused to intein A or B. Intein A and B represent a fusion of a split intein pair from the DnaE family for EPL. In other words, each fused split intein A and B will form one whole fragment that will perform cis-splicing (indicated as Int in Figure 3), which is the goal of the second strategy. After expression and purification of RNase A residues 1-110 with a fused intein A or B, a synthetic peptide representing residues 111-124 of the RNase A was added together with a thiol reagent which will initiate the cis-splicing as it generates a thiolysis reaction. The intein will excise itself and a reactive thioester group will be generated at the C-terminal of the fragment with

residues 1-110. The thioester group will react with the thiol group of the N-terminal cysteine in the peptide with RNase A residues 111-124, thus generating a 124 residue fragment that represents the full-length RNase A.

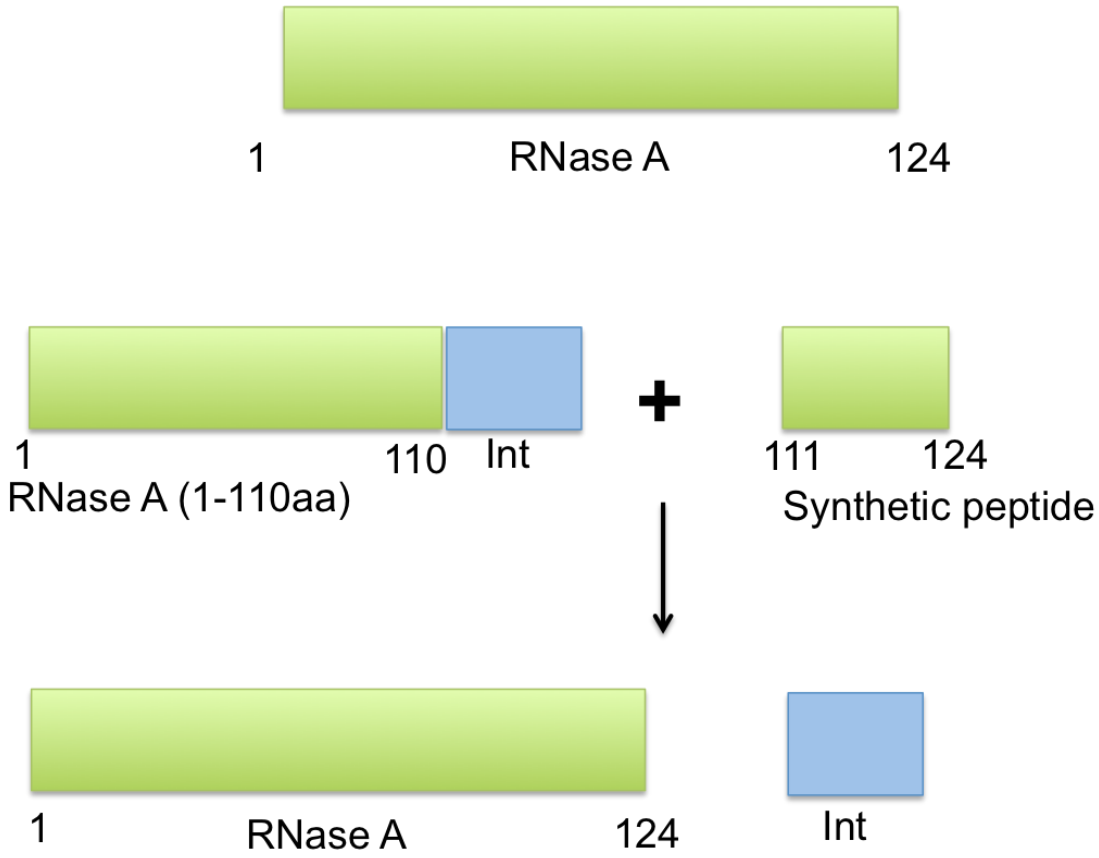


Figure 3 – Illustration of the fused gene encoding for RNase A in the second strategy. The gene encoding residues 1-110 was fused with the gene encoding intein A and B (Int). After expression, a synthetic peptide (111-124) was added. The cis-splicing will be carried out as a thiol group was added and the intein will excise itself and the RNase A (1-110) will be ligated with the synthetic peptide (111-124) generating the full length RNase A.

2 Materials and Methods

2.1 Materials

2.1.1 *E. coli* Strains

Numerous strains were tested in order to find a strain that is suitable to express the different RNase A-intein fusion proteins in soluble form. Table 1 illustrates the strains that have been tested in this project. These strains have been selected based on their features, which includes the ability to express proteins in soluble form, controllable expression level, chaperones co-expression, tight control of expression and correct disulfide bond formation. All these stains belong to lambda DE3 that expresses T7 RNA polymerase controlled by the *lac* promoter.

Table 1 – *E. coli strains* used for test expression. To find an optimal *E. coli* strain for producing an RNase A fusion protein, numerous strains have been tested for expression. The property and specific antibiotic resistance for each tested strain are presented in the table.

Strain	Properties	Antibiotic resistance
SoluBL21 (amsbio)	An optimized strain for expressing insoluble protein in soluble form[A].	None
Arctic Express (Agilent technologies)	A strain engineered that contains chaperon and to improve the expression of protein at a low temperature, which helps the folding of the protein[B].	Gentamycin
Tuner (EMB millipore)	LacZY deletion mutation, which allows different levels of expression depending on the IPTG concentration. Low level of expression may enhance the solubility[C].	None
Shuffle (New England Biolabs)	Enhance the disulfide bond formation in the cytoplasma.	Streptomycin
Rosetta2 plyss (EMB millipore)	A strain suitable for expression of eukaryote proteins that contains rare codons used in *E. coli*[D].	Chloramphenicol
BL21 Star (Thermo Fischer)	Offer enhanced mRNA stability[E].	None

[A] http://www.genlantis.com/solubl21-competent-e-coli.html
[B] http://www.genomics.agilent.com/article.jsp?pageId=468&_requestid=127115
[C] http://www.merckmillipore.com/INTL/en/product/tunerde3plyss-competent-cells,EMD_BIO-70624
[D] https://www.neb.com/products/c3026-shuffle-t7-competent-e-coli
[E] https://www.lifetechnologies.com/order/catalog/product/C602003

2.1.2 Intein Constructs

In the first strategy, where trans protein splicing is used to synthesize RNase A, three pairs of split inteins have been tested (X^N-X^C, Y^N-Y^C and Z^N-Z^C). In the second strategy where cis protein splicing is used, two inteins engineered for EPL have been tested (A and B). All the constructs are listed in Table 2.

To compare the expression and solubility, a positive and a negative protein control will be used in the experiments. The positive control is a protein termed peredox, which is an NADH binding protein fused with a green fluorescent protein variant. The peredox protein is known to be soluble. The full-length RNase A will be used as a negative control since it is known to be insoluble when expressed in *E. coli*.

The fused fragments are cloned into a pET27b plasmid. The vector contains genes coding for antibiotic resistance to kanamycin and the *lacI* gene from the lac operon that codes for the lac repressor, which binds the T7 promoter.

Table 2 – Fused RNase A-intein constructs. The table shows the fused RNase-intein fusion protein tested for each strategy.

Construct name (N=N-terminal intein C=C-terminal intein)	Amino acid residues	Strategy	Mass of protein (kDa)
X^N	1-65	1	20.5
Y^N	1-65	1	20.0
Z^N	1-65	1	21.7
X^C	66-124	1	11.4
Y^C	66-124	1	11.5
Z^C	66-124	1	11.5
A	1-110	2	27.0
B	1-110	2	27.0
Peredox fluorescent protein		Positive control	70.0
Full-length RNase A		Negative control	13.6

2.2 Methods

2.2.1 Transformation using heat shock

Heat shock was used for transforming the plasmids containing truncated RNase A-intein fragments. 0.5µl of fused DNA was mixed with 50µl competent cells. The suspension was kept on ice for 30 minutes. The suspension was heat shocked at 42°C for 30 seconds. Afterward, the suspension was put on ice again for 2 minutes. 800µl of heated SOC medium was added and the culture grew on a shaker for 1 hour at 37°C. 50µl of the suspension was plated on LB agar plates with proper antibiotic at 37°C overnight. Protocols for all the growing solutions and agar plates used in the project are listed in Table 3.

Table 3 – Protocol of the solutions, growing media and agar plates used in the study.

