UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Technology

Sofija Frolova

Effect of protein tag’s codon usage on protein expression in Saccharomyces cerevisiae

Bachelor's Thesis (12 ECTS)

Curriculum Science and Technology

Supervisor(s):
PhD Rait Kivi

Tartu 2021

Effect of protein tag's codon usage on protein expression in Saccharomyces cerevisiae

Abstract:

Recent studies have shown that some codons are preferentially used over others for coding the same sequence polypeptide chains. Synonymous codons usage variety appears among different genomes and even among the same genome. The reasons for codon usage are not completely understood yet. In this study based on Saccharomyces cerevisiae we examined expression of optimized and non-optimized sequences that encode the same protein tags, using western blotting and fluorescence microscopy. Both western blotting and microscopy results showed that codon optimization of the tags didn't have considerable effect on protein expression. On the other hand, small changes in tag amino acid composition had much larger effect on fluorescence signal strength.

Keywords:

Codon usage, codon, synonymous codon, protein tag, western blot, fluorescence microscopy, Saccharomyces cerevisiae

CERCS: P310 Proteins, Enzymology

Valgumärgise koodonkasutuse mõju valguekspressioonile pagaripärmis

Lühikokkuvõte:

Genoomsete järjestuste uurimisel on avastatud, et valgu geenijärjestuses sünonüümsete koodonite kasutamisel eelistatakse mõnda koodonivarianti teisele. Selline nähtus on looduses laialt levinud. Koodonkasutuse erinevuste põhjused ei ole veel täielikult välja selgitatud. Antud töös uuritakse pagaripärmi näitel, kuidas mõjutavad erineva koodonkasutusega, ent identseid valgumärgiseid kodeerivad järjestused, reportervalgu taset. Valgutasemeid mõõdeti wester-blot meetodil ning fluorestsentsmikroskoopiaga. Western-blot näitas, et märgise koodonoptimeermine ei mõjuta oluliselt valguekspressiooni, samas märgise lisamisel oli oluline reportervalgu taset suurendav mõju. Mikroskoopia näitas, et väikesed muutused valgumärgise aminohappejärjestuses võivad oluliselt mõjutada reportervalgu fluorestsents-signaali.

Võtmesõnad:

Koodonkasutus, koodon, sünonüümsed koodonid, valgumärgis, western-bloti meetod, fluorestsentsmikroskoopia, Saccharomyces cerevisiae

CERCS: P310 Proteiinid, ensümoloogia

TABLE OF CONTENTS

TERMS, ABBREVIATIONS AND NOTATIONS

CSM – Complete supplement mixture

CUB – Codon usage bias

EDTA - Ethylenediaminetetraacetic acid

NES – Nuclear export signal

OD – Optical density

RSCU - Relative synonymous codon usage

SAM – S-Adenosyl methionine

TAE buffer - Tris-acetate-EDTA buffer

TE buffer - Tris-EDTA buffer

YPD – Yeast extract peptone dextrose

INTRODUCTION

Ongoing advances in DNA synthesis and the reduction of synthesis cost has created a new branch in biology – synthetic biology. It allows modification of organisms at scale to give them new abilities through tinkering with their genomes, for example adding new genes or regulating their own genome programs through transcription or post-translational modifications. Our lab is working on a protein tag that can be used to control expression of a protein that the tag is attached to through phosphorylation. phosphorylation. In order to make it usable in synthetic biology systems, it is important to understand, how adding this tag to proteins might change the expression of the target protein.

Codon usage means preferences between synonymous codons. Synonymous codons are codons that encode the same amino acid. The codon usage varies between different genomes and even within the same genome. The reasons of these variations are not competely investigated and understood. Several studies made an assumption of some possible factors that affect codon usage, such as metabolic pressures, gene expression controlling through gene sequence, protein folding, adaptation to changing conditions.

The question about the effect that codon usage gives to protein expression is asked and not completely understood too. In this work we would like to study this question. We are planning to construct strains with optimized and not optimized sequences that code for the same amino acid chain. We would like to understand how adding different degron modules to mCherry module affect protein expression. Also, we are planning to compare optimized and not optimized sequences, their composition and variation in protein synthesis.

To answer these questions, we are planning to do nucleotide composition analysis, RSCU analysis, Western Blot experiment and microscopy experiment.

1 LITERATURE REVIEW

1.1 Protein synthesis

Translation is a process through which mRNA that contains genetic information is used for generation of amino acids's sequence in proteins. In bacterial cells 80% of cell energy goes to the protein synthesis. For one protein synthesis over 100 proteins and RNAs coordinated action is required (Alberts et al., 2008).

Ribosomes have two subunits: small and large particles. Prokaryotes have 50S and 30S particles, which form a 70S particle. Eukaryotes have 60S and 40S particles, the complete ribosome is 80S. Also ribosomes have two sites: A site and P site. There are specific mRNA, tRNAs and protein factors binding sites in ribosomes, these bindings are required for synthesis of proteins (Griffiths et al. 2000).

In bacterial ribosomes, the start sites of translation are located between the rRNA in through base-pairing interactions part in the ribosome, the 5' of initiation codon is sequenced immediately (Kozak, 2005; Laursen et al. 2005). Whereas the eukaryotic ribosomes bind close to the 5' cap of the mRNA and scanning happens in a 3' direction, while checking over start codons (Hinnebusch 2011). IF1, IF2, IF3 bacterial translation factors (Laursen *et al.* 2005, Schmeing and Ramakrishnan 2009) are substituted by 11 factors in yeast cells. These 11 factors are: eIF1, eIF1A, eIF2, eIF2B, eIF3, eIF4A, eIF4B, eIF4E, eIF4G, eIF5, eIF5B (Dever et al. 2016). The elongation and termination factors are the same between eukaryotes and prokaryotes, the only difference is in yeast cells. Yeast cells need eEF3 elongation factor, this factor is only required in yeast cells and is not found in bacteria or other eukaryotes (Belfield and Tuite 1993).

Prokaryotic cells will be used as an example for further explanation of protein synthesis.

Figure 1 illustrates a process of amino acid binding to a polypeptide chain. First of all mRNA binds to the 30S ribosome subunit. Then two ribosome sites are overlapped by binding of tRNAs to them. A site accepts the entry of Aminoacyl-tRNA that carries a single amino acid. While peptidil-tRNA is tRNA that carries a growing chain of polypeptide, and its binding site is site P. A new peptide bond is formed by adding a new amino acid to the growing chain. After that tRNA becomes deacylated and is released from site P. Then new peptidyl-tRNA is transferred to the P site and one codon is moved farther by the ribosome. After that, the A site becomes free for receiving incoming aminoacyl-tRNA (Griffiths et al. 2000).

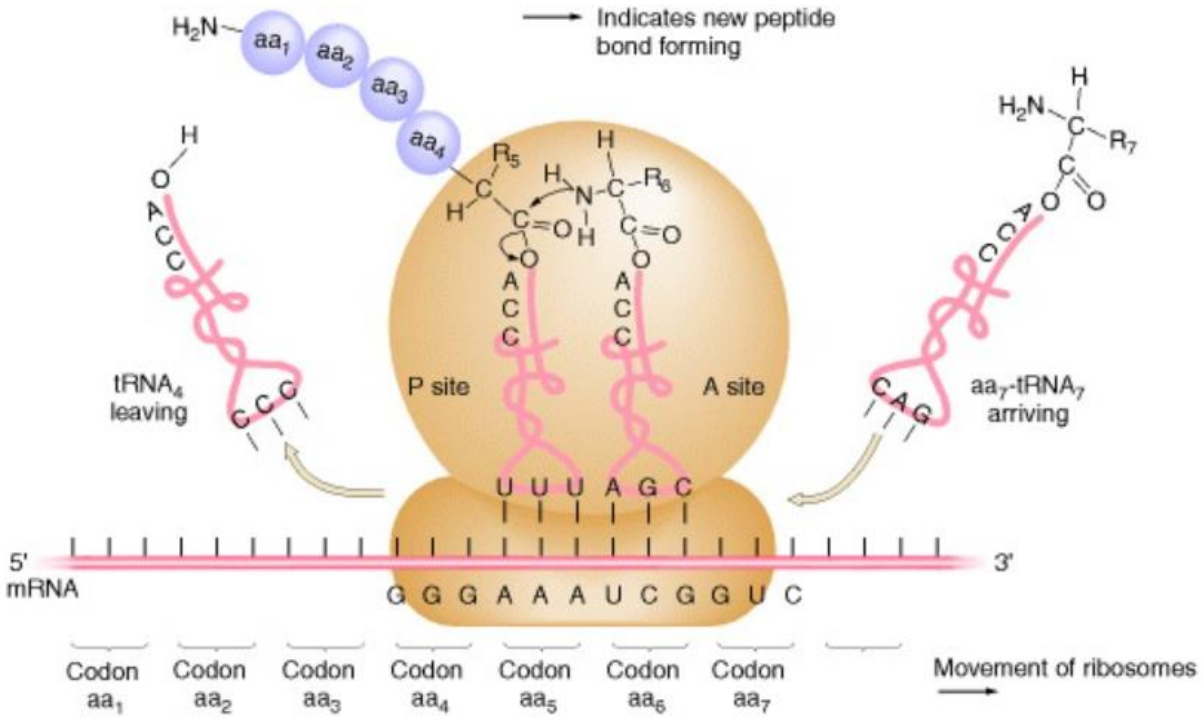

Figure 1. A single amino acid addition to the polypeptide chain during the translation process (Griffiths et al. 2000).

Initiation part in prokaryotes requires mRNA, ribosomes, specific tRNA molecules and initiation factors IF1, IF2, IF3. In E. coli and also mostly in all over prokaryotic cells N-formylmethionine is the first amino acid that is synthesized in polypeptide, $tRNA^{fMET}$ is an initiator tRNA that inserts this first amino acid. AUG, GUG and UUG are initiation codons in E. coli. N-formylMet-tRNA recognizes initiation codons, after that methionine shows up as the chain's first amino acid (Griffiths et al. 2000).

Elongation factors are EF-Tu, EF-Ts, and EF-G. These factors participate in receiving amino acid-tRNA complex into the A site, moving codon along the mRNA in translation 5′ → 3′ direction and releasing uncharged tRNA from the P site (Griffiths et al. 2000).

Three chain-termination codons are UAG, UAA, UGA. These codons are recognized by release factors RF1 and RF2. Release factor RF3 also helps with chain termination catalyzation. The release factors bind to the A site, when peptidyl-tRNA is located in the site P and the chain terminating codons are recognized. Then the polypeptide is released, the ribosome is dissociated into two subunits (Griffiths et al. 2000).

1.2 Codon. Codon table

Codon is a triple-nucleotide code that provides all information to translate the mRNA that was encoded by DNA into amino acid sequence using tRNAs and modification factors (Crick et al. 1961; Kubyshkin et al. 2018; Tamura, 2015).

mRNA is translated into a protein in the same way in all cells, and translation that happens from 4 different polynucleotides (A, U, C, G) into 20 different amino acids is a complex process.

The information is read from mRNA out in the form of a codon. Every codon codes for a particular amino acid. As there are 64 (4x4x4) possible combinations of nucleotides and only 20 amino acids, in some cases there is a necessity of several codons to correspond to one amino acid.

