# DIRECTED EVOLUTION OF SYNTHETIC RIBOSWITCHES AND A LEUCYL tRNA SYNTHETASE

by

ANDREA GABRIELA GUEDEZ PENA

Bachelor of Science, 2008
Universidad de Carabobo
Valencia, Venezuela

**A Dissertation**

Submitted to the Faculty of
The College of Science and Engineering
Texas Christian University
in partial fulfillment of the requirements for the degree of

**Doctor in Philosophy**

**May 2022**

**APPROVAL**

DIRECTED EVOLUTION OF SYNTHETIC RIBOSWITCHES AND A LEUCYL tRNA SYNTHETASE

by

Andrea Gabriela Guedez Pena

Dissertation approved:

Major Professor

MONTCHAMP

For the College

## ACKNOWLEDGEMENTS

To God who has been my guide during my whole life, and I am so thankful for being blessed in so many ways with all the people that surround me and have prayed for me to accomplish this.

I want to thank my family for all their support. They have always been there for me, and I really appreciate their support, prayers, and words of encouragement. To my husband for being willing to take care of our son during my first three years of graduate school and for being a stay-at-home dad so I could go and learn. To my son for all his patience and understanding when mom was tired and overwhelmed. I love you both!

To my advisor, Dr. Youngha Ryu for all his support and patience, for listening to me when I just wanted to break down and cry, and for giving me another perspective. Thank you!

To my committee members: Dr. Stewart, thank you for believing in me and for your insightful scientific and career advice. Dr. Montchamp, for making me appreciate organic chemistry, for taking the time to explain things that were not so obvious to me and for making me an honorary member of his lab. I appreciate it! Dr. Annunziata, thank you for your guidance, and support. I admire your ability retain so much knowledge and information, and to transmit it without having a piece of paper.

To the Chemistry Department Faculty, especially: Dr. Green, you are such an inspiring person and I really appreciate all your advice, and support. Thank you for making academia so much better! Dr. Coffer, thank you so much for believing in me, encouraging me, and pushing me to be a better person. Dr. Dzyuba, thank you for your guidance with NMR and CD. Dr. Fry, thank you for your patience when my English was not so good, and I could not put in words what was in my mind! I also appreciate you letting me teach for you when you were unavailable.

To my dear friends and colleagues: Karen Winters, you were (and still are) such an important

part of my life. Thank you for being so kind, and for showing me a simpler perspective in life, for your guidance while running organic reactions, and for all your help. Marlius, Ely, David, Magy, Katie and Debora: Thank you for your constant prayers and support, for your encouraging notes, and for all the "little things" that made such a difference. Thank you Shamby, Will, Alex, Casey, and Daniel for making the chemistry department a better place. I appreciate you!

Thank you to Nick, Maegyn, Axel, Chelsea, Kristof, and Jason Mars for your help.

Thanks to every senior graduate student that supported me when I was starting: Nguyen, Anne, Stephanie, Aisha: Thank you for your advice, encouragement, support and fun conversations. Marianne, thank you for training me for biochemistry lab and for your encouraging words. Arshad, Nelli, Adam, Samantha, and Charles: thank you for your support. Each one of you were helpful and I am grateful to have gotten to know.

I also want to thank my amazing high school and undergraduate students, especially Mitch, Alex, and Emily. You worked so hard and listened to my suggestions. I am proud of you!

I cannot leave out Page, Rich, Carla, Heidi and Jerry. There are no words to express how thankful I am for every one of you. The department would be so different without you!

To Dr. Fabrizio Donnarumma at LSU Mass Spec facility, for all your support and for making sure I understood the data you sent me.

To International services for all the support answering all my questions.

To the College of Science and Engineering for financing my education and for giving me a SERC grant, and TCU-RCAF and Andrews Institute (to Y.R.) for funding.

Dr. Neilson… You were the first person to answer my emails when I decided to apply to TCU. Thank you! We miss you in the department.

*Thanks to every person that made it possible to be where I am: the ones that inspired me, the ones that supported along the way…*

## PUBLICATIONS INCLUDED IN THIS DISSERTATION

The text and figures in Chapter 5, in part or in full, is a reprint of the material as it appears in:

Guedez, Andrea; Sherman, Matt; Ryu, Youngha. “Dual genetic selection of the theophylline riboswitch with altered aptamer specificity for caffeine”. Biochem. Biophys. Res. Commun. **2021**, *579*, 105-109.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF SCHEMES

## LIST OF ABBREVIATIONS

4-nitro-1-phe: 4-nitro-1-phenylalanine

5'-UTR: 5'- untranslated region

5-FU: 5-Fluorouracil

A: Adenine

aaRS: Aminoacyl synthetases (e.g., IleRS: Isoleucine synthetase)

AdoCbl: Ado cobalamin

Ala (A): Alanine

AMP: Adenosine monophosphate

Arg (R): Arginine

Asn (N): Asparagine

Asp (D): Aspartate

ATP: Adenosine triphosphate

*BamHI:* Restriction enzyme from *Bacillus amyloliquefaciens.* It cleaves the recognition sequence 5'-G/GATCC-3

*BL21(DE3): E. coli* cells used for high level protein expression

BOC-Dap-OH: (S)-2-[(tert-Butoxycarbonyl)amino]-3-aminopropionic acid

C: Cytidine

CaffRS: Caffeine Riboswitch

CAT: Chloramphenicol acetyl transferase protein (*cat:* chloramphenicol acetyl transferase encoding gene)

CD: Circular dichroism

c-di-GMP: Cyclic di guanosine monophosphate

CDR: Chemically rich defined media

c.f.u: colony forming units

*ClaI:* Restriction enzyme derived from *Caryophanon latum L.* It cleaves DNA at 5′-AT/CGAT-3′ sequence

CP1: Editing domain of the leucyl-tRNA synthetases (ΔCP1: lacking the editing domain)

Cys (C): Cysteine

Da: Daltons (kDa: Kilodaltons)

Dansyl-Dap: 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid

dd$H_2O$: MilliQ water

DEPC: Diethyl pyrocarbonate

*DH10B:* A cloning strain of *E. coli.*

dNTPs: deoxynucleoside triphosphates

*DraIII:* Restriction enzyme derived from *Deinococcus radiophilus.* It cleaves the sequence 5'-CACNNN/GTG-3'

DSC: Differential scanning calorimetry

dTMP: deoxy thymidyl monophosphate

dUMP: deoxy uracil monophosphate

*EcoRI:* Restriction enzyme derived from *E. coli.* It cleaves the sequence 5'- G/AATTC-3'

EDTA: Ethylenediaminetetraacetic acid

ESI-MS: Electrospray ionization mass spectrometry

FRET: Fluorescence resonance energy transfer

G: Guanine

*gcvPA:* Gene that encodes probable glycine dehydrogenase

*gcvPB:* Gene that encodes probable glycine decarboxylase

*gcvT:* Gene that encodes GNMT.

GFPuv: Green fluorescent protein (*$gfp_{uv}$:* Green fluorescent protein encoding gene)

*GH371: E. coli* cells that lack the *upp* gene and are used for genetic selection

Gln (Q): Glutamine

Glu (E): Glutamate

Gly (G): Glycine

*glmS:* Glucosamine-6-phosphate riboswitch ribozyme

GMML: Glycerol minimal media containing leucine

GNMT: Glycine-N-Methyl transferase

GTP: guanosine triphosphate

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

*HindIII:* Restriction enzyme derived from *Haemophilus influenzae*, that cleaves the sequence 5'-A/AGCTT-3'

His (H): Histidine

Ile (I): Isoleucine

IPTG: Isopropyl β-D-1-thiogalactopyranoside

ITC: Isothermal titration calorimetry

LacZ*:* β-galactosidase enzyme (*lacZ*: β-galactosidase encoding gene)

LB: Luria Bertani

Leu (L): Leucine

LeuRS: Leucyl-tRNA synthetase

Lys (K): Lysine

Met (M): Methionine

-$Met_1$: Methionine has been cleaved from first position

*MluI:* Restriction enzyme derived from *Micrococcus luteus* that cleaves the sequence 5'-A/CGCGT-3'

MOPS: 3-(N-morpholino)propanesulfonic acid

mRNA: messenger RNA

MtLRS: *Methanobacterium thermoautotrophicum* leucyl-tRNA synthetase

$N_{A:}$ Binding stoichiometry

*NcoI:* Restriction enzyme derived from *Nocardia corallina* that cleaves the sequence 5'-C/CATGG-'3'

*NdeI:* Restriction enzyme derived from *Neisseria denitrificans* that cleaves the sequence 5'-CA/TATG-'3'

NEB: New England Biolab

*NheI:* Restriction enzyme derived from *Neisseria mucosa heidelbergensis* that cleaves the sequence 5'- G/CTAGC-3'

Ni-NTA: Nickel- nitrilotriacetic acid is the aminopolycarboxylic acid conjugated resin used for affinity chromatography to purify His-tagged proteins.

NTPs: Nucleoside triphosphates

$OD_{600}$: Optical density measured at 600nm

OE-PCR: Overlapping extension PCR

ONPG: o-nitrophenyl-β-D-galactopyranoside

PBS: Phosphate buffered saline

PCR: Polymerase chain reaction

PDB: Protein data bank

pET21-Z: Bacterial vector for inducible expression of genes with bacterial translation signals.

Phe (F): Phenylalanine

PPi: Inorganic Pyrophosphate

pRCG: pRep-CatUpp-$GFP_{uv}$ plasmid

pRep-CatUpp: pRep plasmid containing *cat-upp* fusion gene

Pro (P): Proline

proK: Proteinase K

pTrp: Tryptophan promoter

RCSB: Research Collaboratory for Structural Bioinformatics.

RF domain: Rossmann-Fold domain

RNA: Ribonucleic acid

SAM: S-adenosyl methionine

SarRS: Sarcosine riboswitch

SD: Shine Dalgarno sequence (Also called ribosomal binding site or RBS)

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide electrophoresis

SELEX: Systematic Evolution of Ligands by Exponential Enrichment

Ser (S): Serine

SOD: Superoxide dismutase

*SpeI:* Restriction enzyme derived from *Sphaerotilus species* that cleaves the sequence 5′-A/CTAGT-3′

SphI: Restriction enzyme derived from *Streptomyces phaeochromogenes* that cleaves the sequence 5′- GCATG/C-3′

SPR: Surface plasmon resonance

TAA: DNA sequence for the ochre stop codon

TAG: DNA sequence for the amber stop codon

*TetA*: Gene that encodes tetracycline resistant protein.

TGA: DNA sequence for the opal stop codon

TheoRS: Theophylline riboswitch

Thr (T): Threonine

TLC: Thin layer chromatography

tRNA: Transport RNA

Trp (W): Tryptophan

Tyr (Y): Tyrosine

TΨC: Thymidine, pseudouridine, cytidine.

U: Uracil

UAA: Unnatural amino acid

UMP: Uracil monophosphate

UPRT: Uracil phosphoribosyl transferase (*upp:* uracil phosphoribosyl transferase encoding gene)

Val (V): Valine

WT: Wild Type

Xgal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

XL1-Blue: *E. coli* cells that allow blue-white color screening

ΔG: Free energy or Gibb's energy change

ΔH: Enthalpy change

ΔS: Entropy change

# CHAPTER 1

# AMINOACYL SYNTHETASES AND THEIR ROLE IN PROTEIN SYNTHESIS

## 1.1 **Protein translation**

The process that involves the synthesis of proteins from their messenger RNA (mRNA) is called translation. It occurs at the ribosome in three specific steps that are shown on Figure 1.1: (1) Initiation: aminoacyl t-RNA synthetases (aaRS) load methionine (Met) to the acceptor stem of the tRNA. The anticodon loop of the tRNA carrying methionine (Met) binds to the small ribosomal subunit forming a pre-initiation complex that recognizes the mRNA at the 5'GTP cap and then tRNA scans for the mRNA start codon and binds to it through its anti-terminator loop. (2) The elongation is a GTP required process: elongation factors are recruited and the aminoacylated tRNA correspondent to the successive codon enters to the A site of the ribosome promoting the release of the elongation factor along with GDP. The enzyme peptidyl transferase catalyzes the formation of the peptide bond between the amino acids, and the chain is transferred to the carboxylic end of the P site (translocation). (3) When the stop codon is reached, termination occurs. A release factor protein binds to the A site facilitating the hydrolysis of the synthesized protein.

The amino acids are assigned to the protein synthesis based on a genetic code. Although there are some variations, the almost universal genetic code relates amino acids with a single or several trinucleotide sequences. Table 1.1 shows the standard genetic code for the natural twenty amino acids.

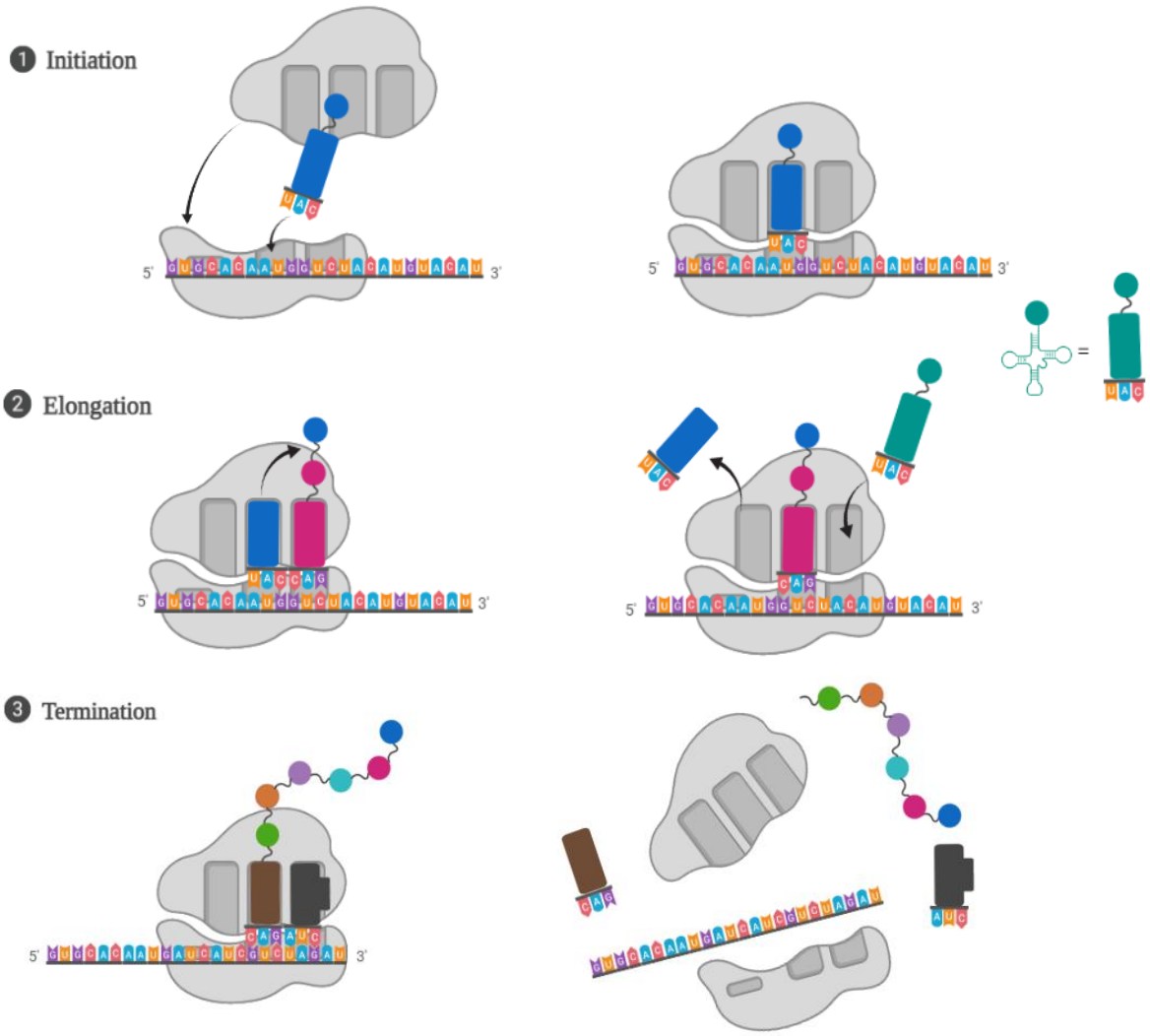
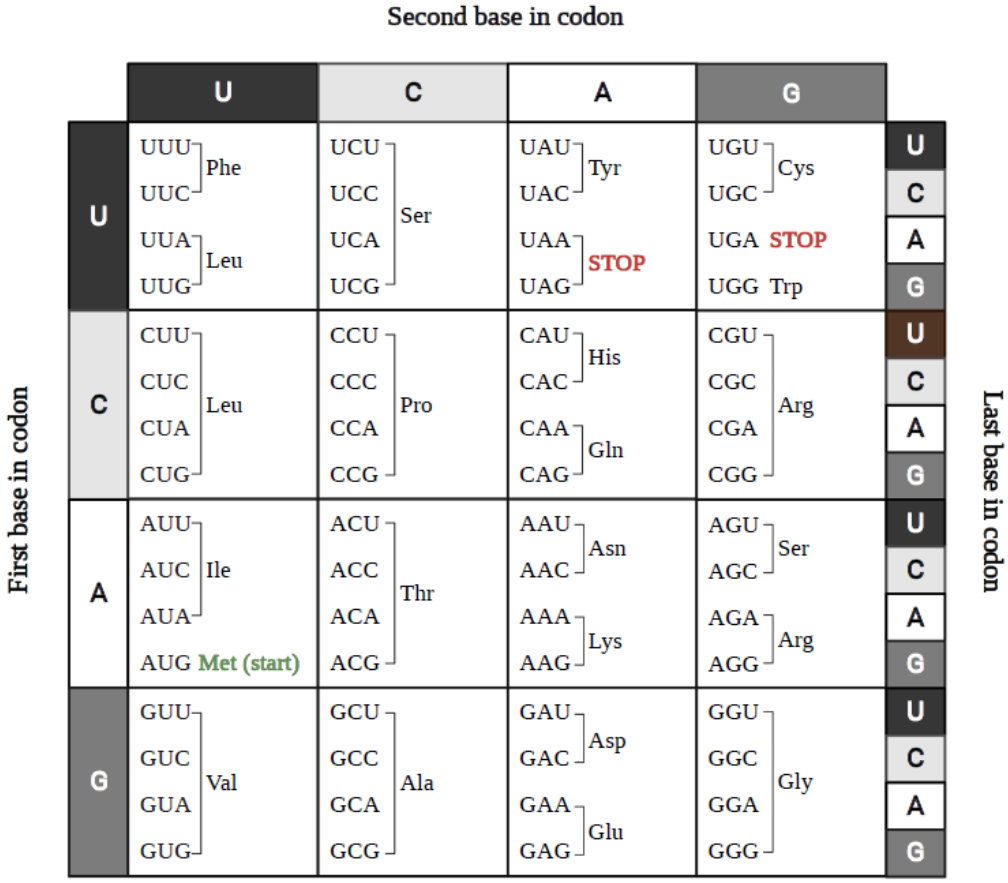



**Figure 1.1** Representation of the translation process at the ribosome and adapted from "Protein Translation". Retrieved from https://app.biorender.com/biorendertemplates (2022).

**Table 1.1** Standard genetic code table. Adapted from "codon chart". Retrieved from https://app.biorender.com/biorendertemplates (2022).




In bacteria, once the ribosome is assembled, the sequence is read starting with the start codon (AUG), then the synthesis of the protein continues until the stop codon is reached. At this point, release factor proteins facilitate the release of the newly synthesized protein.

### 1.2 tRNA: Structure and recognition

Transfer RNAs (tRNAs) are responsible for the delivery of the correspondent amino acids to the ribosome during translation.[2] Figure 1.2 depicts the general structure of tRNA which is mainly composed of an acceptor stem, dihydrouridine arm (D-arm), anticodon loop, variable loop and TΨC-arm (Thymidine, pseudouridine, cytidine).

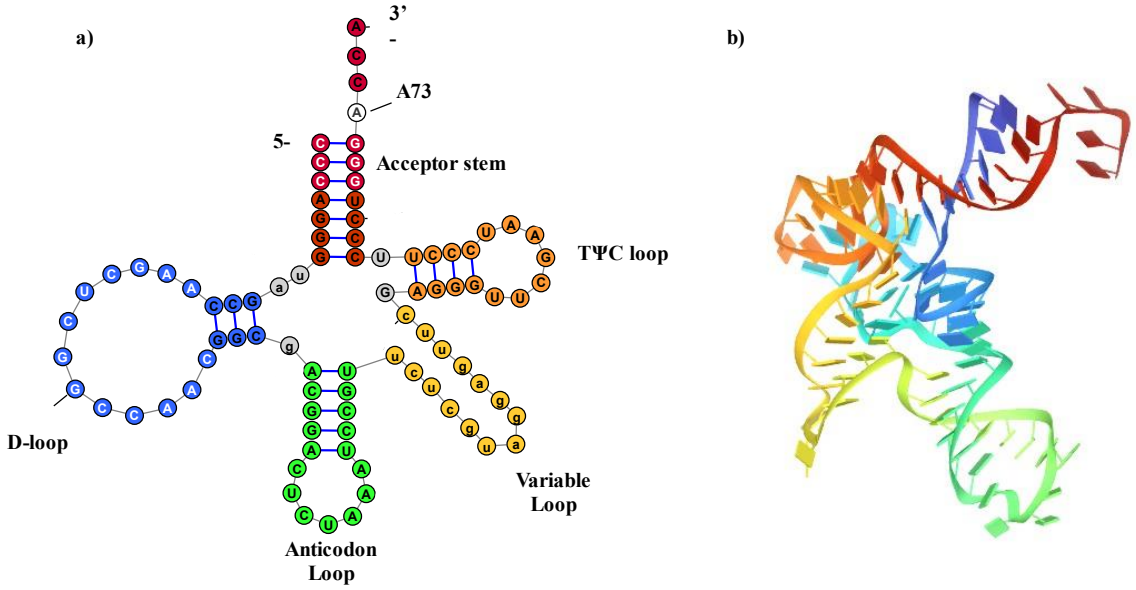



**Figure 1.2** Representation of tRNA: a)cloverleaf diagram of leucyl tRNA from *Halobacterium sp.* NRC-1 and b)three-dimensional structure of leucyl tRNA from *Thermus thermophilus* (PDB: 1WZ2). Adapted from Anderson and Schultz.[3]

tRNA folds in a L-shape in which the D and T loops interact through base pairing creating two sections: the top of the L that interacts with the amino acid, and the bottom which is primarily formed by the anticodon loop and interacts with the ribosome codon during translation. tRNA is the substrate for aminoacyl-tRNA synthetases, the initial binding occurs due to electrostatic interactions with the different motifs of the protein. As the tRNA establishes more contacts with the catalytic site, specific portions of the tRNA called identity elements, play a key role in the aminoacylation process by either inducing the binding of tRNA or promoting the release of the non-cognate tRNA. Identity elements are usually located at the acceptor arm, like the A73 nucleobase at the tRNA, anticodon stem (bases 35-37), or on the variable long arm like in the case of the leucyl-tRNA synthetase which is encoded by more than one codon.[4, 5] This last element is important because some amino acids are encoded by up to six codons which makes

the anticodon loop unsuitable as the only recognition element.[4, 6] We will also explore in the next section how some aaRS have specific editing domains that increase their proofreading ability.

### 1.3 Aminoacyl tRNA-synthetases (aaRS)

The aminoacyl t-RNA synthetases are very specific enzymes that load the amino acids to the corresponding tRNA, which is further delivered to the ribosome by elongation factors.[4, 7] There is a specific aaRS for every amino acid, and they are classified in mainly two classes based on their structure. Both classes are further subdivided in five subclasses based on their ability to recognize structurally similar amino acids as shown in Table 1.2.

**Table 1.2** Structural classification of aminoacyl tRNA synthetases. Adapted from Vijayakumar et al. [8]

| Class | Subclass | aaRS (form) | Organism used to characterize |
|---|---|---|---|
| **I** | a | Isoleucyl tRNA synthetase | *T. thermophilus* |
| | | Methionyl tRNA synthetase | *T. thermophilus* |
| | | Leucyl tRNA synthetase | *T. thermophilus* |
| | | Valyl tRNA synthetase | *T. thermophilus* |
| | b | Cysteinyl tRNA synthetase | *E. coli* |
| | | Glutaminyl tRNA synthetase | *S. cerevisiae* |
| | | Glutamyl tRNA synthetase | *T. thermophilus* |
| | c | Tyrosyl tRNA synthetase | *T. thermophilus* |
| | | Tryptophanyl tRNA synthetase | *H. Sapiens* |
| | d | Arginyl tRNA synthetase | *T. thermophilus* |
| | e | Lysil tRNA synthetase | *P. horikoshii* |
| **II** | a | Glycyl tRNA synthetase | *T. thermophilus* |

| | | |
|---|---|---|
| | Histidyl tRNA synthetase | *E. coli* |
| | Prolyl tRNA synthetase | *P. falciparum* |
| | Serinyl tRNA synthetase | *M. barkeri* |
| | Threonyl tRNA synthetase | *E. coli* |
| b | Asparaginyl tRNA synthetase | *B. malayi* |
| | Aspartyl tRNA synthetase | *S. cerevisiae* |
| | Lysyl tRNA synthetase | *H. sapiens* |
| c | Alanyl tRNA synthetase | *A. fulgidus* |
| | Glycyl tRNA synthetase | *H. sapiens* |
| | Phenylalanyl tRNA synthetase | *T. thermophilus* |
| | Pyrolysil tRNA synthetase | *M. mazei* |

Class I has a catalytic domain or Rossmann-Fold (RF) domain which is composed of alternated α-helices and five-stranded parallel β-sheets that are responsible for ATP binding in the extended conformation. These are usually located at the N-terminal region of the protein. The active site has two motifs denoted as HIGH (His-Ile-Gly-His), and KMSKS (Lys-Met-Ser-Lys-Ser) connected through a peptide called CP1 that act as proofreader.[9] Class I is usually found as monomers and incorporates bulkier amino acids than class II because the active site possesses an open pocket. The class I enzymes interact with the tRNA molecule through the minor groove, and usually aminoacylates the 2'-OH of the adenosyl nucleotides. [8]

The active site of class II synthetases possesses seven stranded β-sheets flanked by α-helices and three motifs that are less conserved than in class I. In contrast to class I, aminoacylation takes place at the 3'-OH of the ribose, and it approaches tRNA through the major groove. The binding

pocket is more compact than in class I and it involves a bent conformation. This section of the dissertation is focused on class I aminoacyl tRNA-synthetases.

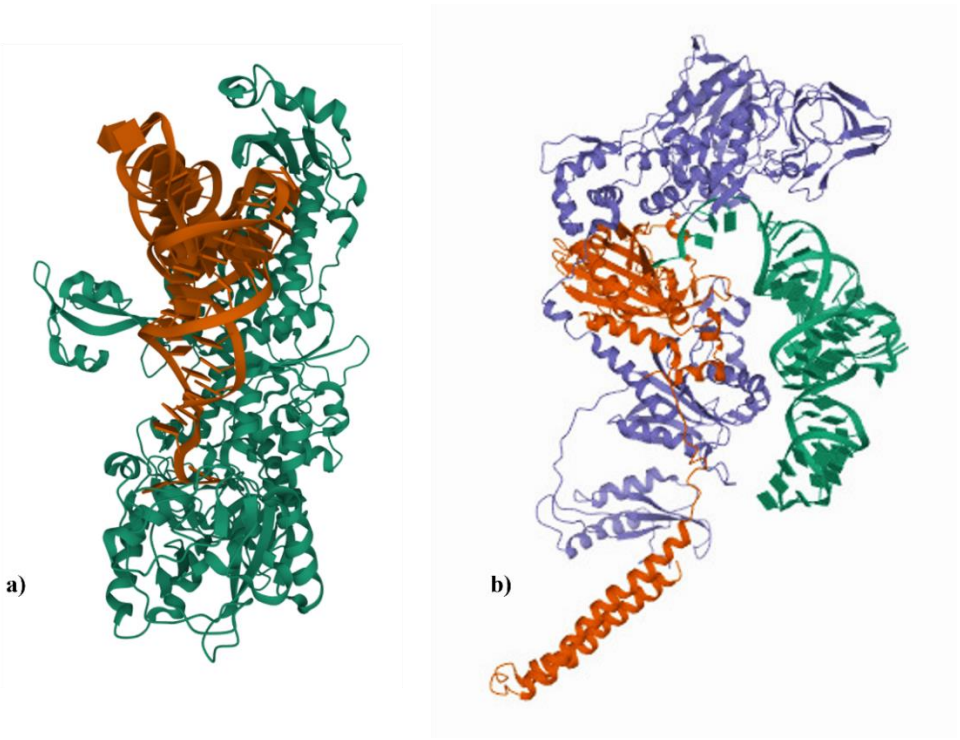
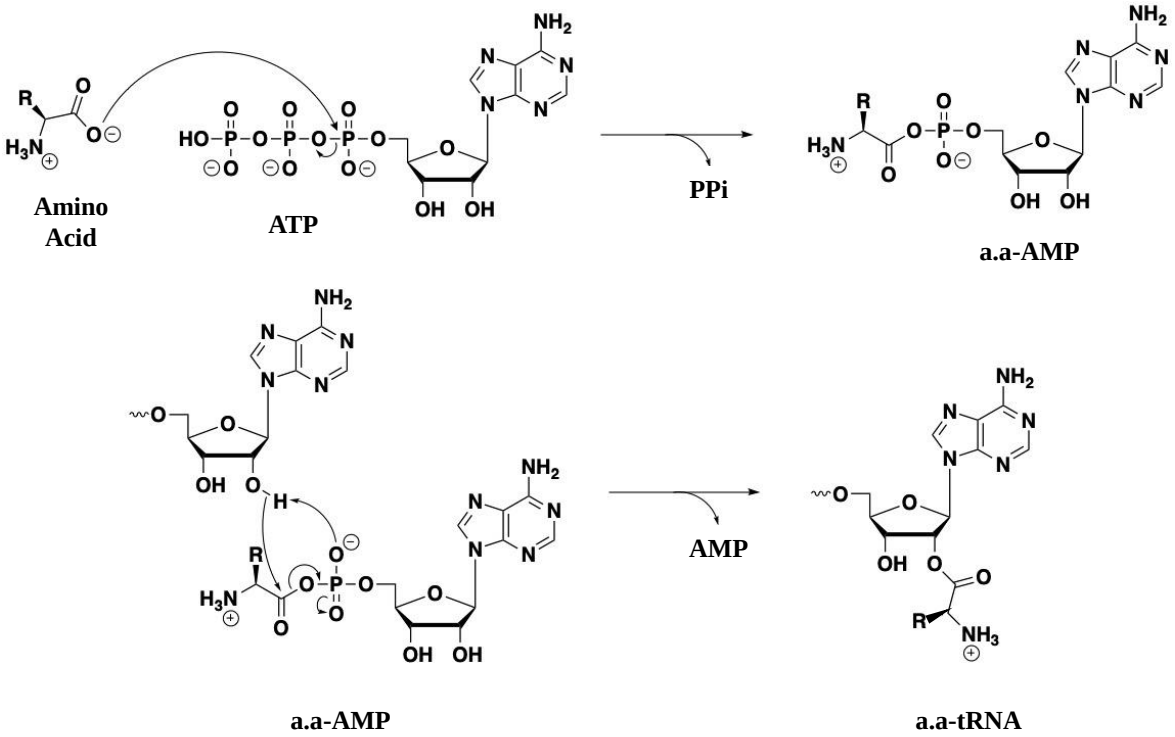



**Figure 1.3** Structure of aminoacyl t-RNA synthetases from *Thermus thermophilus*: a)class I leucyl-tRNA synthetase complexed with leucine-tRNA (RCSB PDB: 2BTE) b)class II phenylalaninyl-tRNA synthetase complexed with phenylalanyl-tRNA (RCSB PDB: 3HFZ).[10]

### 1.4 Aminoacylation of tRNA and proofreading

The general aminoacylation mechanism is shown in Scheme 1.1 and consists of an esterification reaction of the amino acid to tRNA in the presence of adenosine triphosphate (ATP). ATP and the amino acid bind in a concerted way to the catalytic site of the enzyme, followed by a nucleophilic attack of the C-terminus of the amino acid to the phosphate group of the ATP molecule to generate the condensation product (AMP-aa) as well as releasing inorganic pyrophosphate (PPi). In a second step, the hydroxyl group of the adenine 76 from tRNA attacks the electron deficient carbon at the carboxyl group of the AMP-aa to yield the aminoacylation

product (tRNA-aa) and release of AMP. The release of the aminoacyl-tRNA is considered the rate limiting step of aminoacyl synthetases class I.[4]




**Scheme 1.1** General mechanism of aminoacylation by t-RNA synthetases Adapted from Rubio-Gomez et al.[4]

Some tRNA synthetases like IleRS, ValRS, LeuRS possess hydrolytic activity to cleave the non-cognate amino acid from its tRNA through its editing domain or CP1.[4] Leucine is encoded by six different codons, therefore its synthetases (LeuRS) recognize up to six tRNA isoacceptors. Because of the number of codons, the recognition depends on other identity elements: bacterial LeuRS recognizes the D-arm of tRNA as well as the acceptor stem, while the archaeal and some eukaryotic LeuRS recognize the long variable arm and the discriminator A76 at the acceptor stem.[11, 12] Figure 1.4 highlights the recognition sites of bacterial and archaeal leucyl-tRNA synthetases.

In *E. coli*, the flexible domains of the LeuRS are ordered due to a very special type of interactions between the protein and the tRNA molecule. This flexibility gives room to the 3' end of the tRNA to bend into the active site of the synthetase for aminoacylation (Figure 1.4). When a wrong amino acid is loaded, the A76 base at the acceptor stem is no longer recognized and some of these interactions are lost, avoiding the correct arrangement of the tRNA towards the active site as well as the blockage of the catalytic site.[9]

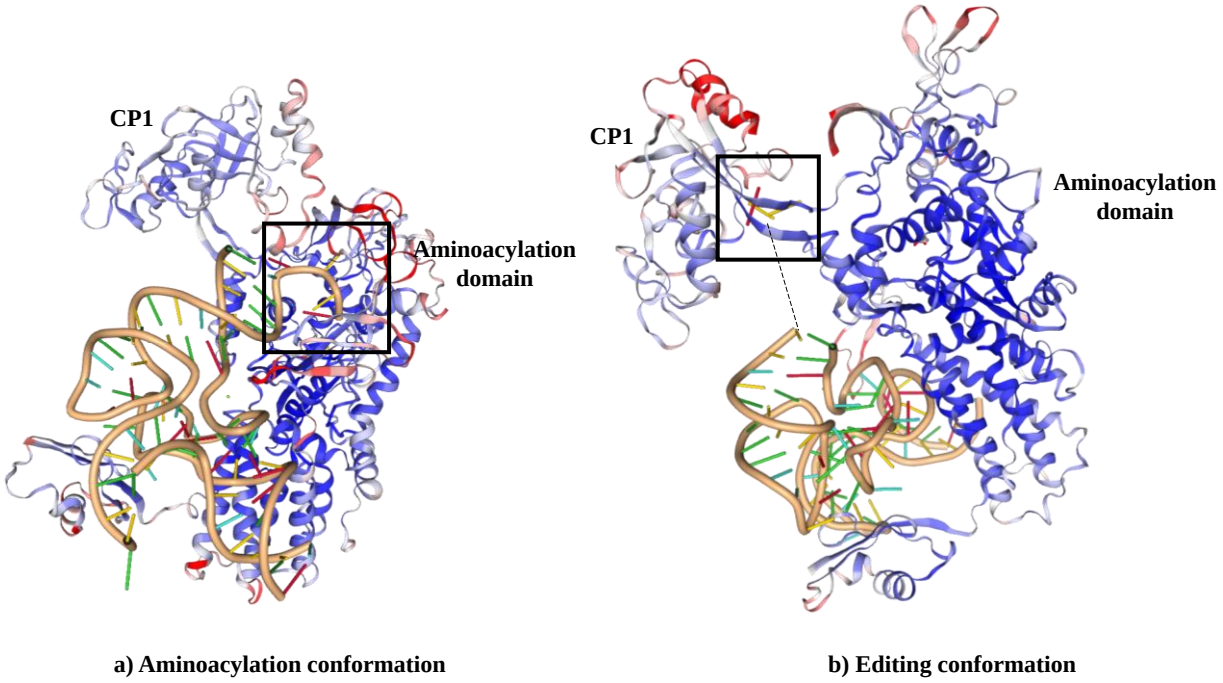



**Figure 1.4** Crystal structure of the *E. coli* leucyl-tRNA synthetases complexed with leucyl-tRNA in: a)aminoacylation and b)editing conformation (PDB: 4AQ7 and 4ARC). Structures retrieved with SWISS MODEL repository.