Solution / **Recipe**	SOB-Super optimal broth Volume: 500mL	LB-Lysogeny broth Volume: 1L pH 7	2xYT Yeast extract & Tryptone Volume: 2L	AIM-Autoinductin medium	LB-Lysogeny Agar 1L pH 7
	2.5g Yeast extract	10g Tryptone	32g Tryptone	16g Tryptone	10g Tryptone
	10g Tryptone	5g Yeast extract	20g Yeast extract	10g Yeast extract	5g Yeast extract
	1ml 5M NaCl	10g NaCl	10g NaCl	3.3 g $(NH_4)2SO_4$	10g NaCl
	1.25 ml 1M KCl			6.8g KH_2PO_4	15g Agar
	5ml 1M $MgCl_2$			0.5g Glucose	
	5ml 1M $MgSO_4$			2g α-lactose	
				0.15g $MgSO_4$	

2.2.2 Protein expression

Small scale
After growing the transformed bacteria on an LB-agar plate, a colony was picked and added to a test tube with 3.5 ml of growing media (LB, 2xYT, AIM) containing specific antibiotics for different strains (see Table 1): Kanamycin 50µg/ml, chloramphenicol 35µg/ml, streptomycin 5µg/ml and gentamycin 10µg/ml. The tube was incubated at 37°C for three hours. The protein was expressed by inducing the cells with 0.25mM IPTG after OD_{600} approximately reached 0.5. Tuner cells were induced with 10µM IPTG. After induction of IPTG the protein was expressed for 16 hours at 18°C.

Large scale
In the purification step, affinity chromatography was used to purify the protein. In this step, a large-scale protein expression was done. The large-scale expression was done expressing the protein with Arctic Express (DE3). Adding 5ml from overnight culture to 500ml of 2xYT medium in a flask with 5µg/ml of kanamycin and 5µl/ml of gentamycin. When OD_{600} reached 0.6 the protein was expressed by induction with 0.2 mM IPTG at 15°C for 16 hours.

2.2.3 Optimizing cell lysis to check the solubility of fusion proteins

After the protein expression, the cells (200µl of *E. coli*) were centrifuged at 8000g for 10 minutes. The supernatant was removed and different lysis methods as listed below were applied to check the efficiency of lysis. The conditions of the lysis methods are presented in Table 4. The lysis buffer used contained 50mM Tris and 50 mM NaCl at pH 7.5.

Table 4 – Conditions tested for the different lysis methods.

Easylyse™	Bug Buster™	Sonication Amplitude 50% On/Off: 50/50	Freeze thaw	IGEPAL[6]	*n*-Dodecyl β-D-maltoside	Glass beads (1:1 ratio)
20 µl	20 µl	10min	3x	1%	1%	5 min
40 µl	40 µl	5 min	6x	2%	2%	10 min
80 µl	80 µl	2 min				

Easylyse™
0.1ml of buffer was used to resuspended the pellet obtained from 200µl of culture. 20µl, 40µl and 80µl of the enzyme mix Easylyse were added to three pellet suspensions. The mixes were incubated for 5 minutes and centrifuged at 8000g for 10 minutes. The supernatant was saved in a new tube.

Bugbuster™
0.1ml of buffer was used to resuspend the pellet obtained from 200µl of culture. 20µl, 40µl and 80µl of Bugbuster were added to three different pellets.
The mixes were resuspended and incubated for 10-20 minutes. The soluble fraction was then removed by centrifugation at 16000g for 20 min at 4°C. The soluble fraction was removed to a new tube and the insoluble fraction was kept in the same tube with an equal amount of buffer as used for resuspension of the pellet.

Freeze thaw
The procedure of freeze thaw was done by repeatedly freezing the pellet in liquid nitrogen and putting it on 37°C for 2 minutes. For one pellet, the procedure was repeated three times and for another pellet six times.

IGEPAL™
1 % and 2% IGEPAL was added for resuspension of two different pellets.

n-Dodecyl β-D-maltoside™
1 % and 2% *n*-Dodecyl β-D-maltoside was added for resuspension of two different pellets.

Glass beads
After centrifugation two pellets were mixed with glass beads to a ratio of 1:1. Two pellets were vortexed with the time span 5 and 10 minutes, respectively using VortexGenie.

Sonication
Three pellets were mixed with lysis buffer, lysozyme, EDTA, and protease. The mixes were incubated for different time spans and the suspensions were removed to a beaker that was placed on ice in the sonicator. The three suspensions were then sonicated with 50% amplitude for 2.5 and 10 minutes, respectively.

2.2.4 Protein solubility test

The lysis step was followed by a solubility test. The lysed cells were centrifuged at 8000g for 10 minutes. The supernatant containing the soluble fraction was transferred to a new tube. The pellet contained the insoluble fraction that was resuspended with an equal volume of buffer. 25µl of each fraction was mixed with 25µl Laemmli protein sample buffer and heated for 10 minutes at 90°C. 25-30µl of the mix and 5µl of the ladder (Roti® mark standard, 14-212kDa) were added to the wells of the SDS-PAGE gel. Table 5 shows the protocol for the gels. The gel ran for 10 minutes at 100 volts and after that, the voltage was increased to 150 V for approximately 40-50 minutes.

Table 5 – Protocol for 1.5 mm SDS-PAGE gels.

14% Gel	6% Stacking gel
3.4 ml Milli-Q water	4.2 ml Milli-Q water
6.53 µl Acrylamide 30%	1 ml Acrylamide 30%
3.5 µl 1.5M Tris pH 8.8	2 µl 1.5M Tris pH 6.8
140 µl 10 % SDS-page	80 µl 10 % SDS-page
140 µl 10% APS	80 µl 10% APS
14 µl TEMED	8 µl TEMED

Staining

Immediately after electrophoresis, the gel was washed with distilled water. After washing, distilled water was added and heated up with the gel. Afterward, the gel was put on the shaker for 5 minutes. This procedure was done 3 times. Fixing solution was added so that it covered the gel and the gel was then put on a shaker for 1 h. The fixing solution was removed and coomassie solution (ROTI-BLUE) was added and the gel was put on the shaker overnight.

2.2.5 Chromatography HiTrap™ TALON® crude

The HiTrap™ TALON® crude 5 ml column that was used in the chromatography step was pre-packed with highly cross-linked agarose beads with an immobilized chelating group. The TALON ligand is a tetra-dentate chelator charged with cobalt ions (Co^{2+}).

Buffers

The chromatography step was repeated two times and the imidazole concentration of the wash buffers and elution buffer was varying for each attempt. Table 6 shows the different concentrations for each trial.

Table 6 – Tested imidazole concentration for the buffers used in the purification step.

Purification attempt / Buffer	First purification	Second purification
Wash buffer 1	20 mM Imidazole	20 mM Imidazole
Wash buffer 2	50 mM Imidazole	40 mM Imidazole
Elution buffer	5 mM Imidazole	20 mM Imidazole

Binding buffer	50mM Sodium phosphate, 300mM NaCl, 10mM Imidazole pH 7.4
Washing buffer 1	50 mM phosphate, 300 mM NaCl, Imidazole, pH 7.4
Washing buffer 2	50 mM phosphate, 300 mM NaCl, Imidazole, pH 7.4
Lysis Buffer	50 mM phosphate, 300 mM NaCl, 5 mM Imidazole, pH 8
Elution buffer	50 mM phosphate, 300 mM NaCl, Imidazole, pH 7.4

The peristaltic pump that was used in this experiment was filled with distilled water and connected to the chromatography system tubing. The flow rate was set at 1ml/minute. The column was washed with five times the column volume (CV) with distilled water. The column was equilibrated with three times the CV with binding buffer. 30ml of the soluble fraction that was saved from the cell lysis was then added with a pump. Five times the CV of wash buffer 1 was added followed by three CV of wash buffer 2. The protein was eluted with an elution buffer with isocratic elution. In order to remove the imidazole from the protein, the elution was added to a PD-10 desalting column and the elution from the PD-10 was collected and pooled.