Transfer RNAs (tRNA) are small RNA molecules that read the code. Amino acids are attached to their specific tRNA molecules to one side, the other side of tRNA transfers anticodons. Anticodons consist of three nucleotides that recognise the mRNA codons via base-pairing (Alberts et al., 2008).

Multiple synonymous codons encode the amino acids. Cells and organelles have a huge variation in the tRNAs relative expression (Dittmar et al. 2006; Goldman, 2011). In humans 13 codons remain without corresponding tRNA genes, and other 48 codons have approximately 500 corresponding tRNAs genes (Goldman, 2011). Interesting is that Chinese hamster ovary cells that are used in therapeutics for protein production are missing the overlapping with human but different tRNA genes set. The absence of these tRNA does not affect synonymous codon usage, because the full complement of codons is used by mRNAs. For instance, aspartic acid is encoded by two codons; despite the absence of GAU codon's corresponding tRNA genes, there is similar codon usage. The 'wobble' enables decoding of both codons by the same tRNA (Mauro and Chappell, 2014).

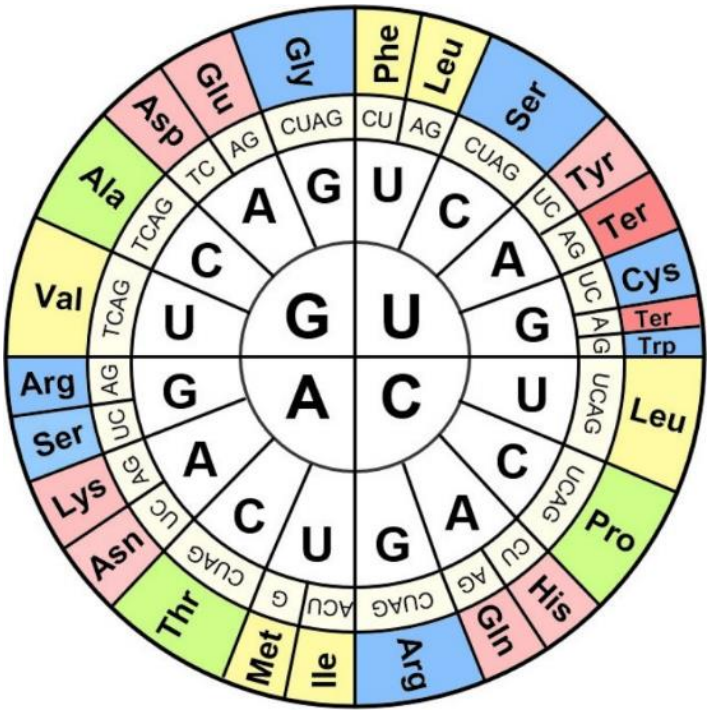

Figure 2. Codon wheel (Saier, 2019).

The codon wheel shown above (Figure 2) is used to determine the amino acid that is encoded by a particular DNA/RNA codon, which consists of small bases (U and C) - pyrimidines, and large bases (G and A) - purines. In the center of the wheel is the first codon position, in the middle of the wheel there is a second codon position and near the periphery of the wheel there is a third codon position (Saier, 2019).

1.3 Codon usage. Why do organisms have different codon usage biases?

The meaning of codon usage bias (CUB) comes from codon preferences over other codons that encode the same amino acid. The CUB is specific for each genome (Bennetzen and Hall, 1982; Ikemura, 1981; Ikemura, 1981; Plotkin et al., 2004). The reasons for CUB are not completely understood, but there can be some major factors that Tylor Ford listed in his article, such as metabolic pressures, gene expression controlling through gene sequence, protein folding, adaptation to changing conditions (Ford, 2018).

The production of recognizing different codons tRNAs, the correct modification of tRNAs and charging the tRNAs with amino acids - all these processes are metabolic pressures that take cellular resources for those actions (Ford, 2018). There was a study conducted by Valur Emilsson and Charles G. Kurland about correlation of tRNA species and their cognation to

major or minor codons. That study showed that there is a preferential upregulation in production of tRNAs that recognize highly expressed genes' codons during high growth rate conditions in *E.coli* (Emilsson and Kurland, 1990).

The rate of protein production depends on abundance and charge of tRNAs, this explains the controlling of gene expression through gene sequence CUB factor. For instance, tRNAs with low abundance encode proteins at lower rates, than those with high abundance (Ford, 2018). *E.coli* and *S.cerevisiae* translation efficiency has good correlation with codon bias (Tuller et al., 2010).

The rate of translation can vary on the different regions of the protein if a mixture of codons with tRNAs that are highly and poorly charged encode a protein. In that case the ribosome will move fast along the abundant, charged tRNAs calling regions and slow along the regions that call for tRNAs with low abundance and poor charge. This stands for protein folding CUB factor explanation, since during slowing a ribosome down period the fast translated regions are able to fold properly (Ford, 2018). This argument was also tested by Sebastian Pechmann and Judith Frydman, they found out that in 10 related yeast strains non-optimal codon tracts are associated with specific secondary structures (Figure 3).

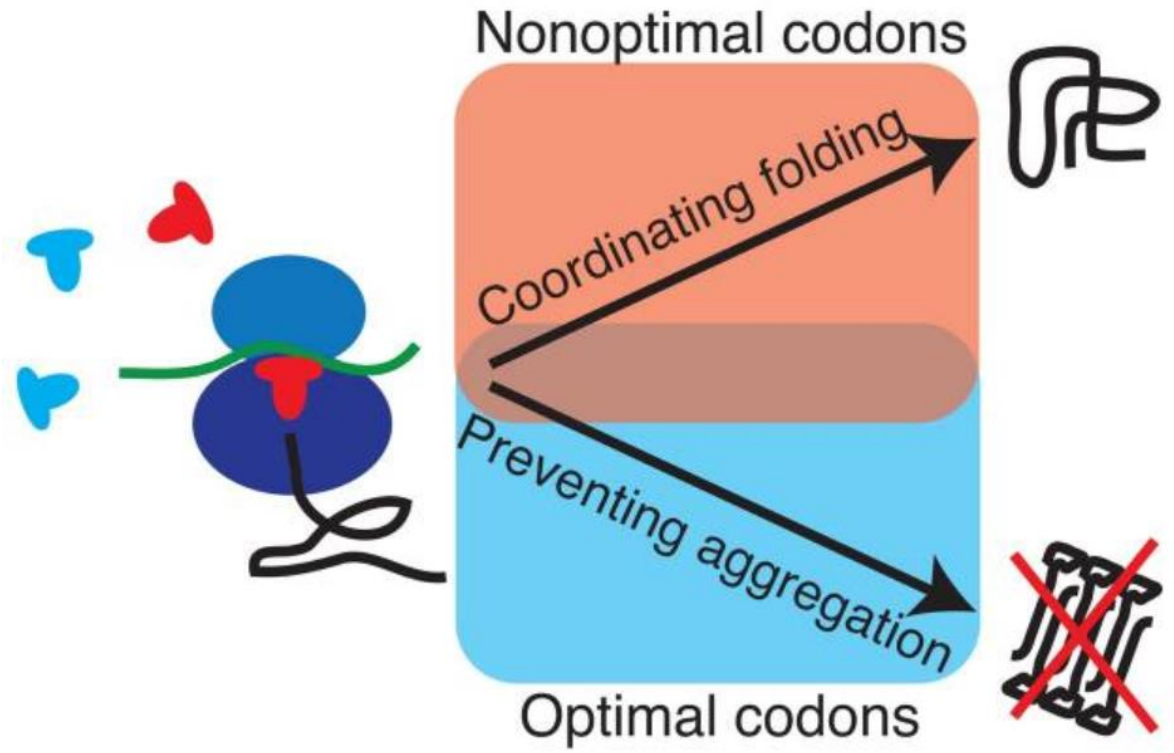

Figure 3. Co-translational folding (Pechmann and Frydman, 2013).

Figure 3 shows that there is more accurate and faster translation of optimal codons, and non-optimal codons cause pausing in translation. Optimal codons that are evolutionary conserved are mostly found at sites where aggregation prevention requires accurate translation, whereas evolutionary conserved non-optimal codons are mostly located in co-translationally folding secondary structure elements, these structures even can fold inside the exit tunnel of ribosome (Pechmann and Frydman, 2013).

Usually, organisms have a necessity of different level gene expression or expression of genes under different conditions, this is possible because of varied codon usage. The level of protein expression, whether it is high or low, can be changed by production and charge of specific tRNA pools by an organism itself. These level changes happen due to the adaptation to changing conditions, that is the factor of CUB (Ford, 2018). As an example, the study of Kimberly A. Dittmar et al. showed that during amino acid starvation tRNAs that are used in amino acid biosynthetic enzymes encoding genes are favourably charged, this leads to higher production of these enzymes (Dittmar et al., 2005).

The Deb et al. article shows such factors, as evolutionary forces, compositional properties, gene expression, protein properties, that contribute to CUB. Their research topic was focused on compositional properties and the extent of CUB investigation of Hepadnaviridae family genomes using bioinformatic tools. They used relative synonymous codon usage analysis (RSCU) for identification of underrepresented and overrepresented codons in each amino acid. The results showed almost equal usage of T, A, C nucleotides, the usage of G nucleotides was different; the percentage of AT was 53.22% and GC percentage was 46.78%, respectively (Figure 4). There were identified overrepresented codons: TCT, AGA, GGA; and underrepresented codons: TCG, AGC, AGT, CCG, CGA, ACG, GCG (Figure 5). The results showed that the codon usage pattern in genes of hepadnavirus genomes appear to be affected by mutational pressure and natural selection (Deb et al., 2020).

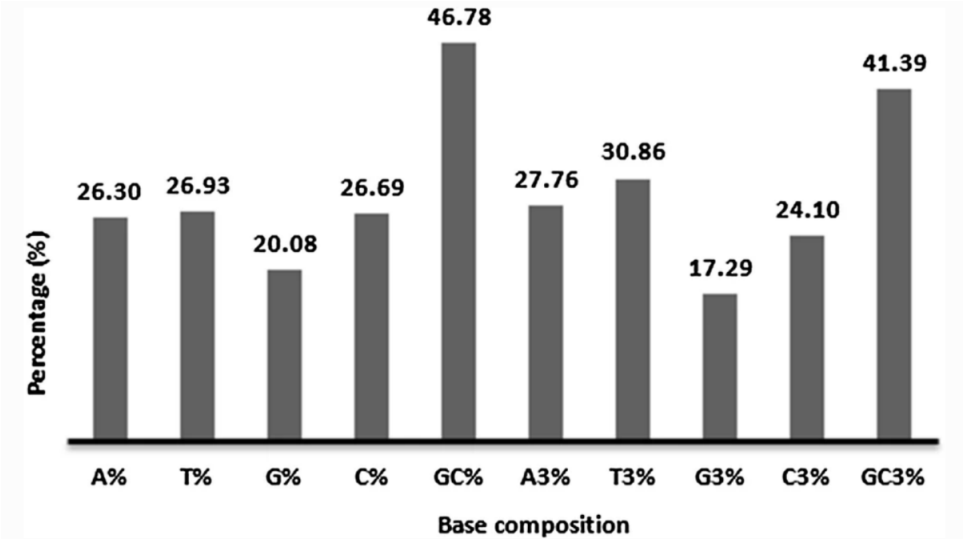

Figure 4. Composition of bases in percentage and composition of bases in percentage at the third codon position of hepadnavirus genomes' genes (Deb et al., 2020).