For some archaea like *Pyrococcus horikoshii*, the C-terminal domain of the tRNA synthetase recognizes the bases at the tip of the long variable arm of the corresponding tRNA, causing a conformational change that induces the movement of the editing domain to open the

aminoacylation binding site. Consequently, the tRNA acceptor stem bends so it can be aminoacylated by the synthetase.[11]

### 1.5 Expanding the genetic code

During the translation process, there are three of the 64 codons that do not code for amino acids but instead they activate release factors that terminate the protein synthesis. These three stop codons are ochre (TAA), opal (TGA) and amber (TAG). The amber stop codon is used only in about 7% of terminations in *E. coli*, and it is not common to terminate essential genes.[13] In addition, the expansion of the genetic code via TAG codon in *E. coli* has been proved to not affect cell growth.[14] To incorporate unnatural amino acids (UAAs) into proteins, several conditions must be met: the tRNA synthetase cannot incorporate endogenous amino acids into endogenous tRNAs, and the tRNA selected to incorporate these amino acids has to be specific for the UAAs. This is usually called an orthogonal pair.[3] Incorporating unnatural amino acids introduces novel functions to proteins with modifications in any part of the protein chain to add functionality like unsaturation that allows click reactions, fluorescence for tracking or sensing, photo crosslinking, protein dynamics studies, among others.[15-19] Tyrosyl-tRNA synthetase from *M. janaschii* and leucyl-tRNA synthetase from *E. coli* have been widely used in *E. coli and* in *S. cerevisiae* using amber stop codon respectively. [15, 20, 21]

These systems were engineered by mutating amino acids in the amino acid binding sites of the synthetases, and the identity elements at the tRNA, and subsequent selection steps in the presence of the unnatural amino acid. [3, 13, 22, 23]

## 1.6 Conclusions

Aminoacyl t-RNA synthetases play a fundamental role in the fidelity of translation and can be used as a tool to incorporate unnatural amino acids into proteins. Although the systems already mentioned have been validated for the incorporation of some UAAs (more than 40 different kinds), the need to gain new functionality into the protein structure justifies the exploration of new systems that can be used as orthogonal pairs to introduce novel unnatural amino acids.[24, 25]

In the next chapter, we will discuss a system composed of leucyl-tRNA synthetase from *Methanobacterium thermoautotrophicum* and tRNA from *Halobacterium sp. NRC-1*.

# CHAPTER 2

# DIRECTED EVOLUTION OF AN ARCHAEAL LEUCYL T-RNA SYNTHETASE TO INCORPORATE UNNATURAL AMINO ACIDS

## 2.1 Structural homology between archaeal and bacterial leucyl tRNA synthetases

Although directed evolution using stop codon approach has been widely used, most of the published studies are based on the incorporation of UAAs using amber stop codon.[26] In 2003, Anderson and Schultz reported a leucyl-tRNA synthetase from *Methanobacterium thermoautotrophicum (MtLRS)* and tRNA from *Halobacterium sp. NRC-1(HhL4)* as an orthogonal pair for amber, opal and four base stop codon suppression in *E. coli*.[3] Several modifications were made to the different recognition sites of the transfer RNA molecule to generate a library, and the selected clone HL(TAG)3 showed about 30% suppression of the amber stop codon. This orthogonal system represents an opportunity to incorporate multiple unnatural amino acids at different positions using a combination of amber, opal and four base stop codons at a time. However, it has not been used to incorporate UAAs into proteins in bacteria to this date. In addition, the structure of MtLRS is not known, but its active site residues were identified by a sequence homology between species with the same domain and structure-based alignments of known LeuRS from *T. thermophilus* and *P. horikoshii.* as shown in Table 2.1 and Figure 2.1.

Previously, our group used positive selection to generate a functional MtLRS whose editing domain was replaced by a six codon linker (L-Y-H-A-V-Y).[27] In this dissertation, we explored the use of this synthetase lacking the editing domain and the HL(TAG)3 tRNA variant from *Halobacterium sp.* NRC-1, to incorporate 2-amino-3-(5-(dimethylamino)naphthalene-1-

sulfonamide)propanoic acid (Dansyl-Dap) into a model protein called Z-domain. A fluorescent amino acid allows the visualization of proteins *in vitro* and *in vivo*, which should be useful for diagnostics and protein dynamics studies among another applications. Dansyl-Dap has been incorporated in yeast, but to this date no reports of incorporations in *E. coli* have been found.[15]

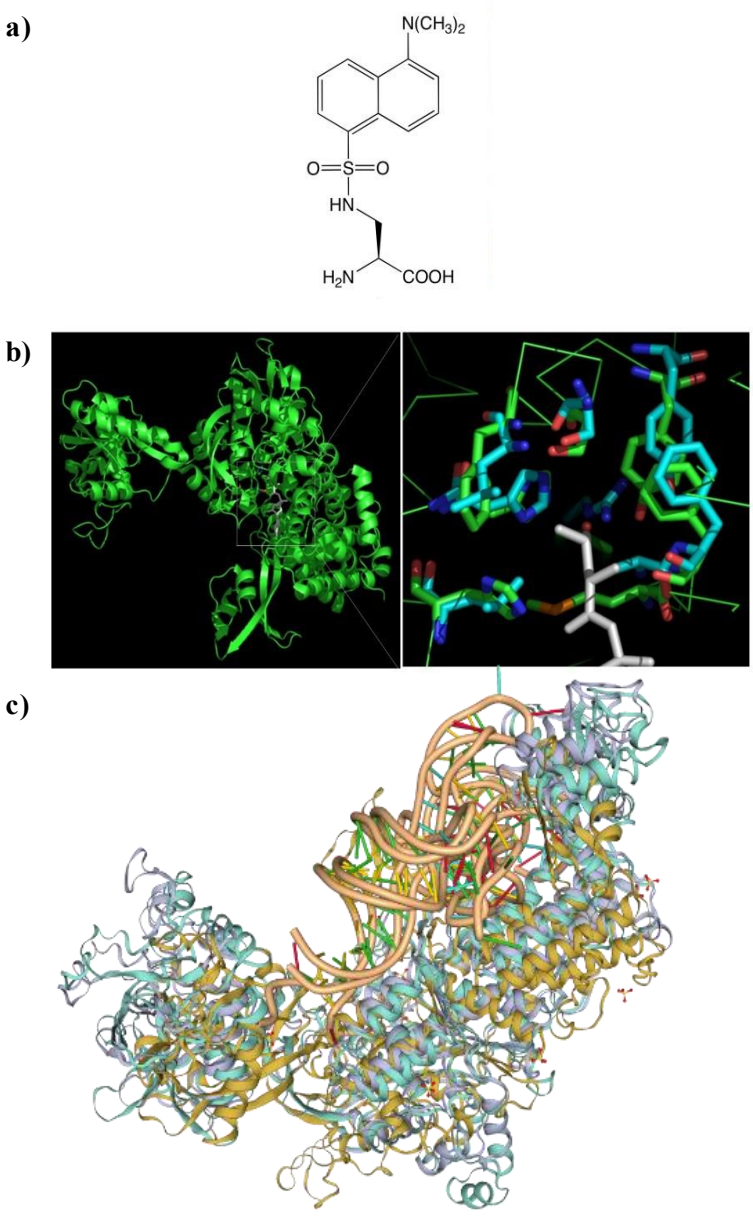



**Figure 2.1** a)chemical structure of Dansyl-Dap, b)structure homology analysis of leucine-tRNA synthetases from *Thermus thermophilus* (green) and *Pyrococcus horikoshii* (cyan) with focus to the amino acid binding site. c)superposed image of MtLRS model (light purple) built from *T.*

*thermophilus* (PDB:2BYT, light brown) and *P. horikoshii* (PDB:1WZ2, light blue) in complex with leucyl-tRNA.

**Table 2.1** Sequence homology analysis of leucyl-tRNA synthetases from *E. coli*, *T. thermophilus*, *P. horikoshii* and *M. thermoautotrophicum.* Analysis performed with clustal omega (see appendix A.1).

| **Bacteria** | | **Archea** | |
|---|---|---|---|
| ***Escherichia coli*** | ***Thermus thermophilus*** | ***Pyrococcus horikoshii*** | ***Methanobacterium thermoautotrophicum*** |
| **M40** | **M40** | **A41** | **A32** |
| **L41** | **F41** | **F42** | **Y33** |
| F493 | F501 | L529 | L510 |
| S496 | S504 | S532 | S513 |
| **Y499** | **Y507** | **Y535** | **Y516** |
| **Y527** | **Y535** | **R607** | **R566** |
| H533 (secondary for binding) | H541 (secondary for binding) | L613 (secondary for binding) | L572 (secondary for binding) |
| **H537** | **H545** | **H617** | **H576** |
| T252 | T252 | T234 | T219 |
| V338 | V340 | V325 | V306 |
| D342 | D344 | D332 | D313 |

## 2.2 Evolving an archaeal leucyl t-RNA synthetase to incorporate Dansyl-Dap in *E. coli*

In 2006, Summerer et al. used *E. coli* leucyl-tRNA synthetase to incorporate Dansyl-Dap into super oxide dismutase (SOD) to study protein folding in yeast. They generated a library in which

five amino acids in the active site (the ones highlighted on table 2.1) were randomized. After a genetic selection, three clones were identified to generate the desired fluorescent protein.[15] We used the sequence homology model to try to obtain the MtLRS variants for the same amino acids as shown in Table 2.2.

**Table 2.2** Intended mutations used to incorporate Dansyl-Dap into Z-domain in *E. coli*

| **Clone** | ***E. coli* (Summerer et al)[15]** | ***M. thermoautotrophicum* (this dissertation)** |
|---|---|---|
| ANIGT | M40A, L41N, Y499I, Y527G, H537T | A32, Y33N, Y516I, R566G, H576T |
| GNIGG | M40G, L41N, Y499I, Y527G, H537G | A32G, Y33N, Y516I, R566G, H576G |
| GLVGE | M40G, L41L, Y499V, Y527G, H537E | A32G, Y33L, Y516V, R566G, H576E |

To make these variants, the genes for MtLRS lacking the editing domain and the tRNA from *Halobacterium sp. NRC-1* (HhL4) were introduced into the pSupK plasmid.[20] The resulting plasmid pSupk-MtLRS-HL(TAG)3-ΔCP1 was mutated at the leucine binding site using overlapping extension polymerase chain reaction (OE-PCR) in two steps with around 20bp overlap between fragments as shown in Figure 2.2. N-terminal mutations (A32X and Y33X, where X represents the intended mutation) were introduced between the *NdeI* and *MluI* restriction sites, while C-terminal mutations (Y516X, R566X, H576X) were introduced between the *MluI* and *DraIII* sites. Additionally, a pET21-Z plasmid encoding a C-terminal his-tagged Z-domain K7TAG mutant in which the lysine codon at the seventh position was changed to the amber stop codon was produced.

Z-domain is a small protein derived from the IgG-binding domain of the staphylococcal protein A, with a molar mass of approximately 7.7kDa when expressed in *E. coli*, that has been previously used to incorporate unnatural amino acids.[28-30] If the synthetase is functional and specific for Dansyl-Dap, it is expected that Dansyl-Dap is incorporated into the seventh position to produce a full-length Z-domain. If the synthetase is functional but nonspecific for Dansyl-Dap, it is expected to form a full-length protein(s) that incorporates endogenous amino acid(s). If the synthetase is not functional, no full-length protein is expected.

To assure that the incorporation is only at the TAG codon and the UAA does not affect growth, pET21 containing a Z-domain without the amber stop codon at the seventh position (WT) was also expressed with and without 1 mM of Dansyl-Dap. BL21(DE3) *E. coli* cells were co-transformed with pSupk-MtLRS-ΔCP1-HL(TAG)3 and pET-21Z K7TAG (or WT), and grown overnight in TB autoinduction media at 37°C. The expressed Z-domain proteins were purified using Ni-NTA affinity chromatography, and subsequently analyzed using SDS-PAGE and mass spectrometry (ESI-MS). Table 2.3 shows the expected molar masses for the Z-domain with Dansyl-Dap incorporated at the seventh position and with endogenous amino acids.

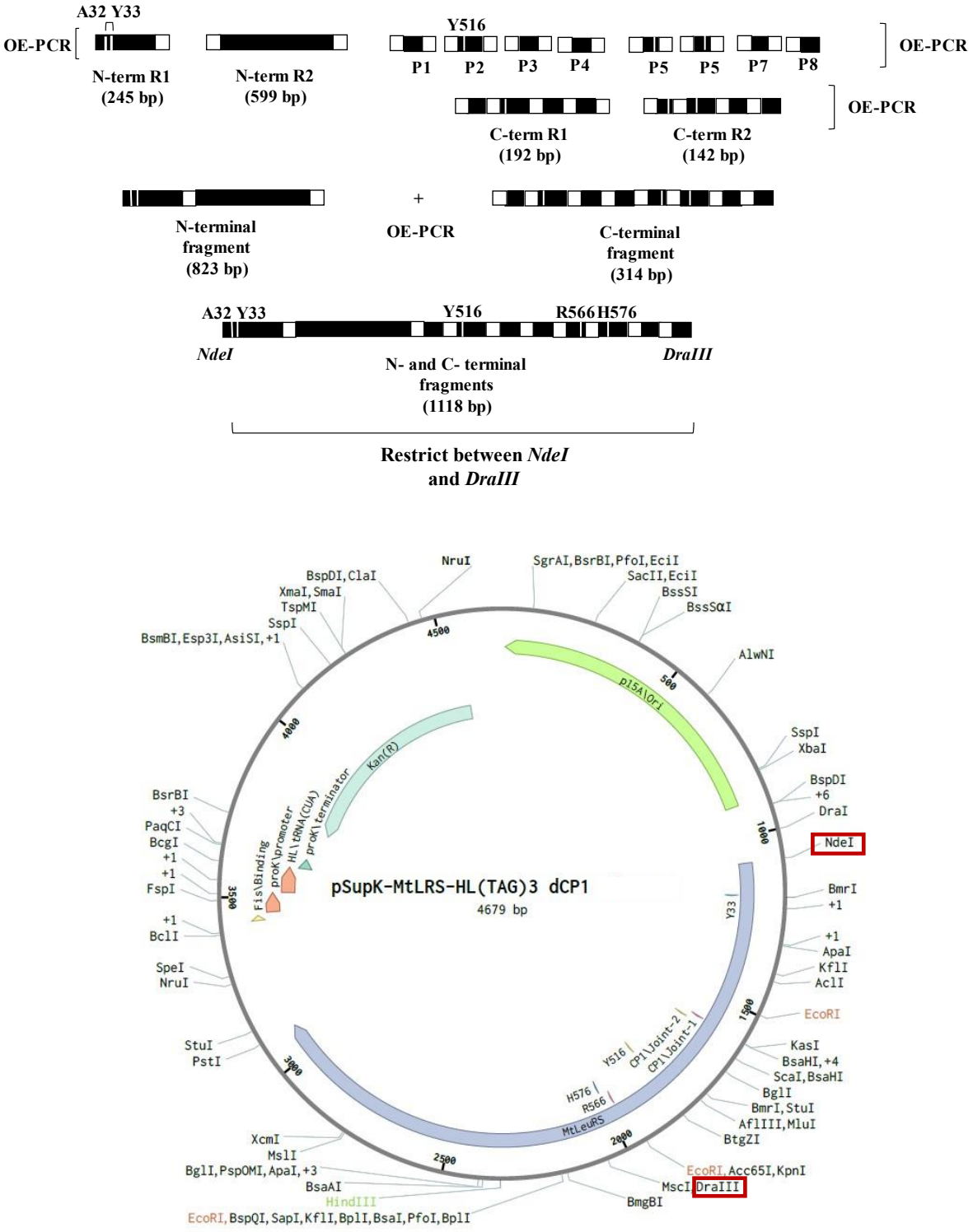



**Figure 2.2** Strategy to construct the pSupK-MtLRS-HL(TAG)3-ΔCP1 mutants using OE-PCR. Top picture shows strategy to incorporate mutations , and bottom picture shows the plasmid components and the restriction sites used highlighted in red boxes. Plasmid is drawn using Benchling.

**Table 2.3** Calculated molar mass of the possible Z-domain mutants obtained after protein expression.

| Mutant | Expected molar mass (Da) | Mutant | Expected molar mass (Da) |
|---|---|---|---|
| Z WT, Z WT(-$Met_1$) | 7893.62, 7762.43 | Z (K7W) | 7820.47 |
| Z (K7E) | 7763.37 | Z (K7F) | 7781.43 |
| **Z (K7DanDap)** | **7953.65** | Z (K7S) | 7721.33 |
| Z (K7G) | 7691.31 | Z (K7T) | 7735.36 |
| Z (K7A) | 7705.33 | Z (K7Y) | 7797.43 |
| Z (K7V) | 7733.39 | Z (K7Q) | 7762.38 |
| Z (K7C) | 7737.44 | Z (K7R) | 7790.44 |
| Z (K7P) | 7731.37 | Z (K7H) | 7771.39 |
| Z (K7L) | 7747.41 | Z (K7D) | 7749.34 |
| Z (K7I) | 7747.41 | Z (K7M) | 7765.45 |

### 2.2.1 Expression and characterization of Z-domain using MtLRS ANIGT clone

The proteins obtained were separated using sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE) and stained with Coomassie Blue Safe Stain. The band corresponding to the full-length Z-domain protein was found for all the species as shown on Figure 2.3. The expression level for the mutant protein was noticeably lower than that for the wild type. Nonetheless, enough protein was obtained for the ESI-MS analysis after purification even for the TAG mutant produced in the absence of Dansyl-Dap whose band was very faint. The UV light-exposed gel showed fluorescent bands for both, wild type and the TAG mutant produced in the presence of Dansyl-Dap.

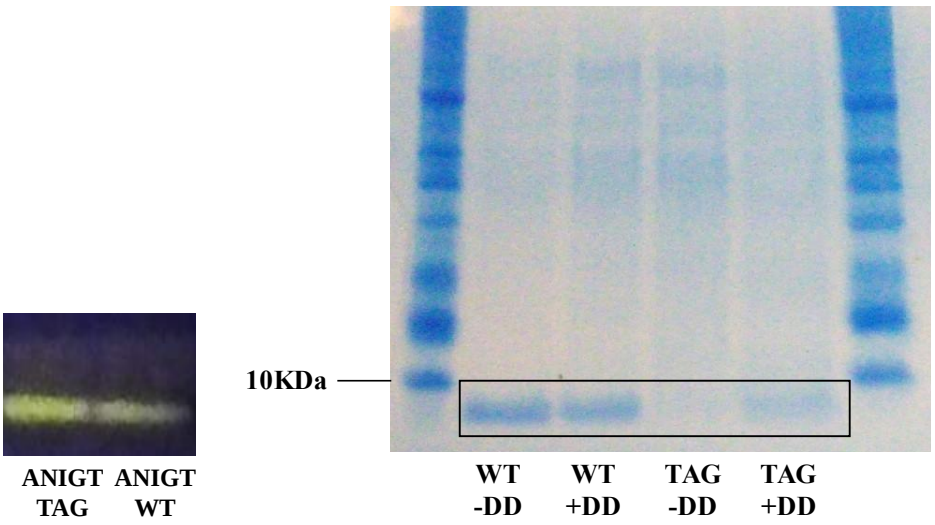



**Figure 2.3** SDS-PAGE analysis of Z-domain expression lysates using MtLRS-HL(TAG)3-ΔCP1 ANIGT clone on wild type pET21-Z and pET21-Z(K7TAG) in the presence and absence of 1 mM of 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid (Dansyl-Dap). Left side shows the UV-exposed unstained gel, and right side shows gel stained with Coomassie Blue Safe Stain.

An ESI-ion trap deconvoluted mass spectrum was generated for the obtained proteins. As shown in Figure 2.4, the wild type Z-domain gave a major peak at 7762.93 Da, which corresponds to the Z-domain without methionine in the first position ($Met_1$).

The molar mass of the minor peak seems to correspond to a small percentage of the Z-domain containing the $Met_1$ residue. Since the fluorescence was observed even for the wild type, we speculated that a residual amount of dansyl chloride from the synthesis of Dansyl-Dap reacted with the N-terminal amino group of the protein to generate a fluorescent band (expected around 7996 Da). However, no higher molecular weight peaks were found. Additionally, when adding wild-type Z-domain expressed in the absence of UAA to a 1mM solution of Dansyl-Dap as a control and analyzed with SDS-PAGE, fluorescent bands were still observed. Therefore, we conclude that the fluorescent bands were observed because of non-covalent interactions between the protein and Dansyl-Dap.

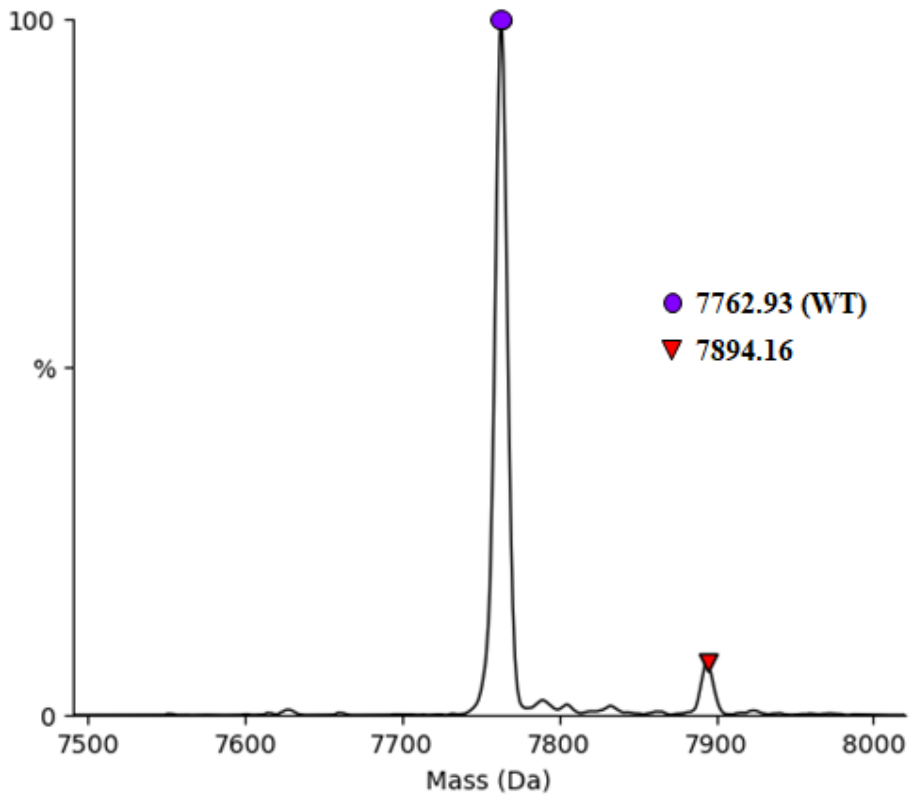



**Figure 2.4** Deconvoluted ESI-MS of Z-domain WT expressed with MtLRS ANIGT clone in the presence of 1 mM Dansyl-Dap. Analyzed with Unidec.

For the expression of Z-domain K7TAG mutant, additional peaks were observed in the mass spectra confirming that amino acids are being incorporated into the protein chain (Figure 2.5). Some peaks were not assigned because the difference between the expected molar mass and the value found is larger than 1 Da. Two major peaks corresponded to the incorporation of lysine, glutamate, and glutamine (7763.18 Da) and tyrosine (7798.27 Da). The incorporation of these results suggests that the MtLRS ANIGT mutant loads natural amino acids to the tRNA molecule.

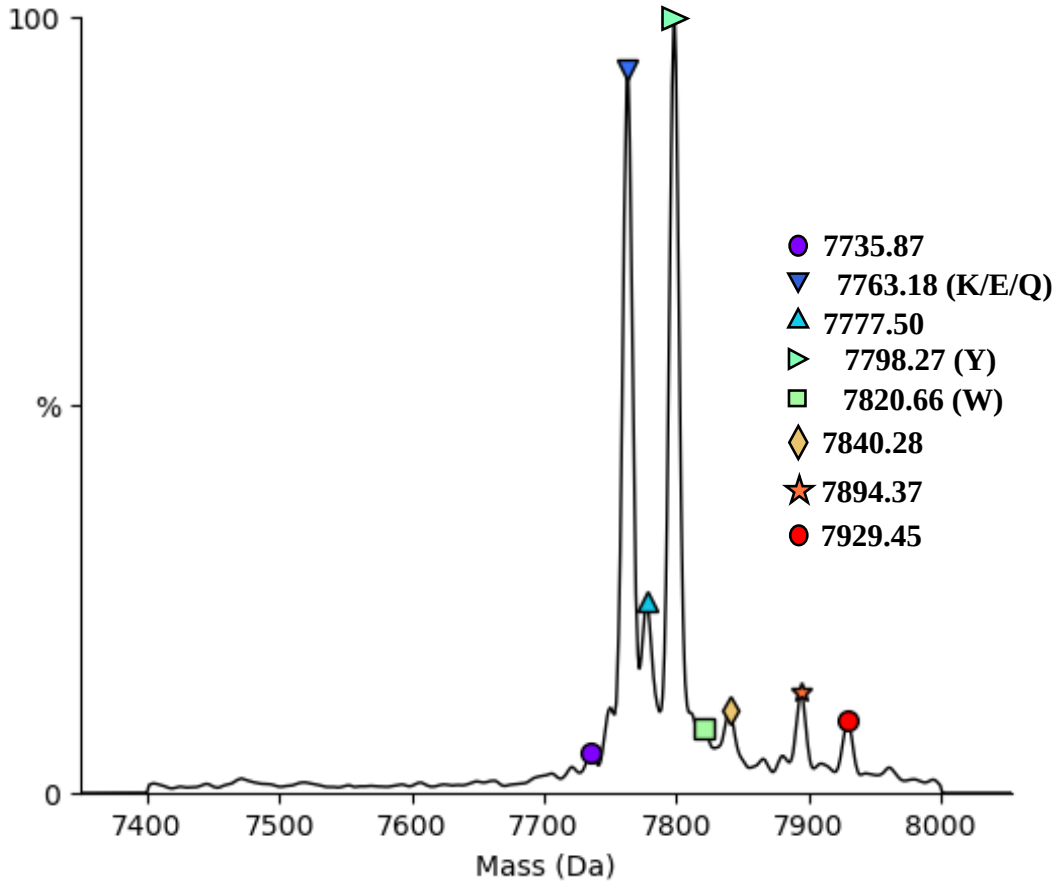



**Figure 2.5** Deconvoluted ESI-MS of Z-domain (K7TAG) expressed with MtLRS ANIGT clone in the presence of 1 mM Dansyl-Dap. Analyzed with Unidec.

As expected, the Z-domain K7TAG expressed in the absence of UAA showed a similar MS spectrum (Figure 2.6). The major peaks were observed at 7761.42Da and 7798.96 Da, which corresponds to the incorporation of lysine/glutamate/glutamine and tyrosine respectively.

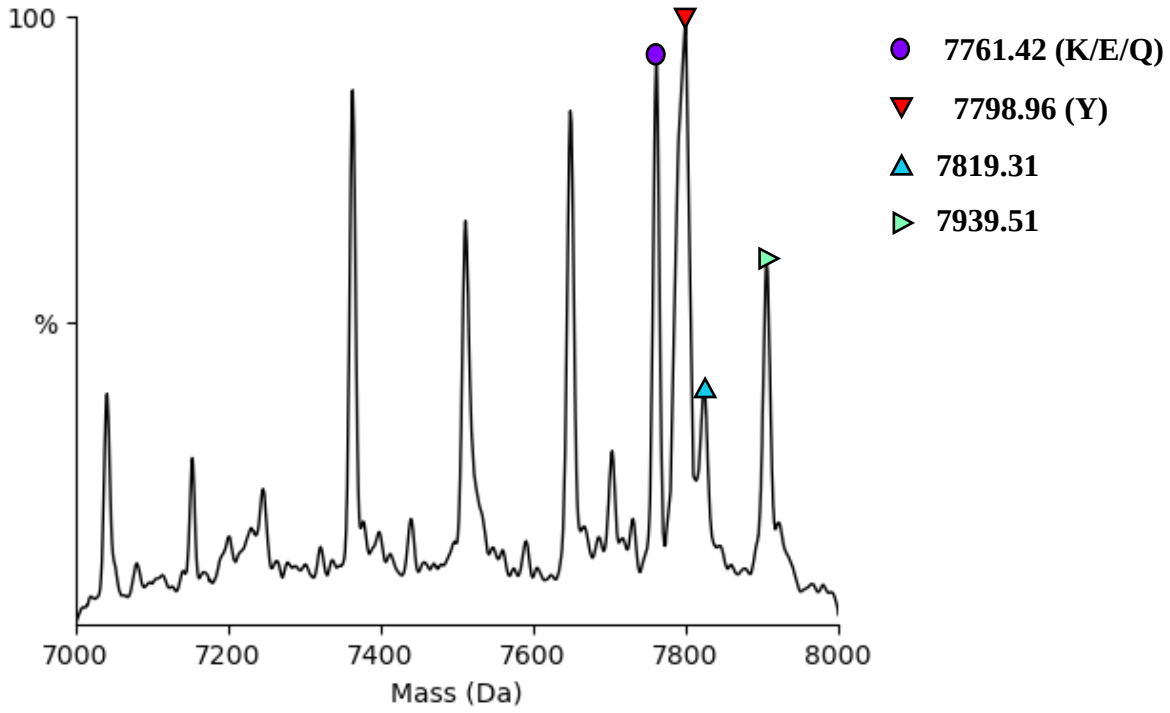



**Figure 2.6** Deconvoluted ESI-MS of Z-domain (K7TAG) expressed with MtLRS ANIGT clone in the absence of Dansyl-Dap. Analyzed with Unidec.

### 2.2.2 Expression and characterization of Z-domain using MtLRS GNIGG clone

When the Z-domain was expressed using the MtLRS GNIGG clone, SDS-PAGE showed similar expression levels of all the proteins in the presence and absence of Dansyl-Dap (Figure 2.7). ESI-MS of the purified wild type Z-domain consistently yielded a major peak at 7763.10 Da as well as the small peak of 7894.48 Da corresponding to acetylated protein. Two additional small peaks were also found as shown in Figure 2.8.

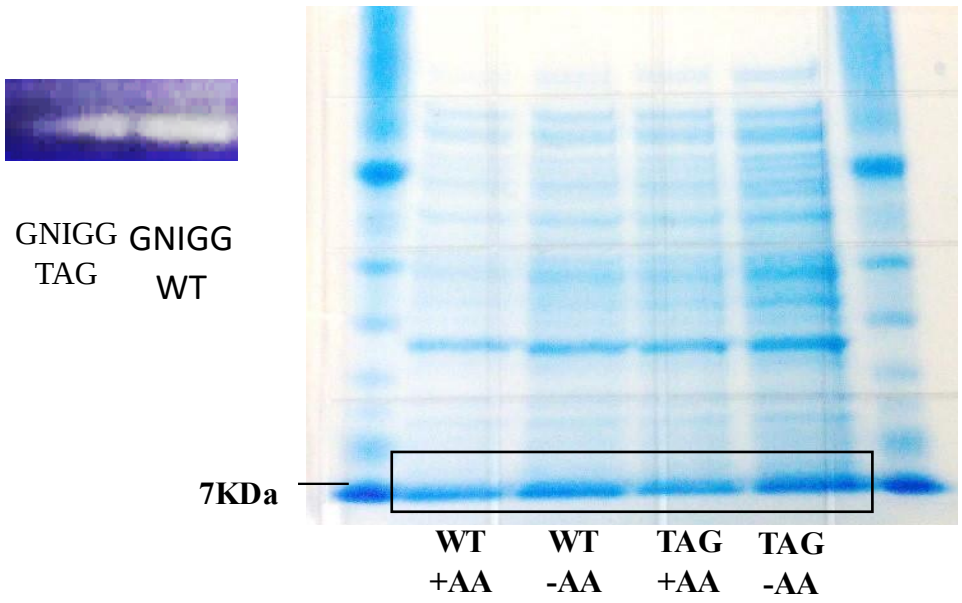



**Figure 2.7** SDS-PAGE analysis of Z-domain expression lysates using MtLRS-HL(TAG)3-ΔCP1 GNIGG clone on wild type pET21-Z and pET21-Z(K7TAG) in the presence and absence of 1 mM of 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid (Dansyl-Dap). Left side shows the UV-exposed unstained gel, and right side shows gel stained with Coomassie Blue Safe Stain.

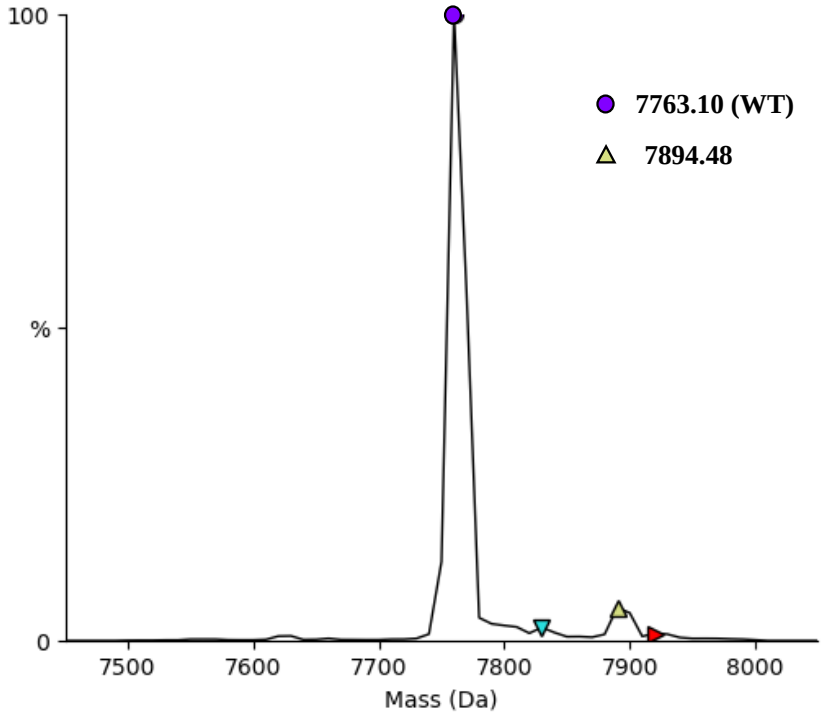



**Figure 2.8** Deconvoluted ESI-MS of wild type Z-domain expressed with MtLRS GNIGG clone in the presence of 1 mM Dansyl-Dap. Analyzed with Unidec.

When the MtLRS GNIGG was used for the expression of Z-domain K7TAG mutant in the presence of Dansyl-Dap, the selectivity towards tyrosine increased compared with ANIGT clone as shown on Figure 2.9. However, no full-length protein with Dansyl-Dap was found. In addition, when no Dansyl-Dap was added, the mass spectrum showed a major peak at 7761.46 Da, which corresponds to the Z-domain K7, Q7 or E7 mutant, and another major at 7798.14 Da, which corresponds to the Z-domain K7Y mutant. Comparable results were obtained for the Z-domain mutants expressed in the absence of Dansyl-Dap as shown in Figure 2.10, confirming the MtLRS GNIGG adds not Dansyl-Dap but some natural amino acids to the tRNA.

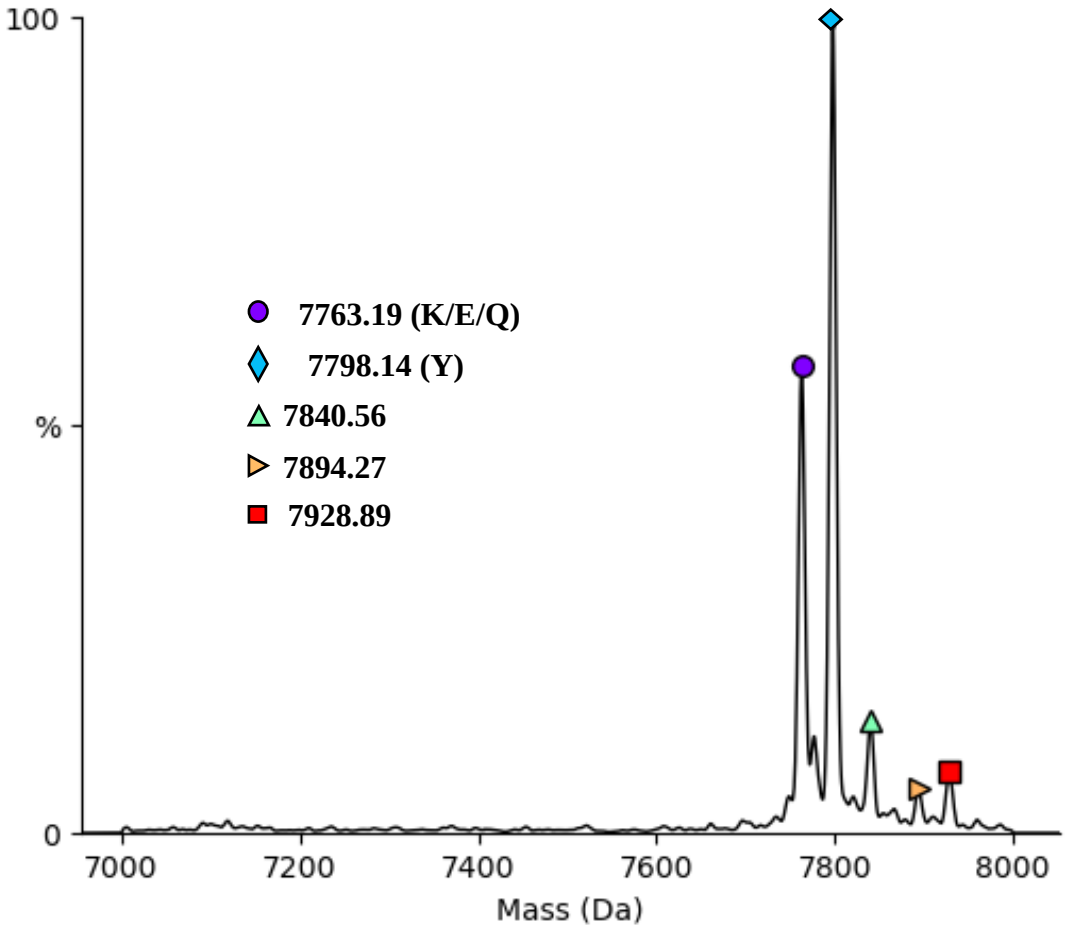



**Figure 2.9** Deconvoluted ESI-MS of Z-domain K7TAG mutant expressed with MtLRS GNIGG clone in the presence of 1 mM Dansyl-Dap. Analyzed with Unidec.