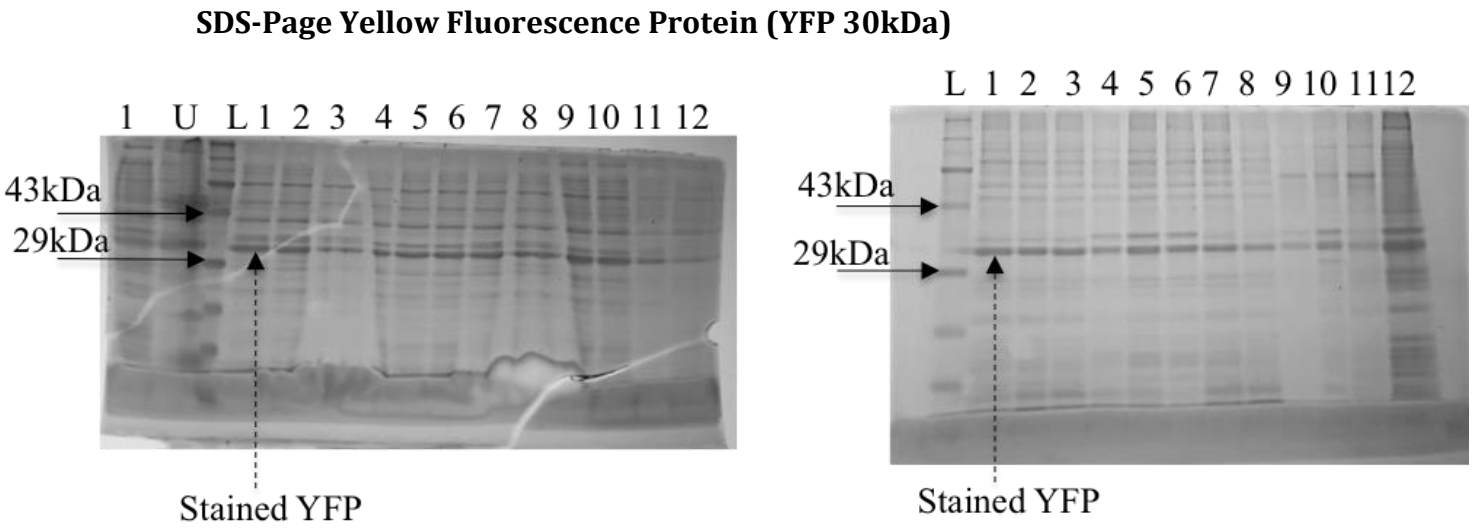
2.2.6 Thiolysis of purified RNase A (1-110)-Intein fusion protein A and B

In this step, the protein splicing was initiated for the expressed and purified fusion proteins RNase A (1-110)-A and RNase A (1-110)-B. 10μM of each fusion protein were mixed with 100mM Mesna (sodium thiomethoxide), 5mM TCEP and 1mM EDTA. The reaction suspension was incubated at 30 °C for 24h. The day after, the synthetic peptide was added (RNase A residues 111-124).

3 Results

3.1 Optimization of lysis method

This experiment was done in order to find a lysis condition that completely lyses the cell pellet for analysis of soluble protein. As shown in Table 4 in the Method section, several lysis methods and conditions were tested with the BL21Star strain expressing Yellow Fluorescent Protein (YFP). The goal of testing all the lysis methods and conditions was to find an optimized lysis method and conditions that could be applied to the RNase A-intein fusion protein. Figure 4 shows the result from expression and lysis of YFP in BL21Star.



A) Insoluble Fraction **B) Soluble Fraction**

Figure 4 – SDS-PAGE results from test expression of YFP with BL21Star and tested lysis methods. Test expression of YFP in BL21Star using different lysis methods. Image A shows the soluble fraction and image B shows the insoluble fraction. The YFP is indicated with black dotted arrows in the images and lanes containing YFP purified with different lysis methods: Bugbuster 10µl (Lane 1), Bugbuster 20µl (Lane 2), Bugbuster 50µl (Lane 3), Easylyse 20µl (Lane 4), Easylyse 30µl (Lane 5), Easylyse 40µl (Lane 6), β-D-maltoside 10 µl (Lane 7), β-D-maltoside 30 µl (Lane 8), Glass beads 2.5 minutes (Lane 9), Glass beads 5 minutes (Lane 10), Sonication 2 minutes (Lane 11), Sonication 4 minutes (Lane 12). Lane U is uninduced sample and lane L is the ladder.

YFP was successfully expressed in BL21Star due to strongly stained protein bands in the soluble and insoluble fraction shown in Figure 4. The black arrows indicate the stained YFP protein in image A and B that corresponds to its mass of 30kDa. Several lysis methods gave strong bands. These lanes are 1-7 and 12 in image A and correspond to Bugbuster 10µl, 20µl and 50µl, Easylyse 20µl, 30µl and 40µl, β-D-maltoside 10µl and sonication 4 minutes. Lanes showing weak (or almost no) bands are lanes 8-11 in image A that correspond to β-D-maltoside 30µl, Glass beads 2.5 minutes, Glass beads 5 minutes and Sonication 2 minutes.

Based on the soluble fraction in image B, Easylyse gave the strongest band, thus good lysis results. Sonication for 4 minutes and Bugbuster also gave strong bands, however, Bugbuster changed the pH value by half a unit. Glass beads were hard to separate from the suspension after usage. IGEPAL precipitated and β-D-maltoside did not show any significant lysis. Thus, lysis using Easylyse seems like an optimal lysis detergent in order to obtain RNase A-intein fusion protein in soluble form.

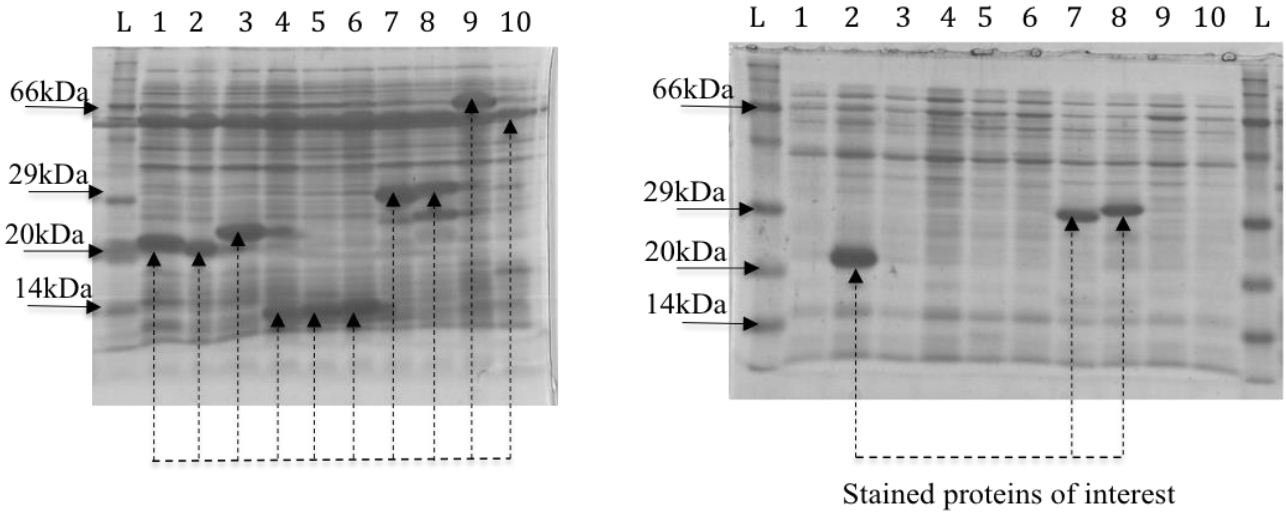
3.2 Test expression with several strains

All the designed RNase A-intein fusion proteins were expressed in different *E. coli* strains proposing to have different features to promote solubility of recombinant proteins. The cells were grown in 2xYT media (except SoluBL21) with a kanamycin concentration of 50µg/ml and cell-specific antibiotic (see Table 1) to OD_{600} of 0.6 and induced with 0.2mM IPTG (except for Tuner which was induced with 10µM IPTG and expressed for 12 hours at 20°C). A soluble fraction was obtained by suspending the pellet obtained from 200µl of cells in a 100µl buffer (50 mM Tris pH 7.5, 50 mM NaCl) containing 30µl lysis agent Easylyse. The fusion protein includes RNase A (1-66)-X^N, RNase A (1-66)-Y^N, RNase A (1-66)-Z^N, RNase A (67-124)-X^C, RNase A (67-124)-Y^C, RNase A (67-124)-Z^C, RNase A (1-110)-A and RNase A (1-110)-B (see Table 2 in Section 2.1.2). The lysis methods and conditions for the different strains are shown in Table 7.

Table 7 – Lysis condition for each of the different strains.