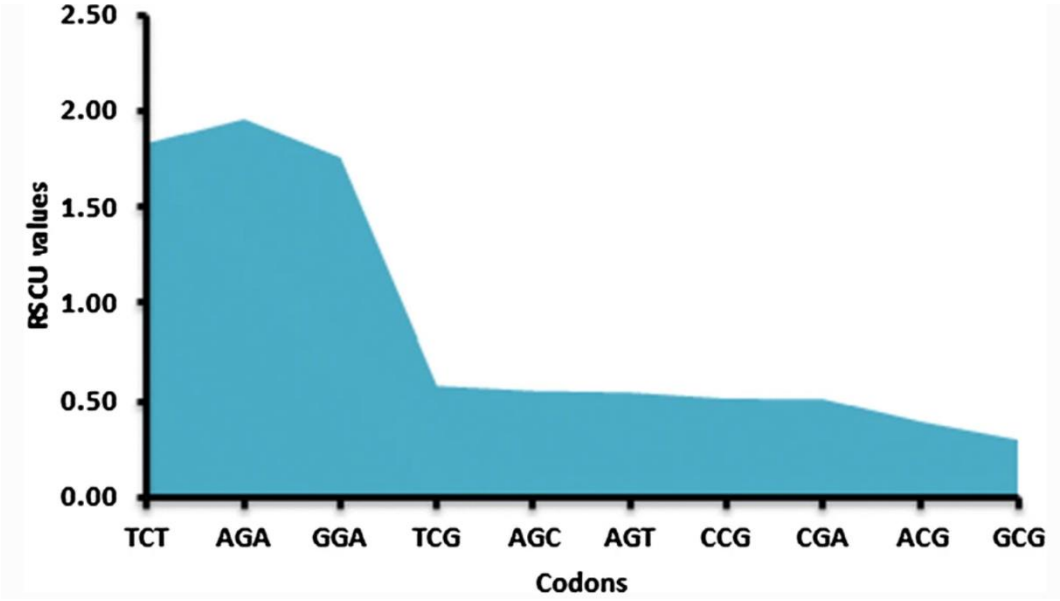

Figure 5. Overrepresented and underrepresented codons in the genomes of hepadnavirus (Deb et al., 2020).

Figure 5 represents overrepresented and underrepresented codons in the genomes of hepadnavirus. Overrepresented codons are codons with RSCU value bigger than 1.6 and underrepresented codons with RSCU value smaller than 0.6 (Deb et al., 2020).

1.4 Effect of codon usage on protein expression

Strategies of codon-optimization for increasing expression of proteins are based on several assumptions (Mauro and Chappel, 2014).

There is an assumption that rare codons slow down translation of RNA at specific sites and that facilitates folding of protein. Germán Rosano and Eduardo Ceccarelli showed in their studies that translation rate is increased by increasing tRNAs expression that correspond to rare codons, also this led to misfolding and aggregation of a protein (Rosano and Ceccarelli, 2009).

In humans, there is a difference up to 10-fold in encoding various amino acids (Figure 6A). However, there are fewer synonymous codons for less frequently encoded amino acids than for more frequently encoded amino acids. When normalization of frequency of amino acid to the synonymous codons per each amino acid is obtained, the differences become equal to 3-fold (Figure 6B). There is a variety up to 10-fold in individual codon usage frequencies (Figure 6C). Rate-limitations of a codon depend on different variables, such as tRNA levels. A various distribution is shown by the normalized codon frequencies, when there is consideration of tRNA genes number (from 0 to 33 genes, based on tRNA isodecoders) and consideration of wobble (Figure 6D) (Mauro and Chappell, 2014).

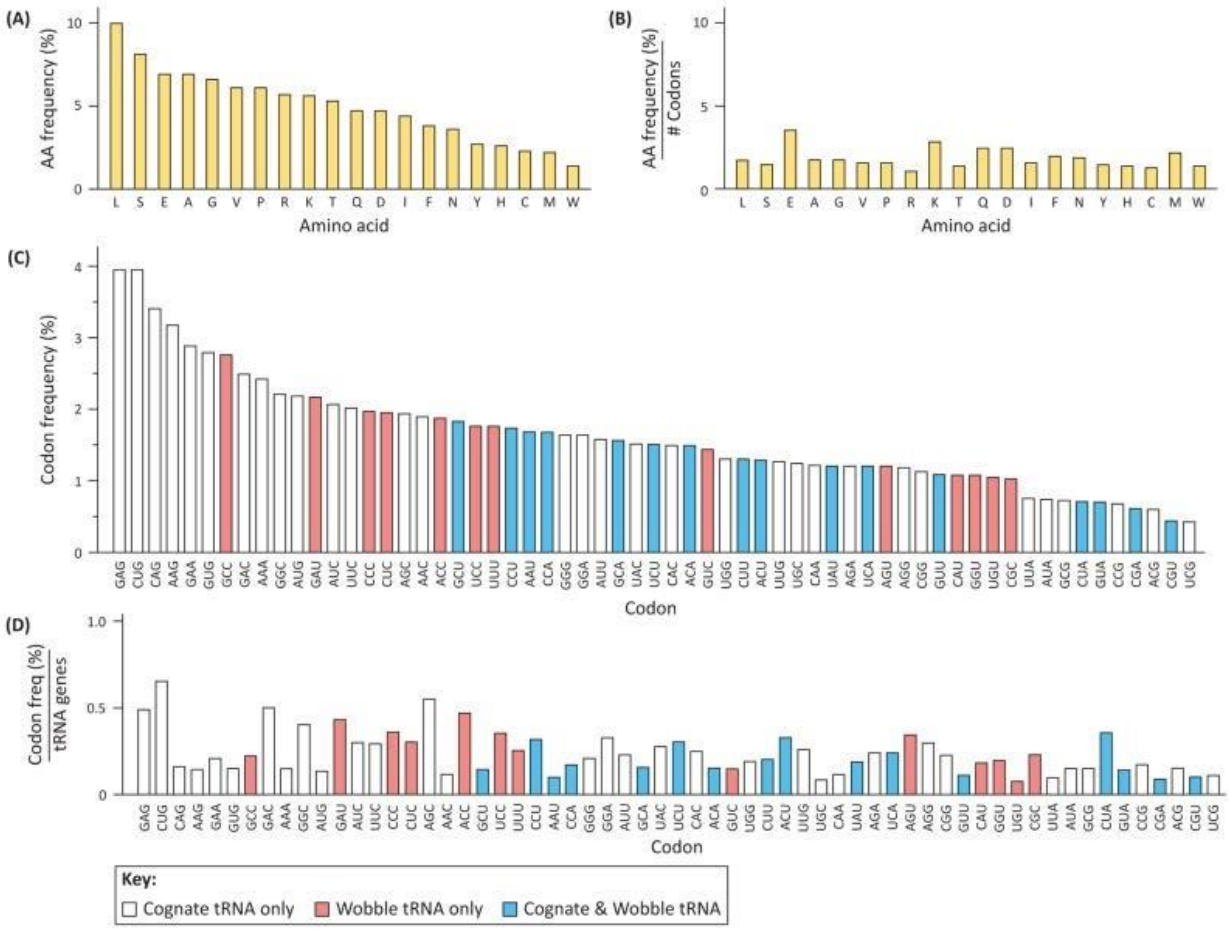

Figure 6. (A) Amino acids frequency. (B) Normalized frequency of amino acids. (C) Codon usage frequency (D) Normalized codon frequences (Mauro and Chappell, 2014).

Another assumption is that synonymous codons can be interchangeable, and this does not affect structure and function of protein (Mauro and Chappell, 2014). The study of Sander et al showed the effects of synonymous codon changes, where they used a fluorescent protein gene in their experiments. This gene was engineered to detect different folding structures by different properties of fluorescence (Sander et al., 2014).They made substitution within 18 synonymous codons. These changes had a small effect on the rate of translation. However, these minor changes were sufficient to make competition between certain structure formations. Thus, the result of this study showed that protein structure can be dramatically altered by subtle changes of synonymous codon (Sander et al., 2014).

2 THE AIMS OF THE THESIS

Our lab is working on protein tags that can be used to degrade protein in S. cereviseae. Main aims of my thesis are to test if:

1) codon usage of the tag have any effect on overall tag-mCherry expression

2) adding this tag to a reporter protein has any effect on overall tag-mCherry expression

To answer these questions, I set out to:

- To prepare 5 yeast containing following reporter sequences:
  1. NES-STST-mCherry-3HA optimized sequence
  2. NES-STST-mCherry-3HA
  3. NES-AAAA-mCherry-3HA optimized sequence
  4. NES-AAAA-mCherry-3HA
  5. NES-mCherry-3HA (control)
- To compare effect of codon usage between optimized, not optimized and control sequences, using western blotting and flurorescence microscopy.

3 EXPERIMENTAL PART

3.1 MATERIALS AND METHODS

3.1.1 Materials

3.1.1.1 A table of used plasmids

Table 1. A table of used plasmids.

Name	Description	Source
pRKI0121	NES-STST-mCherry-3HA	Rait Kivi
pRKI0141	NES-AAAA-mCherry-3HA	Rait Kivi
pSF1	NES-STST-mCherry-3HA optimized sequence	This study
pSF2	NES-AAAA-mCherry-3HA optimized sequence	This study
pSF3	NES-mCherry-3HA	This study

3.1.1.2 A table of used strains

Table 2. A table of used strains.

Strain	Background	Genotype	Source
RKI514	W303	bar1Δ::HISG his3::PACT1(-1-520)-LexA-ER-haB112-TCYC1::HIS3 ura3::insul-(lexA-box)4-NES-Δ130N-Clb3-eGFP::URA3 Clb3::Trp1 pGal1 3HA-Grr1:NatNT	Rait Kivi
SF1	W303	bar1Δ::HISG his3::PACT1(-1-520)-LexA-ER-haB112-TCYC1::HIS3 ura3::insul-(lexA-box)4-NES-Δ130N-Clb3-eGFP::URA3 **leu2::PADH1 optNES-TTT-STST-23PNF-mCherry-3HA::LEU2** Clb3::Trp1 pGal1 3HA-Grr1:NatNT	This study
SF2	W303	bar1Δ::HISG his3::PACT1(-1-520)-LexA-ER-haB112-TCYC1::HIS3 ura3::insul-(lexA-box)4-NES-Δ130N-Clb3-eGFP::URA3 **leu2::PADH1 optNES-TTT-AAAA-23PNF-mCherry-3HA::LEU2** Clb3::Trp1 pGal1 3HA-Grr1:NatNT	This study

SF3	W303	bar1Δ::HISG his3::PACT1(-1-520)-LexA-ER-haB112-TCYC1::HIS3 ura3::insul-(lexA-box)4-NES-Δ130N-Clb3-eGFP::URA3 **leu2::PADH1 NES-mCherry-3HA::LEU2** Clb3::Trp1 pGal1 3HA-Grr1:NatNT	This study
RKI0521	W303	bar1Δ::HISG his3::PACT1(-1-520)-LexA-ER-haB112-TCYC1::HIS3 ura3::insul-(lexA-box)4-NES-Δ130N-Clb3-eGFP::URA3 **leu2::PADH1 NES-TTT-AAAA-23PNF-mCherry-3HA::LEU2** Clb3::Trp1 pGal1 3HA-Grr1:NatNT	Rait Kivi
RKI0519	W303	bar1Δ::HISG his3::PACT1(-1-520)-LexA-ER-haB112-TCYC1::HIS3 ura3::insul-(lexA-box)4-NES-Δ130N-Clb3-eGFP::URA3 **leu2::PADH1 NES-TTT-STST-23PNF-mCherry-3HA::LEU2** Clb3::Trp1 pGal1 3HA-Grr1:NatNT	Rait Kivi

3.1.1.3 A table of used primers

Table 3. A table of used primers.