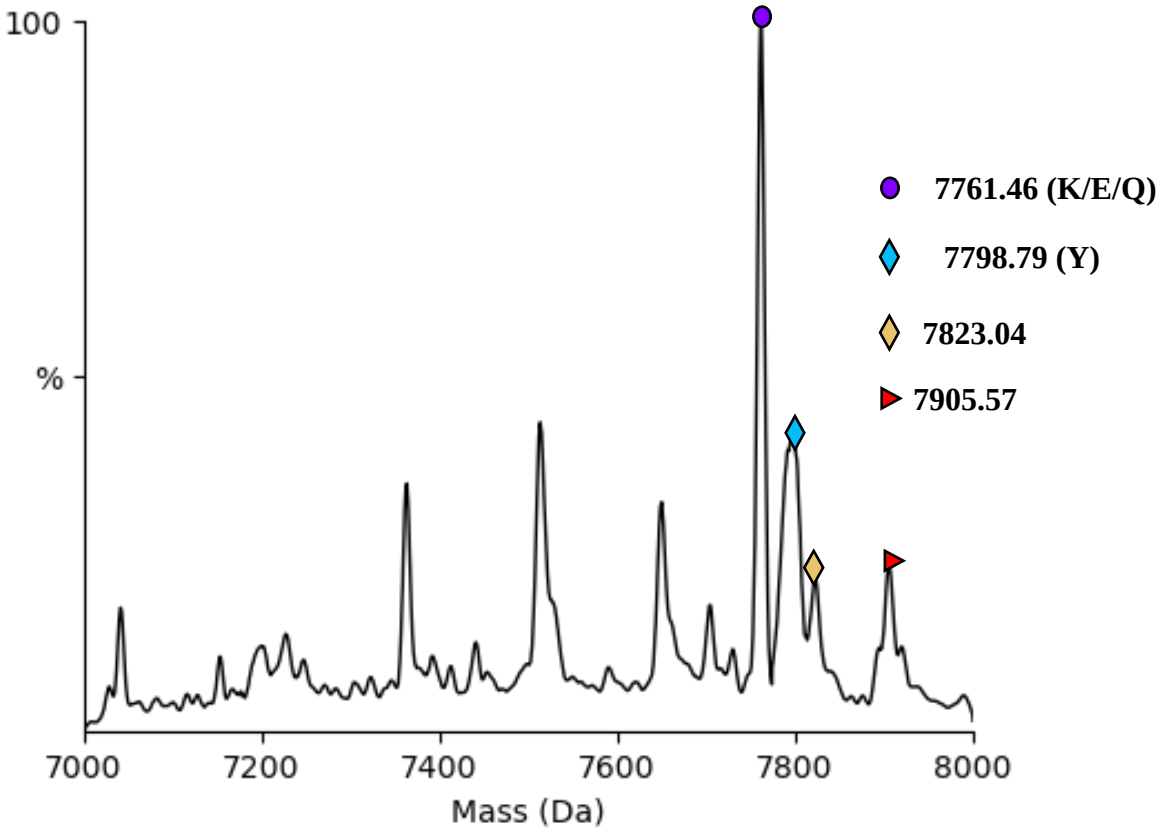



**Figure 2.10** Deconvoluted ESI-MS of Z-domain K7TAG mutant expressed with MtLRS GNIGG clone in the absence of Dansyl-Dap. Analyzed with Unidec.

### 2.2.3 Expression and characterization of Z-domain using MtLRS GLVGE clone

When the Z-domain was expressed with MtLRS GLVGE clone, a protein band was found for the Z-domain wild type and K7TAG mutants in the presence and absence of Dansyl-Dap (Figure 2.11). The ESI-MS spectra of wild type expressed in the presence of the UAA (Figure 2.12) showed that the results were consistent with those for MtLRS ANIGT and GNIGG mutants: A main peak was observed at 7763.10 Da that represents the wild type without $Met_1$.

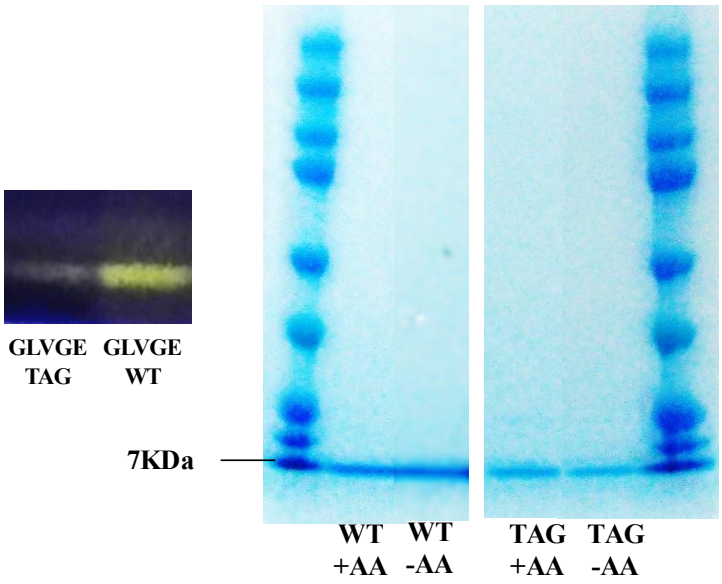



**Figure 2.11** SDS-PAGE analysis of Z-domain expression using MtLRS-HL(TAG)3-ΔCP1 GLVGE clone on wild type pET21-Z and pET21-Z(K7TAG) in the presence and the absence of 1 mM of 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid (Dansyl-Dap). Left side shows the UV-exposed unstained gel, and right side shows gel stained with Coomassie Blue Safe Stain.

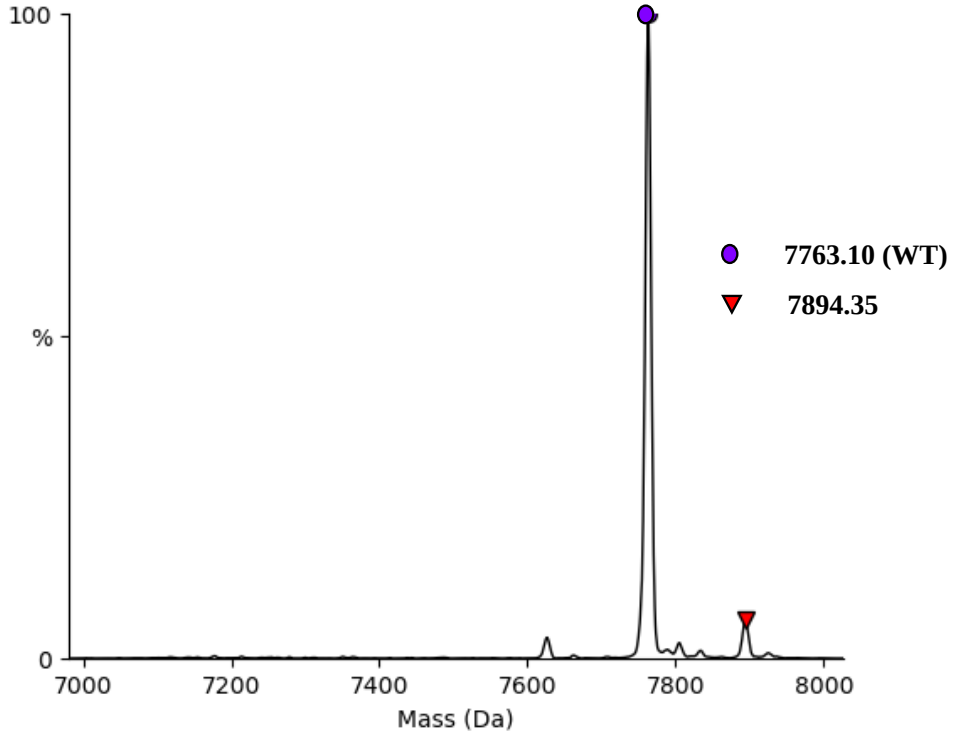



**Figure 2.12** Deconvoluted ESI-MS of wild type Z-domain expressed with MtLRS GLVGE clone in the presence of 1 mM Dansyl-Dap. Analyzed with Unidec.

For the Z-domain K7TAG mutant using MtLRS GLVGE expressed in the presence of Dansyl-Dap, The MS spectrum showed a major peak at 7797.95 Da correlated to the incorporation of tyrosine at the stop codon, and another main peak at 7763.25 Da correlated to the incorporation of lysine, glutamate, or glutamine (Figure 2.13). These peaks also appeared on the ESI-MS spectrum of the K7TAG mutant expressed without Dansyl-Dap (Figure 2.14).

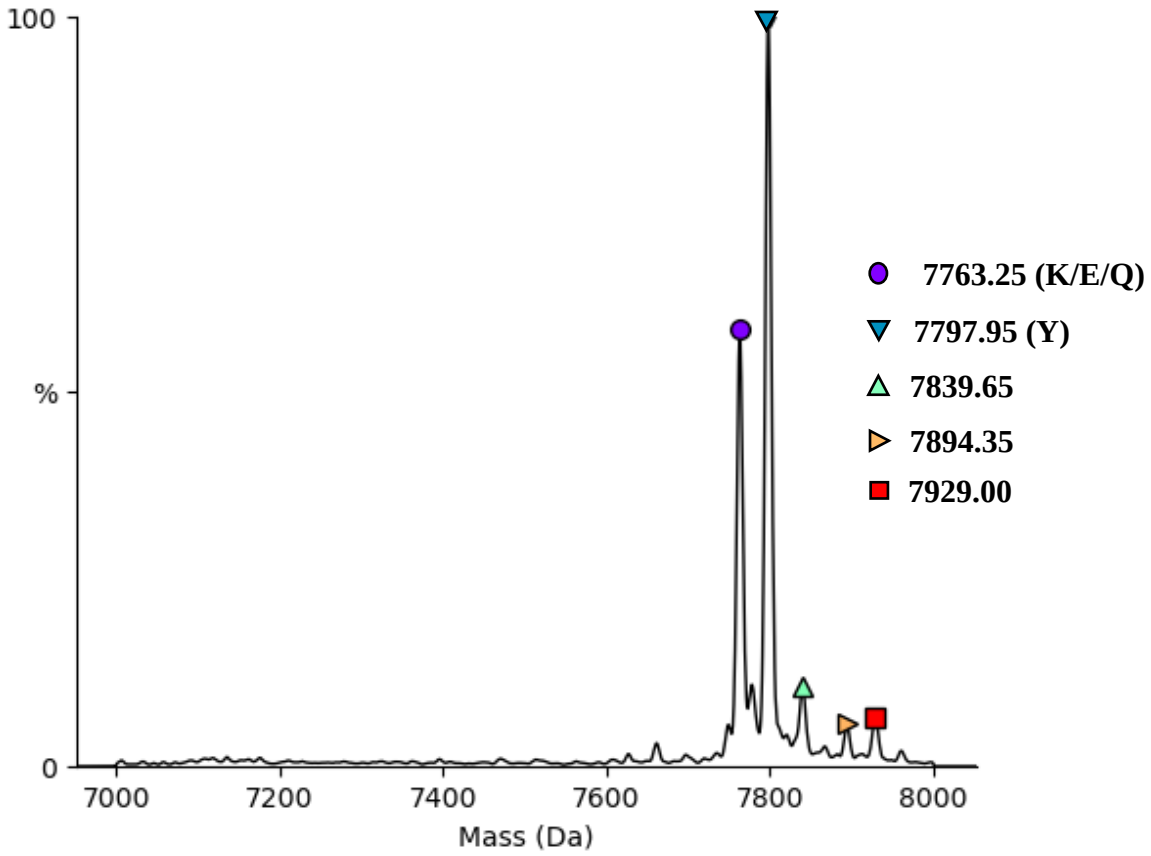



**Figure 2.13** Deconvoluted ESI-MS Z-domain K7TAG mutant expressed with MtLRS GLVGE clone in the presence of 1 mM Dansyl-Dap. Analyzed with Unidec.

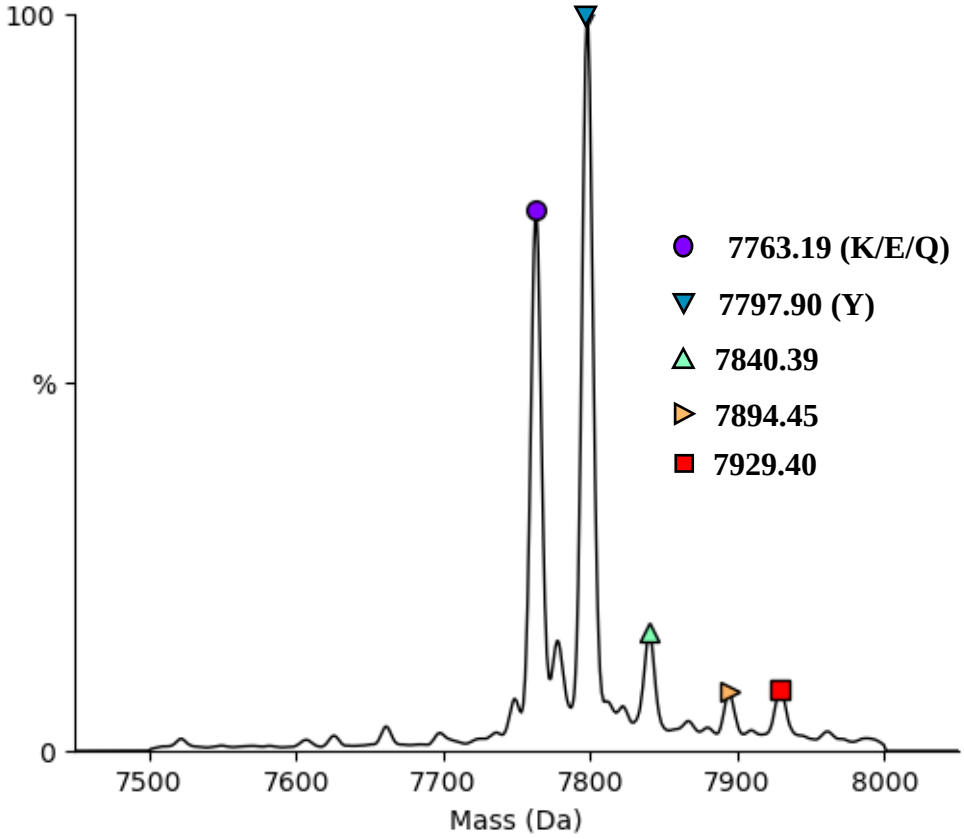



**Figure 2.14** Deconvoluted ESI-MS Z-domain K7TAG mutant expressed with MtLRS GLVGE clone in absence of Dansyl-Dap. Analyzed with Unidec.

## 2.3 Conclusions

The MtLRS-HL(TAG)3-ΔCP1 mutants constructed (ANIGT, GNIGG, GLVGE) showed affinity for incorporation of some natural amino acids, especially tyrosine. However, the incorporation of Dansyl-Dap was not observed for any of the MtLRS mutants. We conclude that the system might be used to explore the incorporation of tyrosine analogs.

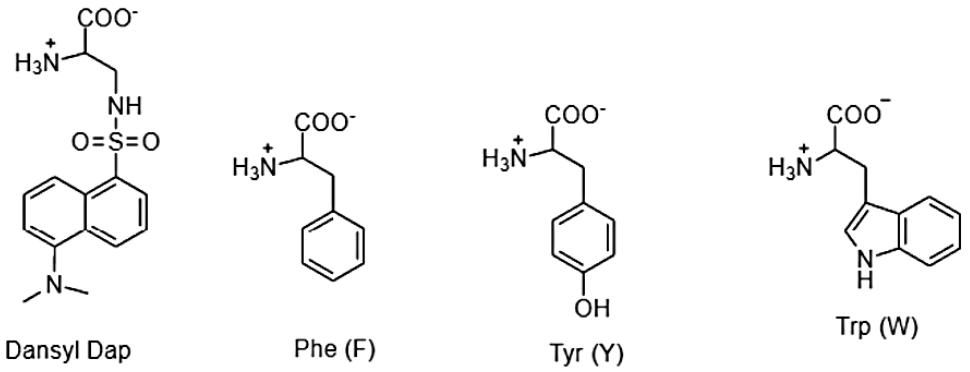



**Figure 2.15** Structures of Dansyl-Dap and aromatic amino acids.

The lack of incorporation of this UAA into the protein can be explained by the fact that although there is a high degree of structural homology between archaea and bacteria, MtLRS may accommodate amino acids in a different way in the active site, being more compact and allowing the incorporation of smaller size aromatic amino acids. Now, the lack of incorporation of phenylalanine makes us speculate that hydrogen bonding may be important in a specific area of the binding site as phenylalanine lacks H-bonding capabilities while in the case of Trp two rings may occupy too much space in the binding site (Figure 2.15). In the next chapter we will explore the use of a library of MtLRS for the genetic selection of MtLRS mutants that could incorporate Dansyl-Dap and other unnatural analogs.

# CHAPTER 3

# DUAL GENETIC SELECTION OF A LEUCYL-TRNA SYNTHETASE

In the previous chapter we demonstrated that the leucyl-tRNA synthetase from *Methanobacterium thermoautotrophicum* (MtLRS) lacking its editing domain and the tRNA from *Halobacterium sp. NRC-1* HL(TAG)3 can efficiently incorporate non-cognate amino acids to the amber stop codon. However, our findings also revealed that the binding site may accommodate amino acids in a different way to the bacterial ones. Therefore, a library of MtLRS lacking its editing domain was designed by randomizing five positions in the amino acid binding site (A32, Y33, Y516, R566 and H576). This library was screened for MtLRS variants that incorporate UAAs through the dual genetic selection including positive and negative selection using the *cat-upp* fusion gene.[26, 31]

## 3.1 Construction of pRCG plasmid

In 2009, Melançon and Schultz developed a dual genetic screening of *Methanococcus jannaschii* tyrosyl-tRNA synthetase that incorporates non-cognate amino acids at the amber stop codon using the *cat-upp* fusion gene.[22] We constructed the pRCG plasmid by combining pBREP and pRep-CatUpp plasmids between *HindIII* and *PciI* restriction sites as shown in Figure 3.1.

pBREP encodes a T7 RNA polymerase which contains two amber stop codons at permissible positions (M1TAG and Q107TAG) , controlling the expression of the *GFPuv* gene. pRep-CatUpp encodes the fusion proteins: chloramphenicol acetyl transferase (CAT) and uracil phosphoribosyl transferase (UPRT). Initially, TAG mutations were introduced at two permissible sites of the c*at* gene: Q98 and D111.

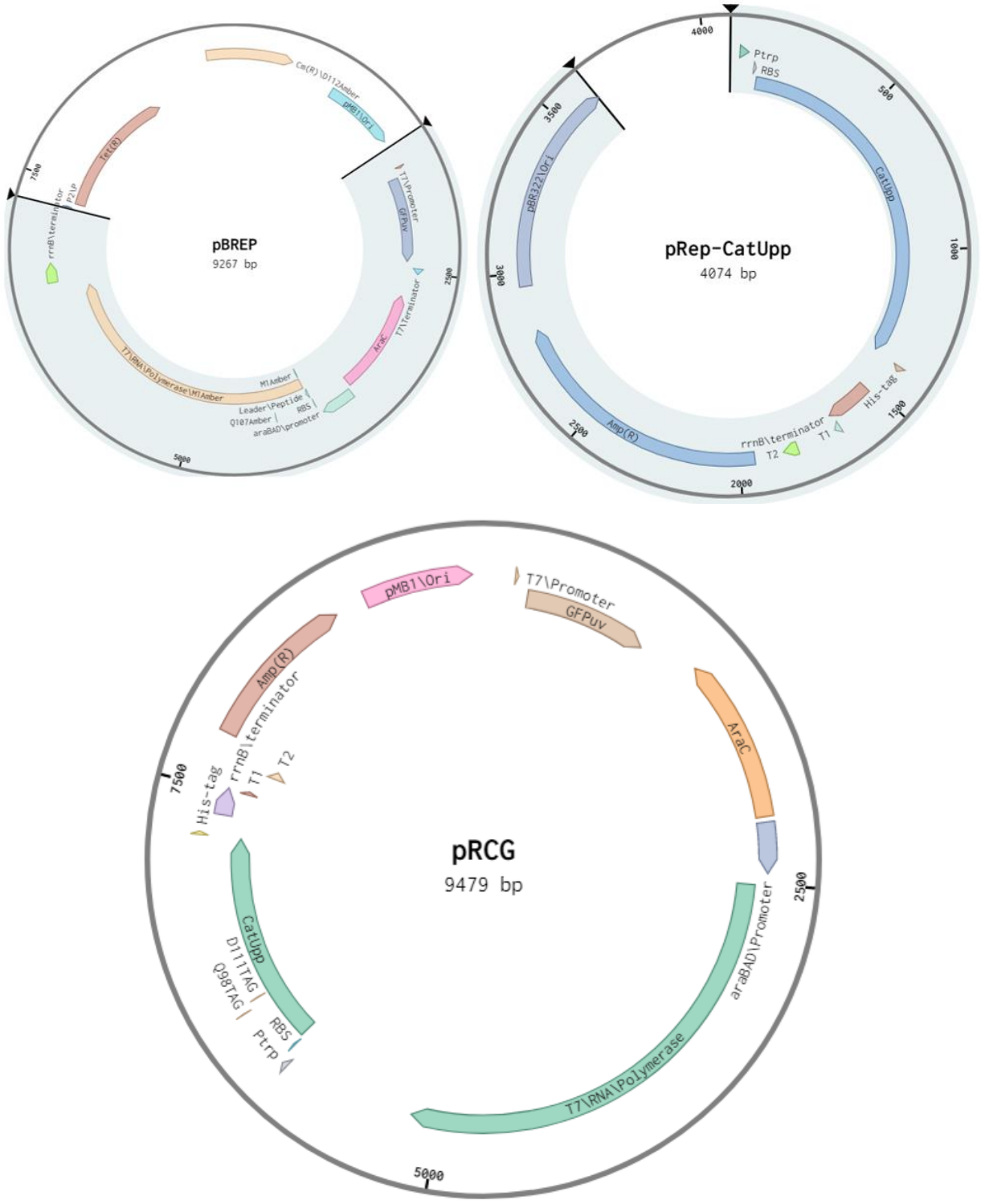



**Figure 3.1** The construction of the pRCG plasmid.

### 3.2 pRCG selection process

The genetic selection was performed in *GH371 E. coli* cells that lack the *upp* gene. During the positive selection, when the selected UAA is present and a functional synthetase loads it to the tRNA, a full-length CAT is expressed resulting in chloramphenicol resistance and yielding colonies. Since the synthetase can also load another amino acid present in the media, a negative selection is necessary.

Negative selection is performed in the presence of 5-fluorouracil (5-FU) without unnatural amino acid. The synthetases that add any natural amino acid will generate a full length UPRT that will form an irreversible covalent interaction with thymidylate synthase (TS), an enzyme necessary for the conversion of dUMP to dTMP, causing the inhibition of TS and resulting in cell death. Only the cells that do not express a full length UPRT will survive. [13, 31] To test the adequacy of the selection system for MtLRS/tRNA pair, the pSupK-MtLRS-HL(TAG)3-ΔCP1 plasmid encoding the wild type synthetase lacking the editing domain was co-transformed with pRCG, and the pre-selection conditions were established.

#### 3.2.1 Amber suppression of MtLRS on pRCG Q98TAG D111TAG

For the positive selection, the transformed cells were plated on GMML media at different concentrations of chloramphenicol and compared with the cells lacking the synthetase. Since the wild type MtLRS is specific for leucine present in the media, it is expected that the full-length CAT is expressed in the presence of the wild type. Therefore, the cells should survive at higher concentrations of chloramphenicol. In addition, because a stop codon is present in the T7 RNA polymerase gene that controls the $gfp_{uv}$ gene, if the MtLRS incorporates an amino acid in

response to the stop codon, green fluorescent protein will be expressed, and the colonies will be fluorescent. Figure 3.2 shows the results for the pre-selection conditions on the pRCG mutant.

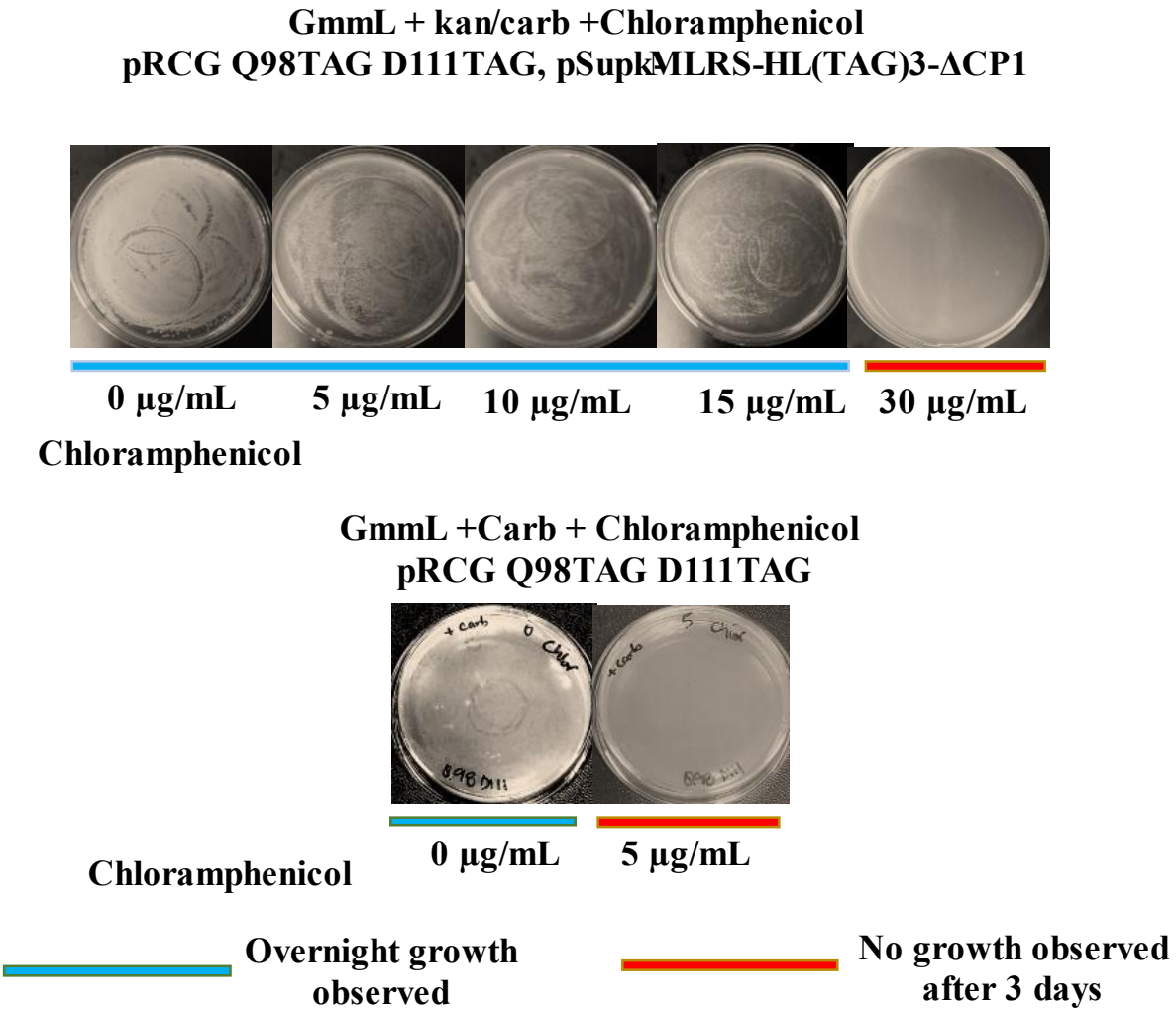
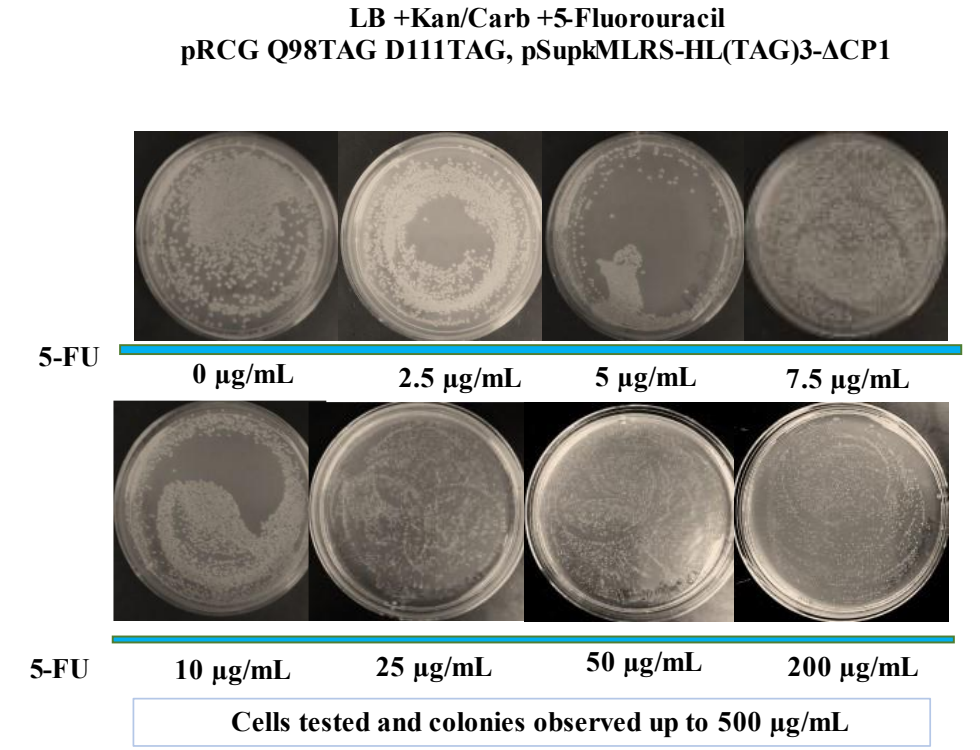



**Figure 3.2** Positive selection conditions obtained when co-transforming *GH371 E. coli* cells with pSupK-MtLRS-HL(TAG)3-ΔCP1 and pRCG (Q98TAG, D111TAG).



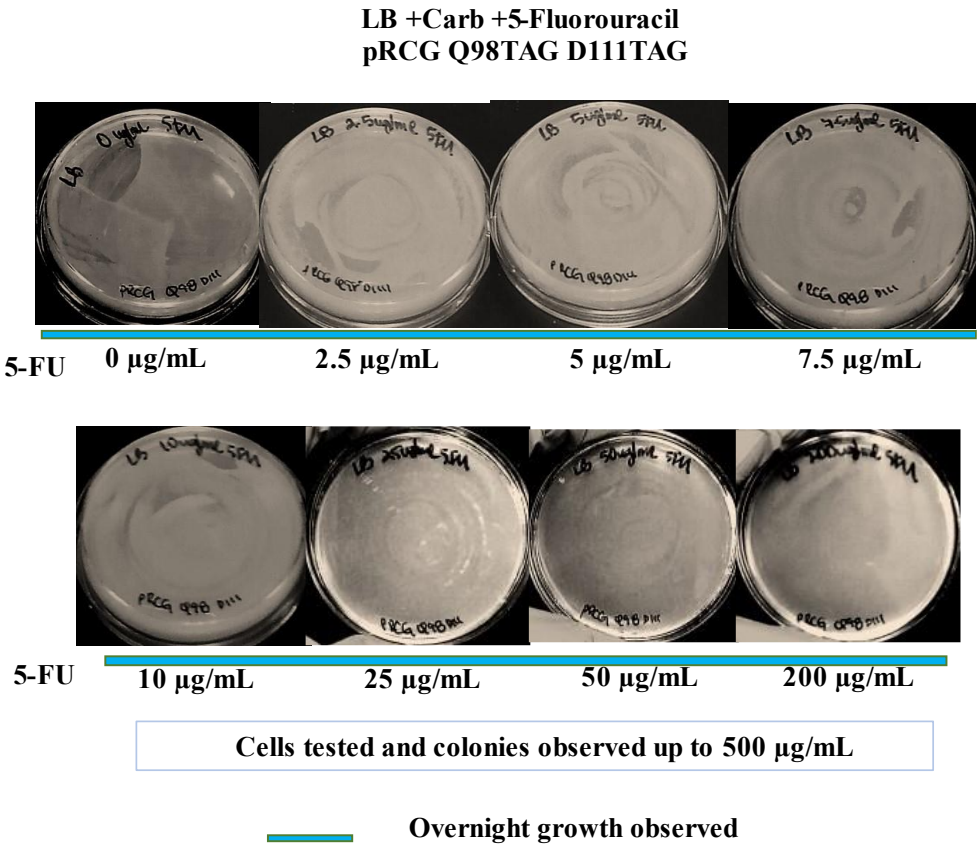




**Figure 3.3** Negative selection conditions obtained when co-transforming *GH371 E. coli* cells with pSupK-MtLRS-HL(TAG)3-ΔCP1 and pRCG (Q98TAG, D111TAG).

The MtLRS was able to incorporate leucine to the CAT Q98TAG D111TAG mutant and confer resistance up to 15 µg/mL of chloramphenicol. In contrast, no colonies were obtained for the cells transformed with the pRCG only in the presence of chloramphenicol.

For the negative selection conditions, in which leucine is also present in the growth media, the cells transformed with both plasmids will die because the full-length UPRT will be expressed. When the double mutant was tested in the presence of 5-FU, no difference was found for the cells with or without wild type MtLRS as shown in Figure 3.3. This result suggested that the negative selection using pRCG double mutant was inefficient. To increase the efficiency of the negative selection conditions, site directed mutagenesis was used to generate two single TAG mutants of pRCG plasmid: pRCG Q98TAG and pRCG D111TAG, because one stop codon should be suppressed more efficiently than two stop codons. However, we found fluorescence on the plates, which indicates that this system was able to suppress both amber stop codons when expressing the $GFP_{uv}$. These results suggest that D111TAG position may play a role in stability of the protein or activity of the *cat-upp* fusion.

### 3.2.2 Amber suppression of MtLRS on pRCG single mutants

Pre-selection conditions were determined again with the wild-type leucine synthetase in GMML media, and the results are shown in Figures 3.5 and 3.6. The clone Q98TAG showed more than 5-fold increase in chloramphenicol resistance compared to the double mutant. Moreover, the Q98TAG mutant showed 1.5-fold higher chloramphenicol resistance than the D111TAG mutant. These results suggest that Q98TAG is suppressed more efficiently than D111TAG although both positions are considered permissible because they are not part of the protein core as shown in Figure 3.4.

For the negative selection conditions, Figures 3.7 and 3.8 showed that single mutants showed no cell growth when 5-FU was above 22.5 µg/mL in the presence of leucine, suggesting the efficient suppression of both Q98TAG and D111TAG.