Strain / Lysis Condition	**Arctic Express**	**Tuner**	**Rosetta**	**SoluBL21**	**Shuffle**
Easylyse	25 µl	30 µl	30 µl	25 µl	25 µl
Bugbuster	5 µl	-	-	5 µl	-
Lysis buffer	100 µl	100 µl	100 µl	100 µl	100 µl

3.2.1 Arctic Express (DE3)

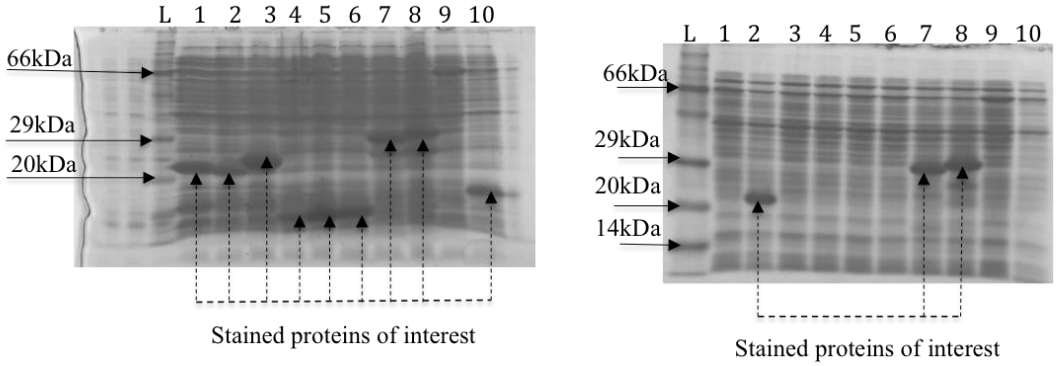


A) Total Fraction **B) Soluble Fraction**

Figure 5 – SDS-PAGE results from test expression of RNase A-intein fusion proteins with Arctic Express. The expression of the eight RNase A-intein constructs, and of a positive and negative control was done in Arctic Express. The cells were lysed with 100µl lysis buffer, 25µl Easylyse and 5µl Bugbuster. The lanes contain Ladder (Lane L), RNase A (1-66)-X^N (Lane 1), RNase A (1-66)-Y^N (Lane 2), RNase A (1-66)-Z^N (Lane 3), RNase A (66-124)-X^C (Lane 4), RNase A (66-124)-Y^C (Lane 5), RNase A (66-124)-Z^C (Lane 6), RNase A (1-110)-A (Lane 7), RNase A (1-110)-B (Lane 8), Peredox fluorescent protein as positive control (Lane 9) and full-length RNase A (1-124) as negative control (Lane 10). Gel image A contains the soluble fraction and image B contains the total fraction.

Expression of all RNase A-intein fusion proteins was successful due to the strongly stained proteins in Figure 5A showing the total fraction. Only three of the fusion proteins were found in soluble form as shown by the black arrows in Figure 5B. These fusion proteins are RNase A (1-66)-Y^N, RNase A (1-110)-A and RNase A (1-110)-B corresponding to lanes 2,7 and 8. The stained protein band in lane 2 corresponds to the mass of RNase A (1-66)-Y^N that is ~21 kDa. The stained proteins in lanes 7 and 8 correspond to the mass of RNase A (1-110)-A and RNase A (1-110)-B that are 29kDa.

3.2.2 Tuner

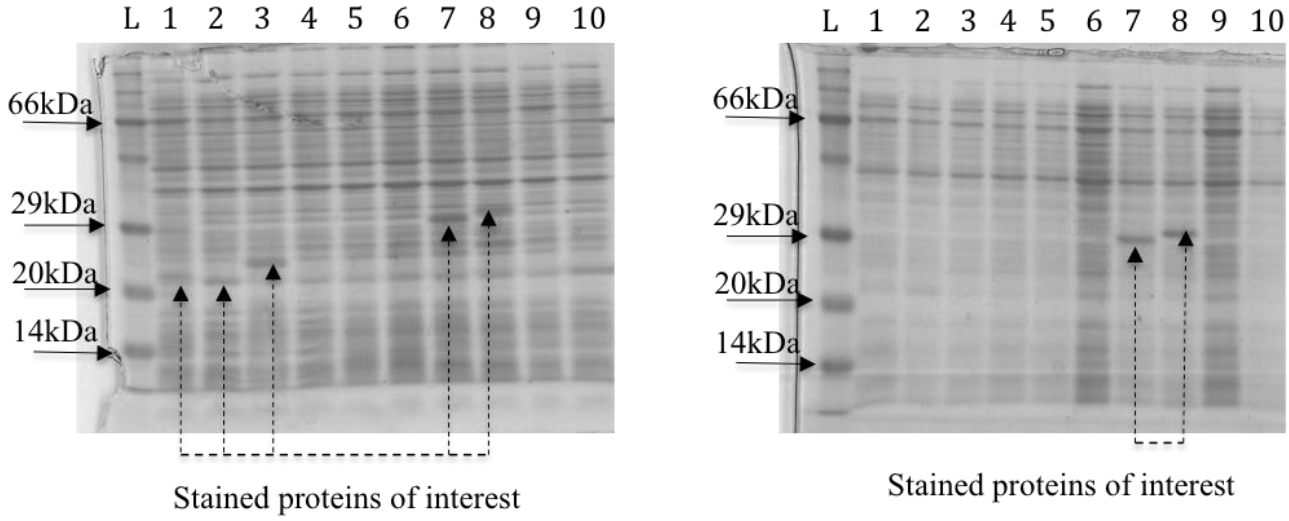


A) Total Fraction **B) Soluble Fraction**

Figure 6 – SDS-PAGE results from test expression of RNase A-intein fusion proteins with Tuner. The expression of the eight RNase A-intein constructs, and of a positive and negative control was done in Tuner. The cells were lysed with 100µl lysis buffer and 30µl Easylyse. The lanes contain Ladder (Lane L), RNase A (1-66)-X^N (Lane 1), RNase A (1-66)-Y^N (Lane 2), RNase A (1-66)-Z^N (Lane 3), RNase A (66-124)-X^C (Lane 4), RNase A (66-124)-Y^C (Lane 5), RNase A (66-124)-Z^C (Lane 6), RNase A (1-110)-A (Lane 7), RNase A (1-110)-B (Lane 8), Peredox fluorescent protein as positive control (Lane 9) and full-length RNase A (1-124) as negative control (Lane 10). Gel image A contains the soluble fraction and image B contains the total fraction.

Expression of all RNase A-intein fusion proteins was successful due to the strongly stained proteins in Figure 6A showing the total fraction. Only three of the fusion proteins were found in soluble form as shown by the black arrows in Figure 6B. These fusion proteins are RNase A (1-66)-Y^N, RNase A (1-110)-A and RNase A (1-110)-B corresponding to lanes 2,7 and 8. The stained protein band in lane 2 corresponds to the mass of RNase A (1-66)-Y^N that is ~21kDa. The stained protein bands in lane 7 and 8 correspond to the mass of RNase A (1-110)-A and RNase A (1-110)-B that are 29kDa.

3.2.3 Rosetta

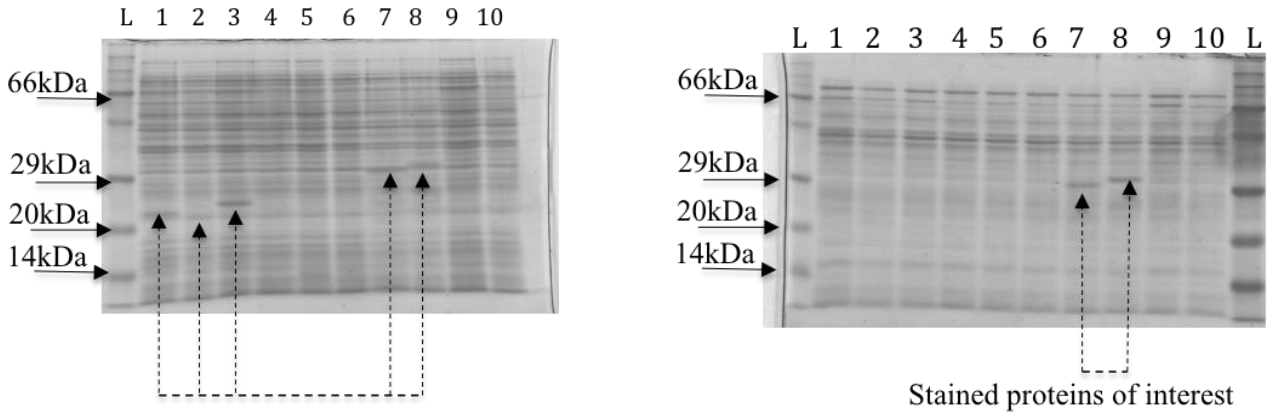


A) Total Fraction **B) Soluble Fraction**

Figure 7 – SDS-PAGE results from test expression of RNase A-intein fusion proteins with Rosetta. The expression of the eight RNase A-intein constructs, and of a positive and negative control was done in Rosetta. The cells were lysed with 100µl lysis buffer and 30µl Easylyse. The lanes contain Ladder (Lane L), RNase A (1-66)-X^N (Lane 1), RNase A (1-66)-Y^N (Lane 2), RNase A (1-66)-Z^N (Lane 3), RNase A (66-124)-X^C (Lane 4), RNase A (66-124)-Y^C (Lane 5), RNase A (66-124)-Z^C (Lane 6), RNase A (1-110)-A (Lane 7), RNase A (1-110)-B (Lane 8), Peredox fluorescent protein as positive control (Lane 9) and full-length RNase A (1-124) as negative control (Lane 10). Gel image A contains the soluble fraction and image B contains the total fraction.