Primer's number from Loog Lab data-base	Sequence
4788	GCA TGG ATC CAT GAA TGA ATT G
3649	GTT ATC CTC CTC GCC CTT GCT CAC CAT TCC AAC CCT AAC AGG GTT A
4523	CAT GCG GCC GCT TAA GCG
4804	TAG CAT AGG ATC CAT GAA TGA ATT GGC
2066	GTA AAA CGA CGG CCA GT

3.1.1.4 Synthetic DNA

Synthetic DNA for both optimized and unoptimized tag sequences were ordered from GeneArt (Thermo Fischer) over 2 year window. We noticed that the optimization algorithm was changed during that time, which promted this investigation on codon usage and its effect on protein expression.

3.1.2 Strains preparation

To analyze the codon usage effect on protein expression we needed to prepare strains with optimized sequence and control strains that do not contain degron parts. Below there is description of NES-AAAA-mCherry-3HA (optimized sequence), NES-STST-mCherry-3HA (optimized sequence), NES-mCherry-3HA colony 1 and 2 strains preparation.

3.1.2.1 Degron amplification PCR

For amplification of optimized degron fragments AAAA/STST we used synthetic DNA as a template for PCR. The volume of the reaction mixture was 50 µl. Reaction mixture consisted of milli-Q H2O, 1x Phusion HF Buffer (Thermo Fisher Scientific), forward (4788) and reverse (3649) primers with final concentration 0,3 µM, 10 ng synthetic DNA template (either AAAA or STST), 250 µM dNTPs and 1 unit of Phusion DNA Polymerase (Thermo Fisher Scientific).

The reaction mixture was placed into the thermocycler with a program shown in Table 4.

Table 4. PCR program.

Step	Temperature	Time	Cycle
Initial Denaturation	98°C	1 min	1
Denaturation	98°C	10 s	35
Annealing	58°C	10 s	
Extension	72°C	21 s	
Final Extension	72°C	5 min	1
Hold	15°C	∞	

The oligonucleotide annealing temperature was calculated using Benchling.com platform; it takes in account the length and deoxynucleotide content of primers to calculate the annealing temperature. The extension time was calculated by formula ((length of the fragment)/2000)*60.

3.1.2.2 Agarose gel electrophoresis for degron fragments

DNA fragments from PCR were checked using agarose gel electrophoresis. Electrophoresis bath was filled with a 1xTAE buffer (40 mM Tris-acetate pH 8.3, 1 mM EDTA). We added 4 µl of 6x DNA Loading Dye (Thermo Fisher Scientific) to stain the reaction samples. After that samples (AAAA/STST amplified fragments) and DNA Ladder 1 (Thermo Fisher Scientific) were loaded to the 1 % Agarose gel (40 mM Tris-acetate with pH 8.3, 1 mM EDTA, 1% agarose, 0.05 µl Atlas ClearSight DNA Stain (BioAtlas)). The settings of electrophoresis were set to 170 V and 30 minutes. After the program had finished, we located our gel under UV light (280 nm) and took a picture. DNA Ladder 1 (Thermo Fisher Scientific) was used to assess the length of PCR products. The correctly sized bands with the length of 700 bp were cut out from the gel and placed into new Eppendorf tubes.

DNA fragments were purified from the gel pieces using a protocol provided by FavorPrepTM GEL/PCR Purification Kit (Favorgen).After purification of DNA from gel we measured DNA concentration with NanoDrop 1000 3.8.1 software in NanoDrop 1000 Spectrophotometer using NanoDrop 1000 3.8.1 software (Thermo Fisher).

3.1.2.3 Overlap extension PCR for sewing degron and mCherry fragments and their amplification

For sewing two fragments together, a degron fragment (AAAA/STST) that we amplified during our first PCR and mCherry fragment that Rait Kivi had provided, the overlap extension PCR was done. The reaction mixture volume was 50 µl. First, we added milli-Q H2O, 1x concentration of 5x Phusion HF Buffer (Thermo Fisher Scientific). Then fragments were added: 10 ng of mCherry and 10 ng amplified degron fragments (AAA for one mixture, STST for another mixture). Before starting the program we added 1 unit of Phusion DNA Polymerase (Thermo Fisher Scientific).

After the end of the 1 cycle and before the start of the Denaturation phase of 35 cycles, when two fragments annealed in the overlap area during the first annealing phase and extended to make sewed two fragments together during the extension phase, the program was paused and 4804 (forward) and 4523 (reverse) primers were added. The final concentration of each primer in the reaction mixture was 0,3 µM. After that, the program proceeded till the end.

Table 5. Overlap extension PCR program.

Step	Temperature	Time	Cycle

Initial Denaturation	98°C	3 min	1
Annealing	65°C	1 min	
Extension	72°C	3 min	
Denaturation	98°C	10 s	35
Annealing	63°C	15 s	
Extension	72°C	45 s	
Final Extension	72°C	5 min	1
Hold	15°C	∞	

The annealing temperature was also calculated using the Benchling platform that takes in account the length and content of primers for temperature calculation. The extension time was calculated by the same formula that we used for the first PCR ((length of the fragment)/2000))*60.

After the overlap extension PCR program had been finished we did agarose gel electrophoresis, the technique is described in chapter 3.1.2.2.

3.1.2.4 Restriction of the backbone plasmid and amplified inserts

The enzyme restriction was necessary for further ligation of backbone plasmid and insert. First of all we needed to cut out STST-mCherry fragment from plasmid pRKI0121 (Table 1) that was given by Rait Kivi, for that we made a reaction mixture. The final volume of this mixture was 30 µl. First Milli-Q H2O and 10x FastDigest™ Buffer were added, the final concentration of the buffer was 1x. Then we added 1 µg of pRKI0121 plasmid, 0,5 units of FastAP (Thermo Fisher Scientific) and 5 units of each BamHI and NotI (Thermo Fisher FastDigest™) restriction enzymes.

The second enzyme restriction was done to cut the ends of PCR products that were made in chapter 3.1.2.3 for further ligation, exactly AAAA-mCherry and STST-mCherry. The total volume of each mixture was 20 µl. We added milli-Q H2O, 1 µg of PCR product, 10x

FastDigest™ Buffer with final concentration 1x and 5 units of each restriction enzymes BamHI/NotI (Thermo Fisher FastDigest™).

All mixtures were incubated at 37°C for 40 minutes. The restriction reaction for plasmid linearization was controlled by gel electrophoresis and purified, the exact description is in chapter 3.1.2.2. Restriction enzymes of PCR products restriction were inactivated at 80°C in 5 minutes.

3.1.2.5 Ligation of backbone plasmid and insert

During this step new plasmids were constructed: pSF1, pSF2, pSF3 (Table 1). When restriction was completed we proceeded with a ligation reaction. We ligate together the backbone (pRKI0121 (Table 1) without STST-mCherry fragment) and the insert (restricted AAAA-mCherry/STST-mCherry fragments). The final volume of reaction was 10 µl, and it contained milli-Q H_2O, 10x T4 DNA Ligase Buffer with 1x final concentration, 2:1 ratio of volume of insert and backbone (100 and 50 ng, respectively) and 0.5 units of T4 DNA Ligase (Thermo Fisher Scientific). The reaction mixture was incubated at 16°C during the night.

3.1.2.6 Bacterial transformation

We transformed ligated plasmids into E. coli Turbo competent cells that were taken from the -80°C freezer. 50 µl of cells were mixed with 5 µl of ligation product and incubated on ice for 30 minutes. After that the 40 seconds heat shock at 42°C was done and mixtures were put on ice for 5 minutes. Then 500 µl of LB media (5 g/L yeast extract (Formedium), 10 g/L tryptone (BD Biosciences), 10 g/L NaCl) was added to each mixture and they were incubated in 220 rpm shaker at 37°C for 60 minutes. After incubation we centrifuged the cells at 6000g for 1 minute. Then we removed 350 µl of supernatant and resuspended the pellet in the media that was left. Then we plated cells on LB plates (with 100 µg/mL of ampicillin) and incubated for 15 hours at 37°C.

3.1.2.7 Plasmid control

To control colonies that had been grown on the plates whether they have the correct insert we needed to make MiniPreps and control them by restriction. First of all we chose some colonies on the plates and put them to grow into 5 ml LB media with 100 µg/mL ampicillin overnight in 220 rpm shaker at 37°C. Then the DNA was extracted using FavorPrepTM Plasmid DNA Extraction Mini Kit (Favorgen) protocol and DNA concentration was measured as it is described in part 3.1.2.2.

To control the insert DNA was restricted, 20 µl reaction mixture was made. The mixture was made of milli-Q H2O, 10 ng of extracted DNA, 10x FastDigest™ Buffer with final concentration 1x and 3 units of each restriction enzyme BamHI/NotI (Thermo Fisher FastDigest™). The restriction reactions were incubated at 37°C for 40 minutes and after that run on the gel, as described in chapter 3.1.2.2.

Plasmids with correctly sized inserts were sequenced at Institute of Genomics Core Facility with primer 2066.

3.1.2.8 Plasmid linearisation

Plasmids were linearized for transformation into yeast.The final volume of each reaction was 20 µl and contained milli-Q H2O, 10 ng of plasmid, 10x FastDigest™ Buffer with final concentration 1x, 1 mM of SAM cofactor (Thermo Fisher FastDigest™) and 2,5 units of AjuI restriction enzyme (Thermo Fisher FastDigest™). Then the reaction mixture was incubated for 40 minutes at 37°C. After incubation the enzyme was inactivated at 65°C for 20 minutes.

3.1.2.9 Yeast transformation

Linearized plasmids were transformed into *Saccharomyces cerevisiae* RKI514 (Table 2) which resulted in SF1, SF2, SF3 strains (Table 2).

First of all, the strain RKI514 was taken from the stock and transferred on a YPD plate (10 g/L yeast extract (Formedium), 20 g/L glucose (Oriola), 20 g/L peptone (Formedium)) and incubated overnight at 30°C. Then strain cells were grown in 50 ml YPD media (10 g/L yeast extract (Formedium), 20 g/L glucose (Oriola), 20 g/L peptone (Formedium)) at 30°C shaker and for 8 hours, then optical density (OD) was measured by spectrophotometer Ultrospec 10 (AmershamBiosciences), the wavelength was 600 nm and the calibration sample was 1 ml of YPD media. Cells were harvested for transformation at OD=0.60.