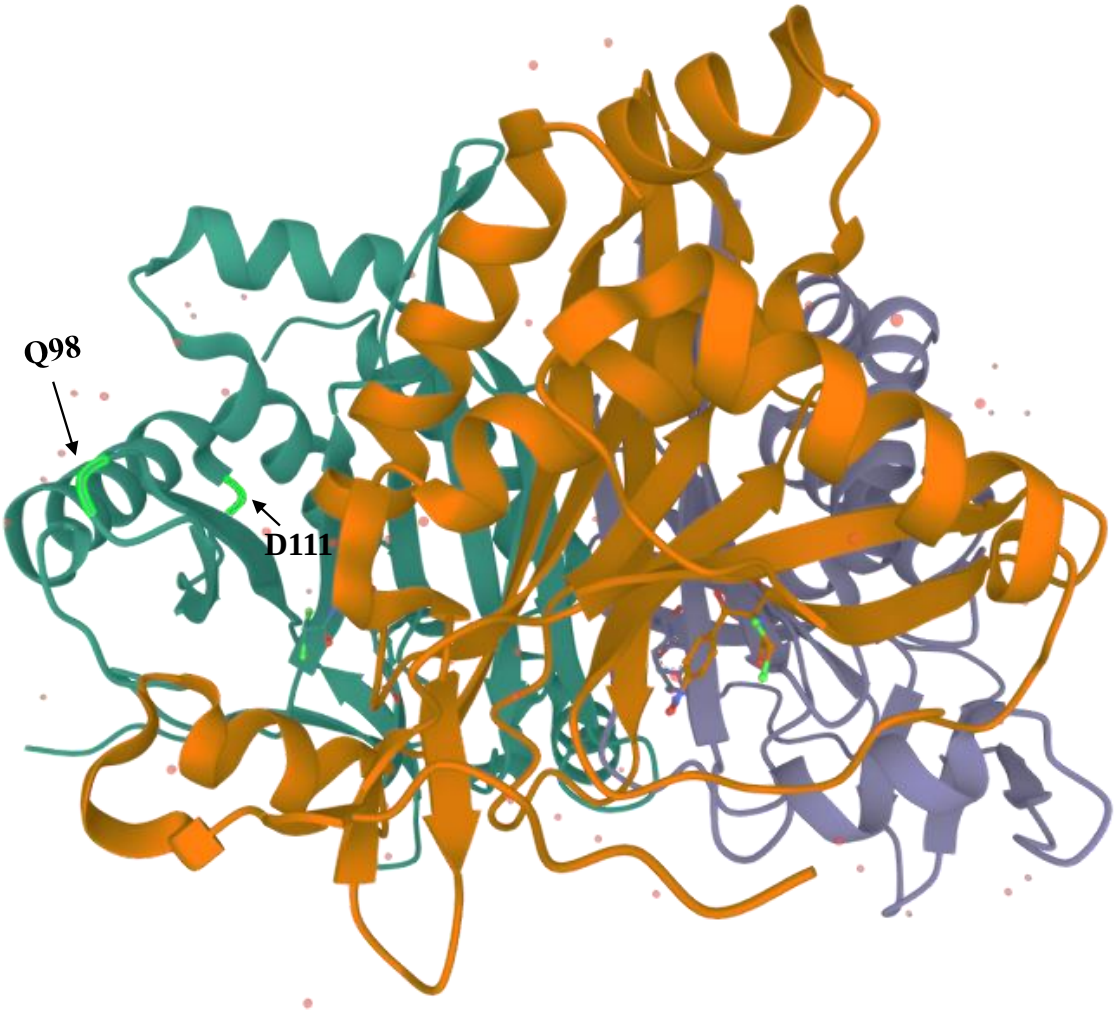



**Figure 3.4** Crystal structure of *E. coli* chloramphenicol acetyltransferase. Permissible positions are highlighted in bright green. (RCSB PDB: 3U9F)

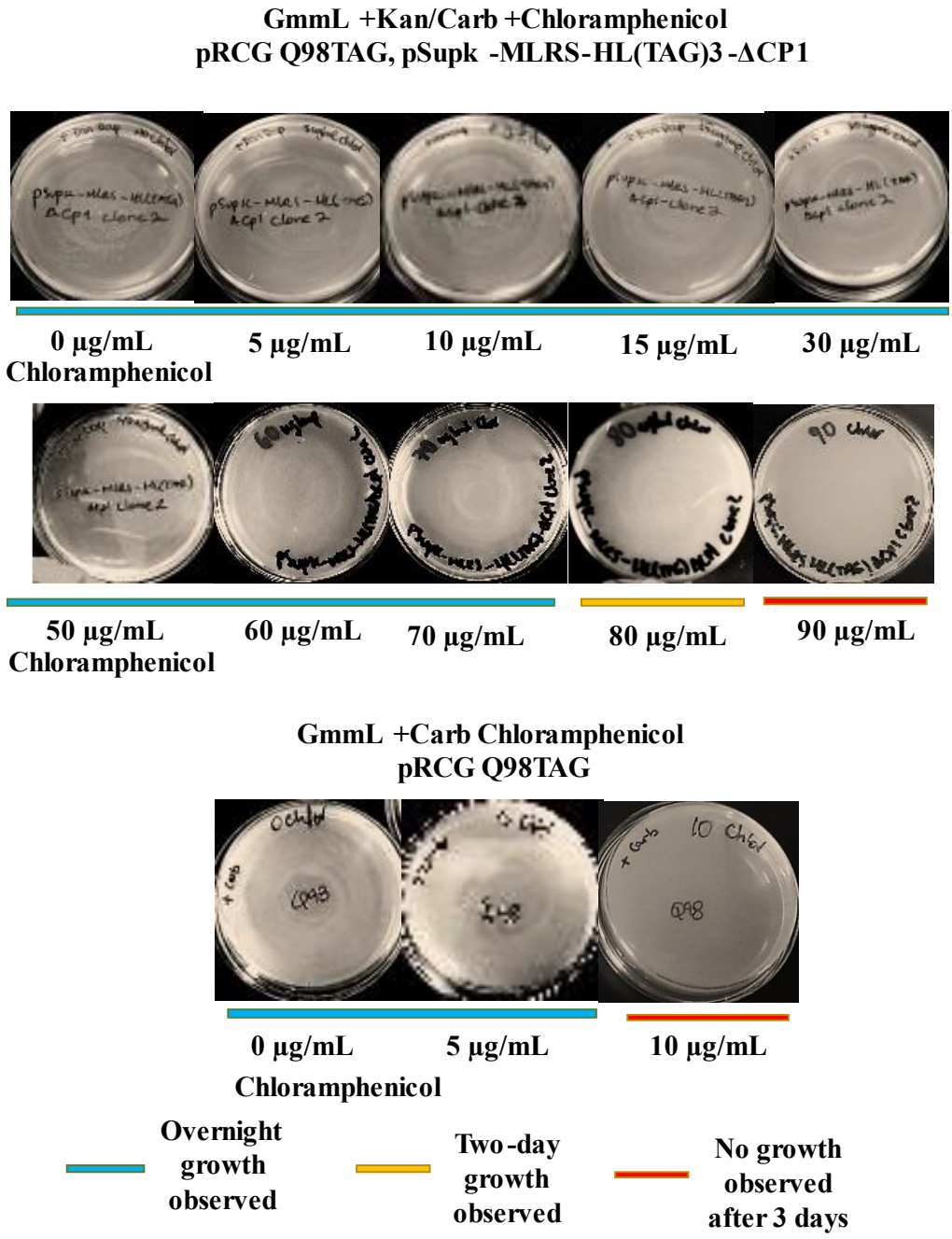



**Figure 3.5** Positive selection conditions obtained when co-transforming *GH371 E. coli* cells with pSupK-MtLRS-HL(TAG)3-ΔCP1 and pRCG Q98TAG.

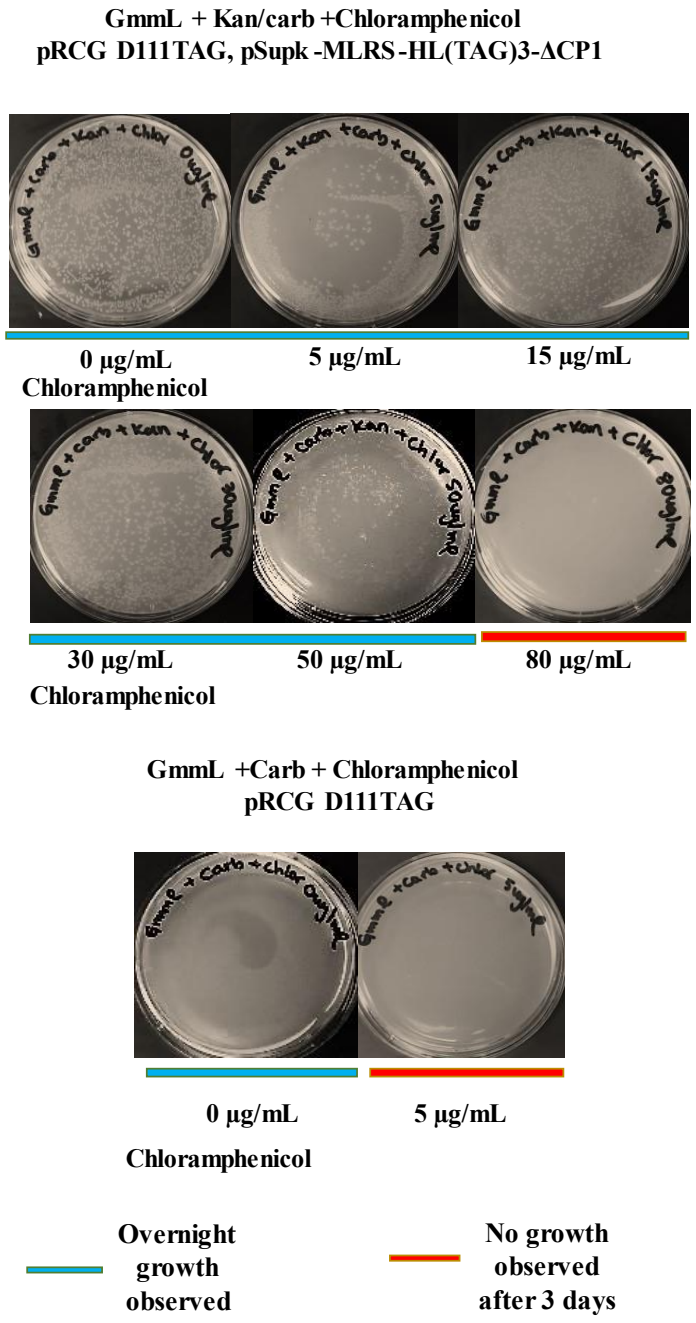



**Figure 3.6** Positive selection conditions obtained when co-transforming *GH371 E. coli* cells with pSupK-MtLRS-HL(TAG)3-ΔCP1 and D111TAG.

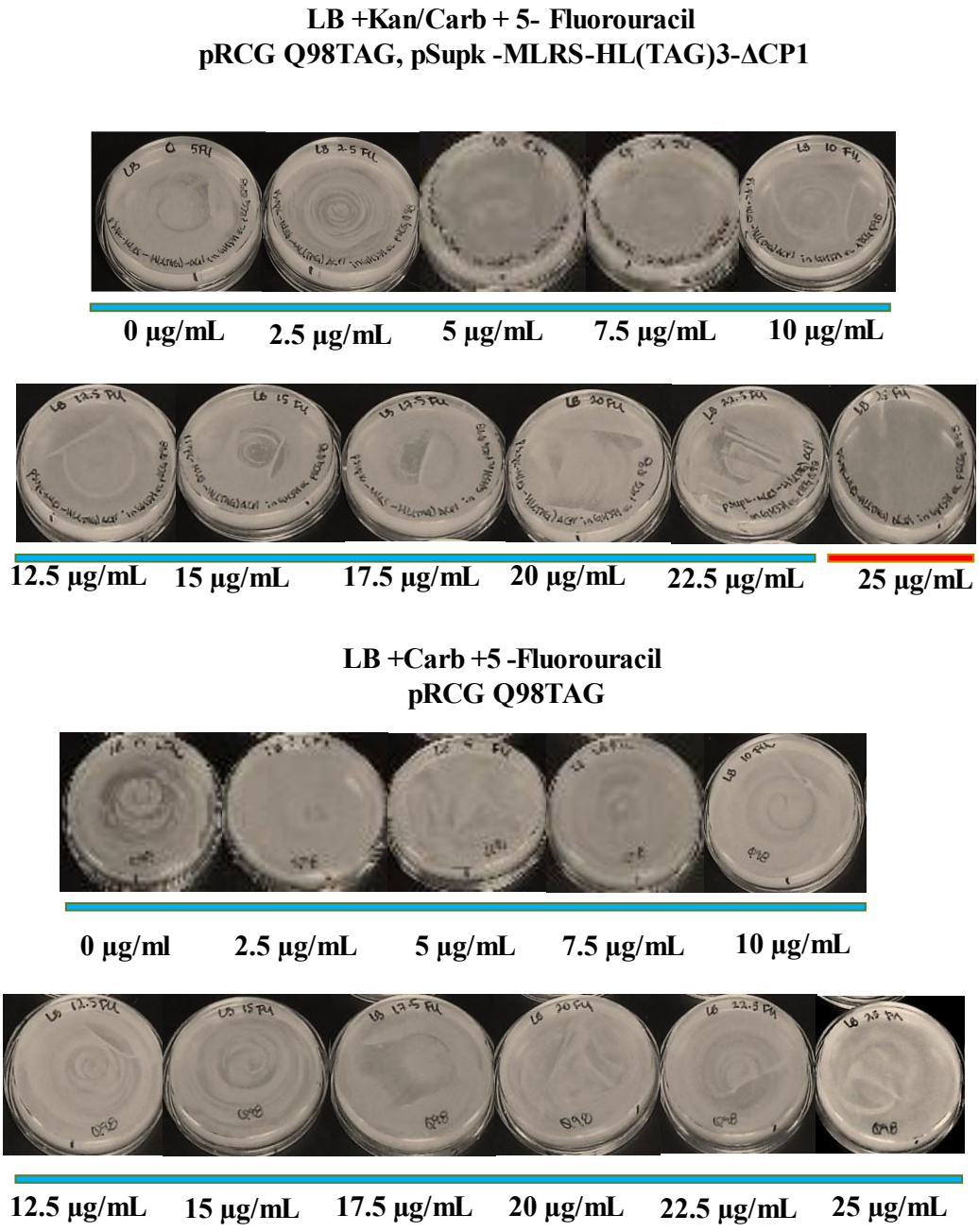



**Figure 3.7** Negative selection conditions obtained when co-transforming *GH371 E. coli cells* with pSupK-MtLRS-HL(TAG)3-ΔCP1 and pRCG Q98TAG.

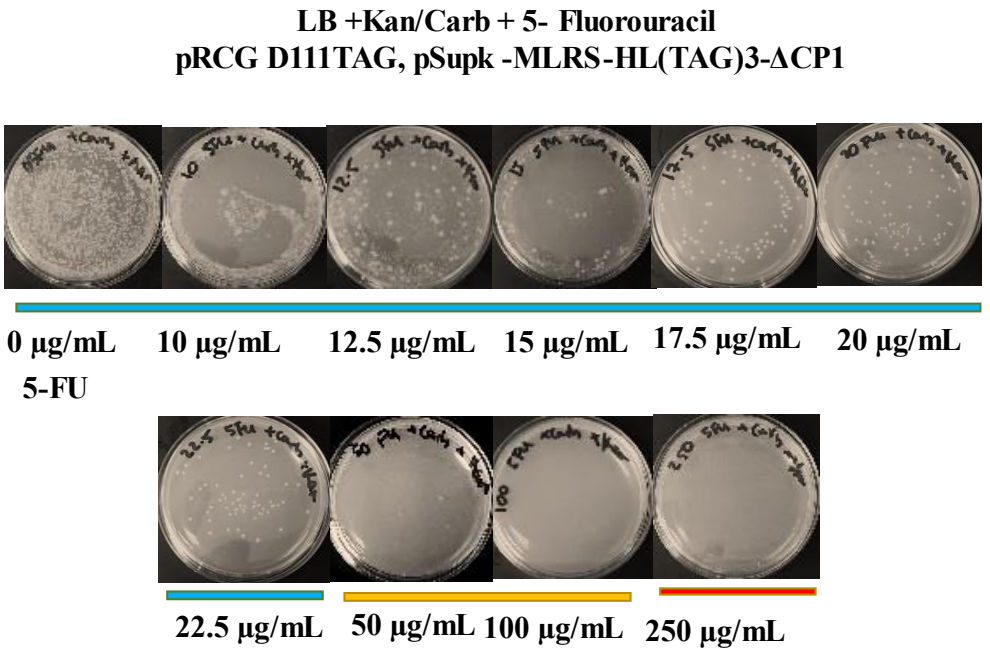



**Figure 3.8** Negative selection conditions obtained when co-transforming GH371 *E. coli* cells with pSupK-MtLRS-HL(TAG)3-ΔCP1 and pRCG Q98TAG and D111TAG single mutants.

### 3.3 Construction of a MtLRS library

Overlapping extension PCR (OE-PCR) was used to make a library with five randomized positions (A32X, Y33X, Y516X, R566X and H576X) in the leucine binding site (Figure 3.9). *GH371 E. coli* cells were co-transformed with pRCG Q98TAG and the MtLRS library. Cells were amplified, and dual genetic screening in the presence of two UAA: Dansyl-Dap and 4-nitro-1-phenylalanine (4-nitro-1-phe) was separately performed.

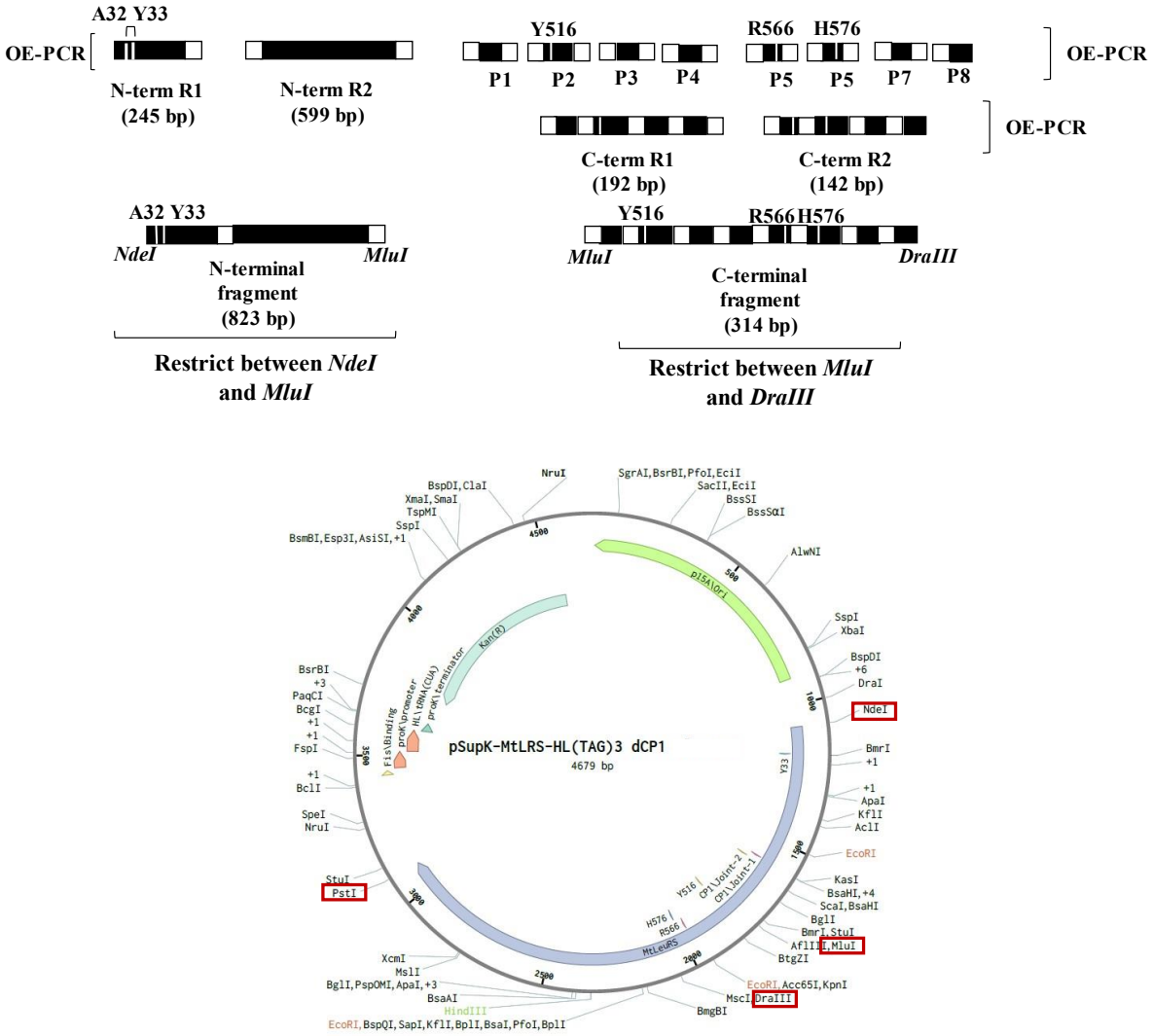



**Figure 3.9** Construction of MtLRS library using OE-PCR.

After three rounds of positive selection and two rounds of negative selections, clones were assayed individually for growth on chloramphenicol containing plates in the presence or absence of Dansyl-Dap and 4-nitro-1-phe. Figure 3.10 shows the results of the first 96 clones tested for each UAA.

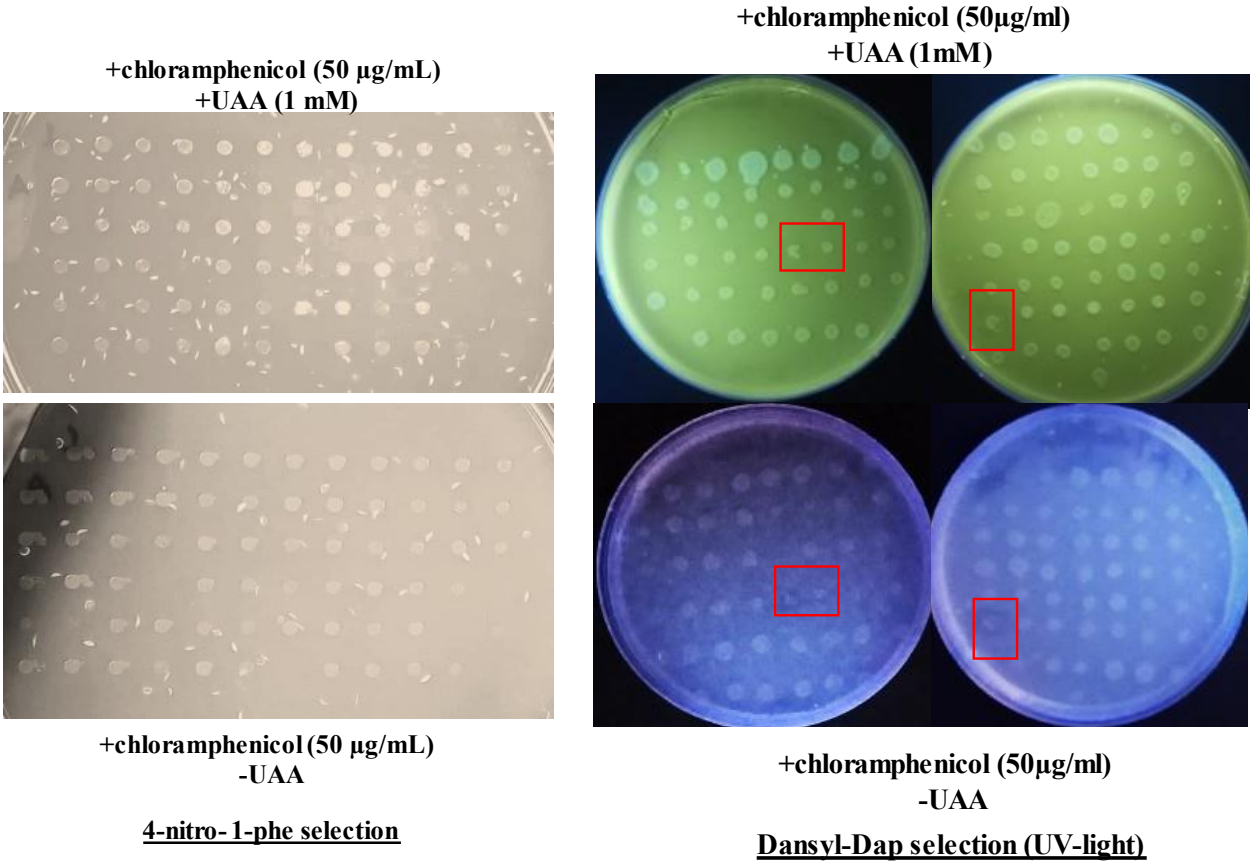



**Figure 3.10** Re-testing of first 96 clones after third round of selection in the presence and absence of 4-nitro-1-phenylalanine (left) and Dansyl-Dap (right) with 50 µg/mL chloramphenicol on GMML growth media.

Out of the first 96 clones assayed for growth on 50 µg/mL of chloramphenicol with and without 4-nitro-1-phe, no clones were specific for UAA only. In the case of Dansyl-Dap, three clones appeared only on the plates containing this UAA, identifying them as possible functional synthetases to incorporate this amino acid.

### 3.4 Conclusions and future directions

The pRCG Q98TAG D111TAG plasmid was proven ineffective for the negative selection because no difference was observed in 5-FU sensitivity between the cells transformed with and without MtLRS- HL(TAG)3-ΔCP1 in GMML media.

Both single mutants (pRCG Q98TAG and pRCG D111TAG) were able to suppress amber stop codon efficiently for both positive and negative selection steps. The Q98TAG showed some background suppression at 5 µg/mL chloramphenicol, but it can be used within 10-80 µg/mL of chloramphenicol and from 22.5 µg/mL of 5-FU. The D111TAG mutant can also be used, but the concentration of chloramphenicol needs to be adjusted within 5-50µg/mL and the concentration of 5-FU needs to be higher than 100 µg/mL.

The MtLRS library in which five amino acids in the active site were randomized was prepared, and the dual genetic screening for Dansyl-Dap and 4-nitro-phe was performed using the pRCG Q98TAG plasmid. After three rounds of positive selection and two rounds of negative selection for each UAA, ninety-six clones were re-tested in the presence and absence of UAA and 50 µg/mL of chloramphenicol. For 4-nitro-1-phe, no differences were found for the first 96 clones, and more clones need to be re-tested. In the case of Dansyl-Dap selection, three clones were identified and are currently being sequenced. These clones will be used to express the Z-domain protein with a TAG stop codon at the $7^{th}$ position for the incorporation of Dansyl-Dap.

# CHAPTER 4

# RIBOSWITCHES AND GENE REGULATION

Riboswitches are untranslated region (5'-UTR) of bacterial and some eukaryotic mRNAs that upon binding to metabolites regulate the expression of proteins that involve the synthesis, transport or degradation of these metabolites.[32, 33] Riboswitches have two main components: an aptamer domain responsible for binding to the target molecule with high specificity, and the expression platform, located in proximity to transcription or translation sites to regulate gene expression.[34] Although most riboswitches are composed of one aptamer, tandem aptamers are also found. [32] The interactions between the metabolite and the aptamer domain, lead the structural changes in the expression platform to release or sequester key gene regulatory elements.

## 4.1 Regulation Mechanisms

Main regulation mechanisms of riboswitches are based on controlling transcription, translation, splicing of genes or altering RNA stability. There are more than forty classes of riboswitches that regulate metabolites such as amino acids, nucleotide bases, coenzymes, signaling molecules, and ions.[34, 35] Four major mechanisms by which riboswitches regulate gene expression are outlined in Figure 4.1.[36]

Figure 4.1a shows "ON" transcription switches like the glycine riboswitch. When the metabolite is not present, the RNA is folded in a way that the U-rich region is sequestered to the transcription terminator sequence forming a stem and preventing the binding of RNA polymerase. Once glycine binds, the RNA structure is rearranged to create an anti-terminator loop which allows transcription to continue.[37] "OFF" transcription switches like the vitamin $B_{12}$

work as shown in Figure 4.1b. When the metabolite is not present, the anti-terminator loop is formed to continue the transcription of genes necessary for the synthesis of the metabolite. When an optimal level of metabolite is reached, conformational changes induce the formation of the terminator loop, shutting down the transcription.[38]

For "ON" translation switches (Figure 4.1c), without target molecule, the Shine-Dalgarno sequence or ribosomal binding site (RBS) is sequestered in the expression platform, resulting in no translation of RNA. Once the metabolite binds to the RNA, the conformational changes induce the release of the RBS to initiate translation. We will discuss this case with the theophylline riboswitch in the next chapter.[31] For "OFF" translation switches (Figure 4.1d) like the AdoCbl riboswitch, RBS is sequestered upon binding to the metabolite to repress translation.[34, 39]

Some uncommon mechanisms of gene regulation by riboswitches include the riboswitch ribozymes involving the cleavage of mRNA upon binding to the metabolite (e.g. *glmS* riboswitch)[40] and the intron self-splicing (e.g. c-di-GMP switch).[41]

**4.2 Riboswitch folding, recognition and binding**

Previous studies have revealed that riboswitches follow some common architectural principles: once the RNA forms base pairs (helices) or loops (non-paired regions), these can form coaxial stacks in which two or more individual helices arrange in a continuous pseudo continuous helix that can be further arranged in parallel or perpendicularly to form D- and T-loops. [42, 43]

More complex architectures in the riboswitch structures include hairpins in which the 3'- and 5'-extremes of the helix are linked and internal loops in which at least one of the strands of RNA is

not base paired. Junctions or multiloops are formed by the interaction of multiple double helices, and specific binding motifs are where the aptamer region is usually located. [43]

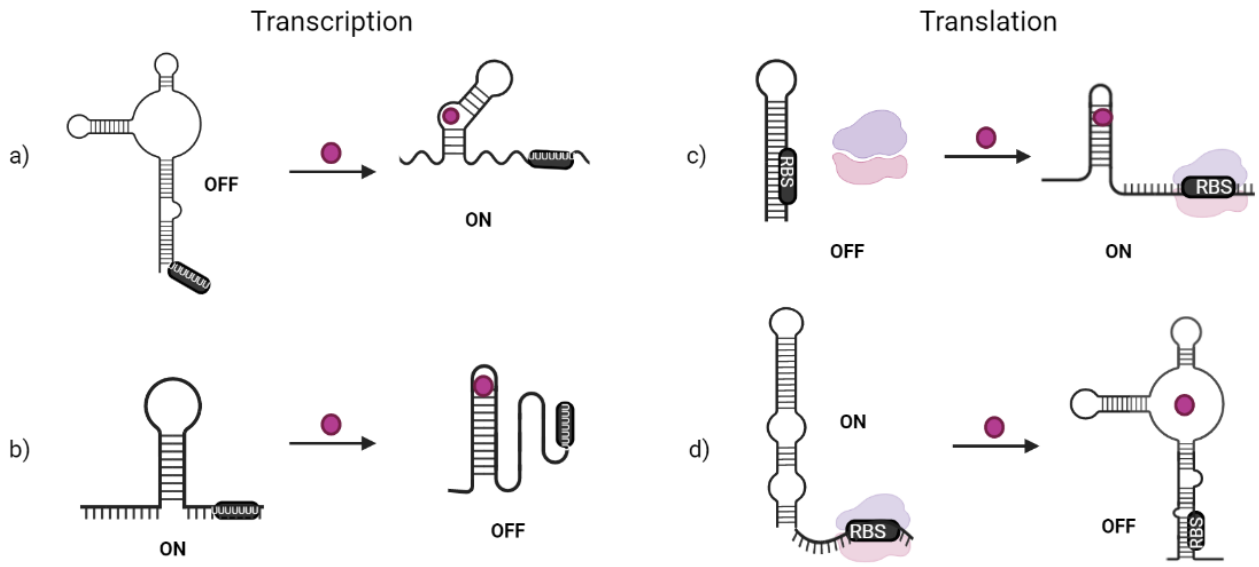



**Figure 4.1** Riboswitch canonical regulation mechanisms: a)single aptamer transcription activation, b)single aptamer transcription attenuation. Anti-terminator is highlighted in black. c)single aptamer translation initiation, d)single aptamer translation repression Ribosomal binding site is highlighted in black.[34, 35]. Created with Biorender.

There is no universal metabolite recognition mechanism among riboswitches, but most riboswitches form tight binding pockets that can bind ligands with affinities as low as 2.1 nM like in the case of the *Thermoanaerobacter tengcongensis* PreQ riboswitch.[44] This high specificity is possible due to a complex network of interactions that involve base pairing, water/salt bridging, electrostatic or van der Waals. As outlined in Figure 4.2, the binding pockets are formed by non-canonical base pairs and unions of stems. In some cases, ions stabilize the charges between ligands and residues responsible for binding.[39] While some riboswitches have well conserved aptamer domains with defined architectures, some do not show

correlation between the structure and ligand binding, and rely more on the RNA residues and interactions with the metabolite to build specific spatial conformations. In fact, there are several S-adenosyl methionine (SAM) switches with different structural features that can bind to the same ligand in different ways.[45] Also, some riboswitches including the glycine riboswitch have more complex structures and can bind more than one ligand.[46]

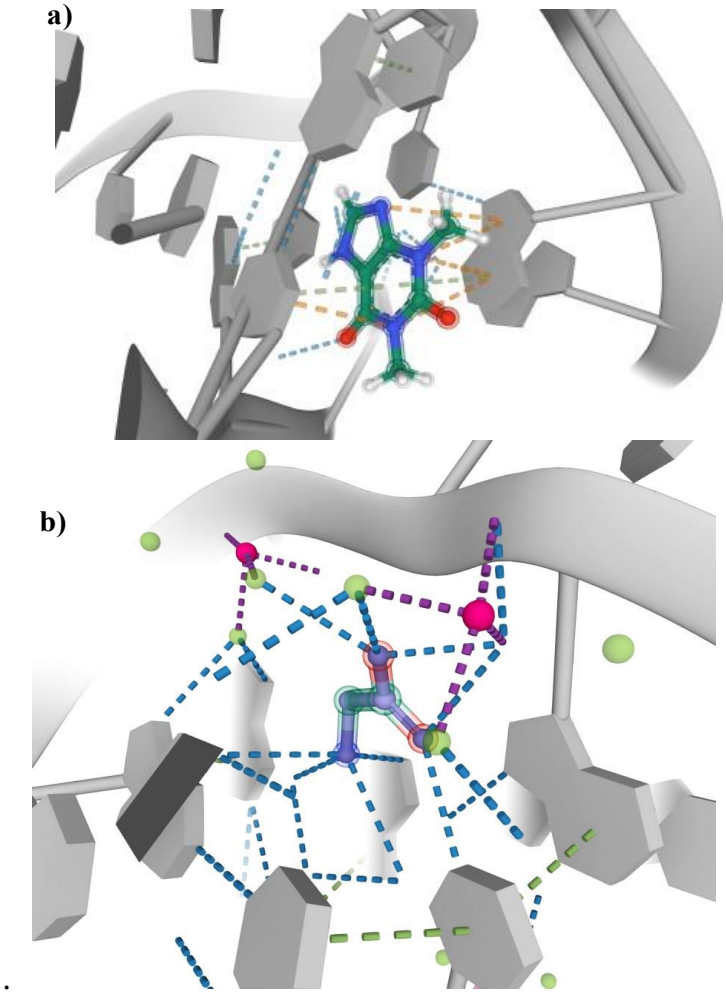



**Figure 4.2** Crystal structure-based network of interactions at a)theophylline-binding RNA and b)glycine riboswitch. Interactions denoted by color scheme: blue (hydrogen bonding), green (π-stacking), orange (cation-π), purple (metal coordination). Green and pink spheres represent water and $Mg^{2+}$ ions respectively. Retrieved with Mol* using RSCB PDB: 1O15 and 3OWW as templates.[47]

### 4.3 Synthetic Riboswitches

Synthetic riboswitches are new methods for gene regulation and detection of exogenous molecules. The first synthetic riboswitch corresponded to a theophylline riboswitch which was further optimized to be able to discriminate about 10000-fold against caffeine, which only differs by a methyl group at the N7 position.[48, 49] This aptamer binds to theophylline with a dissociation constant ($K_d$) of 0.32 μM. [49] Naturally occurring riboswitches have been modified at the aptamer domain through the introduction of libraries that are further selected for the molecule of interest.[50, 51] The use of *in silico* high throughput screening is rapidly increasing to narrow down the sequences to be considered for selection. In this method, a pool of RNA sequences is put into computer software to identify aptamer-like motifs, which are a result of machine-learning programs that consider the sequence of DNA or RNA and its comparison towards an aptamer database, X-ray and 2D structure prediction software, and NMR binding studies.[52-56] In addition, some of these *in silico* analysis can include algorithms to calculate predicted binding energies between the aptamer and the small molecule or target protein.[57, 58]

Some of the current applications of riboswitches and aptamers include the design of biosensors to detect small molecules, functionalization of membranes for aptamer-based ultrafiltration systems, drug delivery, therapeutics, among others.[59-65] In addition, a riboswitch *in vivo* activation may be useful to control gene expression in a similar way that lactose in the *lac* operon system.[66]

Once the library of specific size and characteristics is designed, synthetic riboswitches are selected from the library in the presence of the target of interest.[67, 68] Some selection processes are highlighted below.