Expression of RNase A-intein fusion proteins was only accomplished for RNase A (1-66)-X^N (Lane 1), RNase A (1-66)-Y^N (Lane 2), RNase A (1-66)-Z^N (Lane 3), RNase A (1-110)-A (Lane 7) and RNase A (1-110)-B (Lane 8) as indicated by the black arrows in Figure 7A representing the total fraction. Only two of the fusion proteins were found in soluble form as shown by the black arrows in Figure 7B. These fusion proteins are RNase A (1-110)-A and RNase A (1-110)-B corresponding lane 7 and 8. The stained protein bands in lanes 7 and 8 correspond to the mass of RNase A (1-110)-A and RNase A (1-110)-B that are 29kDa.

3.2.4 SoluBL(DE3)21

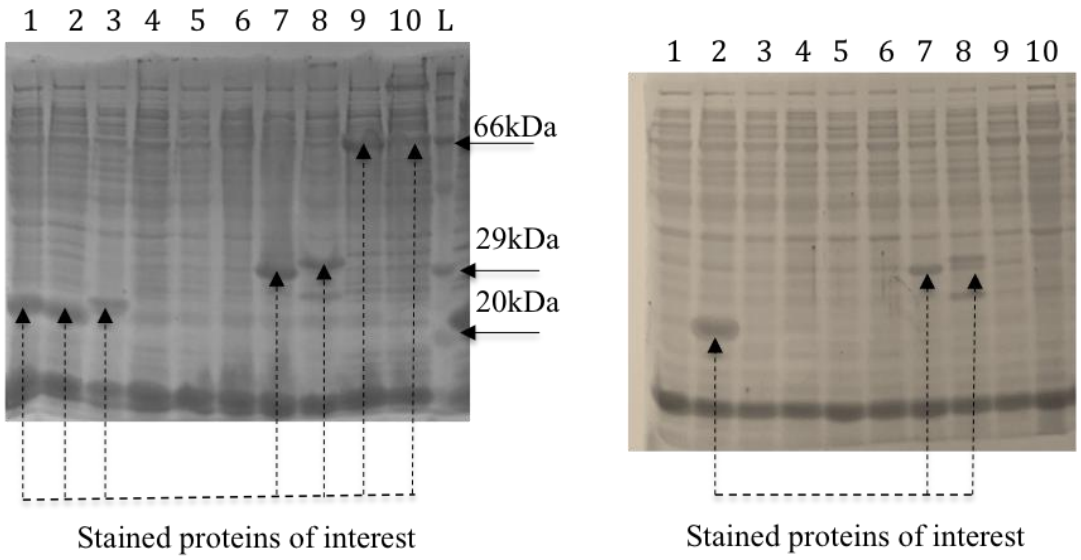


A) Total Fraction **B) Soluble Fraction**

Figure 8 – SDS-PAGE results from test expression of RNase A-intein fusion proteins with SoluBL(DE3)21. The expression of the eight RNase A-intein constructs, and of a positive and negative control was done in SoluBL21. The cells were lysed with 100µl lysis buffer, 25µl Easylyse and 5µl Bugbuster. The lanes contain Ladder (Lane L), RNase A (1-66)-X^N (Lane 1), RNase A (1-66)-Y^N (Lane 2), RNase A (1-66)-Z^N (Lane 3), RNase A (66-124)-X^C (Lane 4), RNase A (66-124)-Y^C (Lane 5), RNase A (66-124)-Z^C (Lane 6), RNase A (1-110)-A (Lane 7), RNase A (1-110)-B (Lane 8), Peredox fluorescent protein as positive control (Lane 9) and full-length RNase A (1-124) as negative control (Lane 10). Gel image A contains the soluble fraction and image B contains the total fraction.

Expression of RNase A-intein fusion proteins was only accomplished for RNase A (1-66)-X^N (Lane 1), RNase A (1-66)-Y^N (Lane 2), RNase A (1-66)-Z^N (Lane 3), RNase A (1-110)-A (Lane 7) and RNase A (1-110)-B (Lane 8) as indicated by the black arrows in Figure 8A representing the total fraction. Only two of the fusion proteins were found in soluble form as shown by the black arrows in Figure 8B. These fusion proteins are RNase A (1-110)-A and RNase A (1-110)-B corresponding lanes 7 and 8, although they show extremely weak staining. The stained protein bands in lanes 7 and 8 correspond to the mass of RNase A (1-110)-A and RNase A (1-110)-B that are 29kDa.

3.2.5 Shuffle



A) Total Fraction **B) Soluble Fraction**

Figure 9 – SDS-PAGE results from test expression of RNase A-intein fusion proteins with Shuffle. The expression of the eight RNase A-intein constructs, and of a positive and negative control was done in Shuffle. The cells were lysed with 100µl lysis buffer and 25µl Easylyse. The lanes contain Ladder (Lane L), RNase A (1-66)-X^N (Lane 1), RNase A (1-66)-Y^N (Lane 2), RNase A (1-66)-Z^N (Lane 3), RNase A (66-124)-X^C (Lane 4), RNase A (66-124)-Y^C (Lane 5), RNase A (66-124)-Z^C (Lane 6), RNase A (1-110)-A (Lane 7), RNase A (1-110)-B (Lane 8), Peredox fluorescent protein as positive control (Lane 9) and full-length RNase A (1-124) as negative control (Lane 10). Gel image A contains the soluble fraction and image B contains the total fraction.

Expression of RNase A-intein fusion proteins was only accomplished for RNase A (1-66)-X^N (Lane 1), RNase A (1-66)-Y^N (Lane 2), RNase A (1-66)-Z^N (Lane 3), RNase A (1-110)-A (Lane 7) and RNase A (1-110)-B (Lane 8) as indicated by the black arrows in Figure 9A representing the total fraction. RNase A (66-124)-X^C (Lane 4), RNase A (66-124)-Y^C (Lane 5), RNase A (66-124)-Z^C (Lane 6) might also be expressed but they are not seen in the respective lanes as they have equal mass corresponding to stained lysozyme at 14 kDa. Only three of the fusion proteins were found in soluble form as shown by the black arrows in Figure 9B. These fusion proteins are RNase A (1-66)-Y^N, RNase A (1-110)-A and RNase A (1-110)-B corresponding to lanes 2, 7 and 8. However, lane 7 and 8 show extremely weak protein staining bands. The stained protein band in lane 2 corresponds to the mass of RNase A (1-66)-Y^N that is ~21 kDa. The stained protein bands in lane 7 and 8 correspond to the mass of RNase A (1-110)-A and RNase A (1-110)-B that are 29kDa.

3.2.6 Evaluation of the tested strains

Arctic Express generated good expression results as seen from clearly stained bands (see Figure 6). Shuffle also gave strong protein bands, however, since Arctic Express is suitable for expression at low temperature it was chosen for large-scale expression. Expressing protein at a low temperature is preferable for insoluble proteins. Other strains such as Tuner showed to have problems in terms of a leaky promoter, Rosetta and SoluBL21 showed stained protein bands but these were slightly weaker than those observed in Arctic Express. Therefore, Arctic Express will be used for large-scale expression.