We transferred grown culture into a 50 ml falcon tube and centrifuged 1 minute at 3500 G. Then supernatant was discarded and the pellet was resuspended in 1 ml of milli-Q H2O and then transferred into a new Eppendorf tube using 1 ml of milli-Q H2O. After that we repeated centrifugation at 3500 g for 1 min and discarded supernatant. Then pellet was resuspended in 1 mL sterile P.L.I buffer (100mM of lithium acetate (LiAc) in 0.5xTE buffer (5 mM Tris-HCl (pH 8), 0.5 mM EDTA (pH 8)) and centrifuged again during 1 minute. The addition of the P.L.I buffer and further centrifugation was repeated one more time. After last

centrifugation the pellet was resuspended in 1x of the cell volume with P.L.I buffer. These mixtures were incubated for 10 minutes at room temperature.

During the incubation period the Salmon Sperm DNA (ssDNA) was boiled at 100°C for 10 minutes. And after that we put it to cool down on the ice. The next step was to mix 50 µl of linearized plasmids (approximately 1 ug), 100 µl of yeast competent cells and 10 µl of ssDNA. After that we added 700 µL of sterile PEG/LiAc (100mM lithium acetate, 10 mM Tris-HCl (pH 8), 1 mM EDTA, 40% PEG 3350) solution and 48 µl of DMSO. Then the mixtures were gently resuspended by pipetting and placed at 42°C thermostat for 40 minutes to incubate. After incubation mixtures were placed on ice for 2 minutes, then centrifuged at 6000G for 30 seconds. The supernatant was discarded and 1 ml of sterile 1x TE buffer (10 mM Tris-HCl (pH 8), 1 mM EDTA) was added. Then we repeated centrifugation at the same time and speed and discarded the supernatant. At the end 200 µl of sterile 1x TE buffer (10 mM Tris-HCl (pH 8), 1 mM EDTA) was added, then cells were gently resuspended by pipetting and plated on -LEU plates (0,69 g/L -URA powder (Formedium), 20 g/L glucose (Oriola), 6.7 g/L yeast nitrogen base without amino acids (Formedium), 20 g/L agar)). The plates were placed into an incubator at 30°C.

After two days of incubation we could see colonies on the plates. We picked several colonies from each plate and streaked to another -LEU plate and put it into an incubator at 30°C overnight.

3.1.3 Analysis of the strains

To analyze different strains cells' protein expression levels we did microscopy analysis for measuring and comparing mCherry signals and western blot analysis for identification of the amount of protein synthesized.

3.1.3.1 Preparation of cells for microscopy

To obtain cells in the logarithmic growth phase, a day before microscopy analysis we put cells growing into a 96-well microplate at different concentrations. CSM media (20 g/L glucose (Oriola), 6.7 g/L yeast nitrogen base without amino acids (Formedium), 0.79 CSM (Formedium), 20 g/L agar) with 2% of glucose was used. Cells were grown at 1x, 20x and 40x dilutions. We prepared media, distributed to wells, resuspended cells at different concentrations and put the plate to a 1200 rpm incubator shaker at 30°C overnight.

Additionally to strains of interest, RKI514 strain needed to be grown to assess background autofluorescence.

3.1.3.2 Microscopy experiment

The aim of the microscopy experiment was to measure mCherry signals of the cells and compare the results between SF1, SF2, SF3, RKI0521 and RKI0519 strains' cells. This analysis tracked cell signals every 3 minute during 87 minutes in real time during cell growth.

After overnight incubation the 250 µl of cells were transferred into Eppendorf tubes, then we added 250 µl of CSM media with 2% of glucose and sonicated the cells by Sonopuls ultrasonic homogenizer Bandelin 2450 to break down cell clumps. When sonication was done we proceeded with centrifugation at 3,5g for 1 minute. Then 450 µl of supernatant was removed and pellet was resuspended in the rest of the supernatant. Then we pipetted 0,6 µl of cells from each strain to 24 x 50 x 0.08 mm micro cover glass, then each strain cells' were covered with small pieces of CSM with 2%, glucose 1.5 % and agarose gel (7 g/L yeast nitrogen base (BD Biosciences, 0.79 g/L CSM powder (Formedium), 20 g/L glucose (Oriola), 1.5% NuSieve GTG agarose (Lonza)). When cells were covered, the longer pieces of CSM glucose agarose gel were used to place around the smaller pieces that cover cells. After that, we placed 20 x 20 mm micro cover glass on top of the gell pieces and covered this construction with a small plastic box, to avoid gel pieces drying out.

During this experiment we used Zeiss Axio Observer.Z1 microscope with 63x/1.4 Oil M27 Plan-Apochromat objective (Zeiss). To perform imaging of fluorescence Colibri LED modules at 25% power were used, the exposure time of mCherry was 750 ms at 555 nm. This experiment lasted for 87 minutes, it had 30 time points and 3 minutes in between each time point. During the experiment temperature was controlled by Tempcontrol 37-2 digital (Pecon) and was 30°C. We chose 25 position and they were automatically followed by automated stage and ZEN software. Images of cells were exported and analyzed for single cell mCherry values using MatLab with custom scripts for cell segmentation and tracking. Data was visualized using Microsoft Excel.

3.1.3.3 Preparation of cells for Western Blot

Cells were put to grow into glass tubes containing 5 ml of YPD media at 30°C shaker overnight. After 8 hours we measured OD by spectrophotometer Ultrospec 10 (Amersham Biosciences) at 600 nm wavelength (0.1of OD600 = 1*106 cells/mL), 1 ml of YPD was used as a blank of measurements. The OD of cultures was between 0,6-0,8 and we transferred

cultures into 15 ml tubes. Then we proceeded with centrifugation at 3500g during 1 minute, removed supernatant and froze the cells in liquid nitrogen. Then cells were stored in the freezer at -80°C.

3.1.3.4 Western Blot

15 ml tubes that contained cells were taken from the -80°C freezer and put on ice. After that, we resuspended cell pellet in 200 µl of urea lysis buffer (20 mM Tris (pH 7.4), 8 M Urea, 2M Thiourea, 4% CHAPS, 1% DTT, 50 mM NaF, 89 mM BGP, 1 mM Na3VC4), transferred this mixture into Eppendorf tubes and added 200 µl of glass beads. Then we placed tubes into a FastPrep-24 bead beater (MB Biomedicals) and ran it at 4 meters/second for 40 seconds to disrupt the cells.

The next step was to transfer lysate into another Eppendorf tube, so we prepared new Eppendorf tubes. Then using a needle we made a hole at the bottom of the tubes with beads and placed them on top of new tubes. To transfer the lysate completely we spinned down the cells for a few seconds. When lysate was transferred to new tubes, we centrifuged them at 11000g for 10 minutes. Then 2 µl of supernatant was used to perform Bradford assay, as it is described in chapter 3.1.6.5. Bradford assay was necessary to load the gel with the equal amount of protein in lysate. When Bradford assay was completed, we stained supernatant (volume was defined according to Bradford assay results) with 3x SDS buffer (375 mM Tris-HCl 9% SDS 50% Glycerol 0.03% Bromophenol blue) with final concentration of 1x and loaded the 10% acrylamide gel (separating gel: 0.375 M Tris-HCl (pH 8,8), 10% acrylamide (29:1 acrylamide:bis-acrylamide), 0,1% SDS; stacking gel: 0.125M Tris-HCl (pH 6,8), 5% acrylamide (29:1 acrylamide:bis-acrylamide), 0,1% SDS). For being able to measure protein sizes 4 µL of PageRulerTM Prestained Protein Ladder was mixed with 5 µL of 3x SDS buffer and loaded into the gel. The gel was run for 60 minutes at 15 mA current.

When the gel was finished running, we placed it into the Semi-Dry buffer (25 mM Tris-HCl, 192 mM glycine, 0,1% SDSI) for 15 minutes. During this time 4 filter papers and 0.45 µm nitrocellulose blotting membrane were cut out and placed to soak with a gel to Semi-Dry buffer. The size of filter papers and membrane were identical with gel size. When 15 minutes were over, we constructed a “sandwich” made of 2 layers of filter papers on the bottom, nitrocellulose membrane, gel on top of a membrane and 2 layers of filter papers on the top. Then we used Pierce G2 Fast Blotter (Thermo Scientific) with the standard semi-dry transfer program for 1 hour for transferring proteins to the membrane.

When the program finished, membrane was transferred into a blocking solution (5% milk powder, 1x TBS-T buffer (20 mM Tris-HCl, 150 mM NaCl, 0.1% Tween20)) and kept overnight on a tilting shaker at 4°C room.

The next day we placed the membrane into the room temperature for incubation in the primary antibody solution (3% milk powder, 1x TBS-T buffer (20 mM Tris,150 mM NaCl, 0.1% Tween20), 1:500 dilution of mouse anti-HA antibody (Biolegend), 0.03% N3Na). The incubation period was 30 minutes and took place on the tilting shaker at room temperature. Then the membrane was washed 1x15 minute and 3x5 minutes in 1x TBS-T buffer on a shaker. When was washing was finished, we incubated membrane in secondary antibody solution (3% milk powder, 1x TBS-T (20 mM Tris-HCl, 150 mM NaCl, 0.1% Tween20), 1:7500 dilution of goat anti-mouse HRP-conjugated antibody (LabAs)) for 30 minutes on a tilting shaker at room temperature. After incubation the washing with 1x TBS-T buffer was repeated the same amount of time.

After washing the membrane we prepared a 1:1 solution of SuperSignalTM West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific) and rinsed the membrane with it. This step enables visualization of protein bands on the film. Then we covered the membrane with cellophane.

The next step was to detect the exact protein signals, for that we used Odyssey FC imaging system, where chemilumescence signal was quantified for 2 minutes.

During the final step we exposed membrane to autoradiography film (AGFA Medical X- ray film blue, Belgium) for 30 seconds, this step took place in a dark room. Then we placed film into a G150 developer (AGFA) bath for 1 minute and dried it for 15 seconds. After that the film was washed in water for 10 seconds, then dried for 15 seconds. Finally, we placed the film into a G354 fixer (AGFA) bath for 20 seconds and dried for 20 seconds afterwards. Then we rinsed the film with distilled water and placed it inside a fume cup board for drying out. The final step was to mark the PageRulerTM Prestained Protein Ladder bands.

3.1.3.5 Bradford Assay

To perform Bradford assay we diluted PierceTM Coomassie Plus (Bradford) Assay Reagent (Thermo Scientific) 5 times and filled 96-well microplate with 200 µl of solution. Then we resuspended 2 µl of lysate and incubated the plate for 5 minutes at room temperature. Absorbance at 600 nm was measured using Magellan ™ microplate reader.