### 4.3.1 Systematic Evolution of Ligands by Exponential Enrichment (SELEX):

First reported in 1990, SELEX is a technique for *in vitro* selection of a library of aptamers based on binding to the molecule of interest. Figure 4.3 outlines the process in SELEX. The target molecule is bound to a solid surface and the pool of aptamers is passed through. The sequences with affinity for the target compound will remain bound while the rest will be washed away. The nucleic acid bound to the target ligand are collected, amplified, tested for affinity, and sequenced. Finally, the obtained clones are sequenced.[67, 69, 70]

SELEX is a cell-free method that can be used to select aptamers for molecules that are toxic to cells. However, it requires the immobilization of the target to a solid support, large amount of reagents, and a different protocol for different targets to be tested.[70] Several approaches to overcome these limitations include *in silico* pre-screening, use of restriction enzymes to verify DNA or RNA hybridization, and capillary electrophoresis to separate ligand-free aptamers from ligand-bound aptamers, etc.[60, 71-74] Recently, magnetic beads have been functionalized with the target molecule in a chip able to apply highly localized magnetic fields that facilitate the separation of bound and unbound aptamers, however most of the techniques are still expensive.[71]

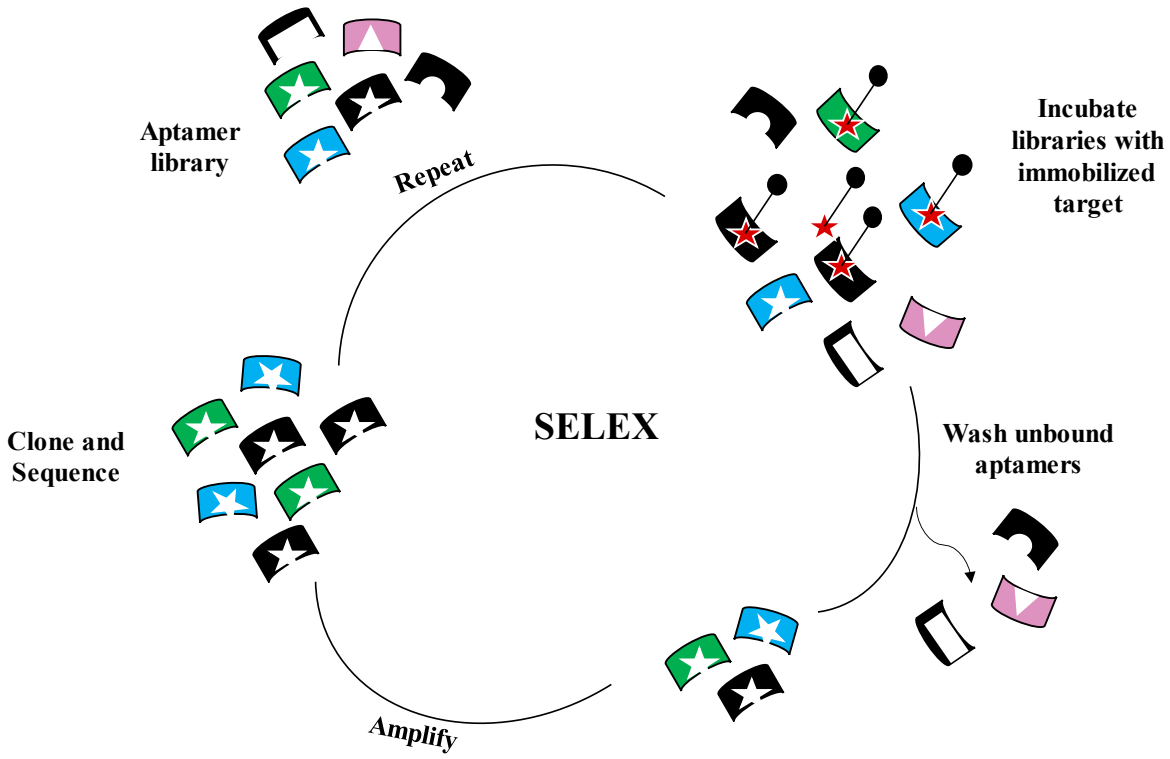



**Figure 4.3** Graphic representation of aptamer selection process using SELEX.

### 4.3.2 Dual genetic screening

Synthetic riboswitches can be selected by the activation of a selectable gene upon binding to the target ligand in living cells. The transformation efficiency needs to be considered during the design of the library because the number of variants tested can be limited by the number of transformed cells.[52] However, one advantage of *in vivo* selection is that the aptamer does not necessarily bind to the target *in vitro* in the same way as in the cellular environment.[50] When a library of aptamers is introduced at the upstream of selection marker genes, functional riboswitch variants can activate the gene in response of the target molecule (Figure 4.4). [75]

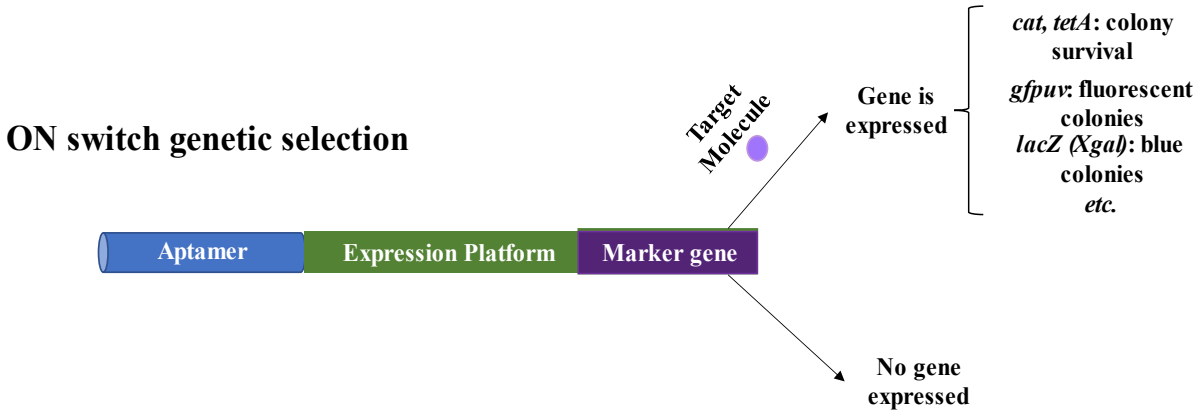



**Figure 4.4** Outline of genetic selection of ON Riboswitches.

While most aptamer selection methods are based on a combination of *in silico* and *in vitro* approaches, there are a few examples of *in vivo* riboswitch selection methods based on the activation of $gfp_{uv}$, *lacZ* and antibiotic resistance genes (*cat* and *tetA-gfp*).[50, 75-78]

### 4.4 Riboswitch binding studies

Reporter genes are used to elucidate the correlation between the concentration of the target molecule and the signal obtained. There are several ways to determine the binding affinities between the aptamers and the molecules of interests and how well the riboswitch discriminates other molecules.

#### 4.4.1 Isothermal Titration Calorimetry (ITC)

It provides a label-free way to measure the equilibrium association constants between the aptamer and the targets of interest. The ligand is titrated with the RNA solution through serial injections and the heat released or absorbed is measured. This method requires high concentrations of RNA (mM or above), and both RNA and the target ligand should be prepared in the same buffer. Thermodynamic parameters like $\Delta H$, $\Delta S°$ and $\Delta G°$ or the binding stoichiometry ($N_A$) can be determined for this method.[79, 80]

### 4.4.2 In-line probing

Taking advantage of the intrinsic instability of RNA, this method is used to elucidate the secondary structure and binding site of aptamers.[81] The RNA molecules labeled with fluorophores or $^{32}P$ are incubated in the presence of ligand and subjected to partial alkaline digestion. The regions that are not organized are more likely to be digested than the ones forming structures with the ligand. Fragments are then analyzed using gel electrophoresis.[82]

### 4.4.3 Fluorescence

The aptamer or the ligand is labeled with a fluorophore and the fluorescence is measured by increasing the concentration of ligand.[60, 83] The changes in the fluorescence are determined between the ligand-free aptamer and the ligand-bound form. Fluorescence resonance energy transfer (FRET) between the aptamer and ligand can also be measured. This method is widely used but fluorophores can alter the binding properties of RNA or ligand, so it must be performed over a range of pH to assure that the changes in fluorescence are a consequence of binding and not changes in the species.[84]

### 4.4.4 Circular dichroism

Because of the presence of asymmetric carbons in RNA, it is possible to study the structural changes in RNA by circular dichroism (CD).[85] This technique is based on the difference in absorbance of left and right circularly polarized light.[84] Aptamers are titrated with increasing concentrations of ligands and the change in the CD spectrum is used to determine the structural changes and binding parameters such as the dissociation constants.[86] This methods requires low concentrations and volume of the aptamers.

Other methods include differential scanning calorimetry (DSC), surface plasmon resonance (SPR), equilibrium dialysis and ultrafiltration for which membranes of 3500 kDa are used for fragments with sizes around 180bp.[84, 87]

### 4.5 Conclusions

Riboswitches are highly adaptable molecules because of the specific interactions and spatial arrangement upon binding to the molecule of interest.[49, 67] Their gene regulation mechanisms could be used for biosensing, gene expression reprogramming, imaging, and gene therapy.[66, 64] Although *in vitro* methods are widely used for the selection of aptamers, there is plenty of opportunity in selecting aptamers based on gene-screening systems *in vivo*. We will explore the use of the *cat-upp* gene fusion for the dual genetic selection of synthetic riboswitches in the next chapters.[69, 88]

# CHAPTER 5

# DUAL GENETIC SELECTION OF THE THEOPHYLLINE RIBOSWITCH WITH ALTERED APTAMER SPECIFICITY FOR CAFFEINE

## 5.1 Design of a riboswitch aptamer library from a synthetic theophylline riboswitch

As a proof of concept, we designed a synthetic theophylline aptamer library from a theophylline riboswitch (TheoRS) with an optimized expression platform for *E. coli* that was previously reported by Lynch et al.[89, 90] Theophylline riboswitch (TheoRS) is an "ON" switch that activates translation. Upon binding of theophylline to the aptamer region, the ribosomal binding site or Shine-Dalgarno sequence is exposed, activating the downstream genes. The aptamer region in this riboswitch was reported to interact with the ligand through eleven bases: 9-12 (5'-CAGC-3'), and 24-30 (5'-GCCCUUG-3'). These positions were randomized as outlined in Figure 5.1 by introducing degenerated NNN codons (randomized A, G, C, T) through PCR, generating a library of $4^{11}$ possible variants (≈4.2 millions). This library was inserted into the pTrp-TheoRS-CatUpp between the *trp* promoter and the *cat-upp* gene fusion gene.[31] The *trp* promoter assures high riboswitch activation and protein expression.[91, 92]

## 5.2 Dual genetic screening to select aptamers

The selection scheme is outlined in Figure 5.2. The conditions were previously determined as 30 μg/mL of chloramphenicol and 2.5 mM of caffeine for the positive selection, and as 0.5 μg/mL of 5-fluorouracil (5-FU) for the negative selection. *GH371 E. coli* cells lacking *upp* gene were used in the genetic selection.[93] When the aptamer binds to caffeine or any other endogenous metabolite, it triggers a conformational change in the RNA structure that induces the release of

the RBS and subsequent expression of chloramphenicol acetyl transferase (CAT), allowing cells to survive in the presence of chloramphenicol. To assure that the aptamer only binds to caffeine and not any endogenous ligands, negative selection plates contain 5-FU. Any endogenous metabolite that binds to the riboswitch will express the UPRT protein, which in the presence of 5-FU leads to the inhibition of thymidylate synthase, resulting in cell death.[94]

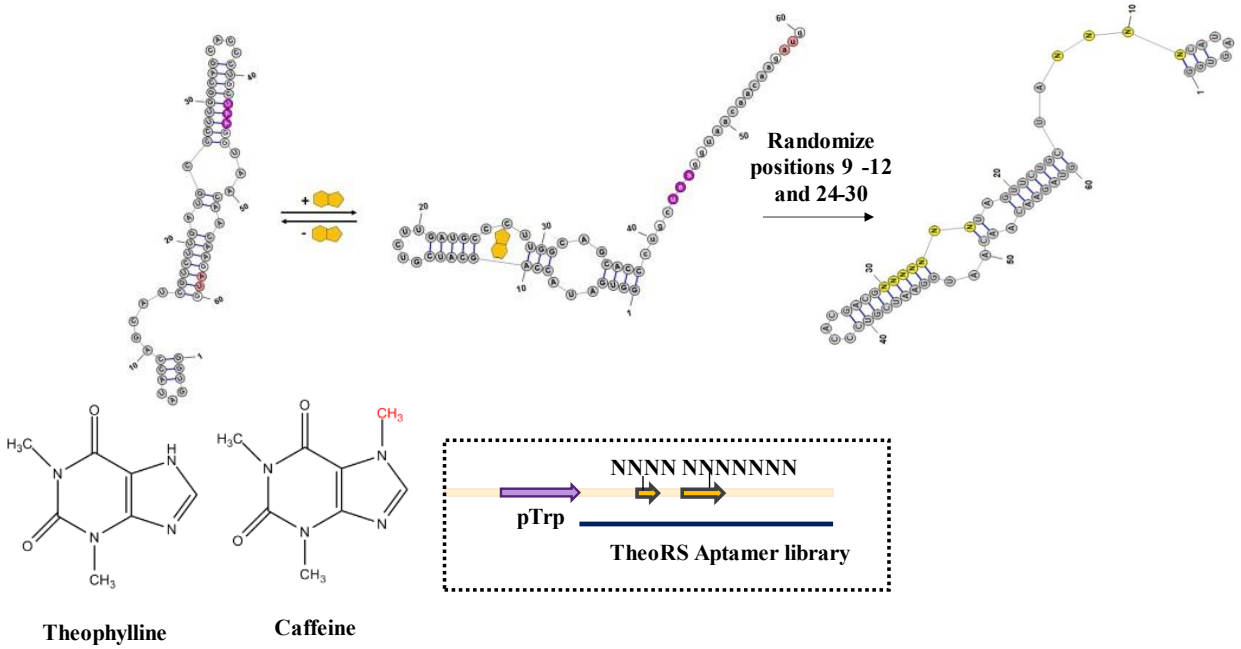



**Figure 5.1** Strategy to design the TheoRS library. Figure shows original binding to theophylline (yellow structure) and randomization of positions 9-12 and 24-30 using PCR. RNA 2D structures were modeled using RNAStructure 6.3 software.[95, 96]

This process was repeated three times for the positive selection and two times for the negative selection. One hundred and ninety-two clones were randomly picked, grown, and stamped on plates containing 50 µg/mL chloramphenicol in the presence and absence of caffeine. One clone was selected for the growth advantage in the plate containing caffeine, and it was tested under the same conditions as shown in Figure 5.3.

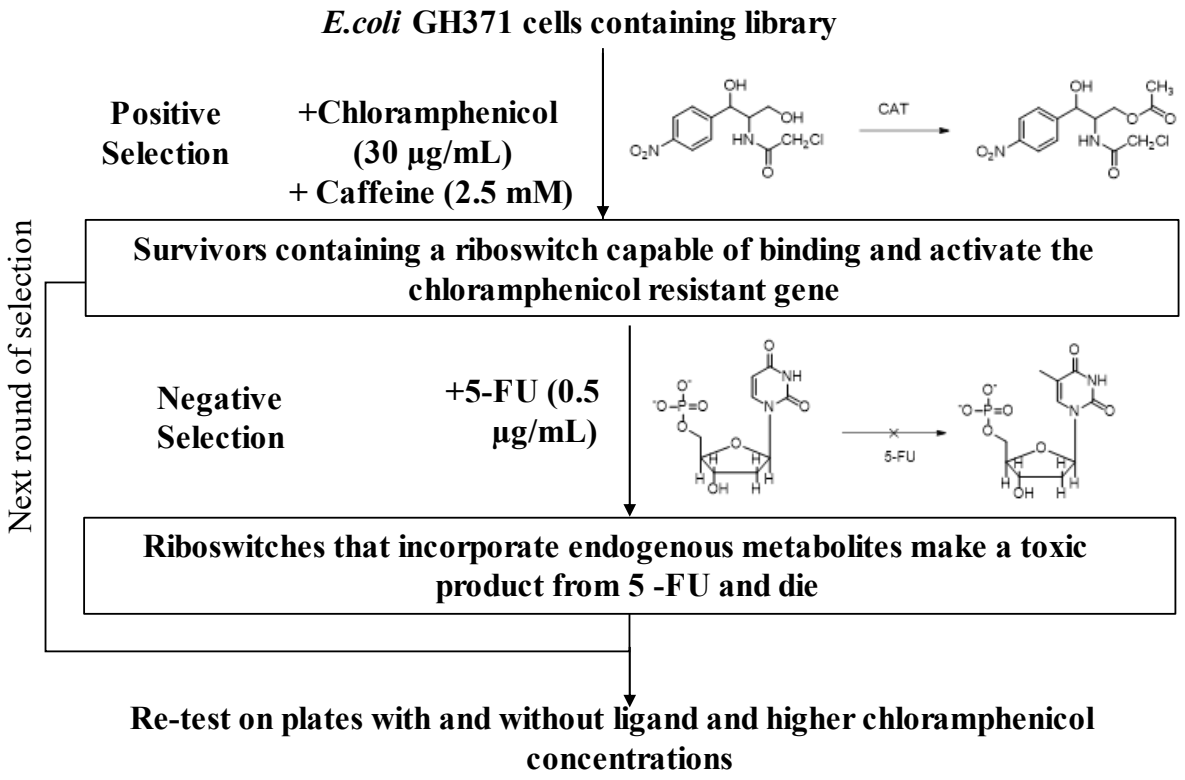



**Figure 5.2** Dual genetic selection for caffeine riboswitches using *cat-upp* fusion gene.

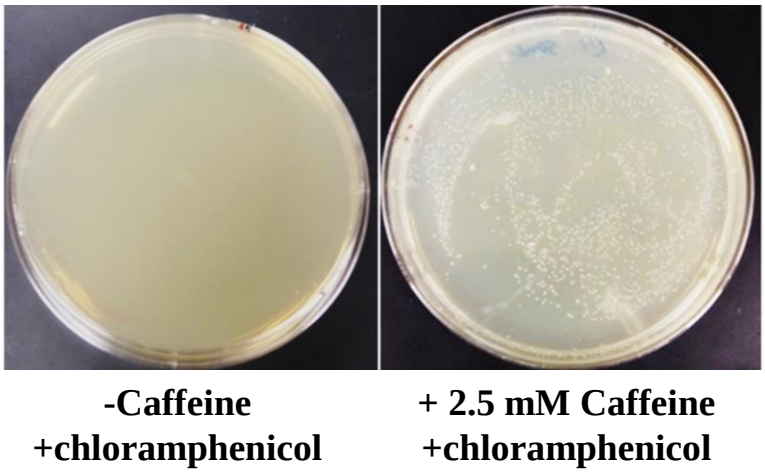



**Figure 5.3** Caffeine induced expression of chloramphenicol acetyl transferase (CAT) in *GH371 E. coli* cells.[31]

The selected clone was sequenced and the bases at the aptamer domain were: 5'-ATTT-3' and 5'-TGTGTAC-3' for the positions 9-12 and 24-30 respectively. This clone was named caffeine riboswitch (CaffRS).

### 5.3 Caffeine induced expression of β-galactosidase (LacZ) and green fluorescent protein ($GFP_{uv}$)

LacZ and GFP are widely used as reporter proteins to measure the levels of gene expression by colorimetric and fluorescence assays.[88, 96-99] In order to study whether CaffRS could activate the expression of these proteins, CaffRS was inserted between the *trp* promoter and *lacZ* or $gfp_{uv}$ genes to produce the pTrp-CaffRS-LacZ and pTrp-CaffRS-GFPuv plasmids, respectively. *XL1-Blue E. coli* cells were transformed with the respective plasmids and were grown overnight at 37°C in Luria Bertani media (LB) containing 2.5 mM caffeine. Blue colonies were formed by LacZ expression in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and fluorescent colonies were formed by $GFP_{uv}$ expression and as shown in Figure 5.4. In contrast, for the cells grown in the absence of caffeine, white colonies were observed for the LacZ expression, and little fluorescence was detected for the GFP expression.

When caffeine-induced cells are spread in culture media containing X-gal and IPTG, β-galactosidase is expressed and X-gal is hydrolyzed to galactose (1) and 5-bromo-4-chloro-3-indole (2), which further dimerizes into (3) in the presence of oxygen, forming an intense blue precipitate as shown in Scheme 5.1. The intensity of the blue color is correlated with the level of expression of LacZ.[100]

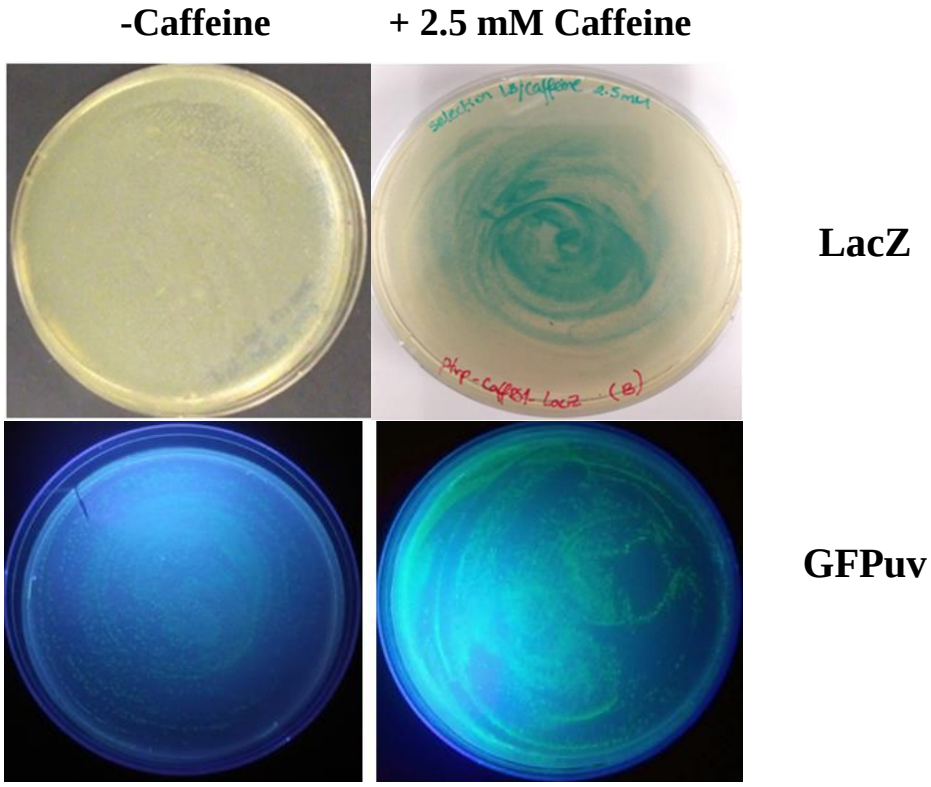
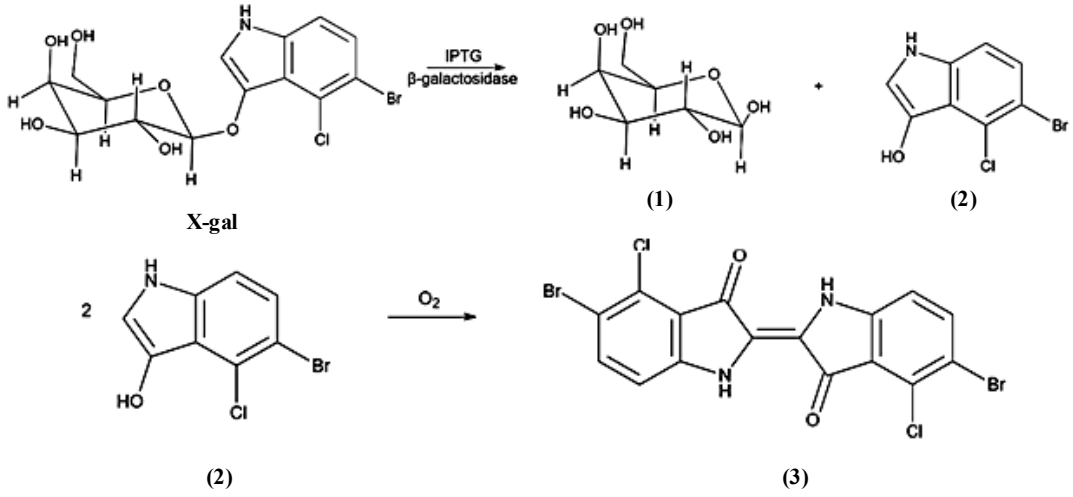


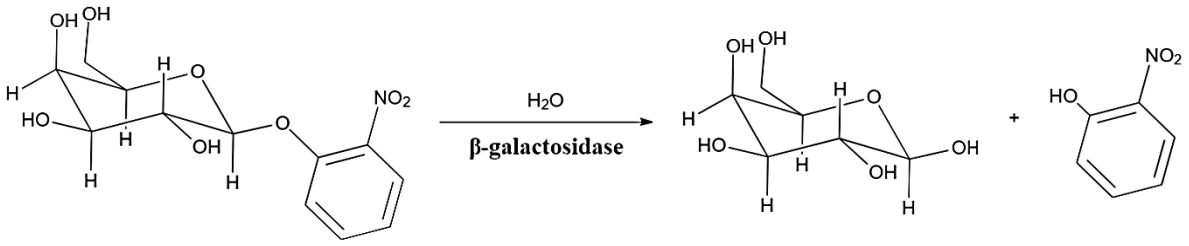

**Figure 5.4** Caffeine induced expression of LacZ (top) and GFPuv (bottom) in *XL1-Blue E. coli* cells.




**Scheme 5.1** Hydrolysis of X-gal by β-galactosidase.

### 5.4 β-galactosidase assay

To determine the concentration-signal dependency and specificity of CaffRS respect to caffeine, a β-galactosidase assay was performed. Both pTrp-CaffRS-LacZ and pTrp-TheoRS-LacZ were used to transform XL1-Blue *E. coli* cells and the cells were grown overnight in the presence of increasing concentrations of caffeine and theophylline within 0-5 mM range.




**Scheme 5.2** Hydrolysis of ONPG by β-galactosidase.

For the β-galactosidase assay, cells are lysed using sonication and the lysate is reacted with o-nitrophenyl-β-D-galactopyranoside (ONPG), a galactose analog that colors the solution yellow by the release of o-nitrophenol (Scheme 5.2). The time of the reaction is determined from the moment that the substrate is added to the development of yellow color, in which sodium carbonate or bicarbonate is added to stop the reaction.[96, 101] The absorbance of the solution at 420 nm is determined and the Miller units are calculated using the formula:

$$\text{M.U} = \frac{A_{420\text{nm}} \times 1000}{OD_{600} \times t \times V}$$

Where:

MU: Miller units

A420nm: Absorbance of the samples at 420nm.

OD600: Optical density of the cell culture

t: time in minutes

V: volume in mL.

As expected, the cells containing CaffRS showed an increase in the β-galactosidase activity proportional to the concentration from 0-5 mM (Figure 5.5) and the increase in the activity corresponded to 3-fold when the concentration of caffeine was 5 mM respect to no caffeine in solution.[31]

When cells containing CaffRS were grown in the presence of theophylline, the LacZ assay showed a proportional increase in the β-galactosidase activity as the concentration of theophylline increases. This result suggests an expansion of the selectivity rather than a change in the selectivity, probably due to the small difference between caffeine and theophylline (theophylline is smaller by one methyl group at nitrogen atom N-7 than caffeine).

In the case of TheoRS, our findings are consistent with the specificity of TheoRS reported in the literature.[89] When comparing both riboswitches, CaffRS showed a higher background signal when no caffeine was present, which suggests that CaffRS has the equilibrium shifted further than TheoRS towards its ligand-free "ON" structure from its ligand-free "OFF" conformation where the RBS is exposed or sequestered, respectively.[89]

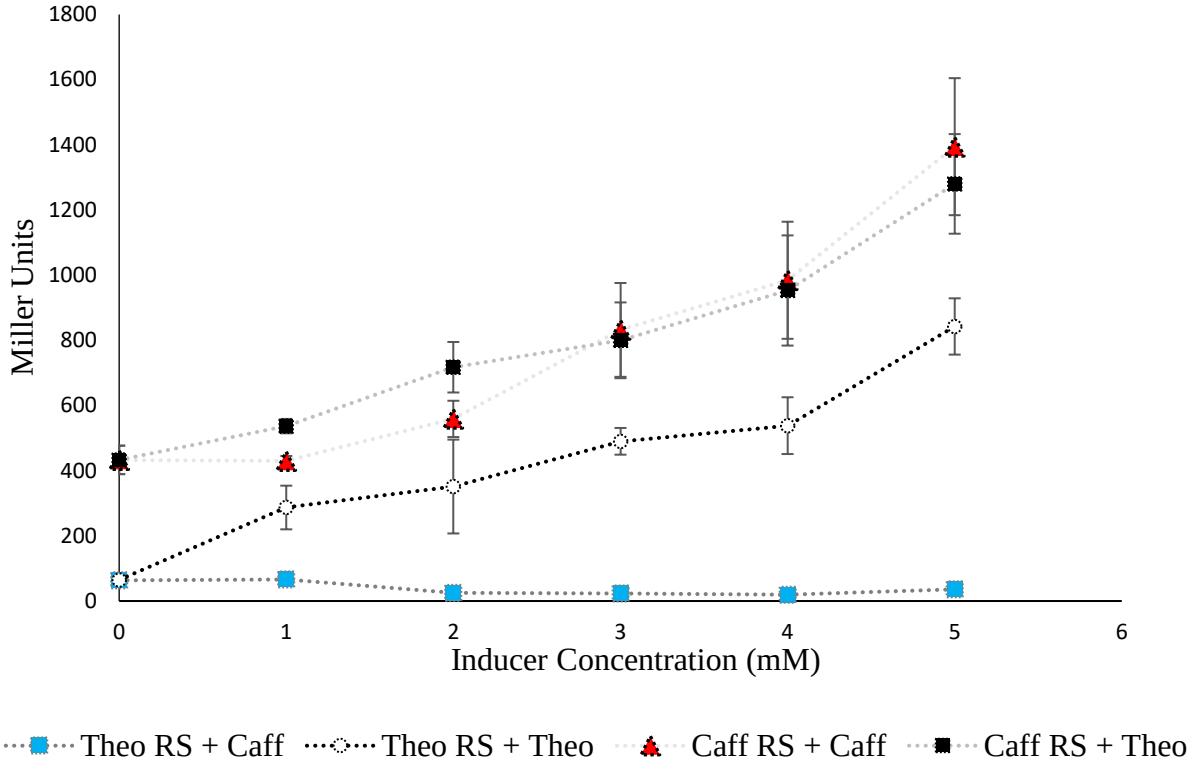



**Figure 5.5** β-galactosidase activities of CaffRS and TheoRS in the presence of 0-5 mM of caffeine and theophylline. Values represent the mean of three independent measurements and their respective standard deviations (s.d).

We used RNA structure 6.3, a 2D RNA structure software, to predict the secondary structures of CaffRS and TheoRS (Figure 5.6). TheoRS has six ligand-free conformations with similar energies, but only one exposes the RBS. However, CaffRS has four different conformations with similar energies and two of them have exposed RBS.

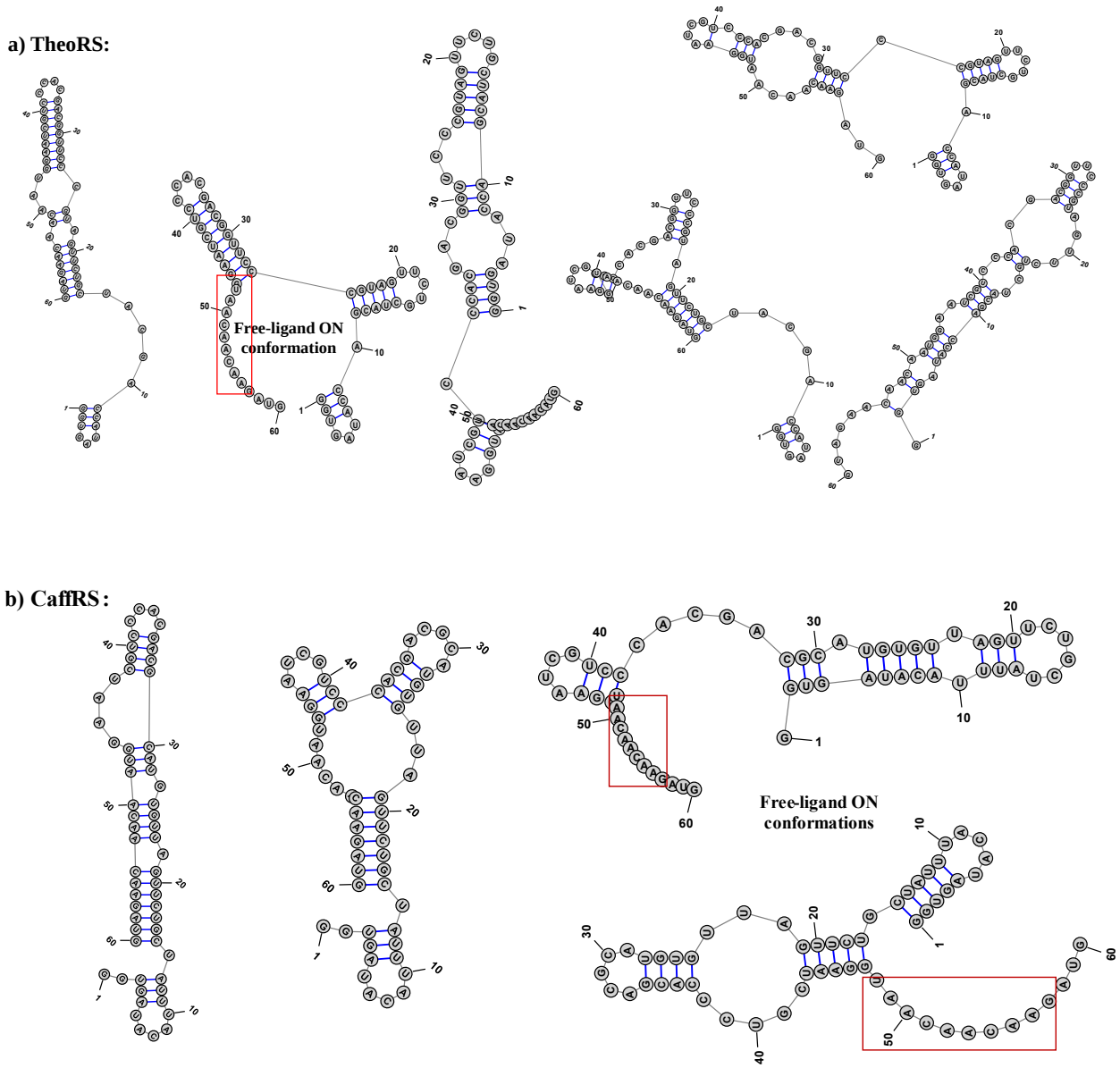



**Figure 5.6** 2D structures determined for a) TheoRS and b) CaffRS in ligand free state using RNAStructure 6.3.[95] Red box represents the beginning of the RBS.

## 5.5 Conclusions

We demonstrated that *cat-upp* fusion gene can be used to select synthetic riboswitches *in vivo*. CaffRS enhanced the expression of GFPuv and LacZ as well as expanded selectivity for theophylline and caffeine. This simple cell viability-based selection could be useful to reengineer

existing riboswitches and to select new synthetic riboswitches, in addition to the use of exogenous substances to control gene expression.

# CHAPTER 6

## DUAL GENETIC SELECTION OF A SARCOSINE SPECIFIC APTAMER FROM THE GLYCINE RIBOSWITCH

In 2004, Breaker et al. identified a glycine-sensing RNA in *Vibrio cholerae* and *Bacillus subtilis* responsible for the regulation of a three gene operon (*gcvT, gcvPA, gcvPB*) that controls the use of glycine as carbon source. The glycine riboswitch is composed of two different aptamers (type I and type II) adjacent to each other and followed by a single expression platform.[37, 102] When glycine binds to the glycine aptamer, the anti-terminator loop is formed to allow transcription to occur. Although it was initially suggested that both aptamers were necessary for binding, the new evidence indicated that both tandem and single aptamers are found in glycine regulation systems and they bind with similar affinity to glycine.[103-105] Since aptamer type II is directly located upstream from the expression platform without the need to have a ghost aptamer that is usually needed to stabilize the structure of aptamer type I, we decided to create a library of glycine aptamers from the *B. subtilis'* glycine riboswitch type II aptamer to select synthetic riboswitches that bind to small molecules of biological relevance, including sarcosine and other glycine analogs. [105]

Sarcosine or N-methyl glycine is a naturally occurring metabolite synthesized by the enzyme glycine N-methyl transferase (GNMT), which is over-expressed in prostate cancer tissues (PCa) and indicates poor prognosis.[106, 107] Elevated levels of sarcosine in urine samples have been suggested as a non-invasive method to diagnose prostate cancer. Therefore, we intended to select synthetic sarcosine riboswitches as a potential biosensor for sarcosine.

## 6.1 Sarcosine aptamer library design

Applying the same principle from the previous chapter, we randomized seventeen positions at the glycine binding site (*gcvT*) using polymerase chain reaction (PCR) to produce a gcvT aptamer library. The library was inserted into the pTrp-gcvT-CatUpp plasmid (Figure 6.1).