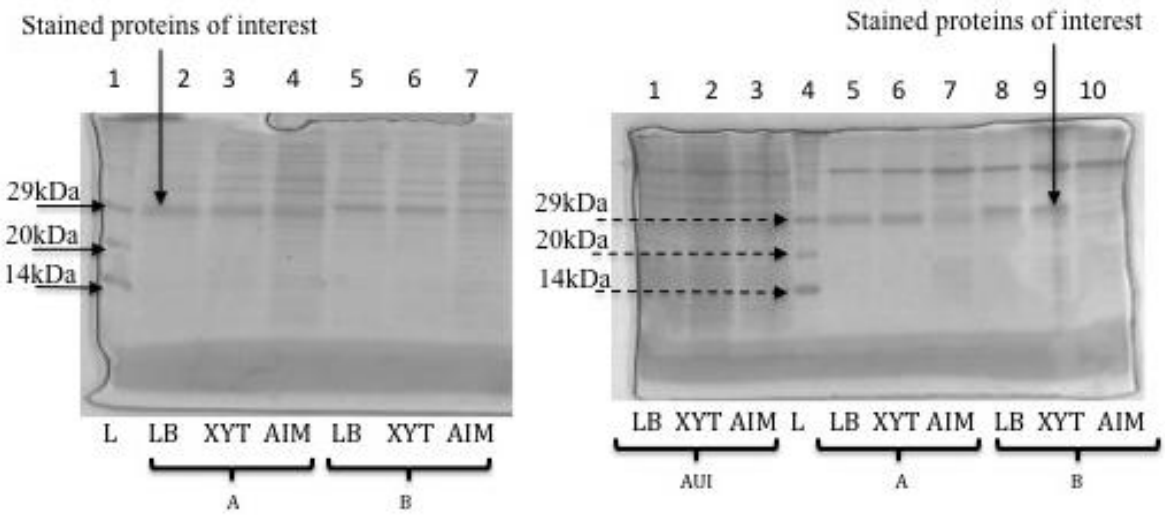
The fusion proteins that were accomplished to be expressed in soluble form were RNase A (1-66)-Y^N, RNase A (1-110)-A and RNase A (1-110)-B. For further experiments, the fusion protein RNase A (1-110)-A and RNase A (1-110)-B will be used, which are the fusion proteins of the second strategy.

The fusion protein RNase A (1-66)-Y^N was also generated in its soluble form, however, in order to continue the experiment and trans-splicing with its proper counterpart, RNase A (1-66)-Y^C will also be needed in its soluble form. Since the counterpart was not obtained in soluble form the trans-splicing reaction cannot be performed. None of the pairs of constructs of the first strategy were managed to be produced in soluble form and therefore, the experiments for strategy 1 could not be continued.

3.3 Expressed protein ligation of construct A and B

3.3.1 Test of media for large-scale test expression

Before purification of fusion proteins RNase A (1-110)-A and RNase A (1-110)-B for the Expressed Protein Ligation, several growth media were tested. Fusion proteins RNase A (1-110)-A and RNase A (1-110)-B were test-expressed in LB, 2xYT and AIM followed by lysis with 200µl of the sample, 150µl lysis buffer and 40µl Easylyse. Figure 10 shows the result from the solubility tests of the fusion protein using different growth media. The masses of the fusion proteins RNase A (1-110)-A and RNase A (1-110)-B are 29 kDa.



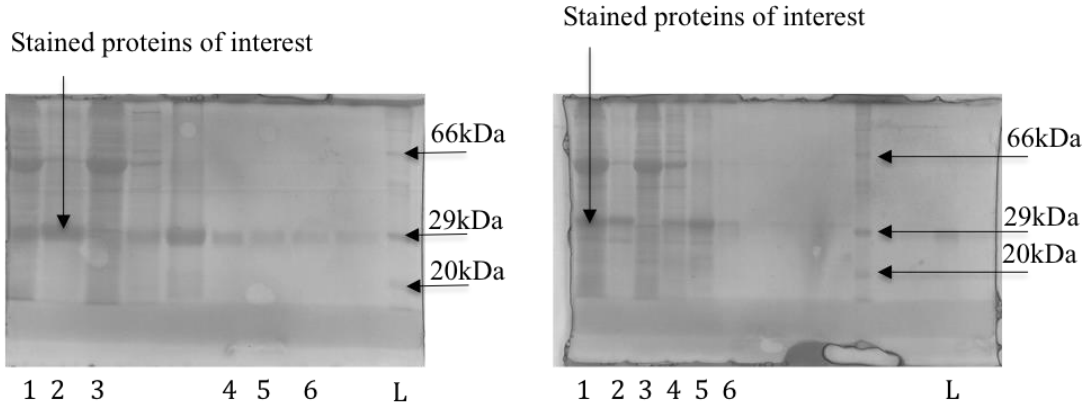
A) Soluble Fraction **B) Insoluble Fraction**

Figure 10 – SDS-PAGE results from test expression of fusion proteins RNase A (1-110)-A and RNase A (1-110)-B with Arctic Express in different growth media. The images show the result of the fusion proteins RNase A (1-110)-A and RNase A (1-110)-B tested with different growth media. The lanes are labeled with respective growth media LB, 2xYT and AIM. Lane containing RNase A (1-110)-A is labeled with A and lane containing RNase A (1-110)-B is labeled with B. Lane containing uninduced sample are labeled AUI. The masses of the constructs are 29 kDa.

The fusion designs RNase A (1-110)-A and RNase A (1-110)-B for EPL were well expressed in Arctic Express (DE) when growing the cells in the three growing media LB, 2xYT and AIM. Strongly stained protein bands are shown in lanes 2-4 representing RNase A (1-110)-A and lanes 5-6 representing RNase A (1-110)-B in Figure 10A, soluble fraction. The strong bands correspond to the fusion protein mass of 29kDa.

Thus, all three tested media yielded a soluble form of the fusion proteins. These fusion proteins are known to be partly soluble when expressed in cells growing in 2xYT that is shown in lanes 6 and 9 in Figure 10B representing the insoluble fraction. In contrast to the cells growing in AIM media, the band corresponding to the fusion proteins are weakly stained in the insoluble fraction as shown in lanes 7 and 10 in Figure 10 B representing the insoluble fraction. This indicates that AIM promotes solubility of these fusion proteins when expressed in Arctic Express (DE3).

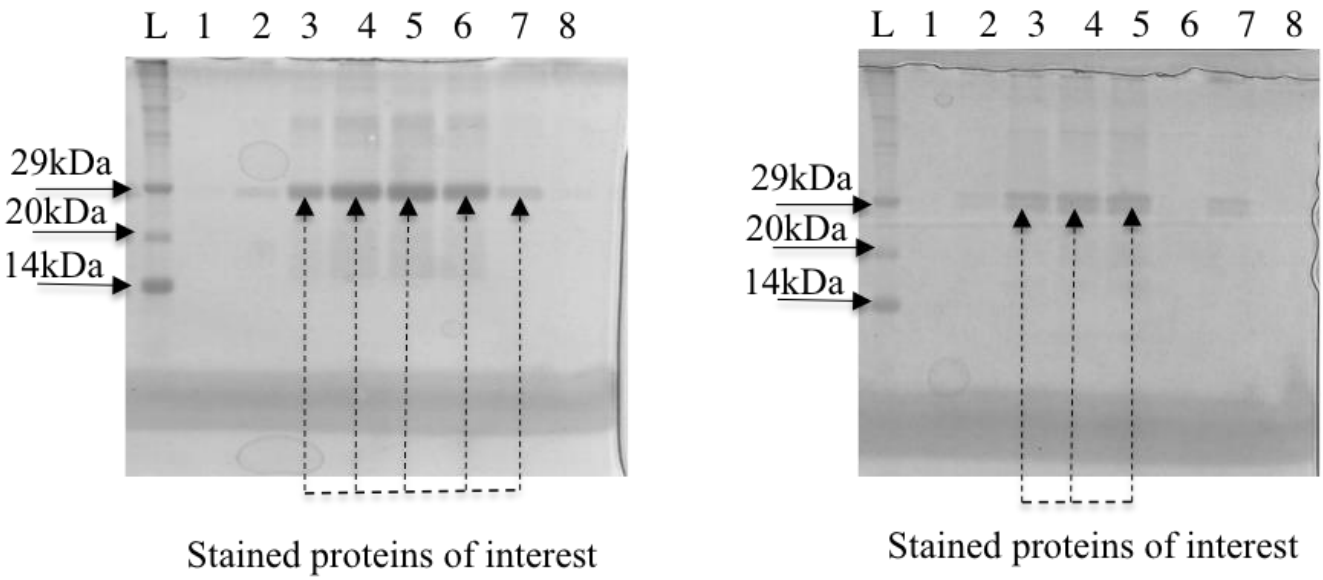
3.3.2 First purification attempt with HiTrap™ TALON® crude and PD-10 desalting column



A) Construct RNase A (1-110)-A **B) Construct RNase A (1-110)-B**

Figure 11 –SDS-PAGE results from the first purification attempt of RNase A (1-110)-A and RNase A (1-110)-B expressed in Arctic Express. The lane contains soluble fraction (Lane 1), the insoluble fraction (Lane 2), Flow through (Lane 3), Washing steps 1 and 2 (Lane 4 and 5), Elution (Lane 6) and Ladder (Lane L). The masses of the fusion proteins are 29 kDa.

The purification of the expressed fusion proteins RNase A (1-110)-A and RNase A (1-110)-B generated a small amount of protein as shown by the weakly stained protein bands in Figure 11 that correspond to the mass of 29kDa. The proteins were washed out already in the washing step, which explains the weak bands. During this purification attempt, the imidazole concentrations in the wash solutions 1 and 2 were 20mM and 50mM as recommended in the purification protocol.