3.2 RESULTS

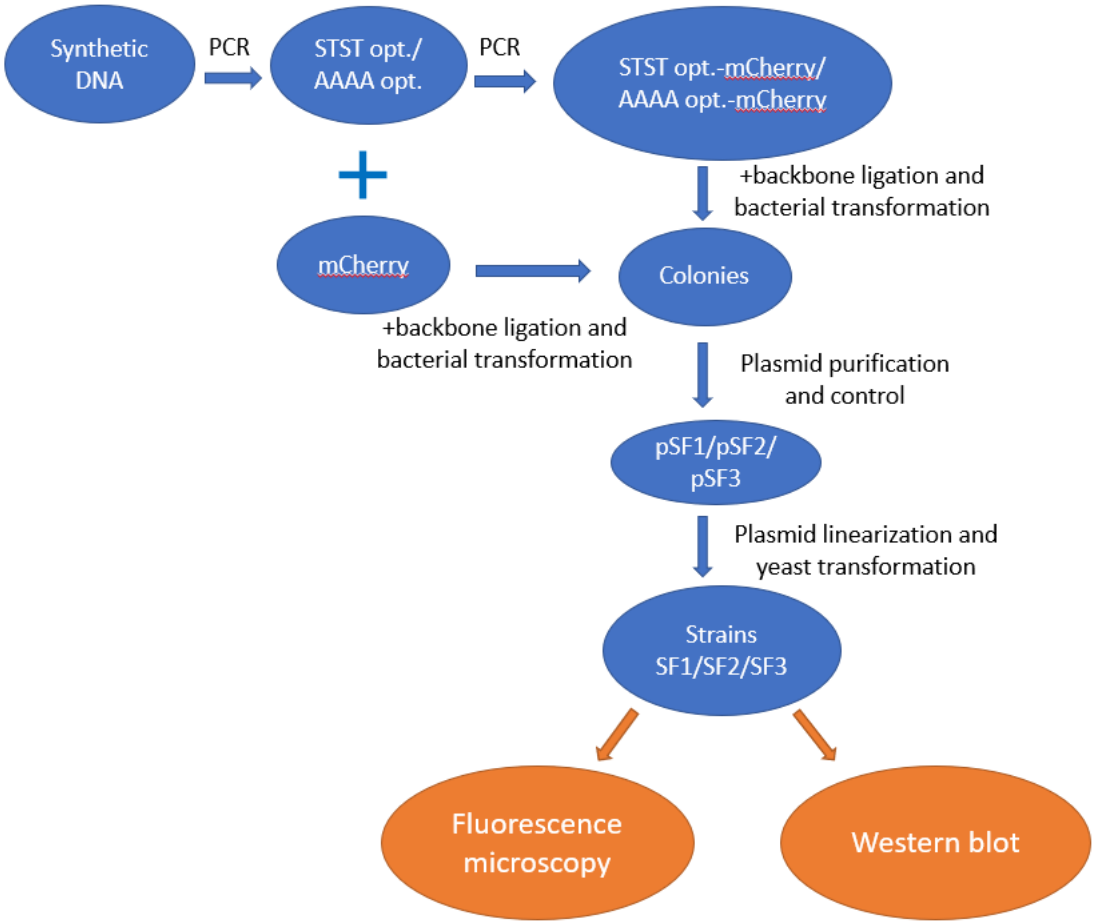

Figure 7. Workflow.

In this section the results of the work are provided. During our work (Figure 7) the next strains

were made:

- SF1 (STST opt.-mCherry)
- SF2 (AAAA opt.-mCherry)
- SF3 (mCherry)
- RKI0519 (STST-mCherry)
- RKI0521 (AAAA-mCherry)

Codon optimized STSTopt.-mCherry sequence and STST-mCherryhave the same amino acid sequence. AAAAopt.-mCherry optimized sequence and AAAA-mCherry have also the same sequence of amino acids. However, there is a difference in codons that code the same sequences.

In this paragraph we describe microscopy and western blot experiments results. Also we describe analysis of the nucleotide composition, RSCU values and codon usage. For these calculations we used the tool at http://genomes.urv.es/CAIcal/ website (Puigbo et al., 2008). The graphs, tables and histograms were made using Microsoft Excel.

3.2.1 Sequences analysis

3.2.1.1 RSCU analysis results

RSCU stands for Relative Synonymous Codon Usage. We wanted to identify underrepresented and overrepresented codons, as was shown in Deb et al. studies with Hepadnaviridae family. The codons with RSCU value over 1,6 are overrepresented and with RSCU value under 0,6 are underrepresented (Deb et al., 2020).

In our studies we would like to show the usage of codons using RSCU values. We compared STSTopt.-mCherry optimized sequence and STST-mCherry sequences. We did not compare sequences with AAAA degron between each other, since they differ between STST degron sequences only by 4 amino acids. In STSTopt.-mCherry optimized sequence we found 12 overrepresented codons in (TTG, GTT, TCT, CCA, GCT, AGA, ACT, GGT, TGT, ATT, CAA) and 10 underrepresented codons (ATC, ATA, GAC, CGT, TTA, CTG, GGA, GAG, CCC, CAG). In not optimized sequence STST-mCherry there are 7 overrepresented codons (TTA, AGA, TAC, TGT, CCA, GTT, GCC) that is 3 codon less than in optimized sequence and 11 underrepresented codons (ATC, CCG, CAT, CGG, TTG, CTC, AAA, ACA, GGG, GCA, TCG), 1 codon more that optimized sequence has (Figure 8). That means STSTopt.-mCherry optimized sequence uses more identical codons to code for amino acids than not optimized one.

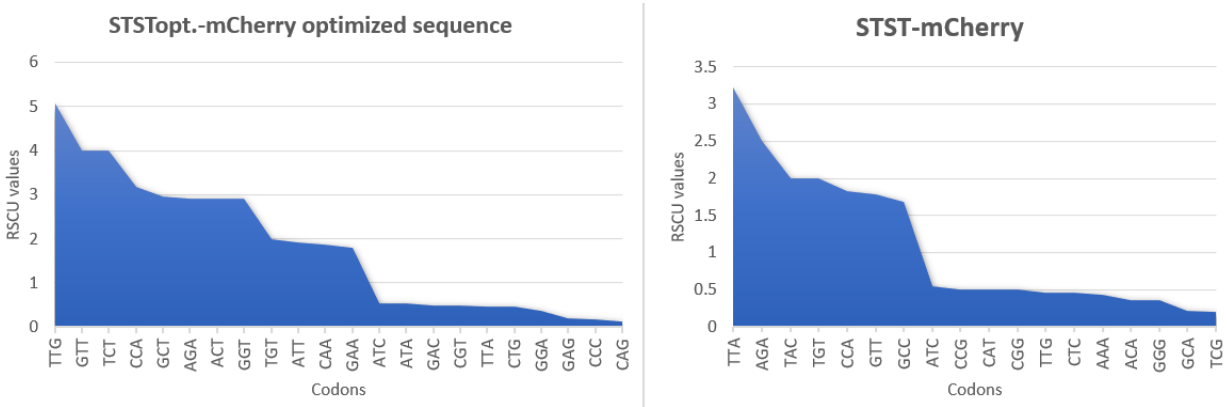

Figure 8. Overrepresented and underrepresented codons of STSTopt.-mCherry optimized sequence.

3.2.1.2 Codon usage

The table below (Table 6) represents codon usage of STSTopt.-mCherry optimized sequence and STST-mCherry sequences. It shows all existing codons, amino acids that are coded by these codons and numbers show the amount of codons represented in the sequence. Each amino acid associated codon and number of codons are shown in different colors.

From the table below (Table 6) we can see that STST-mCherrysequence uses a bigger selection of codons. The optimized sequence has less variety of codons, but uses one codon more times. STSTopt.-mCherry optimized sequence uses most often TCT codon (20 times) that codes for serine while STST-mCherry sequence uses GAA codon more frequently that encodes g*lutamic acid.* We can also see that STSTopt.-mCherry optimized sequence does not use such codons, as CTT, CTC, CTA, TCG, AGT, AGC, GGG, GTC, GTA, GTG, CCG, ACA, ACG, GCA, GCG, CGC, CGA, CGG, TGC. And STST-mCherry does not use CTA, TAT, GTA, CGC, CGA and TGC codons.

Taking these results into account, we can see that although the amino acid sequence is the same, codon usage is different between these two sequences.

Table 6. Codon usage of STSTopt.-mCherry optimized sequence and STST-mCherry sequence.

CODON USAGE															
CODONS	TTT	TTC	TTA	TTG	CTT	CTC	CTA	CTG	TCT	TCC	TCA	TCG	AGT	AGC	GAA
Aminoacids	F	F	L	L	L	L	L	L	S	S	S	S	S	S	E
STSTopt.-mCherry optimized sequence	2	4	1	11	0	0	0	1	20	7	3	0	0	0	17
STST-mCherry	2	4	7	1	2	1	0	2	8	5	7	1	4	5	13
CODONS	ATT	ATC	ATA	GGT	GGC	GGA	GGG	ATG	TAT	TAC	CAT	CAC	CAA	CAG	GAG
Aminoacids	I	I	I	G	G	G	G	M	Y	Y	H	H	Q	Q	E
STSTopt.-mCherry optimized sequence	7	2	2	8	2	1	0	7	2	2	2	2	16	1	2
STST-mCherry	5	2	4	4	2	4	1	7	0	4	1	3	8	9	6
CODONS	GTT	GTC	GTA	GTG	CCT	CCC	CCA	CCG	AAT	AAC	AAA	AAG	GAT	GAC	TGT
Aminoacids	V	V	V	V	P	P	P	P	N	N	K	K	D	D	C
STSTopt.-mCherry optimized sequence	9	0	0	0	4	1	19	0	6	12	8	6	9	3	1
STST-mCherry	4	3	0	2	5	5	11	3	6	12	3	11	5	7	1
CODONS	ACT	ACC	ACA	ACG	GCT	GCC	GCA	GCG	CGT	CGC	CGA	CGG	AGA	AGG	TGC
Aminoacids	T	T	T	T	A	A	A	A	R	R	R	R	R	R	C
STSTopt.-mCherry optimized sequence	8	3	0	0	14	5	0	0	1	0	0	0	9	2	0
STST-mCherry	2	4	1	4	6	8	1	4	3	0	0	1	5	3	0

3.2.1.3 Nucleotide composition analysis

For sequence comparison we used Clustal 2.1 Multiple Sequences Alignments software. The results showed that there is 79.29% of identity between NES-STST-mCherry-3HA optimized and NES-STST-mCherry-3HA sequences. And 79.01% of identity between AAAAopt.-mCherry optimized and AAAA-mCherry sequences.

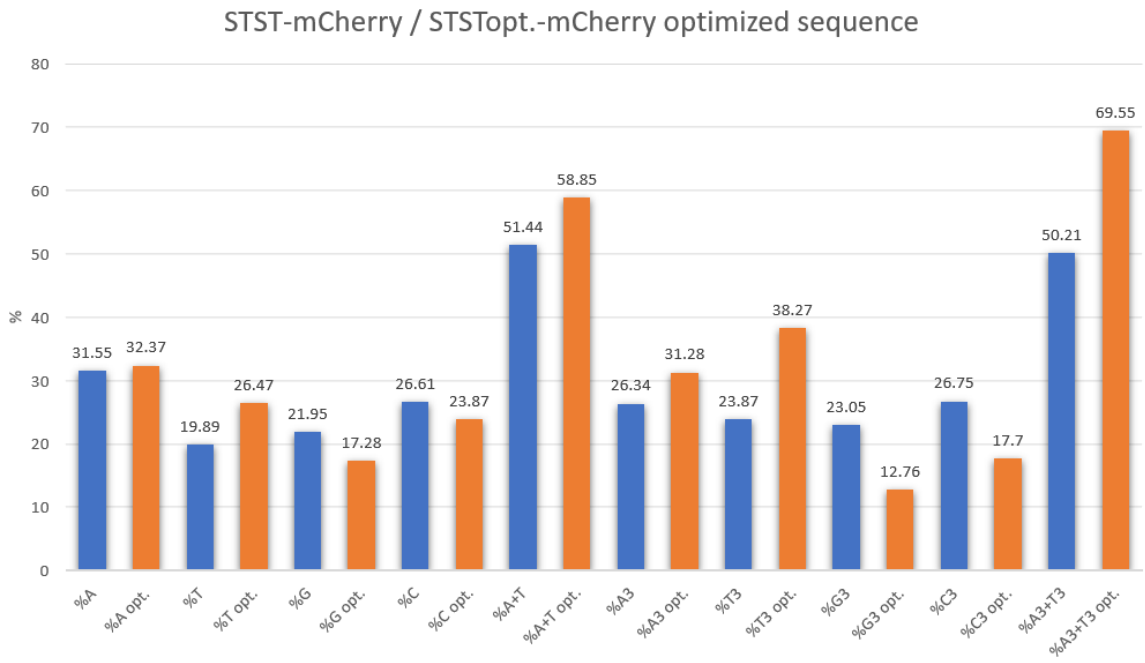

Figure 9. Percentage of nucleotides, percentage of nucleotides at 3rd position in codon, percentage of A-T nucleotide pairs and percentage of A3-T3 nucleotide pairs located at 3rd position of codon in STSTopt.-mCherry optimized sequence (blue bars) and STST- mCherry sequence (orange bars).