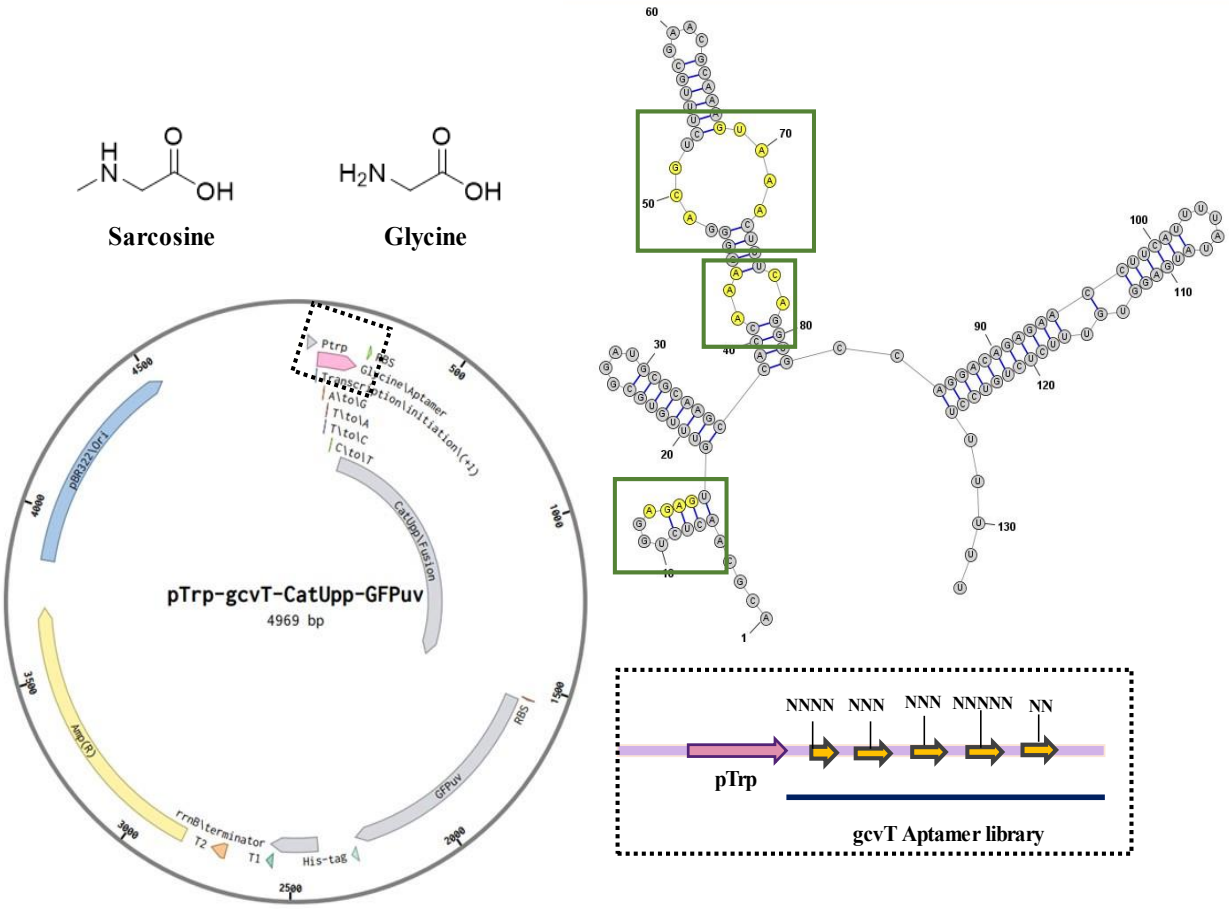



**Figure 6.1** Strategy to design the gcvT library. Positions 13-16, 42-44, 49-51, 68-72 and 77-78 were randomized using PCR. RNA 2D structures were modeled using RNAStructure 6.3 software

### 6.2 Dual genetic screening of sarcosine riboswitches

The selection scheme is outlined in Figure 6.2. Chemically defined rich (CDR) media was used to select aptamers against sarcosine in the presence of 30 µg/mL chloramphenicol for the positive selection and 0.5 µg/mL 5-FU for the negative selection.

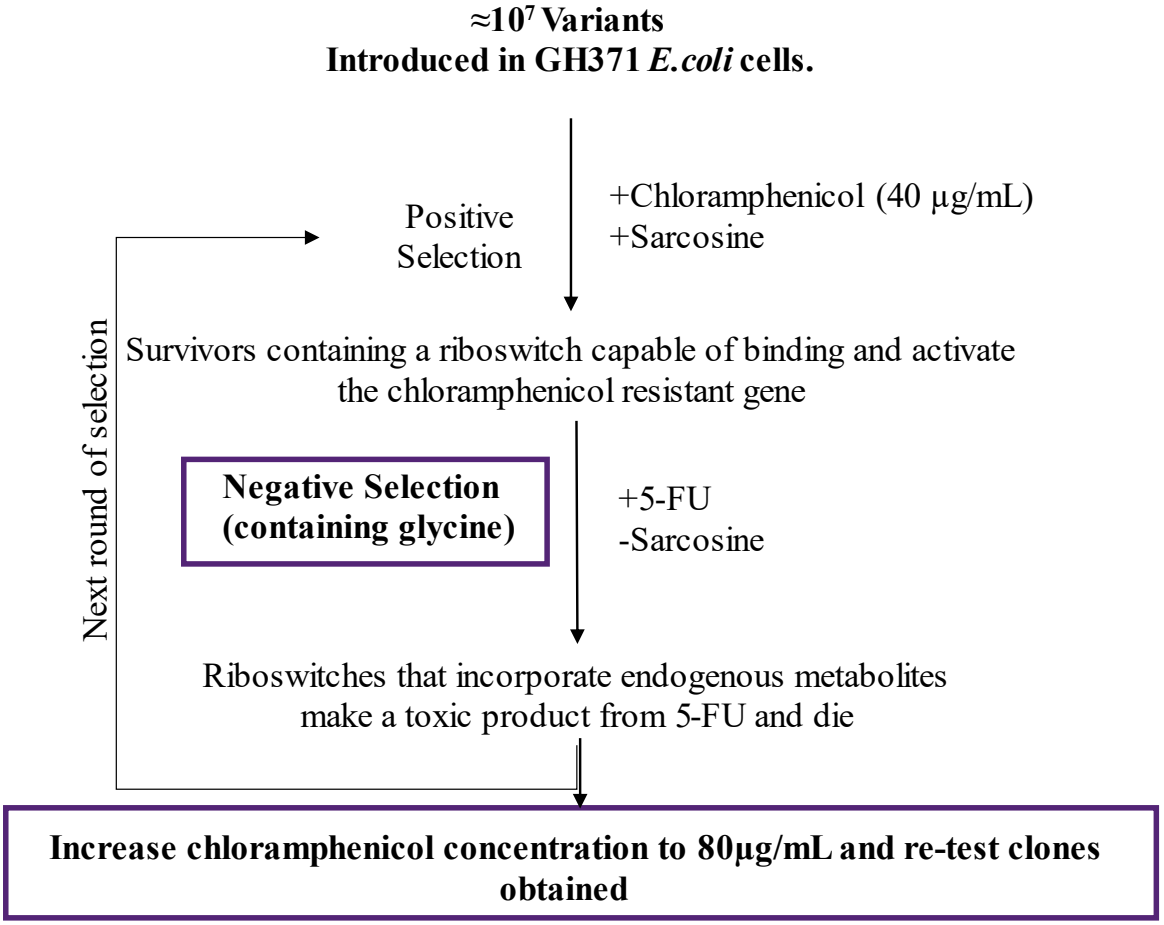



**Figure 6.2** Dual-genetic screening used to select riboswitches against sarcosine (5 mM).

It should be noted that glycine is a natural amino acid, and it is present in media during both positive and negative selection steps. The cells containing any aptamer that binds to glycine would die during the negative selection process, enriching the riboswitches that are only

activated by sarcosine. Three rounds of selection were performed, and 192 clones were re-tested at 40 µg/mL of chloramphenicol.

Initially, no difference in cell growth was observed for the chloramphenicol plates with and without sarcosine, possibly due to higher chloramphenicol tolerance by the gcvT riboswitches compared to the TheoRS. Therefore, the chloramphenicol tolerance of the library was tested again after the third round of selection and colonies were able to survive up to 80 µg/mL chloramphenicol (Figure 6.3). Sixty clones were re-tested under these conditions, and twenty-four colonies survived at 80 µg/mL of chloramphenicol when 5 mM of sarcosine was added. As a result of the enrichment process, several clones had the same sequence, as shown in Table 6.1 and an extra base was introduced to clone D4 yielding eleven different variants.

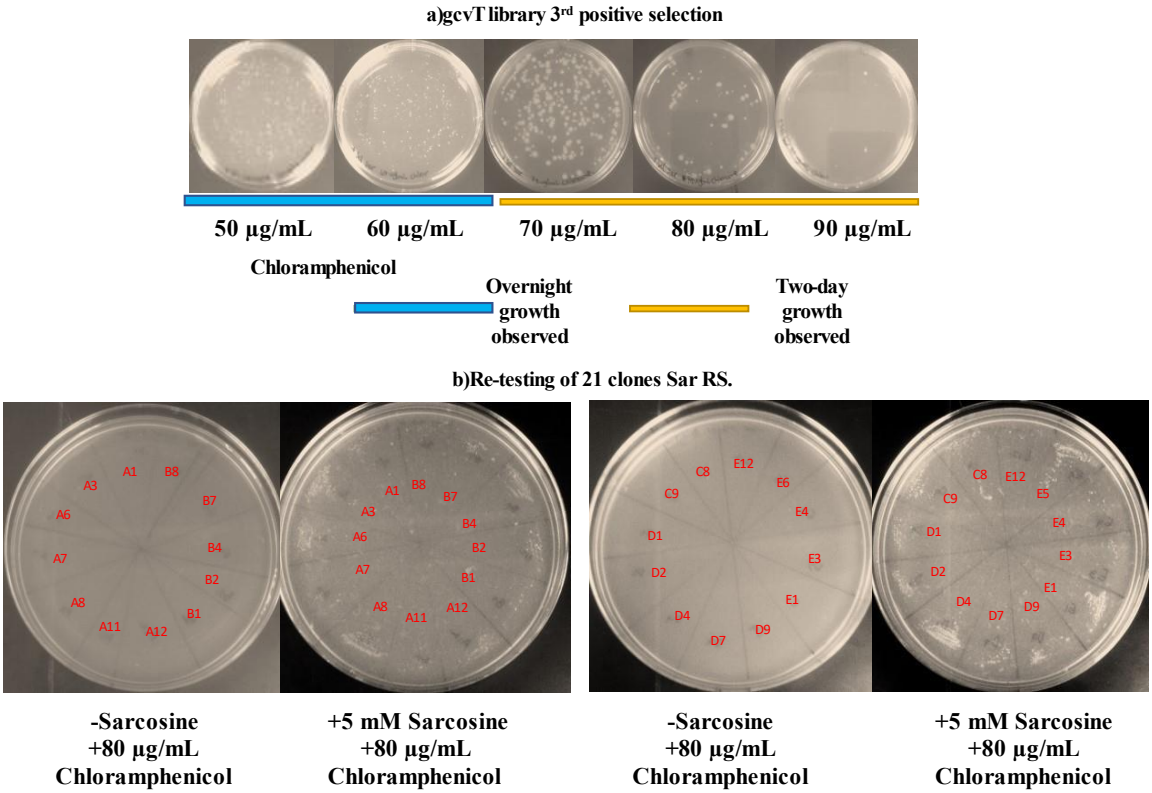



**Figure 6.3** a)chloramphenicol tolerance test of gcvT library after 3rd selection. b)re-testing of the 3rd positive selection of the gcvT library with higher chloramphenicol concentrations and 5 mM sarcosine.

**Table 6.1** DNA sequences of the sarcosine specific riboswitches derived from the glycine riboswitch (gcvT). Bases underlined represent positions mutated, while highlighted bases represent changes in sequence within the intended mutations.

| Mutant | Sequence 5' to 3' |
|---|---|
| WT (*gcvT*) | ACG CAA CTC TGG AGA GTG TTT GTG CGG ATG CGC AAG CCA CCA AAG<br>GGG ACG TCT TTG CGT ATG CAA AGT AAA CTT TCA GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| A1 | ACG CAA CTC TGG **GA**A GTG TTT GTG CGG ATG CGC AAA CCA CCA **TG**G<br>GGG A**AA** TCT TTG CGT ATG CAA A**AC** A**CT** CTT TCA GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| A7, B1, D1, D7, D9, E12 | ACG CAA CTC TGG AGA GTG TTT GTG CGG ATG CGC AAA CCA CCA **TC**G<br>GGG A**AA** TCT TTG CGT ATG CAA AG**G TGG** CTT T**TT** GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| A8 | ACG CAA CTC TGG AGA **T**TG TTT GTG CGG ATG CGC AAA CCA CC**T G**AG<br>GGG **GTA** TCT TTG CGT ATG CAA A**TA C**AA CTT TCA GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| C9 | ACG CAA CTC TGG **TTT T**TG TTT GTG CGG ATG CGC AAA CCA CCA **C**AG<br>GGG A**AA** TCT TTG CGT ATG CAA AGT **TAC** CTT T**GT** GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| A6 | ACG CAA CTC TGG AG**T T**TG TTT GTG CGG ATG CGC AAA CCA CC**G TA**G<br>GGG **C**CG TCT TTG CGT ATG CAA AG**A TCG** CTT TCA GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| D4 | ACG CAA CTC TGG **TCC T**TG TTT GTG CGG ATG CGC AAA CCA CC**C CT**G<br>GGG **CCA** TCT TTG CGT ATG CA**C CAG GCG** CCT T**TG** CGG TGC CAG GAC<br>AGA GAA CCT TCA TTT TAC ATG AGG TGT TTC TCT GTC CTT TTT T |

| | |
|---|---|
| A3 | ACG CAA CTC TGG **TTT C**TG TTT GTG CGG ATG CGC AAA CCA CC**C** AAG<br>GGG A**A**G TCT TTG CGT ATG CAA AG**C TG**A CTT T**GG** GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| B7 | ACG CAA CTC TGG **C**GA GTG TTT GTG CGG ATG CGC AAA CCA CC**G GG**G<br>GGG A**TT** TCT TTG CGT ATG CAA A**AG TGC** CTT T**TT** GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| A11, B4 | ACG CAA CTC TGG **GC**A **A**TG TTT GTG CGG ATG CGC AAA CCA CC**T G**AG<br>GGG **TAC** TCT TTG CGT ATG CAA AGT AAA CTT TCA GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| A12, B8 | ACG CAA CTC TGG **TA**A GTG TTT GTG CGG ATG CGC AAA CCA CC**G** A**C**G<br>GGG A**AA** TCT TTG CGT ATG CAA AG**C** A**GC** CTT TCA GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |
| C8 | ACG CAA CTC TGG AG**T T**TG TTT GTG CGG ATG CGC AAA CCA CC**G GG**G<br>GGG **TAT** TCT TTG CGT ATG CAA A**TT AT**A CTT T**CC** GGT GCC AGG ACA<br>GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT |

### **6.**3 **Sarcosine induced expression of β-galactosidase (LacZ)**

The selected gcvT riboswitch variants were introduced into pTrp-gcvT-LacZ between the *SpeI* and *NcoI* restriction sites to produce the pTrp-SarRS-LacZ plasmid. The resulting plasmids were used to transform *XL1-Blue E. coli* cells, spread on LB plates containing Xgal/IPTG/sarcosine, and blue colonies were isolated. β-galactosidase assay was performed for the cells transformed with different clones of SarRS. Figure 6.4 shows the results obtained when all the variants expressed β-galactosidase with different sarcosine concentrations.

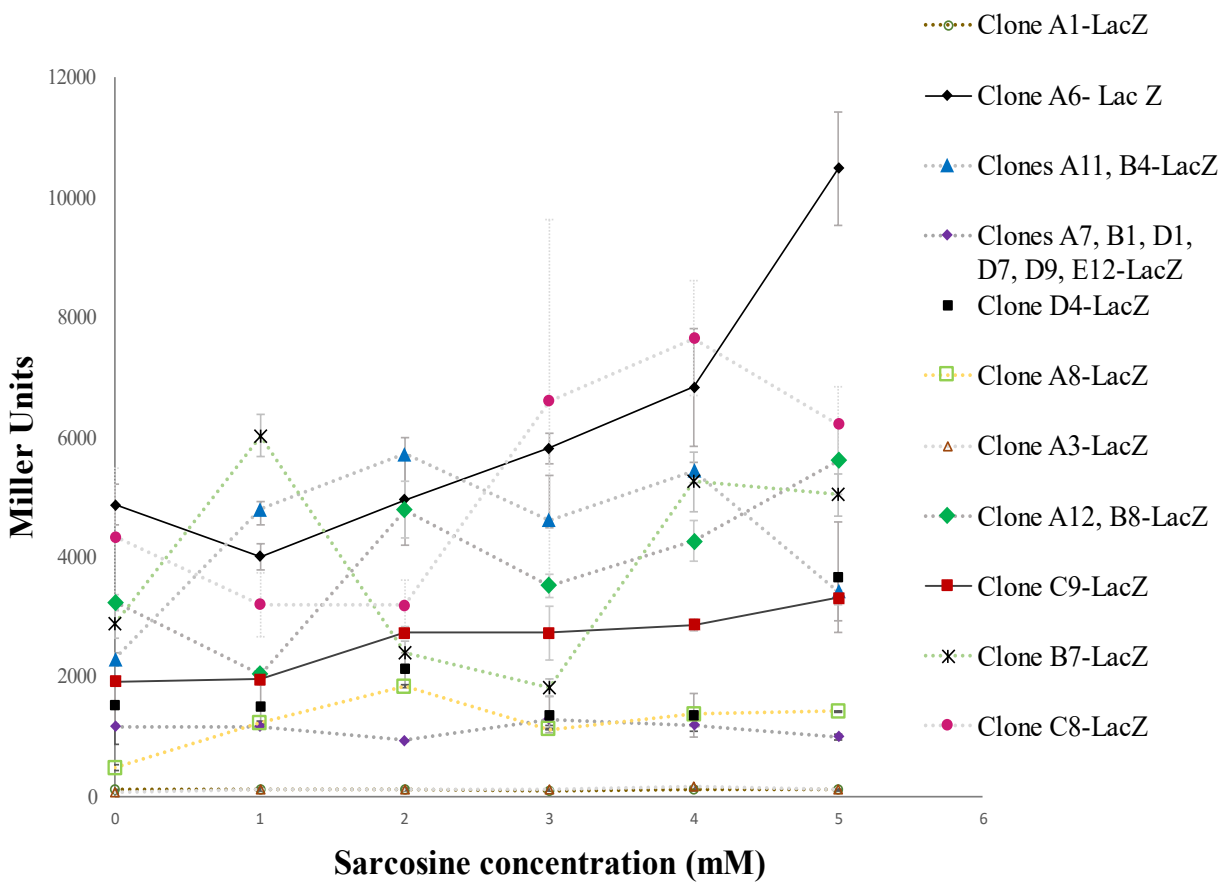



**Figure 6.4** β-Galactosidase activities for SarRS clones in the presence of 0-5 mM of sarcosine. Measures represent three biological replicates with their respective standard deviation (s.d). Analyzed with Microsoft excel.

From all the clones tested, A3, A1, A7(B1, D1, D7, D9, E12) and A8 showed slight increase in the β-galactosidase activities within their determined standard deviations compared with the rest of the clones which indicated a low level of gene activation. Interestingly, some of these clones were highly enriched within the pool selected, suggesting that they were likely a consequence of the selection when the lower concentration of chloramphenicol was used.

Clones D4, A12 (B8) and B7 did not show a concentration-dependent increase in the β-galactosidase activities. Although more information is needed to justify this behavior, some of

the possible reasons can be multiple specificity (competitivity), low affinity, complex binding mechanisms, and even errors associated with the assay performed.

Clone C8 exhibited promising results, with β-galactosidase activities about 1.5-fold with respect to the background signal, however the values obtained were not proportional to the increase in the concentrations. It is likely that this riboswitch is activated by sarcosine, but the binding mechanism may be different, and some cooperation may be required.

Two clones showed a concentration-dependent response within their respective standard deviations and are likely to be specific for sarcosine only: A6 and C9. In both cases the signal was increased about 1.5 and 2-folds when the concentration of ligand went from 0 to 5 mM.

Although the background signal found at 0 mM is attributed to the presence of a possible ON free ligand conformation, when studying the 2D possible structures of the RNA (Figure 6.5), no conformations were found exposing the anti-terminator loop, even for the glycine riboswitch that previously showed low background signal when no glycine is added.[37, 108] Lipfert et al. suggested the existence of a transition state between unbound (U) and glycine-bound (B) called M, in which magnesium plays a role in the formation of a less stable but possible structure in *Vibrio cholerae* glycine aptamers, specially Type II. The small angle X-ray scattering (SAXS) data suggests that the riboswitch is partially folded at millimolar concentrations of $Mg^{2+}$ and that those contacts are further enhanced when glycine is added.[109, 110] Additionally, the riboswitch might still be reacting with glycine from the culture media when no sarcosine is present, but once it is added the specificity may shift towards it. *In vitro* assays may be necessary to determine the selectivity of SarRS.

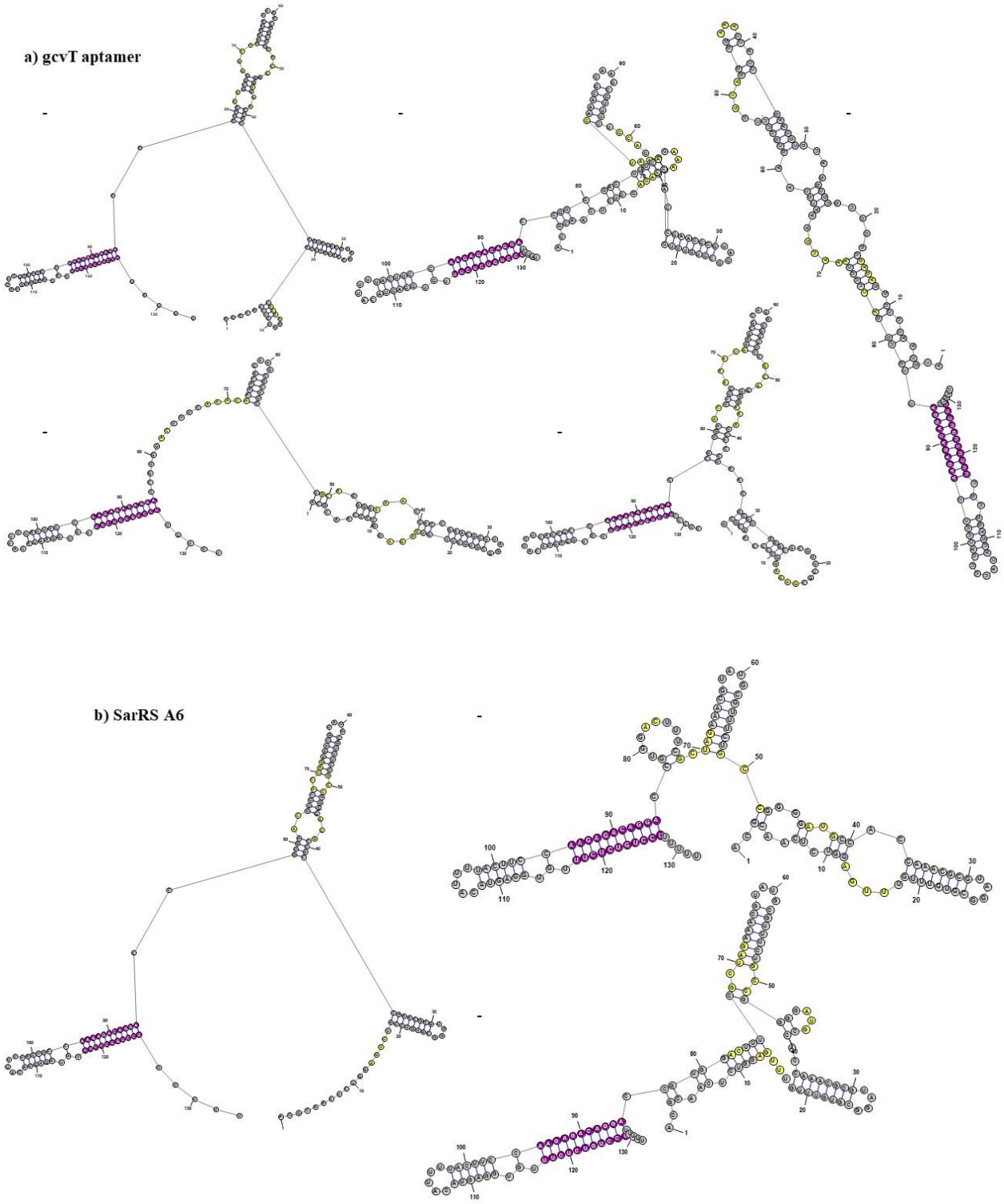



**Figure 6.5** 2D structures determined for a)glycine RS and b)sarcosine RS A6 and c)sarcosine RS C9 in ligand free state using RNAStructure 6.3.Yellow bases denote randomized positions, purple bases represent the formation of the terminator loop (OFF conformation)[95]

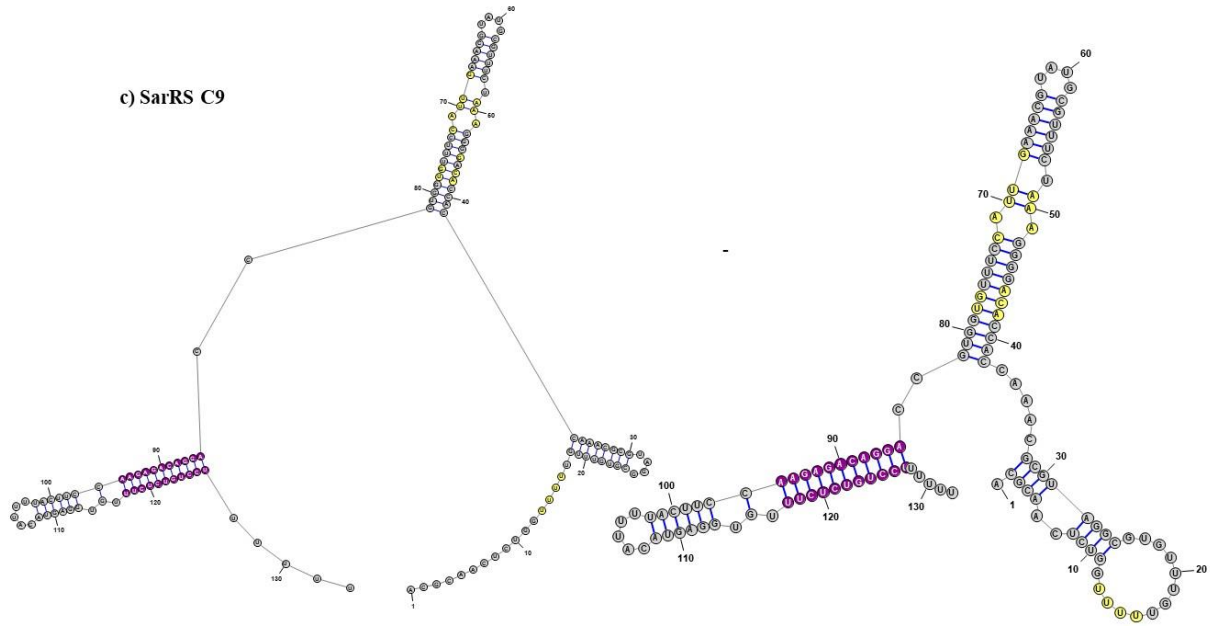



**Figure 6.5 (Continued)** 2D structures determined for a)glycine RS and b)sarcosine RS A6 and c)sarcosine RS C9 in ligand free state using RNAStructure 6.3.Yellow bases denote randomized positions, purple bases represent the formation of the terminator loop (OFF conformation)[95]

### 6.4 *In vitro* transcription of sarcosine riboswitches

Sarcosine riboswitches A6 and C9 were amplified by PCR and the T7 promoter was added at the upstream of the riboswitch sequence. The resulting DNA was inserted between *EcoRI* and *HindIII* restriction sites of the pUC-19 plasmid. The DNA was amplified to generate a linear 166bp DNA (Figure 6.6a). *In vitro* transcription was performed using T7 RNA polymerase. The transcripts were analyzed with TBE-Urea gel electrophoresis (Figure 6.6b), visualized through UV-shading, extracted with n-butanol, and purified using ethanol precipitation.

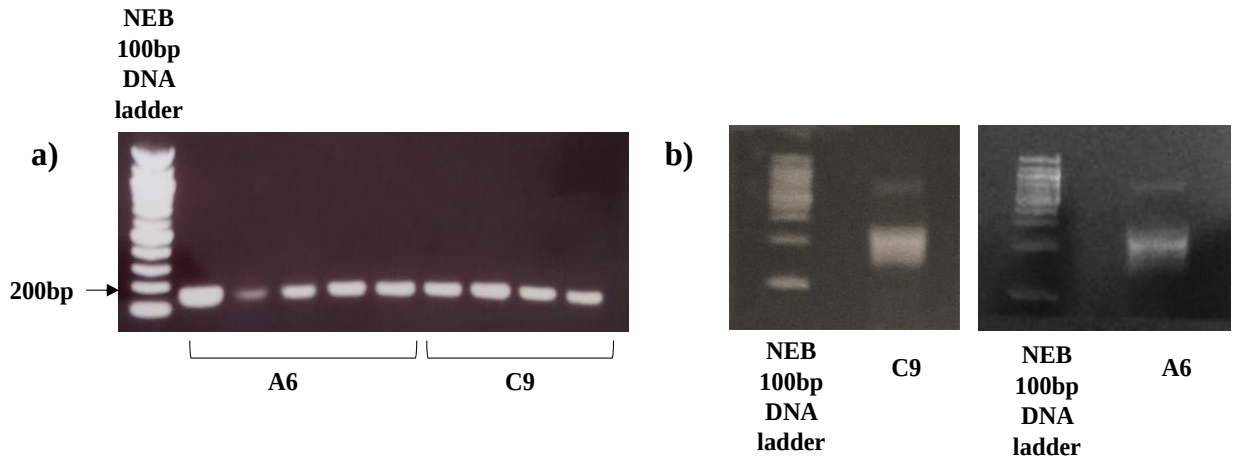



**Figure 6.6** a)amplification of linear DNA of T7SarRS clones A6 and C9. b)*in vitro* transcription of SarRS clones A6 and C9. NEB DNA 100bp ladder used as a reference.

### 6.5 Sarcosine riboswitch binding studies by ITC

The RNA was allowed to fold at 60°C for five minutes, cooled down on ice, and dialyzed overnight using a 3.5 kDa cut-off dialyzer tube with a buffer containing 4-(2-hydroxyethyl)-1-piperazine ethane sulphonic acid (HEPES), NaCl and $MgCl_2$ as suggested by the literature.[111] Several trials using RNA concentrations in the micromolar range were done, but the ITC signals obtained were not strong enough for binding studies. The amount of RNA in millimolar concentrations required for the ITC measurements is too large to be produced. Also, because the glycine riboswitch is $Mg^{2+}$ dependent for binding, the 10 mM $Mg^{2+}$ concentration used may have not been enough to generate a well-resolved ITC curve. As suggested by some studies, we plan to use 40 mM of $MgCl_2$.[46]

### 6.6 Conclusions and future directions

We used *in vivo* selection to generate two riboswitches that respond to the increment in concentrations of sarcosine and that can be potentially used to sense sarcosine in biological samples. To evaluate their specificity and binding, it is necessary to use an *in vitro* assay. We

are currently studying the use of circular dichroism as an alternative to study binding of SarRS A6 and C9 when titrated with different concentrations of sarcosine and glycine. We also used the library to generate synthetic riboswitches for creatine, another glycine analog. The variants obtained are currently being sequenced and their β-galactosidase activity is being determined.

## EXPERIMENTAL SECTION

### 1. Synthesis of Dansyl-Dap

2-Amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid was synthesized according to the protocol published by Summerer et al.[15] Dansyl chloride and triethylamine were purchased from Sigma, and BOC-Dap-OH from Fluka. $^{1}$H-NMR and $^{13}$C-NMR were recorded on a 400 MHz Bruker Avance spectrometer at 400 and 101MHz respectively. The spectra are shown on Figures S1 and S2.

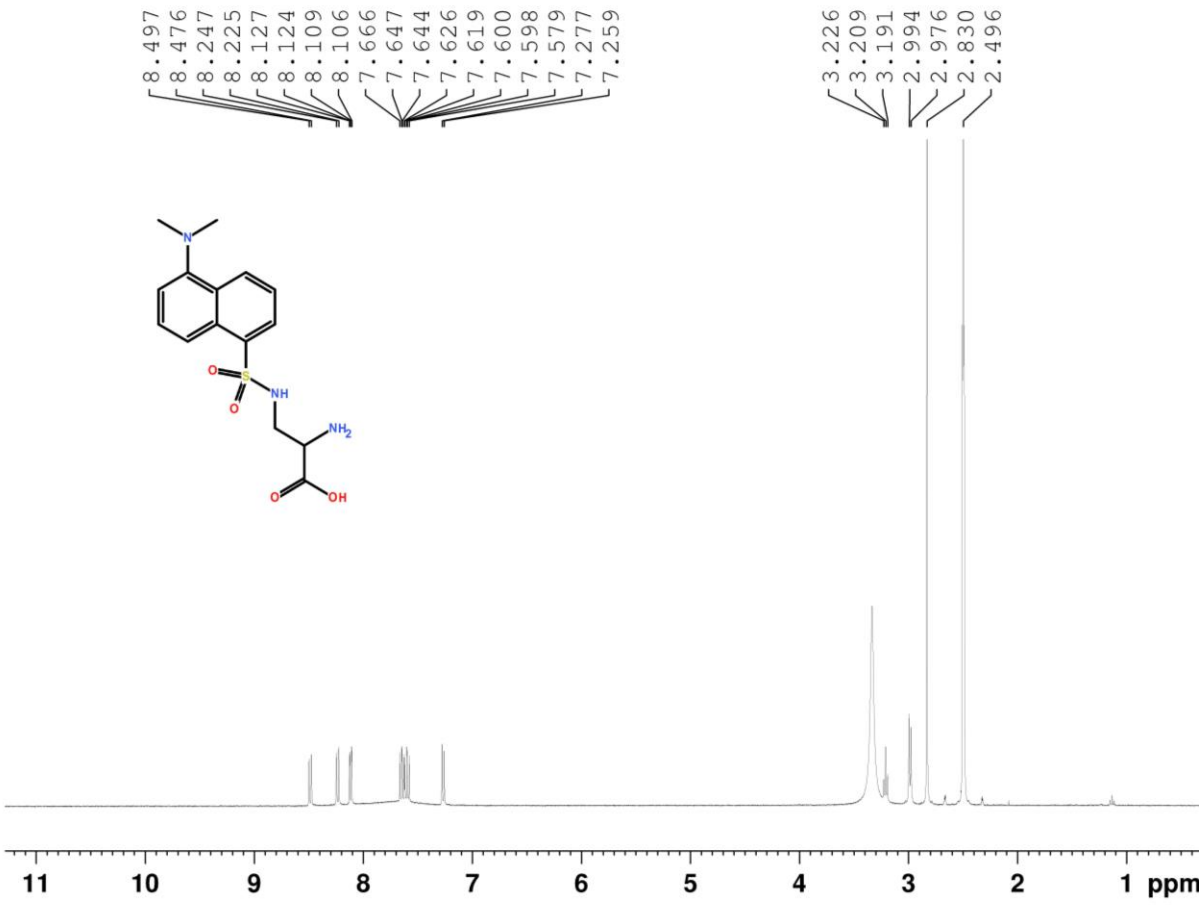



**Figure S. 1** $^{1}$H spectrum of synthesized Dansyl-Dap in DMSO-d$^{6}$.

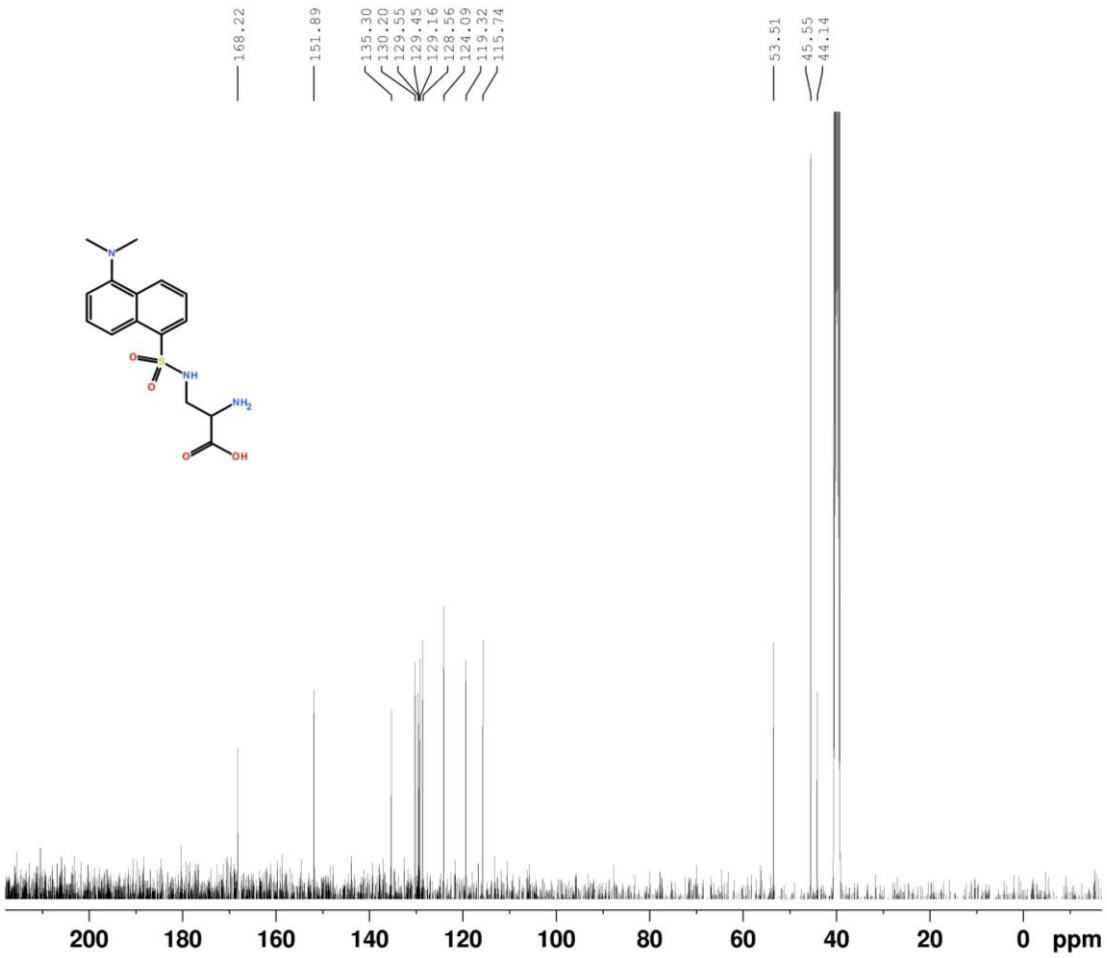



**Figure S. 2** $^{13}$C NMR spectrum of synthesized Dansyl-Dap in DMSO-d$^{6}$.

## 2. General biological procedures

PCR was performed with 10 ng of DNA template, 1X iproof reaction buffer, 10 nmol dNTPs, 1U iproof DNA polymerase (Bio-Rad), and 25 pmol primers in 50 µL of total volume. Thermal conditions depended on the size and primers used and will be described in the subsequent sections.