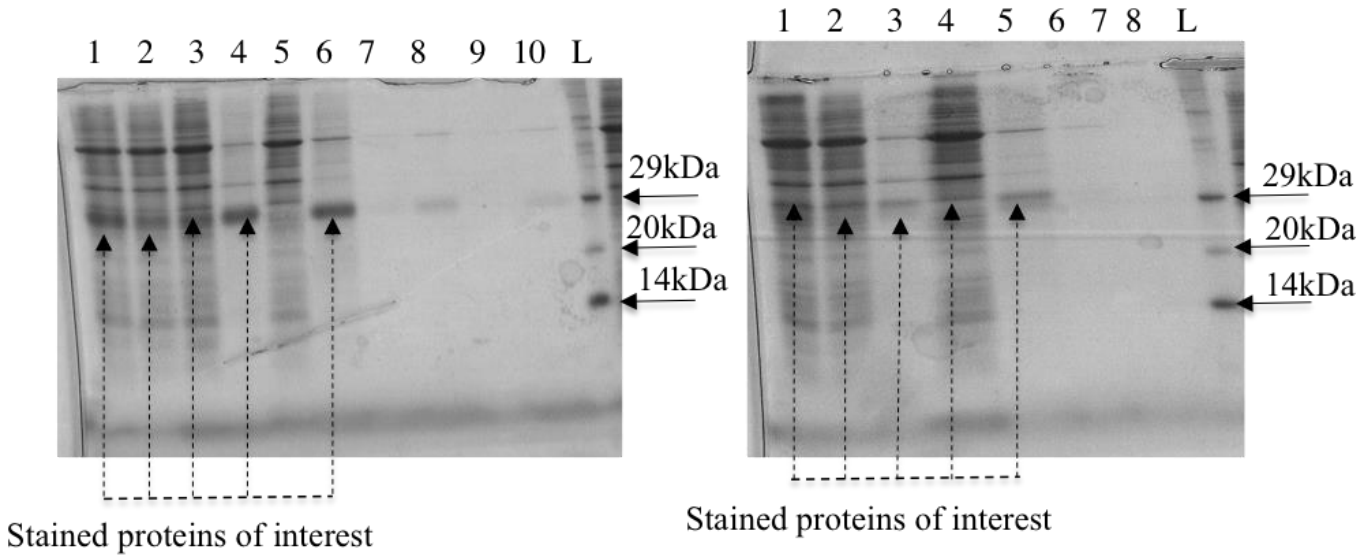


A) Construct RNase A (1-110)-A **B) Construct RNase A (1-110)-B**

Figure 12 –SDS-PAGE results from PD-10 desalting column flow through from the first purification attempt. Lane 1-8 contains the flow though of the PD-10 column. Lane L contains ladder. The mass of the protein is 29kDa.

Figure 12 shows the results after the PD-10 column. The diluted protein was concentrated and added to the PD-10 column which are resulted in strongly stained protein bands in lane 3-7 in Figure 12A and lane 3-5 in Figure 12B indicated by black arrows. Since a large amount of the protein was washed out during the washing step the purification needed to be done once more.

3.3.3 Second purification attempt with HiTrap™ TALON® crude and PD-10 desalting column



A) Construct RNase A (1-110)-A **B) Construct RNase A (1-110)-B**

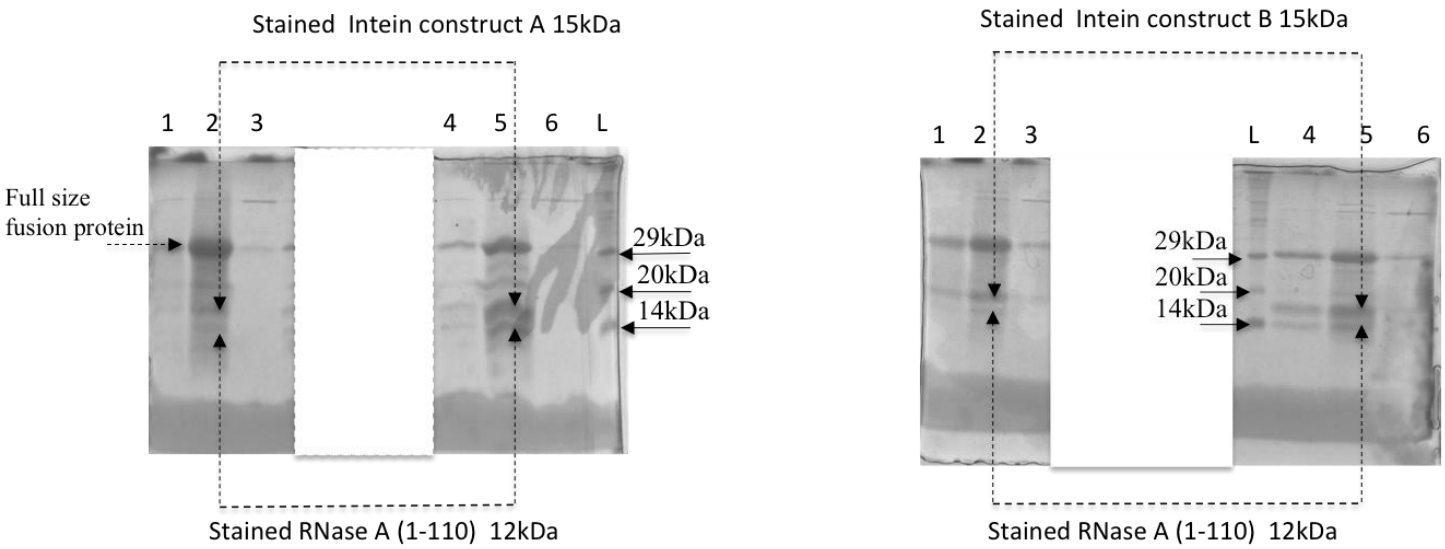
Figure 13 –SDS-PAGE results from the second purification attempt of RNase A (1-110)-A and RNase A (1-110)-B expressed in Arctic Express. The lanes in image A contain Total fraction (Lane 1), Soluble fraction 1 (Lane 2), Soluble fraction 2 (Lane 3), Insoluble fraction (Lane 4), Flow through (Lane 5), Washing step (Lane 6), Elution (Lane 7), PD-10 Flow through (Lane 8-10). The lanes in image B contain Total fraction (Lane 1), Soluble fraction (Lane 2), Insoluble fraction (Lane 3), Flow through (Lane 4), Washing step (Lane 5), Elution (Lane 6), PD-10 Flow through (Lane 7) Ladder (Lane 8). The masses of the fusion proteins are 29 kDa.

In the second purification, the washing solutions 1 and 2 had an Imidazole concentration of 20mM and 40mM, respectively. The stained protein bands from the different chromatography steps are indicated with black arrows showing the protein of interest. The lower concentration of Imidazole in washing buffers in the second purification attempt was more suitable for purification of the fusion protein, although still some protein was found in the eluate of the washing step. The fact that a large amount of protein was obtained by elution can not be seen from the gel, since for unknown reasons upon elution following the affinity step, the yielded protein aggregated, which was tentatively attributed to high protein yield following a high-level expression. Lane 7 (Figure 13A) and lane 6 (Figure 13B) correspond to the clear supernatant of the elution step without the precipitated protein and therefore these lanes show no protein band. The fact that there was protein remaining in the soluble state can be estimated from the protein bands in lanes 8-10 (Figure 13A) and lane 7 (Figure 13B). The remaining soluble protein loaded on a PD-10 column was not very

concentrated which explains the faint protein bands in lane 7-10 in image A and lane 7-8 in image B. For the usage of the protein in the coming thiolysis step, the purified protein was concentrated using a spin column.

3.3.4 Thiolysis of purified RNase A-Intein fusion protein

The thiolysis was done with purified fusion proteins RNase (1-110)-A and RNase (1-110)-B from the first and second purification attempt. The concentration of Mesna and TCEP was 100mM and 5mM. The final concentration of the purified protein was 10µM. Figure 14 illustrates the thiolysis results obtained on thiolysis reactions containing different concentrations of protein, peptide and Mesna in different lanes.