Figure 9 shows nucleotide composition of STSTopt.-mCherry optimized sequence and STST-mCherry sequence. Also, it illustrates the percentage of each nucleotide at 3rd position in codon.

We can see that both sequences are A-T nucleotide pair rich. STSTopt.-mCherry optimized sequence has more A-T nucleotide pairs than STST-mCherry sequence. 69,55% of A3-T3 nucleotide pairs are located at codon's 3rd position in optimized sequence, while non optimized sequence has only 50,21% of A3-T3 pairs at 3rd position. Also the graph shows that both sequences are rich with A nucleotides. In STSTopt.-mCherry optimized sequence at the codon's 3rd position T3 nucleotide appears more frequently and for non optimized sequence C3 shows up more times.

3.2.2 Protein expression measurement

To test, if different codon usage of tag has an effect on protein expression, we measured relative protein amount in cells by Western blotting and fluorescence microscopy.

3.2.2.1 Western Blot results

Figure 10 represents western blot results. Samples are listed below:

1. AAAAopt.-mCherry optimized sequence
2. STSTopt.-mCherry optimized sequence
3. AAAA-mCherry
4. STST-mCherry
5. mCherry colony 1
6. mCherry colony 2

The protein molecular weights are shown by a ladder on the right side of Figure 12 and are represented in kDa. From this figure we can see that AAAAopt.-mCherry optimized sequence, STSTopt.-mCherry optimized sequence, AAAA-mCherry, STST-mCherry have the same molecular weight and it is equal to 72 kDa. mCherry colony 1 and mCherry colony 2 have the same molecular weight approximately equal 40 kDa.

Figure below (Figure 10) shows equal signal between AAAAopt.-mCherry optimized sequence, STSTopt.-mCherry optimized sequence, AAAA-mCherry, STST-mCherry and between mCherry colony 1, mCherry colony 2.

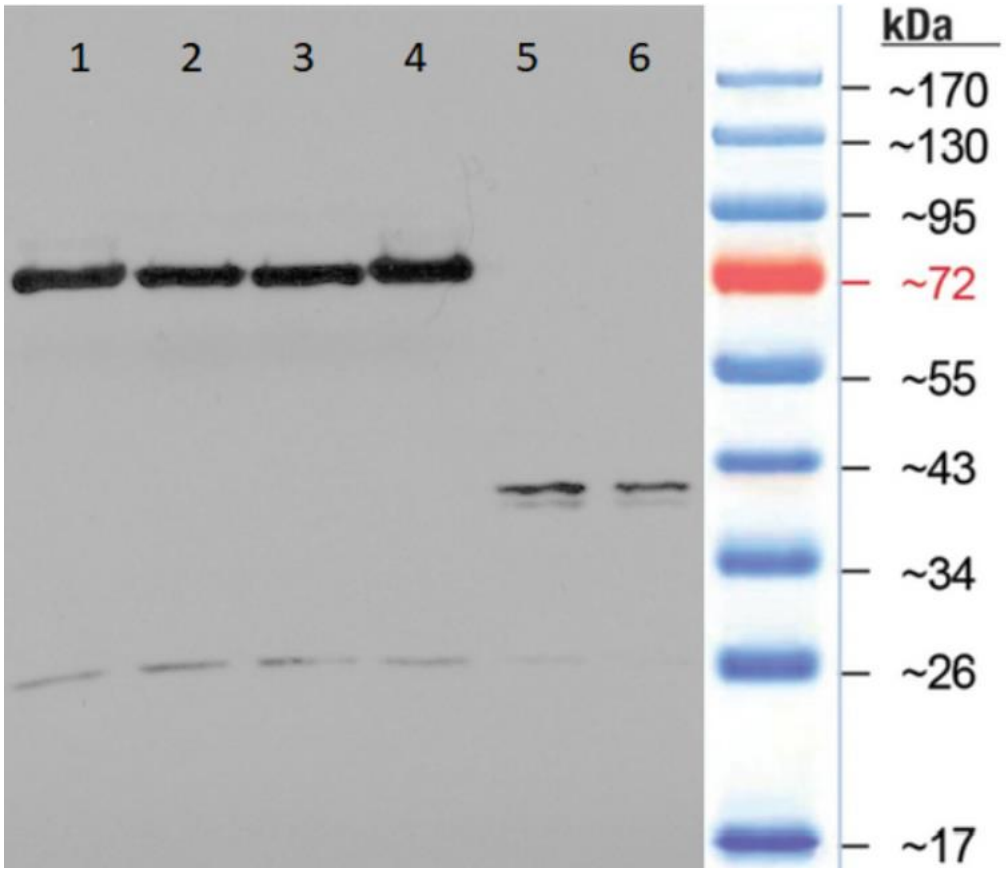

Figure 10. Western Blot results. (1) AAAA opt.-mCherry (2) STST opt. -mCherry (3) AAAA opt.-mCherry (4) STST-mCherry (5) mCherry colony 2 (6) mCherry colony 1

For quantification of protein signals we used Odyssey FC imaging system. Figure 11 shows the results.

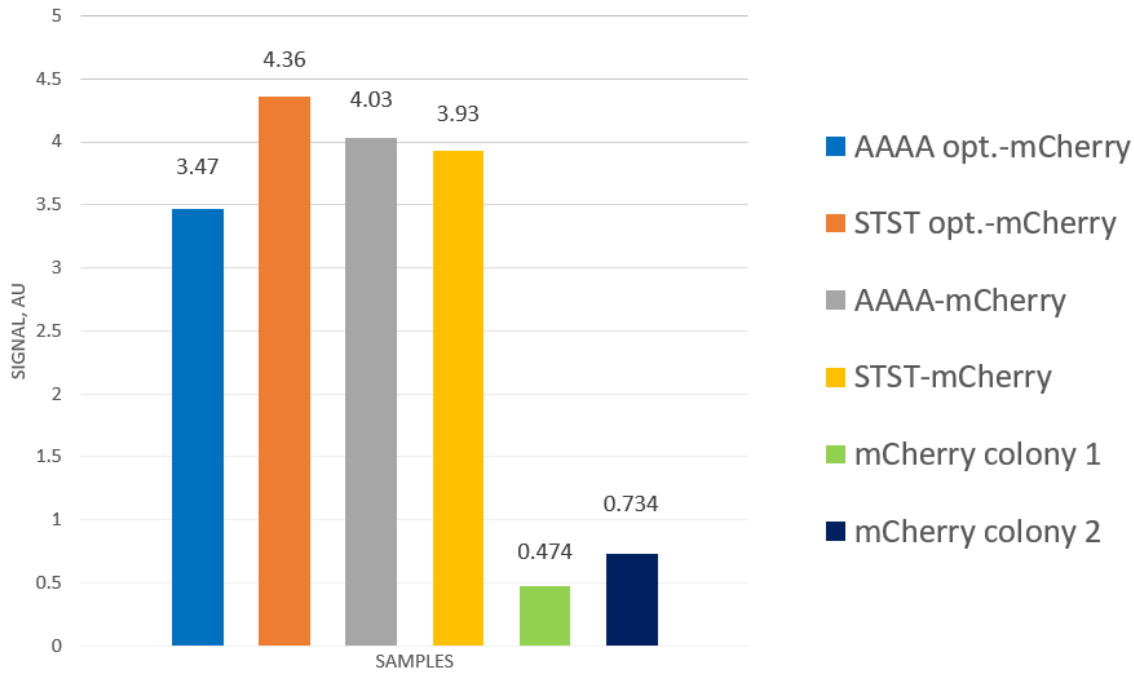

Figure 11. Western blot analysis results obtained by Odyssey FC imaging system.

Figure above (Figure 11) shows that protein expression between AAAAopt.-mCherry optimized sequence, STSTopt.-mCherry optimized sequence, AAAA-mCherry, STST-mCherry is slightly different. The higher signal is obtained by STSTopt.-mCherry optimized sequence. Also we can see that STSTopt.-mCherry optimized sequence has higher signal than STST-mCherry, while AAAA-mCherry has higher signal than AAAAopt.-mCherry optimized sequence. The signal of mCherry colony 1 and mCherry colony 2 is much lower than the signal of sequences that contain degron and slightly varies between not containing degron sequences.

3.2.2.2 Fluorescence microscopy results

Microscopy experiment was done to see the expression of mCherry in different strains. Figure 12 illustrates changes of mCherry expression over time obtained by AAAAopt.-mCherry optimized sequence, AAAA-mCherry, STSTopt.-mCherry optimized sequence, STST-mCherry, mCherry colony 1 and mCherry colony 2. And Figure 13 shows the average mCherry signal obtained from the same strains.

From these 2 figures (Figure 12, Figure 13) we can see that AAAA-mCherry optimized and not optimized sequences have much higher mCherry signal compared to STST-mCherry optimized and not optimized sequences. Also optimized sequences have slightly higher mCherry signal than not optimized sequences. The sequences that do not have degron part mCherry colony 1 and mCherry colony 2 are located in between degron containing sequences by mCherry signal expression.