In every cloning step, DNA size was confirmed using 1% agarose electrophoresis. After PCR, DNA was purified using QIA quick PCR purification kit (Qiagen). Restricted fragments were purified using QIAquick gel extraction kit (Qiagen).

Ligation was performed with T4 DNA ligase (NEB). DNA was purified using Qiagen minelute kit.

Sequencing was done using BigDye terminator sequencing kit in 20 µL total volume: 200 ng DNA, 3.2 pmol primer, 1.75 µL Big Dye sequencing buffer. DNA was incubated in compel beads in a glycol solution (BET), magnetized and washed twice with 70% ethanol and eluted with 40 µL of $ddH_2O$. Sequence was analyzed with ABI 3130 Genetic analyzer, SeqA6 (ThermoFisher) and Benchling.

Plasmid DNA was isolated (miniprepped) by resuspending cells in Qiagen P1 buffer, lysed with Qiagen P2 buffer, and neutralized with Qiagen N3 buffer. Lysate was centrifuged at 13500 rpm for 10 min to separate the bacterial debris. Supernatant was purified using QIAprep Spin DNA miniprep kit (Qiagen).

**3. Construction of the MtLRS ANIGT, GNIGG, GLVGE**

The fragment containing N-terminal mutations was design in two steps:

1. Fragment N1 (245 bp) and N2 (599 bp) were made with pSupk-MtLRS-HL(TAG)3-ΔCP1 as a template, using the primers shown in Table E.1. Thermal conditions are shown in Table E.2.

2.The N-terminal fragment (823 bp) was assembled using equimolar amounts of fragments N1 and N2 (40 ng DNA total), MLRS_F1 F and MLRS PCR *MluI* R primers, and the components described in section 2. Thermal conditions are shown in Table E.3.

**Table E.1** Primers used to introduce N-terminal mutations on MtLRS ANIGT, GNIGG and GLVGE mutants.

| Reaction | Mutant | Primer forward | Primer Reverse |
|---|---|---|---|
| N1 | ANIGT | MLRS_F1 F: 5'-CCC GGT CAT CAA TCA TCC CC-3' | MLRS A32N33 OE_PCR R: 5'-GCA TCG CAC CAC TGG GGT AGG GGT TCG CGA CTG TGA GGA ATA TCT TTT CTC TGT C-3' |
| N1 | GNIGG | MLRS_F1 F (see ANIGT) OE | MLRS G32N33 OE_PCR R: 5'-GCA TCG CAC CAC TGG GGT AGG GGT TGC CGA CTG TGA GGA ATA TCT TTT CTC TGT C-3' |
| N1 | GLVGE | MLRS_F1 F (see ANIGT) | MLRS G32L33 OE_PCR R: 5'- GCA TGC CAC CAC TGG GGT AGG GCA GGC CGA CTG TGA GGA ATA TCT TTT CTC TGT C-3' |
| N2 | All clones | MLRS Y33X OEPCR F: 5'-CCC TAC CCC AGT GGT GCG ATG C-3' | MLRS *MluI* R: 5'-GGG CAG ACG CGT TCC AAG GC-3' |

**Table E.2** Thermal conditions used for the N-terminal reactions 1 and 2 during the design of the MtLRS ANIGT, GNIGG and GLVGE mutants.

| Step | Temperature (°C) | Time | |
|---|---|---|---|
| 1 | 98 | 30s | Initial denaturation |
| 2 | 98 | 5s | Denaturation |

| | | | |
|---|---|---|---|
| 3 | 50 | 10s | Annealing |
| 4 | 72 | 15s | Extension, go to step 2 34 times |
| 5 | 72 | 5min | Final extension |

**Table E.3** Thermal conditions used for the N-terminal fragment assembly during the design of the MtLRS ANIGT, GNIGG and GLVGE mutants.

| **Step** | **Temperature (°C)** | **Time** | |
|---|---|---|---|
| 1 | 98 | 30s | Initial denaturation |
| 2 | 98 | 10s | Denaturation |
| 3 | 60 | 30s | Annealing |
| 4 | 72 | 30s | Extension, go to step 2, 34 times |
| 5 | 72 | 5min | Final extension |

For C-terminal mutations, no template was used. Instead, two fragments (C1 and C2) were made with equimolar amounts of the primers shown in Tables E.4-E.6 (for a total of 50 pmol) and the components in section 2.

**Table E.4** Primers used to introduce C-terminal mutations on MtLRS ANIGT mutant.

| **Mutant** | **Reaction** | **Primers** |
|---|---|---|
| ANIGT | C1 | MLRS OE_PCR *MluI* P1: 5'-CTT GGA ACG CGT CTG CCC TGG GAT GAG AGG TGG ATC ATC GAA CCC CTC ACA GAC TC-3' |
| | | MLRS OE_PCR I516 P2: 5'-CAT CGA ACC CCT CAC AGA CTC AAC AAT CAT TAT GGC ATA TTA CAC CAT CGC A-3' |

| | | |
|---|---|---|
| | | MLRS lib P3: 5’-CAA AGA ACT CAT CGT CCA TCT CCC CGG CAT CCA TCT CCC TGA GGC GGT GTG CGA TGG TGT AAT ATG CC-3’ |
| | | MLRS lib P4: 5’-CCG GAA TTC CTC CCT GAG ATC CTC AAA GGT TCC TGA ATC ATC TAG GAA TAT GGC ATC AAA GAA CTC ATC GTC CAT CTC-3’ |
| | C2 | MLRS OE_PCR G566 P5: 5’-GGA TCT CAG GGA GGA ATT CCG GTA CTG GTA CCC CCT TGA CTG GGG CCT CTC TGC AAA-3’ |
| | | MLRS OE_PCR T576 P6: 5’-CTC TGC AAA GGA CCT CAT AGG CAA TAC CCT GAC ATT CCA TAT ATT CCA CCA CTC AGC-3’ |
| | | MLRS lib P7: 5’- CCA CAG CAC CCC GTG GCC ACC CTG ACT CAG GGA ATA TGG CTG AGT GGT GG-3’ |
| | | MLRS lib P8: 5’-CCA CAG CAC CCC GTG GCC ACC C-3’ |

**Table E.5** Primers used to introduce C-terminal mutations on MtLRS GNIGG mutant.

| **Mutant** | **Reaction** | **Primers** |
|---|---|---|
| GNIGG | C1 | MLRS OE_PCR *MluI* P1: 5’-CTT GGA ACG CGT CTG CCC TGG GAT GAG AGG TGG ATC ATC GAA CCC CTC ACA GAC TC-3’ |
| | | MLRS OE_PCR I516 P2: 5’-CAT CGA ACC CCT CAC AGA CTC AAC AAT CAT TAT GGC ATA TTA CAC CAT CGC A-3’ |
| | | MLRS lib P3: 5’-CAA AGA ACT CAT CGT CCA TCT CCC CGG CAT CCA TCT CCC TGA GGC GGT GTG CGA TGG TGT AAT ATG CC-3’ |
| | | MLRS lib P4: 5’-CCG GAA TTC CTC CCT GAG ATC CTC AAA GGT TCC TGA ATC ATC TAG GAA TAT GGC ATC AAA GAA CTC ATC |

| | | |
|---|---|---|
| | | GTC CAT CTC-3’ |
| | C2 | MLRS OE_PCR G566 P5: 5’-GGA TCT CAG GGA GGA ATT CCG GTA CTG GTA CCC CCT TGA CTG GGG CCT CTC TGC AAA-3’ |
| | | MLRS OE_PCR G576 P6: 5’-CTC TGC AAA GGA CCT CAT AGG CAA TGG CCT GAC ATT CCA TAT ATT CCA CCA CTC AGC-3’ |
| | | MLRS lib P7: 5’- CCA CAG CAC CCC GTG GCC ACC CTG ACT CAG GGA ATA TGG CTG AGT GGT GG-3’ |
| | | MLRS lib P8: 5’-CCA CAG CAC CCC GTG GCC ACC C-3’ |

**Table E.6** Primers used to introduce C-terminal mutations on MtLRS GLVGE mutant.

| **Mutant** | **Reaction** | **Primers** |
|---|---|---|
| GLVGE | C1 | MLRS OE_PCR *MluI* P1: 5’-CTT GGA ACG CGT CTG CCC TGG GAT GAG AGG TGG ATC ATC GAA CCC CTC ACA GAC TC-3’ |
| | | MLRS OE_PCR V516 P2: 5’-CAT CGA ACC CCT CAC AGA CTC AAC AAT CGT GAT GGC ATA TTA CAC CAT CGC A-3’ |
| | | MLRS lib P3: 5’-CAA AGA ACT CAT CGT CCA TCT CCC CGG CAT CCA TCT CCC TGA GGC GGT GTG CGA TGG TGT AAT ATG CC-3’ |
| | | MLRS lib P4: 5’-CCG GAA TTC CTC CCT GAG ATC CTC AAA GGT TCC TGA ATC ATC TAG GAA TAT GGC ATC AAA GAA CTC ATC GTC CAT CTC-3’ |
| | C2 | MLRS OE_PCR G566 P5: 5’-GGA TCT CAG GGA GGA ATT CCG GTA CTG GTA CCC CCT TGA CTG GGG CCT CTC TGC AAA-3’ |
| | | MLRS OE_PCR E576 P6: 5’-CTC TGC AAA GGA CCT CAT AGG |

| | | |
|---|---|---|
| | | CAA TGA ACT GAC ATT CCA TAT ATT CCA CCA CTC AGC-3’ |
| | | MLRS lib P7: 5’- CCA CAG CAC CCC GTG GCC ACC CTG ACT CAG GGA ATA TGG CTG AGT GGT GG-3’ |
| | | MLRS lib P8: 5’-CCA CAG CAC CCC GTG GCC ACC C-3’ |

C-terminal fragment (314 bp) was assembled using equimolar amounts of fragments C1 (192 bp) and C2 (142 bp) for 20 ng total template, MLRS OE_PCR *MluI* P1 and pmol MLRS lib P8 primers, and the components shown in section 2. Thermal conditions are shown in Table E.7.

**Table E.7** Thermal conditions used for the C-terminal fragment assembly during the design of the MtLRS ANIGT, GNIGG and GLVGE mutants.

| **Step** | **Temperature (°C)** | **Time** | |
|---|---|---|---|
| 1 | 98 | 30s | Initial denaturation |
| 2 | 98 | 10s | Denaturation |
| 3 | 50 | 30s | Annealing |
| 4 | 72 | 30s | Extension, go to step 2 34 times |
| 5 | 72 | 5min | Final extension |

The last step of overlapping PCR was performed: Equimolar amounts of both N- and C- terminal fragments (to 35 ng of DNA total) were reacted MLRS_F1 F and MLRS lib P8 primers and the components at section 2. Thermal conditions are shown in Table E.8.

**Table E.8** Thermal conditions to generate the inserts for the MtLRS ANIGT, GNIGG and GLVGE mutants.

| Step | Temperature (°C) | Time | |
|---|---|---|---|
| 1 | 98 | 30s | Initial denaturation |
| 2 | 98 | 10s | Denaturation |
| 3 | 60 | 30s | Annealing |
| 4 | 72 | 45s | Extension, go to step 2<br>34 times |
| 5 | 72 | 5min | Final extension |

The fragment containing both N- and C-terminal mutations (1118 bp) was restricted between *NdeI* and *DraIII* and inserted into pSupK-MtLRS-HL(TAG)3-ΔCP1.

*DH10B E. coli* cells were transformed by electroporation and grown overnight at 37°C on Luria Bertani (LB) plates containing 50 µg/mL kanamycin.

**4. Expression of Z-domain WT and TAG mutant**

*BL21(DE3) E. coli* cells were transformed with pET-21Z WT or K7TAG and spread on LB plates containing 100 µg/mL carbenicillin. Electrocompetent cells were prepared by standard procedure and transformed with the pSupK-MtLRS HL(TAG)3-ΔCP1 plasmids containing the ANIGT, GNIGG or GLVGE mutations and spread on LB plates containing 100 µg/mL carbenicillin and 50 µg/mL kanamycin.

A few colonies containing both pET-21 Z (WT and K7TAG) and pSupK-MtLRS-HL(TAG)3-ΔCP1 plasmids (ANIGT, GNIGG, GLVGE) were grown overnight in TB instant autoinduction media at 37°C with and without 1 mM Dansyl-Dap.

**5. Z-domain purification**

Cells were harvested at 6000 g and 4°C for 10 min and lysed using sonication (5 pulses of 20 s at 50%). Lysate was purified using Ni-NTA column in non-denaturing conditions: Lysis buffer (300 mM NaCl, 20 mM $NaH_2PO_4$, pH: 7.5), wash buffer (300 mM NaCl, 20 mM $NaH_2PO_4$, 5 mM Imidazole, pH: 7.5), and elution buffer (300 mM NaCl, 20 mM $NaH_2PO_4$, 500 mM Imidazole, pH: 7.5). The purified Z-domain proteins were desalted and concentrated using Amicon 3K ultrafiltration column, lyophilized, and sent to LSU facility for ESI-MS analysis.

**6. SDS-PAGE**

Protein samples were mixed with Nu-PAGE loading buffer and heated at 95°C for 5 minutes. Solutions were loaded on Nu-PAGE Bis-Tris 4-12% and ran at 200 V for 50 min in 1X MOPS SDS running buffer. The gel was washed three times with water and stained overnight using Simply Blue Safe Stain (ThermoFisher).

**7. Construction of pRCG plasmid**

**7.1 pRCG Q98TAG D111TAG and pRCG D111TAG**

pBREP and pRep-CatUpp plasmids containing Q98TAG and D111TAG or D111TAG mutations were digested with *HindIII* and *PciI* restriction enzymes. DNA was purified from agarose gel and ligated as described in section 2. The purified DNA was used to transform NEB 5α *E. coli* cells by electroporation.

Single colonies were grown overnight at 37°C in LB containing 100 µg/mL of carbenicillin and harvested at 6000 g for 10 min. The plasmid DNA were isolated and sequenced as described in section 2

### 7.2 pRCG Q98TAG

Quick Change site directed mutagenesis kit (Agilent) was used to remove the D111TAG mutation from pRCG plasmid using two primers: CAT QC TAG111D F: 5'-GGA GTG AAT AAC ACG ACG ATT TCC GGC AGT TTC-3' and CAT QC TAG111D R: 5'- GAA ACT GCC GGA AAT CGT CGT GGT ATT CAC TCC-3'. *DpnI* enzyme was used to destroy the parental DNA and NEB 5α *E. coli* cells were transformed using chemical transformation. The plasmid DNA was isolated, and sequencing as described in section 2.

## 8. Pre-Selection conditions

Positive selection plates were made with Bacto agar, GMML media (1X of M9 salts, 0.1 mM $CaCl_2$, 1% glycerol, 0.3 mM leucine, 100 µg/mL carbenicillin, 50 µg/mL kanamycin, and varying concentrations of chloramphenicol from 0-200 µg/mL. When only pRCG plasmid was tested, no kanamycin was added. Negative selection plates were made using LB agar, 100 µg/mL of carbenicillin, 50 µg/mL kanamycin, and varying concentrations of 5-Fluorouracil from 0-500 µg/mL. When only pRCG plasmid was tested, no kanamycin was added.

*GH371 E. coli* cells were transformed with pRCG alone or both pRCG and pSupK-MtLRS-HL(TAG)3-ΔCP1 plasmids. The cells (40µL) were plated on positive and negative selection plates. Cells were grown for 3 days at 37°C.

## 9. Construction of the MtLRS library

Mutations at the N-terminal were introduced as described in experimental section 3. The primers used are shown in Table E.9.

**Table E.9** Primers used to introduce N-terminal mutations on MtLRS library.

| Reaction | Primer forward | Primer Reverse |
|---|---|---|
| N1 | MLRS_F1 F: 5'-CCC GGT CAT CAA TCA TCC CC-3' | MLRS A32X Y33X OE_PCR R: 5'-GCA TCG CAC CAC TGG GGT AGG GMN NMN NGA CTG TGA GGA ATA TCT TTT CTC TGT C-3' |
| N2 | MLRS Y33X OEPCR F: 5'-CCC TAC CCC AGT GGT GCG ATG C-3' | MLRS *MluI* R: 5'-GGG CAG ACG CGT TCC AAG GC-3' |

The same method was used for the introduction of the C-terminal mutations. Primers are shown in Table E.10.

**Table E.10** Primers used to introduce C-terminal mutations on MtLRS library.

| Reaction | Primers |
|---|---|
| C1 | MLRS OE_PCR *MluI* P1: 5'-CTT GGA ACG CGT CTG CCC TGG GAT GAG AGG TGG ATC ATC GAA CCC CTC ACA GAC TC-3' |
| | MLRS OEPCR Y516X P2: 5'-CAT CGA ACC CCT CAC AGA CTC AAC AAT CNN KAT GGC ATA TTA CAC CAT CGC A-3' |
| | MLRS lib P3: 5'-CAA AGA ACT CAT CGT CCA TCT CCC CGG CAT CCA TCT CCC TGA GGC GGT GTG CGA TGG TGT AAT ATG CC-3' |
| | MLRS lib P4: 5'-CCG GAA TTC CTC CCT GAG ATC CTC AAA GGT TCC TGA ATC ATC TAG GAA TAT GGC ATC AAA GAA CTC ATC |

| | |
|---|---|
| | GTC CAT CTC-3’ |
| C2 | MLRS Library P5: 5’-GGA TCT CAG GGA GGA ATT CCG GTA CTG GTA CCC CCT TGA CTG GNN KCT CTC TGC AAA GGA CCT CAT AGG C-3’ |
| | MLRS library P6: 5’-CTC TGC AAA GGA CCT CAT AGG CAA TNN KCT GAC ATT CCA TAT ATT CCA CCA CTC AGC CAT ATT CCC-3’ |
| | MLRS lib P7: 5’- CCA CAG CAC CCC GTG GCC ACC CTG ACT CAG GGA ATA TGG CTG AGT GGT GG-3’ |
| | MLRS lib P8: 5’-CCA CAG CAC CCC GTG GCC ACC C-3’ |

The N-terminal fragment (823 bp) was assembled as described in section 3, but the first 10 cycles were done without primers, and 25 cycles after adding the primers as described in the literature.[112] The DNA was digested with the *NdeI* and *MluI* restriction enzymes, and introduced into pSupk-MtLRS-HL(TAG)3-ΔCP1 by ligation (Section 2). *DH10B E. coli* cells were transformed by electroporation and plated on LB plates containing 50 μg/mL kanamycin. The pool of cells was collected, and DNA was isolated and sequenced as described in part 2. DNA was digested with the *MluI* and *DraIII* restriction enzymes, and the larger band from the 1% agarose electrophoresis was used as a vector to insert the C-terminal mutations.

The C-terminal fragment (314 bp) was assembled as it was done for the ANIGT, GNIGG and GLVGE mutants, and it was digested with the *MluI* and *DraIII* restriction enzymes and introduced into pSupK-MtLRS-HL(TAG)3-ΔCP1 containing N-terminal mutations by ligation (Section 2). DNA was isolated by ethanol precipitation using yeast tRNA and ethanol, as

described by Zhu et al.[113] *DH10B E. coli* cells were electroporated and grown overnight at 37°C. The DNA was isolated and sequenced (Section 2) to verify the randomization at the intended positions. The transformation efficiency of the library achieved was approximately $10^6$ c.f.u.

**10. Dual genetic screening of the MtLRS library**

Dansyl-Dap was synthesized and characterized as described in section 1 of this experimental section. 4-Nitro-1-phenylalanine was purchased from Aroz Technologies.

Positive and negative selection plates were made as described in section 3 with 50 µg/mL chloramphenicol or 22.5 µg/mL 5-FU, respectively. 1 mM Dansyl-Dap or 4-Nitro-1-phenylalanine was added to the positive selection plates.

GH371 *E. coli* cells were transformed with pSupK-MtLRS-HL(TAG)3-ΔCP1 library plasmid and pRCG Q98TAG and amplified using standard procedures. An aliquot of 40 µL cells with an $OD_{600}$ 210 was diluted with 100 µl of 1X phosphate buffered saline solution (PBS) and plated on positive selection plates.

After 72h of growth at 37°C, the pool of colonies obtained were collected in PBS and centrifuged at 6000 g, and 4°C for 10 min. DNA was isolated as described in section 7.2 and used to transform GH371 *E. coli* cells containing pRCG Q98TAG. The transformed DNA was plated on negative selection plates and grown at 37°C for 72h.

From the last positive selection step, 96 colonies were grown in a 96 well plate with 100 µg/mL carbenicillin and 50 µg/mL kanamycin for 5h. The colonies were stamped on GMML plates containing 50 µg/mL chloramphenicol with or without 1 mM UAA using an autoclaved toothpicks or a replicator. Positive and negative plates were compared after 72 h of growth, and

the clones whose colonies only grew when UAA was present were cultured at 37°C overnight in LB containing carbenicillin and kanamycin, miniprepped and sequenced.

**11. Construction of the aptamer library of the theophylline riboswitch**

The aptamer library was constructed by polymerase chain reaction (PCR) of pTrp-TheoRS-CatUpp using two synthetic oligonucleotides: 5'- CAT CGA ACT AGT TTA ATG TGT GGA AGG TGA TAC NNNN ATC GTC TTG AT NNNNNNN GCA GCA CCC TGC TAA GGT AAC -3'and 5'- CCG TAT GGC AAT GAA AGA CGG -3'. The PCR product was inserted into pTrp-TheoRS-CatUpp between *SpeI* and *EcoRI* restriction sites to produce the library plasmid.

**12. Dual genetic selection of the CaffRS**

The library plasmid was introduced into *E. coli* GH371 cells using electroporation. A chemically defined rich (CDR) medium containing 0.2% glucose, 0.2% casamino acids, 50 mM $Na_2HPO_4$, 50 mM $KH_2PO_4$, 25 mM $(NH_3)_2SO_4$, 2 mM $MgSO_4$, 1x vitamins solution, 1x trace metals solution, and carbenicillin (100 µg/mL) was used for *E. coli* cell culture throughout the selection process.[114] The 10,000x trace metals solution and 1,000x vitamin solution were prepared as previously described.[115] For positive selection, the transformed cells were spread on CDR agar plates containing chloramphenicol (30 µg/mL) and caffeine (2.5 mM) and incubated at 37°C overnight. The cells were collected in Luria-Bertani (LB) medium and plasmid DNA was isolated using QIAprep Spin DNA miniprep kit (Qiagen). For negative selection, the isolated DNA was used to transform the GH371 *E. coli* cells by electroporation. The transformed cells were spread on CDR agar plates containing 5-fluorouracil (0.5 µg/mL) and incubated at 37°C overnight. The plasmid DNA was isolated from the collected cells and used in the subsequent round of positive selection. The positive and negative selection steps were repeated once. From

the third positive selection agar plate, individual colonies were randomly picked and suspended in individual wells of 96-well plates containing CDR medium with chloramphenicol (50 µg/mL) and caffeine (2.5 mM). The plates were incubated at 37°C overnight. A 96-pin replicator was used for replica printing of individual colonies from the 96 well plates onto CDR agar plates containing chloramphenicol (50 µg/mL) with and without 2.5 mM caffeine. Plasmid DNA was isolated and sequenced from one clone that showed growth advantage in the presence of caffeine.

**13. Construction of the p-Trp-CaffRS-GFPuv plasmid**

The selected clone was amplified by PCR using two synthetic oligonucleotides: 5'-GGT ACC ATC GAT GAG CTG TTG ACA ATT AAT CAT CG-3' and 5'-TCC TTT GCT AGC CAT CTT GTT GTT ACC TTA GCA GGG-3'. The PCR product was inserted between *NheI* and *ClaI* restriction sites of pGLO plasmid. The plasmid was introduced into *E. coli* XL1-Blue cells using electroporation, and fluorescent colonies were selected using LB plates containing 2.5 mM of caffeine. The caffeine riboswitch was confirmed by sequencing. The same procedure was used to construct p-Trp-TheoRS-GFPuv.

**14. Construction of the pTrp-CaffRS-LacZ plasmid**

The selected clone was amplified by PCR using two synthetic oligonucleotides: 5'-ATC GAT GCA TGC GAG CTG TTG ACA ATT AAT CAT GG-3' and 5'-GGC CGT TAC TAG TGG ATC CAT CTT GTT GTT ACC TTA GCA GGG-3'. The PCR product was inserted between *SpeI* and *SphI* restriction sites of pBAD-LacZ plasmid. The plasmid was introduced into *E. coli XL1-Blue* cells using electroporation, and blue colonies were selected using LB plates containing 2.5 mM of caffeine, 100 µg/mL of carbenicillin, Isopropyl β-D-1-thiogalactopyranoside (IPTG)

and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). The same procedure was used to construct p-Trp-TheoRS-LacZ.

**15. β-Galactosidase Assay**

*XL1-Blue E. coli* cells containing the plasmid p-Trp-CaffRS-LacZ or p-Trp-TheoRS-LacZ were incubated overnight in carbenicillin/LB culture induced with different concentrations of caffeine (0-5 mM) until reaching log phase. Cell permeabilization procedure was performed according to van Vlack and Seeliger.[96] $OD_{600}$ was determined for the induced cultures. 1 mL of cells was centrifuged for 5 min at 12,000 g, removing supernatant. Cells were resuspended in 1mL of Z buffer (60 mM $Na_2HPO_4$, 40 mM $NaH_2PO_4$, 10 mM KCl, 1 mM $MgSO_4$ and 50 mM of β-ME, adjusted at pH=7) and lysed with sonication using 3 pulses of 20s at 50%. Cell debris was removed by centrifugation, and the lysate was transferred to a new tube. O-nitrophenyl-β-D-galactopyranoside solution (2 mg/mL) was added at t=0 to the lysate and to the blank (Z Buffer) until yellow was visible. The reaction was stopped with $NaHCO_3$ 1M solution and the time was recorded. Absorbance at 420 nm was recorded using Agilent Technologies Cary 60 UV-Vis Spectrophotometer System and Miller Units were calculated according with the equation shown in Chapter 5, section 5.4.

**7.16 Construction of the aptamer library of the glycine riboswitch**

Two rounds of PCR were used for the construction of the library. In the first round, five primers were used and in the second round two additional primers were used. Library was inserted into pTrp-gcvT-CatUpp-GFPuv between *BamHI* and *NcoI* restriction sites. Table E.11 shows the sequences of the primers used.

**Table E.11** Primers used for the design of the gcvT riboswitch library.

| Reaction | Primer |
|---|---|
| 1 | pTrp P1:<br>5'- TAG CGG ATC CGA GCT GTT GAC AAT TAA TCA-3' |
| | pTrp P2:<br>5'- CCA GAG TTG CGT TCC ACA CAT TAA ACT AGT TCG ATG ATT AAT TGT CAA CAG CTC-3' |
| | pTrp P3:<br>5'- GTG TGG AAC GCA ACT CTG GNN NNT GTT TGT GCG GAT GCG CAA ACC ACC-3" |
| | gcvT P4:<br>5'- GGT TCT CTG TCC TGG CAC CNN AAA GNN NNN TTT GCA TAC GCA AAG ANN NCC CCN NNG GTG GTT TGC GCA TCC GCA C-3' |
| | gcvT P6:<br>5'- GGT GCC AGG ACA GAG AAC CTT CAT TTT ACA TGA GGT GTT TCT CTG TCC TTT TTT GTA TG-3' |
| 2 | gcvT pTrp PCR *NcoI* R1:<br>5'- CTT ATG AAA GCG ACC GCA GCT AAA AAA CAT ACA AAA AAG GAC AGA GA-3' |
| | gcvT pTrp PCR *NcoI* R2:<br>5'- TTC TCC ATG GGC GAC CTC CTT ATG AAA GCG ACC GCA GC-3' |

**17. Dual genetic screening of the sarcosine riboswitch**

Plates were made as described in section 12, but the concentration of chloramphenicol was increased to 40 µg/mL for the positive selection, and up to 80 µg/mL after adjustments were made. Also, 5 mM of sarcosine was added instead of caffeine.

No adjustments were made for negative selection plates.

**18. Construction of the pTrp-SarRS-LacZ plasmid**

pTrp-SarRS-LacZ was constructed by introducing the Sar RS hits between *SpeI* and *NcoI* restriction sites of pBAD-LacZ using T4 ligase (NEB). *XL1-Blue E. coli* cells were transformed with the pTrp-SarRS-LacZ plasmids, grown overnight at 37°C on LB plates containing IPTG, Xgal and 5 mM of sarcosine. The DNA from blue colonies were isolated and sequenced as described in section 2.

**19. β-galactosidase assay (modified)**

Because of the high β-galactosidase activity by SarRS variants, a modified assay was performed: Single colonies were grown overnight at 37°C in LB containing 100 µg/mL carbenicillin and 0-5 mM of sarcosine respectively. $OD_{600}$ was measured using biowave cell densimeter. Cells were harvested at 12000 rpm for 10 min, 4°C and sonicated in Z buffer (60 mM $Na_2HPO_4$, 40 mM $NaH_2PO_4$, 10 mM KCl, 1 mM $MgSO_4$ and 50 mM of β-ME, adjusted at pH=7) at 50% (20 s on, 2 s off per 1minute). Cells were centrifuged again to remove bacterial debris, and 100 µL supernatant was mixed with 1 mL water for colorimetric assay. 50 µl of 2 mg/mL 2-nitrophenyl-galactopyranoside was added to the diluted lysates and once yellow was developed, reaction was stopped by adding 80 µL 1 M $Na_2CO_3$ solution. Absorbance at 420 nm was monitored and recorded using Agilent Technologies Cary 60 UV-Vis Spectrophotometer System. Z buffer was

used as control. This assay was done three times per sample. Miller units were calculated following the same equation as section 15. Dilution factors were considered.

**20. Introduction of T7 promoter for *in-vitro* transcription**

T7 promoter was introduced by PCR at the upstream of SarRS clones A6 and C9 using two primers: pTrp-gcvT-CatUpp PCR T7 *EcoRI* F: 5'-GCC AGT GAA TTC TAA TAC GAC TCA CTA TAG GAA CGC AAC TCT GG-3' and pTrp-gcvT-CatUpp PCR *HindIII* R: 5'-TAC GCC AAG CTT AAA AAA GGA CAG AGA AAC ACC TC-3'. The 166 bp fragment obtained was introduced into pUC-19 plasmid between *EcoRI* and *HindIII* restriction sites by ligation. *XL1-Blue E. coli* cells were transformed with the ligation products and grown overnight at 37°C on LB plates containing 100 µg/mL carbenicillin. DNA was isolated and sequenced (Section 7.2). To generate a linear DNA to be used for the *in vitro* transcription, DNA was amplified by PCR using pTrp-gcvT-CatUpp PCR T7 *EcoRI* F and gcvT PCR R (5'-AAA AAA GGA CAG AGA AAC ACC-3') and DNA was purified.

**21. *In-vitro* transcription of T7 SarRS clones A6 and C9**

Linear DNA (1µg) containing T7 promoter was reacted with 1X of T7 RNA polymerase buffer (NEB) with 0.5 mM NTPs, 1 U/µL RNAse inhibitor (NEB), 5 mM Dithiothreitol (DTT), and 10 U T7 RNA polymerase in 50 µL of solution completed with DEPC treated water. Reaction was incubated for 6 h at 37°C, followed by addition of 1µL of DNAse I and incubation at 37°C for 1 h.