A) Construct RNase A (1-110)-A **B) Construct RNase A (1-110)-B**

Figure 14 – Separation of RNase A (1-110) and inteins (A and B) after thiolysis attempts.
Lane 1-3 in image A shows the first purification attempt of RNase A(1-110)-A fusion protein. Lane 1 contains only the protein and TCEP, lane 2 contains protein, TCEP and Mesna and lane 3 contain protein, TCEP, Mesna and the synthetic peptide (111-124). Lane 4-6 in image A shows the second purification attempt of RNase A (1-110)-A fusion protein. Lane 4 contains only the protein and TCEP, lane 5 contains protein, TCEP and Mesna and lane 6 contains protein, Mesna, TCEP and the synthetic peptide (111-124). Lane 1-3 in image B shows the first purification attempt of RNase A(1-110)-B fusion protein. Lane 1 contains only the protein, lane 2 contains protein, TCEP and Mesna and lane 3 contains protein, TCEP, Mesna and the synthetic peptide (111-124). Lane 4-6 in image B shows the second purification attempt of RNase A (1-110)-B fusion protein. Lane 4 contains only the protein and TCEP, lane 5 contains protein, TCEP and Mesna and lane 6 contains protein, TCEP Mesna and the synthetic peptide (111-124). The RNase (1-110) protein is 12kDa and intein A and B are 15kDa.

Although the different lanes in the images contain the same amounts of protein and TCEP but different amounts of Mesna and synthetic peptide, the staining is varying as seen by the bands at 29kDa corresponding to the full size of the fusion protein. This cannot be explained and the experiment needs to be repeated in order to be able to draw conclusions. In the presence of 100mM Mesna, the thiolysis was very limited, and it is difficult to assess whether the product bands at 14kDa and 20kDa are indeed stronger in the presence of Mesna compared to the sample without Mesna (no thiolysis). The arrows at 12kDa and 15kDa indicate weak bands corresponding to the mass of RNase A (1-110) and inteins A and B (lane 2 and 5, Figure 14A and 14B). As the efficiency of thiolysis was apparently very weak, the addition of synthetic peptide had no visible effect (lane 3 and 6, Figure 14A and 14B) and no signature of a full-length RNase A protein could be observed.

4 Discussion and Conclusions

In strategy 1, the construct Y^N was found to be obtained in a soluble form in all *E. coli* strains tested. However, in order to move forward and test the protein trans-splicing reaction for synthesis of full-length RNase A the corresponding construct Y^C is also needed in its soluble form. In this project, several methods were tested but in order to make insoluble RNase-intein fusion protein soluble, additional expression conditions and solubility buffers need to be tested.

When continuing the project, one possible way for overcoming the insolubility would be to fuse the insoluble protein to an elastin-like peptide, since this strategy has shown to increase the solubility of otherwise insoluble protein [2,16], but there is no guarantee that this method would work. The reason why constructs corresponding to X^N-X^C, Z^N-Z^C and Y^C could not be obtained in soluble form could be due to many reasons. Since it is known that the RNase A molecule has four interwoven disulfide bonds, there is a chance that improper bonding has occurred within the protein which created misfolding and insolubility.

If one complete split intein pair (X^N-X^C, Y^N-Y^C and Z^N-Z^C) fused with an RNase A variants is made soluble, the trans-splicing procedure might be achieved by mixing the proteins, and due to affinity of the split inteins, the protein will be ligated. Y^N was the only split intein fused with protein that was generated in its soluble form.

If fusion proteins, as described in strategy 1 and 2, are expressed in soluble form then these strategies could be used to express and splice other recombinant proteins that are challenging to express in soluble form in *E. coli*. When designing constructs for protein trans-splicing the splitting site should not be close to the active site and it should contain the recommended mutation AEY at its N-terminal extein and CFN at the C-terminal extein to start with. By adding these mutations in a splitting site away from the active site the activity of the protein could be preserved.

The *E. coli* expression strains that gave the best results were Arctic Express, Tuner and Shuffle. It was decided to continue the large-scale expression with Arctic Express. There might be several reasons for why Arctic Express gave the best expression of the fusion protein. The protein was expressed at 18°C and keeping a low temperature is recommended in order to express a soluble form of the protein. Arctic Express is well suitable for protein expression at a low temperature, which I believe is the main reason why this strain worked successfully when expressing the fusion proteins RNase A (1-110)-A and RNase A (1-110)-B required in EPL.

Cell lysis methods and conditions that were tested showed a wide variation of results. We chose to continue using Easylyse for small-scale expression and sonication for large-scale expression. Both Easylyse and sonication gave equally good results but were used for different scales of production for convenience.

The purification step for the fusion proteins RNase A (1-110)-A and RNase A (1-110)-B was repeated two times. In the first attempt, the highest concentration of imidazole in the wash buffers was 50mM that resulted in the loss of protein in the washing step during affinity purification. In the second attempt, the imidazole concentration was decreased. A large amount of protein was obtained upon elution, however, this was not shown in the gel image (Figure 13) since the yielded protein was aggregated tentatively due to high protein yield from the high-level expression.

In the thiolysis step some indications for a separation of the RNase A fragment with residues 1-110 from the fused intein A and B were seen. The thiolysis was done in order to see if the fusion protein yielded two separate bands. However, due to the differences in the amount of loaded protein, no firm conclusion can be drawn here, whether thiolysis occurred at all. In any case, the thiolysis reaction was far from complete and additional conditions need to be tried in order to improve the yield of the thiolysis step, and, eventually, ligation to the 111-124 peptide fragment to produce full-length RNase A.

5 References

1. Nambiar KP, Stackhouse J, Presnell SR, Benner SA (1987) Expression of bovine pancreatic ribonuclease A in *Escherichia coli*. European Journal of Biochemistry 163: 67-71.

2. Kopito RR (2000) Aggresomes, inclusion bodies and protein aggregation. Trends in Cell Biology 10: 524-530.

3. Raines RT (1998) Ribonuclease A. Chemical Review 98: 1045-1065.

4. Liu XQ (2000) Protein-Splicing Intein: Genetic Mobility, Origin and Evolution. Annual Review of Generics 34: 61-76.

5. Novikova O, Topilina N, Belfort M (2014): Enigmatic Distribution, Evolution and Function of Inteins. J Biol chem 289(21):14490-14497.

6. Dassa B, Pietrokoski S (2005): Origin and Evolution of Inteins and other Hint Domain. In Belfort M, Wood D.W, Stoddard B.L, Debyshire V:Homing Endonucleases and Inteins. New York: 211-231.

7. Shah NH, Muir TW (2014) Inteins: Nature's Gift to Protein Chemists. Chemical Science 5: 446-461.

8. Shah NH, Eryilmaz E, Cowburn D, Muir TW (2013) Extein residues Play an Intimate Role in the Rate-Limiting Step of Protein Trans-Splicing. Journal of the American Chemical Society 135: 5839-5847.

9. Schwarzer D, Ludwig C, Thiel IV, Mootz HD (2012) Probing Intein-Catalyzed Thioester Formation by Unnatural Amino Acid Substitution in the Active Site. Biochemistry Journal 51: 233-42.

10. Lockless SW, Muir TW (2009) Traceless protein splicing utilizing evolved split inteins. PNAS 106 (27).

11. Perler FB (2005) Protein Splicing Mechanisms and Application. IUBMB Life 57(7): 469-476.

12. Shah NH. Shah, Geoffrey DP, Vila-Perello M, Liu Z,Muir TW (2012) Ultrafast Protein Splicing is Common Among Cyanobacterial Split Inteins: Implications for Protein Engineering. Journal of the American Chemical society 134(28)**:** 11338-11341.

13. Severinov K, Muir TW (1998) Expressed Protein Ligation, a Novel Method for Studying Protein-Protein Interaction in Transcription. The Journal of Biological Chemistry 273(26): 16205-16209.

14. Vila-Perello M, Liu Z, Shah NH, Willis JA, Idoyaga J, Muir TW (2013) Streamlined expressed protein ligation using split inteins. Journal of Biological Chemistry society 135(1): 286-92.

15. Raines R.T (2004): Active site of Ribonuclease A. In Zenkova M: Artificial Nucleases. Madison: (19-32).

16. Hassouneh W, MacEwan SR, Chilkoti A (2012) Fusion of elastin-like polypeptides to pharmaceutical proteins. Methods in Enzymology 502: 215-237.

6 Acknowledgements

I would like to express my sincere gratitude to my supervisor Professor Thomas Friedrich for offering me the opportunity to do my master thesis at the Department of chemistry at the Technical University of Berlin. I am very grateful for being offered to participate in the project and for all the help and support.

I would also like to convey special thanks to Vijay Tejwani who has been the supervising Ph.D. student for this project and for going beyond his formal duties to support me and guide me through the project. Lastly, I would like to thank Lars-Göran Josefsson and Lena Henriksson for reviewing my work.

7 Appendix

Mass of the proteins in the ladder used for SDS-PAGE, Roti® mark standard, 14-212kDa from Carl Roth.