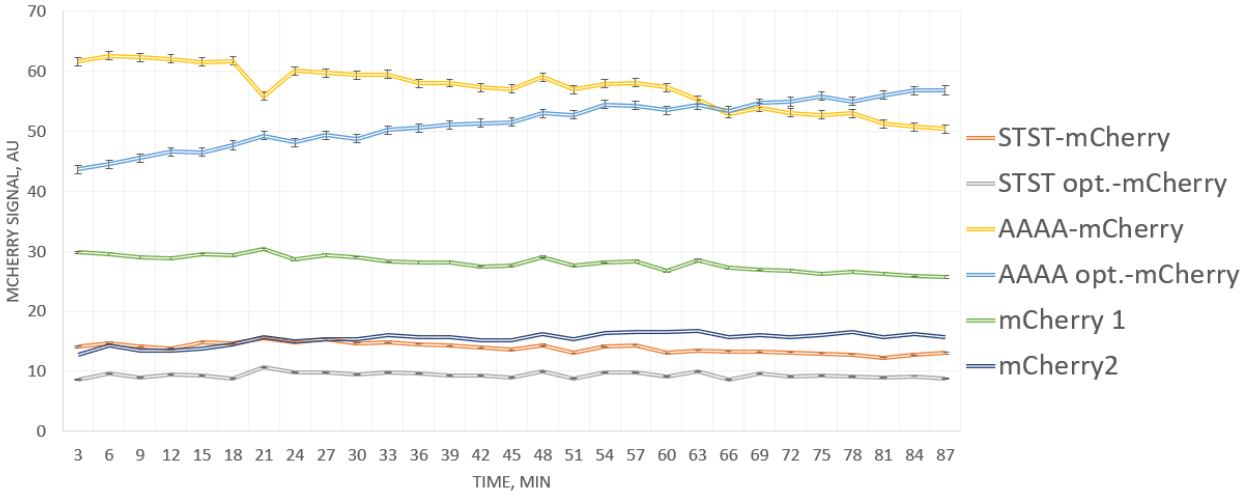

Figure 12. Timelapse fluorescence microscopy results. Changes of mCherry signals over time.

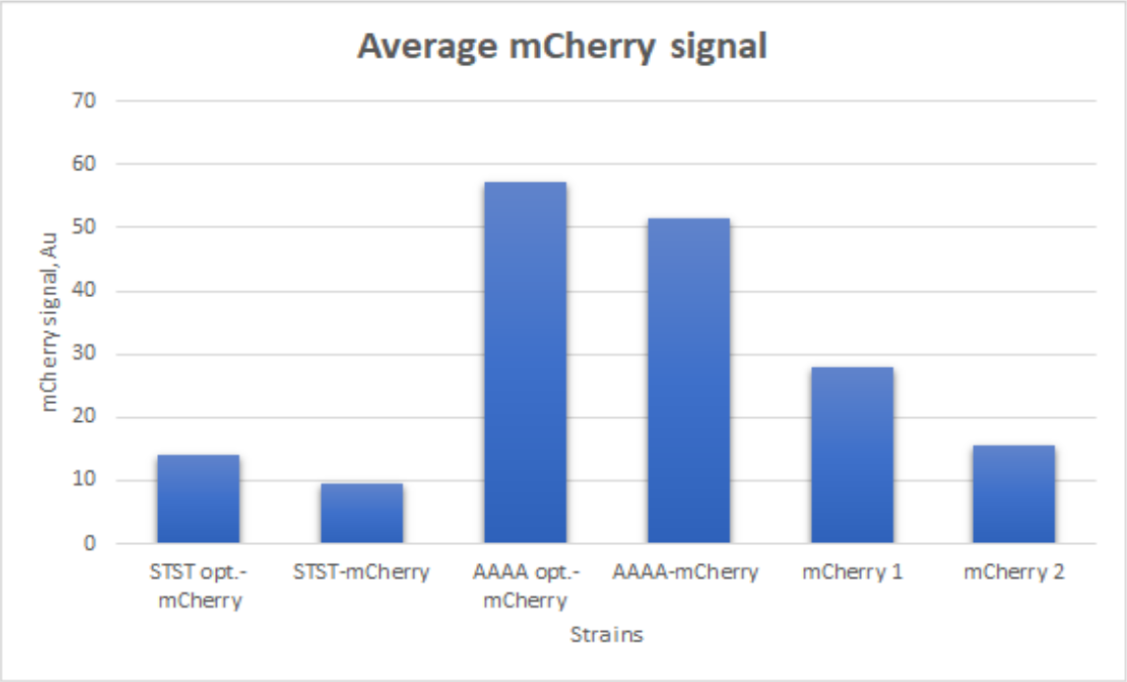

Figure 13. Fluorescence microscopy average mCherry signal from Figure 12.

3.3 DISCUSSION

In this study we prepared strains that have AAAAopt.-mCherry optimized sequence, STSTopt.-mCherry optimized sequence and mCherry sequence and compared their protein expression levels to non-optimized sequences in yeast. The first two sequences differ only by 4 amino acids and the last one is a control sequence that does not contain a degron module.

We did RSCU analysis, as it was described in Deb et al. Studies (Deb et al., 2020). Although our work goals were not associated with CUB factors contribution, we were interested in sequences' composition and RSCU. RSCU results showed that there is higher usage of particular codons in optimized sequences (more overrepresented codons), while not optimized sequences contain more underrepresented codons, that means this sequence has more codon variety, using larger amount of synonymous codons compared to optimized sequence.

The nucleotide composition of these sequences showed that there is 79.29% of overlap between optimized and not optimized STST-mCherry containing sequences, which translates to quite different codon usage, as is evident from Figure 8 and Table 6. There is no big difference in nucleotide composition. However, the 3rd position of codon in optimized sequence is A and T nucleotide rich, while not optimized sequence contains more G and C nucleotides at 3rd position.

We got mixed results of protein expression efficiency obtained from microscopy and Western Blot experiments. The microscopy experiment showed that the highest fluorescence signal comes from the AAAAopt.-mCherry optimized sequence, while the same sequence shows the lowest amount of protein in Western Blot experiment. Also, STSTopt.-mCherry optimized sequence is on the lower side by mCherry signal emission, but has the highest protein level in Western Blot assay. The sequences that do not contain the degron module showed lowest signal during Western Blot, but microscopy results showed average amount of signal compared to different degron-containing sequences. This shows that tagging of mCherry with degron module has an effect on its fluorecence properties and this has to be taken into account during future experiments. Therefore in this context for assessing protein levels western blotting is prefered method and further conclusions will be based on Western Blot results.

According to Western Blot experiment results we can assume that addition of degron modules to mCherry either accelerates protein expression rate or increases its stability. It

remains to be tested, if this phenomenon carries over to other proteins in yeast. There are minor differences in protein levels between two degrons modules, and between optimized and not optimized sequences (Figure 11). Therefore codon optimization doesn't have any meaningful effect for protein expression in our degron tag design.

SUMMARY

The topic of codon usage is not completely understood and requires more detailed studies between various genomes for better understanding of its working algorithms. There are different assumptions of its effect on protein expression and results varies from study to study.

In this work we tried to understand how adding degron modules affect protein expression. Also, we wanted to find if there is difference between optimized and not optimized sequences on protein expression levels.

During these studies we prepared strains that contain sequences of interest: STSTopt.-mCherry optimized sequence, STST-mCherry, AAAAopt.-mCherry optimized sequence, AAAA-mCherry and mCherry. And the protein expression was compared using Western Blot, fluorescence microscopy and sequence composition analysis.

In summary, western blotting experiment demonstrated that addition of the degron tag increased protein expression. Surprisingly, small changes in protein sequence had effect on mCherry fluorescence properties.

REFERENCES

Belfield G.P.; Tuite M.F. Translation elongation factor 3: A fungus-specific translation factor? *Molecular Microbiology*, 1993, vol. 9, no. 3, pp. 411-418.

Bennetzen J.L.; Hall B.D. Codon selection in yeast. *Journal of Biological Chemistry*, 1982, vol. 257, no. 6, pp. 3026-3031.

Crick F.H.C.; Barnett L.; Brenner S.; Watts-Tobin R.J. General nature of the genetic code for proteins. *Nature*, 1961, vol. 192, pp. 1227-1232.

Deb B.; Uddin A.; Chakraborty S. Codon usage pattern and its influencing factors in different genomes of hepadnaviruses. *Archives of Virology*, 2020, vol. 165, pp. 557-570.

Dittmar K.A.; Goodenbour J.M.; Pan T. Tissue-specific differences in human transfer RNA expression. *PloS Genet*., 2006, vol. 2, no. 12, pp. 221.

Emilsson V.; Kurland C.G. Growth rate dependence of transfer RNA abundance in Escherichia coli. *EMBP J*., 1990, vol. 9, no. 13, pp.4359-4366.

Ford T. Addgene Blog. https://blog.addgene.org/plasmids-101-codon-usage-bias (27.09.2018)

Goldman E. tRNA and the Human Genome. *John Wiley &* Son, 2011.

Griffiths A.J.F.; Miller J.H.; Suzuki D.T.; et al. An Introduction to Genetic Analysis. 2000, vol. 7.

Hinnebusch A.G. Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiol Mol Biol Rev*., 2011, vol. 75, no. 3, pp. 434–467.

Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. Journal of Molecular Biology, 1981, vol. 151, no. 3, pp. 389-409.

Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. Journal of Molecular

Biology, 1981, vol. 151, no. 3, pp. 389-409.

Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 2005, vol. 361, no. 21, pp. 13-37.

Kubyshkin V.; Acevedo-Rocha C.G.; Budisa N. On universal coding events in protein Biogenesis. Biosystems, vol. 164, pp. 16-25.

Laursen B.S.; Sorensen H.P.; Mortensen K.K; Sperling-Petersen H.U. Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev.* 2005, vol. 69, no. 1, pp. 101–123.

Mauro P.M.; Chappell S.A. A critical analysis of codon optimization in human therapeutics. *Trends in Molecular Medicine,* 2014, vol. 20, no. 11, pp. 604-613.

Alberts B.; Johnson A.; Lewis J.; Raff M.; Roberts K.; Walter P. Molecular Biology of the Cell. 2008, vol. 5.

Pechmann S.; Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol*., 2013, vol. 20, no. 2, pp. 237-243.

Plotkin J.B; Robins H.; Levine A.J. Tissue-specific codon usage and the expression of human genes. *PNAS*, 2004, vol. 101, no. 34, pp. 12588-12591.

Puigbo P.; Bravo I.G.; Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. Biology Direct, 2008, vol. 3, no. 38.

Rever, Kinzy, Pavitt. Mechanism and Regulation of Protein Synthesis in *Saccharomyces cerevisiae. Genetics,* 2016, vol. 203, no. 1, pp. 65-107.

Rosano G.L.; Ceccarelli E.A. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microbial Cell Factories,* 2009, vol. 8, no. 41.

Saier M.H. Understanding the Genetic Code. *Journal of Bacteriology*, 2019, vol. 201, no. 15, pp. 91-119.

Sander I.M.; Chaney J.L; Clark P.L. Expanding Anfinsen's Principle: Contributions of Synonymous Codon Selection to Rational Protein Design. *Journal of the American Chemical Society*, 2014, vol. 136, no. 3, pp. 858-861.

Schmeing T.M.; Ramakrishnan V. What recent ribosome structures have revealed about the mechanism of translation. Nature, 2009, vol. 461, pp. 1234-1242.

Tamura K. Origins and early evolution of the tRNA molecule. *Life*, 2015, vol. *5*, no. 4, pp. 1687-1699;

Tuller T.; Waldman Y.Y.; Kupiec M.; Ruppin E. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A*, 2010, vol. 107, no. 8, pp. 3645-3650.

Wright F. The 'effective number of codons' used in a gene. Gene, 1989, vol. 87, no. 1, pp. 23-29.

NON-EXCLUSIVE LICENCE TO REPRODUCE THESIS AND MAKE THESIS PUBLIC

I, Sofija Frolova,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work from 20/05/2024 until the expiry of the term of copyright,

Effect of protein tag's codon usage on protein expression in Saccharomyces cerevisiae,

supervised by Rait Kivi PhD.

2. I am aware of the fact that the author retains the rights specified in p. 1.

3. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Sofija Frolova

20/05/2021