RNA was analyzed using 10% TBE-Urea PAGE, bands were visualized by UV-shadowing, cut minced, extracted overnight with RNA extraction buffer (100 mM NaCH3COO, 1 mM EDTA and 0.1% SDS, pH= 4.5) and filtered. RNA was concentrated with n-butanol, followed by

ethanol precipitation. The RNA pellets were dissolved in DEPC water and stored at -20°C until used.[83]

# APPENDIX

## A.1 Homology results analysis between archaea and bacteria LeuRS

```
CLUSTAL O(1.2.4) multiple sequence alignment

WP_063123617.1: E. coli
SP|O27552: M.thermoaut.
BAW01681.1 T. thermoph.
WP_010885055.1 P. horikoshii

WP_063123617.1      MQEQYRPEEIESKVQLHWDEKRTFEVTE-DE-SKEKYYCLSMLPYPSGRLHMGHVRNYTI 58
SP|O27552|SYL_METTH -------MDIERKWRDRWRDAGIFQADPDD---REKIFLTVAYPYPSGAMHIGHGRTYTV 50
BAW01681.1          -MEKYNPHAIEAKWQRFWEEKGFMKAKDLPG-GRGKQYVLVMFPYPSGDLHMGHLKNYTM 58
WP_010885055.1      -MAELNFKAIEEKWQKRWLEAKIFEPNIRDKPKEKKFYITVAFPYLSGHLHVGHARTYTI 59
                             ** * :  * :   ::        . * :     ** ** :*:** :.**:

WP_063123617.1      GDVIARYQRMLGKNVLQPIGWDAFGLPAEGAAVKNNTA--PAPWTYDNI----------- 105
SP|O27552|SYL_METTH PDVYARFKRMQGYNVLFPMAWHVTGAPVIGIARRIQRKDPWTLKIYREVHRVPEDELERF 110
BAW01681.1          GDVLARFRRMQGYEVLHPMGWDAFGLPAENAALKFGVH--PKDWTYANI----------- 105
WP_010885055.1      PDVIARFKRMQGYNVLFPMAWHITGSPIVGIAERIKNRDPKTIWIYRDVYKVPEEILWTF 119
                     ** **::** * :** *:.*.  * *  . * :          * ::

WP_063123617.1      -----------AYMKNQLKMLGFGYDWSRELATC--TPEYYRWEQKFFTELYKKGLVYKK 152
SP|O27552|SYL_METTH SDPEYIVEYFSREYRSVMEDMGYSIDWRREFKTT--DPTYSRFIQWQIRKLRDLGLVRKG 168
BAW01681.1          -----------RQAKESLRLMGILYDWDREVTTC--EPEYYRWNQWIFLKMWEKGLAYRA 152
WP_010885055.1      EDPINIVKYFMKAAKETFIRAGFSVDWSREFYTTSLFPPFSKFIEWQFWKLKEKGYIVKG 179
                                 :. :   *   ** **. *    * : :: :  : :: . *   :

WP_063123617.1      TSAVNWCPNDQTVLANEQVIDGCCWRCD-TKVERKEIPQWFIKITAYADELLNDLDKLDH 211
SP|O27552|SYL_METTH AHPVKYCPECENPVGDHDLLEGEGVAINQLTLLK------------------------- 202
BAW01681.1          KGLVNWCPKCQTVLANEQVVEGRCWRHEDTPVEKRELEQWYLRITAYAERLLKDLEGL-D 211
WP_010885055.1      AHRVRWDPVVGTPLGDHDLMEGEDVPILDYIIIK------------------------- 213
                       *.: *   . :.:.::::*         : :

WP_063123617.1      WPDTVKTMQRNWIGRSEGVEITFNVNDYDNTLTVYTTRPDTFMGCTYLAVAAGHPLAQKA 271
SP|O27552|SYL_METTH --------------------FKLGD----SYLVAATFRPETIYGATNLWLNPDEDYVRVE 238
BAW01681.1          WPEKVKAMQRAWIGRSEGAEILFPVEGKEVKIPVFTTRPDTLFGATFLVLAPEHPLTLEL 271
WP_010885055.1      -------------------FELRENGEVIYLPAATLRPETVYGVTNMWVNPNATYVKAK 253
                                     : :       : . * **:*. * * : :      .

WP_063123617.1      A--ENNPEL-AAFIDECRNT--KVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFV 326
SP|O27552|SYL_METTH ----TGGEEWIISRAAVDNLSHQKLDLKVSGDVNPGDLIGMCVENPVTGQEHPILPASFV 294
BAW01681.1          AAPERREEV-LAYVEAAKRK--TEIERQAEGREKTGVFLGAYALNPATGERVPIWTADYV 328
WP_010885055.1      VRRKDKEETWIVSKEAAYKLSFQDREIEVIEEFKGEKLIGKYVRNPVSGDEVIILPAEFV 313
                           *          .     : :     :     *  . :* :*:.  :  *.:*
```

```
WP_063123617.1      LMEYGTGAVMAVPGHDQRDY----------EFASKYGLNIKPVI------LAA-D--GSE 367
SP|O27552|SYL_METTH DPEYATGVVFSVPAHAPADFIALEDLRTDHELLERYGLED-VVADIEPVNVIAVDGYGEF 353
BAW01681.1          LYGYGTGAIMAVPAHDQRDY----------EFARKFGLPIKKVI------ERP-E--DPL 369
WP_010885055.1      DPDNATGVVMSVPAHAPFDHVALEDLKRETEILEKYDIDPRIVENITYISLIKLEGYGDF 373
                        .**.:::**.*   *.          *:  ::.:    *          :  .

WP_063123617.1      P----------------DLSQ-------QALTEKGVLF-NSGEFNGLDHEAAFNAIADKL 403
SP|O27552|SYL_METTH PAAEVIEKFGVRNQED-PRLEDATGELYKIEHARGVMSSHIPVYGGMKVSEAREVIADEL 412
BAW01681.1          P----------------EPLE-------RAYEEPGIMV-NSGPFDGTPSEEGKKRVIAWL 405
WP_010885055.1      PAVEEVNKLGIKSQKDKEKLEQATKTIYKAEYHKGIF--KVPPYEGKPVQEVKEAIAKEM 431
                    *                   :      :     *::  :   : *   .   : :   :

WP_063123617.1      TAMGVGERKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQLPVILPEDV----- 458
SP|O27552|SYL_METTH KDQGLADEMYEFAERPVI-------------CRCGGRCVVRVMEDQWFMKYSDDAWKDLA 459
BAW01681.1          EEKGLGKGRVTYRLRDWLISRQRYWGTPIPMVHCEACGVVPVPEEELPVLLPDLK----- 460
WP_010885055.1      LEKGIAEIMYEFAEKNVI-------------SRFGNRAVIKIIHDQWFIDYGNPEWKEKA 478
                       *:..    :  :                      *:   .::  :   :

WP_063123617.1      -----VMDGI----TSPIKADPEWAK-TT--VNGM---PALR-ETDTFDTFMESSWYYAR 502
SP|O27552|SYL_METTH HRCLDGMKIIPEEVRANFEYYIDWLNDWACSRRIGLGTRLPWDERWIIEPLTDSTIYMAY 519
BAW01681.1          -----DVEDIRPKGKSPLEAHPEFYE-TTCPRCGG---PAKR-DTDTMDTFFDSSWYYLR 510
WP_010885055.1      RKALERMKILPETRRAQFEAIIDWLDKKACARKIGLGTPLPWDPEWVIESLSDSTIYMAY 538
                          :. :     : ::   :: .  :                 :: : :*: *


WP_063123617.1      YTCPEYKEGM-----LD---------------------------------------SEAA 518
SP|O27552|SYL_METTH YTIAHRLREM------DAGEMDDEFFDAIFLDDSG---------------TFEDLREEF 557
BAW01681.1          YTDPHNDRLP-----FD---------------------------------------PEKA 526
WP_010885055.1      YTISRHINKLRQEGKLDPEKLTPEFFDYIFLEEFSEDKEKELEKKTGIPAEIIHEMKEEF 598
                    **  .  .       *                                          *

WP_063123617.1      NYWLPVDIYIGGIEHAIMHLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMVLADAFYYV- 577
SP|O27552|SYL_METTH RYWYPLDWRLSAKDLIGNHLTFHIFHHSAIFP---------------------------- 589
BAW01681.1          NAWMPVDQYIGGVEHAVLHLLYSRFFTKFLHDLGMVKVEEPFQGLFTQGMVLAWTDFGPV 586
WP_010885055.1      EYWYPLDWRCSGKDLIPNHLTFFIFNHVAIFR---------------------------- 630
                    . * *:*   .. :    ** :  *    :

WP_063123617.1      GENGERNWV-SPVDAIVERDEKGRIVK--------AKDAAGHELVYTGMSKMSKSKNNGI 628
SP|O27552|SYL_METTH ----ESGWP--------------------------RGAVVFGMGLLEGNKMSSSKGNVI 618
BAW01681.1          EVEGEVVRLPEPTRIRLEIPESALSLEDVRKMGAELRPHEDGTLHLWKPAVMSKSKGNGV 646
WP_010885055.1      ----EEHWP--------------------------KGIAVNGFGTLEGQKMSKSKGNVL 659
                        *                               :     :      **.**.* :
```

```
WP_063123617.1      DPQVMVERYGADTVRLFMMFASPADMTLEWQESGVEGANRFLKRVWKLVYEHTAKG---- 684
SP|O27552|SYL_METTH LLRDAIEKHGADVVRLFLMSSAEPWQDFDWRESEVIGTRRRIEWFREFGERVSGILDGRP 678
BAW01681.1          MVGPFVKEQGADIARITILFAAPPENEMVWTEEGVQGAWRFLNRIYRRVAEDREAL---L 703
WP_010885055.1      NFIDAIEENGADVVRLYIMSLAEHDSDFDWRRKEVGKLRKQIERFYELISQFAEY----- 714
                         ::. *** .*: ::  :     : * .. *    : :: . .   .

WP_063123617.1      -DVAALNVDALTEDQKALRRDVHKTIAKVTDDIGRRQTFNTAIAAIMELMNKLAKAPT-- 741
SP|O27552|SYL_METTH VLSEVTPAEPESFIGRWMMGQLNQRIREATRALESFQTRKAVQEALYLLKKDVDHYLKRV 738
BAW01681.1          ATSEVFQAEALEGKDRELYGKLHETLKKVTEDLE-ALRFNTAIAALMEFLNALYEYRK-- 760
WP_010885055.1      ---EVKGNVELKDIDRWMLHRLNKAIKETTNALEEFRTRTAVQWAFYSIMNDLRWYLRRT 771
                        .          : :   ::: : :.*  :      .:.  *:  : : :

WP_063123617.1      ---DGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELKGEGDIDNAPWPVADEKAMEED 798
SP|O27552|SYL_METTH EGRVDDEVKSVLANVLHAWIRLMAPFIPYTAEEMWERYGGEGFVAEAPWPDFSDDAESRD 798
BAW01681.1          ---DRPV-TPVYRTALRYYLQMLFPFAPHLAEELWHWFWPD-SLFQAGWPELDEKALEKD 815
WP_010885055.1      EGRDDEAKRYVLRTLADVWVRLMAPFTPHICEELWEKLGGEGFVSLAKWPEPVEEWWNET 831
                              :        :::: ** *: .  :*.    :  :  * **   :.  ..

WP_063123617.1      STLVVVQVNGKVRAKITVPVDAT----------------EEQVRERAGQEH--------- 833
SP|O27552|SYL_METTH VQVAEEMVQNTVRDIQEIMKILG-STPERVHIYTSPKWKWDVLRVAAEVGKLDMGSIMGR 857
BAW01681.1          VVEVAVQVNGRVRGTIRIPKDAP---------------LEVARAEALKVR--------- 850
WP_010885055.1      IEAEEEFIRSVMEDIKEIIEVAKIENAKRAYIYTAEDWKWKVAEVVSEKRDFK--SSMEE 889
                           :.. :.    :                     .  .  :
```

## REFERENCES


1. Simonetti, A.; Marzi, S.; Jenner, L.; Myasnikov, A.; Romby, P.; Yusupova, G.; Klaholz, B. P.; Yusupov, M., A structural view of translation initiation in bacteria. *Cellular and Molecular Life Sciences* **2008,** *66* (3), 423.

2. Berg, M. D.; Brandl, C. J., Transfer RNAs: diversity in form and function. *RNA Biology* **2021,** *18* (3), 316-339.

3. Anderson, J. C.; Schultz, P. G., Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair for Four-base, Amber, and Opal Suppression. *Biochemistry* **2003,** *42* (32), 9598-9608.

4. Rubio Gomez, M. A.; Ibba, M., Aminoacyl-tRNA synthetases. *RNA (New York, N.Y.)* **2020,** *26* (8), 910-936.

5. Liu, C.; Sanders, J. M.; Pascal, J. M.; Hou, Y.-M., Adaptation to tRNA acceptor stem structure by flexible adjustment in the catalytic domain of class I tRNA synthetases. *RNA* **2012,** *18* (2), 213-221.

6. Beuning, P. J.; Musier-Forsyth, K., Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers* **1999,** *52* (1), 1-28.

7. Yan, W.; Tan, M.; Eriani, G.; Wang, E.-D., Leucine-specific domain modulates the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase. *Nucleic Acids Research* **2013,** *41* (9), 4988-4998.

8. Vijayakumar, R.; Parismita, K.; Harish, S.; Awanish, K.; Timir, T., Aminoacyl-tRNA synthetases: Structure, function, and drug discovery. *International Journal of Biological Macromolecules* **2018,** *111*, 400-414.

9. Palencia, A.; Crépin, T.; Vu, M. T.; Lincecum, T. L.; Martinis, S. A.; Cusack, S., Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase. *Nature Structural and Molecular Biology* **2012,** *19* (7), 677-84.

10. Sehnal, D.; Bittrich, S.; Deshpande, M.; Svobodová, R.; Berka, K.; Bazgier, V.; Velankar, S.; Burley, S. K.; Koča, J.; Rose, A. S., Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research* **2021,** *49* (W1), W431-W437.

11. Fukunaga, R.; Yokoyama, S., Aminoacylation complex structures of leucyl-tRNA synthetase and tRNALeu reveal two modes of discriminator-base recognition. *Nature Structural* and *Molecular Biology* **2005,** *12* (10), 915-922.

12. Tukalo, M.; Yaremchuk, A.; Fukunaga, R.; Yokoyama, S.; Cusack, S., The crystal structure of leucyl-tRNA synthetase complexed with tRNALeu in the post-transfer–editing conformation. *Nature Structural and Molecular Biology* **2005,** *12* (10), 923-930.

13. Ovaa, H.; wals, k., Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Frontiers in Chemistry* **2014,** *2*: 15.

14. Normanly, J.; Kleina, L. G.; Masson, J.-M.; Abelson, J.; Miller, J. H., Construction of Escherichia coli amber suppressor tRNA genes: III. Determination of tRNA specificity. *Journal of Molecular Biology* **1990,** *213* (4), 719-726.

15. Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G., A genetically encoded fluorescent amino acid. *Proceedings of the National Academy of Sciences* **2006,** *103* (26), 9785.

16. Kim, C. H.; Axup, J. Y.; Schultz, P. G., Protein conjugation with genetically encoded unnatural amino acids. *Current Opinion in Chemical Biology* **2013,** *17* (3), 412-419.

17. Ge, Y.; Fan, X.; Chen, P. R., A genetically encoded multifunctional unnatural amino acid for versatile protein manipulations in living cells. *Chemical Science* **2016,** *7* (12), 7055-7060.

18. Wang, J.; Xie J Fau - Schultz, P. G.; Schultz, P. G., A genetically encoded fluorescent amino acid. *Journal of the American Chemical Society* **2006**, *128* (27), 8738-8739.

19. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G., A Genetically Encoded Photocaged Amino Acid. *Journal of the American Chemical Society* **2004,** *126* (44), 14306-14307.

20. Ryu, Y.; Schultz, P. G., Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nature Methods* **2006,** *3* (4), 263-265.

21. Italia, J. S.; Addy, P. S.; Wrobel, C. J. J.; Crawford, L. A.; Lajoie, M. J.; Zheng, Y.; Chatterjee, A., An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nature Chemical Biology* **2017,** *13* (4), 446-450.

22. Melançon, C. E., 3rd; Schultz, P. G., One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorganic Medicinal Chemistry Letters* **2009,** *19* (14), 3845-7.

23. Krishnakumar, R.; Ling, J., Experimental challenges of sense codon reassignment: An innovative approach to genetic code expansion. *FEBS Letters* **2014,** *588* (3), 383-388.

24. Nikić-Spiegel, I., Expanding the Genetic Code for Neuronal Studies. *ChemBioChem* **2020,** *21* (22), 3169-3179.

25. De la Torre, D.; Chin, J. W., Reprogramming the genetic code. *Nature Reviews Genetics* **2021,** *22* (3), 169-184.

26. Wang, L.; Xie J Fau - Schultz, P. G.; Schultz, P. G., Expanding the genetic code. *Annual Reviews of Biophysical and Biomolecular Structure* **2006**, *35*, 225-249.

27. Fung, E. A Leucyl-Trna Synthetase Without The Editing Domain And Its Directed Evolution For Unnatural Protein Engineering. Honors Thesis, Texas Christian University, 2016.

28. Lee, B. S.; Choi, W. J.; Lee, S. W.; Ko, B. J.; Yoo, T. H., Towards Engineering an Orthogonal Protein Translation Initiation System. *Frontiers in chemistry* **2021,** *9*, 772648-772648.

29. Cellitti, S. E.; Jones, D. H.; Lagpacan, L.; Hao, X.; Zhang, Q.; Hu, H.; Brittain, S. M.; Brinker, A.; Caldwell, J.; Bursulaya, B.; Spraggon, G.; Brock, A.; Ryu, Y.; Uno, T.; Schultz, P. G.; Geierstanger, B. H., In vivo Incorporation of Unnatural Amino Acids to Probe Structure, Dynamics, and Ligand Binding in a Large Protein by Nuclear Magnetic Resonance Spectroscopy. *Journal of the American Chemical Society* **2008,** *130* (29), 9268-9281.

30. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G., Addition of the keto functional group to the genetic code of Escherichia coli. *Procedings of the National Academy of Science U S A* **2003,** *100* (1), 56-61.

31. Guedez, A.; Sherman, M.; Ryu, Y., Dual genetic selection of the theophylline riboswitch with altered aptamer specificity for caffeine. *Biochemical and Biophysical Research Communications* **2021,** *579*, 105-109.

32. Bédard, A.-S. V.; Hien, E. D. M.; Lafontaine, D. A., Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2020,** *1863* (3), 194501.

33. Winkler, W. C.; Breaker, R. R., REGULATION OF BACTERIAL GENE EXPRESSION BY RIBOSWITCHES. *Annual Review of Microbiology* **2005,** *59* (1), 487-517.

34. Breaker, R. R., Riboswitches and Translation Control. *Cold Spring Harbor perspectives in biology* **2018,** *10* (11), a032797.

35. Ariza-Mateos, A.; Nuthanakanti, A.; Serganov, A., Riboswitch Mechanisms: New Tricks for an Old Dog. *Biochemistry (Mosc)* **2021,** *86* (8), 962-975.

36. Smith, A. M.; Fuchs, R. T.; Grundy, F. J.; Henkin, T., Riboswitch RNAs: Regulation of gene expression by direct monitoring of a physiological signal. *RNA Biology* **2010,** *7* (1), 104-110.

37. Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R., A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science* **2004,** *306*, 275-279.

38. Mandal, M.; Breaker, R. R., Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology* **2004**, *5*(6), 451-463.

39. Serganov, A.; Nudler, E., A Decade of Riboswitches. *Cell* **2013,** *152* (1), 17-24.

40. Winkler, W. C.; Nahvi A Fau - Roth, A.; Roth A Fau - Collins, J. A.; Collins Ja Fau - Breaker, R. R.; Breaker, R. R., Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **2004**, *428*, 281-286.

41. Lee Elaine, R.; Baker Jenny, L.; Weinberg, Z.; Sudarsan, N.; Breaker Ronald, R., An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. *Science* **2010,** *329* (5993), 845-848.

42. Garst, A. D.; Edwards, A. L.; Batey, R. T., Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol* **2011,** *3* (6), a003533.

43. Hendrix, D. K.; Brenner, S. E.; Holbrook, S. R., RNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics* **2005,** *38* (3), 221-43.

44. Jenkins, J. L.; Krucinska, J.; McCarty, R. M.; Bandarian, V.; Wedekind, J. E., Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *Journal of Biological Chemistry* **2011,** *286* (28), 24626-24637.

45. Batey, R. T., Recognition of S-adenosylmethionine by riboswitches. *Wiley Interdisciplinary Reviews RNA* **2011,** *2* (2), 299-311.

46. Huang, L.; Serganov, A.; Patel, D. J., Structural Insights into Ligand Recognition by a Sensing Domain of the Cooperative Glycine Riboswitch. *Molecular Cell* **2010,** *40* (5), 774-786.

47. Tanida, Y.; Matsuura, A., Alchemical free energy calculations via metadynamics: Application to the theophylline-RNA aptamer complex. *Journal of Computational Chemistry* **2020,** *41* (20), 1804-1819.

48. Suess, B.; Fink, B.; Berens, C.; Stentz, R.; Hillen, W., A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Research* **2004,** *32* (4), 1610-1614.

49. Jenison Robert, D.; Gill Stanley, C.; Pardi, A.; Polisky, B., High-Resolution Molecular Discrimination by RNA. *Science* **1994,** *263* (5152), 1425-1429.

50. Page, K.; Shaffer, J.; Lin, S.; Zhang, M.; Liu, J. A.-O., Engineering Riboswitches in Vivo Using Dual Genetic Selection and Fluorescence-Activated Cell Sorting. *ACS Synthetic Biology* **2018**, *7*(9), 2000-2006.

51. Topp, S.; Gallivan, J. P., Emerging applications of riboswitches in chemical biology. *ACS Chemical Biology* **2010,** *5* (1), 139-148.

52. Etzel, M.; Mörl, M., Synthetic Riboswitches: From Plug and Pray toward Plug and Play. *Biochemistry* **2017,** *56* (9), 1181-1198.

53. Lee, G.; Jang, G. H.; Kang, H. Y.; Song, G., Predicting aptamer sequences that interact with target proteins using an aptamer-protein interaction classifier and a Monte Carlo tree search approach. *PLOS ONE* **2021,** *16* (6), e0253760.

54. Chen, Z.; Hu, L.; Zhang, B.-T.; Lu, A.; Wang, Y.; Yu, Y.; Zhang, G., Artificial Intelligence in Aptamer-Target Binding Prediction. *International journal of molecular sciences* **2021,** *22* (7), 3605.

55. Laserson, U.; Gan, H. H.; Schlick, T., Predicting candidate genomic sequences that correspond to synthetic functional RNA motifs. *Nucleic Acids Research* **2005,** *33* (18), 6057-6069.

56. Lee, J. F.; Hesselberth, J. R.; Meyers, L. A.; Ellington, A. D., Aptamer database. *Nucleic Acids Research* **2004,** *32* (Database issue), D95-D100.

57. Zulkeflee Sabri, M.; Azzar Abdul Hamid, A.; Mariam Sayed Hitam, S.; Zulkhairi Abdul Rahim, M., In-Silico Selection of Aptamer: A Review on the Revolutionary Approach to Understand the Aptamer Design and Interaction Through Computational Chemistry. *Materials Today: Proceedings* **2019,** *19*, 1572-1581.

58. Sepehri Zarandi, H.; Behbahani, M.; Mohabatkar, H., In Silico Selection of Gp120 ssDNA Aptamer to HIV-1. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* **2020,** *25* (9), 1087-1093.

59. Romero-Reyes, M. A.; Heemstra, J. M., Sequestration and Removal of Multiple Small-Molecule Contaminants Using an Optimized Aptamer-Based Ultrafiltration System. *Bioconjugate Chemistry* **2021,** *32* (9), 2043-2051.

60. Sanford, A. A.; Rangel, A. E.; Feagin, T. A.; Lowery, R. G.; Argueta-Gonzalez, H. S.; Heemstra, J. M., RE-SELEX: restriction enzyme-based evolution of structure-switching aptamer biosensors. *Chemical Science* **2021,** *12* (35), 11692-11702.

61. Romero-Reyes, M. A.; Heemstra, J. M., Small-Molecule Sequestration Using Aptamer-Functionalized Membranes. *ACS Materials Letters* **2019,** *1* (5), 568-572.

62. Pellitero, M. A.; Arroyo-Currás, N., Study of surface modification strategies to create glassy carbon-supported, aptamer-based sensors for continuous molecular monitoring. *Analytical and Bioanalytical Chemistry* **2022**.

63. Shahrad, S.; Rajabi, M.; Javadi, H.; Karimi Zarchi, A. A.; Darvishi, M. H., Targeting lung cancer cells with MUC1 aptamer-functionalized PLA-PEG nanocarriers. *Scientific Reports* **2022,** *12* (1), 4718.

64. Yang, C.; Zhao, H.; Sun, Y.; Wang, C.; Geng, X.; Wang, R.; Tang, L.; Han, D.; Liu, J.; Tan, W., Programmable manipulation of oligonucleotide–albumin interaction for elongated circulation time. *Nucleic Acids Research* **2022,** *50* (6), 3083-3095.

65. Liu, Y.; Peng, C.; Zhang, H.; Li, J.; Ou, H.; Sun, Y.; Wen, C.; Qi, D.; Hu, X.; Wu, E.; Tan, W., DNA aptamer S11e recognizes fibrosarcoma and acts as a tumor suppressor. *Bioactive Materials* **2022,** *12*, 278-291.

66. Gallivan, J. P., Toward reprogramming bacteria with small molecules and RNA. *Current Opinion in Chemical Biology* **2007,** *11* (6), 612-619.

67. Wu, Y. X.; Kwon, Y. J., Aptamers: The "evolution" of SELEX. *Methods* **2016,** *106*, 21-28.

68. Dixon, N.; Duncan Jn Fau - Geerlings, T.; Geerlings T Fau - Dunstan, M. S.; Dunstan Ms Fau - McCarthy, J. E. G.; McCarthy Je Fau - Leys, D.; Leys D Fau - Micklefield, J.; Micklefield, J., Reengineering orthogonally selective riboswitches. *Procedures of the National Academy of Science* **2020**, *107*(7), 2830-2835.

69. Stoltenburg, R.; Reinemann, C.; Strehlitz, B., SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* **2007,** *24* (4), 381-403.

70. Liu, H.; Yu, J., Challenges of SELEX and Demerits of Aptamer-Based Methods. *Aptamers for Analytical Applications* **2018**, 345-364.

71. Lou, X.; Qian, J.; Xiao, Y.; Viel, L.; Gerdon Aren, E.; Lagally Eric, T.; Atzberger, P.; Tarasow Theodore, M.; Heeger Alan, J.; Soh, H. T., Micromagnetic selection of aptamers in microfluidic channels. *Proceedings of the National Academy of Sciences* **2009,** *106* (9), 2989-2994.

72. Liu, Y.; Wang, C.; Li, F.; Shen, S.; Tyrrell, D. L. J.; Le, X. C.; Li, X.-F., DNase-Mediated Single-Cycle Selection of Aptamers for Proteins Blotted on a Membrane. *Analytical Chemistry* **2012,** *84* (18), 7603-7606.

73. Drabovich, A.; Berezovski, M.; Krylov, S. N., Selection of Smart Aptamers by Equilibrium Capillary Electrophoresis of Equilibrium Mixtures (ECEEM). *Journal of the American Chemical Society* **2005,** *127* (32), 11224-11225.

74. McKeague, M.; Wong, R. S.; Smolke, C. D., Opportunities in the design and application of RNA for gene expression control. *Nucleic Acids Research* **2016,** *44* (7), 2987-2999.

75. Muranaka, N.; Sharma, V.; Nomura, Y.; Yokobayashi, Y., An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Research* **2009,** *37* (5), e39-e39.

76. Desai, S. K.; Gallivan, J. P., Genetic Screens and Selections for Small Molecules Based on a Synthetic Riboswitch That Activates Protein Translation. *Journal of the American Chemical Society* **2004,** *126* (41), 13247-13254.

77. Nomura, Y.; Yokobayashi, Y., Dual genetic selection of synthetic riboswitches in Escherichia coli. *Methods in Molecular Biology* **2014**, *1111*, 131-130.

78. Hong, K.-Q.; Zhang, J.; Jin, B.; Chen, T.; Wang, Z.-W., Development and characterization of a glycine biosensor system for fine-tuned metabolic regulation in Escherichia coli. *Microb Cell Fact* **2022,** *21* (1), 56-56.

79. Liberman, J. A.; Bogue, J. T.; Jenkins, J. L.; Salim, M.; Wedekind, J. E., Chapter Eighteen - ITC Analysis of Ligand Binding to PreQ1 Riboswitches. In *Methods in Enzymology*, Burke-Aguero, D. H., Ed. Academic Press: 2014; Vol. 549, pp 435-450.

80. Amano, R.; Furukawa, T.; Sakamoto, T., ITC Measurement for High-Affinity Aptamers Binding to Their Target Proteins. In *Microcalorimetry of Biological Molecules: Methods and Protocols*, Ennifar, E., Ed. Springer New York: New York, NY, 2019; pp 119-128.

81. Regulski, E. E.; Breaker, R. R., In-line probing analysis of riboswitches. *Methods in Molecular Biology* **2008**, *419*, 53-67.

82. Mlýnský, V.; Bussi, G., Understanding in-line probing experiments by modeling cleavage of nonreactive RNA nucleotides. *RNA (New York, N.Y.)* **2017,** *23* (5), 712-720.

83. Budhathoki, P.; Bernal-Perez, L. F.; Annunziata, O.; Ryu, Y., Rationally-designed fluorescent lysine riboswitch probes. *Organic* and *Biomolecular Chemistry* **2012,** *10* (39), 7872-7874.

84. Jing, M.; Bowser, M. T., Methods for measuring aptamer-protein equilibria: a review. *Analytica Chimica Acta* **2011**, *686* (1-2), 9-18.

85. Sun, L.; Jin, H.; Zhao, X.; Liu, Z.; Guan, Y.; Yang, Z.; Zhang, L.; Zhang, L., Unfolding and Conformational Variations of Thrombin-Binding DNA Aptamers: Synthesis, Circular Dichroism and Molecular Dynamics Simulations. *ChemMedChem* **2014,** *9* (5), 993-1001.

86. Poniková, S.; Antalík M Fau - Hianik, T.; Hianik, T., A circular dichroism study of the stability of guanine quadruplexes of thrombin DNA aptamers at presence of K+ and Na+ ions. *General Physiology and Biophysics* **2008**, *27*(4), 271-277.

87. Tang, Q.; Su X Fau - Loh, K. P.; Loh, K. P., Surface plasmon resonance spectroscopy study of interfacial binding of thrombin to antithrombin DNA aptamers. *Journal of Colloids and Interface Science* **2007**, *315* (1), 99-106.

88. Machtel, P.; Bąkowska-Żywicka, K.; Żywicki, M., Emerging applications of riboswitches - from antibacterial targets to molecular tools. *Journal of Applied Genetics* **2016,** *57* (4), 531-541.

89. Lynch, S. A.; Desai, S. K.; Sajja, H. K.; Gallivan, J. P., A High-Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights into Their Function. *Chemistry and Biology* **2007,** *14* (2), 173-184.

90. Lynch, S. A.; Topp S Fau - Gallivan, J. P.; Gallivan, J. P., High-throughput screens to discover synthetic riboswitches. *Methods in Molecular Biology* **2009**, *540,* 321-333.

91. Bass, S. H.; Yansura, D. G., Application of the E. coli trp promoter. *Molecular Biotechnology* **2000**, *16* (3), 253-260.

92. Hallewell Ra Fau - Emtage, S.; Emtage, S., Plasmid vectors containing the tryptophan operon promoter suitable for efficient regulated expression of foreign genes. *Gene* **1980**, *9* (1-2), 24-47.

93. Sherman, M. Synthetic caffeine riboswitches as genetical and analytical tools. Honors Thesis, Texas Christian University, 2014.

94. Peters, G. J.; Backus, H. H. J.; Freemantle, S.; van Triest, B.; Codacci-Pisanelli, G.; van der Wilt, C. L.; Smid, K.; Lunec, J.; Calvert, A. H.; Marsh, S.; McLeod, H. L.; Bloemena, E.; Meijer, S.; Jansen, G.; van Groeningen, C. J.; Pinedo, H. M., Induction of thymidylate synthase as a 5-fluorouracil resistance mechanism. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2002,** *1587* (2), 194-205.

95. 95. Reuter, J. S.; Mathews, D. H., RNAstructure: software for RNA secondary structure prediction and analysis. *Nucleic Acids Research* **2013**, *41*(W), 471-474.

96. Seeliger, J. C.; Topp, S.; Sogi, K. M.; Previti, M. L.; Gallivan, J. P.; Bertozzi, C. R., A Riboswitch-Based Inducible Gene Expression System for Mycobacteria. *PLOS ONE* **2012,** *7* (1), e29266.

97. Shanidze, N. A.-O.; Lenkeit, F.; Hartig, J. S.; Funck, D. A.-O., A Theophylline-Responsive Riboswitch Regulates Expression of Nuclear-Encoded Genes. *Plant Physiology* **2020**, *182* (1), 123-135.

98. Topp, S.; Reynoso, C. M. K.; Seeliger, J. C.; Goldlust, I. S.; Desai, S. K.; Murat, D.; Shen, A.; Puri, A. W.; Komeili, A.; Bertozzi, C. R.; Scott, J. R.; Gallivan, J. P., Synthetic riboswitches that induce gene expression in diverse bacterial species. *Appl Environ Microbiol* **2010,** *76* (23), 7881-7884.

99. Reynoso Colleen, M. K.; Miller Mark, A.; Bina James, E.; Gallivan Justin, P.; Weiss David, S.; Sansonetti Philippe, J., Riboswitches for Intracellular Study of Genes Involved in Francisella Pathogenesis. *mBio 3* (6), e00253-12.

100. Burn, S. F., Detection of β-galactosidase activity: X-gal staining. *Methods in Molecular Biology* **2012**, *886*, 241-250.

101. Smale, S. T., Beta-galactosidase assay. *Cold Harbor Spring Protocols* **2010**, *5:pdb.prot5423.*

102. Barrick, J. E.; Breaker, R. R., The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biology* **2007,** *8* (11), R239.

103. Ruff, K. M.; Muhammad, A.; McCown, P. J.; Breaker, R. R.; Strobel, S. A., Singlet glycine riboswitches bind ligand as well as tandem riboswitches. *RNA* **2016,** *22* (11), 1728-1738.

104. Crum, M.; Ram-Mohan, N.; Meyer, M. M., Regulatory context drives conservation of glycine riboswitch aptamers. *PLOS Computational Biology* **2019,** *15* (12), e1007564.

105. Torgerson, C. D.; Hiller, D. A.; Stav, S.; Strobel, S. A., Gene regulation by a glycine riboswitch singlet uses a finely tuned energetic landscape for helical switching. *RNA* **2018,** *24* (12), 1813-1827.

106. Khan, A. P.; Rajendiran, T. M.; Bushra, A.; Asangani, I. A.; Athanikar, J. N.; Yocum, A. K.; Mehra, R.; Siddiqui, J.; Palapattu, G.; Wei, J. T.; Michailidis, G.; Sreekumar, A.; Chinnaiyan, A. M., The Role of Sarcosine Metabolism in Prostate Cancer Progression. *Neoplasia* **2013,** *15* (5), 491-IN13.

107. Cernei, N.; Heger, Z.; Gumulec, J.; Zitka, O.; Masarik, M.; Babula, P.; Eckschlager, T.; Stiborova, M.; Kizek, R.; Adam, V., Sarcosine as a potential prostate cancer biomarker--a review. *Int J Mol Sci* **2013,** *14* (7), 13893-908.

108. Babina, A. M.; Lea, N. E.; Meyer, M. M., In Vivo Behavior of the Tandem Glycine Riboswitch in Bacillus subtilis. *mBio* **2017,** *8* (5), e01602-17.

109. Lipfert, J.; Das, R.; Chu, V. B.; Kudaravalli, M.; Boyd, N.; Herschlag, D.; Doniach, S., Structural transitions and thermodynamics of a glycine-dependent riboswitch from Vibrio cholerae. *Journal of molecular biology* **2007,** *365* (5), 1393-1406.

110. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D., Understanding nucleic acid-ion interactions. *Annual Reviews Biochemistry* **2014,** *83*, 813-841.

111. Liberman Joseph, A.; Suddala Krishna, C.; Aytenfisu, A.; Chan, D.; Belashov Ivan, A.; Salim, M.; Mathews David, H.; Spitale Robert, C.; Walter Nils, G.; Wedekind Joseph, E., Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proceedings of the National Academy of Sciences* **2015,** *112* (27), E3485-E3494.

112. Shevchuk, N. A.; Bryksin, A. V.; Nusinovich, Y. A.; Cabello, F. C.; Sutherland, M.; Ladisch, S., Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Research* **2004,** *32* (2), e19-e19.

113. Zhu, H.; Dean, R. A., A novel method for increasing the transformation efficiency of Escherichia coli—application for bacterial artificial chromosome library construction. *Nucleic Acids Research* **1999,** *27* (3), 910-911.

114. Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **2005**, *41*(1), 207-234.

115. Sreenath, H. K.; Bingman, C. A.; Buchan, B. W.; Seder, K. D.; Burns, B. T.; Geetha, H. V.; Jeon, W. B.; Vojtik, F. C.; Aceti, D. J.; Frederick, R. O.; Phillips, G. N.; Fox, B. G., Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expression and Purification* **2005,** *40* (2), 256-267.

# VITA

Born and raised in Valencia, Venezuela. Andrea Guedez received her Bachelor of Science degree in Chemistry in 2008 from The University of Carabobo in Naguanagua, Venezuela. Andrea served as Teaching Assistant of several undergraduate chemistry laboratories, as well as Instructor of Chemistry at the same university. In 2007, Andrea participated in an internship in Pinturas Flamuko, a paint formulation company and worked there as formulator after graduating. In 2009, she moved to the food industry and worked at GRUMA Venezuela for five years as external operations supervisor, in parallel with her teaching career at The University of Carabobo, where she taught undergraduate level chemistry courses as a part time instructor.

Andrea moved to Texas in the Fall of 2016, to pursue a doctoral degree in Biochemistry at Texas Christian University in Fort Worth, TX, working with Professor Youngha Ryu. In April 2018, she received the Dean's teaching award from the College of Science and Engineering, as well as the teaching award for an advanced laboratory course from the Department of Chemistry and Biochemistry. Three years later in 2021, she received an ACS BIOL Division Travel Award to attend to the Fall 2021 ACS National Meeting *and* Expo, as well as the People's Choice award in the 3MT competition and the College of Science and Engineering and SciComm Graduate Award at the Michael and Sally McCracken annual Student Research Symposium (SRS). Andrea obtained the Doctor of Philosophy degree in Chemistry in May 2022.

# ABSTRACT

DIRECTED EVOLUTION OF SYNTHETIC RIBOSWITCHES AND A LEUCYL tRNA SYNTHETASE
by

**Andrea Gabriela Guedez Pena**

Ph. D. in Chemistry, 2022

Advisor: Dr. Youngha Ryu, Associate Professor

An orthogonal pair composed by a leucyl-tRNA synthetase from *Methanobacterium thermoautotrophicum* (MtLRS) and tRNA from *Halobacterium sp. NRC-1 (Hhl4)* was used to incorporate UAAs in bacteria. Using a sequence homology analysis between LeuRS from bacteria and archaea, we determined the amino acid residues responsible for binding in the active site of the MtLRS. We generated three variants containing five mutations in the active site to incorporate 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid (Dansyl-Dap), and used these variants in the expression of Z-domain containing an amber stop codon (K7TAG).[15] The variants did not incorporate Dansyl-Dap at the TAG codon of the Z-domain gene. We designed a library of MtLRS variants randomizing the same positions in the active site. For dual genetic selection we constructed the pRCG plasmid, which contains a *cat-upp* fusion gene with TAG codons at permissible sites of the chloramphenicol acetyl transferase (CAT) gene. Three variations of the pRCG plasmid were tested for amber suppression using the wild type MtLRS lacking the editing domain: Q98TAG, D111TAG, and a double mutant containing both mutations. The Q98TAG showed the highest amber stop codon suppression efficiency for the selection experiment, so this variant was used in the genetic selection of the MtLRS library for the incorporation 4-nitro-1-phenylalanine and Dansyl-Dap. The obtained

variants are currently under study to evaluate their ability to incorporate these UAAs to the Z-domain protein.

Dual genetic selection was also used to select synthetic riboswitches. As a proof of concept, we randomized eleven bases at the binding site of the synthetic theophylline riboswitch and used the *cat-upp* fusion gene to select a riboswitch specific for caffeine. The selected variant (CaffRS) was tested for β-galactosidase activity, a colorimetric assay to study the concentration-signal dependency.[31] We then used this approach to select riboswitches for sarcosine, a prostate cancer biomarker. Two clones showed β-galactosidase activities proportional to the concentration of sarcosine (A6 and C9) and are currently tested for *in vitro* binding assays.